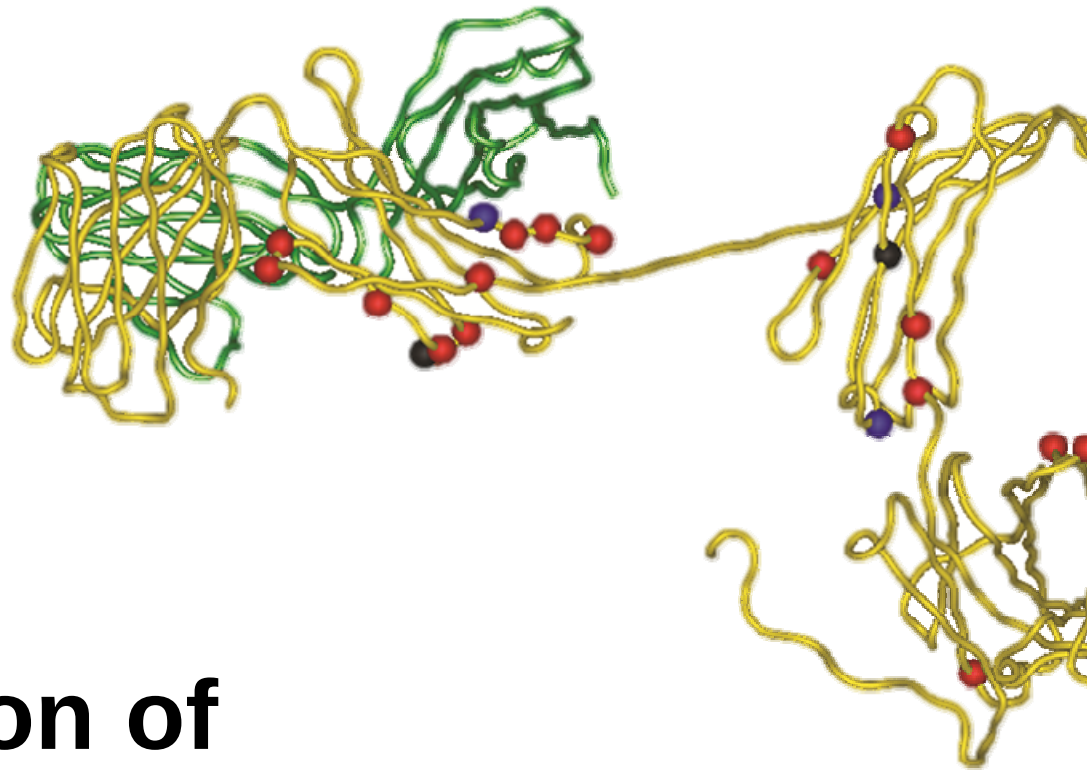




Evolution of immunoglobulins in Leporids: inferences of genetic resistance against pathogens

Ana Maria Pinheiro Rombert Pinhão

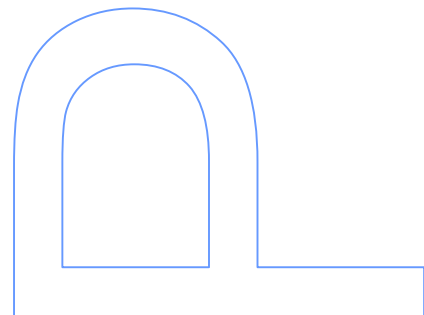
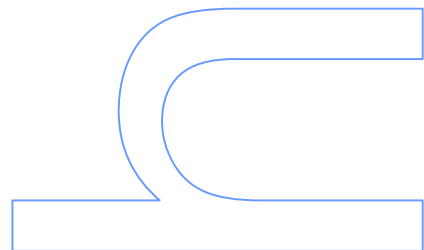
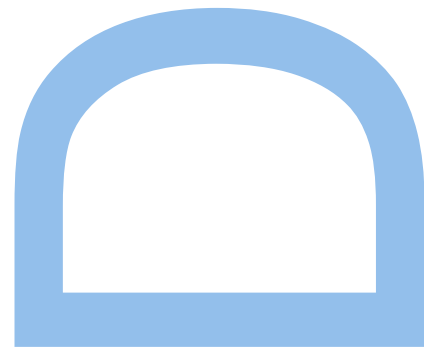
Doutoramento em Biodiversidade, Genética e Evolução
Departamento de Biologia
2016

Orientador

Paulo Célio Alves, Professor Associado, Faculdade de Ciências da Universidade do Porto

Coorientador

Christian Gortazar Schmidt, Profesor titular, SaBio (Sanidad y Biotecnología) research group, National Wildlife Research Institute IREC (CSIC-Universidad de Castilla – La Mancha)



Ao Periquito e à Pintassilga

Nota prévia

Na elaboração desta dissertação, e nos termos do número 2 do Artigo 4º do Regulamento Geral dos Terceiros Ciclos de Estudos da Universidade do Porto e do Artigo 31º do D.L. 74/2006, de 24 de Março, com a nova redação introduzida pelo D.L. 230/2009, de 14 de Setembro, foi efetuado o aproveitamento total de um conjunto coerente de trabalhos de investigação já publicados ou submetidos para publicação em revistas internacionais indexadas e com arbitragem científica, os quais integram alguns dos capítulos da presente tese. Tendo em conta que os referidos trabalhos foram realizados com a colaboração de outros autores, o candidato esclarece que, em todos eles, participou ativamente na sua conceção, na obtenção, análise e discussão de resultados, bem como na elaboração da sua forma publicada.

Este trabalho foi apoiado pela Fundação para a Ciência e Tecnologia (FCT) através da atribuição de bolsa de doutoramento (SFRH/BD/71252/2010) e de dois projectos de investigação (PTDC/BIA-ANM/3963/2012; PTDC/BIA-BEC/103158/2008) e co-financiado pelo Programa Operacional Potencial Humano (POPH), financiado pela União Europeia através do Fundo Social Europeu (ESF) e pelo Governo de Portugal.



UNIÃO EUROPEIA
Fundo Social Europeu

Agradecimentos/Acknowledgments

A todos que de uma forma ou outra contribuíram para que esta tese fosse possível o meu muito obrigada! Aos que, por acaso, me terei esquecido de mencionar desde já as minhas desculpas pela omissão.

Ao meu orientador, Prof. Dr. Paulo Célio Alves, agradeço ter aceite orientar esta tese, por todo o apoio científico, entusiasmo, colaboração, disponibilidade e amizade demonstrados ao longo do trabalho.

To my co-supervisor, Prof. Dr. Christian Gortázar, I thank for welcoming me in SaBio-IREC and for all scientific discussions, availability and kindness. Thank you also for having opened the possibility of working in your field of wildlife epidemiology and disease control, though it has not yet come to be I have learned a lot from your work and talks.

To Prof. Dr. José de la Fuente I would like to thank having welcomed me in his lab.

To all SaBio-IREC personel that made me feel at home. To Encarni, for all your help with burocracies, housing and your patience with my enquiries. To Isabel, for sharing with me your knowledge and all your help with lab work, without you some tasks would have been much harder to complete, and also, and very importantly, for making mine and my baby stay at Ciudad Real much more pleasant.

To Jenny Woof thank you for having received me in her lab in Dundee. Thank you so much for sharing with me your knowledge on immunology, for your guidance, support, kindness and patience. Thank you for discussing my results with me, your input greatly improved my work and this dissertation.

To Laurent Abi-Rached thank you for having welcomed me in Marseille, for all your teachings and help with natural selection analysis, and for discussing my results. Your input greatly improved my work and this dissertation.

To Rose Mage thank you for your immense and contagious enthusiasm regarding rabbit immunogenetics. Thank you for your support and for believing in my work. You have contributed greatly to my growth as a researcher and I have learned a lot from all the discussions and inputs from you.

To Martin Flajnick thank you for having discussed my results and work plans with me and also for having reviewed the Leporid IgG manuscript.

To Peter Parham, Dennis Lanning, Katherine L. Knight and Wessel van der Loo, I thank for discussing and improving my manuscripts.

Ao Zef a preciosa ajuda e explicações com as análises do BEAST.

À Joana Abrantes pelo interesse, discussões científicas e revisão de manuscritos.

Às Imídias, Joana, Helena, Fabi, Guida, Patrícia, Maria e Tereza, pela amizade, companheirismo, risota e momentos bem passados. À Helena que tão bem me recebeu e me reensinou a estar em laboratório passados 3 anos de ausência. À Tereza em especial muito obrigada pela grande ajuda com o trabalho laboratorial nas últimas fases deste doutoramento. À Teresa, Diana e Sandra Afonso por tão bem me receberem no início deste projecto e pela amizade, companheirismo e momentos de descontração e risota ao longo destes anos.

À Fundação para a Ciência e a Tecnologia (FCT), agradeço a concessão de uma Bolsa de Doutoramento (SFRH/BD/71252/2010).

Ao Centro de Investigação em Biodiversidade e Recursos Genéticos (CIBIO) e ao Centro de Testagem Molecular (CTM), agradeço a ajuda técnica proporcionada por muitos dos seus elementos e as condições laboratoriais proporcionadas. À Susana pelos conselhos técnicos sobre sequenciação, à Sara João pela disponibilidade e ajuda com burocracias, à Sandra Afonso toda a partilha e lições sobre técnicas de genética e biologia molecular. A todos pelos momentos de descontração, boa disposição, conversas várias

À Sara Ferreira, Sandra Rodrigues e Maria Sant'ana pela disponibilidade e apoio na resolução de burocracias várias.

A todos os que mais directamente contribuíram para o meu percurso académico e que me fez chegar até aqui. Ao Henrique Cabral que orientou o meu estágio e ao Prof. Dr. Paulo Alexandrino que orientou o meu mestrado, obrigada pelos ensinamentos. À Célia pela amizade, apoio, disponibilidade e pelo muito que me ensinaste na ciência e

na vida. Ao Rui Faria pelo apoio, disponibilidade e tudo o que me ensinaste quer no laboratório quer com as análises de dados.

Aos meus pais, à minha irmã e ao meu irmão muito obrigada por tudo, sem o vosso incentivo, apoio, carinho e paciência não teria chegado até aqui. Ao meu Pai por me ter transmitido o gosto pela bicharada e a curiosidade de querer saber mais. Mãe, pelo que sempre nos exigiste e me fez querer mais, aqui fica: mais tarde do que previsto, mas lá me doutorei!

Ao Pedro, sem ti esta tese não era possível. Muito obrigada pelo apoio incondicional, por me contagiares com o teu entusiasmo pela ciência e por acreditares em mim. Muito obrigada por estares ao meu lado há 10 anos (quase 11), por aturares as minhas idiossincrasias, pelo carinho, compreensão e amor, e mais que tudo pelo nosso projecto de vida em comum e a mais que dois.

Aos meus filhos, Dinis e Eva, tão pequenos e que tanto me exigem e ensinam: a ser mais gentil, a ver a maravilha das mais pequenas coisas, o valor de um sorriso e um abraço... o vosso amor é maravilhoso, preenche-me a mim e à minha vida!

À Mariana, quase minha filha, por teres aceite que faça parte da tua família.

Abstract

The European rabbit, *Oryctolagus cuniculus*, has long been an important model for immunological surveys and, thus, its immunoglobulin genes have been, to some extent, well studied. Regarding the immunoglobulins heavy chain genes, particularities have been revealed that set the European rabbit apart from all studied mammals. Among these are the preferential usage of one *VH* gene in *VDJ* rearrangements, the existence of trans-species polymorphism in the *IGHV* locus, the unusually large genetic distances between *VH* allelic lineages and the fact that the antibody repertoire is diversified in this species only after birth. Additionally, the European rabbit has a single IgG, contrarily to most mammals that have multiple IgG subclasses, and 13 IgA subclasses, being the most complex IgA system of all studied mammals. Despite all these surprising facts study of other Leporids *IGH* genes are limited. Since the 1950's the European rabbit populations have been decimated by two viral fatal diseases, Myxomatosis and Rabbit haemorrhagic disease, the second closely related to the European brown hare syndrome that has been causing population declines in hares populations, another leporid. Given this context, the study of leporids immune system genes is of utmost importance. In this thesis we aimed at extending knowledge of immunoglobulin genes, both variable and constant regions of the Heavy chain, to other Leporids

The study of rabbit *IGHV* genes has unveiled unique features and contributed to the current knowledge of Ig structure and function. Contrary to human and mice, the rabbit expresses preferentially one *VH* gene, the most *D*-proximal *VH1*, in the generation of its antibody repertoire. The *VH1* is used in 80-90% of *VDJ* rearrangements and in the remaining 10-20% are used the *VHn* genes, that map at least 100 Kb upstream of *VH1*. The maintenance of the *VHn* genes usage in low frequency in *VDJ* rearrangements has been suggested to represent a relic of an ancestral immunologic response to pathogens. To address this question, we sequenced *VDJ* genes for another Leporid, genus *Lepus*, which separated from European rabbit 12 million years ago. We found that *Lepus* also uses the *VHn* genes in 5-10% of its *VDJ* rearrangements. Our results show that the *VHn* genes are a conserved ancestral polymorphism that has been maintained in the Leporids genome and is being used for the generation of *VDJ* rearrangements by both modern *Lepus* and *Oryctolagus*. Furthermore the leporid *VH* genes allelic lineages have unusually large genetic distances and remain one of the few reported examples of trans-species polymorphism. The *VH* lineages seem to have been retained differently in the European rabbit and hares (*Lepus* sp.) and this led to the hypothesis that the *VH*

specificities could be associated with different environments. To address this hypothesis we sequenced *VDJ* genes from a third leporid genus, *Sylvilagus*. We found a fifth and equally divergent *VHa* lineage, the a5, and an ancient lineage, the sS, related to the hares sL, but failed to obtain *VHn* genes. These results show that the studied leporids employ different *VH* lineages in the generation of the antibody repertoire, suggesting that the leporid *VH* genes are subject to strong selective pressure likely imposed by specific pathogens.

Of the five mammalian immunoglobulin isotypes, IgA is likely the one in which production a higher investment is made, as IgA is the predominant immunoglobulin isotype in mucosal tissues and external secretions as well as being the second most prevalent immunoglobulin in serum. Considering that rabbit has the most complex IgA system observed with 13 *IGHA* genes and 11 expressed IgA subclasses which are differentially distributed among mucosal tissues, we can speculate that IgA is of high importance for the rabbit immune system. To achieve its function of eliminating pathogens IgA binds through its Fc to several receptors, of which Fc α RI, pIgR and Fc α / μ R are the most relevant. On the other hand, numerous pathogens trying to evade elimination by the immune system have evolved proteins targeting IgA, either IgA-binding proteins that block IgA coupling to IgA receptors or proteases that cleave the IgA hinge uncoupling the recognition of foreign antigens from the effector functions that eliminate them. Considering the complexity of its interactions with the surrounding environment, IgA is a likely target for natural selection. To investigate this possibility, the action of natural selection on IgA was examined in depth with six different Maximum likelihood methods. The analysis of mammalian sequences identified 18 positively selected sites: ten in C α 1, five in C α 2 and three in C α 3. All but one of these positions display variation in polarity and charge. Their structural locations suggest they indirectly influence the conformation of sites on IgA that are critical for interaction with host IgA receptors and also with proteins produced by mucosal pathogens that prevent their elimination by IgA-mediated effector mechanisms. That IgA residues subject to positive selection impact sites targeted both by host receptors and subversive pathogen ligands highlights the evolutionary arms race playing out between mammals and pathogens, and further emphasizes the importance of IgA in protection against mucosal pathogens.

The European rabbit IgG has been of significant importance in immunological research and is well characterized. The two loci identified by serology, the *d* and *e loci* that locate to the hinge and CH2 domains, respectively, were used to assess European rabbit wild populations and domestic breeds IgG genetic diversity. These two *loci* were also used to screen other Leporids IgG. With these two loci in view, nucleotide

sequencing studies of leporids IgG have targeted the hinge and CH2 domains for a restricted number of *Sylvilagus* and *Lepus* species, and complete *IGHG* nucleotide sequences are available for only a few European rabbits. Thus, there remains a void in the knowledge of the whole IgG genetic diversity for European rabbit wild populations and domestic breeds as well as of the evolution of this gene in Leporids. We sequenced IgG from genomic DNA of 30 wild European rabbits, including both subspecies (*O. c. algirus* and *O. c. cuniculus*) from three geographic regions, and 15 domestic *O. c. cuniculus*. Genetic diversity was highest within Iberian wild populations. Only two new amino acid polymorphisms were found, both in *O. c. algirus*. To evaluate the evolution of this gene in leporids, we sequenced the complete *IGHG* for six other genera: *Bunolagus*, *Brachylagus*, *Lepus*, *Pentalagus*, *Romerolagus* and *Sylvilagus*. A gradient in leporid IgG constant domain diversity was observed with the CH1 being the most conserved and the CH3 the most variable domain. Positive selection was found to be acting on all constant domains but with a greater incidence in the CH3 domain, where a cluster of three positively selected sites was identified. In the hinge region, only three polymorphic positions were observed. The same hinge length was observed for all leporids. Unlike the variation observed for the European rabbit, all 11 *Lepus* species studied share exactly the same hinge motif, suggesting its maintenance as a result of an advantageous structure or conformation.

Keywords

Leporids, European rabbit, *Lepus*, Adaptive immunity, Immunoglobulins, *IGHV*, IgA, IgG, Evolution, Co-evolution, Genetic diversity patterns, Positive selection.

Resumo

O coelho-bravo, *Oryctolagus cuniculus*, tem sido um importante modelo para a investigação na área da imunologia e, como tal, os genes que sintetizam as imunoglobulinas têm sido bastante estudados. Nos genes envolvidos na produção da cadeia pesada das imunoglobulinas foram descobertas várias características que distinguem o coelho dos restantes mamíferos. Entre elas podemos destacar o uso preferencial de apenas um gene *VH* nos rearranjos dos genes *VDJ*, a existência de polimorfismos trans-específicos no *locus IGHV*, as distâncias genéticas elevadas entre linhagens alélicas dos genes *VH* e ainda o facto de a diversificação dos anticorpos ocorrer apenas nas semanas após o nascimento. Para além destas características, o coelho-bravo apresenta ainda outras peculiaridades como ter uma só IgG, quando a maioria dos mamíferos tem várias subclasses de IgG, e 13 subclasses de IgA, tendo o mais complexo sistema de IgA de todos os mamíferos estudados. Apesar de se conhecerem estas particularidades nas imunoglobulinas do coelho-bravo, os genes *IGH* de outros Leporídeos têm sido pouco estudados. Desde a década de 1950 as populações de coelho-bravo têm sido dizimadas por duas doenças virais que apresentam mortalidades muito elevadas, a Mixomatose e a Doença Hemorrágica do Coelho. O vírus que causa a Doença Hemorrágica do Coelho pertence à mesma família do vírus que causa a Síndrome da lebre-comum, doença que tem causado declínios em populações de lebres (*Lepus* sp.), outro Leporídeo. Neste contexto, o estudo dos genes do sistema imunitário dos Leporídeos assume particular importância. Assim, o objectivo desta tese é desenvolver o conhecimento dos genes das imunoglobulinas, das regiões variáveis e constante da cadeia pesada, para outros Leporídeos.

O estudo de genes *IGHV* do coelho contribuiu para o conhecimento atual da estrutura e função das imunoglobulinas. Ao contrário do observado para o Homem e rato, o coelho expressa preferencialmente um gene *VH*, o *VH1*, na geração do seu repertório de anticorpos. O *VH1* é utilizado em 80 a 90% dos rearranjos de genes *VDJ*; nos restantes 10-20% são utilizados os genes *VHn*, que se localizam pelo menos 100 kb a montante do *VH1*. Foi sugerido que a conservação do uso de genes *VHn* em baixa frequência nos rearranjos de genes *VDJ* pode representar uma relíquia de uma resposta imunológica ancestral a agentes patogénicos. Para responder a esta questão, sequenciámos genes *VDJ* para outro leporídeo, género *Lepus*, que se separou do coelho-bravo há 12 milhões de anos, e descobrimos que o género *Lepus* também utiliza os genes *VHn* em 5-10% dos rearranjos de genes *VDJ*. Os resultados obtidos mostram que os genes *VHn* são um polimorfismo ancestral conservado que foi

mantido no genoma dos Leporídeos e está a ser utilizado para a geração de rearranjos de genes *VDJ* pelas espécies actuais dos géneros *Lepus* e *Oryctolagus*. Para além disto, os genes *VH* dos Leporídeos apresentam linhagens alélicas que têm distâncias genéticas invulgarmente grandes e constituem um dos poucos exemplos relatados de polimorfismo trans-específico. As linhagens *VH* parecem ter sido mantidas de forma diferente no coelho-bravo e nas lebres o que levou à hipótese de que as especificidades das diferentes linhagens de *VH* podem estar associadas a diferentes ambientes. Como abordagem a esta hipótese, sequenciámos genes *VDJ* de um terceiro leporídeo nativo do continente americano, do género *Sylvilagus*. Encontrámos uma quinta linhagem *VHa*, a $\alpha 5$, e uma linhagem antiga, a *sS*, relacionada com a linhagem *sL* das lebres, mas não encontramos genes *VHn*. Estes resultados mostram que os leporídeos estudados usam diferentes linhagens *VH* na geração do repertório de anticorpos, o que sugere que os genes *VH* dos leporídeos estão sujeitos a uma forte pressão selectiva provavelmente imposta por agentes patogénicos específicos.

A IgA é, provavelmente, a classe de imunoglobulina na produção da qual os organismos biológicos fazem um maior investimento, uma vez que a IgA é a classe predominante nas mucosas e secreções externas, bem como é a segunda classe mais prevalente no plasma. Considerando que o coelho-bravo tem o mais complexo sistema de IgA observado com 13 genes *IGHA* e 11 subclasses de IgA expressas que apresentam uma distribuição diferencial pelos diferentes tecidos mucosos, podemos especular que IgA é de grande importância para o sistema imunológico do coelho. Para cumprir a sua função de eliminar agentes patogénicos, a IgA liga-se, através do seu Fc, a vários receptores, dos quais o $Fc\alpha RI$, $pIgR$ e $Fc\alpha/\mu R$ são os mais relevantes. Por outro lado, vários agentes patogénicos que tentam evitar a eliminação pelo sistema imunitário desenvolveram estratégias dirigidas à IgA, tanto proteínas que ligando-se ao IgA-Fc bloqueiam o seu acoplamento a receptores ou proteases que cortam a *hinge* da IgA separando o reconhecimento de antígenos estranhos das funções efectoras que os eliminam. Assim, considerando a complexidade das suas interações com o meio envolvente, a IgA é um alvo provável para a acção da selecção natural. Para investigar esta possibilidade, a acção da selecção natural na evolução da IgA foi examinada usando seis métodos de máxima verossimilhança. A análise de sequências de mamíferos identificou 18 codões sob selecção positiva: dez no $C\alpha 1$, cinco no $C\alpha 2$ e três no $C\alpha 3$. Todos estes codões, com excepção de um, exibem variação de polaridade e carga. A sua localização na estrutura da IgA sugere que podem influenciar indirectamente a conformação de zonas da IgA que são críticas para a interacção com os receptores de IgA e também com as proteínas produzidas por agentes patogénicos para impedir a sua eliminação. O facto de os resíduos da IgA

que estão sob selecção positiva terem impacto sobre regiões importantes tanto para a interacção com receptores do hospedeiro e ligandos dos agentes patogénicos dá ênfase à co-evolução protagonizada por mamíferos e patógenos, e realça ainda mais a importância da IgA na protecção contra agentes patogénicos da mucosa.

A IgG do coelho-bravo tem tido uma importância significativa na investigação imunológica e está bem caracterizada. Os dois *loci* identificados por serologia, os *loci d* e *e* que se localizam na *hinge* e domínio CH2, respectivamente, foram utilizados para avaliar a diversidade genética de populações selvagens e raças domésticas de coelho-bravo. Estes dois *loci* foram também utilizados para estudar a IgG de outros Leporídeos. Os estudos de sequenciação nucleotídica da IgG dos Leporídeos têm tido como foco apenas a *hinge* e domínio CH2 e para um número restrito de espécies dos géneros *Sylvilagus* e *Lepus*, e, assim, só estão disponíveis sequências completas de *IGHG* para um reduzido número de coelhos-bravos. Como tal, persiste ainda hoje uma lacuna no que respeita ao conhecimento da diversidade genética da IgG de coelho-bravo, tanto em populações selvagens como domésticas, bem como da evolução deste gene noutros Leporídeos. Neste trabalho, sequenciámos a IgG a partir do DNA genómico de 30 coelhos-bravos selvagens das subespécies *O. c. algirus* e *O. c. cuniculus* e 15 animais domésticos da subespécie *O. c. cuniculus*. A diversidade genética observada foi maior nas populações selvagens ibéricas. Encontrámos dois novos polimorfismos de aminoácidos na subespécie *O. c. algirus*. Para avaliar a evolução deste gene nos Leporídeos, sequenciámos o *IGHG* completo para seis outros géneros: *Bunolagus*, *Brachylagus*, *Lepus*, *Pentalagus*, *Romerolagus* and *Sylvilagus*. Identificámos um gradiente na diversidade dos domínios constantes da IgG dos Leporídeos, sendo o CH1 o domínio mais conservado e o CH3 o domínio mais variável. Encontrámos, ainda, evidências da acção de selecção positiva em todos os domínios constantes, mas com maior incidência no domínio CH3, onde identificámos um conjunto de três codões sob selecção positiva. Na região da *hinge* foram observadas apenas três posições polimórficas e para todos os Leporídeos se observou o mesmo comprimento desta região. Ao contrário da variação observada para o coelho-bravo, as 11 espécies estudadas do género *Lepus* partilham o mesmo motivo na *hinge*, o que sugere a sua manutenção como resultado de uma estrutura ou conformação vantajosa.

Palavras chave

Leporídeos, Coelho-bravo, *Lepus*, Imunidade adaptativa, Imunoglobulinas, *IGHV*, IgA, IgG, Evolução, Co-evolução, Padrões de diversidade genética, Selecção positiva.

Table of contents

AGRADECIMENTOS/ACKNOWLEDGMENTS	III
ABSTRACT	VII
RESUMO	XI
TABLE OF CONTENTS	XV
LIST OF TABLES	XVII
LIST OF FIGURES	XIX
 CHAPTER 1: GENERAL INTRODUCTION	1
1. The Immune system: innate and adaptive immunity	3
1.1 Immunoglobulins	4
1.1.1. Immunoglobulin isotypes	4
1.1.2 Heavy chain Variable region	5
1.1.3 Light chain	7
2. Lagomorph evolution	7
3. Leporid Immunoglobulin Heavy chain (IGH) locus	9
3.1. The Immunoglobulin Heavy chain Constant region (IGHC) <i>locus</i>	9
3.2. The Immunoglobulin Heavy chain Variable region (IGHV) locus	12
4. Objectives and thesis organization	13
5. References	16
 CHAPTER 2: VH EVOLUTION IN LEPORIDS.....	23
PAPER 1	
MOLECULAR BASES OF GENETIC DIVERSITY AND EVOLUTION OF THE IMMUNOGLOBULIN	
HEAVY CHAIN VARIABLE REGION (<i>IGHV</i>) GENE LOCUS IN LEPORIDS.	25
PAPER 2	
SEQUENCING OF MODERN <i>LEPUS</i> VDJ GENES SHOWS THAT THE USAGE OF <i>VHN</i> GENES	
HAS BEEN RETAINED IN BOTH <i>ORYCTOLAGUS</i> AND <i>LEPUS</i> THAT DIVERGED 12 MILLION	
YEARS AGO.	51
Supplementary Material	67
PAPER 3	
SEQUENCING OF <i>SYLVILAGUS</i> VDJ GENES REVEALS A NEW <i>VH_A</i> ALLELIC LINEAGE AND	
SHOWS THAT ANCIENT <i>VH</i> LINEAGES WERE RETAINED DIFFERENTLY IN LEPORIDS. ...	69
 CHAPTER 3: IGA EVOLUTION IN MAMMALS.....	85
PAPER 4	
COMPUTATIONAL ANALYSES OF AN EVOLUTIONARY ARMS RACE BETWEEN MAMMALIAN	
IMMUNITY MEDIATED BY IMMUNOGLOBULIN A AND ITS SUBVERSION BY BACTERIAL	
PATHOGENS	87

CHAPTER 4: IGG EVOLUTION IN LEPORIDS	109
PAPER 5	
LEPORID IGG SHOWS EVIDENCE OF STRONG SELECTIVE PRESSURE ON THE HINGE AND CH3 DOMAINS	111
PAPER 6	
SURVEY OF GENETIC DIVERSITY FOR IGG IN WILD AND DOMESTIC RABBITS.....	135
 CHAPTER 5: GENERAL DISCUSSION	 145
1. Diversifying selection acting on the Ig-Fc fragment.....	147
2. Insights in the <i>IGHV</i> genes evolution	148
3. Major conclusions.....	150
4. Future Perspectives	151
5. References	153

List of Tables

CHAPTER 2: VH EVOLUTION IN LEPORIDS

PAPER 1

Table 1- Gene frequencies obtained in populations of <i>Lepus granatensis</i> and <i>L. europaeus</i> at IGHV.....	35
---	----

PAPER 2

Table 1- Vh genes sequences accession numbers for sequences used in the phylogenetic analysis.	55
Table 2- Genetic distances between hare and rabbit VH nucleotide sequences.....	58
Supp. table 1- Alignment of European rabbit and hare VH protein sequences...	67

PAPER 3

Table 1- Vh genes sequences accession numbers for sequences used in the phylogenetic analysis.	73
Table 2- Genetic distances between leporids VH nucleotide sequences.....	76

CHAPTER 3: IGA EVOLUTION IN MAMMALS

PAPER 4

Table 1 - Phylogenetic tests of positive selection in primates	95
Table 2 - Characterization of natural amino acid variation for each residue identified under positive selection in primate IgA.....	96
Table 3 - Phylogenetic tests of positive selection in mammals.....	98
Table 4 - Characterization of natural amino acid variation for each residue identified under positive selection in mammal IgA.	99

CHAPTER 4: IGG EVOLUTION IN LEPORIDS

PAPER 5

Table 1- IGHG sequences accession numbers for sequences used in this study.	117
---	-----

PAPER 6

Table 1- Immunoglobulin gamma nucleotide sequence polymorphism and estimates of haplotype diversity, haplotype diversity variance and haplotype standard deviation in the European rabbit populations/domestic animals studied.	138
---	-----

Table 2- Amino acid polymorphic positions for European rabbit IgG populations/domestic animals.	140
--	-----

List of Figures

CHAPTER 1: GENERAL INTRODUCTION

Fig. 1 – Genomic organization of immunoglobulin genes.	6
Fig. 2 - Evolutionary topology reflecting the relationships between the Lagomorpha group based in a molecular super matrix.	8
Fig. 3 – European rabbit IGHG locus from VH haplotype (a2 F-I) for which nucleotide sequence is available from sequenced, assembled BAC clones.	10
Fig. 4 – The European rabbit IGHA region.	12

CHAPTER 2: VH EVOLUTION IN LEPORIDS

PAPER 1

Fig. 1 - Evolutionary topology reflecting the relationships within the Lagomorpha group based on a molecular super matrix (adapted from Matthee et al. 2004)...	29
Fig. 2 - Rabbit IGHV (regions from VH haplotype (a2 F-I)) for which nucleotide sequence is available from sequenced, assembled BAC clones.	32
Fig. 3 - Amino acid sequences of european rabbit (<i>O. cuniculus</i>) proteins (encoded by VH1a1-a3, VH4.1-4.2, VHx, VHy, VH _z) and hare (VHa2L, VHsLe, VHsLg) VH genes.	33
Fig. 4 - Geographic distribution of <i>VHa</i> locus allotype frequencies, serologically determined for populations of European wild rabbit (<i>O. cuniculus</i>) from the Iberian Peninsula and France.	34
Fig. 5 - Neighbour-joining tree for <i>Oryctolagus cuniculus</i> , <i>Lepus europaeus</i> (Le), and <i>Lepus granatensis</i> (Lg) VH genes, showing the transpecies polymorphism of VHa2.	37

PAPER 2

Fig. 1. Phylogenetic tree of European rabbit and hare VH genes.	57
Fig. 2. Alignment of European rabbit and hare VH protein sequences.	58

PAPER 3

Fig. 1. Phylogenetic tree of Leporid VH genes.	75
Fig. 2. Alignment of leporid VH protein sequences.	79

CHAPTER 3: IGA EVOLUTION IN MAMMALS

PAPER 4

Figure 1. Residues encoded by PSC in human IgA1 C α 1, C α 2 and C α 3 domains.	97
--	----

Figure 2. Residues encoded by PSC in the three-dimensional structure of human IgA1 and their relationship to sites of biological interest.....	100
---	-----

CHAPTER 4: IGG EVOLUTION IN LEPORIDS

PAPER 5

Figure 1. European rabbit (<i>Oryctolagus cuniculus</i>) IgG DNA sequence of constant and hinge regions (Genbank accession number: L29172).	114
Figure 2. Variable codons and amino acid physicochemical properties for leporid IGHG CH1 and CH2 domains and hinge region.....	120
Figure 3. Leporids residues encoded by positively selected codons in the three-dimensional structure of rabbit IgG.	123

PAPER 6

Figure 1- Map with the geographical locations of the populations used in this study.....	138
---	-----

Chapter 1

General Introduction

1. The Immune system: innate and adaptive immunity

The Immune system is a complex arrangement of mechanisms designed to prevent invasion and damage to the organism caused by pathogens. Traditionally it is divided into innate and adaptive immunity, where the first is the genetically pre-selected immune system providing the first line of defense and the second has a randomly generated antigen–receptor diversity and is able to remember and improve at each encounter with the same pathogens. Components of the innate immune system, such as the pattern recognition receptors (PRRs, that include the toll like receptors and scavenger receptors) and the complement system, are found throughout the animal kingdom (Flajnik and Du Pasquier 2004). Adaptive immunoglobulins (Igs), T-cell receptors (TCRs) and Major histocompatibility complex (MHC) molecules are found only in jawed vertebrates (Flajnik and Kasahara 2010). In fact, until ten years ago the adaptive immune system was thought to exist only in jawed vertebrates but the discovery of a convergent adaptive immune system based on variable lymphocyte receptors (VLRs) in lampreys, a jawless vertebrate, (Pancer et al. 2004), and of somatic diversification in mollusk immune genes (Zhang et al. 2004) have come to question what is adaptive or innate and the origins of adaptive immunity (Flajnik 2014; Flajnik and Du Pasquier 2004; Flajnik and Kasahara 2010).

The immune response involves always two steps: recognizing the foreign antigens and mounting a response to eliminate the invader. The vertebrate adaptive immune system uses two classes of molecules in the first step: the TCRs, which are present in mature T lymphocytes membrane, and Igs, which are displayed and secreted by B lymphocytes. Then, Igs contribute to pathogen elimination either by direct neutralization or by eliciting a variety of effector mechanisms, for example antibody-dependent cellular cytotoxicity, antibody-dependent cell phagocytosis, and complement-dependent cytotoxicity, whereas TCRs recruit a variety of cell mediated immune responses, such as destruction of infected/dysfunctional somatic cells by cytotoxic T lymphocytes, increasing the B cell antibody response by helper T cells and recruitment of macrophages to the site of infection (Roitt et al. 2001).

To be able to circumvent and evade these immune mechanisms pathogens are continuously acquiring new mutations exerting selective pressure on the host that must also acquire new mutations to be able to recognize the ‘new’ pathogen phenotype. This never-ending arms race drives the host-pathogen co-evolution in that the host immune system continuously evolves to escape the pathogen and the

pathogen continuously evolves to infect the host. This process illustrates the evolutionary law proposed by Van Valen (1973), the 'Red Queen hypothesis'.

1.1 Immunoglobulins

Immunoglobulins (Igs) antibodies (Ab) or B cell receptors (BCR) are, perhaps, one of the most emblematic components of the jawed vertebrate immune system. Typically Y-shaped, Igs are glycoproteins that have a basic structure of two light and two heavy chains, each having a variable (V) and a constant (C) region, linked together by disulphide bridges. Each chain is organized in globular domains consisting of approximately 110-130 amino acids. The light chains (VL and CL domains) and the variable (VH) and first constant domain of the heavy chain (CH1) constitute the two Fab regions that bind antigens. The remaining constant domains of the heavy chain (CH) constitute the Fc region, which elicits mechanisms that lead to pathogen elimination, the so called effector functions. Differences in the CH chain define Ig isotypes and, thus, each isotype has specific effector functions. Eutherian mammals have five Ig isotypes, IgM, IgD, IgE, IgG and IgA, and a few other isotypes have been described in gnathostomes. The CH chains that define the five mammalian isotypes are designated by μ , δ , ϵ , γ and α , respectively.

1.1.1. Immunoglobulin isotypes

Immunoglobulin M (IgM), the most ancient Ig isotype, is structurally conserved and has the same function in all gnathostomes (reviewed in (Flajnik and Kasahara 2010)). IgM, in its transmembrane form, is the first isotype to be expressed during B-cell development. As such the VH and VL regions associated with this Ig have not undergone much antigenic driven diversification being able to recognize a wider range of antigens than other isotypes (Schroeder and Cavacini 2010). After B cell maturation and antigenic stimulation IgM is secreted in multimeric form. However, in cartilaginous fish IgM is secreted abundantly as a monomer (reviewed in (Flajnik and Kasahara 2010)).

Immunoglobulin D (IgD), like IgM, is an ancient isotype but its function remains elusive. Until late XX century IgD was believed to be present only in primates and rodents (Zhao et al. 2002) but the discovery of an ortholog in bony fish (Wilson et al.

1997) and amphibians (Zhao et al. 2006) proved otherwise. Furthermore, there seems to be a phylogenetic relationship between IgD and IgW, an isotype found in cartilaginous fish and lungfish (Ota et al. 2003). Thus, even though an IgD/IgW precursor was present in the ancestor of all gnathostomes this Ig has been subject to large structure variations and evolutionary losses (reviewed in Flajnik and Kasahara 2010).

Immunoglobulin G (IgG) is the predominant serum immunoglobulin and also has the longest serum half-life. IgG participates directly to immune responses, including neutralization of toxins and viruses, and also indirectly through its effector functions, prompted through neutralization, binding to Fcγ receptors and activation of the complement pathway. Differences in the CH domains have led to the identification of IgG subclasses, each subclass differing in effector functions, which in mammals have been found to range from 1-7 (Schroeder and Cavacini 2010).

Immunoglobulin E (IgE) is a very potent immunoglobulin associated with hypersensitivity, allergic reactions and response to parasitic worm infections (Schroeder and Cavacini 2010). Both IgG and IgE have evolved from IgY, an isotype found on amphibians, reptiles and birds. IgY is the functional equivalent to IgG and IgE in these *taxa* and its structure is similar to that of IgE (Warr et al. 1995).

Immunoglobulin A (IgA), the mucosal Ig, is secreted in large amounts in mucosal tissues and external secretions where it plays a major role as a first line of defense against pathogens. This isotype is found in mammals, birds and reptiles and seems to have the same function in all *taxa* (reviewed in Flajnik and Kasahara 2010). In amphibians mucosal immunity is provided by IgX an isotype that seems to have a different origin from IgA (Warr et al. 1995).

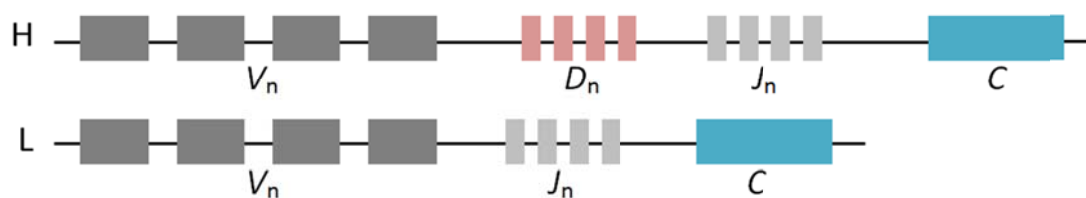
Other isotypes include the Platypus IgO (Zhao et al. 2009), bony fish IgT and IgZ (Danilova et al. 2005; Hansen et al. 2005) and cartilaginous fish IgNAR (New Antigen Receptor), a homodimer of two identical heavy chains that does not associate with light chains (Roux et al. 1998). Curiously, camelids have a modified form of IgG that also does not associate with light chains (Nguyen et al. 2000).

1.1.2 Heavy chain Variable region

The Ig VH region results of the assembly of three different genes: V (Variable), D (Diversity) and J (Joining). Multiple copies of each of these genes exist in the vertebrate's germline. In most vertebrates these genes are arranged in a translocon

configuration, where multiple *V* occur upstream of multiple *D* and multiple *J* followed by the CH region genes. The exception are cartilaginous fishes in which the Ig genes are arranged in clusters, each one having unique *V*, *D*, *J* and CH genes (Figure 1). In the generation of the primary antibody repertoire vertebrates have developed two differing strategies. Some, as human, mice and amphibians, rely on the combinatorial rearrangement of a large number of *V*, *D*, and *J* genes. Others, most mammals and birds, diversify rearranged *VDJ* genes by somatic hypermutation and/or gene conversion in gut associated lymphoid tissue (GALT). These species possess or use a limited number of germline *V* segments and use the somatic hypermutation and gene conversion mechanisms to overcome this limitation. Among these, one of the best studied models is the European rabbit (Flajnik 2002).

a) Translocon configuration



b) Cluster configuration

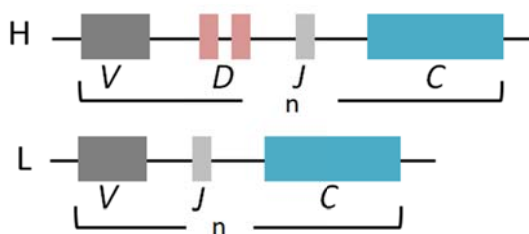


Fig. 1 – Genomic organization of immunoglobulin genes. a) Mammals, birds, reptiles and amphibians have all immunoglobulin genes in the so-called translocon configuration, whereas b) cartilaginous fish have all immunoglobulin genes in the cluster configuration. H- heavy chain, L – light chain, V – variable region genes, D- diversity region genes, J- joining region genes, C- constant region genes. (adapted from (Flajnik 2002)).

The VH domain, and also the VL domain, is divided into sections according to sequence variability. There are three hypervariable regions that interact with the antigen and are, thus, termed complementarity-determining regions (CDRs). These are situated between four regions of relatively invariable sequence that are termed framework regions (FRs). The combination of the CDRs of both VH and VL domains create the antigen-binding sites.

1.1.3 Light chain

In mammals the Light chain (L) is encoded by two isotypes, kappa (κ) and lambda (λ), which are functionally indistinguishable. A third L isotype, sigma (σ), has been identified in frogs (Schwager et al. 1991). All three isotypes have been shown to have arisen in the ancestor of living vertebrates but have been retained differently (reviewed in Flajnik and Kasahara 2010).

The L is encoded by three genes: *VL* (variable), *JL* (joining) and *CL* (constant region), and each isotype has its set of *VL*, *JL* and *CL* elements. In the human and mouse one C κ gene and multiple copies of C λ , each one associated with its own J λ , exist. In the rabbit the κ locus has suffered a duplication giving rise to two L κ , *IG κ 1* and *IG κ 2*, of which the *IG κ 2* is poorly expressed (reviewed in Mage et al. 2006; Pinheiro et al. 2015).

2. Lagomorph evolution

In the early 1900's Gridley (1912) proposed that the duplicidentata group, which included rabbits, hares and pikas, be raised to the order Lagomorpha, thus separating the lagomorphs from the rodents. The lagomorphs are thought to have had their origin 62-100 million years ago (Benton and Donoghue 2007) in Asia (Asher et al. 2005; Rose et al. 2008) and are divided into two families, Ochotonidae (pikas) and Leporidae (hares and rabbits). The divergence time between these two families is not consensual, being placed at 30-40 Million years ago according to the fossil data (Dawson, 1981) and at 50-55 Million years ago by molecular data (Springer et al. 2003).

The family Ochotonidae originated in Asia, with diversity and distribution range peaking between 23.03-5.33 million years ago in the Miocene, when their range expanded to Europe, South Africa and North America (Dawson, 1967; Ge et al. 2013). Nowadays the Ochotonids include one single genus, *Ochotona*, with 25 extant species subdivided into four subgenera: Pika, Logotona, Conothoa and Ochotona (Lanier and Olson 2009; Lissovsky 2014; Melo-Ferreira et al. 2015), and are present in Asia and North America.

The family Leporidae, according to the fossil data, originated in North America and Asia around 20 Million years ago (Dawson 1981; Lopez-Martinez 1989). Later, at

around 8 Million years ago, an apparently rapid diversification and range expansion to Europe and South Africa occurred (Flynn et al 2013). Molecular data suggests that this rapid diversification, the great Leporid radiation, is older and happened at 14 Million years ago (Matthee et al 2004). Leporids are commonly divided into two groups, hares and rabbits. Hares are represented by a single genus, *Lepus*, comprising 24 to 30 currently recognized species with a worldwide distribution (Alves et al. 2008; Corbet and Hill 1980; Flux and Angermann 1990). The rabbit group includes 10 genera (*Brachylagus*, *Bunolagus*, *Caprolagus*, *Nesolagus*, *Oryctolagus*, *Pentalagus*, *Poelagus*, *Pronolagus*, *Romerolagus* and *Sylvilagus*) most of which are monotypic. *Nesolagus*, *Pronolagus* and *Sylvilagus*, which comprise two, four and more than 16 species, respectively, are the exceptions (Can et al. 2001; Chapman et al. 1992; Frey et al. 1997; Matthee and Robinson 1996; Surridge et al. 1999; Whiteford, 1995). Though most rabbit genera have restricted geographical distributions, when taken together all 10 rabbit genera have a worldwide distribution. Leporid taxonomy remains unclear due to discrepancies between fossil and molecular data. A supermatrix based on nuclear and mitochondrial markers suggests that a group that encompasses the genera *Nesolagus*, *Poelagus* and *Pronolagus* was the first to diverge from the remaining genera at 14 million years ago (Matthee et al. 2004; Figure 2). Adding morphological and geographical data as well as unique insertions/deletions information to this supermatrix did not alter the obtained topology (Robinson and Matthee, 2005).

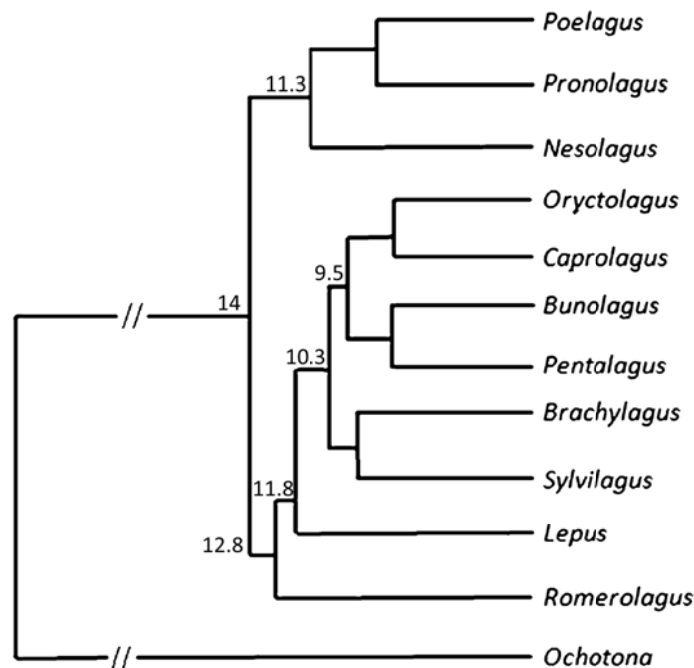


Fig. 2 - Evolutionary topology reflecting the relationships between the Lagomorpha group based in a molecular super matrix. Divergence times are indicated in million years at the nodes of the branches (adapted from Matthee et al., 2004).

3. Leporid Immunoglobulin Heavy chain (IGH) locus

The European rabbit (*Oryctolagus cuniculus*) whole genome sequence was completed in 2009 (OryCun2.0; Accession AAGW02000000) and its annotated assembly is available at NCBI, UCSC and Ensembl. Despite the huge amount of information only a portion of the genes of the Immunoglobulin Heavy Chain *locus* (IGH) are found in unplaced scaffolds in this whole genome sequence. Thus, the complete map of the rabbit IGH *locus* is still missing and the existing knowledge on this *locus* comes from more targeted analysis. The most comprehensive map of rabbit IGH *locus* was obtained from BAC library analysis and encompasses 34 V elements, 11 D elements, six J elements and the C μ , C γ , C ϵ and four C α genes (Figure 3) (Ros et al. 2004). The number of VH genes present in the rabbit germline is estimated to be over 200 (Currier et al. 1988; Gallarda et al. 1985; Ros et al. 2004) and 13 C α genes are known (Burnett et al. 1989). Furthermore, it was not until recently that the chromosomal location IGH *locus* was mapped to the q-telomere of chromosome 20 (Gertz et al. 2013).

3.1. The Immunoglobulin Heavy chain Constant region (IGHC) *locus*

The European rabbit IGH *locus* has some unique features. As some other mammals the European rabbit has no IgD and no *IGHD* has been found in the rabbit germline (Lanning et al. 2003; Ros et al. 2004). IgM and IgE are encoded by single genes, *IGHM* and *IGHE*, as well as IgG, a surprising fact among mammals that typically have multiple *IGHG* copies encoding multiple IgG subclasses (Ros et al. 2004; Gertz et al. 2013). For IgA, on the other hand, the European rabbit has 13 *IGHA* genes encoding 13 IgA subclasses (Burnett et al. 1989) being the most complex IgA system of all studied mammals.

