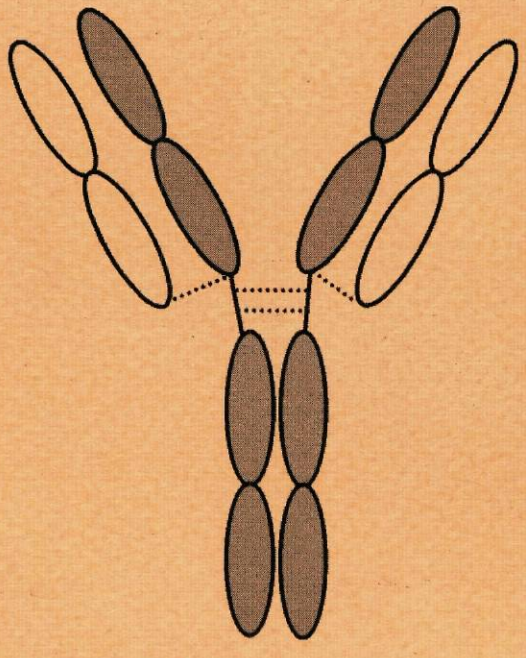
Pedro José de Castro Esteves

# Molecular and population genetic analysis of polymorphism at the antibody loci *IgGCH2* and *IgVH* in lagomorphs



Faculdade de Ciências
Universidade do Porto
2003

Pedro José de Castro Esteves

# Molecular and population genetic analysis of polymorphism at the antibody loci *IgGCH2* and *IgVH* in lagomorphs

Dissertação apresentada à Faculdade de Ciências da Universidade do Porto para obtenção do grau de Doutor em Biologia

Faculdade de Ciências
Universidade do Porto
2003

À minha filha Mariana e aos meus pais

# Declaração

Na elaboração desta dissertação e nos termos do nº 2 do artigo 8 do decreto-lei nº 388/70, foi feito o aproveitamento total dos resultados de trabalhos já publicados ou submetidos para publicação, que fazem parte integrante de alguns capítulos da presente dissertação. Em todos estes trabalhos, o candidato participou na obtenção, análise e discussão dos resultados, assim como na redacção dos manuscritos.

# Acknowledgements

Many thanks to my supervisor, Professor Wessel van der Loo for his considerable scientific input in this project. I have appreciated his dedication and friendship throughout my PhD. His many constructive comments helped me to grow as a researcher. I hope to continue to work with him to answer the many remaining questions raised by this thesis.

Thanks to Professor Katherine Knight for giving me the opportunity to work in her laboratory in the United States. Thanks also to all the people from her lab who made my stay very pleasant. Especial thanks to Dennis for many enjoyable conversations and friendship, and for all his help with the work. Thanks also to Mae and Rod for their excellent accommodation and friendship.

Thanks to the many other people who have helped during my thesis, including James Harris, Steve Weiss and Guillaume Queney.

# Agradecimentos

Para a concretização deste trabalho colaboraram muitas pessoas e instituições, às quais expresso todo o meu agradecimento.

Em primeiro lugar, quero agradecer ao meu co-orientador, Professor Doutor Nuno Ferrand de Almeida todo o empenho e entusiasmo que revelou no decorrer deste trabalho, a enorme confiança que sempre depositou em mim, bem como todas as discussões cientificas tão enriquecedoras e importantes para a conclusão deste trabalho. Agradeço-lhe ainda todas as facilidades que me proporcionou nos aspectos científicos e logísticos, bem como a forma correcta e franca com que sempre se relacionou comigo.

Ao Paulo Célio quero agradecer toda a ajuda que me deu na árdua tarefa de obter amostras, bem como todo o apoio que sempre demonstrou. Ao Professor Doutor Jorge Rocha agradeço os primeiros ensinamentos na área da genética, bem como a sua

contribuição no artigo de Porto Santo. À Madalena Branco agradeço a disponibilidade demonstrada em me ajudar na análise de dados.

À Fundação para a Ciência e Tecnologia agradeço a atribuição da bolsa de Doutoramento, sem a qual teria sido impossível a concretização desta dissertação.

À Catarina agradeço todo o carinho, entusiasmo, paciência e amizade demonstrada, bem como toda a imprescindível ajuda na redacção e discussão da tese.

À Rita agradeço o sacrifício, solidariedade, carinho e amizade demonstrada todos estes anos.

A todos os colegas de Vairão, agradeço toda a camaradagem e o bom ambiente de trabalho que sempre proporcionaram, bem como o saudável modo como discutiram os meus resultados. Em particular ao Rui Faria, ao Armando Geraldes, ao Pedro Cardia e ao Sequeira quero agradecer toda a amizade e companheirismo que sempre demonstraram, bem como todas as deliciosas horas de amena cavaqueira.

A todos os colegas do FCAV (clube da bola) agradeço todos os bons momentos e o facto de terem possibilitado que eu demonstrasse todas as minhas enormes (!!!) capacidades futebolísticas. À Teresa todos os cafezinhos, tão necessários para acabar a tese.

A todos os meus amigos agradeço o enorme interesse e ânimo que sempre demonstraram. Em especial a Sara, o Sequeira e o Rui "Spider" pelo enorme apoio que me deram em momentos muito difíceis da minha vida. Ao Niná agradeço o companheirismo, a amizade e o enorme interesse que sempre demonstrou pelo meu trabalho.

À minha família, os meus pais, as minhas irmãs, os meus sobrinhos e cunhado agradeço todo o incentivo e apoio que sempre demonstraram.

À Marianinha agradeço a paciência que teve comigo, sempre pronta a perdoar-me as minhas ausências. Os seus carinhos e beijos tornaram bem mais fácil a execução desta tese.

# Table of contents


Summary 1

Resumo 5

Résumé 9

Introduction 13

Chapter 1. Essentials 17

Basic Immunology 19

Structure and function of Antibodies 20
The antigen-binding sites 22
Fc region 23
Hinge region 24
Heavy-light domains interactions 24
Immunoglobulin classes 25

Genomic organization of Ig genes 26
Heavy Chain 26
Light Chain 29
Immunoglobulin class switch 30

Generation of antigen-binding repertoire 31
Post-rearrangement diversification 33

Rabbit taxonomy and genetics 35
Taxonomy of Lagomorpha group 35
The genetic diversity of European rabbit 36
Mitochondrial DNA (mtDNA) 37
Genetic variation of Proteins 37
Rabbit Ig allotypes 38
Microsatellite markers 39
Domestication 39

The rabbit in Immunological research 40

Citations 42

Chapter 2. Genetic variation at the constant region of the IgG antibody in leporids 53

Introduction 55
Rabbits possess only one IgG class 55
Indirect evidence for adaptive variation at the rabbit IgG CH2 domain (*e*-locus) 55

Article 1
P.J. Esteves, P.C. Alves, N. Ferrand and W. van der Loo (2002). Restriction fragment alleles of rabbit IGHG genes with reference to the rabbit *IGHGCH2* or *e* locus polymorphism. *Animal Genetics* **33**: 309-311. 59

Genetic diversity at the *IgGCH2* region in *Oryctolagus cuniculus* 65

The phylogeny of *IgGCH2* 69

Article 2
P.J. Esteves, P.C. Alves, N. Ferrand and W. van der Loo (2002). Hotspot variation at the CH2-CH3 interface of Leporid IgG antibodies (*Oryctolagus*, *Sylvilagus* and *Lepus*). *European Journal of immunogenetics* **29** (6): 529-36. 73

Variability at position 309 among mammalian species 83

Citations 86

Chapter 3. Studies of gene diversity at the Heavy chain variable region gene locus (*a* locus or $IgV_H$ locus) in leporids 91

Introduction 93
Population genetic studies 95
Indications that the $V_H1$ polymorphism can be trans-specific 96
The diversification of Ab primary repertoire in the rabbit 96
Evolution of $V_H$ gene family 98
Different hypotheses to explain the large inter-allelic differences among $V_H1$ genes 100

Material and Methods 101
Serological analyses 101
Analysis of genetic diversity 102
Sequencing of *VDJ* gene segments and germline $V_H$ gene segments 102
Phylogenetic analysis 103

Results and Discussion 105
Genetic diversity in Iberian rabbit populations 105
Analysis of genetic diversity 108
Cytonuclear disequilibria (*a* locus *vs* mtDNA) 108

Article 3
P.J. Esteves, D. Lanning, S.K. Zhai, N. Ferrand, K.L. Knight and W. van der Loo (submitted). Allelic variation at the $V_H a$ locus in natural populations of rabbit (*Oryctolagus cuniculus*, L.) 111

Genetic diversity in *Lepus* species 139

Article 4
P.J. Esteves, D. Lanning, S.K. Zhai, N. Ferrand, K. L. Knight and W. van der Loo (submitted). The evolution of immunoglobulin heavy chain variable region ($IgV_H$) in Leporids: a case of trans-species polymorphism 141

Understanding the large inter-allelic differences at the $V_H 1$ locus 157
Evolution of $V_H$ genes in the lagomorph group 157

Citations 161

CHAPTER 4. Evaluation of the bottleneck effect in the Porto Santo wild rabbit population 167

Introduction 169

Material and Methods 171
Assessment of genetic diversity 171
Statistical tests to detect bottleneck situations 173

Results and Discussion 175
Assessment of genetic diversity 175
Phylogeographic origin of Porto Santo rabbit population 177
Nuclear markers 177
Mitochondrial DNA 179

Article 5
P.J. Esteves, S. Weiss, J. Rocha, G. Queney, M. Branco, W. van der Loo and N. Ferrand (submitted). The "Rabbit-Eve" of Porto Santo: A 500 year-old mammalian population derived from a single female 183

Evaluating bottlenecks in wild rabbit populations 191
The likelihood that one single pregnant female originated the Porto Santo rabbit population 195
Implications for the trans-generic origin of the complex Ig allotypes 195

Citations 196

Chapter 5. Conclusions 201

Gene diversity at the *e* locus (*IgGCH2*) in *Oryctolagus cuniculus* 203

Gene diversity at the *IgVH* in leporids 204

Evaluating the bottleneck effect in the Porto Santo wild rabbit population 207

Appendix 1 209

Appendix 2 217

# Summary

Rabbits present a variety of allelic variants at different loci that encode the different constituents of the antibody molecule. The structure of the alleles are well described for domestic rabbit, but, except for the $C_{K1}$ L chain locus, have not been studied in wild populations. By sequence determinations of "novel" allelic variants, we have substantially contributed to a better knowledge of these immunological polymorphisms. In these studies we also included species of other lagomorph genera, namely *Lepus*, *Sylvilagus* and *Ochotona*. The comparison of serological and DNA sequence data allowed a better comprehension of the evolutionary processes that underlie the unusual divergence patterns of these polymorphisms.

As a first step we have studied a polymorphism that is both simple and recent. It concerns the genetic variation at the second domain of Ig Gamma H chain (*IgGCH2*) or *e* locus, a domain that mediates important Ab effector functions, and is a known target of pathogen attack. We sequenced the *IgGCH2* domain of 11 different lagomorph species: one of *Oryctolagus*, eight of *Lepus* (*europaeus*, *granatensis*, *timidus*, *americanus*, *capensis*, *calotis*, *californicus* and *castroviejoi*) and two of *Sylvilagus* (*floridanus* and *cunicularis*). The analyses of the obtained data revealed a "hotspot" of variability at the amino acid position H309, where six different triplets were observed, corresponding to five different amino acids (Met, Lys, Val, Ala and Thr). The extension of this analysis to other mammal groups (*Primata, Metatheria, Rodentia, Carnivora, Perissodactyla* and *Artiodactyla*) also revealed high variability levels at this position. We could present indirect evidence that this variation is due to diversity enhancement selection. Furthermore, based on the sequence data, we have developed a PCR/RFLP methodology that distinguishes the rabbit *IgGCH2* alleles *e14* and *e15,* which is a useful alternative to the currently used serological methods.