Despite its importance as a model for immunological surveys, the study of European rabbit immunoglobulins has focused on the IgG disregarding other Ig isotypes. In fact, the information available for European rabbit IgM and IgE is extremely scarce with only a few sequences available in public databases. On the other hand, the European rabbit, and to some extent other leporids, IgG has been well studied. Serologic studies described two *loci* for the European rabbit IgG each with two segregating alleles: the *locus d* with *d11* and *d12* allotypes located to position 9 in the hinge region and the *locus e* with *e14* and *e15* allotypes located to position 92 of the *IGHG* CH2 (Hamers and Hamers-Casterman 1965; IMGT numbering,

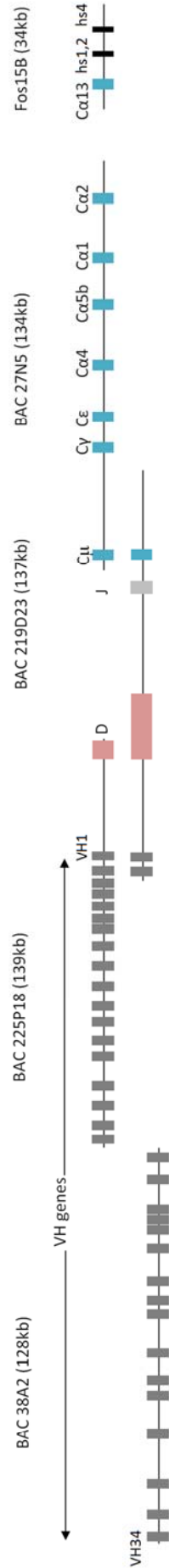


Fig. 3 – European rabbit IGHG locus from VH haplotype (a2 F-I) for which nucleotide sequence is available from sequenced, assembled BAC clones. Depicted are V genes, D and J regions and also C_μ, C_γ, C_ε and 5 C_α genes, as well as two 3'α enhancer regions, hs1,2 and hs4. The D-proximal VH gene, VH1, used in the majority of VDJ rearrangements is indicated. Identified above are the non-overlapping BAC clones 38A2 and 225P18 (GenBank Accession Numbers AY386694 and AY386697), and 219D23 and 27N5 (GenBank Accession Number AY386695 and AY386696). The 3' end of BAC 225P18 overlaps the 5' end of BAC 219D23 and the 3' end of BAC 219D23 overlaps the 5' end of BAC 27N5. The fosmid clone Fos 15B (GenBank Accession Number AY386698), that does not overlap with any of the BAC clones, is also identified above (adapted from Ros et al. 2004).

Lefranc et al. 2005). These two *loci* were used to assess the IgG genetic diversity in wild European rabbit populations and domestic breeds (Hamers and Hamers-Casterman 1965; Dubiski 1969; Mage 1981; Esteves et al. 2002, 2006). Serologic studies further looked into other Lagomorphs IgG genetic diversity and found the *e15* allotype in species of the genera *Lepus*, *Sylvilagus*, *Romerolagus*, and *Ochotona* (van der Loo and Hamers-Casterman 1979; Cazenave et al. 1987;) but could not identify the *d11* and *d12* allotypes in *Lepus* and *Sylvilagus* genera (Hamers-Casterman et al. 1979). At the same time, the relationship between serology and protein variation at the *e locus* was studied by amino acid sequencing of tryptic peptides in various Lagomorph species (Appella et al. 1971; Aggarwal and Mandy 1976; Teherani et al. 1979, 1982; Cazenave et al. 1987). Later, the sequencing of the *IGHG CH2* domain for 12 species of *Oryctolagus*, *Sylvilagus*, and *Lepus* genera showed a hotspot of variation at the *e locus* with four different amino acid residues in *Lepus* species (Esteves et al. 2002). Sequencing of the *IGHG* hinge domain for six species of the same *genera* showed that the leporid IgG hinge is conserved with no variation in amino acid length and only two polymorphic sites, positions 8 and 9, with four amino acid residues that can occur at position 9, the *d locus*. As this variation occurs at sites of potential O-glycosylation and/or proteolytic cleavage, the authors suggest that the underlying genetic diversity could be the outcome of selection (Esteves et al. 2006). Although a great effort has been done in the genetic diversity characterization of leporid IgG, nucleotide sequencing data of the complete *IGHG* molecule is available for only a few European rabbits (Bernstein et al. 1983; Martens et al. 1984; Ros et al. 2004).

The European rabbit 13 *IGHA* genes have also deserved some studies. Early serologic studies identified two IgA subclasses (Conway et al. 1969; Knight and Hanly 1975) and later a third IgA subclass was identified, which suggested that multiple *Cα* genes existed (Muth et al. 1983). Shortly after this a Southern blot analysis indicated that as many as ten *Cα* genes were present in the rabbit germline (Knight et al. 1985). With the advances in cloning and sequencing technology Burnett et al (1989) firstly described the 13 *Cα* genes, all showing a normal intron-exon organization and apparently functional. Given the context of mice having one *IGHA* gene (Auffray et al. 1981) and primates having two *IGHA* genes (Kawamura et al. 1992) this was a remarkable and intriguing finding that raised many questions on how and why did such a complex system arose. Despite the nucleotide sequence and *in vitro* expression analyses indicated that all 13 *IGHA* copies were functional (Burnett et al. 1989; Schneiderman et al. 1989), searching for the different IgA isotypes in different mucosal tissues failed to identify the *Cα3* and *Cα8* transcripts (Spieker-Polet et al. 1993). Analysis of the *Iα* promoter regions, which controls transcription of the *Cα* genes prior

to class switch recombination, showed that these are defective for $C\alpha 3$ and $C\alpha 8$ genes and as a result these are not expressed *in vivo* (Spieker-Polet et al. 2002). The remaining 11 rabbit IgA subclasses are expressed at varying levels and are differentially distributed among mucosal tissues (Spieker-Polet et al. 1993). The mechanism for this differential tissue expression is still not known, but the finding that IgA4 is preferentially expressed in the proximal ileum as opposed to the distal ileum suggests that intestinal microflora may influence IgA subclass distribution (Mage et al. 2006). Searching for the rabbit 3' α enhancer regions, which in human are present downstream of each $C\alpha 1$ and $C\alpha 2$ genes (Mills et al. 1997), Volgina et al. (2000) obtained the *IGHA* locus map (Figure 4). In an effort to answer the question of how did such a complex IgA system arise in rabbit, Burnett et al (1989) performed Southern blots on other Lagomorphs DNA and found multiple IgA copies for *Sylvilagus*, *Lepus*, *Pentalagus* and *Ochotona* genera, showing that the duplication of the C_α genes began in a Lagomorph ancestor. Therefore, it seems likely that all Lagomorphs have multiple IgA subclasses.

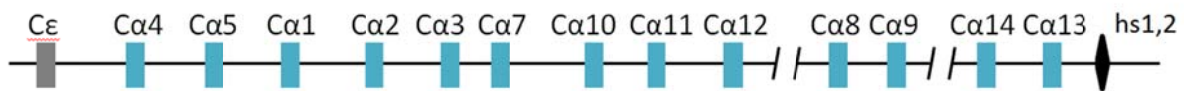


Fig. 4 – The European rabbit *IGHA* region. Represented are the 13 non allelic $C\alpha$ genes, $C\alpha 6$, not shown, is allelic to $C\alpha 5$. This region is flanked 5' by $C\epsilon$ gene and 3' by the 3' α enhancer region $hs1,2$ (adapted from Volgina et al. 2000).

3.2. The Immunoglobulin Heavy chain Variable region (*IGHV*) locus

The study of the European rabbit *IGHV* genes has revealed a number of particularities, making the rabbit a unique model among all studied mammals. Among these are the preferential usage of one *VH* gene in *VDJ* rearrangements, the existence of trans-species polymorphism in the *IGHV* locus, the unusually large genetic distances between allelic lineages and the fact that the antibody repertoire is diversified in this species only after birth. These subjects, as well as the relevance of the European rabbit in the history of immunology, are reviewed in detail in a paper presented in Chapter 2,

Paper 1: Pinheiro A, Lanning D, Alves PC, Mage RG, Knight KL, van der Loo W, Esteves PJ. (2011). Molecular bases of genetic diversity and evolution of the

Immunoglobulin Heavy chain Variable region (IGHV) gene locus in leporids. Immunogenetics, 63, 397-408.

4. Objectives and thesis organization

At the beginning of this thesis, knowledge of the leporid adaptive immune system and particularly immunoglobulin genes was mainly restricted to the European rabbit. Moreover, after having played such a relevant role in the development of modern immunology foundations during the 20th century and despite its current importance as a model for studies of human diseases and as a source of highly specific polyclonal and monoclonal antibodies of high affinity, the study of the European rabbit immune system seemed to be somewhat forgotten. Furthermore, since the 1950's the European rabbit populations have been decimated first by Myxomatosis, a fatal disease caused by the Myxoma virus that causes a benign localized cutaneous fibroma in its natural host, the Brushrabbit (*Sylvilagus* sp.) (Fenner and Ratcliffe 1965), and since the 1980's also by Rabbit haemorrhagic disease, another lethal disease caused by a Calicivirus, the Rabbit haemorrhagic disease virus (Abrantes et al. 2012). Hares (*Lepus* sp.) have also suffered population declines since the 1980's due to infection with another calicivirus, the European brown hare syndrome virus (Lopes et al. 2014). Given this context, the study of leporids immune system genes is of utmost importance. Thus, the main goal of this thesis was to extend the knowledge of European rabbit immunoglobulin genes, both variable and constant regions of the Heavy chain, to other Leporid species. This was achieved by pursuing the following specific objectives:

- i) to broaden the knowledge of Leporid VH genes evolution by answering to two hypotheses: does the maintenance of the *VHn* genes usage in low frequency in *VDJ* rearrangements represent a relic of an ancestral immunologic response to pathogens? Are the VH genes allelic lineages specificities associated with different environments?;
- ii) to investigate the action of natural selection on the evolution of IgA;
- iii) to study the whole IgG genetic diversity for European rabbit wild populations and domestic breeds, as well as the evolution of this gene in other Leporids.

The thesis is organized into five chapters and six scientific papers already published in journals indexed in the Science Citation Index (SCI) are included.

In *Chapter 1: General Introduction*, basic immunology principles are presented as well as is given the state of the art on the knowledge of European rabbit and other leporids immunoglobulins. Aspects such as the basic immunoglobulins structure, Lagomorphs taxonomy and European rabbit IGH locus are reviewed.

The European rabbit has long been an important model for immunological studies. Particularly, the study of rabbit IGHV genes has unveiled unique features and contributed to the current knowledge of Ig structure and function. Contrary to human and mice, the rabbit expresses preferentially one *VH* gene, the most *D*-proximal *VH1*, in the generation of its antibody repertoire. The *VH1* is used in 80-90% of *VDJ* rearrangements and in the remaining 10-20% are used the *VHn* genes, that map at least 100 Kb upstream of *VH1*. The maintenance of the *VHn* genes usage in low frequency in *VDJ* rearrangements has been suggested to represent a relic of an ancestral immunologic response to pathogens. Furthermore the leporid *VH* genes allelic lineages have unusually large genetic distances and remain one of the few reported examples of trans-species polymorphism outside MHC. The *VH* lineages seem to have been retained differently in the European rabbit and hares (*Lepus* sp.) and this led to the hypothesis that the *VH* specificities could be associated with different environments. In *Chapter 2: VH evolution in Leporids*, the most relevant findings achieved so far concerning the rabbit *IGHV* are reviewed and are presented the studies conducted to answer these two hypotheses. This *chapter* comprises three papers:

Paper 1: Pinheiro A, Lanning D, Alves PC, Mage RG, Knight KL, Loo Wvd, Esteves PJ (2011). Molecular bases of genetic diversity and evolution of the Immunoglobulin Heavy chain Variable region (*IGHV*) gene locus in leporids. *Immunogenetics*, 63(7), 397-408.

Paper 2: Pinheiro A, de Mera IG, Alves PC, Gortázar C, de la Fuente J, Esteves PJ (2013) Sequencing of modern *Lepus* *VDJ* genes shows that the usage of *VHn* genes has been retained in both *Oryctolagus* and *Lepus* that diverged 12 million years ago. *Immunogenetics*, 65(11), 777-84.

Paper 3: Pinheiro A, Melo-Ferreira J, Abrantes J, Martinelli N, Lavazza A, Alves PC, Gortázar C, Esteves PJ. (2014) Sequencing of *Sylvilagus* *VDJ* genes reveals a new *VHa* allelic lineage and shows that ancient *VH* lineages were retained differently in leporids. *Immunogenetics*, 66(12), 719-26.

Of the five mammalian immunoglobulin isotypes IgA is likely the one in which production a higher investment is made, as IgA is the predominant immunoglobulin isotype in mucosal tissues and external secretions as well as being the second most

prevalent immunoglobulin in serum. More in more, in addition to its role as a major line of defense against pathogens, IgA is also crucial to the maintenance of the commensal microbiota in the intestinal tract. Considering that rabbit has the most complex IgA system observed with 13 *IGHA* genes and 11 expressed IgA subclasses which are differentially distributed among mucosal tissues, we can speculate that IgA is of high importance for the rabbit immune system. To achieve its function of eliminating pathogens IgA binds through its Fc to several receptors, of which Fc α RI, pIgR and Fc α / μ R are the most relevant. On the other hand, numerous pathogens trying to evade elimination by the immune system have evolved proteins targeting IgA, either IgA-binding proteins that block IgA coupling to IgA receptors or proteases that cleave the IgA hinge uncoupling the recognition of foreign antigens from the effector functions that eliminate them. Considering the complexity of its interactions with the surrounding environment, IgA is a likely target for natural selection. In *Chapter 3: IgA evolution in Mammals*, an in depth analysis of the action of natural selection on IgA was performed using maximum likelihood methods to detect signatures of positive selection. The results are presented in one paper:

Paper 4: Pinheiro A, Woof JM, Abi-Rached L, Parham P, Esteves PJ (2013) Computational analyses of an evolutionary arms race between mammalian immunity mediated by immunoglobulin A and its subversion by bacterial pathogens. *PLoS One*, 8(9), e73934.

The European rabbit IgG has been of significant importance in immunological research and is well characterized. The two loci identified by serology, the *d* and *e loci* that locate to the hinge and CH2 domains, respectively, were used to assess European rabbit wild populations and domestic breeds IgG genetic diversity. These two *loci* were also used to screen other Leporids IgG. With these two loci in view nucleotide sequencing studies of leporids IgG have targeted the hinge and CH2 domains for a restricted number of *Sylvilagus* and *Lepus* species and complete *IGHG* nucleotide sequences are available for only a few European rabbits. Thus, there remains a void in the knowledge of the whole IgG genetic diversity for European rabbit wild populations and domestic breeds as well as of the evolution of this gene in Leporids. In *Chapter 4: IgG evolution in Leporids*, the studies pursued to answer to this gap in the knowledge of Leporids IgG are presented in two scientific papers:

Paper 5: Pinheiro A, Woof JM, Almeida T, Abrantes J, Alves PC, Gortázar C, Esteves PJ (2014) Leporid immunoglobulin G shows evidence of strong selective pressure on the hinge and CH3 domains. *Open Biology*, 4(9), 140088.

Paper 6: Pinheiro A, Almeida T, Esteves PJ (2015) Survey of genetic diversity of IgG in wild and domestic rabbits. International Journal of Immunogenetics 42(5), 364-7.

Finally in *Chapter 5: General Discussion*, the major findings and conclusions of this thesis are discussed and future research lines are presented.

5. References

- Abrantes J, van der Loo W, Le Pendu J, Esteves PJ (2012) Rabbit haemorrhagic disease (RHD) and rabbit haemorrhagic disease virus (RHDV): a review Vet Res 43:12 doi:10.1186/1297-9716-43-12
- Aggarwal SJ, Mandy WJ (1976) Lagomorph IgG hinge region: allotype associated amino acid sequence variations Immunochemistry 13:215-220
- Alves PC, Ferrand N, Hackländer K (2008) Lagomorph biology : evolution, ecology, and conservation. Springer, Berlin ; New York
- Appella E, Chersi A, Mage RG, Dubiski S (1971) Structural basis of the A14 and A15 allotypic specificities in rabbit immunoglobulin G Proc Natl Acad Sci U S A 68:1341-1345
- Asher RJ, Meng J, Wible JR, McKenna MC, Rougier GW, Dashzeveg D, Novacek MJ (2005) Stem Lagomorpha and the antiquity of Glires Science 307:1091-1094 doi:10.1126/science.1107808
- Auffray C, Nageotte R, Sikorav JL, Heidmann O, Rougeon F (1981) Mouse immunoglobulin A: nucleotide sequence of the structural gene for the alpha heavy chain derived from cloned cDNAs Gene 13:365-374
- Benton MJ, Donoghue PC (2007) Paleontological evidence to date the tree of life Mol Biol Evol 24:26-53 doi:10.1093/molbev/msl150
- Bernstein KE, Alexander CB, Mage RG (1983) Nucleotide sequence of a rabbit IgG heavy chain from the recombinant F-I haplotype Immunogenetics 18:387-397
- Burnett RC, Hanly WC, Zhai SK, Knight KL (1989) The IgA heavy-chain gene family in rabbit: cloning and sequence analysis of 13 C alpha genes EMBO J 8:4041-4047
- Can DN, Abramov AV, Tikhonov AN, Averianov AO (2001) Annamite striped rabbit *Nesolagus timminsi* in Vietnam. *Acta Theriol.* **46**, 437–440.

- Cazenave PA, Bennamar A, Sogn JA, Kindt TJ (1987) Immunoglobulin genes in feral population. In: (ed. S. Dubiski) The rabbit in contemporary immunological research, p 148. Longman Scientific & Technical Publications.
- Chapman JA, Cramer KL, Dippenaar NJ, Robinson TJ (1992) Systematics and biogeography of the New England cottontail, *Sylvilagus transitionalis* (Bangs, 1895), with the description of a new species from the Appalachian Mountains. *Proc. Biol. Soc. Washington* **105**, 841–866.
- Conway TP, Dray S, Lichter EA (1969) Identification and genetic control of the f4 and f5 rabbit gamma-A immunoglobulin allotypes. Phenogroups at the f locus J Immunol 103:662-667
- Corbet GB, Hill JE (1980) A world list of mammalian species. British Museum (Natural History); Comstock Pub. Associates, London, Ithaca, N.Y.
- Currier SJ, Gallarda JL, Knight KL (1988) Partial molecular genetic map of the rabbit VH chromosomal region J Immunol 140:1651-1659
- Danilova N, Bussmann J, Jekosch K, Steiner LA (2005) The immunoglobulin heavy-chain locus in zebrafish: identification and expression of a previously unknown isotype, immunoglobulin Z Nat Immunol 6:295-302 doi:10.1038/ni1166
- Dawson MR (1967) Lagomorph history and the stratigraphic record. U. Kansas Geol. Spec. Pub. 2, 287–316.
- Dawson M (1981). Evolution of the modern lagomorphs. In: (eds K Myers and CD McInnes) Proceeding of the World Lagomorph Conference. Univ. Guelph, Ontario, pp 1-16
- Drummond AJ, Ho SY, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence PLoS Biology 4:e88
- Dubiski S (1969) Immunochemistry and genetics of a "new" allotypic specificity Ae14 of rabbit gamma-G immunoglobulins: recombination in somatic cells J Immunol 103:120-128
- Esteves PJ, Alves PC, Ferrand N, van der Loo W (2002) Hotspot variation at the CH2-CH3 interface of leporid IgG antibodies (*Oryctolagus*, *Sylvilagus* and *Lepus*) Eur J Immunogenet 29:529-535
- Esteves PJ, Carmo C, Godinho R, van der Loo W (2006) Genetic diversity at the hinge region of the unique immunoglobulin heavy gamma (IGHG) gene in leporids (*Oryctolagus*, *Sylvilagus* and *Lepus*) Int J Immunogenet 33:171-177 doi:10.1111/j.1744-313X.2006.00588.x
- Fenner F, Ratcliffe F (1965) Myxomatosis. University Press, Cambridge
- Flajnik MF (2002) Comparative analyses of immunoglobulin genes: surprises and portents Nat Rev Immunol 2:688-698 doi:10.1038/nri889

- Flajnik MF (2014) Re-evaluation of the immunological Big Bang Curr Biol 24:R1060-1065 doi:10.1016/j.cub.2014.09.070
- Flajnik MF, Du Pasquier L (2004) Evolution of innate and adaptive immunity: can we draw a line? Trends Immunol 25:640-644 doi:10.1016/j.it.2004.10.001
- Flajnik MF, Kasahara M (2010) Origin and evolution of the adaptive immune system: genetic events and selective pressures Nat Rev Genet 11:47-59 doi:10.1038/nrg2703
- Flux JEC, Angermann R (1990) The hares and jackrabbits. In: (eds. J. A. Chapman, J.E.C. Flux) Rabbits, Hares and Pikas: Status Conservation Action Plan, pp 61-94. Gland, Switzerland: International Union for Conservation of Nature and Natural Resources.
- Flynn LJ, Winkler AJ, Erbaeva M, Alexeeva N, Anders U, Angelone C, Čermák S, Fladerer FA, Kraatz B, Ruedas LA, Ruf I, Tomida Y, Veitschegger K, Zhang Z (2014) The Leporid Datum: a late Miocene biotic marker. Mammal Review, 44: 164–176. doi: 10.1111/mam.12016
- Frey JK, Fisher RD, Ruedas LA (1997) Identification and restriction of the type locality of the Manzano Mountains cottontail, *Sylvilagus cognatus* Nelson, 1907 (Mammalia: Lagomorpha: Leporidae). Proc. Biol. Soc. Washington 110, 329–331.
- Gallarda JL, Gleason KS, Knight KL (1985) Organization of rabbit immunoglobulin genes. I. Structure and multiplicity of germ-line VH genes J Immunol 135:4222-4228
- Ge D, Wen Z, Xia L, Zhang Z, Erbaeva M, Huang C, Yang Q (2013) Evolutionary history of lagomorphs in response to global environmental change PLoS One 8:e59668 doi:10.1371/journal.pone.0059668
- Gertz EM, Schaffer AA, Agarwala R, Bonnet-Garnier A, Rogel-Gaillard C, Hayes H, Mage RG (2013) Accuracy and coverage assessment of *Oryctolagus cuniculus* (rabbit) genes encoding immunoglobulins in the whole genome sequence assembly (OryCun2.0) and localization of the IGH locus to chromosome 20 Immunogenetics 65:749-762 doi:10.1007/s00251-013-0722-9
- Gridley J (1912) The lagomorphs and independent order. Science 36:285–286
- Hamers-Casterman C, Wittouck E, van der Loo W, Hamers R (1979) Phylogeny of the rabbit gamma-chain determinants: a d12-like antigenic determinant in *Pronolagus rupestris* J Immunogenet 6:373-381
- Hamers R, Hamers-Casterman C (1965) Molecular localization of A chain allotypic specificities in rabbit IgG (7S gamma-globulin) J Mol Biol 14:288-289

- Hansen JD, Landis ED, Phillips RB (2005) Discovery of a unique Ig heavy-chain isotype (IgT) in rainbow trout: Implications for a distinctive B cell developmental pathway in teleost fish *Proc Natl Acad Sci U S A* 102:6919-6924 doi:10.1073/pnas.0500027102
- Kawamura S, Saitou N, Ueda S (1992) Concerted evolution of the primate immunoglobulin alpha-gene through gene conversion *J Biol Chem* 267:7359-7367
- Knight KL, Burnett RC, McNicholas JM (1985) Organization and polymorphism of rabbit immunoglobulin heavy chain genes *J Immunol* 134:1245-1250
- Knight KL, Hanly WC (1975) Genetic control of alpha chains of rabbit IgA: allotypic specificities on the variable and the constant regions *Contemp Top Mol Immunol* 4:55-88
- Lanier HC, Olson LE (2009) Inferring divergence times within pikas (*Ochotona* spp.) using mtDNA and relaxed molecular dating techniques *Mol Phylogenet Evol* 53:1-12 doi:10.1016/j.ympev.2009.05.035
- Lanning DK, Zhai SK, Knight KL (2003) Analysis of the 3' Cmu region of the rabbit Ig heavy chain locus *Gene* 309:135-144
- Lefranc MP et al. (2005) IMGT unique numbering for immunoglobulin and T cell receptor constant domains and Ig superfamily C-like domains *Dev Comp Immunol* 29:185-203 doi:10.1016/j.dci.2004.07.003
- Lissovsky A (2014) Taxonomic revision of pikas *Ochotona* (Lagomorpha, Mammalia) at the species level. *Mammalia* 78:199–216
- Lopes AM et al. (2014) Molecular evolution and antigenic variation of European brown hare syndrome virus (EBHSV) *Virology* 468-470:104-112 doi:10.1016/j.virol.2014.08.002
- Lopez-Martinez N (1989) Revision sistemática y biostratigrafca de los lagomorphos (Mammalia) del Terciário y Cuaternario de Espana. *Memorias del Museo Paleontologico de la Universidad de Zaragoza*, n°3. Diputacion General de Aragon.
- Mage R (1981) The phenotypic expression of rabbit immunoglobulins: A model of complex regulated gene expression and cellular differentiation *Contemp Top Mol Immunol* 8:89-112
- Mage RG, Lanning D, Knight KL (2006) B cell and antibody repertoire development in rabbits: the requirement of gut-associated lymphoid tissues *Dev Comp Immunol* 30:137-153 doi:10.1016/j.dci.2005.06.017
- Martens CL, Currier SJ, Knight KL (1984) Molecular genetic analysis of genes encoding the heavy chains of rabbit IgG *J Immunol* 133:1022-1027

- Matthee CA, Robinson TJ (1996) Mitochondrial DNA differentiation among geographical populations of *Pronolagus rupestris*, Smith's red rock rabbit (Mammalia: Lagomorpha) *Heredity* (Edinb) 76 (Pt 5):514-523
- Matthee CA, van Vuuren BJ, Bell D, Robinson TJ (2004) A molecular supermatrix of the rabbits and hares (Leporidae) allows for the identification of five intercontinental exchanges during the Miocene *Syst Biol* 53:433-447
- Melo-Ferreira J, Lemos de Matos A, Areal H, Lissovsky AA, Carneiro M, Esteves PJ (2015) The phylogeny of pikas (*Ochotona*) inferred from a multilocus coalescent approach *Mol Phylogenet Evol* 84:240-244 doi:10.1016/j.ympev.2015.01.004
- Mills FC, Harindranath N, Mitchell M, Max EE (1997) Enhancer complexes located downstream of both human immunoglobulin Calpha genes *J Exp Med* 186:845-858
- Muth KL, Hanly WC, Knight KL (1983) Serologic cross-reactions among rabbit secretory IgA molecules: evidence for multiple subclasses of secretory IgA-f molecules *Mol Immunol* 20:989-999
- Nguyen VK, Hamers R, Wyns L, Muyldermans S (2000) Camel heavy-chain antibodies: diverse germline V(H)H and specific mechanisms enlarge the antigen-binding repertoire *EMBO J* 19:921-930 doi:10.1093/emboj/19.5.921
- Ota T, Rast JP, Litman GW, Amemiya CT (2003) Lineage-restricted retention of a primitive immunoglobulin heavy chain isotype within the Dipnoi reveals an evolutionary paradox *Proc Natl Acad Sci U S A* 100:2501-2506 doi:10.1073/pnas.0538029100
- Pancer Z, Amemiya CT, Ehrhardt GR, Ceitlin J, Gartland GL, Cooper MD (2004) Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey *Nature* 430:174-180 doi:10.1038/nature02740
- Pinheiro A, Neves F, Lemos de Matos A, Abrantes J, van der Loo W, Mage R, Esteves PJ (2015) An overview of the lagomorph immune system and its genetic diversity *Immunogenetics* doi:10.1007/s00251-015-0868-8
- Rambaut A, Drummond AJ (2007) Tracer v1.4: Available from <http://beast.bio.ed.ac.uk/Tracer>
- Roitt IM, Brostoff J, Male DK (2001) *Immunology*. 6th edn. Mosby, Edinburgh ; New York
- Robinson TJ, Matthee CA (2005), Phylogeny and evolutionary origins of the Leporidae: a review of cytogenetics, molecular analyses and a supermatrix analysis. *Mammal Review*, 35: 231–247. doi: 10.1111/j.1365-2907.2005.00073.x

- Ros F, Puels J, Reichenberger N, van Schooten W, Buelow R, Platzner J (2004) Sequence analysis of 0.5 Mb of the rabbit germline immunoglobulin heavy chain locus Gene 330:49-59 doi:10.1016/j.gene.2003.12.037
- Rose KD, DeLeon VB, Missiaen P, Rana RS, Sahni A, Singh L, Smith T (2008) Early Eocene lagomorph (Mammalia) from Western India and the early diversification of Lagomorpha. Proc Biol Sci 275:1203-1208 doi:10.1098/rspb.2007.1661
- Schneiderman RD, Hanly WC, Knight KL (1989) Expression of 12 rabbit IgA C alpha genes as chimeric rabbit-mouse IgA antibodies. Proc Natl Acad Sci U S A 86:7561-7565
- Schroeder HW, Jr., Cavacini L (2010) Structure and function of immunoglobulins. J Allergy Clin Immunol 125:S41-52 doi:10.1016/j.jaci.2009.09.046
- Schwager J, Bürckert N, Courtet M, Du Pasquier L (1991) The ontogeny of diversification at the immunoglobulin heavy chain locus in Xenopus. EMBO J. 10(9):2461-70
- Spieker-Polet H, Yam PC, Knight KL (1993) Differential expression of 13 IgA-heavy chain genes in rabbit lymphoid tissues. J Immunol 150:5457-5465
- Spieker-Polet H, Yam PC, Knight KL (2002) Functional analysis of I alpha promoter regions of multiple IgA heavy chain genes. J Immunol 168:3360-3368
- Springer MS, Murphy WJ, Eizirik E, O'Brien SJ (2003) Placental mammal diversification and the Cretaceous-Tertiary boundary Proc Natl Acad Sci U S A 100:1056-1061 doi:10.1073/pnas.0334222100
- Surridge AK, Timmins RJ, Hewitt GM, Bell DJ (1999) Striped rabbits in Southeast Asia. *Nature* **400**, 726.
- Teherani J, Capra JD, Aggarwal S, Mandy WJ (1979) Amino acid sequence analysis of group e allotype-related peptides derived from lagomorph IgG Eur J Immunol 9:690-695 doi:10.1002/eji.1830090906
- Teherani J, Capra JD, Mandy WJ (1982) Amino acid sequence of the CH2 domain from various lagomorph IgGs Mol Immunol 19:841-846
- van der Loo W, Hamers-Casterman C (1979) Phylogeny of the rabbit immunoglobulin allotypes of the a-locus. Evidence for polymorphic variable region genes in a Paleolagina Romerolagus diazi [proceedings] Arch Int Physiol Biochim 87:212-213
- Van Valen L (1973) A new evolutionary law. Evolutionary Theory 1:1-30.
- Volgina VV, Kingzette M, Zhai SK, Knight KL (2000) A single 3' alpha hs1,2 enhancer in the rabbit IgH locus J Immunol 165:6400-6405
- Warr GW, Magor KE, Higgins DA (1995) IgY: clues to the origins of modern antibodies Immunol Today 16:392-398 doi:10.1016/0167-5699(95)80008-5

- Whiteford CEM (1995) Molecular phylogeny of the genus *Pronolagus* (Mammalia: Lagomorpha) and the use of morphological and molecular characters in the delineation of *P. rupestris*. MSc Thesis, University of Pretoria.
- Wilson M, Bengten E, Miller NW, Clem LW, Du Pasquier L, Warr GW (1997) A novel chimeric Ig heavy chain from a teleost fish shares similarities to IgD Proc Natl Acad Sci U S A 94:4593-4597
- Zhang SM, Adema CM, Kepler TB, Loker ES (2004) Diversification of Ig superfamily genes in an invertebrate Science 305:251-254 doi:10.1126/science.1088069
- Zhao Y et al. (2009) Ornithorhynchus anatinus (platypus) links the evolution of immunoglobulin genes in eutherian mammals and nonmammalian tetrapods J Immunol 183:3285-3293 doi:10.4049/jimmunol.0900469
- Zhao Y et al. (2002) Artiodactyl IgD: the missing link J Immunol 169:4408-4416
- Zhao Y et al. (2006) Identification of IgF, a hinge-region-containing Ig class, and IgD in *Xenopus tropicalis* Proc Natl Acad Sci U S A 103:12087-12092 doi:10.1073/pnas.0600291103

Chapter 2

VH evolution in Leporids

Paper 1

Molecular bases of genetic diversity and evolution of the immunoglobulin heavy chain variable region (*IGHV*) gene locus in leporids.

Ana Pinheiro, Dennis Lanning, Paulo C. Alves, Rose G. Mage, Katherine L. Knight, Wessel van der Loo and Pedro J. Esteves

Immunogenetics, 63(7), 397-408. doi: 10.1007/s00251-011-0533-9.

Abstract

The rabbit has long been a model for studies of the immune system. Work using rabbits contributed both to the battle against infectious diseases such as rabies and syphilis, and to our knowledge of antibodies' structure, function, and regulated expression. With the description of rabbit Ig allotypes, the discovery of different gene segments encoding immunoglobulins became possible. This challenged the "one gene-one protein" dogma. The observation that rabbit allotypic specificities of the variable regions were present on IgM and IgG molecules also led to the hypothesis of Ig class switching. Rabbit allotypes contributed to the documentation of phenomena such as allelic exclusion and imbalance in production of allelic gene products. During the last 30 years, the rabbit Ig allotypes revealed a number of unique features, setting them apart from mice, humans and other mammals. Here, we review the most relevant findings concerning the rabbit *IGHV*. Among these are the preferential usage of one *VH* gene in *VDJ* rearrangements, the existence of trans-species polymorphism in the *IGHV* locus

revealed by serology and confirmed by sequencing *IGHV* genes in *Lepus*, the unusually large genetic distances between allelic lineages and the fact that the antibody repertoire is diversified in this species only after birth. The Whole Genome Sequence of rabbit, plus re-sequencing of additional strains and related genera, will allow further evolutionary investigations of antibody variation. Continued research will help define the roles that genetic, allelic and population diversity at antibody loci may play in host-parasite interactions.

Keywords

leporids, rabbit, *IGHV*, evolution, immunoglobulin repertoire

Historic perspective

Studies of the rabbit immune system have greatly contributed to our knowledge of the structure, function and regulation of antibodies. During the late 19th and early 20th centuries, the use of rabbit in immunological research was crucial to the development of the rabies vaccine by Louis Pasteur and to the battle against syphilis (reviewed in Dubiski 1987). In the first half of the 20th century, the foundations of molecular immunology were laid with almost exclusive use of the rabbit. The antigenic polymorphism of serum components was described in this species as early as 1902, although the antigens involved were characterized only 50 years later (see Kelus and Gell 1967). In 1956, Oudin demonstrated and defined allotypy of immunoglobulins in the rabbit. Allotypes were proven to be hereditary traits in rabbit by Dubiski and coworkers (1959) prior to the establishment of the two major linkage groups “a” and “b”, corresponding to the H and light chain loci, respectively (Oudin 1960), and even before the heterodimeric structure of the antibody molecule was clearly established.

The existence of genetic markers of the different gene fragments encoding the antibody H chain was unique to this species and the rabbit Ig allotypes challenged the “one gene-one protein” dogma. Indeed, Todd (1963) and others found that rabbit allotypic specificities of the variable regions were present on both IgM and IgG molecules, suggesting that the same *VH* gene segment can be translocated to different constant region genes. This observation opened the road to confirmation of the concept of multiple “germline” *VH* gene segments that can be joined to a limited number of genes encoding the different heavy chain constant regions (Dreyer and Bennett 1965), and led to the hypothesis of Ig class switching (Kearny et al. 1976; van der Loo et al. 1979; reviewed in Severinson et al. 1982). Hamers and co-workers (1966) reported cis-expression of the *VH* and *CH* genes by using allelic markers on both protein domains. The markers in the CH2 domain were re-discovered by Dubiski (1969) and renamed the e14 and e15 allotypes. Mage and co-workers (1971) reported the first of a number of crossing-over events observed by laboratories during breeding. Findings of genetic recombination between the genes controlling the rabbit *VH* and *CH* (estimated recombination frequency of 0.1%; (Mage 1979; Kelus and Steinberg 1991)) confirmed the model of Dreyer and Bennett (1965).

Documentation of the current *V-D-J-C* model therefore originated through studies of the rabbit allotypes. In addition to genetic linkage of *VH* and *CH* genes, and the expression of apparently identical *VH* regions on different classes of Ig, it was through studies of rabbit immunoglobulin markers that phenomena such as allelic

exclusion and imbalance in production of allelic gene products were described (e.g. Davie et al. 1971; Looor and Kelus 1978; Schmale et al. 1969; Wolf et al. 1971). During the last 30 years, the rabbit Ig allotypes revealed a number of unique features, setting them apart from mice, humans and other mammals. Among these are the preferential usage of one *VH* gene in *VDJ* rearrangements, the evidence for the existence of trans-species polymorphism in the *IGHV*, *IGHG* and *IGKC1* loci, the unusually large genetic distances between the allelic lineages and the fact that the antibody repertoire is diversified in this species only after birth.

Lagomorph taxonomy

The order LAGOMORPHA comprises two families: OCHOTONIDAE and LEPORIDAE. The family OCHOTONIDAE includes 30 species restricted to the genus *Ochotona* (Pikas). The family LEPORIDAE can be divided into two groups: hares and rabbits. According to Chapman and Flux (1990), the hare group encompasses a single genus, *Lepus*, whereas the rabbit group includes 10 genera (*Brachylagus*, *Bunolagus*, *Caprolagus*, *Nesolagus*, *Oryctolagus*, *Pentalagus*, *Poelagus*, *Pronolagus*, *Romerolagus* and *Sylvilagus*) and 25 species (Alves and Hacklander, 2008). There is not a consensus estimate of the divergence time between rabbits and hares. Analyses of several nuclear and mitochondrial markers suggest that the genera *Oryctolagus* and *Lepus* diverged between 6 and 20 Mya (Biju-Duval et al. 1991; Halanych and Robinson 1999; Matthee et al. 2004) (Fig. 1), although fossil data suggests an earlier divergence (between 2.5 and 3.5 million years ago (Mya) (Lopez-Martinez 2008).

The genus *Oryctolagus* is monospecific (*O. cuniculus*). In the Iberian Peninsula, where the fossil record suggests the European rabbit originated (Pages 1980; Lopez-Martinez 1989, 2008; Corbet 1994), two morphologically differentiated subspecies of European rabbit have been distinguished: *O. cuniculus algirus* and *O. cuniculus cuniculus* (Cabrera 1914). These two subspecies diverged ~1.8 Mya (reviewed in Carneiro et al. 2009). *O. c. algirus* inhabits the southwestern Iberian Peninsula, while *O. c. cuniculus* is present in the northeastern Iberian Peninsula. *O. c. cuniculus* later expanded its range North towards France, likely after the last glacial peak (Queney et al. 2001), where it still remains present. The Pleistocene Ice ages forced many temperate species to retreat into southern refugia, leading to high levels of diversity and endemism in these areas (Hewitt 1996). The Iberian Peninsula was one of these refugia in Europe (Taberlet et al. 1998). This is reflected in the significantly lower

genetic diversity of the wild French *O. cuniculus* populations compared to the Iberian populations (van der Loo et al. 1991, 1999; Queney et al. 2001; Esteves et al. 2004; Ferrand and Branco 2007; Surridge et al. 2008; Carneiro et al. 2011). The European rabbit gene pool has been manipulated through a recent single domestication event of French origin, and therefore all domestic rabbits belong to subspecies *O. c. cuniculus* (reviewed in Ferrand and Branco 2007, Carneiro et al. 2011). Today, through mediated anthropic dispersal, the European rabbit is present in Continental Europe, England, Australia, New Zealand, North and South America, and North Africa.

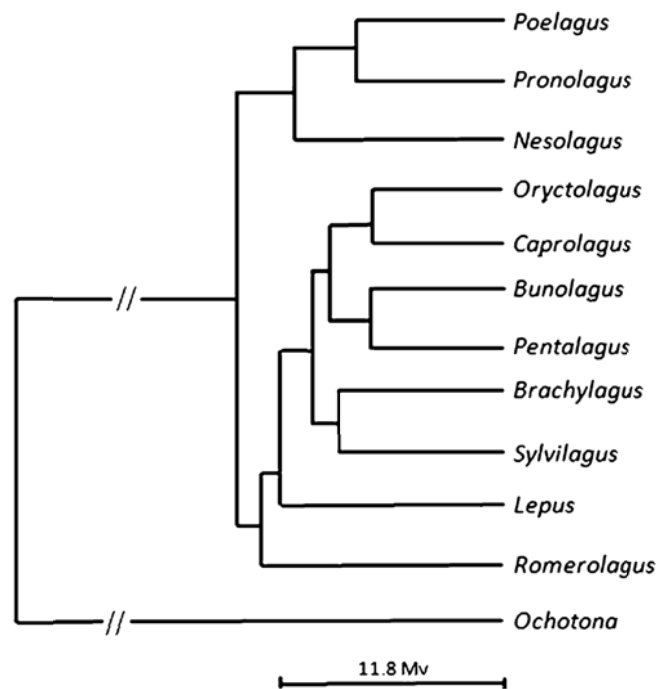


Fig. 1 - Evolutionary topology reflecting the relationships within the Lagomorpha group based on a molecular super matrix (adapted from Matthee et al. 2004).

For other leporids, the study of immunoglobulin genes is mainly restricted to the genera *Lepus* and *Sylvilagus* (Cazenave et al. 1977, Teherani and Mandy 1976a, 1976b; Teherani et al. 1979; Teherani et al. 1982; Bouton and van der Loo 1997; Esteves et al. 2002a, 2005, 2006). *Lepus* is a polytypic, cosmopolitan genus, comprising 24 to 30 currently recognized species (Corbet and Hill 1980; Flux and Angermann 1990; Alves and Hacklander 2008), that, like *Sylvilagus*, most probably originated in North America (Lopez-Martinez 2008, Matthee et al. 2004). *Lepus* species that have been used in immunogenetic studies include *L. americanus*, *L. timidus*, *L. granatensis* and *L. europaeus*. *L. americanus* is distributed throughout North America and occupies a basal position in the *Lepus* group (Matthee et al. 2004). *L. granatensis*

is endemic to the Iberian Peninsula, covering the whole peninsula except the Northeast where it is replaced by *L. europaeus*, which spans a wide range throughout Europe, with several introduced populations in different regions, like South America, Australia, New Zealand and several islands including Barbados, Reunion and Falklands (reviewed in Alves and Hacklander 2008). *L. timidus* is a Palearctic relict species with an extensive distribution area in Europe and Asia, from Ireland and Scandinavia in the west, across Siberia, Mongolia and China to the northern islands of Japan in the east. *Sylvilagus* is restricted to the American continent and comprises seventeen species (reviewed in Alves and Hacklander 2008). Within this genus, the most studied species is *S. floridanus*, a species that occupies Southern and Western North America.

Rabbit preferentially expresses *IGHV1*

The rabbit *IGHV* allotypes are highly divergent and behave as Mendelian alleles (Oudin 1956, 1960; Dubiski et al. 1959; Dray et al. 1963; Kim and Dray 1972; Mage et al. 1984a). Serological surveys of domestic rabbits defined three allotypic lineages, the so-called *VHa* allotypes *a1*, *a2* and *a3*. *VH* allotypic markers have not been found in other species, and the Mendelian behavior of rabbit *IGHV* allotypes was puzzling for many years, as rabbit serological data revealed a complex situation where homozygous *VHa* rabbits were found to express distinct *VH* genes. Some of them were devoid of *VHa* allotype-specific determinants (the so-called *VHa* negative or *VHn*, alias *VHx*, *VHy* and *VHz*; Horng et al. 1976). The *VHa* positive molecules (i.e. displaying motifs of *VHa*-characteristic determinants) also showed variation which revealed the existence of multiple *VHa*-related gene fragments (Brezin and Cazenave 1975, Mage et al. 1976, van der Loo et al. 1977). This was later confirmed by genomic mapping of the rabbit *VH* region which revealed the existence of over 200 *VH* elements in the germline (Currier et al. 1988; Gallarda et al. 1985; Ros et al. 2004). This may, in part, explain why for many years it was so difficult to understand the presence of allelic allotypic markers on most serum immunoglobulins of all classes (and probably the skepticism of researchers less acquainted with serological methods).

An important contribution to understanding the mechanisms underlying preferential *VDJ* rearrangement in the rabbit was the study of a rabbit strain, called “Alicia”, detected as a mutant during breeding at the Basel Institute, Switzerland (Kelus and Weiss 1986). The mutant rabbits derived from a *VH1a2* parental rabbit in which a 10 Kb segment of genomic DNA containing the *VH1* gene was deleted (Knight and

Becker 1990, Allegrucci et al. 1990, Ros et al. 2004). In contrast to normal individuals of the *a2* lineage, the young homozygous *alilali* mutant Alicia rabbits produced only trace amounts of *a2* molecules and their serum contained mostly Ig resulting from rearranged genes not encoding *VHa* allotype-associated epitopes (*VHa* negative or *VHn* genes) (Kelus and Weiss 1986; Di Pietro et al. 1990; Chen et al. 1993). As the genes encoding the *VHn* allotypes map at least 100 Kb upstream of *VH1* (Mage et al. 2006), it is intriguing that *VDJ* rearrangements of these appeared to be more prevalent than *VDJ* rearrangements of the *D*-proximal functional *VHa* positive *VH4* gene in young Alicia rabbits. Further analysis showed that the *VH4* gene is in fact the predominantly rearranged *VH* gene found in the bone marrow of young Alicia rabbits, but for some reason most *VH4*-utilizing B-cells are eliminated. Thus, the immunoglobulins of the young rabbits are produced by B-cells that utilize *VHn*. As the young "Alicia" rabbit ages, the proportions of B cells expressing serological *a2* specificities increase (Kelus and Weiss 1986; Di Pietro et al. 1990; Allegrucci et al. 1990; Chen et al. 1993; Pospisil et al. 1995). Analysis of nucleotide sequences of the promoter region showed that more than 80% of the *VDJ* rearrangements in older Alicia rabbits utilize either the functional *VH4* or *VH7* genes localized upstream of *VH1*. The *VH4* and the *VH7* genes have 7 (out of 11) specific nucleotides associated with the allotype *a2*, while the other nucleotides that characterize *a2* are gained through somatic gene conversion using *VH9* or a *VH9*-like germline gene as donor (Sehgal et al. 1998; Zhu et al. 1999).

The study of this mutant rabbit strain showed that the Mendelian inheritance of the *VHa* allotypes in normal rabbit is explained by the preferential usage of only one *VH* gene in *VDJ* rearrangements, i.e. the *D*-proximal *VH* gene segment *VH1* (see Figure 2), which is deleted from the "ali" genome (Knight and Becker 1990; Knight 1992, Allegrucci et al. 1990, Ros et al. 2004). Despite having more than 200 *VH* genes (Currier et al. 1988; Gallarda et al. 1985; Ros et al. 2004), over 50% are "non-functional" and, apart from *VH1*, an even smaller fraction seems to encode *a*-positive proteins (Ros et al. 2004). About 80 to 90% of circulating Ig molecules are derived from the *VH1* gene and express the *VHa* allotypic markers (e.g. Kindt 1975; Margolies et al. 1977). The *VH* regions of the remaining 10-20% of Ig molecules are encoded by the *VHn* genes (Kim and Dray 1973; Roux 1981), which are localized at least 100 Kb upstream of *VH1* (Mage et al. 2006). Thus far, the basis for the preferential usage of the *VH1* gene in *VDJ* rearrangements remains unanswered. Enhanced chromatin accessibility of this region of the DNA may be responsible (Mage et al. 2006). Furthermore, we can speculate that the retention and occasional usage of the *VHn* genes in 10-20% of *VDJ* rearrangements may represent an evolutionary relic of Lagomorphs ancestral immunologic response to pathogens.

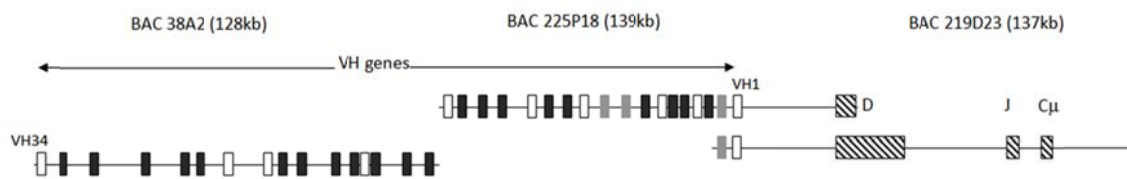


Fig. 2 - Rabbit IGHV (regions from V_H haplotype (a2 F-I)) for which nucleotide sequence is available from sequenced, assembled BAC clones. Depicted are V genes, D and J regions and also C_μ. V_H functional genes are represented in white rectangles, V_H possible pseudogenes in grey and V_H pseudogenes in black, identified according to criteria described in Ros et al. (2004). The D-proximal V_H gene, V_H1, used in the majority of VDJ rearrangements is indicated. Identified above are the non-overlapping BAC clones 38A2 and 225P18 (GenBank Accession Numbers AY386694 and AY386697), and 219D23 (GenBank Accession Number AY386695) from a different BAC library. The 3' end of BAC 225P18 overlaps the 5' end of BAC 219D23 (adapted from Ros et al. 2004).