A major part of our work concerns the *IgVH* locus. In contrast to the *e* locus the $V_H$ polymorphism is complex and ancient. The VH region diversity has been thoroughly studied in the domestic rabbit. Three serologically defined allotypic lineages are described, the so-called $V_Ha$ allotypes *a1, a2* and *a3*. These allotypes are inherited as mendelian alleles because of the preferential usage of the $V_H1$ gene segment in the *VDJ* rearrangements. We have extended these studies to wild rabbit. The *IgVH1* locus was serologically analysed in

18 Iberian wild rabbit populations (402 specimens). Besides the previously described alleles (*a1*, *a2*, *a3* and *a3v*), we found two more alleles, *a1v* and *"a-blank"*, that are endemic to the Iberian Peninsula. One of these alleles (*"a-blank"*) was found to be present in all Iberian populations. This allotype was however more frequent in populations of the subspecies *O. c. algirus*. Sequencing of $V_H$ gene expressed in wild rabbit specimens typing *"a-blank"* allotype showed that these rabbits use preferentially $V_H$ genes that, although clearly related to the known $V_Ha$ genes, define a new major allotypic lineage, called here *a4*.

The large sequence difference between $V_H1$ alleles has suggested lineage coalescence times of 50 Million years, which is much longer than the time of separation between rabbits and hares. We have tested the trans-species nature of the polymorphism via sequence determination of $V_H$ gene segments that are expressed in specimens of *Lepus* (*L. europaeus* and *L. granatensis)*. The results revealed the existence in this genus of two distinct $V_H$ lineages, which we called *a2L* and *nL*. Phylogenetic analysis showed that the *nL* lineage was not closely related to this group nor to any other group described so far. In contrast, the lineage *a2L* clusters within the rabbit $V_Ha$ lineages, being closely related to rabbit allele *a2*, with which it forms a monophyletic group designated as "$V_Ha2$-like". These results support the trans-species polymorphism hypothesis, implying that very long lineage persistence times can at least in part explain the high inter-allelic divergence.

The data concerning the $V_H1$ locus may seem contradictory. On one hand, the association of specific alleles with subspecies markers suggests that the large differences between $V_Ha$ locus alleles may have accumulated after the separation of the subspecies. On the other hand, the confirmation of the trans-specific nature of $V_H1$ polymorphism supports the hypothesis of very long lineage persistence times. However, we have also shown that evolutionary rates can differ significantly among lineages, which allows the reconciliation of the apparent contradictions mentioned above. The different lines of evidences obtained in this study suggest therefore that allelic lineages have been maintained throughout successive speciation steps while evolutionary rates can differ among allelic lineages.

We further extended our knowledge of the $V_H$ gene segment to other lagomorph species (genera *Sylvilagus* and *Ochotona*) by sequencing germline $V_H$ gene segments. The comparison with other vertebrates suggests that in lagomorphs all $V_H$ genes belong to group C. Our data allow the definition of four clusters within lagomorphs ($V_Hn$, $V_HnL$, $V_Ha$ and $V_HP$). While some groups only contain $V_H$ genes of one species ($V_HnL$ and $V_HP$), some $V_H$ genes from different species cluster together. This suggests that lagomorph $V_H$ gene

segments evolved under the model of birth-and-death evolution (i.e. some $V_H$ gene segments diverged after duplication while other were maintained with high homology, explaining the grouping of $V_H$ genes from different species).

The hypothesis that at least four allotypic lineages have been maintained during 50 Million years of lagomorph speciation raises the question of the minimal size of founder populations. This question is addressed in the study of the genetic composition of the population of Porto Santo. The rabbit was introduced in this Atlantic island during the XV$^{th}$ century. Several historical documents suggest that a single pregnant female was released on this island, which was followed by a sudden increase in population numbers. By using a set of 34 markers (mtDNA, microsatellite, protein and immunogenetic loci) we showed that the Porto Santo rabbit population originated from Southwestern Portugal, and had, as expected, the signature of a recent bottleneck. A detailed analysis of the genetic data reported was in strong support of the hypothesis that one single pregnant female founded the contemporary Porto Santo rabbit population.

# Resumo

Os diferentes *loci* que controlam a biossíntese dos anticorpos apresentam no coelho uma grande variedade alélica. No que diz respeito ao coelho doméstico, a estrutura desses alelos encontra-se bem estudada. Porém, até à data esses marcadores não foram caracterizados em populações de coelho bravo, com excepção do *locus* $C_{K1}$ da cadeia leve. Através da determinação da sequência codificante de novos alótipos detectados na Península Ibérica, contribui-se substancialmente para um conhecimento mais alargado destes polimorfismos imunológicos. Nesta caracterização foram também incluídos outros lagomorfos, nomeadamente pertencendo aos géneros *Lepus*, *Sylvilagus* e *Ochotona*. A comparação de dados serológicos com a determinação das sequências nucleotídicas permitiu uma compreensão mais ampla dos processos evolutivos que são responsáveis pelos padrões de divergência pouco usuais encontrados nestes polimorfismos.

Numa primeira fase, estudou-se um polimorfismo que é simultaneamente simples e recente. Trata-se da variação genética do segundo domínio da região constante da cadeia pesada das imunoglobulinas da classe Gamma (*IgGCH2* ou *CHγ2*, também denominado como "locus *e*"). Este domínio desempenha um papel preponderante nas múltiplas funções dos anticorpos e é um dos alvos preferenciais do ataque de agentes patogénicos. Sequenciou-se o domínio *IgGCH2* em onze espécies diferentes: oito do género *Lepus* (*europaeus*, *granatensis*, *timidus*, *americanus*, *capensis*, *calotis*, *californicus* e *castroviejoi*) e duas do género *Sylvilagus* (*floridanus* e *cunicularis*), para além do coelho (género *Oryctolagus*). A análise dos dados obtidos revelou a existência de um "hotspot" de variabilidade na posição aminoacídica H309, onde foram observados cinco resíduos diferentes (Met, Lys, Val, Ala e Thr), ao passo que nas restantes posições a variabilidade se revelou praticamente inexistente. A extensão desta análise a outros grupos de mamíferos (*Primata, Metatheria, Rodentia, Carnivora, Perissodactyla* e *Artiodactyla*) demonstrou que este fenómeno não está limitado apenas aos lagomorfos. Foi possível mostrar de forma indirecta que esta variabilidade é promovida por selecção, que actua favorecendo a diversidade nesta posição em particular. Os dados obtidos relativamente à variação deste marcador na Península Ibérica confirmaram a tese de que o polimorfismo *e14/e15* do coelho doméstico, ausente das populações Ibéricas, se fica a dever a uma mutação recente de um gene ancestral *e15*, provavelmente de origem Ibérica. Com base na determinação das

sequências de DNA subjacentes a esta mutação, foi ainda possível desenvolver uma metodologia de PCR/RFLP que permite distinguir os dois alelos referidos, constituindo assim uma alternativa útil aos métodos serológicos normalmente usados.

Uma parte importante deste trabalho foi consagrada ao estudo da variação alélica do locus *IgVH*, que codifica a porção variável da cadeia pesada dos anticorpos. Contrariamente ao marcador anteriormente abordado (*IgGCH2*), no qual o polimorfismo se limita à variação de um só aminoácido, a variabilidade deste locus é complexa e filogeneticamente muito antiga. Este polimorfismo foi extensivamente estudado no coelho doméstico, onde foram descritas três linhagens alotípicas (*a1, a2* e *a3*). Estes alótipos comportam-se como alelos mendelianos devido ao facto de os coelhos utilizarem preferencialmente apenas um gene $V_H$ durante o rearranjo *VDJ*, nomeadamente o gene $V_H1$. Com este trabalho, ampliou-se o estudo deste polimorfismo a populações Ibéricas de coelho bravo. Nestas populações foram detectados, para além dos já descritos (*a1, a2, a3* e *a3v*), dois novos alótipos (*a1v* e "*a-blank*"), um dos quais ("*a-blank*") está presente em todas as populações Ibéricas e parece estar associado à subespécie *O. c. algirus*. A determinação da sequência do RNA mensageiro demostrou que os indivíduos com o alótipo "*a-blank*" utilizam quase exclusivamente um gene $V_H$ que, embora filogeneticamente relacionado com os genes $V_Ha$ do coelho doméstico, define uma nova linhagem alotípica que se designou por *a4*.

Baseando-se nas elevadas distâncias que separam os diferentes alelos $V_H1$, alguns autores sugeriram um tempo de coalescência das linhagens da ordem dos 50 milhões de anos, o que é muito superior ao tempo de separação entre *Lepus* e *Oryctolagus*. Para testar a hipótese da natureza trans-específica deste polimorfismo, determinou-se a sequência nucleotídica dos segmentos $V_H$ expressos em espécies do género *Lepus* (*L. europaeus* e *L. granatensis*). Nestas espécies foram detectadas duas linhagens, designadas por *a2L* e *nL*. As análises filogenéticas demonstram que a linhagem *nL* não se agrupa dentro do grupo $V_Ha$ nem de qualquer outro grupo de genes $V_H$ descrito até à data. Pelo contrário, a linhagem *a2L* é aparentada com a linhagem *a2* do coelho e as distâncias entre o alelo $V_H$-*a2* e os seus homólogos $V_H$-*a1,-a3* e *-a4* são notoriamente mais elevadas do que as distâncias detectadas entre as linhagens *a2* e *a2L*. Estes resultados apoiam a hipótese do polimorfismo dos genes $V_Ha$ ser trans-específico, uma vez que implicam que as várias linhagens tenham iniciado a sua divergência antes da separação das duas espécies.

Os resultados relativos ao locus $V_H1$ aparentam ser contraditórios. A associação de diferentes alelos com subespécies distintas não se coaduna, aparentemente, com tempos de coalescência das linhagens que são claramente anteriores à data da especiação. No entanto, foi possível demonstrar que as taxas de evolução não são as mesmas para todas as linhagens, o que poderá em parte ajudar a conciliar as duas observações.

Adicionalmente, alargou-se a análise através da sequenciação de genes $V_H$ da linha germinal de espécies pertencentes aos géneros *Sylvilagus* e *Ochotona*. A comparação das sequências obtidas com as de outros vertebrados sugere que todos os genes $V_H$ dos lagomorfos pertencem ao grupo C, onde se repartem por quatro subgrupos: $V_Hn$, $V_HnL$, $V_Ha$ e $V_HP$. Apesar de os subgrupos $V_HnL$ e $V_HP$ conterem apenas sequências pertencentes a um só género (*Lepus* e *Ochotona*, respectivamente), os outros agrupamentos reúnem genes de espécies e géneros diferentes. Este cenário sugere que os genes $V_H$ dos lagomorfos seguem um modelo de evolução das famílias multigénicas denominado "birth-and-death" (ou seja, certos genes sofreram duplicação e divergiram desde então, enquanto outros mantêm um elevado grau de homologia).