Genetic diversity at the heavy chain variable region *IGHV* locus (a locus)

In the domestic rabbit, the three serologically defined allotypic lineages, *VHa* allotypes *a1*, *a2* and *a3*, are highly divergent ($\pm 20\%$ amino acid sequence differences). The allelic specificities of *a1* and *a2* are correlated with several amino acid differences in framework regions 1 (positions 5, 8, 11, 13, 14, 17 and 18) and 3 (positions 74, 76, 79, 80, 83, 84, 96 and 97) (Fig. 3) (Tonnellet et al. 1983; Mage et al. 1984a; Knight and Becker, 1990). Serologically distinct *VHa* allotypes were discovered in wild rabbits and named VHa100-VHa109 by Cazenave and coworkers (reviewed in Cazenave et al. 1987). Partial protein sequences of purified antibodies of the VHa100 type (Tonnellet et al. 1983), suggest these are related to the more recently described and fully sequenced *VHa* allele(s) corresponding to the *a4* lineage (Esteves et al. 2004). Studies by double immunodiffusion of sera from domestic breeds and wild rabbits of Continental Europe (North of Pyrenees Mountains), Great Britain and overseas showed the presence of allotypes *a1*, *a2* and *a3*, as well as a rare allotype with partial reaction against *a3*-specific antiserum, found in the French populations. In all studied populations, the gene frequencies of *a1* and *a3* were similar ($\pm 40\%$) and higher than *a2* ($\pm 20\%$) (van der Loo, 1993). Interestingly, these allelic frequencies reflect the “pecking order” of their relative expression in heterozygous animals (Mage, 1967; Lummus et al. 1967). Quite surprisingly, the study of wild rabbit populations from the Iberian Peninsula belonging to the subspecies *O. c. algirus* and *O. c. cuniculus* revealed a different picture with a much higher prevalence of the “a-blank” allele. Also, a gradient in the gene frequencies of this allotype was noted, decreasing from

southwestern to northeastern populations (Fig. 4) (Esteves et al. 2004). Sequencing of rearranged *VH* genes expressed in *O. c. algerus* rabbits that were typed as “a-blank” showed that these rabbits preferentially use *VH* genes that, although clearly related to the known *VHa* genes, defined a new major allotypic lineage, designated *a4*.

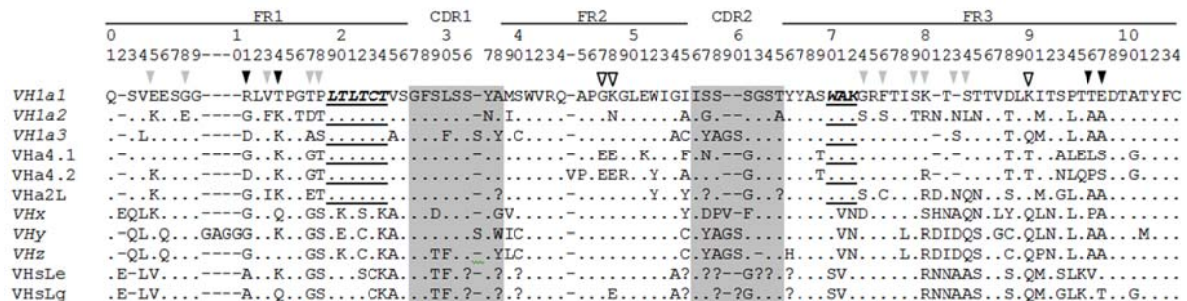


Fig. 3 - Amino acid sequences of european rabbit (*O. cuniculus*) proteins (encoded by VH1a1-a3, VH4.1-4.2, VHx, VHy, VHz) and hare (VHa2L, VHsLe, VHsLg) VH genes. Germline elements are *italicized* (Accession numbers: *VH1a1*-M93171; *VH1a2*-M93172; *VH1a3*-M93173; *VHx*-L03846; *VHy*-L03890; *VHz*-AF264469.). VHa4.1-4.2, VHa2L, VHsLe and VHsLg represent consensus protein sequences obtained by Esteves et al. 2004, 2005. Framework regions (FR1, FR2, FR3) and each CDR (CDR1 and CDR2) are defined according to the IMGT numbering system (also shown). CDRs are shaded. The amino acids that are characteristic of *a1*, *a2* and *a4* lineages are marked with black, grey and white arrows, respectively. Hallmark rabbit VH α residues are highlighted in bold and underlined. Note that the *Lepus* VHa2L shares these residues, as well as five of the VH1a2 lineage characteristic amino acids. Dots (.) indicate identity with *VH1a1*, dashes (-) represent indels and question marks (?) represent ambiguous positions in consensus sequences.

The *a4* sequences display the rabbit *VHa* hallmark residues together with a number of unprecedented amino acid changes in FR2 and FR3, which may explain the lack of reactivity with known alloantisera (Fig. 3). The net protein distances between the VH-a4 and the VH-a1, a2, and a3 lineages, was 20%, 29%, and 21% respectively, similar to the net protein distances between VH-a1, a2, and a3 lineages. As such, a new *VHa* allele was found in Iberian *O. c. algerus* populations, constituting the fourth of the distantly related lineages of the rabbit *VHa*-locus, one of which, the *a4* lineage, seems to be endemic in the Iberian range (Esteves et al. 2004). The data for the *VHa* lineages also suggested that some lineages, namely *VHa a1*, *a2* and *a4*, showed higher evolutionary rates than the *VHa a3* and *VHn* lineages. This increased evolutionary rate may account for the accumulation of a considerable fraction of the amino acid differences characterizing the *a4* lineage after the separation of the two *O. cuniculus* subspecies, explaining its association with subspecies *O. c. algerus* (Esteves et al. 2004).

Studies on the *IGHV* locus diversity for other leporids are limited to a few data obtained for some *Lepus* species, namely *L. americanus*, *L. europaeus* and *L. granatensis*. For *Lepus americanus*, only serological analyses have been conducted;

these showed cross-reaction with rabbit anti-a1, -a2 and -a3 antisera and the absence of individuals with no cross-reaction to rabbit antisera (De Poorter, 1984). As for *L. europaeus* and *L. granatensis*, a serological analysis of several populations using allo-antisera against rabbit a1, a2 and a3 allotypes revealed only two phenotypes: partial reaction to anti-a2 antisera and no reaction to any rabbit antiserum (Table 1) (Esteves, 2003). Sequences of expressed *VH* genes were obtained for only five *Lepus* specimens of both species, three of which expressed a2-cross-reacting serum proteins, and these could be grouped into two different lineages (see below 'The trans-specific polymorphism of leporids *IGHV*'). Extending the study of *IGHV* diversity to more *Lepus* individuals and species could unravel more *VH* lineages, as shown by the rabbit *VH1-a4* lineage found only in *O. c. algirus*.

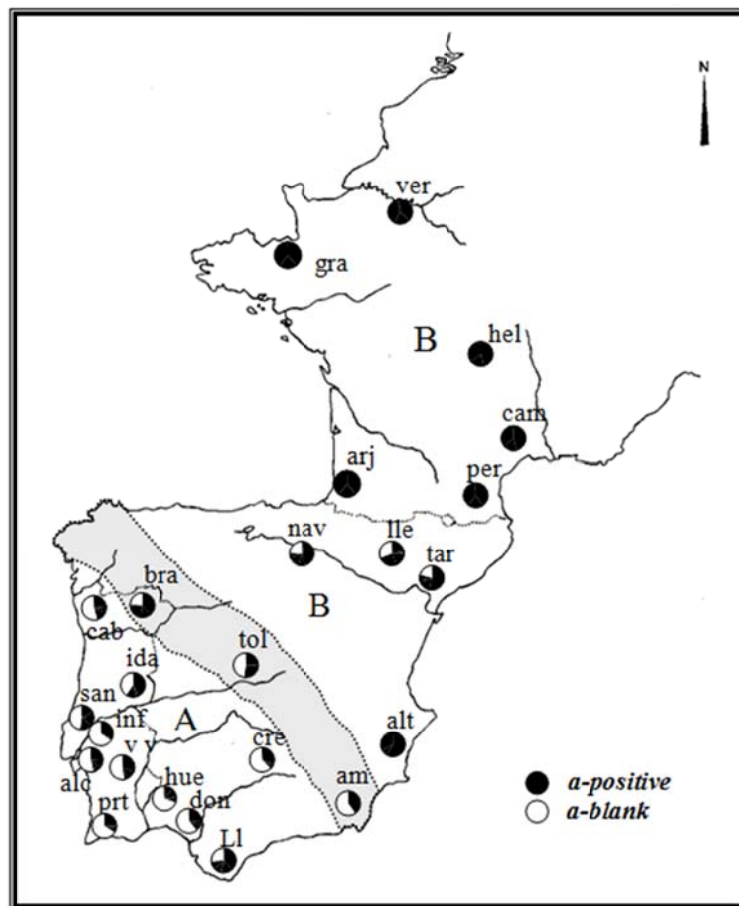


Fig. 4 - Geographic distribution of *VHa* locus allotype frequencies, serologically determined for populations of European wild rabbit (*O. cuniculus*) from the Iberian Peninsula and France. The subspecies *O. c. algirus* occupies the Southwestern part of Iberia, which is indicated by A. Rabbits of the Northwestern areas of Iberia and of the rest of Europe (indicated by B) belong to the subspecies *O. c. cuniculus*. The contact zone between the two subspecies is indicated in gray. The coloring of the disks reflects the relative allele frequencies per locality, analyzed as a two-allele locus with a-blank (white) and a-positive (black) alleles. Locality abbreviations are as follows: prt - Portimão; san - Santarém; alc - Alcochete; ida - Idanha; cab - Cabreira; LI - Las Lomas; hue - Huelva; don - Doñana; cre - Ciudad Real; inf - Infantado; vv - Vila Viçosa; am - Amoladeras; tol - Toledo; bra - Bragança; tar-Tarragona; alt- Alicante; nav - Navarra; lle - Lerida; per - Perpigan; cam - Camargue; ver - Versailles; arj - Arjuzanx; gra - Grax; hel - Helene.

Table 1 - Gene frequencies obtained in populations of *Lepus granatensis* and *L. europaeus* at IGHV. *Lepus* individuals were tested for cross-reaction with rabbit anti -a1, -a2 and -a3 antisera. Serotype codes are shown in parentheses: 1 indicates complete identity with rabbit allo-antisera, p partial identity and 0 no reaction (Esteves 2003).

	n	a1 (100)	a2 (010)	a3 (001)	a1v (p00)	a2v (0p0)	a3v (00p)	blank (000)
<i>Lepus granatensis</i>								
Portugal								
Bragança	13	-	-	-	-	0.08	-	0.92
Santarém	20	-	-	-	-	0.41	-	0.59
Idanha	20	-	-	-	-	0.29	-	0.71
Pancas	20	-	-	-	-	0.45	-	0.55
Aljustrel	20	-	-	-	-	0.55	-	0.45
Alcochete	7	-	-	-	-	0.08	-	0.92
Spain								
Granada	20	-	-	-	-	0.19	-	0.81
<i>average unweighted</i>	120	-	-	-	-	0.29	-	0.71
<i>Lepus europaeus</i>								
Lass (France)	20	-	-	-	-	0.5	-	0.5

The trans-specific polymorphism of leporid *IGHV*

Such large interallelic distances as those observed at the rabbit *VHa* locus can be the outcome of unusually long allele persistence times and/or increased mutant recruitment rates. The first indications that the different *VH1* polymorphisms could be trans-specific came from studies showing serological cross-reactivity of *Lepus* sera with rabbit *VHa*-allotype-specific allo-antisera (van der Loo et al. 1977, Horng et al. 1980, van der Loo, 1987). Indeed, for *Lepus americanus* large scale serological analysis, comprising several hundreds of specimens, showed cross-reaction with rabbit anti-a1, -a2 and -a3 antisera, revealing at least three alleles with hierarchical frequency distributions similar to those observed for feral rabbits (De Poorter, 1984; van der Loo, 1987). Su and Nei (1999) compared the extent of sequence divergence between the rabbit *a1*, *a2* and *a3* allotypes with that between human and mouse *VH* gene sequences and concluded that, assuming a “normal” mutation rate, the rabbit *VH1* polymorphism has persisted for about 50 My. Since the *Oryctolagus* and *Lepus* genera diverged 12 Mya (Matthee et al. 2004), the allelic lineages present in one leporid species should be more related to some of the alleles expressed in the other species than to their conspecific allelic counterparts. Serological cross-reactivity can depend

upon a small number of amino acid replacements, and allo-antisera may reveal plesiomorphic character states (i.e., which became antigenic in the recipient strain when they were altered), and so the hypothesis that the *VH1* polymorphism preceded the rabbit-hare split remained controversial until sequence data was obtained.

Focusing on this question, Esteves and co-workers (2005) sequenced rearranged *VH* gene products from five *Lepus* specimens of both *L. granatensis* and *L. europaeus* species, three of which expressed $\alpha 2$ -cross-reacting serum proteins, and compared them to known rabbit *VH* sequences of different allotypes. Within the *Lepus* *VH* genes, two lineages were observed, one of which ($\alpha 2L$) showed some of the most outstanding characteristics of the rabbit *VH α* genes, furthermore possessing five out of eleven amino acid residues that characterize the allotype *VH α 2* (Mage et al. 1984a). This lineage was only obtained from specimens showing rabbit $\alpha 2$ antisera cross-reaction. A second lineage displayed more similarity to the rabbit *VH n* gene fragments (*sL*). Phylogenetic inference methods clearly placed the hare $\alpha 2L$ sequences into the same monophyletic group with the rabbit $\alpha 2$ sequences (Fig. 5), thus showing that these sequences were more closely related to the rabbit *VH1- $\alpha 2$* allele than *VH1- $\alpha 2$* was to its allelic counterparts *VH1- $\alpha 1$* and *VH1- $\alpha 3$* . Also, the genetic distance measured between different rabbit genomic *VH α* sequences was approximately 1.5 times larger than between the *Lepus* $\alpha 2L$ consensus and genes of the rabbit $\alpha 2$ allotype. Together, these findings strongly suggest that the allotype split predates the *Oryctolagus-Lepus* species split, supporting previous evidence of the trans-species nature of the *VH1* polymorphism. The alternative interpretation, that the evolutionary rate of the rabbit $\alpha 2$ lineage has been significantly slower than at the other lineages, was not supported by the data, as the $\alpha 2$ sequences were rather more derived compared to the other allotypes, in particular, the $\alpha 3$ lineage.

The large inter-allelic differences at the *VH1* locus

Regarding the two different hypotheses invoked to explain the very large inter-allelic distances observed at the *IGHV1* locus, the results obtained seem to be contradictory. On the one hand, the association of specific alleles with subspecific markers suggests that the large differences between $\alpha 4$ and domestic alleles ($\alpha 1$, $\alpha 2$ and $\alpha 3$) may have accumulated after the separation of the subspecies, with different evolutionary rates among lineages. On the other hand, the confirmation of the trans-specific nature of the *VH1* polymorphism between *Oryctolagus* and *Lepus* supports the

hypothesis that very long lineage persistence times have contributed to allelic divergence. The evidence obtained points to two major conclusions, 1) the allelic lineages can be maintained in the genome for a long time and 2) mutation rates can differ between allelic lineages. The origin of the selection forces that regulate the *VH1* polymorphism is unknown. However, the requirement for exogenous factors, such as the intestinal microbiota, for diversification of the primary antibody repertoire in rabbits suggests that this mechanism might impose constraints on the evolution of the *VH1* alleles.

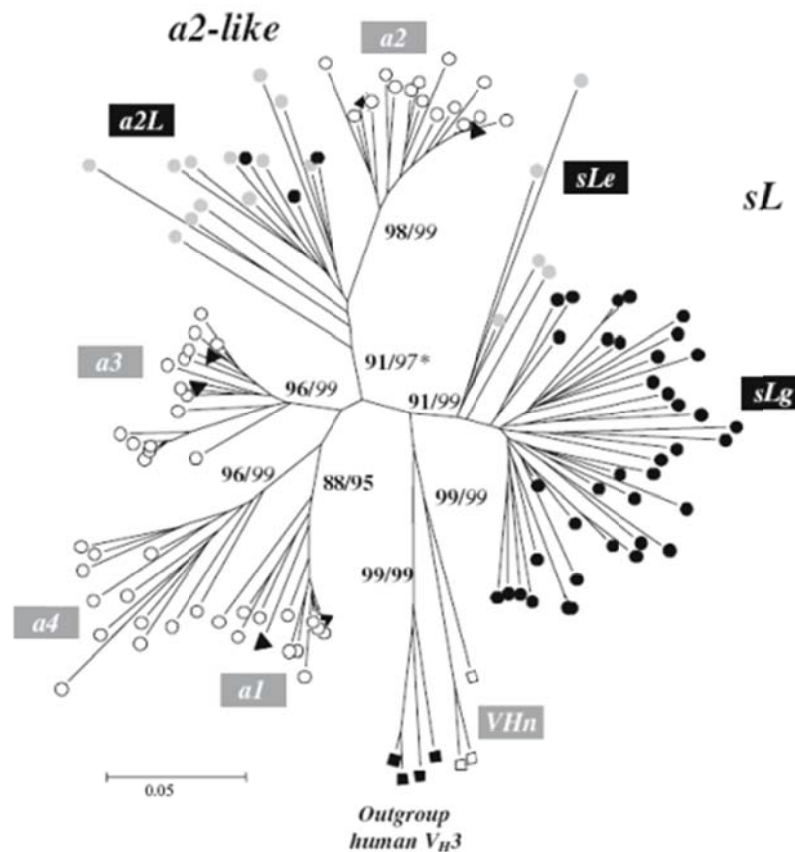


Fig. 5 - Neighbour-joining tree for *Oryctolagus cuniculus*, *Lepus europaeus* (Le), and *Lepus granatensis* (Lg) *VH* genes, showing the transspecies polymorphism of *VHa2*. *VH* genes from human class 3 (*VH3*) determine the root of the tree. Numbers represent the bootstrap values (BP, in bold), and the confidence probability values (CP, in italics). Expressed *VH* genes (cDNA) are labeled by black (Le), gray (Lg), and white (*O. cuniculus*) circles. Germline *VHa* gene sequences (*VH1-VH4*) of rabbit allotypes *a1*, *a2*, and *a3* are marked by black triangles. White squares indicate *VHn* germline genes. Black squares indicate human *VH3* germline genes. "With kind permission from Springer Science+Business Media: Immunogenetics, The evolution of the immunoglobulin heavy chain variable region (IgVH) in leporids: an unusual case of transspecies polymorphism., 57(11), 2005, 874, Esteves PJ, Lanning D, Ferrand N, Knight KL, Zhai SK, van der Loo W, Figure 1."

Diversification of the primary Ab repertoire in the rabbit

Among vertebrate species, two different strategies to generate the primary Ab repertoire have been adopted. Human and mouse preferentially use combinatorial rearrangements of a large number of V, D, and J gene segments. In contrast, several species possess or use only a limited number of germline V segments. Birds, together with several mammalian species (rabbit, sheep, pig and bovine), use strategies of primary repertoire development that overcome this limitation. In these species, the primary antibody repertoire is diversified post-rearrangement by one or both of two mechanisms: somatic gene conversion and somatic hypermutation.

In chicken, somatic gene conversion is the major mechanism of diversification of immunoglobulin heavy and light chain variable regions. A set of pseudogenes is used as donors and the unique rearranged V(D)J gene acts as acceptor (Reynaud et al. 1985; 1989). In sheep, B cell diversification occurs in early development in the ileal Peyer's patches and results from somatic hypermutation rather than somatic gene conversion (Reynaud et al. 1991; 1995). Both processes are used to diversify the antibody repertoire in cattle (Parng et al. 1996; Lucier et al. 1998).

Becker and Knight (1990) showed that the rearrangement process in the rabbit involves a single preferentially rearranged *VH* gene that is diversified mainly through somatic gene conversion, using other upstream *VH* genes as donors (Becker and Knight, 1990; reviewed in Knight and Crane, 1994). Somatic hypermutation also takes place, distributing point mutations throughout the entire *VDJ* gene (Short et al. 1991; Weinstein et al. 1994, Sehgal et al. 2000). In rabbits, B cells generated in the fetal liver and bone marrow subsequently migrate to the appendix and other gut-associated lymphoid tissues (GALT), where they proliferate and are intensively diversified. The *VH* genes upstream of *VH1* contribute to the expressed *VH1* diversification by somatic gene conversion. They evolved to some extent in concert, which resulted in the presence of the allotypic motifs on other *VH* gene segments of the Ig heavy chain locus (haplotype polymorphism).

In contrast to species such as chickens (e.g. Reynaud et al. 1989), sheep (Reynaud et al. 1995) and cattle (e.g. Lucier et al. 1998), the diversification of the primary antibody repertoire in rabbits is not developmentally programmed. It has been shown that the peripheral Ab repertoire is diversified between 4 weeks and 2 months after birth (Cooper et al. 1968; Weinstein et al. 1994). Surgical removal of the appendix and sacculus rotundus on the day of birth and the Peyer's patches at three weeks of

age (Cooper et al. 1968) more recently showed that GALT is essential for the generation of the primary antibody repertoire (Vadhy et al. 1998). It has also been shown that exogenous factors, such as the intestinal microbiota, are required for Ab repertoire diversification in the rabbit (Lanning et al. 2000a, 2000b; Sehgal et al. 2002). The rabbit is a species with altricial young (naked and blind at birth). Other lagomorph species, such as hares (genus *Lepus*) have precocial young. The unique situation of post-natal diversification in rabbit could be related to the shortening of gestation time (four weeks in rabbit and six weeks in hares). Considering the differences in reproductive biology between the genera *Oryctolagus* and *Lepus*, it would be interesting to investigate whether in *Lepus* the Ab repertoire is diversified neonatally or during foetal life.

Evolution of *VH* gene families

The study of *VH* genes from several vertebrate species showed that *VH* genes from the same species could belong to different groups. In tetrapods, the *IGHV* sequences have been classified into three clans, clans I, II and III (Schroeder et al. 1990; Kirkham et al. 1992; Das et al. 2008). The human and mouse possess *VH* genes that are very diversified and may cluster with either clans I, II or III (polyphyletic), whereas *VH* genes of chicken, rabbit, horse, and artiodactyls (cattle, sheep, and swine) are monophyletic. Chicken, rabbit and swine *VH* genes each form distinct clusters within group III, whereas those of horse, cattle and sheep each form distinct clusters within group II. Since *VH* genes of different artiodactyl species cluster within different groups (II and III), their common ancestor must have possessed *VH* genes from both groups (Sitnikova and Su, 1998). Thus, the presence of *IGHV* genes of only one clan in several tetrapod species reflects an evolutionary loss of *IGHV* genes.

Monofunctional multigene families are generally believed to undergo processes of genetic exchange, like gene conversion and unequal crossover, which homogenize the DNA sequences (Smith et al. 1971; Smith, 1974; Zimmer et al. 1980). This has been called “concerted” evolution. The concerted evolution model has been invoked to explain the evolution of *VH* genes (Hood et al. 1975; Ohta, 1983). However, the studies of *VH* gene families in human and mouse (Gojobori and Nei, 1984; Tutter and Riblet, 1989), and studies on much longer evolutionary time scales, showed that the pattern of *VH* gene evolution could be better explained by the so-called “birth-and-death” model of evolution (Ota and Nei, 1994; Nei et al. 1997; Sitnikova and Su, 1998; Das et al.

2008), in which the number of genes in a family (or “library”) are allowed to expand and contract. This model is similar to the “accordion model” of MHC evolution proposed by Klein et al. (1993) and postulates that, depending upon the need to protect the host from ever-changing groups of parasites, some *VH* gene libraries are duplicated and can diverge functionally, while others become pseudogenes and/or are deleted from the genome. The end result of this process is a mixture of divergent and highly homologous groups of genes, and the maintenance of a substantial number of pseudogenes (Ota and Nei, 1994; Nei et al. 1997). According to this model, the tetrapod ancestor possessed three *VH* gene groups (I, II, and III) and library contraction events occurred in several lineages independently. In species that have inherited only one group of *VH* genes, showing low levels of *IGHV* gene sequence variation, antibody diversification is mainly due to somatic hypermutation and/or somatic gene conversion. It appears that *VH* gene library contraction is associated with the development of a specific organ for extensive somatic diversification of the Ab repertoire (bursa of Fabricius in chicken; ileal Peyer’s patches in sheep, cattle and probably swine and horse; appendix in rabbit).

Genome Sequencing and future prospects

Whole genome sequencing (WGS) at ~7x coverage and assembly of a high quality draft rabbit genome sequence was recently completed at the Broad Institute, Boston (OryCun2.0). The ENCODE Project, with ~ 1% of rabbit genomic sequence from a different, normal NZW animal, includes some genes of immunological interest but none of the regions with immunoglobulin heavy or light chain loci. The Thorbecke rabbit strain chosen by Broad for WGS was partially inbred. Although the rabbits accepted skin grafts, they were still segregating for the *VH*a allotypes a1 and a2. The rabbit chosen for sequencing was heterozygous at the immunoglobulin heavy chain locus. The *IGH* locus is not assigned to a chromosome in the Broad assembly. The *IGHV* allele encoding a1 (*VH*1a1) is located in unplaced scaffold chrUn0742 and the *IGHV* allele encoding a2 (*VH*1a2) is located in unplaced scaffold chrUn0439. Unfortunately, there are missing traces or traces of low quality between the telomere and the *IGH*. Until new sequences are obtained, the current available traces from Broad OryCun2 may not allow further assembly. There are ~75 additional unplaced scaffolds that appear to contain *VH* genes.

Little is known about the evolution of the *IGHV* genes in the Lagomorph group. Thus far, only the rabbit *O. cuniculus* has been extensively studied and some sequences of *IGHV* genes have been obtained for two *Lepus* species, *L. europaeus* and *L. granatensis*. The study of these species has yielded some interesting results and led to new insights on the evolution of *IGHV* genes in the Lagomorph group. Also, the few data obtained so far for the Lagomorph *IGHG* hinge region and CH2 domain has shown some particular patterns of genetic diversity possibly linked to resistance against pathogens (Esteves et al. 2002a; Esteves et al. 2002b; Esteves et al. 2006). Extending our knowledge about both the *IGHV* genes and *IGH* constant regions to other Lagomorph species would certainly contribute further to our understanding of the evolution of *IGHV* genes and the roles that genetic diversity of immunoglobulins may play in host resistance to pathogens.

Acknowledgements

This work was supported by grants of Foundation for Science and Technology-Portugal to A.P. and P.J.E. (SFRH/BD/71252/2010 and SFRH/BPD/27021/2006, respectively, and the intramural program of the NIAID, NIH (R.G.M.) .

References

- Allegrucci M, Newman BA, Young-Cooper GO, Alexander CB, Meier D, Kelus AS, Mage RG (1990) Altered phenotypic expression of immunoglobulin heavy-chain variable-region (VH) genes in Alicia rabbits probably reflects a small deletion in the VH genes closest to the joining region. *Proc Natl Acad Sci U S A* 87 (14):5444-5448
- Alves, PC, Hackländer, K (2008) Lagomorph species: Geographical distribution and conservation status. In: Alves PC, Ferrand N, Hackländer K (eds) *Lagomorph Biology*. Springer Berlin Heidelberg, pp 395-405.
- Becker RS, Knight KL (1990) Somatic diversification of immunoglobulin heavy chain VDJ genes: evidence for somatic gene conversion in rabbits. *Cell* 63 (5):987-997.
- Biju-Duval C, Ennafaa H, Dennebouy N, Monnerot M, Mignotte F, Soriguer RC, Gaaïed AE, Hili AE, Mounolou J-C (1991) Mitochondrial DNA evolution in lagomorphs: Origin of systematic heteroplasmy and organization of diversity in European rabbits. *Journal of Molecular Evolution* 33 (1):92-102.

- Bouton C, van der Loo W (1997) The trans-species nature of rabbit b locus polymorphism is supported by studies on the snow-shoe hare. *Immunogenetics* 45 (6):444-446.
- Brezin C, Cazenave PA (1975) Cross reaction between allotype Aa1 of rabbit immunoglobulins and antibodies derived against allotype Aa3: participation of variants of Aa1 specificity in this cross reaction. *Immunochemistry*. 12(3):241-247.
- Cabrera (1914) *Fauna Ibérica*. Museo Nacional de Ciencias Naturales de Madrid, Madrid
- Carneiro M, Afonso S, Geraldès A, Garreau H, Bolet G, Boucher S, Tircazes A, Queney G, Nachman MW, Ferrand N (2011) The Genetic Structure of Domestic Rabbits. *Molecular Biology and Evolution*. (Epub ahead of print) doi:10.1093/molbev/msr003
- Carneiro M, Ferrand N, Nachman MW (2009) Recombination and Speciation: Loci Near Centromeres Are More Differentiated Than Loci Near Telomeres Between Subspecies of the European Rabbit (*Oryctolagus cuniculus*). *Genetics* 181 (2):593-606.
- Cazenave PA, Bennamar A, Sogn JA, Kindt TJ (1987) Immunoglobulin genes in feral population. In: Dubiski S (ed) *The rabbit in contemporary immunological research*. Longman Scientific & Technical, p 148.
- Cazenave PA, Brézin C, Puget A, Mandy W. (1977) Phylogeny of the e allotypic system of rabbit immunoglobulins: study of determinants recognized on ochotona IgG by anti-e15 antisera. *Ann Immunol*. 128(1-2):323-327.
- Chapman J, Flux J (1990) Introduction and overview of the Lagomorphs. In: Chapman J, Flux J (eds) *Rabbits, Hares and Pikas, Status, survey and conservation action plan* UICN/SSC Lagomorph Specialist Group, Switzerland, Gland, pp 1-6
- Chen HT, Alexander CB, Young-Cooper GO, Mage RG (1993) VH gene expression and regulation in the mutant Alicia rabbit. Rescue of VH α 2 allotype expression. *J Immunol* 150 (7):2783-2793
- Cooper MD, Perey DY, Gabrielsen AE, Sutherland DE, McKneally MF, Good RA (1968) Production of an antibody deficiency syndrome in rabbits by neonatal removal of organized intestinal lymphoid tissues. *Int Arch Allergy Appl Immunol* 33 (1):65-88
- Corbet GB (1994) Taxonomy and origins. In: Thompson HV, King CM (eds) *The european rabbit*. Oxford Science Publications, Oxford, pp 1-6
- Corbet GB, Hill JE (1980) *A World list of mammalian species*. Oxford University Press, Oxford

- Currier SJ, Gallarda JL, Knight KL (1988) Partial molecular genetic map of the rabbit VH chromosomal region. *J Immunol* 140 (5):1651-1659
- Das S, Nozawa M, Klein J, Nei M (2008) Evolutionary dynamics of the immunoglobulin heavy chain variable region genes in vertebrates. *Immunogenetics* 60 (1):47-55.
- Davie JM, Paul WE, Mage RG, Goldman MB (1971) Membrane-associated immunoglobulin of rabbit peripheral blood lymphocytes: allelic exclusion at the b locus. *Proc Natl Acad Sci U S A* 68 (2): 430-434
- De Poorter M (1984) An experimental test of predictions from different hypothesis of self regulation in the snowshoe hare (*Lepus americanus* Erxleben, 1777). Vrije University, Brussels
- DiPietro LA, Short JA, Zhai SK, Kelus AS, Meier D, Knight KL (1990) Limited number of immunoglobulin VH regions expressed in the mutant rabbit "Alicia". *Eur J Immunol* 20 (6):1401-1404.
- Dray S, Young GO, Nisonoff A (1963) Distribution of Allotypic Specificities among Rabbit Gamma-Globulin Molecules Genetically Defined at Two Loci. *Nature* 199:52-55
- Dreyer WJ, Bennett JC (1965) The molecular basis of antibody formation: a paradox. *Proc Natl Acad Sci U S A* 54 (3):864-869
- Dubiski S (1969) Immunochemistry and genetics of a "new" allotypic specificity Ae14 of rabbit gamma-G immunoglobulins: recombination in somatic cells. *J Immunol* 103 (1):120-128
- Dubiski S (1987). Foreword. In: Dubiski S (ed) *The rabbit in contemporary immunological research*. Longman Scientific & Technical, John Wiley, pp. pp ix-xii
- Dubiski S, Dudziak Z, Skalba D (1959) Serum groups in rabbits. *Immunology* 2 (1):84-92
- Esteves PJ (2003) Molecular and population genetic analysis of polymorphism at the antibody loci IgGCH2 and IgVH in lagomorphs Faculdade de Ciências Universidade do Porto, Porto
- Esteves PJ, Alves PC, Ferrand N, van der Loo W (2002a) Hotspot variation at the CH2-CH3 interface of leporid IgG antibodies (*Oryctolagus*, *Sylvilagus* and *Lepus*). *Eur J Immunogenet* 29 (6):529-535.
- Esteves PJ, Alves PC, Ferrand N, van der Loo W (2002b) Restriction fragment alleles of the rabbit IGHG genes with reference to the rabbit IGHGCH2 or e locus polymorphism. *Anim Genet* 33 (4):309-311.

- Esteves PJ, Carmo C, Godinho R, van der Loo W (2006) Genetic diversity at the hinge region of the unique immunoglobulin heavy gamma (IGHG) gene in leporids (*Oryctolagus*, *Sylvilagus* and *Lepus*). *Int J Immunogenet* 33 (3):171-177.
- Esteves PJ, Lanning D, Ferrand N, Knight KL, Zhai SK, van der Loo W (2004) Allelic variation at the VHa locus in natural populations of rabbit (*Oryctolagus cuniculus*, L.). *J Immunol* 172 (2):1044-1053
- Esteves PJ, Lanning D, Ferrand N, Knight KL, Zhai SK, van der Loo W (2005) The evolution of the immunoglobulin heavy chain variable region (IgVH) in leporids: an unusual case of transspecies polymorphism. *Immunogenetics* 57 (11):874-882.
- Ferrand N, Branco M (2007) The evolutionary history of the european rabbit (*Oryctolagus cuniculus*): major patterns of population differentiation and geographic expansion inferred from protein polymorphism. In: Weiss S, Ferrand N (eds) *Phylogeography of Southern Europe Refugia*. Springer, pp 207-235
- Flux JEC, Angermann R (1990) The hares and jackrabbits. In: Chapman JA, Flux JEC (eds) *Rabbits, Hares and Pikas: Status Conservation Action Plan*. International Union for Conservation of Nature and Natural Resources, Gland, Switzerland, pp 61-94
- Gallarda JL, Gleason KS, Knight KL (1985) Organization of rabbit immunoglobulin genes. I. Structure and multiplicity of germ-line VH genes. *J Immunol* 135 (6):4222-4228
- Gojobori T, Nei M (1984) Concerted evolution of the immunoglobulin VH gene family. *Mol Biol Evol* 1 (2):195-212
- Halanych KM, Robinson TJ (1999) Multiple substitutions affect the phylogenetic utility of cytochrome b and 12S rDNA data: examining a rapid radiation in leporid (Lagomorpha) evolution. *J Mol Evol* 48 (3):369-379
- Hewitt GM (1996) Some genetic consequences of ice ages, and their role, in divergence and speciation. *Biological Journal of the Linnean Society* 58 (3):247-276.
- Hood L, Campbell JH, Elgin SC (1975) The organization, expression, and evolution of antibody genes and other multigene families. *Annu Rev Genet* 9:305-353.
- Horng WJ, Knight KL, Dray S (1976) Heavy chain variable region allotypic sub-specificities of rabbit immunoglobulins. I. Identification of three subpopulations of a1 IgG molecules. *J Immunol*. 116(1):117-125.
- Horng WJ, Papagiannes E, Dray S, Rodkey LS (1980) Expression of cross-reacting determinants of the immunoglobulin heavy chain variable region a3 allotype in *Oryctolagus* and *Lepus*. *Mol Immunol* 17 (1):111-117

- Kearney JF, Cooper MD, Lawton AR (1976) B cell differentiation induced by lipopolysaccharide. IV. Development of immunoglobulin class restriction in precursors of IgG-synthesizing cells. *J Immunol.* 117(5 Pt 1):1567-72.
- Kelus AS, Gell PG (1967) Immunoglobulin allotypes of experimental animals. *Prog Allergy* 11:141-184
- Kelus AS, Steinberg CM (1991) Is there a high rate of mitotic recombination between the loci encoding immunoglobulin VH and CH regions in gonial cells? *Immunogenetics* 33 (4):255-259
- Kelus AS, Weiss S (1986) Mutation affecting the expression of immunoglobulin variable regions in the rabbit. *Proc Natl Acad Sci U S A* 83 (13):4883-4886
- Kim BS, Dray S (1972) Identification and genetic control of allotypic specificities on two variable region subgroups of rabbit immunoglobulin heavy chains. *Eur J Immunol* 2 (6):509-514.
- Kim BS, Dray S (1973) Expression of the a, x, and y variable region genes of heavy chains among IgG, IgM, and IgA molecules of normal and a locus allotype-suppressed rabbits. *J Immunol* 111 (3):750-760
- Kindt TJ (1975) Rabbit immunoglobulin allotypes: structure, immunology, and genetics. *Adv Immunol* 21:35-86
- Kirkham PM, Mortari F, Newton JA, Schroeder HW, Jr. (1992) Immunoglobulin VH clan and family identity predicts variable domain structure and may influence antigen binding. *EMBO J* 11 (2):603-609
- Klein J, Ono H, Klein D, O'hUigin C (1993) The accordion model of MHC evolution. *Progress in immunology* 8:137-143
- Knight KL (1992) Restricted VH gene usage and generation of antibody diversity in rabbit. *Annu Rev Immunol* 10:593-616.
- Knight KL, Becker RS (1990) Molecular basis of the allelic inheritance of rabbit immunoglobulin VH allotypes: implications for the generation of antibody diversity. *Cell* 60 (6):963-970.
- Knight KL, Crane MA (1994) Generating the antibody repertoire in rabbit. *Adv Immunol* 56:179-218
- Lanning D, Sethupathi P, Rhee KJ, Zhai SK, Knight KL (2000a) Intestinal microflora and diversification of the rabbit antibody repertoire. *J Immunol* 165 (4):2012-2019.
- Lanning D, Zhu X, Zhai SK, Knight KL (2000b) Development of the antibody repertoire in rabbit: gut-associated lymphoid tissue, microbes, and selection. *Immunol Rev* 175:214-228

- Loor F, Kelus AS (1978) Allelic exclusion in the B lineage cells of the rabbit. *Eur J Immunol* 8 (5):315-324.
- Lopez-Martinez N (1989) Revisión sistemática y biostratigráfica de los Lagomorpha (Mammalia) del Terciario y Cuaternario de España vol 3. Memorias del Museo Paleontológico de la Universidad de Zaragoza. Diputación General de Aragón,
- Lopez-Martinez N (2008) The Lagomorph Fossil Record and the Origin of the European Rabbit. In: Alves PC, Ferrand N, Hackländer K (eds) *Lagomorph Biology*. Springer Berlin Heidelberg, pp 27-46.
- Lucier MR, Thompson RE, Waire J, Lin AW, Osborne BA, Goldsby RA (1998) Multiple sites of V lambda diversification in cattle. *J Immunol* 161 (10):5438-5444
- Lumms Z, Cebra JJ, Mage R (1967) Correspondence of the relative cellular distribution and serum concentration of allelic allotypic markers in normal and "allotype-suppressed" heterozygous rabbits. *J Immunol* 99 (4):737-743
- Mage R, Young GO, Dray S (1967) An effect upon the regulation of gene expression: allotype suppression at the a locus in heterozygous offspring of immunized rabbits. *J Immunol* 98 (3):502-509
- Mage RG (1979) A new look at the biological and genetic significance of rabbit heavy chain allotypes. *Ann Immunol (Paris)* 130 (2):105-114
- Mage RG, Bernstein KE, McCartney-Francis N, Alexander CB, Young-Cooper GO, Padlan EA, Cohen GH (1984) The structural and genetic basis for expression of normal and latent VHa allotypes of the rabbit. *Mol Immunol* 21 (11):1067-1081
- Mage RG, Lanning D, Knight KL (2006) B cell and antibody repertoire development in rabbits: the requirement of gut-associated lymphoid tissues. *Dev Comp Immunol* 30 (1-2):137-153.
- Mage RG, Young-Cooper GO, Alexander C (1971) Genetic control of variable and constant regions of immunoglobulin heavy chains. *Nat New Biol* 230 (10):63-64
- Mage RG, Young-Cooper GO, Rejnek J, Ansari AA, Alexander CB, Appella E, Carta-Sorcini M, Landucci-Tosi S, Tosi RM (1976) Rabbit Immunoglobulin Allotypes: complexities of their genetics, expression, structural basis and evolution. *Cold Spring Harbor Symp. Quant. Biol.* 41:677-686.
- Margolies MN, Cannon LE, Kindt TJ, Fraser B (1977) The structural basis of rabbit VH allotypes: serologic studies on a1 H chains with defined amino acid sequence. *J Immunol* 119 (1):287-294
- Matthee CA, van Vuuren BJ, Bell D, Robinson TJ (2004) A molecular supermatrix of the rabbits and hares (Leporidae) allows for the identification of five intercontinental exchanges during the Miocene. *Syst Biol* 53 (3):433-447

- Nei M, Gu X, Sitnikova T (1997) Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proc Natl Acad Sci U S A* 94 (15):7799-7806
- Ohta T (1983) On the evolution of multigene families. *Theor Popul Biol* 23 (2):216-240.
- Ota T, Nei M (1994) Divergent evolution and evolution by the birth-and-death process in the immunoglobulin VH gene family. *Mol Biol Evol* 11 (3):469-482
- Oudin J (1956) [The allotype of certain blood protein antigens]. *C R Hebd Seances Acad Sci* 242 (21):2606-2608
- Oudin J (1960) Allotypy of rabbit serum proteins. I. Immuno-chemical analysis leading to the individualization of seven main allotypes. *J Exp Med* 112:107-124
- Pages MV (1980) Essai de reconstitution de l'histoire du lapin de garenne en Europe. *Bull Mens Off Natl Chasse Sp. Scien. Techn*:13-21
- Parng CL, Hansal S, Goldsby RA, Osborne BA (1996) Gene conversion contributes to Ig light chain diversity in cattle. *J Immunol* 157 (12):5478-5486
- Pospisil R, Young-Cooper GO, Mage RG (1995) Preferential expansion and survival of B lymphocytes based on VH framework 1 and framework 3 expression: "positive" selection in appendix of normal and VH-mutant rabbits. *Proc Natl Acad Sci U S A* 92 (15):6961-6965
- Queney G, Ferrand N, Weiss S, Mougel F, Monnerot M (2001) Stationary Distributions of Microsatellite Loci Between Divergent Population Groups of the European Rabbit (*Oryctolagus cuniculus*). *Molecular Biology and Evolution* 18 (12):2169-2178
- Reynaud CA, Anquez V, Dahan A, Weill JC (1985) A single rearrangement event generates most of the chicken immunoglobulin light chain diversity. *Cell* 40(2):283-291.
- Reynaud CA, Dahan A, Anquez V, Weill JC (1989) Somatic hyperconversion diversifies the single Vh gene of the chicken with a high incidence in the D region. *Cell* 59 (1):171-183.
- Reynaud CA, Garcia C, Hein WR, Weill JC (1995) Hypermutation generating the sheep immunoglobulin repertoire is an antigen-independent process. *Cell* 80 (1):115-125.
- Reynaud CA, Mackay CR, Muller RG, Weill JC (1991) Somatic generation of diversity in a mammalian primary lymphoid organ: the sheep ileal Peyer's patches. *Cell* 64 (5):995-1005.
- Ros F, Puels J, Reichenberger N, van Schooten W, Buelow R, Platzer J (2004) Sequence analysis of 0.5 Mb of the rabbit germline immunoglobulin heavy chain locus. *Gene* 330:49-59.

- Roux KH (1981) A fourth heavy chain variable region subgroup, w, with 2 variants defined by an induced auto-antiserum in the rabbit. *J Immunol* 127 (2):626-632
- Roux KH, Greenberg AS, Greene L, Strelets L, Avila D, McKinney EC, Flajnik MF (1998) Structural analysis of the nurse shark (new) antigen receptor (NAR): molecular convergence of NAR and unusual mammalian immunoglobulins. *Proc Natl Acad Sci U S A*. 95(20):11804-9.
- Schmale J, Costea N, Dray S, Heller P, Yakulis V (1969) Allelic exclusion of light chain allotypes in rabbit IgM cold agglutinins. *Proc Soc Exp Biol Med* 130 (1):48-50
- Schroeder HW, Jr., Hillson JL, Perlmutter RM (1990) Structure and evolution of mammalian VH families. *Int Immunol* 2 (1):41-50
- Sehgal D, Mage RG, Schiaffella E (1998) VH mutant rabbits lacking the VH1a2 gene develop $\alpha 2+$ B cells in the appendix by gene conversion-like alteration of a rearranged VH4 gene. *J Immunol* 160 (3):1246-1255
- Sehgal D, Obiakor H, Mage RG (2002) Distinct clonal Ig diversification patterns in young appendix compared to antigen-specific splenic clones. *J Immunol* 168 (11):5424-5433
- Sehgal D, Schiaffella E, Anderson AO, Mage RG (2000) Generation of heterogeneous rabbit anti-DNP antibodies by gene conversion and hypermutation of rearranged VL and VH genes during clonal expansion of B cells in splenic germinal centers. *Eur J Immunol* 30 (12):3634-3644.
- Severinson E, Bergstedt-Lindqvist S, van der Loo W, Fernandez C (1982) Characterization of the IgG response induced by polyclonal B cell activators. *Immunol Rev* 67:73-85
- Short JA, Sethupathi P, Zhai SK, Knight KL (1991) VDJ genes in VHa2 allotype-suppressed rabbits. Limited germline VH gene usage and accumulation of somatic mutations in D regions. *J Immunol* 147 (11):4014-4018
- Sitnikova T, Su C (1998) Coevolution of immunoglobulin heavy- and light-chain variable-region gene families. *Mol Biol Evol* 15 (6):617-625
- Smith GP (1974) Unequal crossover and the evolution of multigene families. *Cold Spring Harb Symp Quant Biol* 38:507-513
- Smith GP, Hood L, Fitch WM (1971) Antibody diversity. *Annu Rev Biochem* 40:969-1012.
- Su C, Nei M (1999) Fifty-million-year-old polymorphism at an immunoglobulin variable region gene locus in the rabbit evolutionary lineage. *Proc Natl Acad Sci U S A* 96 (17):9710-9715

- Surridge AK, van der Loo W, Abrantes J, Carneiro M, Hewitt GM, Esteves PJ (2008) Diversity and evolutionary history of the MHC DQA gene in leporids. *Immunogenetics* 60 (9):515-525.
- Taberlet P, Fumagalli L, Wust-Saucy A-G, Cosson J-F (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology* 7 (4):453-464.
- Teherani J, Capra JD, Aggarwal S, Mandy WJ (1979) Amino acid sequence analysis of group e allotype-related peptides derived from lagomorph IgG. *Eur J Immunol* 9 (9):690-695.
- Teherani J, Capra JD, Mandy WJ (1982) Amino acid sequence of the CH2 domain from various lagomorph IgGs. *Mol Immunol.* 19(7):841-846.
- Teherani J, Mandy WJ (1976a) Constant region IgG allotypes in cottontail rabbits: group E allelic polymorphism. *Immunochemistry* 13 (3):221-227
- Teherani J, Mandy WJ (1976b) Constant region IgG allotypes in hares: group e allelic polymorphism. *Immunochemistry* 13 (5):401-406
- Todd CW (1963) Allotypy in rabbit 19S protein. *Biochem Biophys Res Commun* 11:170-175
- Tonnelle C, Cazenave PA, Brezin C, Moinier D, Fougereau M (1983) Structural correlates to the rabbit immunoglobulin heavy chain a100 allotype. *Mol Immunol* 20 (7):753-761
- Tutter A, Riblet R (1989) Evolution of the immunoglobulin heavy chain variable region (Igh-V) locus in the genus *Mus*. *Immunogenetics* 30 (5):315-329
- Vajdy M, Sethupathi P, Knight KL (1998) Dependence of antibody somatic diversification on gut-associated lymphoid tissue in rabbits. *J Immunol* 160 (6):2725-2729
- van der loo W (1987) "Population genetical studies on the adaptive significance of the immunoglobulin allotypes in wild rabbit" In: Dubiski S (Ed) "The Rabbit in Contemporary Immunological Research", Longman Scientific & Technical, John Wiley, pp. 101-126
- van der Loo W, (1993) Variance analysis of immunoglobulin alleles in natural populations of rabbit (*Oryctolagus cuniculus*): the extensive interallelic divergence at the b locus could be the outcome of overdominance-type selection. *Genetics* 135(1):171-187.
- van der Loo W, Arthur CP, Richardson BJ, Wallage-Drees M, Hamers R (1987) Nonrandom allele associations between unlinked protein loci: are the polymorphisms of the immunoglobulin constant regions adaptive? *Proc Natl Acad Sci U S A* 84 (9):3075-3079

- van der Loo W, De Baetselier P, Hamers-Casterman C, Hamers R (1977) Evidence for quasi-silent germline genes coding for phylogenetically ancient determinants of the rabbit a locus allotypes. *Eur J Immunol* 7 (1):15-22.
- van der Loo W, Ferrand N, Soriguer RC. (1991) Estimation of gene diversity at the b locus of the constant region of the immunoglobulin light chain in natural populations of European rabbit (*Oryctolagus cuniculus*) in Portugal, Andalusia and on the Azorean Islands. *Genetics* 127(4):789-799.
- van der Loo W, Gronowicz ES, Strober S, Herzenberg LA. (1979) Cell differentiation in the presence of cytochalasin B: studies on the "switch" to IgG secretion after polyclonal B cell activation. *J Immunol.* 122(4):1203-8.
- van der Loo W, Mougél F, Sanchez MS, Bouton C, Castien E, Fonseca A, Ferrand N, Soriguer R, Monnerot M (1999) Cytonuclear disequilibria in wild populations of rabbit (*Oryctolagus cuniculus* L.) suggest unequal allele turnover rates at the b locus (IGKC1). *Immunogenetics* 49 (7-8):629-643.
- Weinstein PD, Anderson AO, Mage RG (1994) Rabbit IgH sequences in appendix germinal centers: VH diversification by gene conversion-like and hypermutation mechanisms. *Immunity* 1 (8):647-659.
- Wolf B, Janeway CA, Jr., Coombs RR, Catty D, Gell PG, Kelus AS (1971) Immunoglobulin determinants on the lymphocytes of normal rabbits. 3. As4 and As6 determinants on individual lymphocytes and the concept of allelic exclusion. *Immunology* 20 (6):931-944
- Zhu X, Boonthum A, Zhai SK, Knight KL (1999) B lymphocyte selection and age-related changes in VH gene usage in mutant Alicia rabbits. *J Immunol* 163 (6):3313-3320.
- Zimmer EA, Martin SL, Beverley SM, Kan YW, Wilson AC (1980) Rapid duplication and loss of genes coding for the alpha chains of hemoglobin. *Proc Natl Acad Sci U S A* 77 (4):2158-2162

Paper 2

Sequencing of modern *Lepus* VDJ genes shows that the usage of *VHn* genes has been retained in both *Oryctolagus* and *Lepus* that diverged 12 million years ago.