A hipótese da manutenção de um número de linhagens alotípicas igual ou superior a quatro por um período de pelo menos 50 milhões de anos levanta a questão do tamanho mínimo das populações fundadoras. Esta questão é abordada no estudo da composição genética da população de Porto Santo. O coelho foi introduzido nesta ilha atlântica no século XV. Diversos documentos históricos sugerem que essa introdução consistiu na libertação de uma só fêmea prenha, à qual se sucedeu um rápido aumento do efectivo populacional. O estudo de uma bateria de 34 marcadores moleculares (incluindo loci proteicos, microssatélites, marcadores imunogenéticos e de DNA-mitocondrial) revelou que a população de Porto Santo se originou a partir do Sudoeste de Portugal e resulta de um recente efeito de fundador. Uma análise mais detalhada dos dados genéticos demonstra que as observações estão de acordo com a tese que propõe que esta população tenha sido fundada por uma só fêmea prenha.

# Résumé

Les différents loci contrôlant la biosynthèse de l'anticorps, sont chez le lapin, munis d'un large éventail d'allèles. La structure de ces allèles est bien connue, en ce qui concerne le lapin domestique. Par contre ces allèles n'ont pas été étudiés dans des populations sauvages à l'exception près du locus $C_{K1}$ de la chaîne L. Nous avons contribué de manière substantielle à une meilleure connaissance de ces polymorphismes d'anticorps, en déterminant les séquences codantes d'un nombre d'allotypes nouvellement découverts dans des populations ibériques. Nous avons étendu ces études à d'autres espèces de lagomorphes, notamment des espèces appartenant aux genres *Lepus*, *Sylvilagus* et *Ochotona*.

Dans un premier temps, nous avons étudié un polymorphisme qui peut être qualifié de simple et récent. Il s'agit de la variation génétique du deuxième domaine de la partie constante de la chaîne H de l'anticorps de la classe Gamma (*IgGCH2*, ou $CH_\gamma 2$, aussi connue sous le nom de "locus *e*"). Ce domaine joue un rôle prépondérant dans les multiples fonctions d'anticorps, et est une des cibles préférées d'agents pathogènes. Nous avons séquencé le domaine *IgGCH2* dans onze espèces différentes: huit du genre *Lepus* (*europaeus*, *granatensis*, *timidus*, *americanus*, *capensis*, *calotis*, *californicus* et *castroviejoi*), deux du genre *Sylvilagus* (*floridanus* et *cunicularis*), en plus du lapin de garenne (genre *Oryctolagus*). Ceci a permit de constater une variation "hotspot" à la position H309, où cinq différents résidus ont été observés (Met, Lys, Val, Ala et Thr), tandis que la variation était quasiment inexistante dans le restant du domaine. En incluant d'autres ordres de mammifères (*Primata, Metatheria, Rodentia, Carnivora, Perissodactyla* et *Artiodactyla*) l'étude montrait que ce phénomène n'est pas limité aux seuls lagomorphes. On a pu indirectement établir que cette variabilité est promue par une force déterministe qui favorise la diversité ou le changement à cette position particulière. Nos données sur la variation à l'intérieur de la Péninsule Ibérique ont confirmé la thèse que le polymorphisme *e14/e15* du lapin domestique, absent de la Péninsule, est dû à une mutation récente d'un gène ancestral e15, probablement d'origine ibérique. La connaissance des séquences d'ADN sous-jacentes à cette mutation, a rendu possible la mise au point d'une méthode de PCR/RFLP, permettant la détection des allèles *e14* et *e15* du lapin, sans le recours d' antisérum spécifique.

Une partie plus importante de nos travaux est consacrée à l'étude de la variation allélique au locus *IgVH*, qui gouverne la partie variable de l'anticorps. Notons d'emblée, que cette variabilité concerne les régions charpentes (FR), et non pas les régions qui sont impliquées dans la reconnaissance de l'antigène (CDR). Contrairement au polymorphisme précédant ($CH_\gamma 2$), qui se limite à la variation à un seul résidu, ce polymorphisme est complexe (à savoir les allèles diffèrent par des substitutions d'acide aminé multiples), et phylogénétiquement très ancien (plus ancien que les espèces). Ce polymorphisme a été étudié en détail dans les lapins domestiques, où on distingue trois lignées allotypiques (*a1*, *a2*, *a3*). Leur transmission mendélienne est due au fait que le lapin, lors du réarrangement *VDJ*, n'emploie en substance qu'un seul gène $V_H$, c'est à dire le gène $V_H1$. Nous avons élargit ces connaissances aux populations ibériques, où nous avons mis en évidence deux allotypes nouveaux, dont un semble être associé à la sous-espèce *O. c. algirus*. Un de ces allèles se distingue par l'absence de réactions croisées ("*a-blank*"). Le séquençage de l'ARN messager a permis de conclure que les lapins "*a-blank*" utilisent quasiment exclusivement un gène $V_H$, qui s'apparente phylogénétiquement aux $V_H1$ du lapin domestique, et constitue une nouvelle lignée allotypique, appelée "*a4*".

Les grandes distances séparant les différents allèles $V_H1$ ont été interprétées comme une indication du fait que leur coalescence pourrait se situe dans un temps très éloigné (50 million d'années, d'après M. Nei). Nous avons évalué cette hypothèse par l'étude des gènes $V_H$ exprimés dans des lièvres (*Lepus europaeus* et *Lepus granatensis*). On a su établir dans ces espèces deux lignées de $V_H$, nommé *a2L* et *nL*. Il apparaît de façon claire, que la lignée *a2L* du lièvre s'apparente à la lignée *a2* du lapin. Les distances entre les allèles $V_H$-*a2* de ses homologues $V_H1$-*a1*, -*a3* et -*a4* sont nettement plus importantes que les distances entre les lignées *a2* et *a2L*. Ceci démontre que les gènes $V_H$ qui, dans le lapin d'aujourd'hui se comportent comme des allèles mendéliens, ont débuté leur divergence dans une période qui est antérieure à la séparation des deux espèces.

Ces données concernant le locus $V_H1$ semblent contradictoires. L'association de différents allèles avec des sous-espèces distinctes ne s'accorde pas aisément avec des temps de coalescence des lignées qui sont nettement plus longs que les temps de spéciation. Nous avons néanmoins pu établir que les vitesses d'évolution ne sont pas les mêmes dans les différentes lignées, ce qui pourrait aider à réconcilier les deux observations.

Nous avons étendu ces études aux gènes $V_H$ qui sont enfuis dans les génomes (sans être exprimés) des espèces appartenant aux genres *Lepus*, *Sylvilagus*, et *Ochotona*. Tous

s'apparentent au group C, où ils se répartissent en quatre sous-groupes ($V_Hn$, $V_HnL$, $V_Ha$ et $V_HP$). Tandis que toutes les séquences des groupements $V_HnL$ et $V_HP$ n'appartiennent qu'à une seule espèce, dans d'autres groupements, différentes espèces sont réunies. Ceci est conforme à un modèle d'évolution des familles multigéniques dénommé "birth and death" (c'est à dire certains gènes se répandent par duplication et divergent et se perdent en partie, pendant que d'autres maintiennent un degré plus important d'homologie).

L'hypothèse qu'un nombre de lignées allotypiques, supérieur ou égal à quatre, a pu se maintenir durant 50 millions d'années, pose la question de la taille minime des populations fondatrices. Cette question clé est traitée dans l'étude détaillée de la composition génétique de la population de l'île de Porto Santo. Le lapin a été introduit dans cette île de l'Atlantique durant le XV ième siècle de notre ère. Les documents historiques indiquent qu'une seule femelle en gestation a été lâchée, évènement suivi peu de temps après par une démographie galopante. Notre étude de 34 marqueurs génétiques confirme l'existence d'un fort effet fondateur. Une analyse plus en profondeur démontre que les observations sont en parfait accord avec la thèse d'une seule femelle fécondée fondatrice.

# Introduction

This work is situated at the meeting point of two ongoing research projects focusing on the extraordinary degree of genetic variation in European rabbit (*Oryctolagus cuniculus*). These are the phylogeographical studies of this variation and the evolutionary history of the complex immunoglobulin polymorphisms.

There are two subspecies of the European rabbit (*Oryctolagus cuniculus*): *O. c. algirus* and *O. c. cuniculus*. Both subspecies are present in the Iberian Peninsula where they are separated by a secondary contact zone along a Northeast-Southwest cline. The genetic diversity observed at nuclear and mitochondrial markers revealed that the level of population diversity was much higher in Iberia, than in the more recent area of distribution, Continental Europe (North of Pyrenean Mountains), Great Britain and overseas. The domestication of rabbit is recent (Middle Ages) and all domestic breeds are apparently derived from *O. c. cuniculus* populations. In this context, the wide variety of genetic markers makes the rabbit an ideal model species for theoretical genetic investigations regarding speciation phenomenon, hybridisation and introgression, etc., and for evaluating methods of evolutionary inference based on diversity patterns of autosomal, maternal (mtDNA) and paternal (Y chromosome) markers. The introduction of rabbits to numerous islands also provides favourable means for studies of the consequences of population bottlenecks.

Most of the polymorphisms mentioned above are in essence selectively neutral. Historically, the first polymorphisms studied in rabbits were those of the antibody genes. Their highly unusual patterns of gene diversity at both the molecular and the population genetic level revealed that these polymorphisms are adaptive, and maintained by strong selective forces. These polymorphisms have been mainly studied in laboratory breeds and in populations of feral rabbits from the recent distribution range. While we have a good understanding of the mechanisms of antibody gene organisation, ontogeny and expression of the antibody, we know little about aspects of population diversity of antibody loci. The evolutionary origin of rabbit variable region allotypes remains enigmatic, and has been qualified as "an evolutionary puzzle".

We expected that new insights concerning their biological importance could be obtained by studies of the wild populations from the native area of distribution. We have furthermore approached the evolution of these Immunoglobulin (Ig) polymorphisms by cloning and sequencing of Antibody (Ab) genes expressed in the subspecies *O. c. algirus*, and of *Lepus* species. We have also contributed to an evaluation of the hypothesis of multi-allele trans-species polymorphism, by determining the minimal size of founder populations by studying the rabbits of the Atlantic Island of Porto Santo.

We have subdivided our work into five chapters:

1. Essentials
2. Genetic variation at the constant region of the IgG antibody in leporids
3. Studies of gene diversity at the Heavy chain variable region gene locus (*a* locus or *IgVH* locus) in leporids
4. Evaluation of the bottleneck effect in the Porto Santo wild rabbit population.
5. Conclusions

The abstracts of these chapters are presented below:

## CHAPTER 1

## Essentials

This chapter reviews in short a number of topics that may help understanding of the following chapters. They concern the immune system, antibody structure, function and genomic organisation. The processes by which diversity at the antibody level is generated, are explained in some detail. We give a brief account on the Lagomorph's evolutionary history and systematics and survey the known genetic diversity of the rabbit (for protein loci, mitochondrial DNA, microsatellites and immunological markers). The role of the rabbit in immunological research is presented in a historical perspective.

## CHAPTER 2

## Genetic variation at the constant region of the IgG antibody in leporids

We developed a PCR/RFLP approach for studies of the single amino acid polymorphism (SAP) of the constant region of the rabbit Ig Gamma chain (*IgGCH2* or *e*-locus). It offers a convenient alternative to the arcane serological methods in use. Sequences of the *CH2* exon, which harbours the SAP, were determined for different genetic variants for eleven Leporid species (genera *Oryctolagus*, *Lepus* and *Sylvilagus*). The variability at amino acid position 309 (79 in IMGT numbering) was studied in different leporid species and compared to that in some groups of mammals. These studies have revealed the high variability at one particular amino acid position (i.e. H309).

## CHAPTER 3

## Studies of gene diversity at the Heavy chain variable region gene locus (*a* locus or *IgVH* locus) in leporids

The preferential usage of the rabbit $V_H1$ gene, together with its extensive allelic variation, suggests that different allotypic lineages may generate different antibody repertoires. A key question is whether the distribution of the different allelic forms in rabbit can be influenced by environmental factors (such as microbial flora or disease agents) or only depends upon evolutionary history. A major objective of this work was to initiate the phylogeographic approach of the allele diversity of the rabbit *IgVH* gene clusters in the subspecies *O. c. algirus*.

A second question concerns the length of persistence time of $V_Ha$ allotypic lineages, and their number. We have therefore determined DNA sequences, both for expressed (cDNA) and germline $V_H$ genes. These were obtained from specimens of *O. c. algirus* and of other Lagomorph species (*Lepus*, *Sylvilagus* and *Ochotona*). This allowed the reconstruction of likely phylogenetic relationships. Our studies added new facts that contribute to a better understanding of the evolution of the $V_H$ gene cluster in rabbits. Our results show that the divergence between at least some allelic $V_Ha$ lineages of rabbit was anterior to the *Lepus-Oryctolagus* split; i.e. that the polymorphism is older than the species.

The hypothesis of a multi allele trans-species polymorphism raises the question of the minimal population size of founder populations, which is addressed in Chapter 4.

# CHAPTER 4

## Evaluation of the bottleneck effect in the Porto Santo wild rabbit population

In this study we show that for rabbit, the size of founder populations can be extremely limited. Indeed, our study of the Porto Santo population demonstrate that successful populations can be founded by a single pregnant female, making it appear as highly unlikely that four or more $V_Ha$ lineages have been maintained over millions of years. However, the importance and implications of this study reaches far beyond this particular issue.

One of the most important aspects of the natural history of the rabbit is anthropogenic introduction. This species was introduced to the Atlantic Island of Porto Santo in the XV$^{th}$ century and is described in several historical documents. The rabbit of Porto Santo is referred to by eminent scientists (e.g. Darwin and Hæckel) as an example of morphological divergence since its introduction in the island 600 years ago. In this work three immunoglobulin loci ($V_H1$ or *a*, $C_{K1}$ or *b and CH2* or *e*) and 31 additional "neutral" markers (mtDNA, protein loci, microsatellites) were studied, which contributed to: 1) determine the phylogeographic origin of this population; 2) understand the effects of a severe bottleneck on the population genetic structure; 3) confirm the hypothesis that the Porto Santo population was founded by one single pregnant female.

# CHAPTER 1

## Essentials

# Basic Immunology

Vertebrates have a variety of mechanisms that provide protection against infectious pathogens and other agents (viruses, bacteria, protozoa, infectious and infected cells, foreign cells, etc.). Among these defence mechanisms the immune system is probably the most efficient and certainly the most elaborated. It can be innate or adaptive. The adaptive immune response is highly specific for a particular pathogen. Whereas the innate response does not change on repeated exposure to a given infectious agent, the adaptive response improves ("adapts") at each successive encounter with the same pathogen.

The first step of immune response is the recognition of foreign antigens. The adaptive immune system uses two distinct types of molecules in this process, the T cell antigen receptor (TCRs), which are expressed by mature T lymphocytes, and the immunoglobulins (Ig) or antibodies (Ab), which are produced by B lymphocytes. The immune system generates cells that can recognize an enormous range of antigens but each lymphocyte (B or T cell) is genetically capable of recognizing only one particular antigen. After an antigen binds to the few cells that can recognize it, these cells are stimulated to proliferate and mature. Some of them, called memory cells, are maintained in the plasma and constitute the immunological memory of particular antigens.

The second step of immune response is to destroy the pathogens by what is often called the effector system. There are numerous ways to eliminate the pathogens, such as: a) neutralization (antibodies can combat certain pathogens just by binding to them), b) phagocytosis (antibody is important in activating complement, or acting as an opsonin to promote ingestion by phagocytes that will digest the pathogen), c) cytotoxic reaction (where the target cell is recognized, either by a specific antibody bound to the cell surface or by specific TCRs). The cytotoxic cells can destroy the target cell or can induce the target cell to self-destruction (apoptosis).

The immune system is vital for survival. A deficiency in any part of these systems leaves the individual exposed to a greater risk of infection. These deficiencies can be hereditary or acquired (e.g. acquired immunodeficiency syndrome or AIDS). Particular ways by which the system can fail are autoimmunity, where the immune system lost the capacity to recognize the body's own tissues as "self" and reacts against them (e.g. rheumatoid arthritis and pernicious anaemia) and hypersensitivity, where the immune reactions may cause more damage than the pathogen or antigen itself (e.g. molecules on the surface of pollen grains).

The cells involved in the immune response are organized into tissues and organs referred to as the lymphoid system. The major lymphoid organs and tissues are classified as primary (central) or secondary (peripheral). The primary tissues in the mammals are the thymus, for T cells, and bone marrow and foetal liver, for B cells. In these organs the lymphocytes differentiate from the lymphoid stem cells, proliferate and mature into functional cells. Then the lymphocytes migrate into peripheral or secondary lymphoid organs that comprise the spleen, lymph nodes and mucosa-associated lymphoid tissues (MALT). They provide an environment in which lymphocytes can interact with antigens. Immune responses generated in these secondary lymphoid tissues require phagocytic macrophages, antigen-presenting cells, and mature T and B lymphocytes.

## Structure and function of Antibodies

The antibodies are glycoproteins with typical tetrameric Y-shaped form composed of two distinct types of polypeptide chain. They consist of two identical heavy (H) chains of 50-77 Kilo Daltons (KD) that are structurally distinct for each Ig class or subclass, and two identical light (L) chain of 25 KD that are common to all classes. Two distinct domains are

distinguished, each possessing characteristic functions: a variable (V) region located at the amino-terminal part, which exerts recognition functions and a constant (C) region at the carboxyl-terminal part assuring an effector function. A single disulfide bond connects the L and H chains. The connection between the two H chains is made by one or more disulfide bounds. The Ig molecule is bivalent with two identical sites for potential binding to antigen formed by the amino terminal regions of the H-L chain pair. The schematic representation of one Ig molecule is presented in Figure 1.1.

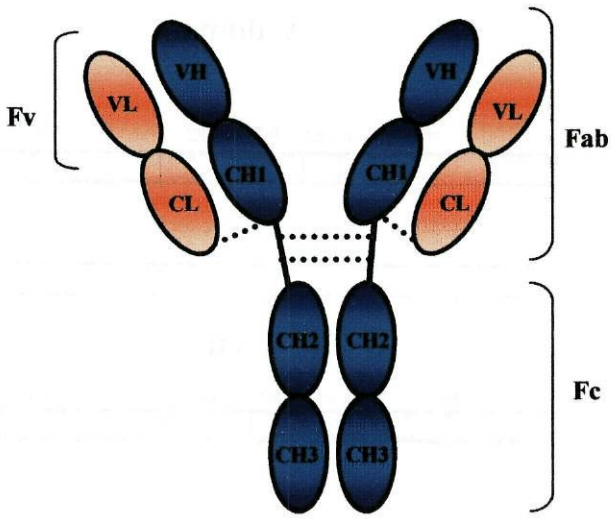



**Figure 1.1.** Schematic representation of a human IgG molecule. The Fv part is composed of the H and L chain V domains (VH and VL) and make up the antigen-binding sites. The Fv and the H chain CH1 and CL domains constitute the Fab part. The C-terminal constant domains of the H chain (CH2-CH3) form the Fc region of the molecule. The hinge region joins Fc and Fab. The inter-chain disulfide bonds are represented by dotted lines.

Each Ig chain is composed of domains with about 110 amino acid residues each. The H chain contains four to five domains (VH, CH1-CH4), while the L chain comprises only two (VL, CL). The amino acid sequence of the amino-terminal segments varies among antibodies, while the carboxyl-terminal domains are much more conserved. The unit composed of the paired H and L chain variable regions (VH and VL), is named Fv (variable Fragment), whereas the fragment containing the whole L chain, the variable (VH) and the first constant (CH1) domains of the H chain is called Fab (antigen-binding Fragment). Finally, the C domains (CH2-CH3, or CH2-CH3-CH4) form the Fc (crystallisable Fragment) part of the Ig, which interacts with effector molecules and cells.

## The antigen-binding sites

Of the domains composing the Fab region (VL, VH, CL and CH1), only the V domains (VL, VH) are involved in antigen binding. The amino acid sequence comparisons within both VH and VL domains revealed in each domain the existence of three hypervariable regions. These are called complementarity-determining regions (CDRs) because they interact with the antigen. The regions separating the CDRs are relatively invariable and are referred to as framework regions (FRs). (see Figure 1.2).

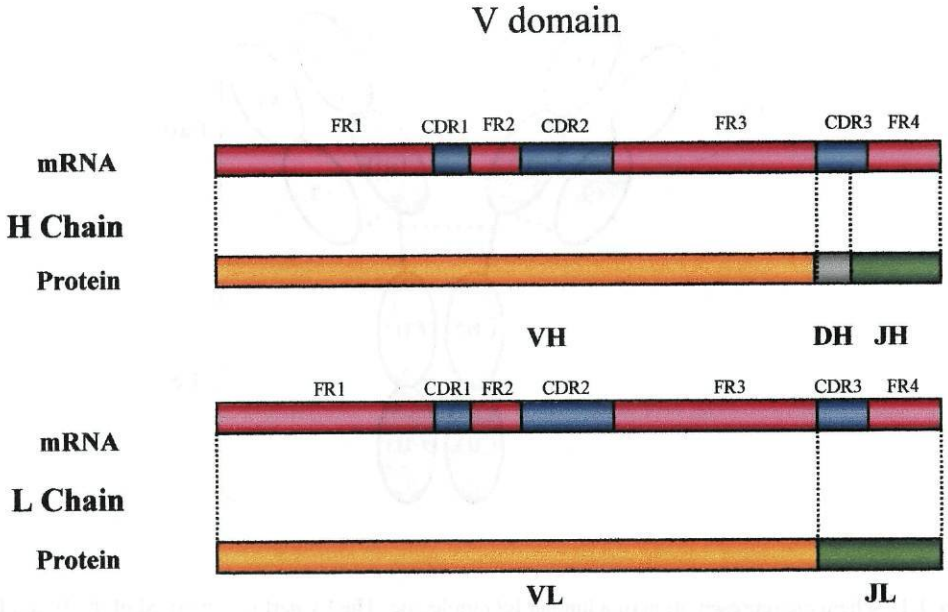



**Figure 1.2.** Schematic representation of the Heavy (H) and Light (L) chain. FR – Framework; CDR – Complementary Determining Region; V – Variable region; D – Diversity region; J – Joining region.

The contribution to antigen binding of the different CDRs in VH and VL domains varies from one antigen-antibody complex to another. Any residues in any of the six hypervariable regions can be actively involved in the molecular interaction with the antigen (see Padlan, 1994). For this reason, the definition of CDR and each FR region are subject to debate. There are different proposals. Most in use are the numbering suggested by Kabat *et al.* (1991) and the unique numbering of IMGT (the international ImMunoGeneTics database http://imgt.cines.fr, Lefranc, 2001). These are presented in Figure 1.3. We adopt the former convention, which is most in use among rabbit immunologists.