Ana Pinheiro, Isabel G. Fernández de Mera, Paulo C. Alves, Christian Gortázar, José de la Fuente and Pedro J. Esteves

Immunogenetics, 65(11), 777-84. doi: 10.1007/s00251-013-0728-3.

Abstract

Amongst mammals the European rabbit (*Oryctolagus cuniculus*) has a unique mechanism of generating the primary antibody repertoire. Despite having over 200 *VH* genes, the *VH1* gene, the most D-proximal *VH* gene, is used in 80-90% of *VDJ* rearrangements, whilst the remaining 10-20% are encoded by the *VHn* genes, that map at least 100 Kb upstream of *VH1*. The maintenance of the *VHn* genes usage in low frequency in *VDJ* rearrangements has been suggested to represent a relic of an ancestral immunologic response to pathogens. To address this question, we sequenced *VDJ* genes for another Leporid, genus *Lepus*, which separated from European rabbit 12 million years ago. Approximately 25 *VDJ* gene sequences were obtained for each one of three *Lepus europaeus* individuals. We found that *Lepus* also uses the *VHn* genes in 5-10% of its *VDJ* rearrangements. Our results show that the

VHn genes are a conserved ancestral polymorphism that has been maintained in the Leporids genome and is being used for the generation of *VDJ* rearrangements by both modern *Lepus* and *Oryctolagus*.

Keywords

IGHV, Leporids, antibody diversification, Vh allotypes, evolution

Introduction

In the generation of the primary antibody repertoire the European rabbit (*Oryctolagus cuniculus*) uses a limited number of germline V segments. Despite having more than 200 *VH* genes (e.g.: Ros et al. 2004), about 80% to 90% of circulating Ig molecules are derived from the *VH1* gene, the most D-proximal *VH* (Knight and Becker 1990; Knight 1992), and express the so called *VHa* allotypic markers (e.g., Kindt 1975; Margolies et al. 1977). Early serological studies described three *a* allotypic lineages, the *a1*, *a2* and *a3*, in the domestic rabbit. These are highly divergent ($\pm 20\%$ amino acid sequence differences), the allelic specificities of *a1* and *a2* being correlated with several amino acid differences in framework regions (FR) 1 (positions 5, 8, 11, 13, 14, 17, and 18) and 3 (positions 74, 76, 79, 80, 83, 84, 96, and 97) (Tonnellet et al. 1983; Mage et al. 1984; Knight and Becker 1990). More recently, a fourth allotypic lineage, the *a4*, was described in wild European rabbit Iberian populations. This lineage displays the rabbit *VHa* hallmark residues together with a number of unprecedented amino acid changes in FR2 and FR3, being equally highly divergent (Esteves et al. 2004). The remaining 10–20% of Ig molecules are encoded by the *VHn* genes (Kim and Dray 1973; Roux 1981). These genes, called *VHx*, *VHy*, and *VHz* (Hornig et al. 1976), do not express *VHa* allotype specific determinants and are localized at least 100 Kb upstream of *VH1* (Mage et al. 2006). It has been suggested that the low frequency usage of the *VHn* genes in VDJ rearrangements may represent an evolutionary relic of an ancestral leporid immunologic response to pathogens (Pineiro et al. 2011).

Studies on the *IGHV* locus diversity for other leporids are limited to a few data obtained for some *Lepus* species. Serological analysis conducted for *L. americanus* showed cross-reaction with rabbit anti-*a1*, anti-*a2*, and anti-*a3* antisera and the absence of individuals with no cross-reaction to rabbit antisera (van der Loo 1987). As for *L. europaeus* and *L. granatensis*, a serological analysis of several populations revealed only two phenotypes: partial reaction to anti-*a2* antisera and no reaction to any rabbit antiserum (Esteves 2003). These studies showing serological cross-reactivity of *Lepus* sera with rabbit *VHa*-allotype-specific allo-antisera were the first indications that the different *VH1* polymorphisms could be trans-specific (Su and Nei 1999). It was only with obtaining sequences of rearranged *VH* genes for *Lepus* specimens, some of which expressed *a2*-cross-reacting serum proteins, that this hypothesis was confirmed (Esteves et al. 2005). Within the *Lepus* *VH* genes, two lineages have been observed, the *a2L* and *sL* (Esteves et al. 2005). The *a2L* lineage shows some of the most outstanding characteristics of the rabbit *VHa* genes,

possessing 5 out of 11 amino acid residues that characterize the allotype VHa2 (Mage et al. 1984), and groups with rabbit a2 sequences, showing that it is more closely related to the rabbit VH1-a2 allele than VH1-a2 was to its allelic counterparts, VH1-a1 and VH1-a3. The sL lineage, apparently a *Lepus* specific lineage, groups distinctively apart from all other VH lineages present in rabbit, showing as distinctive characters the ancestral ₇₀SVK₇₂ motif and A₁₀ and K₉₅ residues (Esteves et al. 2005).

In this study we addressed the hypothesis that the infrequent usage of the VHn genes in VDJ rearrangements is an ancient mechanism of the leporids immunologic response to pathogens. If true, VHn genes should also be found in VDJ rearranged genes for other Leporid species in a similarly low frequency. The family Leporidae comprises eleven genera among which are included *Oryctolagus* and *Lepus* (Alves and Hackländer 2008), which diverged 12 My ago (Mathee et al. 2004). The only study conducted so far on Leporids *IGHV* did not encounter expressed *VHn* genes in *Lepus* but it was based on a rather low number of sequences per individual (approx 12 VDJ sequences/individual were obtained; Esteves et al. 2005). Here, we increased the sequencing effort to obtain *circa* 25 VDJ genes per individual to find that *Lepus* too uses the *VHn* genes in a low percentage of its VDJ rearrangements.

Material and methods

Samples, amplification and sequencing of rearranged VDJ genes

Spleen samples were collected from 3 unrelated specimens of farmed *Lepus europaeus* collected from the Spanish region Cantabria. Total RNA was extracted using RNeasy Mini Kit (Qiagen, Hilden, Germany), following first strand cDNA synthesis using the iScript cDNA Kit (Bio-Rad, Hercules, CA, USA). VDJ rearranged genes were PCR amplified using primers VH (5'GGAGACTGGGCTGCGCTGGCTTCTCCTGGT3'; Esteves et al., 2005) and JH2 (5'TGAGGAGACGGTGACCAGGGTGCCCT3') designed in the conserved region of rabbit JH gene segment. PCR amplification conditions were as follows: 4 min at 95 °C followed by 35 cycles at 95 °C (45 s), 62 °C (45 s) and 72 °C (60 s), with a final extension at 72 °C (5 min). PCR products were purified (MinElute PCR Purification Kit, Qiagen, Hilden, Germany) and cloned into the pGEM-T Easy vector system II (Promega, Madison, WI, USA).

Phylogenetic Analyses

To perform phylogenetic analyses available sequences for leporids VDJ genes were taken from Genbank. Sequences used include rabbit germline VH1 and VH4 genes of a1, a2 and a3 allotypes, rabbit cDNA VH gene sequences representative of allotypes a1, a2, a3, a4.1 and a4.2, rabbit germline and cDNA VH gene sequences of VHx, VHy, and VHz, and *Lepus* cDNA VH gene sequences representative of lineages a2L, sLe and sLg. VH genes of the human VH3 family were used as outgroup. These were aligned with sequences obtained in this study using CLUSTAL W (Thompson et al. 1994) as implemented in BioEdit software (Hall, 1999) and the amino acid sequences were inferred using the software BioEdit (Hall, 1999). Accession numbers for all sequences are given in table 1.

Table 1- Vh genes sequences accession numbers for sequences used in the phylogenetic analysis.

Sequence identification	Accession number
<i>rabbit germline sequences</i>	
VH1a1, VH4a1	M93171, M93181
VH1a2, VH4a2	M93172, M93182
VH1a3, VH4a3	M93173, L03846
Vhx, Vhx2	L03846, M19706
Vhy, Vhy2	L03890, L03874
Vhz	AF264469
<i>rabbit cDNA sequences</i>	
a1_Ocun1-Ocun4	AF029933, AF029934, AF029938, AF029940
a2_Ocun1-Ocun4	L03849, L03851, L03853, L03856
a3_Ocun1-Ocun4	AF029923, AF098235, AF264434, AF264435
a4.1_I- IV	AY207979- AY207982
a4.2_I- III	AY208042, AY208047, AY207967
Vhx_Oca, Vhx_Oca2	AY207986, AY208045
Vhy_Oca	AY208006
<i>hare cDNA sequences</i>	
sLe1_397, sLe1_400	AY288451, AY288453
sLe2_408, sLe2_412	AY288459, AY288462
sLg15_640-642, sLg15_644	AY288464- AY288467
a2_Le1.396, a2_Le1.401	AY288450, AY288454
a2_Lg19C83, a2_Lg19C86	AY288491, AY288494
<i>human VH3 family sequences</i>	
Hsap1- Hsap4	M99666, M99672, M99679, M99682
<i>Current study</i>	
Le5.1-5.5, Le5.7-5.9, Le5.12-5.26	KF460033-KF460055
Le6.1, Le6.3-6.5, Le6.7-6.20, Le6.22, Le6.24	KF460056-KF460075
Le8.1-8.3, Le8.5-8.9, Le8.11-8.13, Le8.16-8.20, Le8.23-8.27	KF460076-KF460096

The computer program jModelTest v2.1.1 (Darriba et al 2012) was used to assess the fit of our dataset to 88 models of sequence evolution considering the

corrected Akaike information criterion. The phylogenetic relationships were analysed in Maximum Likelihood (ML) and Bayesian inference (BI) frameworks. ML analyses were performed using PhyML v3.0 (Guindon and Gascuel, 2003) specifying the optimal mutation model, though not setting up the model parameters. The support of the resulting nodes was estimated using 500 bootstrap replicates as implemented in PhyML. BI analyses were performed using BEAST v1.7.4 (Drummond and Rambaut, 2007). The BEAUti application, part of BEAST package, was used to create the input file for *BEAST again specifying the optimal mutation model, though not setting up the model parameters. Posterior phylogenies were determined using an uncorrelated lognormal relaxed clock (Drummond et al., 2006) and the Yule tree prior. The prior of the relaxed clock standard deviation was set to an exponential distribution with a mean on 0.5 as recommended by the authors. All remaining priors were set to the defaults. Three independent runs of 3×10^6 generations were performed, sampling trees and parameters every 3×10^3 generations. Convergence was checked using Tracer v1.5 (Rambaut and Drummond 2007) and summary trees were generated with TreeAnnotator v1.5.4, part of the BEAST package.

Genetic distances between VH genes among species were calculated using the “compute net average distance between groups” option of MEGA5.0 software (Tamura et al. 2011). This option corrects for variance due to differences within groups, as expected in VH rearranged genes which are somatically diversified.

Results

A total of 64 *VDJ* gene sequences were obtained for the three studied *L. europaeus* individuals and their aminoacid sequences inferred. The phylogenetic relationships of the obtained sequences to known leporids VH sequences were studied using ML and BI methodologies specifying as optimal mutation model TrNef+G. The phylogenies obtained by both methods consistently yielded the same topology, though, in general, the posterior probabilities values lend greater support to the observed groups. The hare sequences obtained in this study fall into one of four groups (Figure 1): (1) the a2-like group, that includes rabbit (a2) and hare (a2L) previously identified VHa2 sequences (37 bootstrap; 1 posterior probability); (2) the VHy-VHz group, that includes rabbit germline and expressed VHy and VHz sequences (90 bootstrap; 1 posterior probability), (3) the VHx group, a heterogeneous group that includes rabbit

germline and expressed VHx sequences (82 bootstrap; 0.83 posterior probability) , and (4) the sL group, containing only *Lepus* VH sequences (74 bootstrap; 0.99 posterior probability).

Of the 23 and 20 VDJ sequences obtained for specimens Le5 and Le6, respectively, two Le5 sequences (8,7%) were classified as VHx, forming a well supported group with a rabbit VHx sequence (81 bootstrap; 0.97 posterior probability), two Le6 sequences (10%) were classified as VHy-VHz and the remaining as a2-like. For specimen Le8 one sequence (4,7%) was classified as VHy-VHz, seven as a2-like and the remaining 13 sequences as sL.

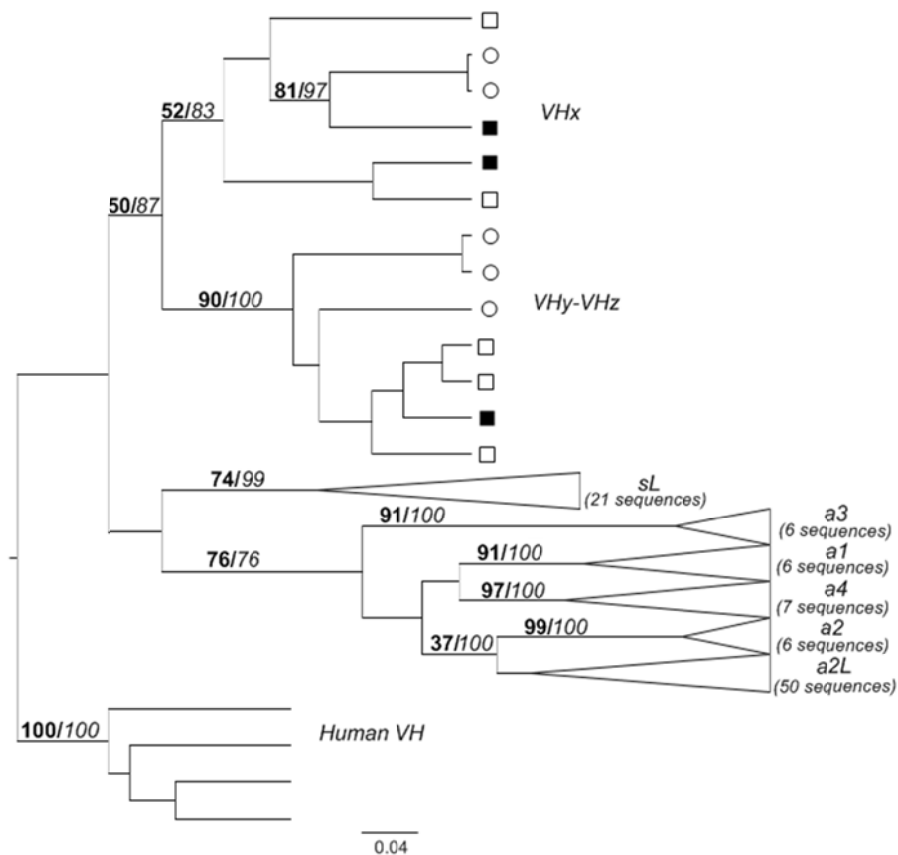


Fig. 1. Phylogenetic tree of European rabbit and hare VH genes. Human VH class 3 genes were used to root the tree. Some groups were collapsed for simplification; the number of sequences in those groups is indicated. ML bootstraps and BI posterior probabilities are depicted in front of each node in bold and italic numbering, respectively. Expressed VHn hare genes obtained in this study are represented by open circles. Rabbit germline *VHn* genes are represented in open squares, Rabbit expressed *VHn* genes are represented in black squares.

Genetic distances between rabbit and hares VH

The genetic distance between rabbit and hares VHx and VHy-VHz genes (0,06 and 0,04, respectively; table 2) is lower than that between VHn genes and all other allelic lineages (0,07-0,21, table 2), being also lower than those observed among the other allelic lineages (0,07-0,15, table 2).

Table 2- Genetic distances between hare and rabbit VH nucleotide sequences

Groups		Hares		Rabbit		sL	a2L	a1	a2	a3	a4
		VHx	VHy-VHz	VHx	VHy-VHz						
Hares	VHx		0.03	0.01	0.02	0.02	0.02	0.02	0.02	0.02	0.02
	VHy-VHz	0.12		0.01	0.01	0.01	0.02	0.02	0.03	0.02	0.02
Rabbit	VHx	0.06	0.05		0.01	0.01	0.02	0.01	0.02	0.01	0.02
	VHy-VHz	0.14	0.04	0.07		0.02	0.02	0.02	0.02	0.02	0.02
	sL	0.11	0.09	0.07	0.12		0.01	0.01	0.02	0.02	0.02
	a2L	0.14	0.12	0.09	0.14	0.10		0.01	0.02	0.02	0.02
	a1	0.13	0.14	0.10	0.16	0.10	0.07		0.02	0.01	0.01
	a2	0.18	0.18	0.15	0.21	0.15	0.07	0.11		0.02	0.02
	a3	0.14	0.12	0.09	0.12	0.10	0.10	0.09	0.14		0.02
	a4	0.17	0.16	0.11	0.18	0.12	0.09	0.08	0.13	0.11	

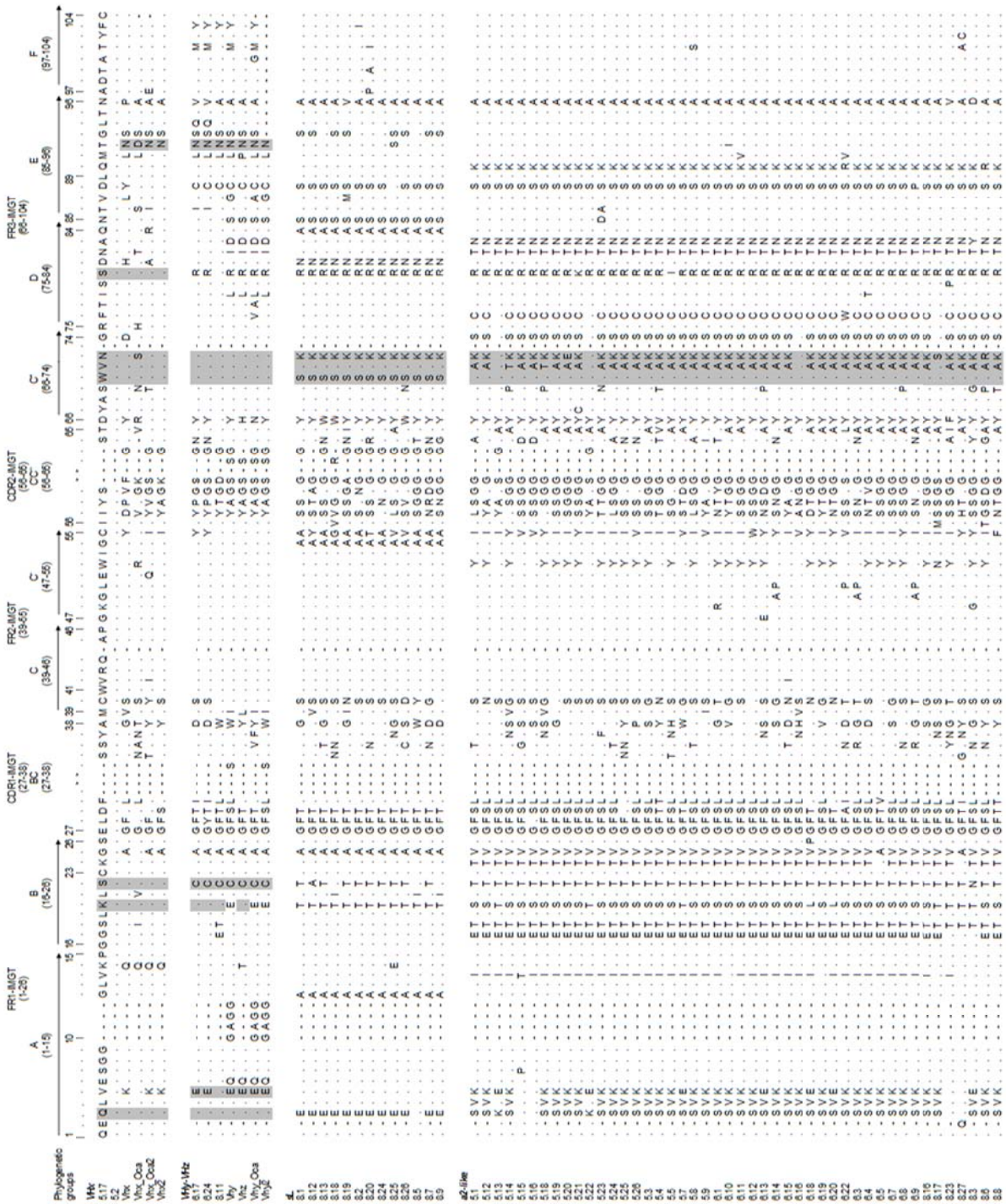
¹Genetic distances between groups of sequences, calculated using the p-distance, are shown below the diagonal. Standard error estimates are shown above the diagonal in italics. All ambiguous positions were removed for each sequence pair.

Characterization of expressed Vh in Lepus

The inferred amino acid sequences were grouped according to phylogenetic clustering and compared to rabbit germline and cDNA VH genes representative of VHn and VH_a allotypes as well as *Lepus* cDNA VH gene sequences representative of lineages a2L and sL (sLe and sLg) (Figure 2; Supplementary Table 1). The rabbit VH_a sequences are characterized by a ₇₀WAK₇₂ motif where ₇₀WVN₇₂ is present in VHn sequences, and hare sL sequences have ₇₀SVK₇₂ (Esteves et al. 2005).

The five sequences obtained in this study that classified as VHn, either as VHx or as VHy-VHz, show the distinctive ₇₀WVN₇₂ motif and the residue Q₃ also characteristic of rabbit VHn peptides, while also showing the K₂₀ featured in rabbit VHx and VHz sequences (Rabbit VHy show E₂₀ instead). Sequences classified as VHy-VHz further show the N₉₂ residue characteristic of rabbit VHn peptides. Additionally, rabbit VHy-VHz peptides show as distinctive the E₅ and C₂₂ residues that were also observed in the majority of hares VHy-VHz sequences. Rabbit VHx peptides share with hare VHx

Fig. 2. Alignment of European rabbit and hare VH protein sequences. Dashes (-) represent alignment gaps. Dots (·) indicate identity with the master sequence except at gap positions. The 'IMGT Protein display for V domain' header is shown (Lefranc et al. 2003). Sequences are grouped into phylogenetic groups according to the tree in Figure 1. Shared characteristic residues of European rabbit and hare VHn genes are highlighted in gray. Also highlighted in gray are the VH_a ₇₀WAK₇₂, VHn ₇₀WVN₇₂ and hares sL sequences ₇₀SVK₇₂ motifs



sequences typical K₂₀, S₂₂ and S₈₀ residues. Hares VHx sequences are further characterized by E₂₇ and contrast with rabbit VHx in residues K₁₄ and ₉₂TG₉₃. Rabbit VHy-VHz sequences uniquely show residues Q₆ and L₇₈ and the motif ₈₂IDQS₈₅, characteristics not found in hares VHy-VHz sequences (which show residues E₆, I₇₈ and the motif ₈₂NAQN₈₅). Thus, these five hare sequences clearly show the rabbit VHn signature, in spite having evolved some particularities.

The sL sequences share the characteristic ₇₀SVK₇₂ motif along with the typical residues E₃, A₁₀, T₂₀ and S/T₂₂. Unlike previously published sL sequences, those obtained in this study show at position 95 T instead of K residues and share the residues ₉₆AA₉₇ with rabbit a2 and VHn sequences.

The majority of a2L sequences show the rabbit VHa ₇₀WAK₇₂ motif, a deletion in position 2 and residues S₃, K₅, T₁₈, S/T₂₀, T₂₂, S₇₄, R₈₀ and N₈₃, as previously described by Esteves and co-workers (2005).

Discussion

The *IGHV* multigene family evolves according to the so called “birth-and-death model, new genes being created by duplication events and divergence of the duplicated genes (Ota and Nei 1994; Das et al. 2008). During this process some genes gain new functions whilst others become pseudogenized or are eliminated from the genome. This process seems to be species-specific, and as a result some species have retained a high level of *IGHV* genes whereas others have a limited *IGHV* repertoire (Das et al. 2008). Human and mice have a large number of *IGHV* genes, relying on the combinatorial joining of V, D and J segments for the generation of a diverse antibody repertoire. On the opposite end, chickens have only one functional *IGHV* and their pre-immune repertoire is generated by the diversification of rearranged *VDJ* genes by somatic hypermutation and/or gene conversion. Most other mammals express a limited number of rearranged *VDJ* genes that are further diversified by somatic hypermutation and/or gene conversion (Flajnick 2002). The rabbit comes as a surprise, for despite the number of functional *IGHV* genes capable of rearranging in the rabbit germline that can be roughly estimated as ≈60 (Ros et al. 2004) it predominantly expresses just one *IGHV*, the most D-proximal *VH1* gene, to generate 80-90% of its antibody repertoire (Knight and Becker 1990; Knight 1992). Furthermore, the remaining 10–20% of Ig molecules are encoded by the VHn genes, VHx, VHy, and VHz (Kim and

Dray 1973; Horng et al. 1976; Roux 1981), that map at least 100 Kb upstream of VH1 (Mage et al. 2006).

A first indication that the usage of the *VHn* genes in VDJ rearrangements could be an alternative mechanism to generate the antibody repertoire comes from the study of the “Alicia” rabbit. In the “Alicia” rabbit genome, a 10 Kb segment containing the VH1 gene was deleted (Knight and Becker 1990; Allegrucci et al. 1990). Instead of the D-proximal functional VHa positive gene present in the Alicia genome, it was found that the serum of young homozygous ali/ali Alicia Rabbits contained mostly Ig resulting from rearranged *VHn* genes (Kelus and Weiss 1986; DiPietro et al. 1990; Chen et al. 1993). In this study, we observed *VH* diversified genes in hares (*Lepus*) that clearly show the characteristics of rabbit *VHn* genes. Phylogenetic inference methods further cluster the hares newly found diversified *VHn* and rabbit *VHn* genes in the same group. These show that the *VHn* lineages are present in the Leporids genome and are being used for the generation of VDJ rearrangements for at least 12 million years, the estimated divergence time between *Oryctolagus* and *Lepus* genera (Mathee et al. 2004). Identically, the a2 allotype lineage is present both in rabbit and hares (Esteves et al. 2005) and serological evidence suggests that the same is true for the a1 and a3 allotype lineages (van der Loo, 1987). On the other hand, considering a neutral substitution rate of 4×10^{-9} for rabbits (Mathee et al. 2004; van der Loo et al. 2009), the genetic distances between rabbit and hares *VHn* would imply a divergence time of $\approx 5\text{-}7$ My (i.e., $0.06/2/4 \times 10^{-9} = 7.5$ My) for these lineages whereas the genetic distances between *VHn* and other allelic lineages would imply a much older divergence time, $\approx 8.75\text{-}26.38$ My. Taken together, these results indicate that the *VHn* genes are a conserved ancestral polymorphism that is being maintained in the Leporids genome. The expression of *VHn* genes in a similarly low frequency both in rabbit (10-20%; Kim and Dray 1973; Roux 1981) and hares (5-10%; this study) further shows that the antibody repertoire in leporids is subject to selective pressure, defying the widely accepted assumption that the adaptive immune system repertoire of receptors responsible for antigen recognition is generated randomly (summarized in Langman and Cohn, 2002). This selective pressure could either be exerted by pathogens or reside on the natural antibody repertoire, reflecting a pressure to maintain a certain set of self-reactivities.

One possible explanation for the maintenance of conserved *VHn* genes and their low frequency usage in leporids would be that these genes represent what were the ancestral leporid *VH* genes. Later in evolution the appearance of a more efficient antigen recognizing *VH*, the *VH1*, led to the preferential expression of a single *VH* in VDJ rearrangements, though not entirely abolishing the *VHn* usage in the generation of

the antibody repertoire. This is in accordance with the hypothesis that host pathogen co-evolution shapes the antibody diversity, acting both at the somatic and genomic level. Due to their ability to accommodate certain antigens, particularly those encountered over a long period of host pathogen interaction, some VH genes are retained in the host germline (Ota and Nei 1994; Rajewsky et al. 1987; Roost et al. 1995). Thus, the low frequency usage of *VHn* genes could be a remnant of an ancient leporids immunologic response to pathogens.

Alternatively, one can put forward the hypothesis that the maintenance of conserved *VHn* genes and their low frequency usage in leporids is required in order to maintain a certain balanced set of the natural antibody repertoire. Three VH clans, I, II and III, have been defined on the basis of FR1 sequence and structure identity (Kirkham et al. 1992; Kirkham and Schroeder 1994), which influences the nature of the elbow joint interaction between the V domain and the CH1 domain (Lesk and Chothia 1988). Within each clan, VH genes are grouped into families based on FR3 sequence and structure identity (Kirkham et al. 1992; Kirkham and Schroeder 1994), which shapes and frames the antigen binding site (reviewed in Chothia et al. 1992; Kirkham et al. 1992). *VH* genes representative of all three clans should have arisen in a tetrapods ancestral with subsequent random losses of *VH* clans along tetrapods evolution (Das et al. 2008), being that in rabbit only clan III *IGHV* genes are found. Curiously, differences that tell apart the *VHn* and *VH1* *VHa* derived genes are characteristic residues on FR1 and FR3 sequences, though the majority of which are located outside the clan specific sequences (FR1codons 6-25 and FR3 codons 67-85; Kirkham et al. 1992; Kirkham and Schroeder 1994) and thus *VHn* sequences clearly belong to clan III. Though clearly clan III members, changes in the *VHn* FR3 structure may affect the binding site conformation and structure, and thus the antibody antigen specificity.

Our results show that the *VHn* genes are not only present in the leporids genome but are also being used in the generation of antibody diversity for at least 12 My. This unusual mechanism, in that the *VHn* genes are used in a low frequency in both *Oryctolagus* and *Lepus* genera, shows that the antibody repertoire is subject to selective pressure.

Acknowledgements

The Portuguese Foundation for Science and Technology supported the doctoral fellowship of Ana Pinheiro (SFRH/BD/71252/2010) and the post-doctoral fellowship of

Pedro J. Esteves (SPRH/BPD/27021/2006). Isabel G. Fernández de Mera is supported by a Juan de la Cierva contract from the Spanish Ministry for Economy and Competitiveness. This work was also supported by a project of the Portuguese Foundation for Science and Technology (PTDC/BIA-BEC/103158/2008).

References

- Allegrucci M, Newman BA, Young-Cooper GO, Alexander CB, Meier D, Kelus AS, Mage RG (1990) Altered phenotypic expression of immunoglobulin heavy-chain variable-region (VH) genes in Alicia rabbits probably reflects a small deletion in the VH genes closest to the joining region. *Proc Natl Acad Sci U S A* 87 (14):5444-5448
- Alves PC, Hackländer K (2008) Lagomorph species: geographical distribution and conservation status. In: Alves PC, Ferrand N, Hackländer K (eds) *Lagomorph biology*. Springer, Berlin Heidelberg, pp 395–405
- Chen HT, Alexander CB, Young-Cooper GO, Mage RG (1993) VH gene expression and regulation in the mutant Alicia rabbit. Rescue of VHa2 allotype expression. *J Immunol* 150 (7):2783-2793
- Chothia C, Lesk AM, Gherardi E, Tomlinson IM, Walter G, Marks JD, Llewelyn MD, Winter G (1992) Structural repertoire of the human VH segments. *J Mol Biol* 227:799-817.
- Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* 9 (8):772
- Das S, Nozawa M, Klein J, Nei M (2008) Evolutionary dynamics of the immunoglobulin heavy chain variable region genes in vertebrates. *Immunogenetics* 60 (1):47-55
- DiPietro LA, Short JA, Zhai SK, Kelus AS, Meier D, Knight KL (1990) Limited number of immunoglobulin VH regions expressed in the mutant rabbit "Alicia". *Eur J Immunol* 20 (6):1401-1404
- Drummond AJ, Ho SY, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biol* 4 (5):e88
- Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7:214
- Esteves PJ (2003) Molecular and population genetic analysis of polymorphism at the antibody loci IgGCH2 and IgVH in lagomorphs. Faculdade de Ciências Universidade do Porto, Porto

- Esteves PJ, Lanning D, Ferrand N, Knight KL, Zhai SK, van der Loo W (2004) Allelic variation at the VHa locus in natural populations of rabbit (*Oryctolagus cuniculus*, L.). *J Immunol* 172 (2):1044-1053
- Esteves PJ, Lanning D, Ferrand N, Knight KL, Zhai SK, van der Loo W (2005) The evolution of the immunoglobulin heavy chain variable region (IgVH) in Leporids: an unusual case of transspecies polymorphism. *Immunogenetics* 57 (11):874-882.
- Flajnik MF (2002) Comparative analyses of immunoglobulin genes: surprises and portents. *Nat Rev Immunol* 2 (9):688-698
- Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52 (5):696-704
- Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41, 95–98.
- Horng WJ, Papagiannes E, Dray S, Rodkey LS (1980) Expression of cross-reacting determinants of the immunoglobulin heavy chain variable region a3 allotype in *Oryctolagus* and *Lepus*. *Mol Immunol* 17 (1):111-117
- Kelus AS, Weiss S (1986) Mutation affecting the expression of immunoglobulin variable regions in the rabbit. *Proc Natl Acad Sci U S A* 83 (13):4883-4886
- Kim BS, Dray S (1973) Expression of the a, x, and y variable region genes of heavy chains among IgG, IgM, and IgA molecules of normal and a locus allotype-suppressed rabbits. *J Immunol* 111 (3):750-760
- Kindt TJ (1975) Rabbit immunoglobulin allotypes: structure, immunology, and genetics. *Adv Immunol* 21:35-86
- Kirkham PM, Mortari F, Newton JA, Schroeder HW Jr. (1992) Immunoglobulin VH clan and family identity predicts variable domain structure and may influence antigen binding. *EMBO J* 11(2):603-9
- Kirkham PM, Schroeder HW Jr. (1994) Antibody structure and the evolution of immunoglobulin V gene segments. *Semin Immunol* 6(6):347-60
- Knight KL (1992) Restricted VH gene usage and generation of antibody diversity in rabbit. *Annu Rev Immunol* 10:593-616
- Knight KL, Becker RS (1990) Molecular basis of the allelic inheritance of rabbit immunoglobulin VH allotypes: implications for the generation of antibody diversity. *Cell* 60 (6):963-970
- Langman RE, Cohn M (2002) If the immune repertoire evolved to be large, random, and somatically generated, then... *Cell Immunol* 216(1-2):15-22
- Lefranc MP, Pommié C, Ruiz M, Giudicelli V, Foulquier E, Truong L, Thouvenin-Contet V, Lefranc G (2003) IMGT unique numbering for immunoglobulin and T cell

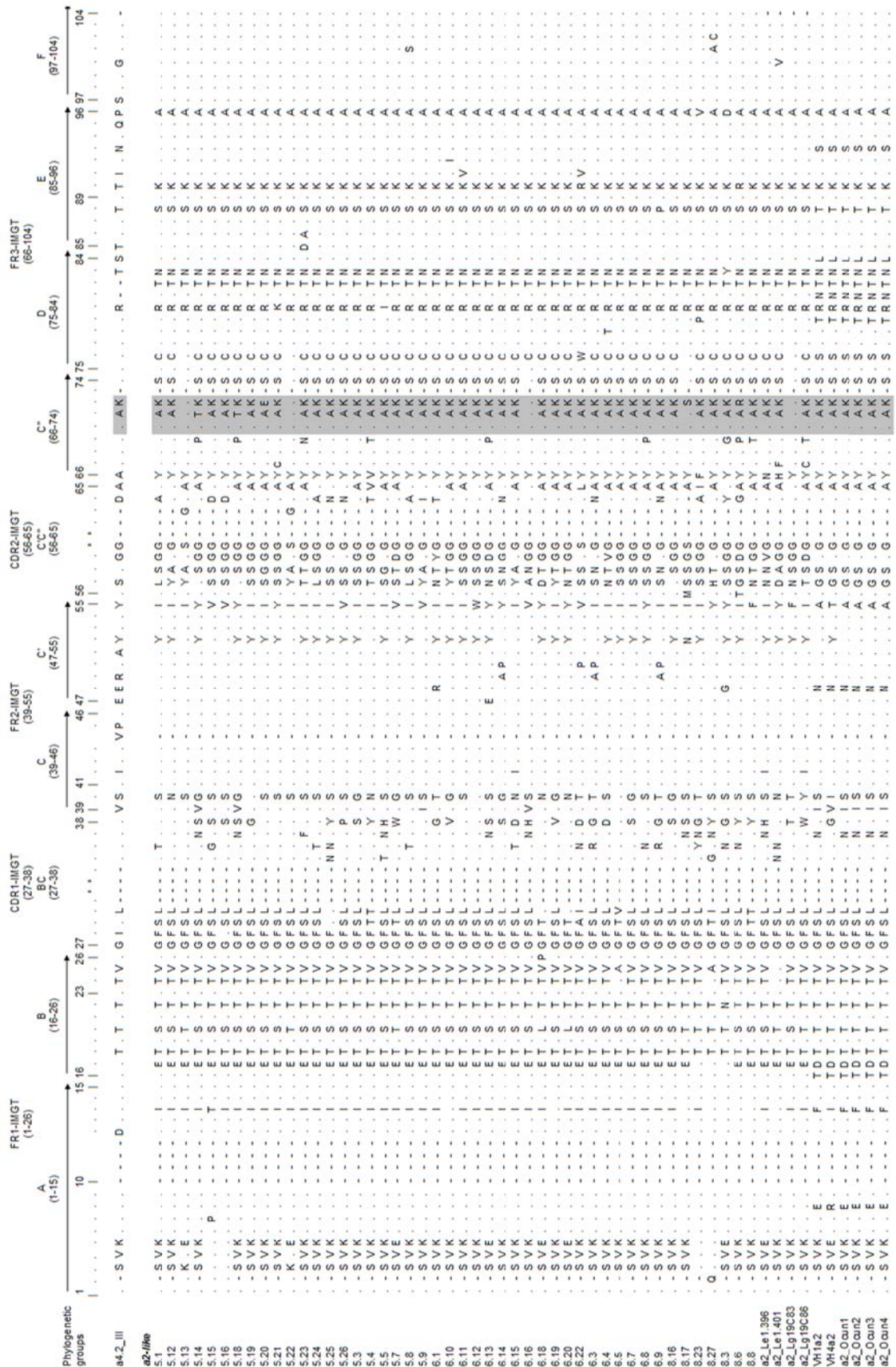
- receptor variable domains and Ig superfamily V-like domains. *Dev Comp Immunol* 27(1):55-77
- Lesk AM, Chothia C (1988) Elbow motion in immunoglobulins involves a molecular ball-and-socket joint. *Nature* 335:188-90
- Mage RG, Bernstein KE, McCartney-Francis N, Alexander CB, Young-Cooper GO, Padlan EA, Cohen GH (1984) The structural and genetic basis for expression of normal and latent VHa allotypes of the rabbit. *Mol Immunol* 21 (11):1067-1081
- Mage RG, Lanning D, Knight KL (2006) B cell and antibody repertoire development in rabbits: the requirement of gut-associated lymphoid tissues. *Dev Comp Immunol* 30 (1-2):137-153
- Margolies MN, Cannon LE, Kindt TJ, Fraser B (1977) The structural basis of rabbit VH allotypes: serologic studies on a1 H chains with defined amino acid sequence. *J Immunol* 119 (1):287-294
- Matthee CA, van Vuuren BJ, Bell D, Robinson TJ (2004) A molecular supermatrix of the rabbits and hares (Leporidae) allows for the identification of five intercontinental exchanges during the Miocene. *Syst Biol* 53 (3):433-447
- Ota T, Nei M (1994) Divergent evolution and evolution by the birth-and-death process in the immunoglobulin VH gene family. *Mol Biol Evol* 11 (3):469-482
- Pinheiro A, Lanning D, Alves PC, Mage RG, Knight KL, van der Loo W, Esteves PJ (2011) Molecular bases of genetic diversity and evolution of the immunoglobulin heavy chain variable region (IGHV) gene locus in leporids. *Immunogenetics* 63 (7):397-408
- Rajewsky K, Forster I, Cumano A (1987) Evolutionary and somatic selection of the antibody repertoire in the mouse. *Science* 238 (4830):1088-1094
- Rambaut A, Drummond AJ (2007) Tracer v1.5. Available from <http://beast.bio.ed.ac.uk/Tracer>.
- Roost HP, Bachmann MF, Haag A, Kalinke U, Pliska V, Hengartner H, Zinkernagel RM (1995) Early high-affinity neutralizing anti-viral IgG responses without further overall improvements of affinity. *Proc Natl Acad Sci U S A* 92 (5):1257-1261
- Ros F, Puels J, Reichenberger N, van Schooten W, Buelow R, Platzer J (2004) Sequence analysis of 0.5 Mb of the rabbit germline immunoglobulin heavy chain locus. *Gene* 330:49-59
- Roux KH (1981) A fourth heavy chain variable region subgroup, w, with 2 variants defined by an induced auto-antiserum in the rabbit. *J Immunol* 127 (2):626-632
- Su C, Nei M (1999) Fifty-million-year-old polymorphism at an immunoglobulin variable region gene locus in the rabbit evolutionary lineage. *Proc Natl Acad Sci U S A* 96 (17):9710-9715

- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, and Kumar S (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol Biol Evol* 28: 2731-2739
- Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680
- Tonnelle C, Cazenave PA, Brezin C, Moinier D, Fougereau M (1983) Structural correlates to the rabbit immunoglobulin heavy chain a100 allotype. *Mol Immunol* 20 (7):753-761
- van der Loo W (1987) Population genetical studies on the adaptive significance of the immunoglobulin allotypes in wild rabbit. In: Dubiski S (eds) *The Rabbit in Contemporary Immunological Research*. Longman Scientific & Technical, John Wiley, pp 101-126
- van der Loo W, Abrantes J, Esteves PJ (2009) Sharing of endogenous lentiviral gene fragments among leporid lineages separated for more than 12 million years. *J Virol* 83 (5):2386-2388

Supplementary Material

Supp. table 1- Alignment of European rabbit and hare VH protein sequences. Dashes (-) represent alignment gaps. Dots (·) indicate identity with the master sequence except at gap positions. The 'IMGT Protein display for V domain' header is shown (LeFranc et al, 2003). Sequences are grouped into phylogenetic groups according to the tree in Figure 1. Shared characteristic residues of European rabbit and hare VHn genes are highlighted in gray. Also highlighted in gray are the VHa 70WAK72, VHn 70WVN72 and hares sL sequences 70SVK72 motifs.

Phylogenetic groups	FR1-IMGT (1-26)					CDR1-IMGT (27-38)					FR2-IMGT (39-55)					CDR2-IMGT (56-65)					FR3-IMGT (66-104)																		
	A (1-15)					B (16-26)					C (39-46)					C' (47-55)					C'' (66-74)					D (75-84)					E (85-96)					F (97-104)			
	1	10	15	16	23	26	27	38	39	41	46	47	55	56	65	66	74	75	84	85	89	96	97	104															
VRh	QEQLVESGG	---	GLVRPGGSLK	LS	KG	SELD	F	---	SS	Y	AM	CWVR	Q	---	AP	KG	LEW	IG	---	---	---	---	---	---															
5.2	K	---	A	GF	L	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
Vh ₁ O _{1a}	K	---	Q	I	V	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
Vh ₂ O ₂	K	---	Q	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
Vh ₂	K	---	Q	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
VRh	E	---	E	---	E	---	---	E	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
6.17	E	---	E	---	E	---	---	E	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
8.11	E	---	E	---	E	---	---	E	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
Vh ₁	E	---	E	---	E	---	---	E	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
Vh ₂ O ₁	E	---	E	---	E	---	---	E	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
Vh ₂ O ₂	E	---	E	---	E	---	---	E	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
Vh ₂	E	---	E	---	E	---	---	E	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
VRh	E	---	E	---	E	---	---	E	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
8.1	E	---	E	---	E	---	---	E	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
8.12	E	---	E	---	E	---	---	E	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
8.13	E	---	E	---	E	---	---	E	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
8.18	E	---	E	---	E	---	---	E	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
8.19	E	---	E	---	E	---	---	E	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
8.20	E	---	E	---	E	---	---	E	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
8.24	E	---	E	---	E	---	---	E	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
8.25	E	---	E	---	E	---	---	E	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
8.26	E	---	E	---	E	---	---	E	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
8.5	E	---	E	---	E	---	---	E	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
8.7	E	---	E	---	E	---	---	E	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
8.9	E	---	E	---	E	---	---	E	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
41.397	E	---	E	---	E	---	---	E	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
41.400	E	---	E	---	E	---	---	E	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
41.408	E	---	E	---	E	---	---	E	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
41.412	E	---	E	---	E	---	---	E	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
41.415	E	---	E	---	E	---	---	E	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
41.415_540	E	---	E	---	E	---	---	E	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
41.415_541	E	---	E	---	E	---	---	E	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
41.415_542	E	---	E	---	E	---	---	E	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
41.415_544	E	---	E	---	E	---	---	E	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
VRh	S	---	S	---	S	---	---	S	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
VRh ₁	S	---	S	---	S	---	---	S	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
VRh ₄	S	---	S	---	S	---	---	S	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
VRh ₄	S	---	S	---	S	---	---	S	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
VRh ₁ O ₁	S	---	S	---	S	---	---	S	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
VRh ₁ O ₂	S	---	S	---	S	---	---	S	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
VRh ₁ O ₃	S	---	S	---	S	---	---	S	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
VRh ₁ O ₄	S	---	S	---	S	---	---	S	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
VRh	S	---	S	---	S	---	---	S	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
VRh ₁	S	---	S	---	S	---	---	S	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
VRh ₃	S	---	S	---	S	---	---	S	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
VRh ₄	S	---	S	---	S	---	---	S	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
VRh ₁ O ₁	S	---	S	---	S	---	---	S	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
VRh ₁ O ₂	S	---	S	---	S	---	---	S	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
VRh ₁ O ₃	S	---	S	---	S	---	---	S	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
VRh ₁ O ₄	S	---	S	---	S	---	---	S	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
VRh	S	---	S	---	S	---	---	S	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
VRh ₁	S	---	S	---	S	---	---	S	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
VRh ₁ O ₁	S	---	S	---	S	---	---	S	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
VRh ₁ O ₂	S	---	S	---	S	---	---	S	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
VRh ₁ O ₃	S	---	S	---	S	---	---	S	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
VRh ₁ O ₄	S	---	S	---	S	---	---	S	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
VRh	S	---	S	---	S	---	---	S	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
VRh ₁	S	---	S	---	S	---	---	S	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
VRh ₁ O ₁	S	---	S	---	S	---	---	S	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
VRh ₁ O ₂	S	---	S	---	S	---	---	S	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
VRh ₁ O ₃	S	---	S	---	S	---	---	S	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
VRh ₁ O ₄	S	---	S	---	S	---	---	S	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
VRh	S	---	S	---	S	---	---	S	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
VRh ₁	S	---	S	---	S	---	---	S	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
VRh ₁ O ₁	S	---	S	---	S	---	---	S	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
VRh ₁ O ₂	S	---	S	---	S	---	---	S	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
VRh ₁ O ₃	S	---	S	---	S	---	---	S	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
VRh ₁ O ₄	S	---	S	---	S	---	---	S	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
VRh	S	---	S	---	S	---	---	S	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
VRh ₁	S	---	S	---	S	---	---	S	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
VRh ₁ O ₁	S	---	S	---	S	---	---	S	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
VRh ₁ O ₂	S	---	S	---	S	---	---	S	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
VRh ₁ O ₃	S	---	S	---	S	---	---	S	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---															
VRh ₁ O ₄	S	---	S	---	S	---	---	S	---	---	---	---	---	---	---	---																							



Paper 3

Sequencing of *Sylvilagus* *VDJ* genes reveals a new *VHa* allelic lineage and shows that ancient *VH* lineages were retained differently in leporids.

Ana Pinheiro, José Melo-Ferreira, Joana Abrantes, Nicola Martinelli, Antonio Lavazza, Paulo C. Alves, Christian Gortázar, and Pedro J. Esteves

Immunogenetics, 66(12), 719-26. doi: 10.1007/s00251-014-0807-0.

Abstract

Antigen recognition by immunoglobulins depends upon initial rearrangements of heavy chain *V*, *D*, and *J* genes. In Leporids a unique system exists for the *VH* genes usage that exhibit highly divergent lineages: the *VHa* allotypes, the *Lepus* sL lineage, and the *VHn* genes. For the European rabbit (*Oryctolagus cuniculus*) four *VHa* lineages have been described, the a1, a2, a3 and a4. For hares (*Lepus* sp.) one *VHa* lineage was described, the a2L, as well as a more ancient sL lineage. Both genera use the *VHn* genes in a low frequency of their *VDJ* rearrangements. To address the hypothesis that the *VH* specificities could be associated with different environments we sequenced *VDJ* genes from a third leporid genus, *Sylvilagus*. We found a fifth and equally divergent *VHa* lineage, the a5, and an ancient lineage, the sS, related to the hares sL, but failed to obtain *VHn* genes. These results show that the studied leporids employ

different *VH* lineages in the generation of the antibody repertoire, suggesting that the leporid *VH* genes are subject to strong selective pressure likely imposed by specific pathogens.

Keywords

IGHV, Leporids, antibody diversification, *VH* allotypes, evolution

Introduction

For the generation of the primary antibody repertoire most mammals express a varying number of rearranged *VDJ* genes (Flajnick 2002). The European rabbit (*Oryctolagus cuniculus*) has a unique system in that, despite having more than 200 *VH* genes (Ros et al. 2004), it predominantly expresses just one *IGHV*, the most D-proximal *VH1* gene, to generate 80-90% of its antibody repertoire (Knight and Becker 1990; Knight 1992). The immunoglobulin (Ig) molecules derived from the *VH1* express the so called *VHa* allotypic markers (e.g., Kindt 1975; Margolies et al. 1977), with three serological lineages having been described in the domestic rabbit, the a1, a2 and a3. These are highly divergent ($\pm 20\%$ amino acid sequence differences), with the allelic specificities of a1 and a2 being correlated with several amino acid differences in framework regions (FR) 1 and 3 (Tonnellet et al. 1983; Mage et al. 1984; Knight and Becker 1990). A fourth equally divergent allotypic lineage, the a4, was described in wild European rabbit Iberian populations (Esteves et al., 2004). The remaining 10–20% of Ig molecules are encoded by the *VHn* genes, *VHx*, *VHy*, and *VHz* (Kim and Dray 1973; Horng et al. 1976; Roux 1981), that map at least 100 Kb upstream of *VH1* (Mage et al. 2006), and do not express *VHa* allotype specific determinants.