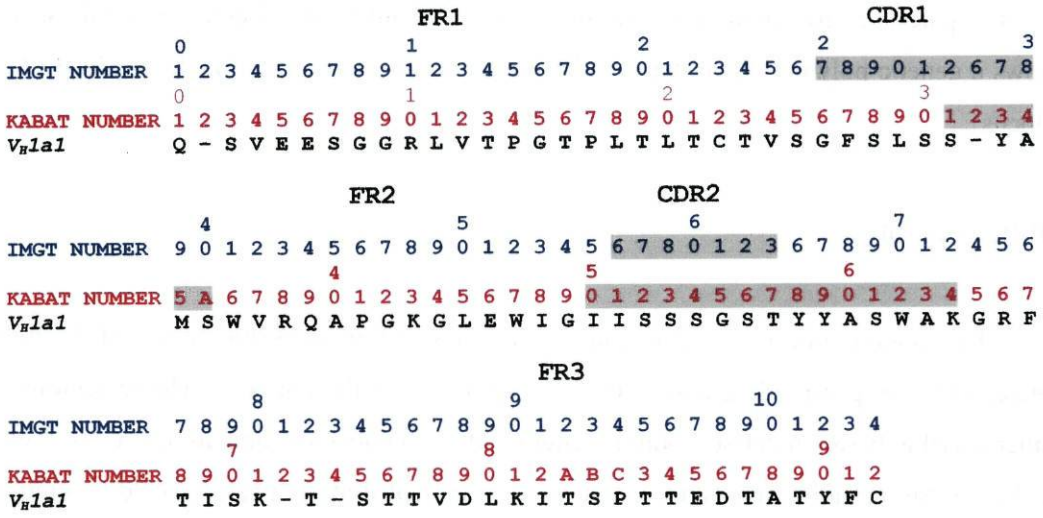



**Figure 1.3.** Amino acid boundaries delineating the sub-domains of the H chain of rabbit *VH1a1* according to Kabat *et al.* (1991) and the unique numbering of IMGT (Lefranc, 2001). The CDRs are grey-shaded.

The overall shapes of the antigen-binding sites created by the juxtaposition of the CDR of the H and L chains vary greatly from clefts to grooves to flat surfaces. In general, small antigens such as haptens or oligo-peptides bind to clefts or pockets lying between the H and L chain V domains, whereas larger protein antigens bind on extended planar surfaces formed by binding sites of antibodies (Webster *et al.*, 1994; Padlan, 1996).

## Fc region

The Fc fragment exposes a variety of epitopes that mediates effector functions, such as complement fixation. The docking of antibody onto Fc receptors of macrophages can either initiate a cascade of reactions leading to the lyses of the foreign cells, or trigger phagocytosis of the antibody-antigen complexes. Each of the five classes of Igs described in the mammals, IgA, IgG, IgD, IgM and IgE are characterized by distinct effector functions. However, all of these classes share the Fab region.

The Fc region of IgA, IgG and IgD consists of CH2-CH3 dimers, while in the case of IgM and IgE is composed of CH2-CH3-CH4. The binding sites of IgG for Fc receptors are localized in the CH2 region (e.g. Tao *et al.*, 1993; Idusogie *et al.*, 2000). In response to the destructive potential of the Fc region, many kinds of microorganisms have developed proteins that are able to bind to, or proteolytically cleave, the Fc part, thus preventing it from functioning. The blocking of IgG functions by staphylococcal protein A and streptococcal protein G are good examples of these strategies (Deisenhofer, 1981). The

rabbit is particular by having no IgG subclasses. The unique CH2 domain of this species shows a genetic polymorphism that will be studied more in detail in this thesis (*IgGCH2* or *e*-locus).

## Hinge region

The hinge region joins the Fc and Fab regions and allows some degree of flexibility between the antigen-binding and effector components of the antibody. Hinge regions are rather variable between classes, both in length and in composition, and are less conserved in evolution. Structurally, hinges are extended segments of dimeric peptides held together by one or more disulfide bridges. It has been shown that the removal of inter-chain disulfide bridges of the hinge abrogates the antibody's effector functions (Coloma *et al.*, 1997). The hinge is the part of an Ig which is most exposed to proteolytic attack. As an example, the IgD immunoglobulins, with their long charged hinge are very susceptible to proteolysis, which may explain their short serum half-life (Spiegelberg, 1977). In the rabbit, 13 IgA isotypes show a high degree of amino acid divergence in the hinge region. The presence of 13 IgA isotypes, each with different patterns of resistance to the various IgA proteases, could provide enhanced protection to the rabbit when exposed to multiple bacterial IgA proteases (Burnett *et al.*, 1989). The IgG hinge region of the rabbit shows two serological allotypes, *d11* and *d12*, which are associated by a Met/Thr interchange at position 209 (Prahl *et al.*, 1969).

## Heavy-Light domains interactions

The H and L chain components of V regions interact by hydrogen bonds (VH-VL), whereas a disulfide bridge covalently links CH1 and CL. The VH-VL association is exclusively mediated by non-covalent interactions. Residues from CDR as well as FR regions participate in this association and the affinity between the two domains is high.

Pairing of monomeric H and L chains obtained from different individuals, or even from different species, was observed under physiological conditions (i.e. between rabbit L chains and human H chains; Kulics *et al.* 1983, reviewed in Frazer and Capra, 1999).

Most of the residues that are important in VH-VL contacts are hydrophobic and well conserved throughout vertebrates (see Hsu and Steiner, 1992). The contribution of the different FRs and CDRs to the VH-VL interaction can vary considerably. The H-CDR3 and

L-CDR3 interaction accounts for more than 60% of the total of the CDR interactions. For the FR-FR interactions more than half of the hydrophobic core contacts involve FR2 of one chain and FR4 of the other (Padlan, 1994). The remaining interactions involve essentially FR2 and FR4 of the L chain with the CDR3 of the H chain while only FR2 of the H chain interacts with the CDR3 of the L chain (Chothia *et al.*, 1985).

## Immunoglobulin classes

Five distinct classes of Ig molecule are recognized in higher mammals, namely IgA, IgD, IgE, IgG and IgM. The H chains that define these five classes are designated by $\alpha$, $\delta$, $\varepsilon$, $\gamma$ and $\mu$, respectively. The sequence difference between these H chains identifies the various isotypes and determines their functional activity. The number and location of inter-chain disulfide bonds, the number of attached oligosaccharide moieties, the number of C domains, and the length and sequence of the hinge region are characteristics of each isotype. There is also heterogeneity within classes. For example, in human and mice there are four subtypes of IgG antibodies that have different physical properties.

IgE and IgG were not identified in reptiles, amphibians or birds and are probably unique to mammals. In fact, in birds and reptiles, antibodies of the IgY isotype exhibit the functional characteristics of both IgE and IgG, whereas the amphibian *Xenopus laevis* expresses the IgX isotype that shows similar tissue distribution as mammalian IgA (Mussmann *et al.*, 1996). Lower vertebrates, such as fish, amphibians and reptiles possess fewer Ig isotypes than mammals. Some produce monomeric and polymeric IgM, truncated and full-length versions of the immunoglobulin polypeptide (Hsu, 1994; Warr, 1995). In addition, new isotypes have been identified in sharks. Indeed, the NAR (New Antigen Receptor), are homodimers consisting of two identical H chains, containing a variable and five constant CH1 type domains. Such homodimeric Igs seem to be restricted to chondrichthyes (Greenberg *et al.*, 1995), with the notable exception of the H chain antibodies of *Camelidae* (Nguyen *et al.*, 2000). The evolutionary trends of diversifying antibody classes suggest that different isotypes serve specific biological functions. B cells can switch from secreting IgM to other isotypes like IgD, IgE or IgG (Ig class switch, see below).

# Genomic organization of Ig genes

An important feature of the immunological system is the fact that the antibody-encoding genes do not exist as such in the germline. The combinatorial rearrangement of the gene segments and the diversity obtained in the segment joining process, together with the diversification process of gene conversion and somatic hypermutation, contribute largely to the impressive capacity of the immune system to generate diversity (see Figure 1.4).

Three clusters of genes, localized on separate chromosomes, code for the different Ig molecule chains (Tonegawa, 1983). These are the gene families coding for H, K-L and λ-L chains, respectively. Each cluster can be composed by gene families encoding constituents of the V and C domains. The germline organization of L and H chains gene segments of higher vertebrates are schematically represented in Figure 1.5.

## Heavy Chain

### *Variable region*

Three types of germline elements encode the Ig H chain V region domain: $V_H$ (Variable), $D_H$ (diversity) and $J_H$ (Joining) segments. For each chain there are several tandem repeats of *V* segments, which are composed of two exons (Leader and *V*-exon) separated by an intron of about 100-bp.

The human H chain gene complex has been localized on the chromosome 14 and consists of 123 $V_H$ segments, classified into seven families, 27 functional $D_H$ genes and nine $J_H$ segments (six are functional) (see Matsuda *et al.*, 1998). The mouse Ig H-chain locus contains about 100-200 $V_H$ genes classified into 15 families, a cluster of four $J_H$ genes and 13 $D_H$ gene segments (see Almagro *et al.*, 1997).

The FR1, CDR1, FR2, CDR2 and FR3 are encoded by the $V_H$ gene segment, while the CDR3 is generated from the *D* and *J* exons. $J_H$ exon encodes part of the CDR3 and the entire FR4 region.