Studies on the *IGHV* locus diversity for other leporids are limited to a few data obtained for some *Lepus* species. Serological analysis conducted for *L. americanus* showed cross-reaction with rabbit anti-a1, anti-a2, and anti-a3 antisera (van der Loo 1987). As for *L. europaeus* and *L. granatensis*, a serological analysis of several populations revealed two phenotypes: partial reaction to anti-a2 antisera and no reaction to any rabbit antiserum (Esteves 2003). Sequencing of rearranged *VH* genes for *Lepus* specimens, some of which expressed a2-cross-reacting serum proteins, proved that the a2 polymorphism is trans-specific. The *Lepus* a2L lineage showed some of the most outstanding characteristics of the rabbit *VHa* genes such as the presence of 5 out of 11 amino acid residues that characterize the allotype *VHa2* (Mage et al. 1984), and being more closely related to the rabbit *VH1*-a2 allele than *VH1*-a2 was to its allelic counterparts *VH1*-a1 and *VH1*-a3 (Esteves et al. 2005). A second lineage, the sL, apparently a *Lepus* specific lineage, groups distinctively apart from all other *VH* lineages present in rabbit, showing as distinctive characters the ancestral $_{70}\text{SVK}_{72}$ motif and residues A₁₀ and K₉₅ (IMGT numbering, Lefranc et al. 2003) (Esteves et al. 2005). More recently, it was shown that, like rabbits, *Lepus* uses the *VHn* genes in 5-10% of its *VDJ* rearrangements revealing that the antibody repertoire is subject to selective pressure and that the low frequency usage of *VHn* genes could be a remnant

of an ancient leporids immunologic response to pathogens (Pinheiro et al. 2011, Pinheiro et al. 2013).

Considering that evidence suggests that the leporid *VH* lineages used to generate the antibody repertoire are subject to selective pressure, in this study we addressed the hypothesis that the *VH* lineages represented in the antibody repertoire could be different in species that inhabit different environments due to exposure to different pathogens. Family Leporidae comprises eleven genera, among which *Sylvilagus* occupies an intermediate position between *Oryctolagus* and *Lepus* (Alves and Hackländer 2008), having diverged from *Oryctolagus* 10 million years ago and from *Lepus* 12 million years ago (Mathee et al. 2004). *Oryctolagus* and the two *Lepus* species, *L. europaeus* and *L. granatensis*, for which *VDJ* genes nucleotide sequences have been obtained, inhabit the Eurasian continent while *Sylvilagus* is a native American leporid (Chapman et al. 1980; Flux and Angerman 1990). Starting in 1966 eastern cottontails were introduced in northern Italy through a series of massive releases for hunting purposes; nowadays eastern cottontails are widespread in this region (Vidus-Rosin et al. 2008). Herein we sequenced *Sylvilagus floridanus* *VDJ* rearranged genes.

Material and methods

Samples, amplification and sequencing of rearranged *VDJ* genes

Spleen samples were collected from four eastern cottontail (*Sylvilagus floridanus*) specimens collected in Pistoia province, Italy. Total RNA was extracted using RNeasy Mini Kit (Qiagen, Hilden, Germany), following first strand cDNA synthesis using the iScript cDNA Kit (Bio-Rad, Hercules, CA, USA). *VDJ* rearranged genes were PCR amplified using primers *VH* (5'GGAGACTGGGCTGCGCTGGCTTCTCCTGGT3'; Esteves et al. 2005) and *JH2* (5'TGAGGAGACGGTGACCAGGGTGCCT3'; Pinheiro et al. 2013). PCR amplification conditions were as follows: 4 min at 95 °C followed by 35 cycles at 95 °C (45 s), 62 °C (45 s) and 72 °C (60 s), with a final extension at 72 °C (5 min). PCR products were purified (MinElute PCR Purification Kit, Qiagen, Hilden, Germany) and cloned into the pGEM-T Easy vector system II (Promega, Madison, WI, USA).

Phylogenetic Analyses

To perform phylogenetic analyses available sequences for Leporids *VDJ* genes were taken from GenBank. Used sequences include rabbit germline *VH1* and *VH4* genes of a1, a2 and a3 allotypes, rabbit cDNA *VH* gene sequences representative of allotypes a1, a2, a3, a4.1 and a4.2, rabbit germline and rabbit and *Lepus* cDNA *VH* gene sequences of *VHx*, *VHy*, and *VHz*, and *Lepus* cDNA *VH* gene sequences representative of lineages a2L, sLe and sLg. *VH* genes of the human *VH3* family were used as an outgroup. These were aligned with sequences obtained in this study using CLUSTAL W (Thompson et al. 1994) as implemented in BioEdit software (Hall, 1999) and the amino acid sequences inferred. Accession numbers for all sequences are given in Table 1.

Table 1- Vh genes sequences accession numbers for sequences used in the phylogenetic analysis

Sequence identification	Accession number
<i>rabbit germline sequences</i>	
VH1a1, VH4a1	M93171, M93181
VH1a2, VH4a2	M93172, M93182
VH1a3, VH4a3	M93173, L03846
Vhx, Vhx2	L03846, M19706
Vhy, Vhy2	L03890, L03874
Vhz	AF264469
<i>rabbit cDNA sequences</i>	
a1_Ocun1-Ocun4	AF029933, AF029934, AF029938, AF029940
a2_Ocun1-Ocun4	L03849, L03851, L03853, L03856
a3_Ocun1-Ocun4	AF029923, AF098235, AF264434, AF264435
a4.1_I-IV, a4.2_I-III	AY207979- AY207982, AY208042, AY208047, AY207967
Vhx_Oca, Vhx_Oca2	AY207986, AY208045
Vhy_Oca	AY208006
<i>hare cDNA sequences</i>	
sLe1_397, sLe1_400, sLe2_408, sLe2_412, Le8.1, Le8.12	AY288451, AY288453, AY288459, AY288462, KF460076, KF460085
sLg15_640-642, sLg15_644	AY288464- AY288467
a2_Le1.396, a2_Le1.401, a2_Lg19C83, a2_Lg19C86, Le5.13, Le6.1, Le8.16	AY288450, AY288454, AY288491, AY288494, KF460042, KF460056, KF460087
VHx Le5.2, Le5.17,	KF460034, KF460046
VHy-VHz Le6.17, Le6.24, Le8.11	KF460070, KF460075, KF460084,
<i>human VH3 family sequences</i>	
Hsap1- Hsap4	M99666, M99672, M99679, M99682
<i>Current study</i>	
Sfl1	KM275500- KM275522
Sfl2	KM275523- KM275542
Sfl3	KM275543- KM275568
Sfl4	KM275569- KM275585

The computer program jModelTest v2.1.1 (Darriba et al. 2012) was used to assess the fit of our dataset to 88 models of sequence evolution considering the corrected Akaike information criterion (AICc). The phylogenetic relationships were analysed in a Bayesian inference (BI) framework, specifying as optimal mutation model K80+G, retrieved from jModelTest. BI analyses were performed using BEAST v1.7.4 (Drummond and Rambaut 2007). The BEAUti application, part of BEAST package, was used to create the input file for *BEAST. Posterior phylogenies were determined using an uncorrelated lognormal relaxed clock (Drummond et al. 2006) and the Yule tree prior. All parameter priors were set to the defaults. Three independent runs of 50×10^6 generations were performed, sampling trees and parameters every 5×10^3 generations, and concatenated using LogCombiner. Convergence was checked using Tracer v1.5 (Rambaut and Drummond 2007) and summary trees were generated with TreeAnnotator v1.5.4, part of the BEAST package.

Genetic distances between *VH* genes among species were calculated using the “compute net average distance between groups” option of MEGA5.0 software (Tamura et al. 2011). This option corrects for variance due to differences within groups, as expected in *VH* rearranged genes which are somatically diversified.

Results

A total of 86 *VDJ* gene sequences were obtained for the four studied *S. floridanus* individuals. Phylogenies were estimated with the whole *VH* sequence and excluding the CDR's, yielding similar clustering of the analysed sequences. Thus, we present here the phylogeny obtained with the whole *VH* sequence (Figure 1). Clusters of the previously described rabbit and hare lineages are shown with good support on the phylogenetic tree: the *VHn* lineage (1.00 posterior probability), the *sL* lineage (1.00 posterior probability) and the *VHa* cluster (0.96 posterior probability) that includes the *VHa1*, *VHa2*, *VHa2L*, *VHa3* and *VH4* lineages (1.00, 1.00, 0.99, 1.00 and 1.00 posterior probabilities, respectively). The *S. floridanus* sequences obtained in this study fall into two groups that include only *Sylvilagus* *VH* sequences: (i) a cluster that is not related to any previously described leporid lineages, the *sS* lineage (1.00 posterior probability); and (ii) a second group that clusters with the rabbit *VHa* lineages, designated here as *a5* lineage (1.00 posterior probability). This *a5* lineage is further

subdivided into two groups, the a5.1 and a5.2 variants (0.98 and 0.95 posterior probabilities, respectively).

Of the 86 obtained sequences, 46 classified as a5.1, 26 classified as a5.2 and 14 sequences classified as sS. These three groups appear to have an allelic distribution as each one of the analysed individuals has sequences belonging to two of the groups: the Sfl1 and Sfl4 specimens have sequences that classified as a5.1 and sS (13-9 and 11-5) and the Sfl2 and Sfl3 specimens have sequences that classified as a5.1 and a5.2 (2-17 and 17-9).

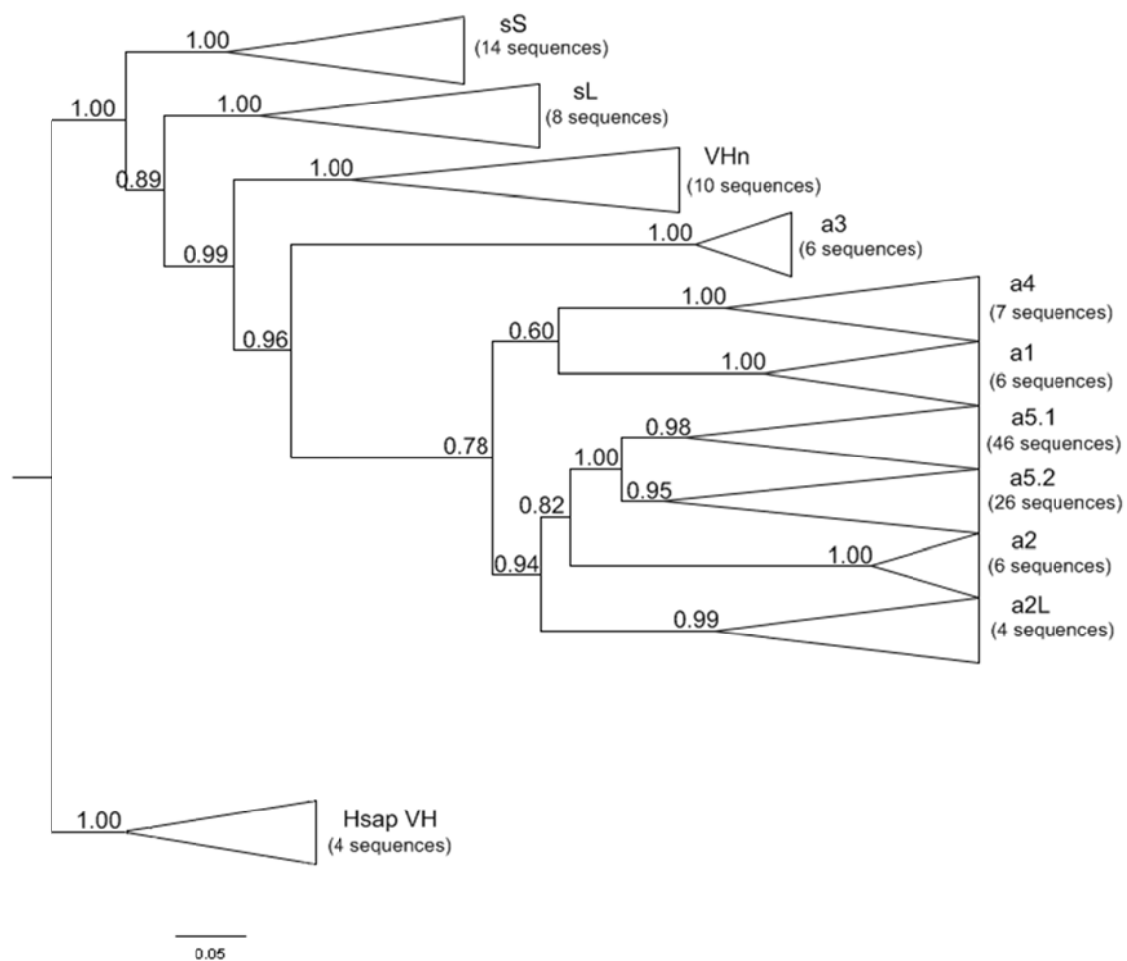


Fig. 1. Phylogenetic tree of Leporid VH genes. Human VH clan 3 genes were used to root the tree. Groups were collapsed for simplification; the number of sequences in each group is indicated. The a1, a2, a3 and a4 groups include only European rabbit sequences, the a2L and sL groups include only hares sequences and the a5.1, a5.2 and sS groups include only *S. floridanus* sequences. The Vhn group includes European rabbit and hares sequences. BI posterior probabilities are depicted in front of each node.

Genetic distances between rabbit, hares and cottontail VH genes

The genetic distances between *S. floridanus* sS lineage and the a1-a4 lineages (0.12-0.16, Table 2) are identical to those between the hares sL lineage and the VHa1-VHa4 lineages (0.10-0.16, Table 2) or between the VHn and the VHa1-VHa4 lineages (0.08-0.15, Table 2). The *S. floridanus* a5 lineage genetic distances to other VHa lineages (0.06-0.10, Table 2) are identical to the genetic distances observed between VHa1-VHa4 lineages (0.07-0.14, Table 2). The genetic distance between the a5.1 and a5.2 variants is lower than the genetic distance between a4.1 and a4.2 variants (0.03 and 0.06, respectively).

Table 2- Genetic distances between leporids VH nucleotide sequences.

Groups	VHn	sL	sS	VHa									
				a1	a2	a2L	a3	a4	a5				
									a4.1	a4.2	a5.1	a5.2	
VHn		0,01	0,01	0,02	0,02	0,01	0,01	0,02	0,02	0,02	0,01	0,01	0,01
sL	0,07		0,01	0,02	0,02	0,02	0,01	0,02	0,02	0,02	0,02	0,02	0,01
sS	0,09	0,05		0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,01	0,02	0,01
VHa	a1	0,10	0,11	0,12		0,02	0,01	0,01	0,01	0,01	0,02	0,01	0,01
	a2	0,15	0,16	0,16	0,11		0,01	0,02	0,02	0,02	0,02	0,01	0,01
	a2L	0,08	0,10	0,14	0,07	0,08		0,02	0,01	0,02	0,02	0,01	0,01
	a3	0,08	0,10	0,13	0,08	0,14	0,10		0,02	0,02	0,02	0,01	0,01
	a4	0,12	0,11	0,14	0,08	0,13	0,09	0,11		---	---	0,01	---
	a4.1	0,12	0,11	0,14	0,08	0,13	0,09	0,10	---		0,01	---	0,02
	a4.2	0,16	0,15	0,19	0,12	0,16	0,11	0,15	---	0,06		---	0,02
	a5	0,10	0,11	0,12	0,07	0,08	0,06	0,10	0,09	---	---		---
	a5.1	0,11	0,12	0,13	0,09	0,09	0,07	0,11	---	0,11	0,13	---	0,01
	a5.2	0,11	0,11	0,12	0,07	0,08	0,06	0,10	---	0,09	0,13	---	0,03

¹Genetic distances between groups of sequences, calculated using the p-distance, are shown below the diagonal. Standard error estimates are shown above the diagonal in italics. All ambiguous positions were removed for each sequence pair.

Characterization of expressed VH in *S. floridanus*

The inferred amino acid sequences were grouped according to the phylogenetic clustering and compared to *Lepus* cDNA and rabbit germline and cDNA VH genes representative of VHn and VHa allotypes as well as sequences representative of lineages a2L and sL (Figure 2).

The rabbit VHa sequences are characterized by the $_{19}\text{LTLTCT}_{24}$ of FR1 and $_{70}\text{WAK}_{72}$ of FR3 motifs as well as showing a deletion at position 2 (Mage et al. 1984). The 72 sequences obtained in this study that classified as VHa have the deletion at position 2 and the majority either has the characteristic $_{70}\text{WAK}_{72}$ or a $_{70}\text{WAQ}_{72}$ motif. The $_{19}\text{LTLTCT}_{24}$ of FR1 is also present though the majority of the sequences have a D₂₀ substitution. These sequences further have as characteristics the motif $_{83}\text{SAG}_{85}$ and residue D₉₆, while also sharing the VHa2 I/F₁₃ and T₇₉ residues. Thus, these sequences have the rabbit VHa hallmark residues but have evolved some particular features which substantiate their inclusion in a new VHa lineage that we designate as a5. According to the phylogenetic tree this a5 lineage has two variants the a5.1 and a5.2 and amino acid differences exist between them that support their distinction. The $_{19}\text{LTLTCT}_{24}$ motif is more conserved in the sequences assigned as a5.2 while those that classified as a5.1 have the D₂₀ as well as K₂₄ substitutions. In FR3 the $_{83}\text{SAG}_{85}$ D₉₆ motif is more conserved in the sequences that classified as a5.1 while those identified as a5.2 have K₈₀, L₈₇, G₉₃ and E₉₇ residues that, although not characteristic to this variant, are not present in the a5.1 variant.

The 14 sequences that classified as a new lineage (sS) have the ancestral $_{70}\text{SVK}_{72}$, also present in the hares sL lineage, and the $_{2}\text{EQ}_3$ motif of rabbit VHx peptides. The FR3 amino acid sequence between residues 83-101 has nine substitutions that are characteristic to this lineage: $_{83}\text{DA}_{84}$, $_{87}\text{LE}_{88}$, D₉₃, $_{95}\text{KTT}_{97}$ and I₁₀₁.

Discussion

The generation of antibody repertoire in the European rabbit has been widely studied and several particularities have been found. It seems to rely on a reduced number of VH genes: the VH1, which expresses the so called VHa allotypic markers, and the VHx, VHy and VHz genes, or VHn genes, which do not have the VHa allotype specificities (Mage et al. 1984; Knight and Becker 1990). For the European rabbit four highly divergent allelic lineages have been described for the VHa allotypes: a1, a2, a3 and a4 (Mage et al. 1984; Knight and Becker 1990; Esteves et al. 2004). In hares

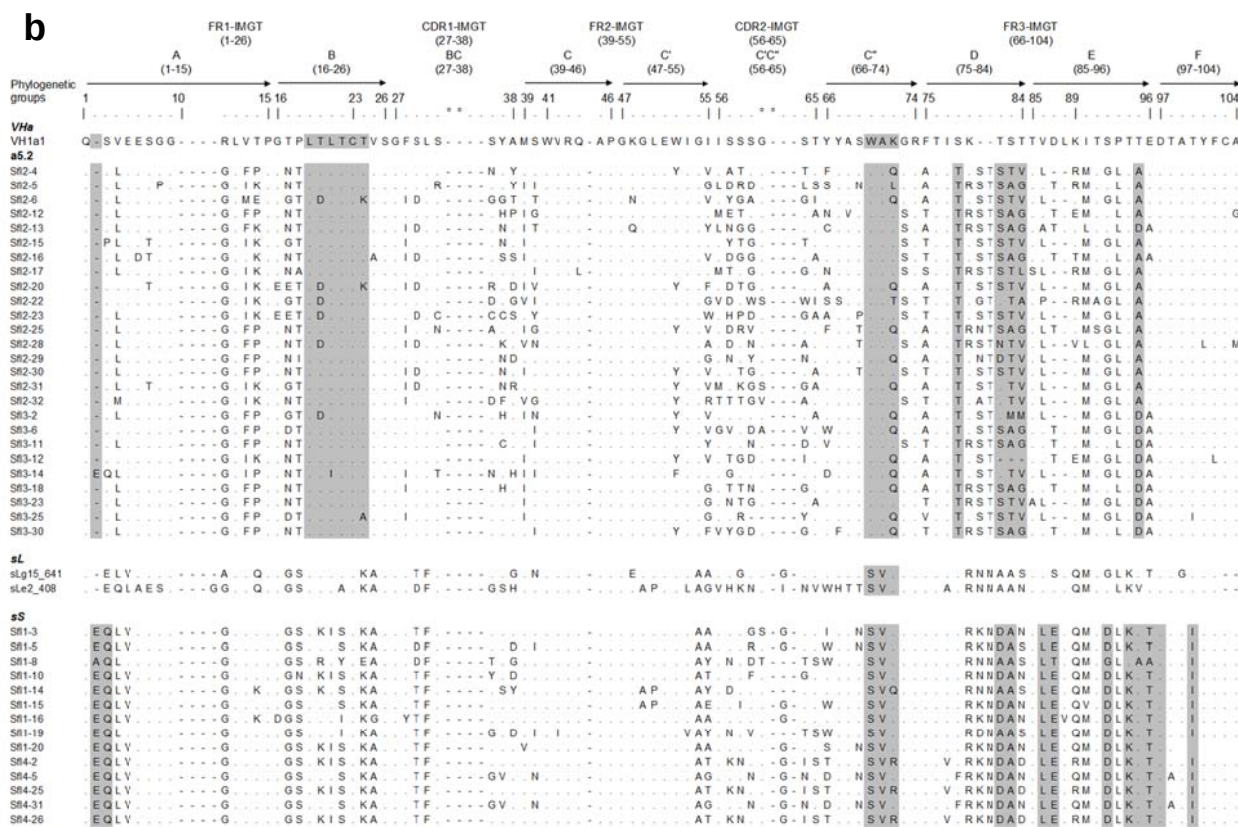
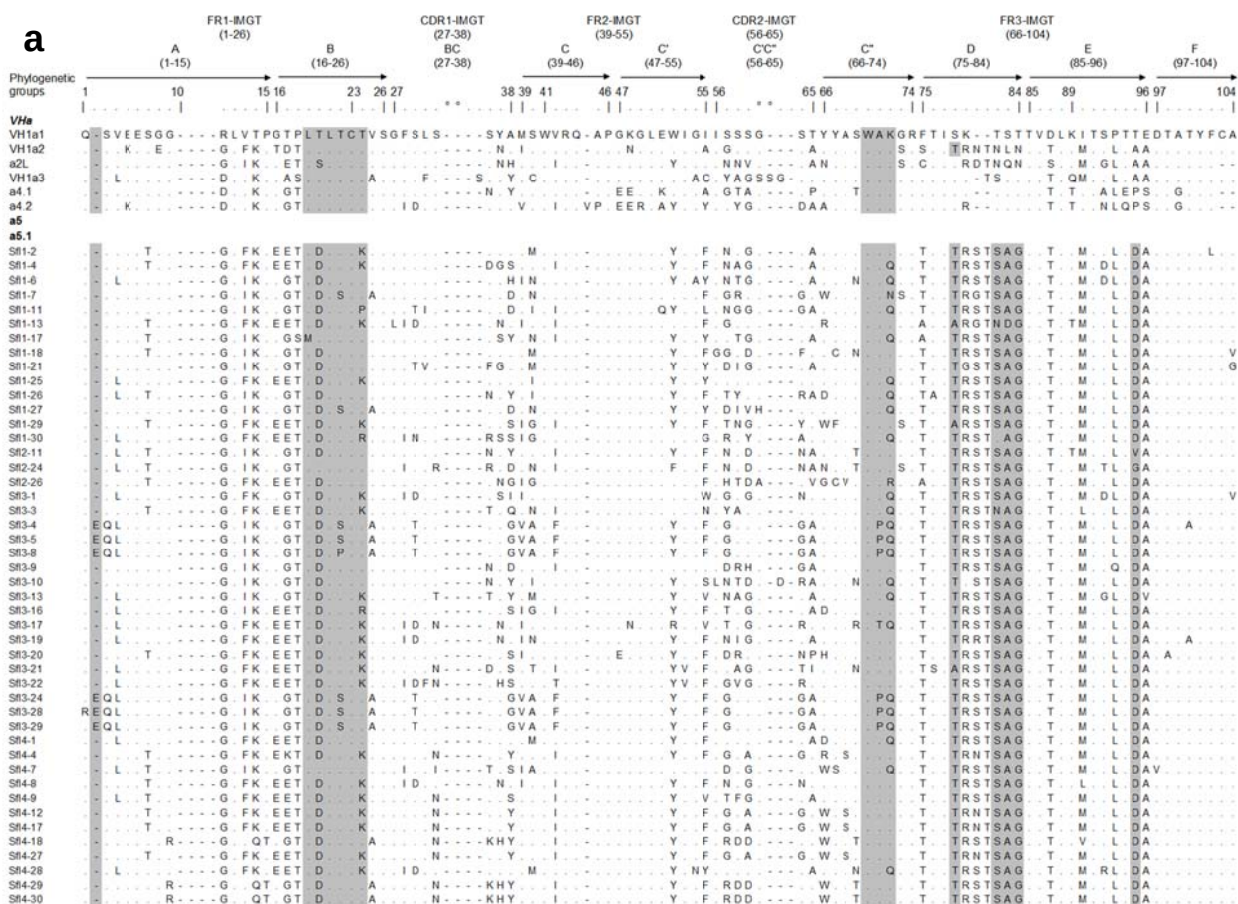


Fig. 2. Alignment of leporid VH protein sequences. a) European rabbit and hares VHa lineages and *S. floridanus* a5.1 lineage, b) *S. floridanus* a5.2 and sS lineages, hares sL lineage and European rabbit and hares VHn lineage. Dashes (-) represent alignment gaps. Dots (·) indicate identity with the master sequence except at gap positions. The 'IMGT Protein display for V domain' header is shown (Lefranc et al, 2003). Sequences are grouped into phylogenetic groups according to the tree in Figure 1. Shown are representatives of European rabbit VH1a1, VH1a2, VH1a3, Vhx, Vhy and Vhz germline sequences (genbank accession numbers M93171, M93172, M93173, L03846, L03890 and AF264469, respectively), European rabbit a4.1 and a4.2 cDNA sequences (genbank accession numbers AY207980 and AY207967), hares a2L and sL cDNA sequences (genbank accession numbers AY288450, AY288465 and AY288459). Also shown are all *S. floridanus* cDNA sequences obtained in this study that group into the a5.1, a5.2 and sS lineages (for genbank accession numbers please see table 1). Highlighted in gray are the VHa deletion at position 2 and motifs ₁₉LTLTCT₂₄ and ₇₀WAK₇₂, and sL/sS sequences ₇₀SVK₇₂ motif. Also highlighted are residues characteristic of a5 and sS lineages.

(*Lepus* sp.) the existence of a VHa lineage, a2L, with nucleotide and amino acid sequence similarity with the rabbit a2 lineage and which shows cross reaction with sera against the rabbit a2 lineage has showed the trans-specific nature of the VHa2 polymorphism (Esteves et al. 2005). A more ancestral lineage, sL, presumably also derived from the VH1 gene, was found to be expressed in hares (Esteves et al. 2005). More recently Pinheiro and co-workers (2013) showed that hares and rabbits share the low frequency usage of the VHn genes in the generation of the antibody repertoire.

In this study, we observed VH expressed genes for *Sylvilagus floridanus* that have the rabbit VHa signature, lineage a5, as well as VH expressed genes that belong to a more ancestral lineage, sS, that seems to be *S. floridanus* specific. In the phylogenetic tree the *S. floridanus* a5 sequences group with rabbit and hare VHa lineages, clearly apart from the *S. floridanus* sS sequences, and the genetic distances also indicate that the a5 lineage is more similar to the VHa lineages than to the sS lineage. These results suggest that the VHa allotypes have an origin that pre-dates the Leporid radiation. Some VHa lineages originated before the species split and are shared between species, such as the a2 and a2L lineages, whereas others apparently are species specific. As Esteves et al. (2004) argued, it is likely that evolutionary modes, and mutational rates, vary among VHa lineages having increased in the a1, a2 and a4 lineages, which would explain the large inter-lineage distances despite a more recent origin. An interesting aspect resides in the differences in number of VHa lineages found in each leporid species studied so far. Apparently the rabbit has retained a greater diversity of VHa lineages, with at least four lineages (Mage et al. 1984; Knight and Becker 1990; Esteves et al. 2004), while hares and cottontail have just one lineage each (Esteves et al. 2005; Pinheiro et al. 2013; present study). This

higher number of *VHa* lineages found in rabbit can be the result of a compensatory mechanism due to the loss of a more ancient lineage of *VH* genes, which would be the rabbit equivalent of the hares sL and cottontail sS. These ancient lineages, characterized by the motif $_{70}\text{SVK}_{72}$ and more highly divergent from the *VHa* lineages, can importantly contribute to the diversity of the antibody repertoire. Its loss may have imposed the diversification of the *VHa* lineages in rabbit. Contrarily to what is widely accepted, increasing evidence suggests the *VH* lineages which are rearranged in different leporid species are under different selective pressures, either driven by pathogens or by the self, though it is not evident how this selection operates (Pinheiro et al., 2013). It has also been shown that in rabbit the diversification of the antibody repertoire requires the interaction with exogenous factors such as the gut microbiota (Lanning et al. 2000a, 2000b; Sehgal et al. 2002), which may also exert selective pressure on the expressed *VH* lineages.

Sequencing of random germline *VH* genes has shown that *VHn* genes are present in *S. floridanus* genome (Esteves, 2003) and thus we hypothesized that, similarly to what was found for rabbit and hares (Pinheiro et al. 2013), *S. floridanus* would use these genes in a low percentage of *VDJ* rearrangements. Interestingly, despite the identical sequencing effort made for *Lepus* and *Sylvilagus* rearranged *VH* (circa 25 sequences *per* individual; Pinheiro et al. 2013; present study), we failed to recover rearranged *VDJ* genes with characteristics of rabbit and hares *VHn* genes in the studied *S. floridanus* specimens. *S. floridanus* may not use the *VHn* genes in the generation of the antibody repertoire or may do so in such a low percentage (under 4%) that the sequencing effort would have to be greatly increased to allow its detection. The usage of *VHn* genes in low frequency both in rabbit (Kim and Dray 1973; Roux 1981) and hares shows that, in leporids, the *VH* lineages used to generate the antibody repertoire are subject to selective pressure (Pinheiro et al. 2013). The rabbit and hare species for which the expression of *VHn* has been shown, *Lepus europaeus*, are European leporids while *S. floridanus* is an American leporid (Chapman et al. 1980; Flux and Angermann 1990). Interestingly two related Caliciviruses affect the European rabbit and two hare species of Eurasian distribution (*L. europaeus* and *L. timidus*) but not cottontails (Abrantes et al. 2012; Lopes et al. 2013). Thus we can put forward the hypothesis that the different continents' pathogen community may impose different selective pressures and favour the expression of *VHn* only in European species.

Acknowledgements

This work was partially funded by FEDER (Fundo Europeu de Desenvolvimento Regional) funds through the Programa Operacional Factores de Competitividade (COMPETE program; FCOMP-01-0124-FEDER-028286) and Portuguese national funds through FCT (Fundação para a Ciência e a Tecnologia; research project PTDC/BIO-ANM/3963/2012)– Quadro de Referência Estratégico Nacional (QREN) funds from the European Social Fund and Portuguese Ministério da Educação e Ciência. FCT also supported the doctoral grant of AP (ref.: SFRH/BD/71252/2010) and the FCT Investigator grant of JA (ref.: IF/01396/2013), and the Post-Doctoral grant of JMF (ref.: SFRH/BPD/43264/2008). “Genomics Applied To Genetic Resources” co-financed by North Portugal Regional Operational Programme 2007/2013 (ON.2 – O Novo Norte), under the National Strategic Reference Framework (NSRF), through the European Regional Development Fund (ERDF), also supported this work.

References

- Abrantes J, van der Loo W, Le Pendu J, Esteves PJ (2012) Rabbit haemorrhagic disease (RHD) and rabbit haemorrhagic disease virus (RHDV): a review. *Vet Res*, 43(1):12.
- Alves PC, Hackländer K (2008) Lagomorph species: geographical distribution and conservation status. In: Alves PC, Ferrand N, Hackländer K (eds) *Lagomorph biology*. Springer, Berlin Heidelberg, pp 395–405
- Chapman JA, Hockman JG, Ojeda MM (1980) *Sylvilagus floridanus*. *Mamm Species* 136: 1-8.
- Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* 9 (8):772
- Drummond AJ, Ho SY, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biol* 4 (5):e88
- Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7:214
- Esteves PJ (2003) Molecular and population genetic analysis of polymorphism at the antibody loci IgGCH2 and IgVH in lagomorphs. Faculdade de Ciências Universidade do Porto, Porto (<http://hdl.handle.net/10216/10598>)
- Esteves PJ, Lanning D, Ferrand N, Knight KL, Zhai SK, van der Loo W (2004) Allelic variation at the VHa locus in natural populations of rabbit (*Oryctolagus cuniculus*, L.). *J Immunol* 172 (2):1044-1053

- Esteves PJ, Lanning D, Ferrand N, Knight KL, Zhai SK, van der Loo W (2005) The evolution of the immunoglobulin heavy chain variable region (IgVH) in Leporids: an unusual case of transspecies polymorphism. *Immunogenetics* 57 (11):874-882.
- Flajnik MF (2002) Comparative analyses of immunoglobulin genes: surprises and portents. *Nat Rev Immunol* 2 (9):688-698
- Flux JEC, Angermann R (1990) The Hares and Jackrabbits. In: JA Chapman, Flux JEC (eds) *Rabbits, Hares and Pikas: Status Survey and Conservation Action Plan*. The World Conservation Union, Gland, Switzerland, pp 61-94
- Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41, 95–98.
- Hornig WJ, Knight KL, Dray S (1976) Heavy chain variable region allotypic sub-specificities of rabbit immunoglobulins. I. Identification of three subpopulations of a1 IgG molecules. *J Immunol.* 116(1):117-125.
- Kim BS, Dray S (1973) Expression of the a, x, and y variable region genes of heavy chains among IgG, IgM, and IgA molecules of normal and a locus allotype-suppressed rabbits. *J Immunol* 111 (3):750-760
- Kindt TJ (1975) Rabbit immunoglobulin allotypes: structure, immunology, and genetics. *Adv Immunol* 21:35-86
- Knight KL (1992) Restricted VH gene usage and generation of antibody diversity in rabbit. *Annu Rev Immunol* 10:593-616
- Knight KL, Becker RS (1990) Molecular basis of the allelic inheritance of rabbit immunoglobulin VH allotypes: implications for the generation of antibody diversity. *Cell* 60 (6):963-970
- Lanning D, Sethupathi P, Rhee KJ, Zhai SK, Knight KL (2000a) Intestinal microflora and diversification of the rabbit antibody repertoire. *J Immunol* 165 (4):2012-2019.
- Lanning D, Zhu X, Zhai SK, Knight KL (2000b) Development of the antibody repertoire in rabbit: gut-associated lymphoid tissue, microbes, and selection. *Immunol Rev* 175:214-228
- Lefranc MP, Pommié C, Ruiz M, Giudicelli V, Foulquier E, Truong L, Thouvenin-Contet V, Lefranc G (2003) IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. *Dev Comp Immunol* 27(1):55-77
- Lopes AM, Gavier-Widén D, Le Gall-Reculé G, Esteves PJ, Abrantes J (2013) Complete coding sequences of European brown hare syndrome virus (EBHSV) strains isolated in 1982 in Sweden. *Arch Virol.* 158(10):2193-6

- Mage RG, Bernstein KE, McCartney-Francis N, Alexander CB, Young-Cooper GO, Padlan EA, Cohen GH (1984) The structural and genetic basis for expression of normal and latent VHa allotypes of the rabbit. *Mol Immunol* 21 (11):1067-1081
- Mage RG, Lanning D, Knight KL (2006) B cell and antibody repertoire development in rabbits: the requirement of gut-associated lymphoid tissues. *Dev Comp Immunol* 30 (1-2):137-153
- Margolies MN, Cannon LE, Kindt TJ, Fraser B (1977) The structural basis of rabbit VH allotypes: serologic studies on a1 H chains with defined amino acid sequence. *J Immunol* 119 (1):287-294
- Matthee CA, van Vuuren BJ, Bell D, Robinson TJ (2004) A molecular supermatrix of the rabbits and hares (Leporidae) allows for the identification of five intercontinental exchanges during the Miocene. *Syst Biol* 53 (3):433-447
- Pinheiro A, Lanning D, Alves PC, Mage RG, Knight KL, van der Loo W, Esteves PJ (2011) Molecular bases of genetic diversity and evolution of the immunoglobulin heavy chain variable region (IGHV) gene locus in leporids. *Immunogenetics* 63 (7):397-408
- Pinheiro A, de Mera IG, Alves PC, Gortázar C, de la Fuente J, Esteves PJ (2013) Sequencing of modern *Lepus* VDJ genes shows that the usage of VHn genes has been retained in both *Oryctolagus* and *Lepus* that diverged 12 million years ago. *Immunogenetics* 65(11):777-84
- Rambaut A, Drummond AJ (2007) Tracer v1.5. Available from <http://beast.bio.ed.ac.uk/Tracer>.
- Ros F, Puels J, Reichenberger N, van Schooten W, Buelow R, Platzer J (2004) Sequence analysis of 0.5 Mb of the rabbit germline immunoglobulin heavy chain locus. *Gene* 330:49-59
- Roux KH (1981) A fourth heavy chain variable region subgroup, w, with 2 variants defined by an induced auto-antiserum in the rabbit. *J Immunol* 127 (2):626-632
- Sehgal D, Obiakor H, Mage RG (2002) Distinct clonal Ig diversification patterns in young appendix compared to antigen-specific splenic clones. *J Immunol* 168 (11):5424-5433
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, and Kumar S (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol Biol Evol* 28: 2731-2739
- Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-

specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680

Tonnelle C, Cazenave PA, Brezin C, Moinier D, Fougereau M (1983) Structural correlates to the rabbit immunoglobulin heavy chain a100 allotype. *Mol Immunol* 20 (7):753-761

van der Loo W (1987) Population genetical studies on the adaptive significance of the immunoglobulin allotypes in wild rabbit. In: Dubiski S (eds) *The Rabbit in Contemporary Immunological Research*. Longman Scientific & Technical, John Wiley, pp 101-126

Vidus Rosin A, Gilio N, Meriggi A (2008) Introduced Lagomorphs as a threat to “native” Lagomorphs: the case of the Eastern cottontail (*Sylvilagus floridanus*) in northern Italy. In: Alves PC, Ferrand N, Hacklander K (eds) *Lagomorph biology. Evolution, ecology and conservation*. Springer, Eidenberg, pp 153–16

Chapter 3

IgA evolution in mammals

Paper 4

Computational analyses of an evolutionary arms race between mammalian immunity mediated by immunoglobulin A and its subversion by bacterial pathogens

Ana Pinheiro, Jenny M. Woof, Laurent Abi-Rached, Peter Parham¹ and Pedro J. Esteves

PLoS One, 8(9), e73934. doi: 10.1371/journal.pone.0073934. eCollection 2013.

Abstract

IgA is the predominant immunoglobulin isotype in mucosal tissues and external secretions, playing important roles both in defense against pathogens and in maintenance of commensal microbiota. Considering the complexity of its interactions with the surrounding environment, IgA is a likely target for diversifying or positive selection. To investigate this possibility, the action of natural selection on IgA was examined in depth with six different methods: CODEML from the PAML package and the SLAC, FEL, REL, MEME and FUBAR methods implemented in the Datamonkey webserver. In considering just primate IgA, these analyses show that diversifying selection targeted five positions of the C α 1 and C α 2 domains of IgA. Extending the analysis to include other mammals identified 18 positively selected sites: ten in C α 1, five in C α 2 and three in C α 3. All but one of these positions display variation in polarity

and charge. Their structural locations suggest they indirectly influence the conformation of sites on IgA that are critical for interaction with host IgA receptors and also with proteins produced by mucosal pathogens that prevent their elimination by IgA-mediated effector mechanisms. Demonstrating the plasticity of IgA in the evolution of different groups of mammals, only two of the eighteen selected positions in all mammals are included in the five selected positions in primates. That IgA residues subject to positive selection impact sites targeted both by host receptors and subversive pathogen ligands highlights the evolutionary arms race playing out between mammals and pathogens, and further emphasizes the importance of IgA in protection against mucosal pathogens.

Keywords

immunoglobulin A, positive selection, coevolution, Fc α RI, pIgR, mucosal pathogens

Introduction

Immunoglobulin A (IgA), in the form of dimers or higher polymers (pIgA) particularly tetramers, is the predominant immunoglobulin isotype in mucosal tissues and external secretions, where it provides a major line of defense against pathogens. In addition, it plays a major role in the maintenance of the commensal microbiota in the intestinal tract, where interplay between commensal microorganisms and IgA promotes a mutually beneficial co-existence [1]. Monomeric IgA is present in serum, being the second most prevalent immunoglobulin after IgG and a critical factor for eliminating pathogens that breach external surfaces [2]. Much energy is expended in producing these serum and mucosal forms of IgA. In humans, for example, more IgA is produced than all the other antibody isotypes combined. Such high investment in IgA is presumably indicative of the key contribution this antibody isotype makes to immune protection. Like all immunoglobulins, IgA displays a basic monomeric structure of two light and two heavy chains, each having a variable and a constant region, linked together by disulphide bridges. Each chain is organized in globular domains consisting of approximately 110-130 amino acids. The light chains (VL and CL domains) and the variable (VH) and first constant domain of the heavy chain (C α 1) constitute the two Fab regions, which bind antigens. The remaining constant domains of the heavy chain (C α 2 and C α 3) constitute the Fc region, responsible for the recruitment of mechanisms that lead to pathogen elimination. Linking the Fab and Fc regions is a flexible hinge region. This basic IgA unit can exist as monomers or be arranged into dimers (dIgA) and higher order multimers in which the monomers are linked by a J (joining) chain. In secretions, IgA is present as secretory IgA (S-IgA), a complex of dIgA or pIgA with another polypeptide chain, the secretory component (SC) [3], which confers some protection from proteolytic cleavage.

IgA has been identified in all mammals and birds studied [3]. In mammals, differences in gene number and molecular forms have been noted, defining different IgA systems. Most mammals have one *IGHA* gene, coding for one IgA isotype, which adopts a dimeric form in serum IgA. Humans, chimpanzees, gorillas and gibbons have, however, two *IGHA* genes, which arose by gene duplication in a common hominoid primate ancestor and code for the IgA1 and IgA2 [4] subclasses. In hominoids serum IgA is mainly monomeric. Rabbit has the most complex IgA system observed, with 13 *IGHA* genes encoding 13 IgA subclasses [5]: of these 13 subclasses, 11 are expressed and are differentially distributed among the mucosal tissues [6]. Mammalian IgA subclasses mainly differ in the length and amino acid sequence of the hinge, which

affects their susceptibility to cleavage by bacterial proteases [5], [7].

Elimination and destruction of pathogens is facilitated by the binding of Ig-antigen complexes to Ig receptors (FcRs) on effector cells and soluble effector molecules such as complement. In most mammals, IgA effector functions appear to be reliant on Fc α RI (CD89), the Fc receptor specific for IgA: binding of the IgA-antigen complex to Fc α RI can lead to phagocytosis, antibody dependent cell-mediated cytotoxicity (ADCC) and release of cytokines and inflammatory mediators. Fc α RI binds to IgA at the C α 2-C α 3 interface [8], [9] an interaction that has been suggested to evolve under pressure from pathogen decoy IgA-binding proteins [10]. Fc α RI appears to be functional in the majority of mammals, but it is notably absent from mice, rabbits and dogs due to either loss of the gene or to its degeneration into a pseudogene.

Other IgA-Fc receptors important for IgA function include the polymeric Ig receptor (pIgR) and the IgA/IgM Fc receptor (Fc α / μ R) [11]. The pIgR is responsible for delivery of the large quantities of pIgA produced in the mucosae across the epithelial cell layer into mucosal secretions. In the process, pIgR is cleaved to yield the SC, which remains covalently complexed with pIgA to form S-IgA. The binding involves interaction of pIgR with J chain and IgA-Fc residues, particularly within the C α 3 domain of IgA. Some of the residues involved are located in the C α 2-C α 3 interface [12] and overlap with residues critical for binding to Fc α RI and Fc α / μ R [3]. In addition to transport of free pIgA, pIgR can also transport polymeric IgA immune complexes, including pIgA complexed with viruses, out across the epithelium [2]. Moreover, pIgA transported via pIgR may intercept and neutralize certain viruses inside epithelial cells [2]. In humans, Fc α / μ R is present on macrophages and plasma cells, and also on follicular dendritic cells in tonsil and in intestinal tissues [11], likely reflecting a role in coordination of the immune response in mucosal tissues. The N-terminal Ig-binding domain of Fc α / μ R shares similarity with domain 1 of pIgR, and the modes of interaction with dIgA are presumed to have similar features. Consistent with this possibility, the results of mutagenesis mapping analysis indicate a critical role for the C α 2-C α 3 domain interface of the IgA heavy chain in the interaction [13].

To evade elimination by the immune system, numerous pathogens have evolved proteins targeting IgA. These include IgA-binding proteins, that by binding to IgA block its coupling to host IgA-receptors, as well as proteases that by cleaving the IgA hinge, uncouple the recognition of foreign antigens from the effector functions that eliminate them. Examples of the former include the Sir22 and Arp4 proteins of *Streptococcus pyogenes*, the β protein of *Streptococcus agalactiae*, and the SSL7 toxin of *Staphylococcus aureus*, all of which bind to residues lying in the C α 2-C α 3 interface

of IgA, thereby blocking access of Fc α RI [14], [15]. Examples of the latter include IgA1 proteases secreted by important bacterial pathogens, such as *Neisseria meningitidis* and *Haemophilus influenzae*, which cleave specifically in the hinge region of IgA1 of humans and great apes. IgA1 proteases are postproline endopeptidases that cleave at either Pro-Ser (type 1 enzymes) or Pro-Thr (type 2 enzymes) peptide bonds within the IgA1 hinge region. To achieve such specific cleavage, these enzymes recognize structural elements within the hinge [16], [17] and some of them also need to contact the Fc region before cleavage can occur [18], [19]. Notably, the type 2 IgA1 protease of *Neisseria meningitidis*, a causative agent of bacterial meningitis, interacts with the C α 3 residues of the C α 2-C α 3 interface also bound by Fc α RI, pIgR and Fc α / μ R, whereas the type 2 IgA1 protease of *Haemophilus influenzae* contacts a different set of C α 3 residues that are implicated in binding to pIgR [19].

Over recent years it has become increasingly apparent that S-IgA contributes to mucosal homeostasis through various mechanisms [20]. For example, coating of commensal bacteria by S-IgA may promote gut colonization and survival through biofilm formation. The role of S-IgA in maintaining the commensal microbiota may depend, at least in part, on interactions between IgA glycans and commensal bacteria [20].

Considering the complex interactions of IgA with other components of the immune system, with commensal microorganisms and the evasion proteins of diverse pathogens, IgA is a likely target for natural selection. Few studies have examined Ig sequences for the impact of natural selection and they have focused on IgA or IgG isotype in a limited number of vertebrate taxa [10], [21], [22]: for example, Abi-Rached et al [10] investigated the pattern of diversification of IgA-Fc using maximum likelihood [23], [24] and pairwise methods, with a focus on primates. To develop deeper understanding of the issue, in this study we took a broader approach that encompasses a wider range of methods and mammalian species. In total, 64 sequences from 28 species representing monotremes, marsupials and eight orders of placental mammals were included in the analyses.

Methods

Primate and mammalian IgA sequences

The complete sequences for primate IgAs used in a previous study [10] were obtained from GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>); accession numbers

are: Human IgA1 and IgA2 - J00220, J00221, M60192 and AJ012264; Chimpanzee IgA1 and IgA2 - X53702 and X53706; Gorilla IgA1 and IgA2 - X53703 and X53707; Gibbon IgA1 and IgA2 - X53708 and X53709; Orangutan IgA -X53704; Rhesus macaque IgA - AY039245 to AY039252, AY294614 and AY294615; Crab-eating macaque IgA - X53705 and Sooty mangabey IgA - AY544580 and AY544581.

Complete sequences for non-primate mammalian IgAs were obtained from IMGT (<http://www.imgt.org/IMGTrepertoire/>), GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>) and Ensembl (<http://www.ensembl.org/index.htm>). In total 64 sequences from 28 species were included in the analyses, representing marsupials, monotremes, and eight orders of placental mammals: primates, artiodactyls, perissodactyls, rodents, carnivores, lagomorphs, chiropters and cetaceans. Accession numbers for the non-primate sequences used are: Cattle IgA - AF109167; Sheep IgA - AF024645; Pig IgA - U12594; Horse IgA - AY247966; Alpaca IgA - AM773729; Mouse IgA - J00475, AF175973 to AF175975, AH011154 to AH011156, and AY045750 to AY045752; Rat - ENSRNOT00000006888 and AY158661; Dog IgA - L36871; Panda IgA - AY818387; Rabbit IgA1 to 13 - X51647, X82108 to X82119; Big brown bat IgA - HM134938; Little brown bat IgA - HM134924; Short-nosed fruit bat IgA - HM134948; Black flying fox IgA - GQ427150; Dolphin IgA - AY621035; Possum IgA - AF091139 and AF027382; Opossum IgA - AF108225 and AF012110; Tasmanian devil - AFEY01402156; Platypus IgA1 and IgA2 - AY055778 and AY055779; Echidna IgA - AF416951. Excluded from the analysis was the recombinant human IgA2(n) allele [25] and the mouse IgA*3 allele for which the sequence has a nucleotide deletion in C α 1 that is presumably a sequencing or typographical error. The monotreme and marsupial C α 1 sequences were not included because of uncertainty in their alignment with the C α 1 domain sequences of placental mammals and also to avoid the risk of saturation that could result from including these highly divergent sequences; likewise, sequences of the rapidly-evolving IgA hinge region were excluded from the analysis.

For the analysis of the primate datasets and of the placental mammal C α 1 dataset, sequences were aligned using CLUSTAL W [27] as implemented in BioEdit [28], and corrected manually; notably, adjustments were made to follow the rigorous IMGT numbering system. For the mammalian C α 2 and C α 3 datasets, amino acid alignments were first generated using MUSCLE [29] and manual corrections, and these alignments were then used as a guide to prepare codon alignments for the same set of sequences.

Codon numbering is according to the Bur IgA1 numbering. IMGT unique numbering for C-DOMAIN [26] is also shown in parenthesis.

Codon-based analyses of positive diversifying selection

To investigate positive selection on IgA, we studied the three constant domains (C α 1, C α 2 and C α 3) separately: for each domain we compared the rate per-site of nonsynonymous substitution (dN) to the rate per-site of synonymous substitutions (dS) in a maximum likelihood (ML) framework, using six different methods. Since each method has strengths and weaknesses, we used the approach of Wlasiuk and Nachman [30] to identify the codons for which the signal of positive selection was strongest: only codons identified by at least two of the ML methods were considered to be positively selected codons (PSC). Unlike pairwise dN/dS analyses, the methods used here rely on phylogenetic approaches and are thus not as sensitive as the pairwise dN/dS methods to differences in the number of sequences present in the taxonomic groups investigated: to increase the resolution of the analysis, we included all available sequences.

We first compared two alternative models implemented in CODEML (PAML 4.4) [23], [24]: M8, which allows for codons to evolve under positive selection (dN/dS>1) and M7, which does not (dN/dS $\leq$ 1). These two nested models were compared using a likelihood ratio test (LRT) with 2 degrees of freedom [31], [32]. The analysis was run twice, and conducted with the F3x4 model of codon frequencies. Codons under positive selection for model M8 were identified using a Bayes Empirical Bayes approach (BEB) [33] and considering a posterior probability of > 90%. For each analysis, a Neighbour-Joining phylogenetic tree was used as the 'working topology', and generated using Mega 5 [34] with the p-distance substitution model and the complete deletion option to handle gaps and missing data. Overall, the tree topologies used reflected the accepted topology for mammals.

We also used the five methods for detecting positive selection available from the DATAMONKEY web server [35]: the Single Likelihood Ancestor Counting model (SLAC), the Fixed Effect Likelihood model (FEL), the Random Effect Likelihood model (REL), the Mixed Effects Model of Evolution (MEME) and the Fast Unbiased Bayesian Approximation (FUBAR). For these analyses, the best fitting nucleotide substitution model was determined through the automatic model selection tool available on the server.

Because recombination can contribute to false inference of positive selection, causing a high rate of false positive detection [36], [37], [38], all datasets were screened for recombination using the GARD [39] method from the DATAMONKEY web server [35]. No evidence of recombination was found.

Location of the PSC in structural models of IgA

A molecular model of human IgA1 (MMDB ID: 10546, PDB ID:1iga [40]) and the three-dimensional X-ray crystal structure of human IgA1-Fc (PDB ID :1OW0 [9]) were used to map the amino acids encoded by PSC onto 3D structures of the protein. To investigate their relation to putative sites of interest, the sites of interaction with host receptors (Fc α RI, plgR and Fc α / μ R [3], [8], [9], [12], [13]) and bacterial proteins (*S. aureus* SSL7 protein, streptococcal IgA binding proteins, *N. meningitidis* and *H. influenzae* type 2 IgA1 proteases [3], [14], [19], [41]) were also mapped onto the 3D structure. For this purpose the NCBI application Cn3D 4.1 (<http://www.ncbi.nlm.nih.gov/Structure/CN3D/cn3d.shtml> [42]) and iMol software [43] were used. Although the molecular model of human IgA1 has the drawbacks of being based on low resolution X-ray and neutron scattering data and of using the X-ray crystal structure of IgG to model the Fc part of IgA (the IgA Fc structure was unavailable at the time), it offered the best means to visualize all PSC in one intact structure. The solved X-ray crystal structure of human IgA1-Fc offers a higher resolution view, and aids understanding of the putative impact of these PSCs on the IgA-Fc ligand interaction.

Results

Natural selection diversified the C α 1 and C α 2 domains of primate IgA sequences

Using the ML approach of PAML [23], [24], evidence for positive diversifying selection was obtained in primates for two of the three IgA constant domains, C α 1 and C α 2, with the model allowing sites to evolve under positive selection (M8) showing a significantly better fit than the model that did not (M7) ($\alpha=0.01-0.05$; Table 1). The other five ML methods also identified positively selected sites for IgA C α 1 and C α 2 but not for IgA C α 3. Comparison of the sites characterised by each method reveals five codons supported with high confidence ($p>0.9$) by at least two methods: of these five positively selected codons (PSC), two are in the C α 1 domain, codons 133 and 166 (C α 1-10 and 45.2), and three others are in the C α 2 domain, codons 296, 319 and 326 (C α 2-84, 100 and 107). Natural amino acid variability and characteristics for each of these codons are given in Table 2: for four of the five positions (133, 166, 319, and 326), changes in amino acid characteristics such as polarity and charge were observed, with potential to alter the protein structure or capacity for protein–protein interaction.

Table 1 - Phylogenetic tests of positive selection in primates

Test of selection		Sites under selection identified by different methods ^a										
IgA dataset	InL M7 (neutral)/InL M8 (selection)	-2lnΔLc	Significance	p _ω , ω _s ^b	PAML M8 (p>0.5; Abi-Rached et al, 2007)	PAML M8 (p>0.9; present study)	SLAC	FEL	REL	MEME	FUBAR	PSC ^c
CH1	-1212.4/-1200.0	24.8	** (p<0.01)	0.03, 3.86	n.a.	133, 166, 197	none	none	133, 137, 166	133, 136, 138, 165	133	133, 166
CH2	-1365.3/-1361.3	8.0	* (p<0.05)	0.02, 2.6	245, 296, 317, 319, 326, 331, 333	319, 326	none	245, 296, 319	273, 293, 296, 317	296, 324, 326, 330	319	296, 319, 326
CH3	-907.08/-907.08	0.0	n.s.	0.00, 1.0	none	none	none	none	None	none	none	-

^aCodons identified by more than one ML method are underlined.

^bp_ω = proportion of the sites under selection, ω_s = estimated dN/dS of the sites under selection in M8.

^cPositively Selected Codons: only the codons identified by at least two of the ML methods were considered to be positively selected codons.

doi:10.1371/journal.pone.0073934.t001

Table 2 - Characterization of natural amino acid variation for each residue identified under positive selection in primate IgA.

Residue ^a	Functional information	Natural amino acid variation			
		H, + ^b	H, - ^b	H, n ^b	HY, n ^b
Primates Cα1					
133	Close to hinge	K	D, E	–	C
166	Exposed	R, K, H	–	S, Q, T	P, I, L
Primates Cα2					
296	Exposed	R, H	–	–	–
319	Exposed; close to SSL7 binding site	K	E	Q	V
326	Exposed	–	E	N	A

^aIgA1 Bur numbering.

^bamino acid characteristics: H- Hydrophilic, HY- Hydrophobic, + positive, – negative, n-neutral.

doi:10.1371/journal.pone.0073934.t002

Mammalian IgA evolution was marked by diversifying selection on the three constant domains

Analysis of IgAs from a much broader range and larger sample of mammals using the ML approach implemented in PAML [23], [24] revealed significant evidence of diversifying selection for two of the three domains investigated (Cα1 and Cα2) (Table 3). However, the other five ML methods clearly identified positively selected sites for all three domains, which is consistent with the three domains having been the targets of diversifying selection. Comparison of the positively selected sites identified by each of the methods led to the identification of eighteen well supported positively selected codons (PSC) (Table 3). Of these eighteen PSC, ten locate to the Cα1 domain, five to the Cα2 domain, and three to the Cα3 domain (Figure 1). All but one of these residues show variations in both polarity and charge, changes that could alter the protein structure or capacity for protein–protein interaction (Table 4); the exception, PSC 431 (Cα3-103), displays a restricted set of residues that share the same polarity and charge, suggesting these characteristics are of value at this position. Of note, two of the changes at PSC in the Cα1 domain, codons 166 and 213 (Cα1-45.2 and 116 IMGT numbering) can generate putative N-glycosylation sites, which could affect protein function: residue 166 is a known N-glycosylation site of primate IgA2 , as well as sheep, panda and alpaca IgA and rabbit IgA7, IgA8, IgA11 and IgA13. In contrast the putative N-glycosylation site at residue 213 appears only in rabbit IgA7, IgA8, IgA11 and IgA13.

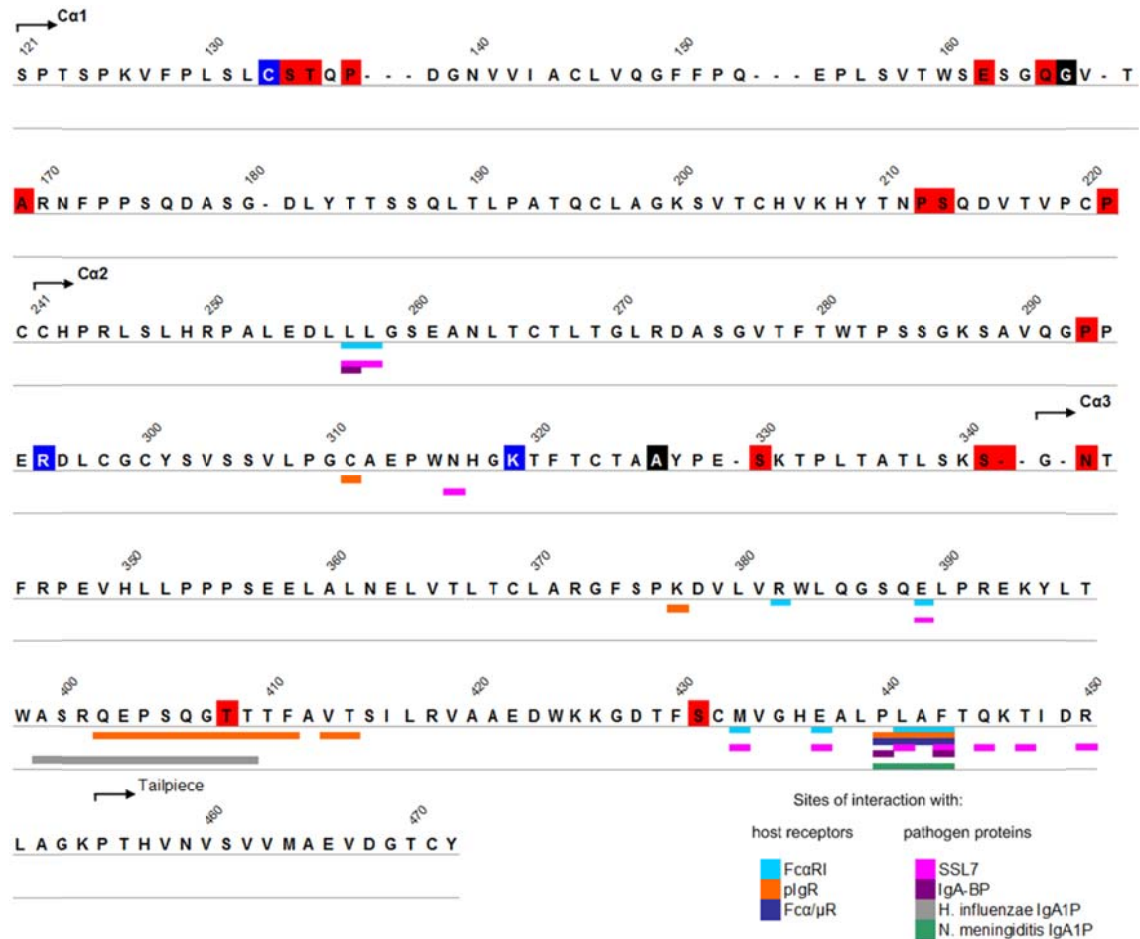


Figure 1. Residues encoded by PSC in human IgA1 Cα1, Cα2 and Cα3 domains. Residues encoded by PSC are highlighted in red (selected in mammals), black (selected in primates and mammals) or blue (selected in primates). Sites of interaction with studied ligands are indicated below the sequence according to the colour panel on the figure. Bur numbering is shown above.

Table 3 - Phylogenetic tests of positive selection in mammals

Test of selection		Sites under selection identified by different methods ^a									
Dataset	InL M7 (neutral)/InL M8 (selection)	-2lnΔLc	Significance	p _{sr} ω _s ^b	PAML M8	SLAC	FEL	REL	MEME	FUBAR	PSC ^c
Placental mammals IgACH1	-6035.6/-6010.7	49.8	** (p<0.01)	0.19, 1.79	121, 126, 134, 135, 136, 137, 138, 139, 157, 162, 166, 170, 194, 198, 199, 200, 201, 217, 221	162, 165, 169, 212, 213	135, 162, 165, 169, 212, 213, 221	221	124, 133, 134, 135, 137, 148, 162, 166, 169, 212, 213, 219, 221	135, 169	134, 135, 137, 162, 165, 166, 169, 212, 213, 221
Mammals IgACH2	-6075.9/-6068.9	14	** (p<0.01)	0.06, 1.34	243, 282, 293, 317, 330, 341a	330	341, 341a	284, 292, 293, 326, 341, 341a	242, 264, 285, 294, 326, 330, 337, 340, 341, 341a	341a	293, 326, 330, 341, 341a
Mammals IgACH3	-5674.91/-5672.44	4.93	n.s. (p<0.1)	0.00, 1.00	none	343, 364, 408	343	431	343, 389, 394, 408, 423, 427, 431, 442, 451	343	343, 408, 431

^aCodons identified by more than one ML method are underlined.^b ω_s = proportion of the sites under selection, ω_s = estimated dN/dS of the sites under selection in M8.^cPositively Selected Codons: only the codons identified by at least two of the ML methods were considered to be positively selected codons.

doi:10.1371/journal.pone.0073934.t003

Table 4 - Characterization of natural amino acid variation for each residue identified under positive selection in mammal IgA.

Residue ^a	Functional Information	Natural amino acid variation			
		H, + ^b	H, - ^b	H, n ^b	HY, n ^b
Placental mammal Cα1					
134	Close to hinge	H, R, K	–	N, Q, S, C, Y	P, L, I, G
135	Close to hinge	H, R	E, D	T, N, Q, S, C	P, V, A, L, I, G
137	Close to hinge	–	D	Q, S, C	P, L, A
162	Exposed	K, H	E, D	Q, T, S	A, P, I, V, G, L
165	Exposed	K	E, D	Q, S	A, P
166	Exposed	–	D	N, S	G, V
169	Exposed	–	D	N, T, S	A, V, I, G
212	Orientated towards variable domains, exposed	K	E	T, S	P, A, V, I
213	Orientated towards variable domains, exposed	R, H	D	N, S	V, G, I, L
221	Bordering hinge	R, K, H	–	S, Q, T	P, I, L
Mammals Cα2					
293	Quite close to hinge, exposed	R, K	E	N, T, S	P, A, L
326	Exposed	K	E	N, T, S	A
330	Exposed	K	E,D	T, S	L, V, A, F
341	Positioned on strand linking Cα2 and Cα3; vicinity of SSL7 binding site	–	D	T, S	P, V, A, L, I
341a	Positioned on strand linking Cα2 and Cα3; vicinity of SSL7 binding site	K, R	–	T, S	L
Mammals Cα3					
343	Positioned on strand linking Cα2 and Cα3; vicinity of SSL7, plgR, and <i>H.influenzae</i> IgA1P binding site	–	E, D	N, T, S	P, V, A, I
408	plgR, and <i>H.influenzae</i> IgA1P binding site	K	–	T, S	I, P, A, V, G
431	Positioned in middle of β-strand, exposed	–	–	N, T, S	–

^aIgA1 Bur numbering.

^bamino acid characteristics: H- Hydrophilic, HY- Hydrophobic, + positive, –negative, n-neutral.
doi:10.1371/journal.pone.0073934.t004

The recently developed MEME methodology [44] can identify both episodic and persistent positive selection, because it allows the distribution of the dN/dS ratio to vary from site to site and also from branch to branch at a site. The additional positively-selected codons identified by MEME and not by the other approaches, are likely to have been subject to episodes of positive selection. Consistent with this interpretation, of 6 such sites detected by MEME in the Cα3 domain residues 389 and 442 are sites targeted by pathogenic IgA-binding proteins (Cα3- 45.2 and 115 IMGT numbering).

Positively selected codons are located near sites of interaction with ligands and bacterial proteases

To understand better the possible biological significance of the detected PSC, we mapped the residues they encode onto a molecular model of human IgA1 and the three-dimensional X-ray crystal structure of IgA1-Fc, along with sites of interaction for host receptors and bacterial proteins (Figure 2). Remarkably, more than half (13 out of 21) of the PSC encode residues located near sites of interaction with ligands and

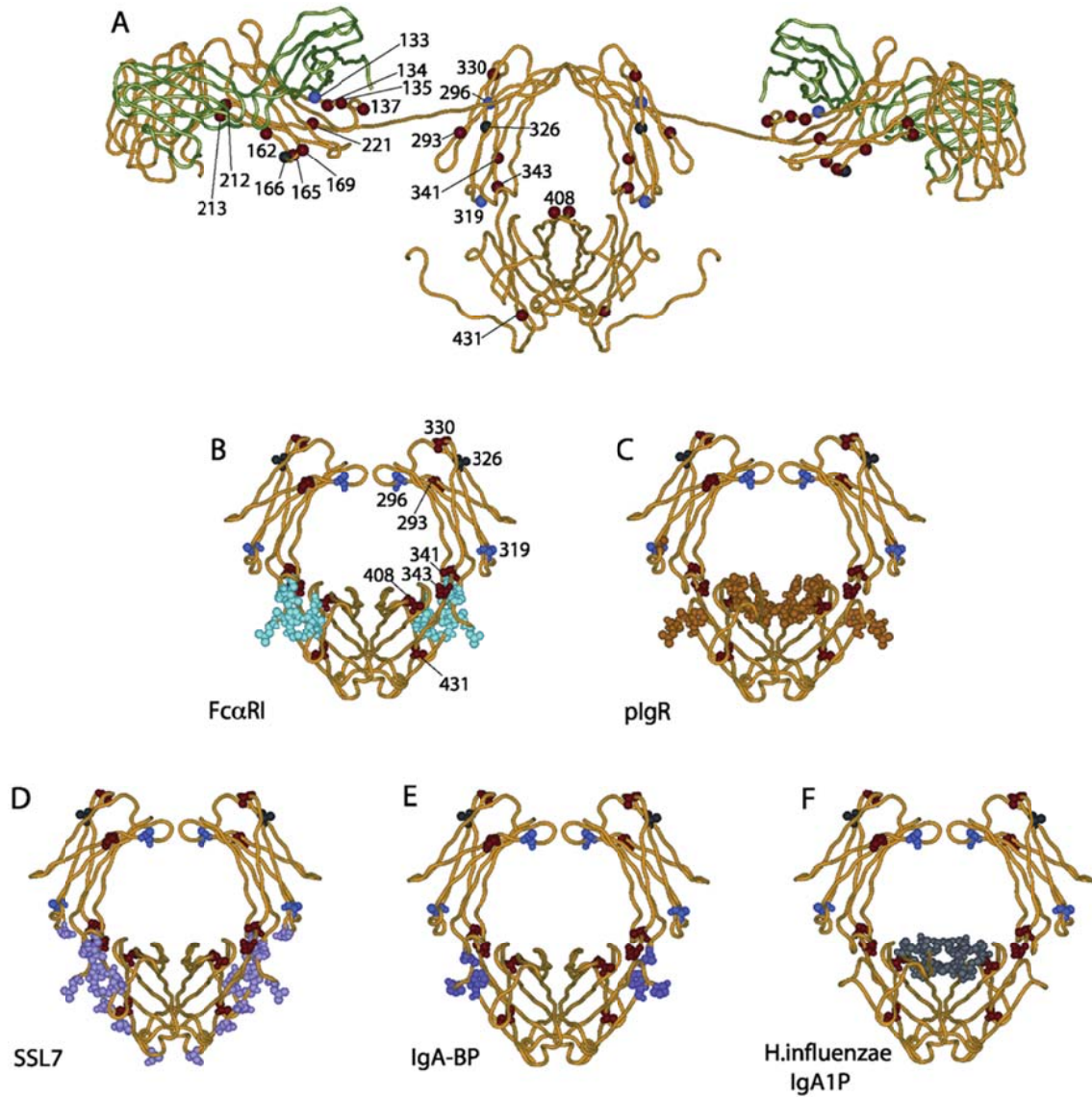


Figure 2. Residues encoded by PSC in the three-dimensional structure of human IgA1 and their relationship to sites of biological interest. **A**, Model of human IgA1 (PDB ID 1iga) with residues encoded by PSC highlighted. The light chains are colored green and the heavy chains colored yellow. Positively selected positions are represented by red dots (selected in mammals), black dots (selected in primates and mammals) or blue dots (selected in primates). **B to F**, Human IgA1-Fc (PDB ID 1OW0) with residues critical for FcαRI interaction (**B**), pIgR interaction (**C**), SSL7 interaction (**D**), streptococci IgA binding proteins interaction (**E**), and *H. influenzae* type 2 IgA1 protease interaction (**F**) highlighted in cyan, orange, pink, purple and grey respectively. Positively selected positions are represented as in panel A. Differences between the Fc structures in **A** and **B-F** reflect the fact that the model of intact IgA1 in **A** is based on low resolution X-ray and neutron scattering modeled based on the X-ray crystal structure of IgG Fc (the closest structure available at the time of modeling), while **B-F** show an X-ray crystal structure of human IgA1-Fc solved as part of a complex with FcαRI (not shown).

bacterial proteases. C α 1 residues 133, 134, 135, 137 and 221 and C α 2 residues 293 and 296 (C α 1-10, 11, 12, 14 and 124 and C α 2-81 and 84, IMGT numbering) are near the hinge region, the preferential target region for some IgA1-specific bacterial proteases. C α 1 residues 212 and 213 (C α 1-115 and 116) have a general orientation towards the variable domains involved in antigen recognition. C α 2 residues 341 and 341a (C α 2-124 and 125) and C α 3 residue 343 (C α 3-1.3) are part of the exposed strand linking the C α 2 and C α 3 domains of IgA1, in the vicinity of the C α 2 NH motif that participates in the binding of *S. aureus* SSL7 molecules to human IgA [25]. Residue 343 also lies close to the putative interaction site for pIgR [12] and a region important for interaction with the type 2 IgA1 protease of *H. influenzae* [19]. Residue 408 (C α 3-85.5) is one of several C α 3 domain residues of human IgA1 that directly influence binding to pIgR; it also lies adjacent to the site where the *H. influenzae* IgA1 protease is believed to bind. Although position 431 (C α 3-103) in the C α 3 domain is positively selected, its location in the IgA molecule is not close to any known interaction sites of IgA-Fc region. Substitutions at this position may exert functional effect by indirectly influencing the conformation of one or more of the interaction sites.

Discussion

Genes involved in host-pathogen interactions are prone to diversifying selection [45], [46]. As pathogens continuously evolve mechanisms to evade host defenses and cause infectious diseases, so must host species evolve counter defense mechanisms if they are to survive. This never-ending arms race subjects those components of the mammalian immune system that recognize pathogens and their products to strong varying selection. IgA, the main Ig isotype present in external secretions and at mucosal surfaces, is uniquely exposed to a wide variety of bacteria, viruses, fungi and other infectious microorganisms, which together exert strong selective pressures on this immunoglobulin isotype. The results obtained in this study demonstrate the considerable impact that positive selection has played in the evolution of IgA in mammals and in the diversity and divergence of IgA among extant mammalian species.

Natural selection diversified IgA in mammals

Consistent with the study of Abi-Rached and coworkers [10], our analysis shows that the C α 2 domain of primate IgA-Fc exhibits evidence of positive diversifying

selection, and the C α 3 domain does not. Making use of six different and complementary methodologies to identify positively selected residues, three C α 2 codons were identified by at least two of the methods used (positions 296, 319 and 326). These three positions also correspond to three of the seven codons identified previously [10] as being positively selected (Table 1). In contrast, the other four positions previously found to be positively selected did not reach the cutoffs for detection used here, even though two of them appeared in individual analyses (positions 245 and 317, Table 1). Because the goals of the two studies were different (sensitive detection in the earlier study versus detection of positions with the strongest signals for selection here), different cutoffs were applied. To reconcile the apparent discrepancies will require analysis of a much larger dataset of IgA sequences.

To develop deeper understanding of IgA evolution, we compared IgA in a broad range of mammalian species. Of eighteen positions selected during mammalian evolution, only two are included in the five positions selected during primate evolution. This difference vividly illustrates the evolutionary plasticity of IgA.

We find that diversifying selection has mainly targeted the C α 1 and C α 2 domains of IgA, and to lesser extent the C α 3 domain. Thus only three of the eighteen selected positions are in the C α 3 domain. One of these, position 431, exhibits relatively conservative variation, having only three alternative amino acids, with similar polarity and charge. The C α 3 domain, along with J chain, plays a key role in binding of pIgA to pIgR. C α 3 is also the main domain of IgA involved in binding to the major IgA-Fc receptors Fc α RI and Fc α / μ R. These crucial roles, along with contributions to the assembly and polymerization of IgA, can explain why C α 3 is the most conserved of the constant region of the IgA heavy chain. In contrast, the ten PSCs detected in C α 1 show more variety in amino acid substitutions, including changes in polarity and charge. Such variation modulates the C α 1 structure, with potential impact on Fab conformation, the antigen-binding site and the hinge region. Substitution at residues 166 and 213 could introduce an additional N-glycosylation site since the corresponding putative site is present in primate IgA2, sheep, panda and alpaca IgA, and some rabbit IgA subclasses. N-linked glycans in the Fab region are known to influence antigen binding, either by increasing affinity for antigen or blocking antigen binding [47]. Since IgA-Fc N-linked glycans could protect IgA from cleavage by bacterial and other proteases [18], we speculate that Fab N-linked glycans can also contribute to such protection from proteases. Furthermore, glycans may impact on interactions of S-IgA with commensal microorganisms, having an effect on the make-up of the microbiota and homeostasis of the gut [1], [20].

IgA diversification in mammals targets sites involved in the interaction with ligands and bacterial proteases

Mapping positively selected sites onto the structures of IgA and IgA-Fc revealed their likely impact on IgA function. Seven such sites, residues 133, 134, 135, 137 and 221 and C α 2 residues 293 and 296, are near the hinge, which links the antigen-recognition function of the Fab arms to the effector-recognition function of the Fc region. Because it is accessible, flexible and essential for antibody function, the hinge is a preferred target for bacterial proteases [3], [48]. Hinge structure varies considerably across mammalian species and between different subclasses and allotypes. For example, the hinge of hominoid IgA1 is 16 amino-acids longer than that of IgA2 and much more susceptible to proteolytic cleavage. The possible advantage of the longer hinge in IgA1 is its greater flexibility and potential for cross-linking antigens on the surface of bacteria and other pathogens [2]. Longer hinges are also a feature of most rabbit IgA subclasses. Thus, any variation that confers protection of the IgA hinges from proteolytic attack is a likely candidate for positive selection. For IgA1 proteases that cleave specifically in the hinge of hominoid IgA1 the distance of the susceptible peptide bond in the hinge from the "top" of the Fc (where the heavy chain enters the globular C α 2 domain) is critical for efficient cleavage [Senior, 2005 #1561]. Indeed, the crystal structure of a bacterial IgA1 protease from *H. influenzae* suggests that an intricate and coordinated association of protease with IgA is essential for optimal orientation of the hinge into the enzyme's active site [49]. Substitutions at residues in and around the hinge could therefore increase resistance to proteolytic attack and become targets for positive selection.

Three positively-selected residues are found near sites of human IgA1 that interact with Fc receptors and bacterial proteins. Residues 341, 341a and 343 are in the strand linking the C α 2 and C α 3 domains, in the vicinity of the C α 2 asparagine-histidine motif that participates in the binding of *S. aureus* SSL7 molecule to human IgA [25]. The CH2-CH3 interface is central to the binding of IgA to several classes of Fc receptor including Fc α RI, Fc α / μ R and pIgR [8], [9], [12], [13], and is also the target of pathogenic mechanisms to obstruct IgA function [14], [15], [19]. Variation at the CH2-CH3 interface could prove adaptive, either by improving the binding of IgA to its Fc receptors or hampering the binding of pathogen decoy molecules, or by achieving both of these effects. Such adaptations could be accomplished by changes in the residues that contact Fc receptors or decoy proteins and also in nearby residues that have conformational impact. Residues under positive selection have been described in the Cy2-Cy3 interface of IgG in leporids [21]. Residue 408, is one of the positively selected

C α 3 residues implicated in the binding of human IgA1 to pIgR [12] and the type 2 IgA1 protease of *H. influenzae*. Substitution at position 408 could therefore provide protection from cleavage by this IgA protease. The results of mutagenesis experiments are consistent with this possibility [19].

MEME methodology, which detects both episodic and persistent positive selection, identified codons in all three IgA analysed domains that were not revealed by the methods detecting only persistent selection. Thus these PSC are candidates for being subject to episodic selection. Among them are residues 389 and 442 in C α 3 that are targets for pathogenic IgA-binding proteins. Residue 442, which was previously shown to be subject to episodes of diversifying selection [10], is a site of N-linked oligosaccharide for IgA in mice. The glycan attached at asparagine 442 of mouse IgA hinders interaction with the *S. aureus* SSL7 decoy protein, but does not affect the binding of IgA to pIgR [50].

In conclusion, this study identified residues under positive selection in all three IgA heavy chain constant region domains. The majority of the identified residues are located in parts of the molecule that are essential for the functions of IgA in resistance to pathogens. This correlation is consistent with the positively-selected residues having influences on the interactions of IgA with immune-system receptors and the microbial proteins that interfere with these interactions. Future functional analyses should determine the mechanisms by which the positively selected residues exert their effect. Such knowledge could assist the design of therapeutic IgA-based monoclonal antibodies that are not susceptible to the pathogenic proteins that obstruct the defense functions of IgA.

Author Contributions

Conceived and designed the experiments: AP PJE. Analyzed the data: AP JMW LAR. Wrote the paper: AP JMW LAR PP PJE.

References

1. Slack E, Balmer ML, Fritz JH, Hapfelmeier S (2012) Functional flexibility of intestinal IgA – broadening the fine line. *Front Immunol* 3:100.
2. Woof JM, Kerr MA (2006) The function of immunoglobulin A in immunity. *J Pathol* 208: 270-282.

3. Woof JM, Russell MW (2011) Structure and function relationships in IgA. *Mucosal Immunol* 4: 590-597.
4. Kawamura S, Saitou N, Ueda S (1992) Concerted evolution of the primate immunoglobulin alpha-gene through gene conversion. *J Biol Chem* 267: 7359-7367.
5. Burnett RC, Hanly WC, Zhai SK, Knight KL (1989) The IgA heavy-chain gene family in rabbit: cloning and sequence analysis of 13 C alpha genes. *EMBO J* 8: 4041-4047.
6. Spieker-Polet H, Yam PC, Knight KL (1993) Differential expression of 13 IgA-heavy chain genes in rabbit lymphoid tissues. *J Immunol* 150: 5457-5465.
7. Schroeder HW, Jr., Cavacini L (2010) Structure and function of immunoglobulins. *J Allergy Clin Immunol* 125: S41-52.
8. Pleass RJ, Dunlop JI, Anderson CM, Woof JM (1999) Identification of residues in the CH2/CH3 domain interface of IgA essential for interaction with the human Fc alpha receptor (Fc alpha R) CD89. *J Biol Chem* 274: 23508-23514.
9. Herr AB, Ballister ER, Bjorkman PJ (2003) Insights into IgA-mediated immune responses from the crystal structures of human Fc alpha RI and its complex with IgA1-Fc. *Nature* 423: 614-620.
10. Abi-Rached L, Dorigi K, Norman PJ, Yawata M, Parham P (2007) Episodes of natural selection shaped the interactions of IgA-Fc with Fc alpha RI and bacterial decoy proteins. *J Immunol* 178: 7943-7954.
11. Monteiro RC, Van De Winkel JG (2003) IgA Fc receptors. *Annu Rev Immunol* 21: 177-204.
12. Lewis MJ, Pleass RJ, Batten MR, Atkin JD, Woof JM (2005) Structural requirements for the interaction of human IgA with the human polymeric Ig receptor. *J Immunol* 175: 6694-6701.
13. Ghumra A, Shi J, McIntosh RS, Rasmussen IB, Braathen R, et al. (2009) Structural requirements for the interaction of human IgM and IgA with the human Fc alpha /mu receptor. *Eur J Immunol* 39: 1147-1156.
14. Pleass RJ, Areschoug T, Lindahl G, Woof JM (2001) Streptococcal IgA-binding proteins bind in the C alpha 2-C alpha 3 interdomain region and inhibit binding of IgA to human CD89. *J Biol Chem* 276: 8197-8204.
15. Wines BD, Willoughby N, Fraser JD, {Lewis, 2008 #1541} PM (2006) A competitive mechanism for staphylococcal toxin SSL7 inhibiting the leukocyte IgA receptor, Fc alpha RI, is revealed by SSL7 binding at the C alpha 2/C alpha 3 interface of IgA. *J Biol Chem* 281: 1389-1393.
16. Batten MR, Senior BW, Kilian M, Woof JM (2003) Amino acid sequence requirements in the hinge of human immunoglobulin A1 (IgA1) for cleavage by streptococcal IgA1 proteases. *Infect Immun* 71: 1462-1469.

17. Senior BW, Woof JM (2005) Effect of mutations in the human immunoglobulin A1 (IgA1) hinge on its susceptibility to cleavage by diverse bacterial IgA1 proteases. *Infect Immun* 73: 1515-1522.
18. Chintalacharuvu KR, Chuang PD, Dragoman A, Fernandez CZ, Qiu J, et al. (2003) Cleavage of the human immunoglobulin A1 (IgA1) hinge region by IgA1 proteases requires structures in the Fc region of IgA. *Infect Immun* 71: 2563-2570.
19. Senior BW, Woof JM (2006) Sites in the CH3 domain of human IgA1 that influence sensitivity to bacterial IgA1 proteases. *J Immunol* 177: 3913-3919.
20. Mantis NJ, Rol N, Corthesy B (2011) Secretory IgA's complex roles in immunity and mucosal homeostasis in the gut. *Mucosal Immunol* 4: 603-611.
21. Esteves PJ, Alves PC, Ferrand N, van der Loo W (2002) Hotspot variation at the CH2-CH3 interface of leporid IgG antibodies (*Oryctolagus*, *Sylvilagus* and *Lepus*). *Eur J Immunogenet* 29: 529-535.
22. Esteves PJ, Carmo C, Godinho R, van der Loo W (2006) Genetic diversity at the hinge region of the unique immunoglobulin heavy gamma (IGHG) gene in leporids (*Oryctolagus*, *Sylvilagus* and *Lepus*). *Int J Immunogenet* 33: 171-177.
23. Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* 13: 555-556.
24. Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24: 1586-1591.
25. Chintalacharuvu KR, Raines M, Morrison SL (1994) Divergence of human alpha-chain constant region gene sequences. A novel recombinant alpha 2 gene. *J Immunol* 152: 5299-5304.
26. Lefranc MP, Pommie C, Kaas Q, Duprat E, Bosc N, et al. (2005) IMGT unique numbering for immunoglobulin and T cell receptor constant domains and Ig superfamily C-like domains. *Dev Comp Immunol* 29: 185-203.
27. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22(22):4673-80.28.
28. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser (Oxf)*. 41:95-9829.
29. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32(5):1792-7.
30. Wlasiuk G, Nachman MW (2010) Adaptation and constraint at Toll-like receptors in primates. *Mol Biol Evol* 27: 2172-2186.

31. Nielsen R, Yang Z (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 148: 929-936.
32. Yang Z, Swanson WJ, Vacquier VD (2000) Maximum-likelihood analysis of molecular adaptation in abalone sperm lysin reveals variable selective pressures among lineages and sites. *Mol Biol Evol* 17: 1446-1455.
33. Yang Z, Wong WS, Nielsen R (2005) Bayes empirical bayes inference of amino acid sites under positive selection. *Mol Biol Evol* 22: 1107-1118.
34. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28: 2731-2739.
35. Pond SL, Frost SD (2005) Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* 21: 2531-2533.
36. Anisimova M, Nielsen R, Yang Z (2003) Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. *Genetics* 164: 1229-1236.
37. Scheffler K, Martin DP, Seoighe C (2006) Robust inference of positive selection from recombining coding sequences. *Bioinformatics* 22: 2493-2499.
38. Shriner D, Nickle DC, Jensen MA, Mullins JI (2003) Potential impact of recombination on sitewise approaches for detecting positive natural selection. *Genet Res* 81: 115-121.
39. Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD (2006) Automated phylogenetic detection of recombination using a genetic algorithm. *Mol Biol Evol* 23: 1891-1901.
40. Boehm MK, Woof JM, Kerr MA, Perkins SJ (1999) The Fab and Fc fragments of IgA1 exhibit a different arrangement from that in IgG: a study by X-ray and neutron solution scattering and homology modelling. *J Mol Biol* 286: 1421-1447.
41. Ramsland PA, Willoughby N, Trist HM, Farrugia W, Hogarth PM, et al. (2007) Structural basis for evasion of IgA immunity by *Staphylococcus aureus* revealed in the complex of SSL7 with Fc of human IgA1. *Proc Natl Acad Sci U S A* 104: 15051-15056.
42. Wang Y, Geer LY, Chappay C, Kans JA, Bryant SH (2000) Cn3D: sequence and structure views for Entrez. *Trends Biochem Sci* 25: 300-302.
43. Rotkiewicz P (2007) iMol Molecular Visualization Program. <http://www.pirx.com/iMol>
44. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, et al. (2012) Detecting individual sites subject to episodic diversifying selection. *PLoS Genet* 8: e1002764.
45. Vallender EJ, Lahn BT (2004) Positive selection on the human genome. *Hum Mol Genet* 13 Spec No 2: R245-254.

46. Kosiol C, Vinar T, da Fonseca RR, Hubisz MJ, Bustamante CD, et al. (2008) Patterns of positive selection in six Mammalian genomes. *PLoS Genet* 4: e1000144.
47. Arnold JN, Wormald MR, Sim RB, Rudd PM, Dwek RA (2007) The impact of glycosylation on the biological function and structure of human immunoglobulins. *Annu Rev Immunol* 25: 21-50.
48. Brezski RJ, Jordan RE (2010) Cleavage of IgGs by proteases associated with invasive diseases: an evasion tactic against host immunity? *MAbs* 2: 212-220.
49. Johnson TA, Qiu J, Plaut AG, Holyoak T (2009) Active site gating regulates substrate selectivity in a chymotrypsin-like serine protease. The structure of *Haemophilus influenzae* IgA1 protease. *J Mol Biol* 389: 559-574.
50. Wines BD, Ramsland PA, Trist HM, Gardam S, Brink R, et al. (2011) Interaction of human, rat, and mouse immunoglobulin A (IgA) with staphylococcal superantigen-like 7 (SSL7) decoy protein and leukocyte IgA receptor. *J Biol Chem* 286: 33118-33124.

Chapter 4

IgG evolution in Leporids

Paper 5

Leporid IgG shows evidence of strong selective pressure on the hinge and CH3 domains

Ana Pinheiro, Jenny M. Woof, Tereza Almeida, Joana Abrantes, Paulo C. Alves, Christian Gortázar, and Pedro J. Esteves

Open Biology, 4(9), 140088. doi: 10.1098/rsob.140088.

Abstract

Immunoglobulin G (IgG) is the predominant serum immunoglobulin and has the longest serum half-life of all the antibody classes. The European rabbit IgG has been of significant importance in immunological research and, thus, is well characterized. However, the IgG of other leporids has been disregarded. To evaluate the evolution of this gene in leporids, we sequenced the complete *IGHG* for six other genera: *Bunolagus*, *Brachylagus*, *Lepus*, *Pentalagus*, *Romerolagus* and *Sylvilagus*. The newly sequenced leporid *IGHG* gene has an organization and structure similar to that of European rabbit IgG. A gradient in leporid IgG constant domain diversity was observed with the CH1 being the most conserved and the CH3 the most variable domain. Positive selection was found to be acting on all constant domains but with a greater incidence in the CH3 domain, where a cluster of three positively selected sites was identified. In the hinge region, only three polymorphic positions were observed. The same hinge length was observed for all leporids. Unlike the variation observed for the European rabbit, all 11 *Lepus* species studied share exactly the same hinge motif, suggesting its maintenance as a result of an advantageous structure or conformation.

Keywords

Rabbit, *IGHG*, genetic diversity, immunoglobulins constant region, hinge region, evolution

Introduction

Immunoglobulin G (IgG) is the predominant serum immunoglobulin. IgG participates directly and indirectly to immune responses through its effector functions, elicited through neutralization and binding to Fcγ receptors and C1q. These functions include antibody-dependent cellular cytotoxicity, antibody-dependent cell phagocytosis, and/or complement-dependent cytotoxicity [1]. IgG also participates in direct neutralization of toxins and viruses. Furthermore IgG is the only Ig class that binds to FcRn, the neonatal Fc receptor, which provides passive immunity to newborns as well as protects this particular Ig class from degradation granting IgG the longest serum half-life of all Ig classes [2]. In mammals the number of IgG subclasses ranges from 1-7, each subclass differing in effector functions [3]. Four IgG subclasses have been found in humans [4], mice [5] and rat [6], three in cattle [7], six in pigs [8] and seven in horses [9]. In bats, the number of IgG subclasses ranges from 1-5 depending on the particular species [10]. Amongst mammals the European rabbit is seemingly unique in that it possesses only one IgG subclass [11].

The European rabbit IgG has been extensively studied and its allelic variation characterized in detail. Two loci were distinguished by serology, the *d* and *e*, each with two segregating alleles, *d11/d12* and *e14/e15*, respectively [12]. The *locus d* is correlated to a Thr-Met change at position 9 (IMGT numbering; [13]) in the *IGHG* hinge region. As for *locus e*, it is correlated to an Ala-Thr change at position 92 in the *IGHG* CH2 (IMGT unique numbering for C-DOMAIN [13]). Serologic studies have found the *e15* allotype in species of the genera *Oryctolagus*, *Lepus*, *Sylvilagus*, *Romerolagus* and *Ochotona* [14,15], but have failed to identify the *d11* and *d12* allotypes in *Lepus* and *Sylvilagus* genera [16]. Protein and nucleotide sequence data for *IGHG* in Leporids are scarce. The relationship between serology and protein variation at the *e locus* or *IGHG* CH2 domain has been studied by amino acid sequencing of tryptic peptides in various lagomorph species [14,17-20]. The nucleotide sequencing data is limited essentially to the *IGHG* molecule sequence for the European rabbit and *IGHG* hinge and *IGHG* CH2 domains for a restricted number of *Sylvilagus* and *Lepus* species. Sequencing of the *IGHG* hinge domain in Leporids showed differences between species at residues 8 and 9 (IMGT numbering, [13]) [21], while in the *IGHG* CH2 a hotspot of variation was found at position 92 (IMGT numbering) [22] (see Fig.1).

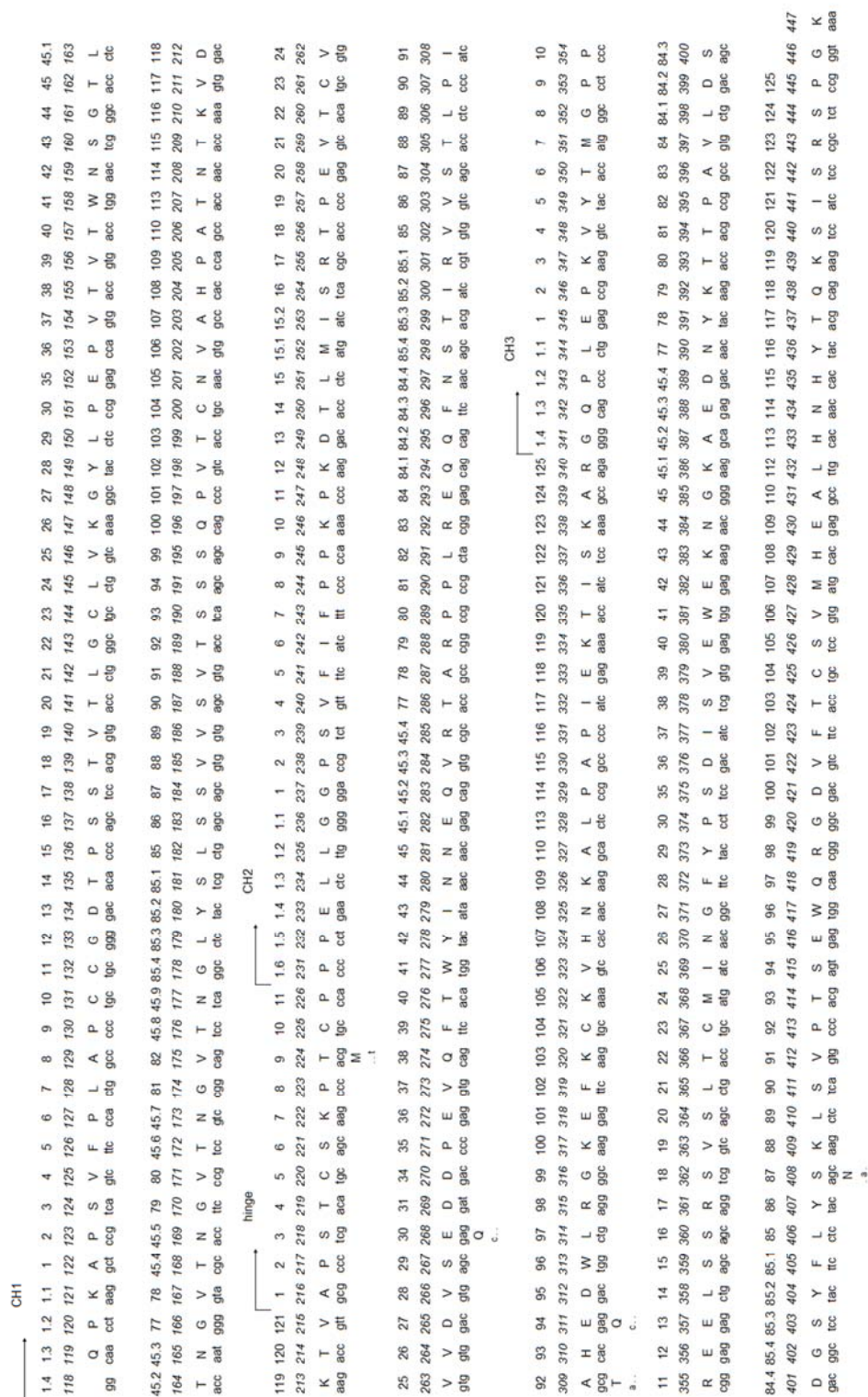


Figure 1. European rabbit (*Oryctolagus cuniculus*) IgG DNA sequence of constant and hinge regions (Genbank accession number: L29172). The amino acid translation is shown above the nucleotide sequence; for polymorphic positions other amino acid possibilities are shown below the nucleotide sequence. The IMGT unique numbering for the constant domain (13) and EU numbering (in *Italic* lettering) are shown on top.

The family Leporidae comprises 11 genera: *Brachylagus*, *Bunolagus*, *Caprolagus*, *Lepus*, *Nesolagus*, *Oryctolagus*, *Pentalagus*, *Poelagus*, *Pronolagus*, *Romerolagus* and *Sylvilagus*) [23,24]. *Lepus* is a polytypic, cosmopolitan genus, comprising 24 to 30 currently recognized species [24-26], that most probably originated in North America [27,28]. The remaining genera have more restricted distributions, though having a nearly worldwide distribution. Most, with the exception of *Nesolagus*, *Pronolagus* and *Sylvilagus*, which comprise two, four and more than 16 species, respectively, are monotypic genera [29-34]. Some discrepancies exist between molecular and fossil data, confusing leporid taxonomy [27,28]. Molecular analyses based on nuclear and mitochondrial markers suggest that within Leporidae two groups diverged 14 million years ago: a first group encompasses the genera *Nesolagus*, *Poelagus* and *Pronolagus* and the second group includes the remaining eight genera. Within this second group *Romerolagus* was the first genus to diverge 13 million years ago, followed by *Lepus* which diverged 12 million years ago. *Brachylagus* and *Sylvilagus* diverged from a group composed of *Bunolagus*, *Caprolagus*, *Oryctolagus* and *Pentalagus* 10 million years ago. The *Brachylagus-Sylvilagus* and *Bunolagus-Caprolagus-Oryctolagus-Pentalagus* splits are estimated at 9 million years ago (approximate times; [28]).

In this study we extend the knowledge on this immunoglobulin class in leporids by sequencing the complete *IGHG* gene for six additional extant leporid genera: *Bunolagus*, *Brachylagus*, *Lepus*, *Pentalagus*, *Romerolagus* and *Sylvilagus*.