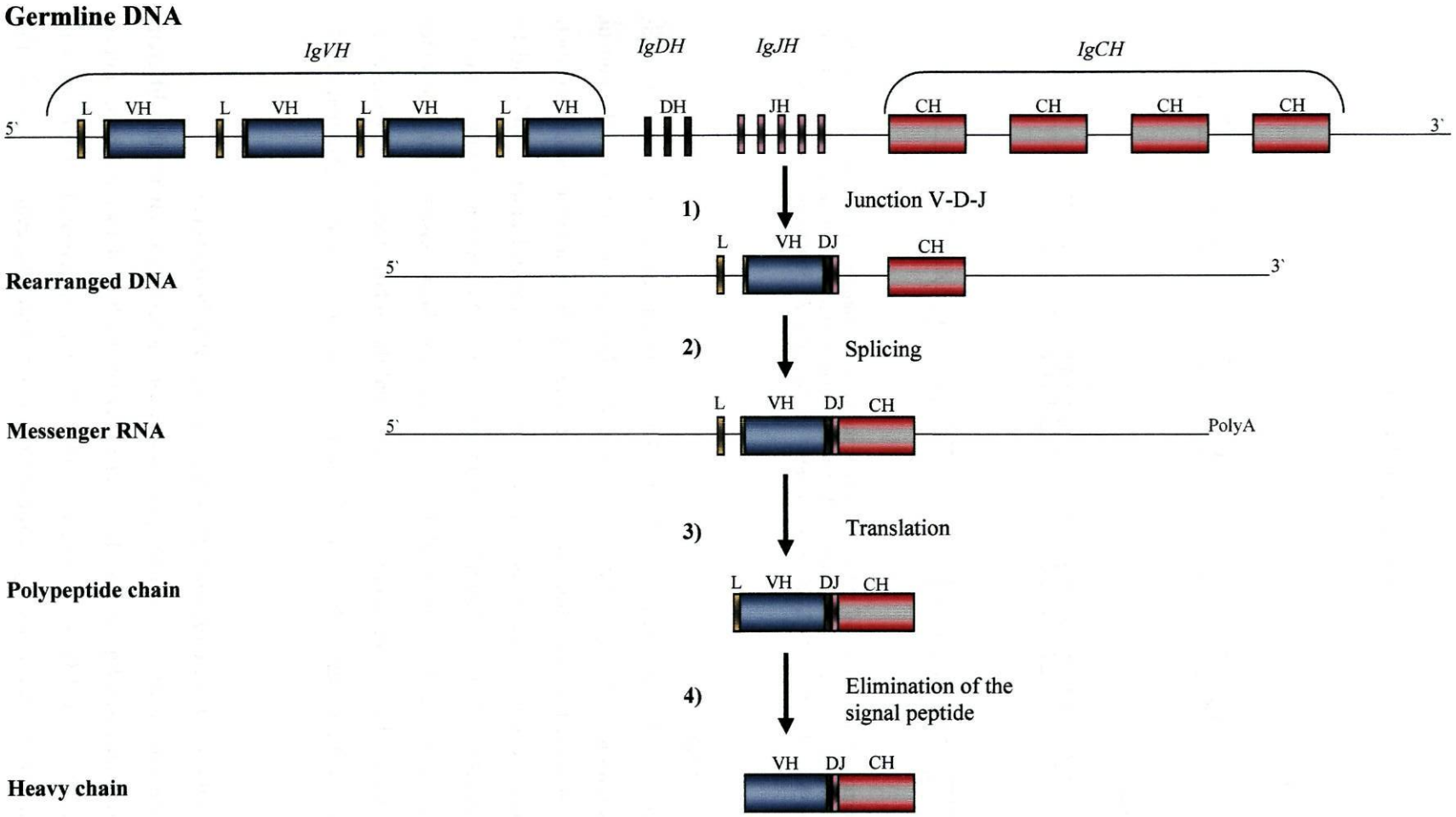



**Figure 1.4:** Synthesis of an immunoglobulin. 1) At the DNA level, one *IgDH* gene is joined to one *IgJH* gene, with deletion of the intermediary DNA, to create a partially rearranged *D-J* gene. 2) In a second step, one of the *IgVH* gene is joined to *D-J*, with the deletion of the intermediary DNA, to generate a completely rearranged *IgVH-D-J* gene. This rearrangement is transcribed with the *IgCH* gene into a *IgVH-D-J-C* pre-messenger RNA, the RNA sequences corresponding to the introns and to the non used *IgJH* genes are excised by splicing, and a mature messenger, which comprises the spliced regions, and the 5`and 3` untranslated sequences, is obtained. 3) The messenger RNA is translated into a polypeptide chain by the ribosome. 4) The signal peptide is cleaved off by a peptidase following the entry of the polypeptide chain in the endoplasmic reticulum, and a mature heavy chain is produced.

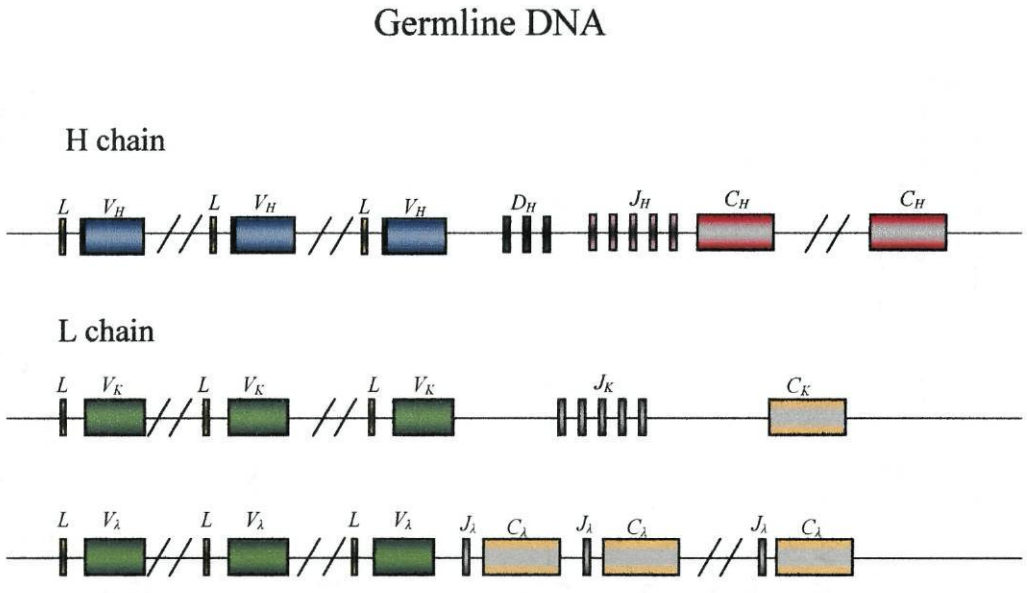



**Figure 1.5.** Schematic representation of the localisation in the human Ig molecule of the gene segments that compose the H and L chain. H chain: *L* – Leader, *V* – Variable, *D* – Diversity, *J* – Joining and *C* – Constant gene segments. Kappa (K) and Lambda (L) L chains: *L* – Leader, *V* – Variable, *J* – Joining and *C* – Constant gene segments.

*Constant region*

The number of genes encoding the H chain constant regions can vary between species. For example, in the mouse it is composed of eight *C* genes whereas in the human this gene region consists of nine functional genes. The rabbit shows an unusual organization of the H chain *C* gene loci, since it possesses only one $C_\gamma$ gene (see Hamers, 1987) and 13 $C_\alpha$ genes (Burnett *et al.*, 1989; Knight and Winstead, 1997). Neither the unicity of the $C_\gamma$ gene, nor the extensive duplication of the $C_\alpha$ genes has been observed in any other mammalian species. The expansion of $C_\alpha$ genes is found throughout the order *Lagomorpha* and suggests that the $C_\alpha$ genes began to duplicate in a common ancestor (Burnett *et al.*, 1989).

*Heavy chain genomic organization in different vertebrate species*

The gene organization of immunoglobulins varies extensively among the different classes of vertebrates (see Figure 1.6). In lower vertebrate such as sharks, the basic unit of repeats is *V-D-D-J-C*, and linked gene groups $(V\text{-}D\text{-}D\text{-}J\text{-}C)_n$ are repeated several times in throughout the chromosomes. In the genome of higher vertebrates the gene segments for the different components of the Ig chains are linked together (*V*n-*D*n-*J*n-*C*n).

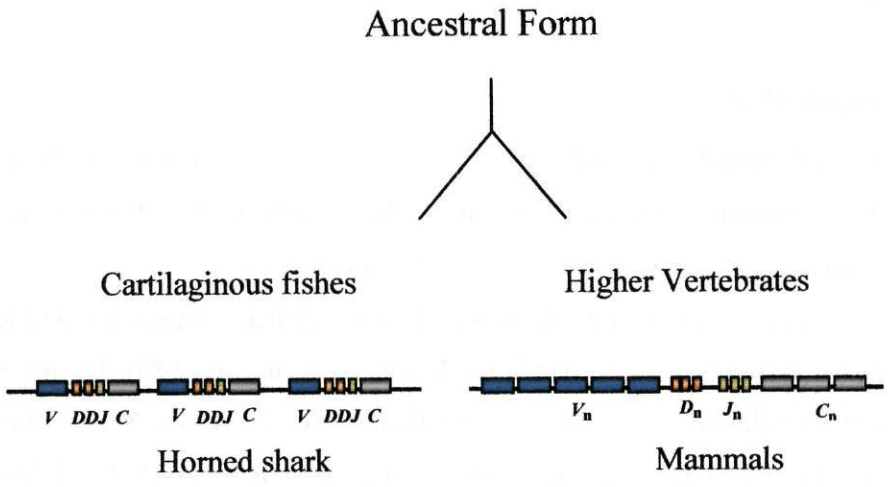



**Figure 1.6.** Genomic organization of immunoglobulins H chain genes. Cartilaginous fishes have the $(V_H\text{-}D_H\text{-}D_H\text{-}J_H\text{-}C_H)$n type organization including germline joined genes, whereas most higher vertebrates, have the $(V_H)$n-$(D_H)$n-$(J_H)$n-$(C_H)$ type gene organization. *V* – Variable segment gene; *D* – Diversity segment genes; *J* – Joining segment gene; *C* – Constant segment gene.

## Light Chain

Two Ig L chain isotypes (K and λ), that are functionally indistinguishable, exist as separate gene loci in mammals.

### *Variable region*

Two types of germline elements encode the Ig V region domain of both Kappa (K) and Lambda (λ) L chains: $V_L$ (Variable) and $J_L$ (Joining) segments. The structure of $V_L$ segments is identical to that described above for $V_H$.

In the mouse, 95% of the L chains are of the K type and only 5% are of the λ type (Cotner *et al.*, 1978). 140 $V_K$ and five $J_K$ genes, localized on chromosome 6, have been identified. The $V_K$ genes were subdivided into 18 gene families (Kirshbaum *et al.*, 1998). The λ locus is localized in chromosome 16 and is composed by 3 *V* segments and four *C* segments (e.g. Sanchez *et al.*, 1991).

In the human only 60% of immunoglobulins are of the K type (Fu *et al.*, 1979). There are 76 $V_K$ genes, from which 32 are potentially functional, and five $J_K$ genes, localized on the chromosome 2 (Malcolm *et al.*, 1982). Human $V_\lambda$ genes have been grouped into 10

families. There have been identified 70 genes, from which 37 are functional (e.g. Stiernholm *et al.*, 1994).

*Constant region*

The mouse $C_L$ gene loci are composed of a single $C_K$ gene and four $C_\lambda$. Each of the $C_\lambda$ is associated with its own $J_\lambda$ segment. The human CK also consists of only one gene, while the $C_\lambda$ locus is composed of seven $J_\lambda$-$C_\lambda$ segments (Figure 1.5)

Rabbit uses almost exclusively the K chain, of which two isotypes, K1 and K2, are described (Benammar and Cazenave, 1982; Heidmann and Rougeon, 1983). The majority of rabbit $V_K$ genes encodes an extra cysteine at position 80 (see for review Mage, 1987). K1 chain is more expressed than K2, and differs from all other known Ig chains by a supplementary disulphide bond between the VK and the CK domain. This is performed by cysteines at positions 80 ($V_K$) and 171 ($C_{K1}$). A schematic representation of rabbit partial *K* chain locus is represented in Figure 1.7.

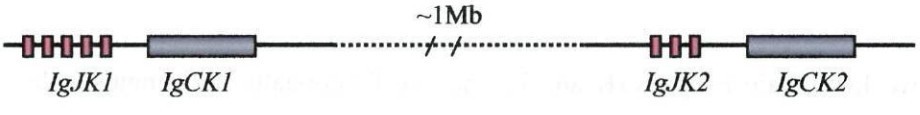



**Figure 1. 7.** Schematic representation of rabbit K Light chain isotypes 1 and 2. $J_K$–Joining K chain genes; $C_K$ – Constant K chain genes.

## Immunoglobulin class switch

During B cell differentiation into memory or Ab secreting cells, the isotype of the Ig H chain changes from μ to γ, ε or α. This important mechanism allows changing the biological effector function of the Ab without changing its specificity. These Ig class switches occur within the cell. It was shown that the same cell that previously secreted IgM could start the secretion of IgG (van der Loo *et al.*, 1979). During the process of class switching $C_H$ genes are replaced through a site-specific DNA recombination process, which takes place between switch regions (S) located upstream from the $C_H$ genes. Switch recombination is accompanied by intramolecular DNA deletion, letting the switched H chain gene on the chromosome (Storb *et al.*, 1981).