Material and methods

Total genomic DNA of specimens of *Bunolagus*, *Brachylagus*, *Lepus*, *Pentalagus*, *Romerolagus* and *Sylvilagus* genera was extracted from frozen liver or ear tissue using an EasySpin Genomic DNA Minipreps Tissue Kit (Citomed). Additionally, six European rabbits were also analysed: two individuals of the subspecies *O. c. algirus*, three individuals of the subspecies *O. c. cuniculus* and a domestic rabbit belonging to the New Zealand White breed. PCR amplification of the four *IGHG* exons was conducted using primers designed on the basis of European rabbit *IGHG* available sequences (GenBank accession number AY386696) [35]. A fragment containing the *IGHG* CH1 and hinge domains was amplified using primers FG12 5' TCAGGCCCGAGACTGTAGACC 3' and RE [21] under the following conditions: 15 min at 95°C followed by 35 cycles at 94°C (30 s), 63°C (30 s) and 72°C (45 s), with a final

extension at 60°C (20 min). Another fragment containing IGHG CH2 and CH3 exons was amplified using primers F3 [22] and RG31 5' TTGGAAGGAATCAGGACAGC 3' under the following conditions: 15 min at 95°C followed by 35 cycles at 94°C (30 s), 60°C (30 s) and 72°C (45 s), with a final extension at 60°C (20 min). Primers were designed so the two fragments overlap in order to obtain the full intron sequence. Sequences were determined by automated sequencing following the Big Dye Terminator Cycle Sequencing protocol (Perkin Elmer, Warrington, UK) using the referred primers.

To confirm the hinge length a fragment of the expressed IgG was also sequenced for one *Lepus granatensis*, one *Lepus europaeus* and one *Sylvilagus floridanus*. Total RNA was extracted from spleen samples using RNeasy Mini Kit (Qiagen, Hilden, Germany), following first strand cDNA synthesis using oligo(dT) as primer (Invitrogen, Carlsbad, CA, USA) and SuperScript III reverse transcription (Invitrogen) as recommended by the manufacturer. A mid CH1 to mid CH2 fragment was PCR amplified using primers FG1int2 (5'CCAGTGACCGTGAAGTGGAA3') and RG2int2 (5'GGACTTTGCACTTGAAGTCC3') designed in conserved regions of the leporid *IGHG* gene segment. A touchdown PCR was performed and the conditions were as follows: 3 min at 98°C followed by 5 cycles at 98°C (30 s), annealing starting at 66°C with a 1°C decrease/cycle until reaching 62°C (30 s) and 72°C (30 s), followed by 30 cycles at 98°C (30 s), 62°C (30 s) and 72°C (30 s), with a final extension at 72°C (5 min).

Sequences obtained in this study were edited and aligned using CLUSTAL W [36] as implemented in BioEdit software [37] and the amino acid sequences were inferred using the software BioEdit [37]. The obtained sequences were also aligned and compared to leporid sequences available in GenBank. Accession numbers for all sequences are given in Table 1. Codon numbering is according to the IMGT unique numbering for C-DOMAIN [13]. Amino acid residue numbering is also defined according to Eu numbering [38]. Sequence nucleotide diversity was estimated using DnaSP version 5.10 [39].

The secondary structure of the leporids IgG heavy chain was analyzed using the DiAminoacid Neural Network Application (DiANNA) (<http://clavius.bc.edu/~clotelab/DiANNA/>) [40-42]. DiANNA predicts the disulphide connectivity using a neural network trained on databases derived from high-quality protein structures that include evolutionary and secondary structure information. First PSIPRED is run to predict the secondary structure and this information is then used to find pairs of cysteines using a maximum weight matching.

Table 1- IGHG sequences accession numbers for sequences used in this study

Sequence identification	Accession number
<i>rabbit germline sequences</i>	AY386696; L29172
<i>rabbit cDNA sequences</i>	DQ402474, M16426, K00752, XM_002723875, J00665
<i>Lepus capensis germline CH2</i>	AJ295218
<i>Lepus californicus germline CH2</i>	AJ295219
<i>Lepus americanus germline CH2</i>	AJ431716
<i>Lepus saxatilis germline CH2</i>	AJ295222
<i>Lepus timidus germline CH2</i>	AJ295216
<i>Sylvilagus cunicularis germline CH2</i>	AJ295223
<i>Sylvilagus floridanus germline hinge</i>	DQ206979
Current study	
germline	
<i>Oryctolagus cuniculus cuniculus</i>	KJ807306-KJ807308
<i>Oryctolagus cuniculus algirus</i>	KJ807309-KJ807310
<i>Oryctolagus cuniculus</i> New Zealand breed	KJ807311
<i>Lepus californicus</i>	KJ807312, KJ807313
<i>Lepus callotis</i>	KJ807314, KJ807315
<i>Lepus castroviejoi</i>	KJ807316
<i>Lepus europaeus</i>	KJ807317-KJ807324
<i>Lepus granatensis</i>	KJ807325-KJ807331
<i>Lepus saxatilis</i>	KJ807332, KJ807333
<i>Lepus townsendi</i>	KJ807334, KJ807335
<i>Lepus americanus</i>	KJ807336, KJ807337
<i>Lepus corsicanus</i>	KJ807338, KJ807340
<i>Lepus capensis</i>	KJ807341, KJ807343
<i>Lepus timidus</i>	KJ807344, KJ807345
<i>Pentalagus furnessii</i>	KJ807346
<i>Sylvilagus floridanus</i>	KJ807347
<i>Sylvilagus bachmanii</i>	KJ807348
<i>Bunolagus monticularis</i>	KJ807351
<i>Brachylagus idahoensis</i>	KJ807350
<i>Romerolagus diazii</i>	KJ807349
cDNA	
<i>Lepus europaeus</i>	KJ807353
<i>Lepus granatensis</i>	KJ807352
<i>Sylvilagus floridanus</i>	KJ807354

The nucleotide sequences alignment was screened for recombination as it can mislead positive selection analysis [43,44]. For this, the software GARD (Genetic Algorithm for Recombination Detection) [45,46] available from the DATAMONKEY web server was used. The best fitting nucleotide substitution model was determined using the automatic model selection tool available on the server.

To identify signatures of selection on leporid IgG we compared the rate per-site of nonsynonymous substitution (dN) to the rate per-site of synonymous substitutions

(dS) in a maximum likelihood (ML) framework, using six different methods. Since each of the methods employs unique algorithms, and as done previously [47-49], we only considered those codons identified by at least two of the ML methods to be positively selected codons. Using the CODEML program of the PAML 4.4 package [50,51] we compared two disparate models, M8, which allows for codons to evolve under positive selection ($dN/dS > 1$) and M7, which does not ($dN/dS \leq 1$), using a likelihood ratio test (LRT) with 2 degrees of freedom [52,53]. Codons under positive selection for model M8 were identified using a Bayes Empirical Bayes approach (BEB) [54] and considering a posterior probability of $>90\%$. Using Mega 5 [55], a Neighbour-Joining phylogenetic tree was used as a working topology, with the p-distance substitution model and the pairwise deletion option to handle gaps and missing data. The obtained tree was in accordance with the lagomorphs accepted phylogeny. The five methods for detecting positive selection available from the DATAMONKEY web server [56] were also used: the Single Likelihood Ancestor Counting model (SLAC), the Fixed Effect Likelihood model (FEL), the Random Effect Likelihood model (REL), the Mixed Effects Model of Evolution (MEME) and the Fast Unbiased Bayesian Approximation (FUBAR). For these analyses, the best fitting nucleotide substitution model was first determined through the automatic model selection tool available on the server.

The location within the IgG structure of the residues under positive selection was analysed by mapping the residues onto the solved crystal structures of rabbit IgG-Fab (PDB ID: 4HBC; [57]) and Fc (PDB ID: 2VUO; [58]). To examine their relation to putative sites of interest, the sites of interaction with Fc γ receptors (Fc γ Rs), FcRn and complement C1q [2,59,60,61] were also mapped onto the 3D IgG-Fc structure. The NCBI application Cn3D 4.1 (<http://www.ncbi.nlm.nih.gov/Structure/CN3D/cn3d.shtml>; [62]) was used to this purpose.

Results

The *IGHG* gene newly sequenced for six leporid genera, *Bunolagus*, *Brachylagus*, *Lepus*, *Pentalagus*, *Romerolagus* and *Sylvilagus*, is similar to the European rabbit (*Oryctolagus*) *IGHG* and, thus, the intron-exon organization was inferred from available published rabbit *IGHG* genes. Splicing signals were present at the intron boundaries and hence it is assumed that all studied leporids share the same *IGHG* exon organization as the European rabbit.

CH1 domain

The CH1 domain is the most conserved among leporid IGHG domains with only 14 amino acid variable positions, the majority of which involve one substitution and half are conservative regarding the amino acid properties. The European rabbit IgG differs from the IgG of other leporids by the Asn/Ser change at position 45.3 and by the Thr/Pro change at position 92. *Bunolagus* IgG has a His for a Glu substitution at position 84.2 and *Lepus* IgG has a Thr for a Ser change at position 90. Only the substitution at position 84.2 involves changes in amino acid characteristics (Fig. 2).

CH2 domain

The CH2 domain of leporid IgGs, though fairly conserved, shows more diversity than the CH1 domain. For the CH2 domain, 15 amino acid variable positions were observed, eleven of which involve one substitution but, contrary to that observed for the CH1 domain, the majority of these polymorphic positions involve changes in amino acid properties. The European rabbit has specific Arg residues at positions 45.4 and 125. The residues Leu17, Pro45.4 and Leu84.1 are specific to *Lepus*, *Brachylagus* and *Bunolagus*, respectively. *Romerolagus* uniquely has Ala1.6, Leu45.4, Val82, and Leu85.2 residues (Fig 2).

CH3 domain

This is the most diverse of the leporid IGHG domains with 36 variable positions, the majority involving changes in amino acid properties, and has the highest number of diagnostic positions. In this domain the European rabbit has specific Pro1.2 and Ala45.2 residues, *Bunolagus* uniquely has Val1.2, Arg3, Thr17 and Pro35 residues and *Brachylagus* has an Asn for a Thr change at position 80. Three diagnostic positions were observed for each of three genera: *Lepus* at Asn15, Thr105 and Leu125 residues, *Pentalagus* at Arg1.3, Ser77 and Ala90, and *Romerolagus* at Lys11, Glu35 and Ala120 as distinctive. Only *Sylvilagus* lacked specific residues (Fig. 2).

Hinge

Three variable positions were found, all involving changes in amino acid properties. Changes at these positions define genera specific motifs, with the exception of *Sylvilagus* and *Romerolagus* that share the same substitutions: Val1, Pro8, Leu9 (Fig. 2). For all studied leporids an eleven residue hinge was observed, which, given the high variability observed in IgG hinge length in other mammals, is surprising. To check if alternative splicing sites could be used by some leporids, the expressed hinge

was sequenced for *Lepus* and *Sylvilagus* individuals and each one had the same eleven residues.

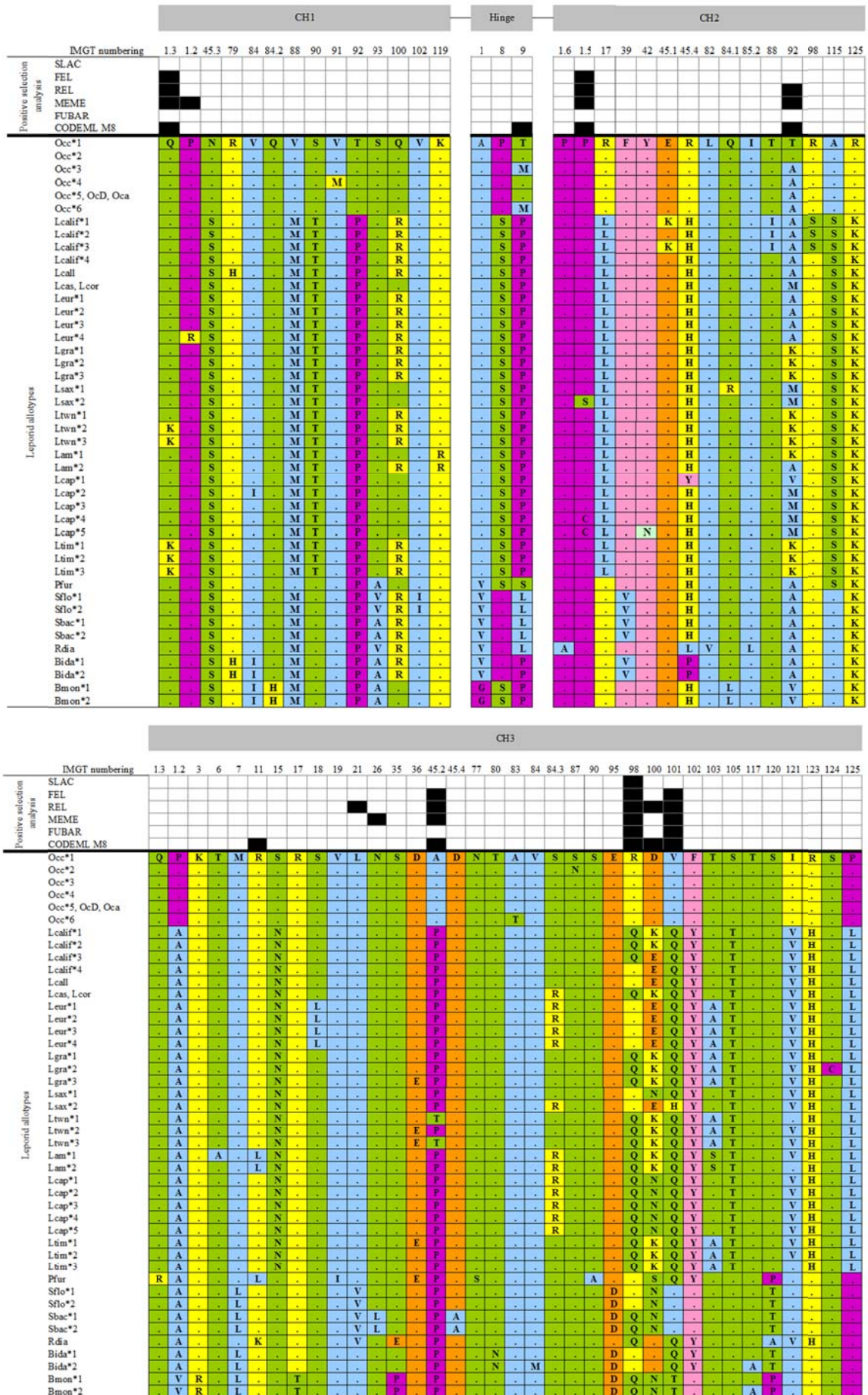
Intron diversity

The intronic regions between CH1-hinge, hinge-CH2 and CH2-CH3 were fully sequenced. Overall nucleotide diversity for these regions is similar to that observed for the coding regions ($Pi_{introns}=0.03704$; $Pi_{exons}=0.03277$). The major differences observed between leporid genera are insertions and deletions (indels), which reflect the accepted leporid phylogeny. Indeed, in the intron between CH1 and hinge exons a 20 bp deletion is shared by *Bunolagus*, *Oryctolagus* and *Pentalagus* while *Romerolagus* has a unique 10 bp deletion. Similarly, in the intron between CH2 and CH3 exons a 1 bp deletion is shared by *Brachylagus* and *Sylvilagus*, while *Oryctolagus* has two characteristic 1 bp deletions and *Romerolagus* has an insertion of 13 bp (Supplementary data).

Leporid IgG heavy chain structure

All immunoglobulin heavy chains are organized into globular domains, a structure stabilized by intra-chain disulphide bonds between conserved cysteines in each domain at positions 23 and 104 in each domain. As expected, all studied leporids share these cysteines. The European rabbit IgG further has two additional cysteines in the CH1 domain at positions 10 and 11 (Cys131 and Cys132; Eu numbering) and a further two in the hinge, at positions 5 and 10. The CH1 domain Cys at position 10 (Cys131; Eu numbering) bonds to the light chain Cys, the CH1 Cys at position 11 (Cys132; Eu numbering) bonds to the Cys at position 5 in the hinge, and the other hinge Cys forms the inter-chain disulphide bridge with its equivalent in the other heavy chain. Again, conserved cysteines at these positions are shared by all studied leporids,

Figure 2. Variable codons and amino acid physicochemical properties for leporid IGHG CH1 and CH2 domains and hinge region. Above are indicated the codons identified as under positive selection by each one of the six methods used; only those codons identified by at least two methods are considered to be positively selected codons. The codon numbering is according to the IMGT unique numbering for the constant domain (13). The colours represent amino acid properties: polar neutral (green), polar positive (yellow), polar negative (orange), non-polar neutral (purple), non-polar aliphatic (blue) and non-polar aromatic (light pink). Occ- *Oryctolagus cuniculus cuniculus*; OcD- Domestic rabbit New Zealand White breed; Oca- *Oryctolagus cuniculus algirus*; Lcalif- *Lepus californicus*; Lcall- *Lepus callotis*; Lcas- *Lepus castroviejo*; Lcap- *Lepus capensis*; Leur- *Lepus europaeus*; Lgra- *Lepus granatensis*; Lsax- *Lepus saxatilis*; Ltwn- *Lepus townsendi*; Lam- *Lepus americanus*; Lcap- *Lepus capensis*; Ltim- *Lepus timidus*; Pfur- *Pentalagus furnessii*; Sflo- *Sylvilagus floridanus*; Sbac- *Sylvilagus bachmanii*; Rdia- *Romerolagus diazii*; Bida - *Brachylagus idahoensis*; Bmon- *Bunolagus monticularis*



suggesting that the IgG heavy chain structure is maintained across leporids. However, extra cysteines were found in the CH2 domain at position 1.5 (Cys232; Eu numbering) of two *L. capensis* alleles and in the CH3 domain at position 124 (Cys444; Eu numbering) for one *L. granatensis* allele. Disulphide bond prediction analysis indicates that the extra CH2 cysteine does not establish any bond. However, for the allele with the additional CH3 Cys, different bonds to those described above are predicted for the CH2 and CH3 domains. The CH2 domain Cys at position 23 (Cys261; Eu numbering) is now predicted to bond with the CH3 domain Cys at position 104 (Cys425; Eu numbering), and the CH3 domain Cys at position 23 (Cys367; Eu numbering) to bond with the CH3 domain Cys at position 124 (Cys444; Eu numbering). However, the physiological relevance of these predictions remains unclear.

Glycosylation is also important for protein structure and function. The European rabbit IgG Fc has an N-glycosylation site at CH2 84.4 (Asn297; Eu numbering), which was found to be conserved in all studied leporids, and indeed all vertebrate IgG. No other N-glycosylation sites were predicted. O-linked glycans, i.e. glycans linked to Ser/Thr residues in Ser/Thr/Pro rich domains, are known on human IgA1 and IgD hinges and also on the rabbit IgG hinge, for which the Thr residue at hinge position 9 of d12 rabbits is O-glycosylated (d11 rabbits hinge position 9 have a Met residue) [16,63,64]. This O-glycan confers protection against cleavage of the rabbit IgG hinge [63]. Putative O-glycosylation sites are present on *Lepus* and *Bunolagus* hinge position 8 and *Pentalagus* hinge positions 8 and 9, all occupied by Ser residues.

Signatures of positive selection on leporid IgG

The comparison of the two disparate models implemented in CODEML revealed evidence for positive selection acting on the leporid IgG, with the model allowing sites to evolve under positive selection (M8) showing a significantly better fit than the model that did not (M7) ($\ln L \text{ M7}/\ln L \text{ M8} = -2850.1/-2817.1$; $-2\Delta\ln L = 33$; $\alpha < 0.01$). Comparing the sites recognized by each of the six employed methods led to the identification of seven positively selected codons. One locates to the CH1 domain, two to the CH2 domain and four to the CH3 domain. All of these residues show changes in amino acid characteristics and occupy exposed positions on the European rabbit IgG structure. Interestingly, four of these codons locate near sites of interaction with ligands: the CH1 residue 1.3 (residue 119; Eu numbering) is in the immediate vicinity of the VH domain, the CH2 residue 1.5 (residue 232; Eu numbering) lies in the region of residues that interact with FcγRs, and the CH2 92 residue (residue 309; Eu numbering) and CH3 45.2 residue (residue 387; Eu numbering) locate on the region of residues that interact with FcRn. Of note, the CH3 domain residues at position 98, 100 and 101 (residues

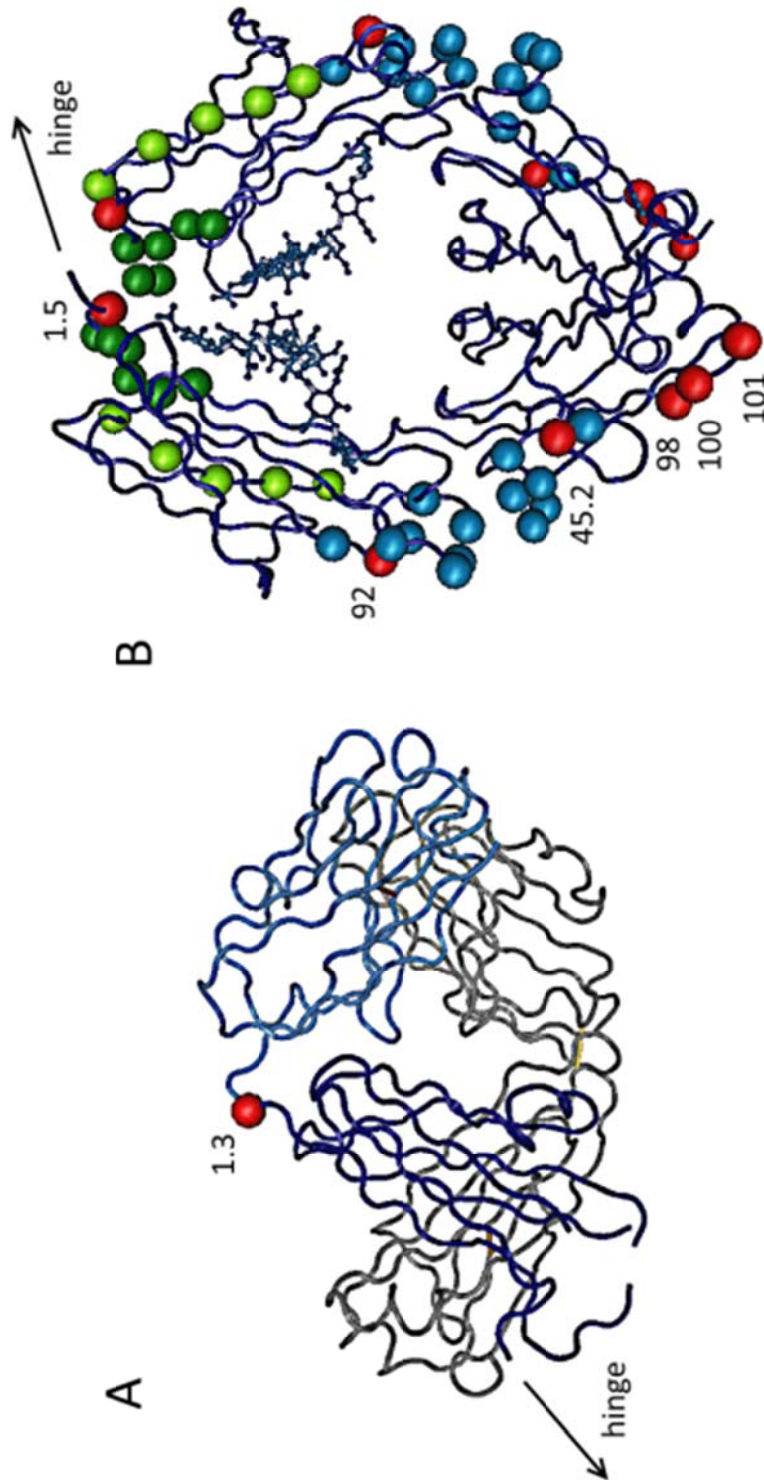


Figure 3. Leporids residues encoded by positively selected codons in the three-dimensional structure of rabbit IgG. **A**, Rabbit IgG-Fab (PDB ID 4HBC) **B**, Rabbit IgG-Fc (PDB ID 2VUO). The light chain is in the background coloured grey, the heavy chain variable domain is in the foreground coloured light blue and the heavy chain constant domains are coloured dark blue. Positively selected codons are represented in red dots. Residues significant for FcγR interaction are highlighted in dark green, residues significant for FcγR interaction are in light blue and residues significant for C1q complement interaction are in light green. Residues numbering is according to IMGT unique numbering for the constant domain [13]. The N-glycan attached to CH2 84.4 (Asn297; Eu number) is shown in ball and stick representation.

419, 421, and 422; Eu numbering), form an exposed cluster in the C-terminal portion of this domain (Fig. 3).

Discussion

Rabbit IgG has been extensively studied and its genetic diversity thoroughly characterised but, despite the relevance of IgG as a crucial component of the host immune response and the uniqueness of rabbit IgG, the extension of this knowledge to other leporids has been neglected. In this work we extended knowledge on the evolution of leporid IgG by analysing six extant genera.

The results obtained in this study reveal that the leporids share considerable sequence similarity for their *IGHG* (approximately 94%). A gradient in constant domain diversity is observed with the CH1 domain being the most conserved of leporid *IGHG* domains and the CH3 domain the most variable. In fact, there are twice as many variable amino acid sites and number of diagnostic substitutions in the CH3 domain compared to the CH1 and CH2 domains, making the CH3 the most informative domain for leporid species identification. The CH3 domain has been identified as the most diagnostic domain to distinguish between IgG isotypes for swine [8] and primate species [65]. The divergence previously found for IgG CH3 domains between macaques and human, species that diverged ~32 million years ago [66], is similar to the divergence found between *Oryctolagus* and *Romerolagus*, genera that separated ~13 million years ago [28], and thus it seems that either leporid IgG CH3 domain is evolving under selective pressure to change or that evolutionary constraints are conserving primate IgG CH3 domains.

Despite the overall conservation found for leporid *IGHG*, hotspots of variability exist in the hinge and CH2 and CH3 domains and we found evidence of positive selection acting on all IgG constant domains. Previous studies of leporid *IGHG* described the hinge position 9 and CH2 position 92 as hotspots of variability [21,22]. Our results, including more genera and species than the former studies, confirm that hinge position 9 is a leporid hotspot of variability while the CH2 position 92 residue (residue 309; Eu numbering) proves to be a *Lepus* specific hotspot, with four different residues in this *genus* but only 2 in other leporids. Additionally we can pinpoint as having high amino acid diversity the CH2 position 45.4 (residue 387; Eu numbering) and CH3 position 100 (residue 421; Eu numbering), each having five different residues in the studied leporids. These hotspots of variability, and also the seven positions

identified as positively selected, exhibit changes in amino acid physicochemical properties, which may impact on the IgG conformation and structure. Changes at the positively selected CH1 1.3 residue (residue 119; Eu numbering) may impact on the antigen binding site conformation. This would possibly improve leporid antigen binding possibilities given that the usage of the VH1 gene in 90% of VDJ rearrangements in leporids confers a somewhat restricted diversity to the leporid VH domain [67-70]. On the other hand, changes at positively selected codons (PSCs) CH2 1.5 (residue 232; Eu numbering), CH2 92 residue (residue 309; Eu numbering) and CH3 45.2 residue (residue 387; EU numbering) may influence binding of IgG to Fc receptors. In particular, the positively selected CH2 position 1.5 (residue 232; Eu numbering) lies in close vicinity to CH2 residues 1.3-1 (234-237 Eu numbering), which in human IgG form the core of the interaction site for FcγR [60,71]. Since IgG from the European rabbit binds human FcγRI with affinity comparable to human IgGs [72] consistent with a similar interaction mode across species, one might speculate that variation at this position may prove adaptive by influencing binding of IgG to rabbit receptors. The changes at PSCs CH2 1.5, CH2 92 and CH3 45.2 may also confer some resistance against proteins produced by some bacterial pathogens that target the lower hinge-proximal CH2 region (e.g. IdeS; [73]) and the CH2-CH3 interface (e.g. Staphylococcal protein A and Streptococcal protein G; [74]). Interestingly, it has been noted that bacterial pathogens in different mammalian species also target this same inter-domain region in immunoglobulins [75] so it is possible that bacterial species evolved to infect leporids may employ a similar evasion strategy of production of proteins that bind the CH2-CH3 interface.

In contrast to what was found for mammalian IgA, where the Ca3 domain showed less evidence of having evolved under positive selection than Ca1 or Ca2 [48], the leporid IgG CH3 domain has the highest number of positively selected sites of all IgG constant domains, showing that this domain is evolving under positive selection. Areas across the surface of the CH3 domain have been implicated in the formation of hexameric IgG that assembles at antigenic cell surfaces, recruits C1q, and activates complement [76]. Thus, the cluster of positively selected codons observed in this region suggests that the C-terminal CH3 has some functional relevance in the leporids and could be related to protective complement-mediated mechanisms against specific pathogens.

The hinge region shows considerable variability both in amino acid composition and length among IgA and IgG subclasses and alleles and across species [e.g. 8,10,77]. The hinge is the preferential target region for proteolytic cleavage by numerous bacterial proteases [discussed in 68,78], which could explain the great

variability observed for this antibody region. Given this context, the lack of variation observed in this study for the *Lepus* IgG hinge is particularly interesting. Previous studies by Esteves et al. [21] indicated that leporids share the same hinge length and have amino acid differences at only two hinge positions, 8 and 9 (IMGT numbering). The European rabbit shows two residues at position 9, Met and Thr, which correlate with the serological allotypes d11 and d12 [12,79]. The d12 Thr residue is O-glycosylated, conferring protection against cleavage of the rabbit IgG hinge [63]. Thus, the glycosylation of rabbit, *Lepus*, *Bunolagus* and *Pentalagus* hinge may protect the IgG from effects of proteases of pathogens and tumor cells. Despite this increased resistance against proteolytic attack of the d12 allotype, both allotypes interact with FcγR equally well, for example [72]. We have confirmed that all seven studied leporid genera have an eleven residue hinge like *Oryctolagus* and that this hinge is expressed by *Lepus* and *Sylvilagus*. Thus, one can assume that most likely all leporids use a short eleven residue hinge. The existence of more than one IgG in leporids other than the European rabbit has, so far, not been assessed. Thus, despite having found no evidence for the existence of more than one *IGHG* copy in the studied leporid genera during the course of this work, we cannot, although highly unlikely, exclude the possibility of additional IgG genes in leporids. Unlike the variation observed for the European rabbit, all 11 *Lepus* species studied share exactly the same hinge motif, indicating that it may be being maintained due to an advantageous structure or conformation.

Acknowledgements

The authors would like to thank Martin Flajnik (University of Maryland) for his helpful comments on this paper and Antonio Lavazza and Nicola Martinelli (IZSLER- Istituto Zooprofilattico Sperimentale della Lombardia e dell'Emilia Romagna) for kindly providing the *Sylvilagus floridanus* samples. The Portuguese Foundation for Science and Technology (FCT) supported the doctoral fellowship of Ana Pinheiro (SFRH/BD/71252/2010); Joana Abrantes is supported by an FCT Investigator grant IF/01396/2013. This work was partially funded by FEDER (Fundo Europeu de Desenvolvimento Regional) funds through the Programa Operacional Factores de Competitividade (COMPETE program; FCOMP-01-0124-FEDER-028286) and Portuguese national funds through FCT (research project PTDC/BIA-ANM/3963/2012) – Quadro de Referência Estratégico Nacional (QREN) funds from the European Social Fund and Portuguese Ministério da Educação e Ciência. “Genomics Applied To

Genetic Resources” cofinanced by North Portugal Regional Operational Programme 2007/2013 (ON.2 – O Novo Norte), under the National Strategic Reference Framework (NSRF), through the European Regional Development Fund (ERDF) supported this work.

References

1. Strohl WR. 2009 Optimization of Fc-mediated effector functions of monoclonal antibodies. *Curr. Opin. Biotechnol.* **20**, 685-691. (doi: 10.1016/j.copbio.2009.10.011)
2. Roopenian DC, Akilesh S. 2007 FcRn: the neonatal Fc receptor comes of age. *Nat. Rev. Immunol.* **7**, 715-725. (doi:10.1038/nri2155)
3. Schroeder HW, Jr., Cavacini L. 2010 Structure and function of immunoglobulins. *J. Allergy Clin. Immunol.* **125**, S41-52. (doi: 10.1016/j.jaci.2009.09.046)
4. Flanagan JG, Rabbitts TH. 1982 Arrangement of human immunoglobulin heavy chain constant region genes implies evolutionary duplication of a segment containing gamma, epsilon and alpha genes. *Nature* **300**, 709-713. (doi:10.1038/300709a0)
5. Shimizu A, Takahashi N, Yaoita Y, Honjo T. 1982 Organization of the constant-region gene family of the mouse immunoglobulin heavy chain. *Cell* **28**, 499-506. (doi: 10.1016/0092-8674(82)90204-5)
6. Bruggemann M, Free J, Diamond A, Howard J, Cobbold S, Waldmann H. 1986 Immunoglobulin heavy chain locus of the rat: striking homology to mouse antibody genes. *Proc. Natl. Acad. Sci. USA* **83**, 6075-6079.
7. Knight KL, Suter M, Becker RS. 1988 Genetic engineering of bovine Ig. Construction and characterization of hapten-binding bovine/murine chimeric IgE, IgA, IgG1, IgG2, and IgG3 molecules. *J. Immunol.* **140**, 3654-3659.
8. Butler JE, Wertz N, Deschacht N, Kacskovics I. 2009 Porcine IgG: structure, genetics, and evolution. *Immunogenetics* **61**, 209-230. (doi: 10.1007/s00251-008-0336-9)
9. Wagner B, Miller DC, Lear TL, Antczak DF. 2004 The complete map of the Ig heavy chain constant gene region reveals evidence for seven IgG isotypes and for IgD in the horse. *J. Immunol.* **173**, 3230-3242. (doi: 10.4049/jimmunol.173.5.3230)
10. Butler JE, Wertz N, Zhao Y, Zhang S, Bao Y, Bratsch S, Kunz TH, Whitaker JO, Jr., Schountz T. 2011 The two suborders of chiropterans have the canonical heavy-chain immunoglobulin (Ig) gene repertoire of eutherian mammals. *Dev. Comp. Immunol.* **35**, 273-284. (doi: 10.1016/j.dci.2010.08.011)

11. Knight KL, Burnett RC, McNicholas JM. 1985 Organization and polymorphism of rabbit immunoglobulin heavy chain genes. *J. Immunol.* **134**, 1245-1250.
12. Hamers R, Hamers-Casterman C. 1965 Molecular localization of A chain allotypic specificities in rabbit IgG (7S gamma-globulin). *J. Mol. Biol.* **14**, 288-289. (doi: 10.1016/S0022-2836(65)80249-2)
13. Lefranc MP, Pommie C, Kaas Q, Duprat E, Bosc N, Guiraudou D, Jean C, Ruiz M, Da Piedade I, Rouard M, Foulquier E, Thouvenin V, Lefranc G. 2005 IMGT unique numbering for immunoglobulin and T cell receptor constant domains and Ig superfamily C-like domains. *Dev. Comp. Immunol.* **29**, 185-203. (doi: 10.1016/j.dci.2004.07.003)
14. Cazenave PA, Bennamar A, Sogn JA, Kindt TJ. 1987 Immunoglobulin genes in feral population. In *The rabbit in contemporary immunological research.* (ed. S. Dubiski) , p 148. Longman Scientific & Technical Publications.
15. van der Loo W, Hamers-Casterman C. 1979 Phylogeny of the rabbit immunoglobulin allotypes: rabbit anti e-locus antisera detect two allelic forms of IgG molecules in *Romerolagus diazi*. *Annals de l'Institut Pasteur, Immunologie C* **130**, 761.
16. Hamers-Casterman C, Wittouck E, van der Loo W, Hamers R. 1979 Phylogeny of the rabbit gamma-chain determinants: a d12-like antigenic determinant in *Pronolagus rupestris*. *J. Immunogenet.* **6**, 373-381. (doi: 10.1111/j.1744-313X)
17. Aggarwal SJ, Mandy WJ. 1976 Lagomorph IgG hinge region: allotype associated amino acid sequence variations. *Immunochemistry* **13**, 215-220. (doi: 10.1016/0019-2791(76)90218-4)
18. Appella E, Chersi A, Mage R, Dubiski S. 1971 Structural basis of the A14 and A15 allotypic specificities in rabbit IgG. *Proc. Natl. Acad. Sci. USA* **68**, 1341-1345.
19. Teherani J, Capra JD, Aggarwal S, Mandy WJ. 1979 Amino acid sequence analysis of group e allotype-related peptides derived from lagomorph IgG. *Eur. J. Immunol.* **9**, 690-695. (doi: 10.1002/eji.1830090906)
20. Teherani J, Capra JD, Mandy WJ. 1982 Amino acid sequence of the CH2 domain from various lagomorph IgGs. *Mol. Immunol.* **19**, 841-846. (doi: 10.1016/0161-5890(82)90349-2)
21. Esteves PJ, Carmo C, Godinho R, van der Loo W. 2006 Genetic diversity at the hinge region of the unique immunoglobulin heavy gamma (IGHG) gene in leporids (*Oryctolagus*, *Sylvilagus* and *Lepus*). *Int. J. Immunogenet.* **33**, 171-177. (doi: 10.1111/j.1744-313X.2006.00588.x)
22. Esteves PJ, Alves PC, Ferrand N, van der Loo W. 2002 Hotspot variation at the CH2-CH3 interface of leporid IgG antibodies (*Oryctolagus*, *Sylvilagus* and *Lepus*). *Eur. J. Immunogenet.* **29**, 529-535. (doi: 10.1046/j.1365-2370.2002.00355.x)

23. Chapman J, Flux J. 1990 Introduction and overview of the lagomorphs. In Rabbits, hares and pikas, status, survey and conservation action plan (eds. J. Chapman, J. Flux), pp 1-6. Gland: UICN/SSC Lagomorph Specialist Group.
24. Alves PC, Hackländer K. 2008 Lagomorph species: geographical distribution and conservation status. In Lagomorph biology (eds. P.C. Alves, N. Ferrand, K. Hackländer), pp 395–405. Berlin Heidelberg: Springer.
25. Corbet GB, Hill JE. 1980 A World list of mammalian species. Oxford: Oxford University Press.
26. Flux JEC, Angermann R. 1990 The hares and jackrabbits. In Rabbits, Hares and Pikas: Status Conservation Action Plan (eds. J.A. Chapman, J.E.C. Flux), pp 61-94. Gland, Switzerland: International Union for Conservation of Nature and Natural Resources.
27. Lopez-Martinez N. 2008 The Lagomorph Fossil Record and the Origin of the European Rabbit. In Lagomorph Biology (eds. P.C. Alves, N. Ferrand, K. Hackländer), pp 27-46. Berlin Heidelberg: Springer.
28. Matthee CA, van Vuuren BJ, Bell D, Robinson TJ. 2004 A molecular supermatrix of the rabbits and hares (Leporidae) allows for the identification of five intercontinental exchanges during the Miocene. *Syst. Biol.* **53**, 433-447. (doi: 10.1080/10635150490445715)
29. Surridge AK, Timmins RJ, Hewitt GM, Bell DJ. 1999 Striped rabbits in Southeast Asia. *Nature* **400**, 726.
30. Can DN, Abramov AV, Tikhonov AN, Averianov AO. 2001 Annamite striped rabbit *Nesolagus timminsi* in Vietnam. *Acta Theriol.* **46**, 437–440.
31. Chapman JA, Cramer KL, Dippenaar NJ, Robinson TJ. 1992 Systematics and biogeography of the New England cottontail, *Sylvilagus transitionalis* (Bangs, 1895), with the description of a new species from the Appalachian Mountains. *Proc. Biol. Soc. Washington* **105**, 841–866.
32. Frey JK, Fisher RD, Ruedas LA. 1997 Identification and restriction of the type locality of the Manzano Mountains cottontail, *Sylvilagus cognatus* Nelson, 1907 (Mammalia: Lagomorpha: Leporidae). *Proc. Biol. Soc. Washington* **110**, 329–331.
33. Whiteford CEM. 1995 Molecular phylogeny of the genus *Pronolagus* (Mammalia: Lagomorpha) and the use of morphological and molecular characters in the delineation of *P. rupestris*. MSc Thesis, University of Pretoria.
34. Matthee CA, Robinson TJ. 1996 Mitochondrial DNA differentiation among geographical populations of *Pronolagus rupestris*, Smith's red rock rabbit (Mammalia: Lagomorpha). *Heredity* **76**, 514–523. (doi:10.1038/hdy.1996.74)

35. Ros F, Puels J, Reichenberger N, van Schooten W, Buelow R, Platzer J. 2004 Sequence analysis of 0.5 Mb of the rabbit germline immunoglobulin heavy chain locus. *Gene* **330**, 49-59. (doi: 10.1016/j.gene.2003.12.037)
36. Thompson JD, Higgins DG, Gibson TJ. 1994 CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Acids Res.* **22**, 4673-4680. (doi: 10.1093/nar/22.22.4673)
37. Hall TA. 1999 bioedit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids Symposium Series* **41**, 95-98.
38. Kabat EA, Wu TT, Bilofsky H, Reid-Miller M, Perry H. 1983 Sequences of proteins of immunological interest. Washington D.C.: US Department of Health and Human Services, National Institutes of Health.
39. Librado P, Rozas J. 2009 DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451-1452. (doi: 10.1093/bioinformatics/btp187)
40. Ferre F, Clote P. 2005 DiANNA: a web server for disulphide connectivity prediction. *Nucl. Acids Res.* **33**, W230-232. (doi: 10.1093/nar/gki412)
41. Ferre F, Clote P. 2005 Disulphide connectivity prediction using secondary structure information and diresidue frequencies. *Bioinformatics* **21**, 2336-2346. (doi: 10.1093/bioinformatics/bti328)
42. Ferre F, Clote P. 2006 DiANNA 1.1: an extension of the DiANNA web server for ternary cysteine classification. *Nucl. Acids Res.* **34**, W182-185. (doi: 10.1093/nar/gkl189)
43. Anisimova M, Bielawski JP, Yang Z (2001) Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Mol Biol Evol* **18**: 1585-1592.
44. Posada D, Crandall KA (2002) The effect of recombination on the accuracy of phylogeny estimation. *J Mol Evol* **54**: 396-402. PubMed: 11847565. (doi: 10.1007/s00239-001-0034-9)
45. Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD (2006) Automated phylogenetic detection of recombination using a genetic algorithm. *Mol Biol Evol* **23**: 1891-1901. (doi: msl051 [pii] 10.1093/molbev/msl051)
46. Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD (2006) GARD: a genetic algorithm for recombination detection. *Bioinformatics* **22**: 3096-3098. (doi: btl474 [pii] 10.1093/bioinformatics/btl474)
47. Areal H, Abrantes J, Esteves PJ. 2011 Signatures of positive selection in Toll-like receptor (TLR) genes in mammals. *BMC Evol. Biol.* **11**, 368. (doi: 10.1186/1471-2148-11-368.)

48. Pinheiro A, Woof JM, Abi-Rached L, Parham P, Esteves PJ. 2013 Computational analyses of an evolutionary arms race between mammalian immunity mediated by immunoglobulin A and its subversion by bacterial pathogens. *PLoS One* **8**:e73934. (doi: 10.1371/journal.pone.0073934)
49. Wlasiuk G, Nachman MW. 2010 Adaptation and constraint at Toll-like receptors in primates. *Mol. Biol. Evol.* **27**, 2172-2186. (doi: 10.1093/molbev/msq104)
50. Yang Z. 1997 PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**, 555-556.
51. Yang Z. 2007 PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586-1591. (doi:10.1093/molbev/msm088)
52. Nielsen R, Yang Z. 1998 Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* **148**, 929-936.
53. Yang Z, Swanson WJ, Vacquier VD. 2000 Maximum-likelihood analysis of molecular adaptation in abalone sperm lysin reveals variable selective pressures among lineages and sites. *Mol. Biol. Evol.* **17**, 1446-1455. (doi: 10.1093/molbev/msi237)
54. Yang Z, Wong WS, Nielsen R. 2005 Bayes empirical bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* **22**, 1107-1118. (doi: 10.1093/molbev/msi237)
55. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011 MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731-2739. (doi: 10.1093/molbev/msr121)
56. Pond SL, Frost SD. 2005 Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* **21**, 2531-2533. (doi: 10.1093/bioinformatics/bti320)
57. Arai H, Glabe C, Luecke H. 2012 Crystal structure of a conformation-dependent rabbit IgG Fab specific for amyloid prefibrillar oligomers. *Biochim. Biophys. Acta* **1820**, 1908-1914. (doi: 10.1016/j.bbagen.2012.08.016)
58. Girardi E, Holdom MD, Davies AM, Sutton BJ, Beavil AJ. 2009 The crystal structure of rabbit IgG-Fc. *Biochem. J.* **417**, 77-83. (doi: 10.1042/BJ20081355)
59. Schneider S, Zacharias W. 2012 Atomic resolution model of the antibody Fc interaction with the complement C1q component. *Mol. Immunol.* **51**, 66-72. (doi: 10.1016/j.molimm.2012.02.111)
60. Sondermann P, Huber R, Oosthuizen V, Jacob U. 2000 The 3.2-Å crystal structure of the human IgG1 Fc fragment-FcγRIII complex. *Nature* **406**, 267-273. (doi:10.1038/35018508)

61. Rayner LE, Kadkhodayi-Kholghi N, Heenan RK, Gor J, Dalby PA, Perkins SJ. 2013 The solution structure of rabbit IgG accounts for its interactions with the Fc receptor and complement C1q and its conformational stability. *J. Mol. Biol.* **425**(3), 506-523. (doi: 10.1016/j.jmb.2012.11.019)
62. Wang Y, Geer LY, Chappey C, Kans JA, Bryant SH. 2000 Cn3D: sequence and structure views for Entrez. *Trends Biochem. Sci.* **25**, 300-302. (doi: 10.1016/S0968-0004(00)01561-9)
63. Goodman, JW. 1965 Heterogeneity of rabbit gamma-globulin. *Biochemistry* **4**, 2350 (doi: 10.1021/bi00887a013)
64. Arnold JN, Wormald MR, Sim RB, Rudd PM, Dwek RA. 2007 The impact of glycosylation on the biological function and structure of human immunoglobulins. *Annu Rev Immunol.* **25**, 21-50 (doi: 10.1146/annurev.immunol.25.022106.141702)
65. Jacobsen FW, Padaki R, Morris AE, Aldrich TL, Armitage RJ, Allen MJ, Lavallee JC, Arora T. 2011 Molecular and functional characterization of cynomolgus monkey IgG subclasses. *J. Immunol.* **186**, 341-349. (doi: 10.4049/jimmunol.1001685)
66. Perelman P, Johnson WE, Roos C, Seuanez HN, Horvath JE, Moreira MA, Kessing B, Pontius J, Roelke M, Rumpler Y, Schneider MP, Silva A, O'Brien SJ, Pecon-Slattery J, 2011 A molecular phylogeny of living primates. *PLoS Genet.* **7**:e1001342. (doi: 10.1371/journal.pgen.1001342)
67. Esteves PJ, Lanning D, Ferrand N, Knight KL, Zhai SK, van der Loo W (2005) The evolution of the immunoglobulin heavy chain variable region (IgVH) in Leporids: an unusual case of transspecies polymorphism. *Immunogenetics* **57**, 874-882. (doi: 10.1007/s00251-005-0022-0)
68. Esteves PJ, Lanning D, Ferrand N, Knight KL, Zhai SK, van der Loo W (2004) Allelic variation at the VH α locus in natural populations of rabbit (*Oryctolagus cuniculus*, L.). *J. Immunol.* **172**, 1044-53. (doi: 10.4049/jimmunol.172.2.1044)
69. Pinheiro A, Lanning D, Alves PC, Mage RG, Knight KL, van der Loo W, Esteves PJ. 2011 Molecular bases of genetic diversity and evolution of the immunoglobulin heavy chain variable region (IGHV) gene locus in leporids. *Immunogenetics* **63**, 397-408. (doi: 10.1007/s00251-011-0533-9)
70. Pinheiro A, de Mera IG, Alves PC, Gortazar C, de la Fuente J, Esteves PJ (2013) Sequencing of modern *Lepus* VDJ genes shows that the usage of VHn genes has been retained in both *Oryctolagus* and *Lepus* that diverged 12 million years ago. *Immunogenetics* **65** (11), 777-784. (doi:10.1007/s00251-013-0728-3)
71. Woof JM, Burton DR. 2004 Human antibody-Fc receptor interactions illuminated by crystal structures. *Nat. Rev. Immunol.* **4**, 89-99. (doi:10.1038/nri1266)
72. Woof JM, Partridge LJ, Jefferis R, Burton DR. 1986 Localisation of the monocyte-

- p>binding region on human immunoglobulin G.
- Mol. Immunol.*
- 23**
- , 319-330. (doi: 10.1016/0161-5890(86)90059-3)
73. Brezski RJ, Jordan RE. 2010 Cleavage of IgGs by proteases associated with invasive diseases: an evasion tactic against host immunity? *MAbs* **2**, 212-220. (doi: 10.4161/mabs.2.3.11780)
74. Sauer-Eriksson AE, Kleywegt GJ, Uhlen M, Jones TA. 1995 Crystal structure of the C2 fragment of streptococcal protein G in complex with the Fc domain of human IgG. *Structure* **3**, 265-278. (doi: 10.1016/S0969-2126(01)00157-5)
75. Lewis MJ, Meehan M, Owen P, Woof JM. 2008 A common theme in interaction of bacterial immunoglobulin-binding proteins with immunoglobulins illustrated in the equine system. *J. Biol. Chem.* **283**, 17615-17623. (doi: 10.1074/jbc.M709844200)
76. Diebolder CA, Beurskens FJ, de Jong RN, Koning RI, Strumane K, Lindorfer MA, Voorhorst M, Ugurlar D, Rosati S, Heck AJ, van de Winkel JG, Wilson IA, Koster AJ, Taylor RP, Saphire EO, Burton DR, Schuurman J, Gros P, Parren PW. 2014 Complement is activated by IgG hexamers assembled at the cell surface. *Science* **343**, 1260-1263. (doi: 10.1126/science.1248943)
77. Burnett RC, Hanly WC, Zhai SK, Knight KL. 1989 The IgA heavy-chain gene family in rabbit: cloning and sequence analysis of 13 C alpha genes. *EMBO J.* **8**, 4041-4047.
78. Woof JM, Russell MW. 2011 Structure and function relationships in IgA. *Mucosal Immunol.* **4**, 590-597. (doi: 10.1038/mi.2011.39)
79. Prahl JW, Mandy WJ, Todd CW. 1969 The molecular determinants of the A11 and A12 allotypic specificities in rabbit immunoglobulin. *Biochemistry* **8**, 4935-4940. (doi: 10.1021/bi00840a042)

Paper 6

Survey of genetic diversity for IgG in wild and domestic rabbits

Ana Pinheiro, Tereza Almeida and Pedro J. Esteves

International Journal of Immunogenetics 42(5), 364-7. doi: 10.1111/iji.12222

Abstract

We sequenced IgG from genomic DNA of 30 wild European rabbits of *O. c. algirus* and *O. c. cuniculus* subspecies from three regions and 15 domestic *O. c. cuniculus*. Genetic diversity was highest within Iberian wild populations. Only two new amino acid polymorphisms were found, both in *O. c. algirus*.

Introduction

Immunoglobulins are a key component of the adaptive immune system, linking antigen recognition to its elimination through several effector functions. IgG is the predominant serum immunoglobulin and has a wide range of functions. The European rabbit (*O. cuniculus*) has played an important role in immunological research (reviewed in Pinheiro *et al.*, 2011) and is unique in having only one IGHG copy (Knight *et al.*, 1985; Hamers, 1987; Ros *et al.*, 2004; Gertz *et al.*, 2013). The assessment of wild populations and domestic breeds IgG genetic diversity was previously done by serological analyses based on two polymorphisms located to the hinge region (d11/d12) and CH2 domain (e14/e15) (Hamers & Hamers-Casterman, 1965; Dubiski, 1969; Mage, 1981; Esteves *et al.*, 2002, 2006). However, nucleotide sequencing data is limited to the complete IGHG molecule for only a few European rabbits (Bernstein *et al.*, 1983; Martens *et al.*, 1984; Ros *et al.*, 2004; Pinheiro *et al.*, 2014).

The European rabbit originated in the Iberian Peninsula (IP), where two subspecies that diverged ~1.8 million years ago (Mya) exist: *O. cuniculus algirus* in the southwest and *O. cuniculus cuniculus* in the northeast. *O. c. cuniculus* later expanded its range North towards France, likely after the last glacial peak (Queney *et al.*, 2001), where it still remains present and from where the species was domesticated in the last 1500 years (Carneiro *et al.*, 2011, 2014).

Samples and Methods

To assess the genetic diversity of wild European rabbit populations we analysed 10 individuals for each of three populations: *O. c. algirus* collected in Portugal and *O. c. cuniculus* collected in Spain and in France (see Fig. 1). We further analysed 15 domestic animals from 6 breeds: French Lop, Flemish Giant, Argent Champagne, Angora, New Zealand White and Netherland Dwarf. Each of these four groups, the three wild populations and domestic breeds, represents a different genetic background and is exposed to different pathogen pressures.

PCR amplifications were performed on gDNA obtained from liver tissue. A fragment containing the IGHG CH1 and hinge domains was amplified using primers FG12 (5'-TCAGGCCCGAGCTGTAGACC-3') (Pinheiro *et al.* 2014) and RE (5'-ACGGTCCCCCAGGAGTTCA-3') (Esteves *et al.* 2006) under the following conditions: 15 min at 95°C followed by 35 cycles at 94°C (30 s), 63°C (30 s) and 72°C

(45 s), with a final extension at 60°C during 10 min. Another fragment containing IGHG CH2 and CH3 exons was amplified using primers F3 (5'-GTGAGTCCCATTAGCCTCAC-3') (Esteves et al. 2002) and RG31 (5'-TTGGAAGGAATCAGGACAGC-3') (Pinheiro et al. 2014) under the following conditions: 15 min at 95°C followed by 35 cycles at 94°C (30 s), 60°C (30 s) and 72°C (45 s), with a final extension at 60°C during 10 min. Sequences were determined by automated sequencing following the Big Dye Terminator Cycle Sequencing protocol (Perkin Elmer, Warrington, UK). Sequences were edited and aligned using CLUSTAL W (Thompson *et al.*, 1994) as implemented in BioEdit software (Hall, 1999) and haplotype reconstruction was obtained for each individual using PHASE version 2.1 (Stephens & Donnelly, 2003) as implemented in DnaSP version 5.10 (Librado & Rozas, 2009). DnaSP was used to estimate population genetic parameters (S , π), Tajima's D , number of haplotypes, haplotype diversity, variance of haplotype diversity and standard deviation of haplotype diversity.

Sequences obtained in this study were deposited in GenBank (accession numbers: KM502985-KM503029).

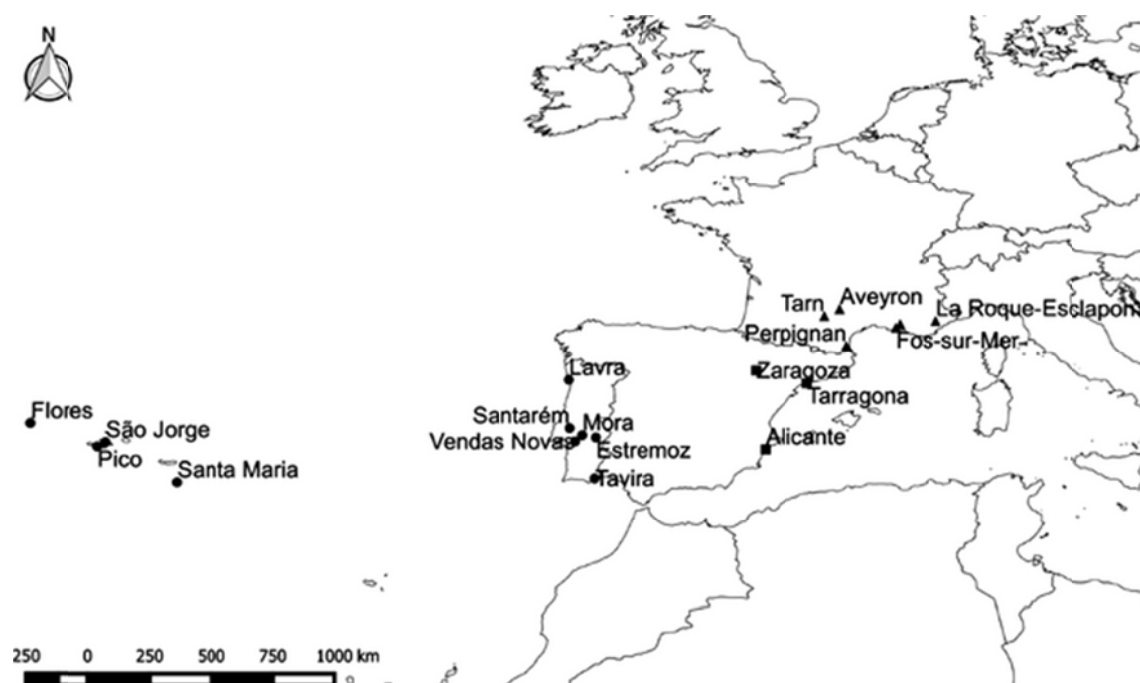


Figure 1- Map with the geographical locations of the populations used in this study. Marked with (●) are *O. c. algirus* populations from Portugal, with (■) are *O. c. cuniculus* from Spain and with (▲) *O. c. cuniculus* from France.

Results and Discussion

Previous studies of European rabbit have shown a decrease in genetic diversity of immunological markers from the wild European rabbit populations of the IP compared to the domestic animals, and that the domestic animals are, to a great extent, a subset of the diversity found in the French wild *O. c. cuniculus* (Abrantes *et al.*, 2013; Esteves *et al.*, 2004; Surridge *et al.*, 2008; van der Loo *et al.*, 1999). Our results showed a greater diversity in the IP populations: of the 18 SNPs found, 13 were observed for the *O. c. algirus* subspecies. On the other hand for French *O. c. cuniculus* only four SNPs were observed and for the Domestic breeds only three SNPs were detected (Table 1). This is confirmed by the nucleotide diversity values (Table 1). Interestingly for the domestic animals the nucleotide diversity is twice as high as for French *O. c. cuniculus*, likely reflecting changes in allele frequencies originated by the founder effects on the origin of domestic breeds (Carneiro *et al.*, 2011) However the Tajima's D test does not reject the null hypothesis that the Domestic breeds populations evolved under neutrality (Table 1).

Table 1- Immunoglobulin gamma nucleotide sequence polymorphism and estimates of haplotype diversity, haplotype diversity variance and haplotype standard deviation in the European rabbit populations/domestic animals studied.

Locus	Subspecies	Population	N	S	π	D	H	Hd	Hd variance	Hd std deviation
IGHG	<i>O. c. algirus</i>	All	45	18	0,0015	-1,6837	14	0,662	0,00289	0,054
		Portugal	10	13	0,0024	-1,2985	7	0,726	0,0084	0,092
		Spain	10	7	0,0014	-1,0339	5	0,616	0,01132	0,106
		France	10	4	0,0006	-1,4084	5	0,442	0,01774	0,133
		Domestic	15	3	0,0012	1,2116	6	0,715	0,00394	0,063

N, sample size; S, segregating sites; π , nucleotide diversity; D, Tajima's D statistical test; H, number of haplotypes; Hd, haplotype diversity; Hd variance, variance of haplotype diversity; Hd std deviation, standars deviation of haplotype diversity

Fourteen haplotypes were recovered (Table 1) with one widespread haplotype shared by all populations (0.567 frequency); remaining haplotypes are rare, occurring once in heterozygosity, or uncommon (0.01-0.10 frequencies). In addition to the most common haplotype, two haplotypes are shared between *O. c. algirus* and Spanish *O. c. cuniculus*, one haplotype is shared between Spanish and domestic *O. c. cuniculus* and

three haplotypes are shared between French and domestic *O. c. cuniculus*. Corroborating the results obtained for the nucleotide diversity, the haplotype diversity estimates show a decrease in genetic variability from the Iberian towards the French populations, and, again, the domestic breeds haplotype diversity is higher than that for the French populations. The higher levels of genetic diversity observed for the wild IP populations are in accordance with observations in previous studies and are likely due to the fact that the IP represents the native species range and that French *O. c. cuniculus* populations had its origin in a range expansion from the Spanish populations (Queney *et al.*, 2001). The high haplotype diversity found for the domestic breeds reinforces the hypotheses that the founder effect on the origin of domestic breeds (Carneiro *et al.*, 2011) caused changes in allele frequencies.

One third of the total number of segregating sites involves non-synonymous substitutions, resulting in six polymorphic amino acid positions (Tables 1 and 2). Unique amino acids for *O. c. algirus* are a Leu located at CH1 residue 85.1 and a Thr hinge residue 6. All studied *O. c. cuniculus* populations may have a Thr at CH3 residue 83. Spanish *O. c. cuniculus* have a unique Met at CH1 position 91 while French and domestic *O. c. cuniculus* share unique residues at hinge position 9 and CH2 residue 92. These two polymorphisms uniquely observed for the French population and domestic breeds correspond to the already described *d* and *e* locus (Hamers & Hamers-Casterman, 1965; Prah1 *et al.*, 1969; Appella *et al.*, 1971; Mage, 1981; Esteves *et al.*, 2002, 2006), which are absent in the IP, the native species range. Previous studies (Pinheiro *et al.*, 2014) have described for the European rabbit the four amino acid polymorphic positions we found for *O. c. cuniculus* wild populations and domestic breeds. However, our results show two new amino acid polymorphic positions in *O. c. algirus*, one in the CH1 domain, the other in the hinge region, the functional meaning of which remains unclear. Noteworthy is also the fact that the highest number of SNPs found for the wild Iberian populations does not translate into a higher number of amino acid polymorphic positions in these populations. The Iberian *O. cuniculus* populations have accumulated a higher number of random mutations, as expected for more ancient populations. On the other hand, the French population and domestic breeds seem to be under a higher selective pressure, possibly due to exposure to different pathogens, leading to the emergence and maintenance of mutations of functional significance. Indeed, van der Loo (1993) hypothesized 'compensatory overdominance' to explain the widespread occurrence of the d11 and e14 allotypes only outside the IP as a mechanism to compensate the lack of diversity shown by these 'new' populations at the a and b loci (*IGHV* and *IGKC*, respectively) when compared to the Iberian Peninsula populations.

Table 2- Amino acid polymorphic positions for European rabbit IgG populations/domestic animals. The codon numbering is according to the IMGT unique numbering for the constant domain (Lefranc et al., 2005)

Subspecies	Population	Domain, amino acid position and most common triplet and amino acid											
		CH1				hinge				CH2		CH3	
		85.1		91		6		9		92		83	
		tcg	Ser	gtg	Val	agc	Ser	acg	Thr	gcg	Ala	gcc	Ala
<i>O. c. algirus</i>	Portugal	...	Ser	...	Val	...	Ser	...	Thr	...	Ala	...	Ala
		.t.	Leu			.c.	Thr						
<i>O. c. cuniculus</i>	Spain	...	Ser	...	Val	...	Ser	...	Thr	...	Ala	...	Ala
				a..	Met							a..	Thr
	France	...	Ser	...	Val	...	Ser	...	Thr	...	Ala	...	Ala
								.t.	Met	a..	Thr	a..	Thr
	Domestic	...	Ser	...	Val	...	Ser	...	Thr	...	Ala	...	Ala
								.t.	Met	a..	Thr	a..	Thr

Acknowledgements

The Portuguese Foundation for Science and Technology (FCT) supported the doctoral fellowship of Ana Pinheiro (SFRH/BD/71252/2010). This work was supported by FCT project PTDC/BIA-ANM/3963/2012. “Genomics Applied to Genetic Resources” cofinanced by North Portugal Regional Operational Programme 2007/2013 (ON.2 – O Novo Norte), under the National Strategic Reference Framework (NSRF), through the European Regional Development Fund (ERDF) supported this work.

References

- Abrantes, J., Areal, H. & Esteves, P.J. (2013) Insights into the European rabbit (*Oryctolagus cuniculus*) innate immune system: genetic diversity of the toll-like receptor 3 (TLR3) in wild populations and domestic breeds. *BMC Genetics*, **21**, 14.
- Appella, E., Chersi, A., Mage, R.G. & Dubiski, S. (1971) Structural basis of the A14 and A15 allotypic specificities in rabbit immunoglobulin G. *Proceedings of the National Academy of Sciences of the USA*, **68**, 1341.
- Bernstein, K.E., Alexander, C.B. & Mage, R.G. (1983) Nucleotide sequence of a rabbit IgG heavy chain from the recombinant F-I haplotype. *Immunogenetics*, **18**, 387.
- Carneiro, M., Afonso, S., Geraldès, A., Garreau, H., Bolet, G., Boucher, S., Tircazes, A., Queney, G., Nachman, M.W. & Ferrand, N. (2011) The genetic structure of domestic rabbits. *Molecular Biology and Evolution*, **28**, 1801.

Carneiro, M., Rubin, C.J., Di Palma, F., Albert, F.W., Alföldi, J., Barrio, A.M. et al. (2014) Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. *Science*, **345**, 1074.

Dubiski, S. (1969) Immunochemistry and genetics of a “new” allotypic specificity Ae14 of rabbit gamma-G immunoglobulins: recombination in somatic cells. *Journal of Immunology*, **103**, 120

Esteves, P.J., Alves, P.C., Ferrand, N. & van der Loo, W. (2002) Hotspot variation at the CH2-CH3 interface of leporid IgG antibodies (*Oryctolagus*, *Sylvilagus* and *Lepus*). *European Journal of Immunogenetics*, **29**, 529.

Esteves, P.J., Lanning, D., Ferrand, N., Knight, K.L., Zhai, S.K. & van der Loo, W. (2004) Allelic variation at the VHa locus in natural populations of rabbit (*Oryctolagus cuniculus*, L.). *The Journal of Immunology*, **172**, 1044.

Esteves, P.J., Carmo, C., Godinho, R. & van der Loo, W. (2006) Genetic diversity at the hinge region of the unique immunoglobulin heavy gamma (IGHG) gene in leporids (*Oryctolagus*, *Sylvilagus* and *Lepus*). *International Journal of Immunogenetics*, **33**, 171.

Gertz, E.M., Schäffer, A.A., Agarwala, R., Bonnet-Garnier, A., Rogel-Gaillard, C., Hayes, H. & Mage, R.G. (2013) Accuracy and coverage assessment of *Oryctolagus cuniculus* (rabbit) genes encoding immunoglobulins in the whole genome sequence assembly (OryCun2.0) and localization of the IGH locus to chromosome 20. *Immunogenetics*, **65**, 749.

Hall, T.A. (1999) bioedit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**, 95.

Hamers, R. (1987) Allotopy of immunoglobulins. In: *The Rabbit in Contemporary Immunological Research* (ed. by S. Dubiski), p. 65. Longman Scientific & Technical Harlow, Essex, England.

Hamers, R., Hamers-Casterman, C. (1965) Molecular localization of A chain allotypic specificities in rabbit IgG (7S gamma-globulin). *The Journal of Molecular Biology*, **14**, 288.

Knight, K.L., Burnett, R.C. & McNicholas, J.M. (1985) Organization and polymorphism of rabbit immunoglobulin heavy chain genes. *Journal of Immunology*, **134**, 1245.

Lefranc, M.P., Pommie, C., Kaas, Q., Duprat, E., Bosc, N., Guiraudou, D., Jean, C., Ruiz, M., Da Piedade, I., Rouard, M., Foulquier, E., Thouvenin, V. & Lefranc, G. (2005) IMGT unique numbering for immunoglobulin and T cell receptor constant domains and Ig superfamily C-like domains. *Developmental and Comparative Immunology*, **29**, 185.

Librado, P. & Rozas, J. (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451.

- Mage, R. (1981) The phenotypic expression of rabbit immunoglobulins: a model of complex regulated gene expression and cellular differentiation. *Contemporary Topics in Molecular Immunology*, **8**, 89.
- Martens, C.L., Currier, S.J. & Knight, K.L. (1984) Molecular genetic analysis of genes encoding the heavy chains of rabbit IgG. *Journal of Immunology*, **133**, 1022.
- Pinheiro, A., Lanning, D., Alves, P. C., Mage, R.G., Knight, K.L., van der Loo, W. & Esteves, P.J. (2011) Molecular bases of genetic diversity and evolution of the immunoglobulin heavy chain variable region (IGHV) gene locus in leporids. *Immunogenetics*, **63**, 397.
- Pinheiro, A., Woof, J.M., Almeida, T., Abrantes, J., Alves, P.C., Gortázar, C. & Esteves, P.J. (2014) Leporid immunoglobulin G shows evidence of strong selective pressure on the hinge and CH3 domains. *Open Biology*, **4**, 140088.
- Prahl, J.W., Mandy, W.J. & Todd, C.W. (1969) The molecular determinants of the A11 and A12 allotypic specificities in rabbit immunoglobulin. *Biochemistry*, **8**, 12.
- Queney, G., Ferrand, N., Weiss, S., Mougél, F. & Monnerot, M. (2001) Stationary distributions of microsatellite loci between divergent population groups of the European rabbit (*Oryctolagus cuniculus*). *Molecular Biology and Evolution*, **18**, 2169.
- Ros, F., Puels, J., Reichenberger, N., van Schooten, W., Buelow, R. & Platzer, J. (2004) Sequence analysis of 0.5 Mb of the rabbit germline immunoglobulin heavy chain locus. *Gene*, **330**, 49.
- Stephens, M. & Donnelly, P. (2003) A comparison of bayesian methods for haplotype reconstruction from population genotype data. *The American Journal of Human Genetics*, **73**, 1162.
- Surridge, A.K., van der Loo, W., Abrantes, J., Carneiro, M., Hewitt, G.M. & Esteves, P.J. (2008) Diversity and evolutionary history of the MHC DQA gene in leporids. *Immunogenetics*, **60**, 515.
- Thompson, J.D., Higgins, D.G. & Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, **22**, 4673.
- van der Loo, W. (1993) Variance analysis of immunoglobulin alleles in natural populations of rabbit (*Oryctolagus cuniculus*) : the extensive interallelic divergence at the b locus could be the outcome of overdominance-type selection. *Genetics*, **134**, 171.
- van der Loo, W., Mougél, F., Sánchez, M.S., Bouton, C., Castien, E., Fonseca, A., Ferrand, N., Soriguer, R. & Monnerot, M. (1999) Cytonuclear disequilibria in wild populations of rabbit (*Oryctolagus cuniculus* L.) suggest unequal allele turnover rates at the b locus (IGKC1). *Immunogenetics*, **49**, 629.

Chapter 5

General Discussion

1. Diversifying selection acting on the Ig-Fc fragment

Immunoglobulins are a key component of the vertebrate immune system. These bifunctional molecules operate as a critical part of the immune response by recognizing and binding to antigens as well as recruiting the effector functions that lead to pathogen elimination. To successfully infect the host, pathogens must evade Igs action. It is not surprising that numerous pathogens have evolved Ig directed strategies to escape the immune system, and similar strategies have been adopted by bacterial pathogens of different mammalian species (Lewis et al. 2008). Among these are Ig-binding proteins and proteases. The Ig-binding proteins bind the Ig-Fc in a location that prevents the coupling of the antigen-Ig complex to its receptors. Proteases, on the other hand, usually cleave the Ig hinge, again uncoupling the recognition of foreign antigens from the effector functions that eliminate them. Examples of Ig-binding proteins include the IgG directed Staphylococcal protein A, Streptococcal protein G (Sauer-Eriksson et al. 1995; Tashiro and Montelione 1995), and *Streptococcus equi* FgBP (Meehan et al. 2001) and IgA directed Sir22 and Arp4 proteins of *Streptococcus pyogenes* (Frithz et al. 1989; Stenberg et al. 1994), the β protein of *Streptococcus agalactiae* (Jerlstrom et al. 1991), and the SSL7 toxin of *Staphylococcus aureus* (Ramsland et al. 2007); curiously all of these proteins bind residues in the CH2-CH3 interface of IgG and IgA blocking access of Fc receptors (Lewis et al. 2008; Pleass et al. 2001; Wines et al. 2006). Examples of proteases include *Streptococcus pyogenes* IdeS and *Staphylococcus aureus* GluV8 that cleave the hinge region of human IgG1 (reviewed in Brezski and Jordan 2010) and IgA1 proteases of *Neisseria meningitidis* and *Haemophilus influenzae*, which cleave the hinge region of IgA1 of humans and great apes (Batten et al. 2003; Senior and Woof 2005).

Given this context of multiple pathogens targeting IgG and IgA Fc to avoid the immune system, it was expected to find signatures of positive selection on these Ig constant region genes. However, studies on how the host-pathogen interaction has impacted Igs evolution are very scarce (e.g. Abi-Rached et al. 2007). Our studies on mammalian IgA (Pinheiro et al. 2013a, paper 4 on Chapter 3 of this thesis) and leporid IgG (Pinheiro et al. 2014a, 2015, papers 5 and 6 on Chapter 4 of this thesis) showed that these Ig isotypes are in fact evolving under strong Darwinian selection most likely imposed by pathogens. For mammalian IgA 18 residues under positive selection were identified in all three IgA heavy chain constant region domains, the majority of which located near sites of functional relevance for pathogen resistance. The results we obtained for leporids IgG also identified seven residues under positive selection located

in parts of the molecule that are essential for IgG functions and genetic characteristics of the leporid IgG molecule that may well be of functional relevance. These findings can be further tested by functional studies to determine their role in resistance against pathogens. Ligands to IgG and IgA, such as Fc receptors and C1q, the first component of complement that binds to IgG should also be investigated. In their study, Abi-Rached et al. (2007) investigated the co-evolution of IgA, its receptor Fc α RI and *Staphylococcus aureus* SSL7 toxin, and found that mammalian IgA is evolving under pathogen pressure to change and then IgA is, in turn, exerting pressure on Fc α RI. We identified as positively selected IgA residues that bind or directly influence binding of pIgR and IgA1 protease of *H. influenza* to IgA Fc and, thus, investigating these IgA ligands may unveil similar stories. From an evolutionary study of receptor-ligand interaction comes a remarkable example of co-evolution, in leporids. Indeed, the study of the chemokine receptor CCR5 in leporids led to the discovery of a gene conversion with CCR2 in European rabbit, riverine rabbit (*Bunolagus monticularis*) and Amami rabbit (*Pentalagus furnessi*) but not in other leporids, such as the eastern cottontail (*Sylvilagus floridanus*) and European and Iberian hares (*Lepus europeaus* and *Lepus granatensis*) (Abrantes et al. 2011; Carmo et al. 2006). These results prompted studies of CCR5 ligands C–C motif chemokines CCL3, CCL4, CCL5, CCL7, CCL8, CCL13 and CCL14 (van der Loo et al. 2012; de Matos et al. 2014; Neves et al. 2015). Among these chemokines, CCL8 was found to be pseudogenized in European rabbit and riverine rabbit while functional in *Lepus* and *Sylvilagus* species suggesting that these two events may well be related (van der Loo et al. 2012).

2. Insights in the *IGHV* genes evolution

As stated in the previous section, it was expected to find selection acting on the Igs constant region. Regarding the variable segment, the heart of the antibody diversity that traditionally is viewed as random, conceptually it is not predictable to be under natural selection. In fact, the vast, supposedly unlimited, Ig repertoire is thought of as randomly generated by first VDJ recombination and then gene conversion and somatic hypermutation, with addition of nucleotides, to allow recognition of a limitless array of antigens. These processes of somatic diversification followed by the somatic selection of immunoglobulins would mean that the germline is not selected (summarized in Langman and Cohn, 2002). However, growing evidence shows otherwise. The first sign that even the V segment is subject to evolutionary constraints was the observation

of sequence homology from sharks to humans that reflects in the sectioning into FRs (framework regions) and CDRs (complementarity-determining regions) of the V sequence. The FRs are four regions of relatively invariable sequence which has been conserved across evolution (Kirkham et al. 1992; Kirkham and Schroeder 1994). Based on the sequence similarity of specific intervals of FR1 and FR3, the vertebrate VH genes have been classified into three clans and, within these clans, into families (Kirkham et al. 1992; Kirkham and Schroeder 1994). Considering the Ig structure, the FRs provide the backbone that lends support to the antigen binding site, created by the juxtaposition of the CDRs. Thus, the conservation of the FRs may be explained by the need to maintain a properly shaped antigen binding site (Kirkham and Schroeder 1994). But also the V genes hypervariable regions, the CDRs, seem to be evolutionary constrained rather than random. In fact, it has been shown that despite all the variability of the CDRs, five of these regions, L1, L2, L3, H1 and H2 (light chains CDR 1, 2 and 3 and heavy chain CDR1 and 2), fit into a rather restricted number of main chain conformations, which have been called canonical structures. It is the length of the loop, as well as residues present at key sites, that determine the conformation of each canonical structure; based on these criteria 18 different canonical structures were described when taken together the CDRs of VL κ and λ as well as VH chains (reviewed in Al-Lazikani et al. 1997). More recently, biases in the amino acid composition of the antigen binding site, particularly of the CDR-H3, were observed and found to be conserved in all jawed vertebrates (Ivanov et al. 2002, 2005). This bias is present at a stage of B-cell development prior to the surface expression of Igs and thus may well be the product of germline selection (Schroeder et al. 2010). In fact, manipulating the germline DH sequence of mice to force the usage of a DH with shifted amino acid bias, i.e. that contains arginines instead of tyrosines, led to an immune deficient phenotype and to the production of auto-antibodies. This shows that natural selection acts to enforce the observed bias in amino acid usage in the germline as to maintain a 'healthy' and effective antibody repertoire (Schroeder et al. 2010, Vale et al. 2013). The results obtained in this thesis (Pinheiro et al. 2013b, 2014b, papers 2 and 3 in Chapter 2) concur to show that natural selection is in fact operating to shape the leporid VH germline diversity. We have showed that the *VHn* genes are a conserved ancestral polymorphism that is being maintained in the leporids genome and are expressed in low frequency both by the European rabbit and hares (Kim and Dray 1973; Roux 1981; Pinheiro et al. 2013b). Interestingly, the *VHn* expression was not observed in the American Leporid *Sylvilagus floridanus* (Pinheiro et al. 2014b, paper 3 in Chapter 2 of this thesis). Our results also indicate that the loss of a more ancient lineage of *VH* genes may have imposed the diversification of the *VHa* lineages in European rabbit

(Pinheiro et al. 2014b, paper 3 in Chapter 2 of this thesis). These results are suggestive of natural selection shaping the leporids germline antibody diversity. In fact, the preferential usage of the D-proximal *VH1* gene in the generation of 80-90% of the rabbit antibody repertoire, in spite possessing several functional *VH* genes, (Knight and Becker 1990; Knight 1992) can be viewed as a first indication that the rabbit antibody repertoire is subject to selective pressures. The specificities carried by this *VH1* gene allelic lineages are so relevant to the proper functioning of the rabbit immune system that, as showed by the study of the Alicia rabbit (Kelus and Weiss 1986), when lost this *VH1* rabbits recreate the *VHa* allelic lineage specificities that was present in their progenitors germline by somatic gene conversion (Sehgal et al. 1998; Zhu et al. 1999).

3. Major conclusions

The studies developed here have contributed to broaden the current understanding of the evolution of Leporid IGH genes, which was the main goal of this thesis. The major findings of this work are:

- i) The *VHn* genes are a conserved ancestral polymorphism that is being maintained in the Leporids genome;
- ii) The *VHn* lineages are being used for the generation of *VDJ* rearrangements for at least 12 million years, the estimated divergence time between *Oryctolagus* and *Lepus* genera;
- iii) The low frequency expression of *VHn* genes both in rabbit and hares, suggests that the antibody repertoire in leporids is subject to selective pressure;
- iv) The *VHa* allotypes have an origin that pre-dates the Leporid radiation: some *VHa* lineages are shared between species, such as the a2 and a2L lineages, whereas others apparently are species specific;
- v) The studied leporids employ different *VH* lineages in the generation of the antibody repertoire, suggesting that the leporid *VH* genes are subject to strong selective pressure likely imposed by specific pathogens;
- vi) During mammalian evolution positive selection has targeted all three IgA constant domains, with a greater incidence in the C α 1 and C α 2 domains. The positively-selected residues have influence on the interactions of IgA with immune-system receptors and the microbial proteins that interfere with these interactions.

- vii) The evolution of IgA in mammals and in the diversity and divergence of IgA among extant mammalian species has been driven to great extent by positive selection;
- viii) The Leporids share considerable sequence similarity for their IGHG (approximately 94%);
- ix) A gradient in leporid IgG constant domain diversity was observed with the CH1 being the most conserved and the CH3 the most variable domain. Positive selection was found to be acting on all constant domains, but with a greater incidence in the CH3 domain, where a cluster of three positively selected sites was identified. Thus, Leporid IgG CH3 domain is evolving under selective pressure to change;
- x) Unlike the variation observed for the European rabbit IgG, all 11 studied *Lepus* species share exactly the same hinge motif, indicating that it may be being maintained due to an advantageous structure or conformation;

4. Future Perspectives

The European rabbit has the most complex IgA system of all studied mammals with 13 *IGHA* genes (Burnett et al. 1989) which remains an intriguing feature. Burnett et al (1989), using southern blot analysis, showed that the duplication of the C_{α} genes began in a Lagomorph ancestral, since *Sylvilagus*, *Lepus*, *Pentalagus* and *Ochotona* genera have multiple IgA copies, but how many copies each Lagomorph has and if copies are shared between species is yet to be determined. Pursuing these questions would contribute to understand the evolutionary dynamics of this Ig isotype in Lagomorphs. Our study (Pinheiro et al. 2013a, paper 4 in Chapter 3) suggests that mammalian IgA is evolving under pathogen pressure, could pathogens be imposing similar pressure on the Lagomorphs IgA? Another aspect that remains to be assessed is whether the single IgG is common to all Lagomorphs. An alternative hypothesis to explain the extreme number of IgA copies is that the duplication of IgA arose as a compensatory mechanism to the absence of multiple IgG subclasses, as most studied mammals have multiple IgG (Flanagan and Rabbitts 1982; Shimizu et al. 1982; Bruggemann et al. 1986; Knight et al. 1988; Wagner et al. 2004; Butler et al. 2009, 2011). Further, our studies show that both IgA and IgG are evolving under selective pressure (Pinheiro et al. 2013a, 2014a, 2015, papers 4, 5 and 6 in Chapters 3 and 4).

Extending the analysis of selection to other Ig isotypes, namely IgM and IgE, would be a logical next step. IgA and IgG are probably the Ig isotypes that have the most complex interactions with their surroundings being exposed to a wider variety of pathogens. IgM is the first isotype to be expressed during B-cell development and after antigenic stimulation isotype switching occurs. As for IgE it is associated with hypersensitivity, allergic reactions and response to parasitic worm infections (Schroeder and Cavacini 2010). As such, analyzing the later Ig isotypes for selection will probably yield contrasting results with those obtained for IgG and IgA.

The studies conducted here have contributed to advance knowledge on the leporids *VH* repertoire and usage. Our results suggest a possible role for the odd low frequency usage of the *VHn* genes and seem to indicate that the leporids *VH* lineages are under selective pressure (Pinheiro et al. 2013b, 2014b, papers 2 and 3 in Chapter 2). To clarify if the *VH* lineages diversification and their usage in leporids are truly a product of natural selection the *VDJ* genes of additional American leporids should be sequenced. Also interesting would be to extend this analysis to ochotonids. Ochotonids, which along with Leporids form the Lagomorphs, belong to a single *genus*, *Ochotona*, present in Asia and North America (Lisovsky, 2014). Apart from a few *Ochotona* germline *VH* sequences (Esteves, 2003) no data is available for this *genus* *VH* genes. Sequencing *VDJ* genes for Asian and American *Ochotona* species will allow comparing the evolutionary history of *VH* genes in both Ochotonidae and Leporidae families. As a first step it is necessary to assess if, like the European rabbit (Knight and Becker 1990; Knight 1992) and apparently other Leporids (Esteves et al. 2005; Pinheiro et al. 2013b, 2014b), *Ochotona* preferentially expresses the *VH1* gene in the generation of its antibody repertoire

Finally, the European rabbit is also unique by the fact that the diversification of the primary Ab repertoire is not developmentally programmed, contrarily to what happens in other vertebrates such as chickens (e.g. Reynaud et al. 1989), sheep (Reynaud et al. 1995) and cattle (e.g. Lucier et al. 1998). In fact, the European rabbit peripheral Ab repertoire is diversified between 4 weeks and 2 months after birth (Cooper et al. 1968; Weinstein et al. 1994) and requires the integrity of the GALT (Vadhy et al. 1998) as well as exogenous factors, such as the intestinal microbiota (Lanning et al. 2000a, 2000b; Sehgal et al. 2002). It has been suggested that the post-natal diversification of the European rabbit Ab repertoire could be related to the shortening of gestation time, since the rabbit is a species with altricial young born after a four weeks gestation as opposed to, e.g., hares that have precocial young and a gestation time of six weeks (Pinheiro et al. 2011). This hypothesis remains to be validated and so investigating whether in other leporids with precocial young the Ab

repertoire is diversified neonatally or during foetal life is an interesting question to pursue.

5. References

- Abi-Rached L, Dorighi K, Norman PJ, Yawata M, Parham P (2007) Episodes of natural selection shaped the interactions of IgA-Fc with FcαRI and bacterial decoy proteins. *J Immunol* 178:7943-7954
- Abrantes J, Carmo CR, Matthee CA, Yamada F, Loo Wvd, Esteves PJ (2011) A shared unusual genetic change at the chemokine receptor type 5 between *Oryctolagus*, *Bunolagus* and *Pentalagus*. *Conserv Genet* 12:325-330 doi:DOI 10.1007/s10592-0099990-1
- Al-Lazikani B, Lesk AM, Chothia C (1997) Standard conformations for the canonical structures of immunoglobulins. *J Mol Biol* 273(4):927-48 doi:10.1006/jmbi.1997.1354
- Batten MR, Senior BW, Kilian M, Woof JM (2003) Amino acid sequence requirements in the hinge of human immunoglobulin A1 (IgA1) for cleavage by streptococcal IgA1 proteases. *Infect Immun* 71:1462-1469
- Brezski RJ, Jordan RE (2010) Cleavage of IgGs by proteases associated with invasive diseases: an evasion tactic against host immunity? *MAbs* 2:212-220
- Bruggemann M, Free J, Diamond A, Howard J, Cobbold S, Waldmann H (1986) Immunoglobulin heavy chain locus of the rat: striking homology to mouse antibody genes. *Proc. Natl. Acad. Sci. USA* 83:6075-6079.
- Burnett RC, Hanly WC, Zhai SK, Knight KL (1989) The IgA heavy-chain gene family in rabbit: cloning and sequence analysis of 13 C alpha genes *EMBO J* 8:4041-4047
- Butler JE, Wertz N, Deschacht N, Kacs Kovics I (2009) Porcine IgG: structure, genetics, and evolution. *Immunogenetics* 61:209-230. (doi: 10.1007/s00251-008-0336-9)
- Butler JE, Wertz N, Zhao Y, Zhang S, Bao Y, Bratsch S, Kunz TH, Whitaker JO, Jr., Schountz T (2011) The two suborders of chiropterans have the canonical heavy-chain immunoglobulin (Ig) gene repertoire of eutherian mammals. *Dev. Comp. Immunol.* 35:273-284. (doi: 10.1016/j.dci.2010.08.011)
- Carmo CR, Esteves PJ, Ferrand N, van der Loo W (2006) Genetic variation at chemokine receptor CCR5 in leporids: alteration at the 2nd extracellular domain by gene conversion with CCR2 in *Oryctolagus*, but not in *Sylvilagus* and *Lepus* species. *Immunogenetics* 58:494-501.

- Cooper MD, Perey DY, Gabrielsen AE, Sutherland DE, McKneally MF, Good RA (1968) Production of an antibody deficiency syndrome in rabbits by neonatal removal of organized intestinal lymphoid tissues. *Int Arch Allergy Appl Immunol* 33 (1):65-88
- de Matos AL, Lanning DK, Esteves PJ (2014) Genetic characterization of CCL3, CCL4 and CCL5 in leporid genera *Oryctolagus*, *Sylvilagus* and *Lepus*. *International Journal of Immunogenetics* 41:154-158 doi:10.1111/iji.12095
- Esteves PJ (2003) Molecular and population genetic analysis of polymorphism at the antibody loci IgGCH2 and IgVH in lagomorphs. Faculdade de Ciências Universidade do Porto, Porto
- Esteves PJ, Lanning D, Ferrand N, Knight KL, Zhai SK, van der Loo W (2005) The evolution of the immunoglobulin heavy chain variable region (IgVH) in leporids: an unusual case of transspecies polymorphism. *Immunogenetics* 57 (11):874-882.
- Flanagan JG, Rabbitts TH (1982) Arrangement of human immunoglobulin heavy chain constant region genes implies evolutionary duplication of a segment containing gamma, epsilon and alpha genes. *Nature* 300:709-713. (doi:10.1038/300709a0)
- Frithz E, Heden LO, Lindahl G (1989) Extensive sequence homology between IgA receptor and M proteins in *Streptococcus pyogenes*. *Mol Microbiol* 3:1111-1119
- Ivanov II, Link JM, Ippolito GC, Schroeder HW Jr. (2002) Constraints on hydropathicity and sequence composition of HCDR3 are conserved across evolution. In: Zanetti M, Capra JD, editors. *The Antibodies*. London: Taylor and Francis Group. pp. 43–67.
- Ivanov II, Schelonka RL, Zhuang Y, Gartland GL, Zemlin M, Schroeder HW Jr. (2005) Development of the expressed immunoglobulin CDR-H3 repertoire is marked by focusing of constraints in length, amino acid utilization, and charge that are first established in early B cell progenitors. *J Immunol* 174:7773–7780.
- Jerlstrom PG, Chhatwal GS, Timmis KN (1991) The IgA-binding beta antigen of the c protein complex of Group B streptococci: sequence determination of its gene and detection of two binding regions. *Mol Microbiol* 5:843-849
- Kelus AS, Weiss S (1986) Mutation affecting the expression of immunoglobulin variable regions in the rabbit. *Proc Natl Acad Sci U S A* 83 (13):4883-4886
- Kim BS, Dray S (1973) Expression of the a, x, and y variable region genes of heavy chains among IgG, IgM, and IgA molecules of normal and a locus allotype-suppressed rabbits. *J Immunol* 111 (3):750-760

- Kirkham PM, Mortari F, Newton JA, Schroeder HW Jr. (1992) Immunoglobulin VH clan and family identity predicts variable domain structure and may influence antigen binding. EMBO J 11(2):603-9
- Kirkham PM, Schroeder HW Jr. (1994) Antibody structure and the evolution of immunoglobulin V gene segments. Semin Immunol 6(6):347-60
- Knight KL, Suter M, Becker RS (1988) Genetic engineering of bovine Ig. Construction and characterization of hapten-binding bovine/murine chimeric IgE, IgA, IgG1, IgG2, and IgG3 molecules. J. Immunol. 140:3654-3659.
- Knight KL (1992) Restricted VH gene usage and generation of antibody diversity in rabbit. Annu Rev Immunol 10:593-616
- Knight KL, Becker RS (1990) Molecular basis of the allelic inheritance of rabbit immunoglobulin VH allotypes: implications for the generation of antibody diversity. Cell 60 (6):963-970
- Langman RE, Cohn M (2002) If the immune repertoire evolved to be large, random, and somatically generated, then... Cell Immunol 216(1-2):15-22
- Lanning D, Sethupathi P, Rhee KJ, Zhai SK, Knight KL (2000a) Intestinal microflora and diversification of the rabbit antibody repertoire. J Immunol 165 (4):2012-2019.
- Lanning D, Zhu X, Zhai SK, Knight KL (2000b) Development of the antibody repertoire in rabbit: gut-associated lymphoid tissue, microbes, and selection. Immunol Rev 175:214-228
- Lewis MJ, Meehan M, Owen P, Woof JM (2008) A common theme in interaction of bacterial immunoglobulin-binding proteins with immunoglobulins illustrated in the equine system. J Biol Chem 283:17615-17623 doi:10.1074/jbc.M709844200
- Lisovsky A (2014) Taxonomic revision of pikas *Ochotona* (Lagomorpha, Mammalia) at the species level. Mammalia 78:199–216
- Lucier MR, Thompson RE, Waire J, Lin AW, Osborne BA, Goldsby RA (1998) Multiple sites of V lambda diversification in cattle. J Immunol 161 (10):5438-5444
- Meehan M, Lynagh Y, Woods C, Owen P (2001) The fibrinogen-binding protein (FgBP) of *Streptococcus equi* subsp. *equi* additionally binds IgG and contributes to virulence in a mouse model. Microbiology 147:3311-3322 doi:10.1099/00221287-147-12-3311
- Neves F, Abrantes J, Lisovsky AA, Esteves PJ (2015) Pseudogenization of CCL14 in the Ochotonidae (pika) family. Innate immunity doi:10.1177/1753425915577455
- Pinheiro A, Lanning D, Alves PC, Mage RG, Knight KL, Loo Wvd, Esteves PJ (2011). Molecular bases of genetic diversity and evolution of the Immunoglobulin Heavy

- chain Variable region (*IGHV*) gene locus in leporids. *Immunogenetics*, 63(7), 397-408. doi: 10.1007/s00251-011-0533-9.
- Pinheiro A, de Mera IG, Alves PC, Gortázar C, de la Fuente J, Esteves PJ (2013a) Sequencing of modern *Lepus* VDJ genes shows that the usage of VHn genes has been retained in both *Oryctolagus* and *Lepus* that diverged 12 million years ago. *Immunogenetics*, 65(11), 777-84. doi: 10.1007/s00251-013-0728-3.
- Pinheiro A, Woof JM, Abi-Rached L, Parham P, Esteves PJ (2013b) Computational analyses of an evolutionary arms race between mammalian immunity mediated by immunoglobulin A and its subversion by bacterial pathogens. *PLoS One*, 8(9), e73934. doi: 10.1371/journal.pone.0073934. eCollection 2013.
- Pinheiro A, Melo-Ferreira J, Abrantes J, Martinelli N, Lavazza A, Alves PC, Gortázar C, Esteves PJ. (2014a) Sequencing of *Sylvilagus* VDJ genes reveals a new VH_a allelic lineage and shows that ancient VH lineages were retained differently in leporids. *Immunogenetics*, 66(12), 719-26. doi: 10.1007/s00251-014-0807-0
- Pinheiro A, Woof JM, Almeida T, Abrantes J, Alves PC, Gortázar C, Esteves PJ (2014b) Leporid immunoglobulin G shows evidence of strong selective pressure on the hinge and CH3 domains. *Open Biology*, 4(9), 140088. doi: 10.1098/rsob.140088.
- Pinheiro A, Almeida T, Esteves PJ (2015) Survey of genetic diversity of IgG in wild and domestic rabbits. *International Journal of Immunogenetics* 42(5), 364-7. doi: 10.1111/iji.12222
- Pleass RJ, Areschoug T, Lindahl G, Woof JM (2001) Streptococcal IgA-binding proteins bind in the Calpha 2-Calpha 3 interdomain region and inhibit binding of IgA to human CD89. *J Biol Chem* 276:8197-8204 doi:10.1074/jbc.M009396200
- Ramsland PA, Willoughby N, Trist HM, Farrugia W, Hogarth PM, Fraser JD, Wines BD (2007) Structural basis for evasion of IgA immunity by *Staphylococcus aureus* revealed in the complex of SSL7 with Fc of human IgA1. *Proc Natl Acad Sci U S A* 104:15051-15056 doi:10.1073/pnas.0706028104
- Reynaud CA, Dahan A, Anquez V, Weill JC (1989) Somatic hyperconversion diversifies the single Vh gene of the chicken with a high incidence in the D region. *Cell* 59 (1):171-183.
- Reynaud CA, Garcia C, Hein WR, Weill JC (1995) Hypermutation generating the sheep immunoglobulin repertoire is an antigen-independent process. *Cell* 80 (1):115-125.
- Roux KH (1981) A fourth heavy chain variable region subgroup, w, with 2 variants defined by an induced auto-antiserum in the rabbit. *J Immunol* 127 (2):626-632

- Sauer-Eriksson AE, Kleywegt GJ, Uhlen M, Jones TA (1995) Crystal structure of the C2 fragment of streptococcal protein G in complex with the Fc domain of human IgG. *Structure* 3:265-278
- Schroeder HW, Jr., Cavacini L (2010) Structure and function of immunoglobulins. *J Allergy Clin Immunol* 125:S41-52 doi:10.1016/j.jaci.2009.09.046
- Schroeder HW Jr., Zemlin M, Khass M, Nguyen HH, Schelonka RL (2010) Genetic control of DH reading frame and its effect on B-cell development and antigen-specific antibody production. *Crit Rev Immunol* 30:327–344.
- Sehgal D, Mage RG, Schiaffella E (1998) VH mutant rabbits lacking the VH1a2 gene develop a2+ B cells in the appendix by gene conversion-like alteration of a rearranged VH4 gene. *J Immunol* 160 (3):1246-1255
- Sehgal D, Obiakor H, Mage RG (2002) Distinct clonal Ig diversification patterns in young appendix compared to antigen-specific splenic clones. *J Immunol* 168 (11):5424-5433
- Senior BW, Woof JM (2005) Effect of mutations in the human immunoglobulin A1 (IgA1) hinge on its susceptibility to cleavage by diverse bacterial IgA1 proteases. *Infect Immun* 73:1515-1522 doi:10.1128/IAI.73.3.1515-1522.2005
- Shimizu A, Takahashi N, Yaoita Y, Honjo T (1982) Organization of the constant-region gene family of the mouse immunoglobulin heavy chain. *Cell* 28:499-506. (doi: 10.1016/0092-8674(82)90204-5)
- Stenberg L, O'Toole PW, Mestecky J, Lindahl G (1994) Molecular characterization of protein Sir, a streptococcal cell surface protein that binds both immunoglobulin A and immunoglobulin G. *J Biol Chem* 269:13458-13464
- Tashiro M, Montelione GT (1995) Structures of bacterial immunoglobulin-binding domains and their complexes with immunoglobulins. *Curr Opin Struct Biol* 5:471-481
- Vajdy M, Sethupathi P, Knight KL (1998) Dependence of antibody somatic diversification on gut-associated lymphoid tissue in rabbits. *J Immunol* 160 (6):2725-2729
- Vale AM, Kapoor P, Skibinski GA, Elgavish A, Mahmoud TI, Zemlin C, Zemlin M, Burrows PD, Nobrega A, Kearney JF, Briles DE, Schroeder HW Jr. (2013) The link between antibodies to OxLDL and natural protection against pneumococci depends on DH gene conservation. *J Exp Med* 210:875–890.
- van der Loo W, Afonso S, de Matos AL, Abrantes J, Esteves PJ (2012) Pseudogenization of the MCP-2/CCL8 chemokine gene in European rabbit (genus *Oryctolagus*), but not in species of Cottontail rabbit (*Sylvilagus*) and Hare (*Lepus*). *BMC genetics* 13:72 doi:10.1186/1471-2156-1372

- Wagner B, Miller DC, Lear TL, Antczak DF. 2004 The complete map of the Ig heavy chain constant gene region reveals evidence for seven IgG isotypes and for IgD in the horse. *J. Immunol* 173:3230-3242. (doi: 10.4049/jimmunol.173.5.3230)
- Weinstein PD, Anderson AO, Mage RG (1994) Rabbit IgH sequences in appendix germinal centers: VH diversification by gene conversion-like and hypermutation mechanisms. *Immunity* 1 (8):647-659.
- Wines BD, Willoughby N, Fraser JD, Hogarth PM (2006) A competitive mechanism for staphylococcal toxin SSL7 inhibiting the leukocyte IgA receptor, Fc alphaRI, is revealed by SSL7 binding at the C alpha2/C alpha3 interface of IgA. *J Biol Chem* 281:1389-1393 doi:10.1074/jbc.M509334200
- Zhu X, Boonthum A, Zhai SK, Knight KL (1999) B lymphocyte selection and age-related changes in VH gene usage in mutant Alicia rabbits. *J Immunol* 163 (6):3313-3320.