# Generation of antigen-binding repertoire

The efficiency of the immune system is in part due to the potential of generating a bewildering variety of receptor molecules. This potential is such that it can match the wide variety of antigens and even anticipate antigen variation. The generation of primary repertoires and Ig gene diversification can occur by many different mechanisms. Furthermore, the anatomical site of primary repertoire diversification varies among species. Indeed, studies in mice and humans identified bone marrow and foetal liver as the site of B cell genesis and development (Claman *et al.*, 1966; Nuñez *et al.*, 1996), whereas the bursa of Fabricius is the site for B cell lymphopoiesis in birds (Cooper and Lawton, 1972). In sheep and cattle, ileal Peyer's patches serve as the main site of B cell diversification (Reynolds and Morris, 1983; Reynolds, 1987). While diversification occur in the bovine spleen (Lucier *et al.*, 1988). In the rabbits, B cell diversification takes place in the appendix (e.g. Archer *et al.*, 1963; Cooper *et al.*, 1968).

*Germline and combinatorial diversity*

The large number of inherited *V*, *D*, and *J* gene segments available for generating the V domain of the Ig chain contributes substantially to antibody diversity. This is well established in the case of mice and humans, where the B cells express different receptor molecules due to numerous functional *V*, *D*, and *J* segments.

Combinatorial diversity results from the random selection for recombination of inherited *V*, *D* and *J* segments in the case of H chain and *V*, and *J* segments in case of L chain. The large primary repertoire produced by the somatic juxtaposition of the different segments is further increased by the random choice of individual H and L chain V domains. In theory, there are $>10^7$ possible combinations.

*Junctional diversity*

Each H and L chain are the outcome of chromosomal rearrangements involving the $V_H$ and $V_L$ gene segments (*V*(*D*)*J* joining). *V* exons are flanked at the 3'end by a recombination signal sequence (RSS) which consists of conserved heptameric (CACAGTG) and nonameric (ACAAAAACC) sequences separated by a spacer sequence of 23-bp in the H and λ L chain and of 12-bp in K L chain (see Figure 1.8). Each *D* exon is flanked on both

sides by an RSS with a short signal element spacer of 12-bp. The $J_H$ and $J_K$ exons possess at its 5`end an RSS with the long spacer of 23-bp, while $J_\lambda$ possesses a spacer of 12 bp. Recombination takes place between two heptamer-nonamer pairs, and can occur only when one pair is separated by a 12-bp spacer, and the other pair is separated by a 23-bp spacer, (12-23 spacer-rule; for review see Akamatsu *et al.*, 1994). The *V(D)J* type recombination is determined by the recombination-activating proteins, RAG1 and RAG2. It has been shown that these proteins are involved in the cleavage reaction recognizing the RSS and in the efficient joining of coding ends (Ramsden *et al.*, 1997; Agrawal *et al.*, 1998).

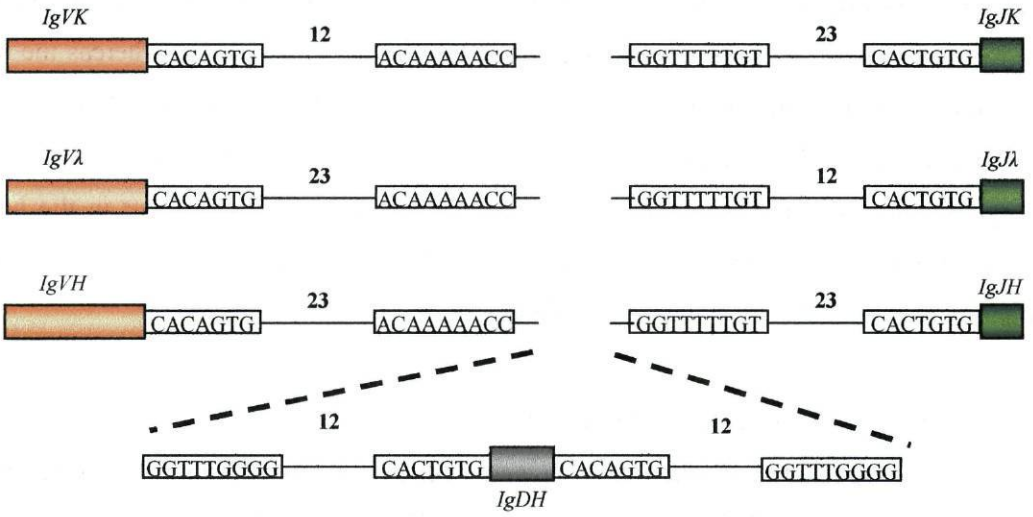



**Figure 1.8.** Schematic representation of *IgV*, *D* and *J* gene recombination signals. The conserved heptameric and nonameric sequences that flank each of the segments are represented in boxes. See details in the text.

Much of the antibody diversity is generated from the imprecise joining of *V*, *D* and *J* gene segments. This somatic process deletes and sometimes inverts short DNA segments in lymphocytes, and has so far been observed in most of the major lineages of jawed vertebrates, including cartilaginous fishes. This random loss or gain of nucleotides increases specifically the diversity of the third hypervariable region (CDR3). Two types of mechanisms are responsible for the incorporation of additional nucleotides at the coding ends: a) introduction of palindromic nucleotide sequences, that incorporate a few extra nucleotides reversed in orientation to the original coding end; b) n-region addition, that consists of a random insertion of non-template nucleotides between the coding ends of H chain gene segments, and is catalysed by terminal deoxynucleotidyl transferase (TdT) (Landau *et al.*, 1987; Meier and Lewis, 1993).

## Post-rearrangement diversification

Although these two mechanisms are able to generate an extensive repertoire of possible Ag-binding sites, several species possess, or use, only a limited number of functional germline *V* segments. For example birds, together with a number of mammal species (rabbit, sheep, pig and bovine), have less extensive germline repertoires than mice and humans and consequently use strategies of primary repertoire development that overcome this limitation. In these cases, the primary antibody repertoire is diversified after *V(D)J* rearrangement by two mechanisms: somatic hypermutation and gene conversion.

*Somatic hypermutation* is characteristic of Ig genes and affects the affinity and the specificity of the antibody repertoire. The mechanism responsible for mutations is targeted to rearranged *V* gene segments, both active and inactive, introducing immediate point mutations at a rate of about $10^{-4}$ to $10^{-3}$ per base pair per cell generation, which is clearly higher than the normal frequency of random mutations. The general pattern of somatic mutations is not random, since replacement mutations are often over-represented in the CDRs, whereas silent mutations are concentrated in the FRs. Like any other mechanism introducing mutations, there are preferred sequences and pronounced hotspots for somatic hypermutation (Betz *et al.*, 1993; Reynaud *et al.*, 1995; Belessi *et al.*, 2001). Indeed, hotspot clusters have been observed in the CDRs 1, 2, and 3 regions as well as in the J-C region intron (Motoyama *et al.*, 1991; González-Fernandez *et al.*, 1994; Lanning and Knight, 1997).

*Gene conversion* is a non-reciprocal transfer of genetic information. In the particular case of Ig diversification, it involves the replacement of segments of the $V_H$ gene engaged in *V(D)J* joining, with homologous sequences from upstream *V* genes. Gene conversion is the major mechanism of the Ab diversification in the chicken, where a set of pseudogenes are used as donors and the unique rearranged $V_H/V_L$-*(D)J* acts as acceptor (Reynaud *et al.*, 1985). The diversification takes place in the Bursa of Fabricius and in the spleen upon immunization (Arakawa *et al.*, 1996, 1998). A similar mechanism is found in the rabbit H chain loci, where the rearrangement involves a single $V_H$ gene that is diversified mainly through gene conversion using other upstream $V_H$ genes (Becker and Knight, 1990; Knight and Crane, 1994). Somatic hypermutation also takes place throughout the entire *VDJ* genes (Short *et al.*, 1991; Weinstein *et al.*, 1994). The rabbit's diversification process is explained

in more detail in chapter 3. In sheep, the B cell diversification occurs before birth in early development in the ileal Peyer's patches as a result of somatic hypermutation and gene conversion (Reynaud *et al.*, 1991). Both processes are used to diversify the Ag-binding repertoire in cattle (Parng *et al.*, 1996; Lucier *et al.*, 1998). Gene conversion mechanisms are possibly the main source of Ab repertoire diversification in swine (Sun and Butler, 1996).

# Rabbit taxonomy and genetics

## Taxonomy of Lagomorpha group

The order LAGOMORPHA was proposed by Gridley (1912) to distinguish the lagomorphs from the rodents. The phylogenetic position of the lagomorph group is not consensual. According to fossil data the lagomorphs are associated with rodents (e.g. Novaceck, 1992; Meng *et al.*, 1994), and classified within the super-order GLIRES. Based on molecular data, two conflicting evolutionary scenarios have been proposed for the mammals, placing the lagomorphs in different phylogenetic positions. For some authors the rodents group separated 110 Million years (My) ago, before the great mammalian radiation (90 My) and in this case the lagomorphs are more related to the primates and artiodactyls (Graur *et al.*, 1996; Kumar and Hedges, 1998; Nei *et al.* 2001). Other authors cluster the rodents with the lagomorphs and propose that the artiodactyls diverged 95 My ago, before the separation of the primates from the ancestral of rodent/lagomorphs (Huchon *et al.*, 1999; Liu and Miyamoto, 1999; Adkins *et al.*, 2001; Murphy *et al.*, 2001; Springer *et al.*, 2003).

The order LAGOMORPHA is divided into two families: OCHOTONIDAE and LEPORIDAE. The family OCHOTONIDAE is restricted to the genus *Ochotona* (Pikas) that includes 25 extant species. These species are of small body size, possess 26 teeth and are adapted to alpine environments. The family LEPORIDAE can be divided into two groups: hares and rabbits. They have well-developed posterior members, elevated eyes, long ears and 28 teeth. According to Chapman and Flux (1990), the hare group encompasses a single genus, *Lepus,* comprising 29 species, whereas the rabbit group includes 10 genera (*Brachylagus*, *Bunolagus*, *Caprolagus*, *Nesolagus*, *Oryctolagus*, *Pentalagus*, *Poelagus*, *Pronolagus*, *Romerolagus* and *Sylvilagus*) and 25 species. The genus *Oryctolagus* is monospecific (*O. cuniculus*) while for genus *Sylvilagus* 13 species are described. The genus *Ochotona* is present in Northern Asia and West North America, the genus *Oryctolagus* is present in Continental Europe, North Africa (Morocco and Algeria), England, islands over the world, Australia, New Zealand, and South America. The genus *Lepus* is distributed in Holartic, Indomalayan and Afro-tropical regions (except tropical forest) and the genus *Sylvilagus* is restricted to the American continent (Corbet, 1994).

*Paleontological data*

The lagomorph fossil data are scarce and fragmented. Some authors suggested that the lagomorphs appeared in Asia during the later Eocene (approximately 45 My ago) (Erbajeva, 1981). The leporids probably separated from ochotonids during the Oligocen or Upper Eocene (30-40 My ago) (Dawson, 1981). The family LEPORIDAE developed and differentiated in North America and Asia (Dawson, 1981; Lopez-Martinez, 1998) and arrived in Europe during the great migration of the upper Miocene (around 20 My ago) with posterior radiation (Lopez-Martinez, 1989).

*Molecular data*

Recently, the study of 19 nuclear and three mitochondrial genes from 42 representative species of all extant orders of placental mammals confirmed previous estimates that the genus *Ochotona* diverged from the rabbit at 50-55 My ago (Springer *et al.*, 2003). The divergence time between rabbits and hares is not consensual. The cytochrome *b* analyses suggest that the genera *Oryctolagus*, *Lepus* and *Sylvilagus* diverged 6-8 MY ago (Biju-Duval *et al.*, 1991). On the other hand the study of 12S rRNA indicated that the most recent common ancestral lived 12.2-16.3 My ago (Halanych and Robinson, 1999). Polychlorinated biphenylbinding protein gene (PCB) studies pointed out a time of divergence between *Oryctolagus* and *Lepus* of 20 My (Su and Nei, 1999).

## The genetic diversity of European rabbit

In the Iberian Peninsula there are two morphologically differentiated subspecies of rabbit (*Oryctolagus cuniculus)*: *O. c. algirus* and *O. c. cuniculus* (Cabrera, 1914). In view of the small and fragmented fossil record, it seems that the European rabbit was probably originated in the Iberian Peninsula during the medium Pleistocene and then expanded towards Southern France (Pages, 1980; Lopez-Martinez, 1989; Corbet, 1994). The Phoenicians 3000 years ago carried out what might be the first human-mediated geographical expansion of the rabbit species by introducing it into Mediterranean Islands (Balearic Islands, Corsica, Sardinia, Tunisian islands) (Vigne, 1988; Hardy *et al.*, 1994). Today, the rabbit is present in Continental Europe, England, Australia, New Zealand, South America, and North Africa (Morocco and Algeria).

Genetic variation in wild and domestic rabbit populations has been studied in detail. We give an overview of molecular genetic studies performed that involved (a)

mitochondrial DNA (mtDNA), (b) protein variation, (c) immunological markers, and (d) microsatellites.

## Mitochondrial DNA (mtDNA)

The study of the mtDNA molecule (cytochrome *b*) in rabbit populations revealed two major maternal lineages, one that is present in wild populations from Southwestern Iberian Peninsula, called lineage *A*, and other, called lineage *B*, present in the Northeast of Iberian Peninsula, in all domestic breeds and in all wild populations of Continental Europe (north of Pyrenean Mountains), Great Britain and overseas. These lineages showed 4% of nucleotide divergence, which suggests that, using the average rate of evolution of mammalian mtDNA, these two groups diverged in the Pleistocene (about 2 million years) (Biju-Duval *et al.*, 1991). Later studies of more French and Spanish wild rabbit populations led to the suggestion of two refugia that separated two MY ago, one in the Southwest of the Iberian Peninsula, and another probably located in Northern Spain (Monnerot *et al.*, 1994). Branco *et al.* (2000, 2002) showed the existence of a contact zone in the Iberian Peninsula suggesting a post-glacial expansion from the Southwestern refugium to North, and from the Northeastern to South or West.

## Genetic variation of proteins

The first works that used the genetic variation of proteins in the characterisation of rabbit populations confirmed the British origin of the feral rabbits of Australia (Richardson *et al.*, 1980). The study of wild rabbit populations from the Iberian Peninsula, Azores (Atlantic islands) and France, using 16 polymorphic markers confirmed the hypothesis that the rabbit was originated in the Iberian Peninsula (Ferrand, 1995). The same author also suggested the existence of two glacial refugia in which the long-term geographical isolation gave rise to two different rabbit subspecies. One refugium was located in Southwestern Iberian Peninsula, where subspecies *O. c. algirus* originated. The other refugium was localized in Northeastern Spain and Southern France, where the rabbit *O. c. cuniculus* originated. The study of numerous protein markers in wild rabbit populations from the Iberian Peninsula revealed a broad and gradual transition between the Southwestern and Northeastern groups of populations, suggesting the absence of intrinsic barriers to gene flow between these two evolutionary units (Branco, 2000).

## Rabbit Ig allotypes

The first genetic studies at the protein level were done on laboratory rabbits (Oudin 1956). By immunizing one rabbit with the antibody of another rabbit, Oudin discovered the Ab allotypes. At a time that the Ig molecule was believed to be monomeric, rabbit genetics showed that the Ab was encoded by two independently segregating loci, called *a* locus and the *b* locus. The *a* locus was identified as the locus encoding the H chain variable region (*Ig*$V_H$*a*-locus), while the *b* locus encodes the constant part of the L chain (*IgCK*1 locus). The variability at both the *a* and the *b* locus is characterized by extremely large distances between alleles. The largest distances exist among *b* locus alleles (up to 35% of amino acid differences; note that the CK regions of human and chimpanzee are identical).

While the *a* and *b* markers concern the Fab region, genetic variation was also observed at the Fc region (H chain constant region). In contrast with the rabbit "complex" allotypes, these markers resemble the Ig allotypes described in species like human or mouse, and relate to single amino acid substitutions. One of them, the *e*-locus, of the *CH2* exon of the IgG class, was studied in this work.

The very large divergence at the $C_{K1}$ locus, which was confirmed as a structural single-gene locus, makes it instrumental for evolutionary studies of the rabbit species. In domestic breeds, wild and feral rabbit populations from the most recent distribution range of the species (Continental Europe North of the Loire river, Great Britain, Australia, Kerguelen islands), four alleles can be distinguished by serological methods at the *b* locus, the so-called *b4*, *b5*, *b6* and *b9* (van der Loo and Arthur, 1986; van der Loo, 1987, 1993). The study of this locus in Iberian populations by serological methods revealed at least 16 alleles (Cazenave *et al.*, 1987; van der Loo *et al.*, 1991, 1999). This high number of alleles is in accordance with the hypothesis that the rabbit originated in the Iberian Peninsula.

Cytonuclear disequilibrium was observed in wild rabbit populations from the Iberian Peninsula. The alleles *b4* and *b5* of *IgCK1* that comprise over 90% of the gene pool in the more recent area of distribution of the species were also predominant in areas of mtDNA type *B* prevalence within the Iberian Peninsula. However, in areas of mtDNA lineage *A* prevalence, the *b4* and *b5* allotypes are rare or absent, the majority of the alleles are serologically related to *b4* and *b5*, but are distinct "endemic" variants (van der Loo *et al.*, 1999).

The DNA sequence comparison of the rabbit *IgCK1* alleles revealed that substitutions rates were significantly larger at amino acid replacement sites than at synonymous sites (van

der Loo and Verdoodt, 1992), while non-coding regions were more conserved than exons (Akimenko *et al.*, 1986). This unusual form of allelic diversity is therefore due to forces favouring variation at the protein level. Furthermore, population genetic studies revealed a generalised excess of heterozygous individuals at *b* locus but not at other loci (van der Loo and Arthur, 1986; van der Loo, 1987). Linkage disequilibrium between the *b* locus and *e locus* (van der Loo *et al.*, 1987, 1996; van der Loo, 1993) strongly suggested that overdominance-type selection is occurring at this locus, because these loci segregate independently.

## Microsatellite markers

In the rabbit, 27 polymorphic microsatellite loci have been described (van Haeringen *et al.*, 1997; Mougel *et al.*, 1997; Surridge *et al.*, 1997; Ferrand *et al.*, 2000). In contrast to the results obtained in the above-mentioned markers, the use of nine microsatellite loci in the genetic characterization of rabbit populations from the Iberian Peninsula did not confirm the occurrence of two well-separated groups. Indeed, these loci revealed severe allele-size homoplasy, which underestimates the divergence time between the main groups of populations in Iberia. However, for the same set of markers it was detected a significant difference in allelic diversity between the French and the Iberian Peninsula populations (Queney *et al.*, 2001).

## Domestication

The rabbit is the only domesticated mammal of Western Europe origin. This process began late in the Middle Ages, probably in southern France monasteries (Robinson, 1984; Callou, 1995). The genetic characterisation of 12 domestic breeds through the study of mtDNA, microsatellites, immunological and protein markers, (Bolet *et al.*, 1999; Queney *et al.*, 2002; Esteves *et al.*, unpublished; van der Loo *et al.*, unpublished) showed that only a very small fraction of the species diversity is present in the domestic breeds. The results obtained in this study are congruent with previous publications on the immunoglobulins polymorphism (van der Loo *et al.*, 1991) and genetic variation of proteins (Ferrand, 1995) and consolidate the viewpoint that domestic breeds originated from France.

# The rabbit in immunological research

The use of rabbit in immunological research is very old. In the nineteenth century, Louis Pasteur used the rabbit spinal cord to prepare rabies vaccine (Pasteur *et al.*, 1881). The study of rabbit was also crucial in the battle against syphilis (Wassermann *et al.*, 1906; Parodi, 1907). In the first half of the twentieth century, the foundations of molecular immunology have been 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). Studies on laboratory rabbit have greatly contributed to our knowledge of the structure, function and regulation of antibodies. In 1956, Oudin demonstrated and defined allotypy of immunoglobulins in the rabbit. Three years later Dubiski *et al.* (1959) showed that the rabbit allotypes were hereditary traits and in 1960, Oudin established the two major linkage groups "*a*" and "*b*" corresponding to the H chain and the K light chain loci, respectively. Hamers and Hamers-Casterman (1965) discovered allotypes of the Ig constant region (*e* locus). The existence of genetic markers of the different gene segments was – and to a large extent still is – unique to this species. The rabbit Ig allotypes allowed challenging the "one gene-one protein" dogma. Indeed, Todd (1963) and others found that in rabbit allotypic specificities of the variable regions presented by IgG or IgM molecules were identical, suggesting that a same $V_H$ gene segment can be translocated to different constant region genes.

This observation lead to the hypothesis of the Ig class switch (Severinson *et al.*, 1982), which opened the road to the concept of multiple "germline" $V_H$ gene segments that can be joined to a limited number of genes encoding the different H constant regions (Dreyer *et al.*, 1965). In 1962, Hamers and co-workers showed the cis-expression of the $V_H$ and $C_H$ genes, by using the associated genetic markers (i.e. $V_Ha$ or *a* locus and *CH2* or *e* locus). Later, Mage *et al.* (1971) confirmed this model by documenting a number of crossing-over events between the genes controlling the rabbit variable and constant heavy chain. Mage (1979) and Kelus and Steinberg (1991) estimated a crossing-over frequency of 0.1%.

The rabbit allotype model is therefore at the origin of the current *V-D-J-C* model. It was through studies of rabbit immunoglobulin markers that phenomena such as allelic exclusion, allelic imbalance in gene expression, genetic linkage of $V_H$ and $C_H$ genes, and the expression of apparently identical VH regions on different classes of Ig were described.

However, the use of the rabbit in several immunological fields was restricted by technical difficulties (e.g. failure to induce rabbit myeloma tumors, and the absence of inbred strains). Nevertheless, the rabbit is still important in transplantation and ophthalmologic research. Rabbit antibodies are more stable than those from rodents and are therefore often used in the production of high affinity antibodies or of anti-idiotype vaccines. During the last 25 years the rabbit Ig allotypes revealed a number of unique features (described in chapters 2 and 3), setting them apart from the allotypes found in mice and human. Most important are the highly unusual large genetic distances between the allelic lineages, which can reach 35% of amino acid divergence between *IgCK1* (*b* locus) alleles.

It appeared that this unusual degree of allotype diversity is the outcome of selection (van der Loo, 1993). The detection of selection on polymorphic genes in natural populations is particularly difficult. The rabbit unites a number of characteristics that make it an ideal species for studies of the biological importance of population diversity at the Ig loci (van der Loo 1987): 1) the rabbit Ig allotypes are well characterized at the genetic, structural and genomic level; 2) the rabbit is a popular laboratory animal and a game species with the physiology and ecology well studied; 3) the population dynamics of this species (high mortality and birth rates), makes it more likely that the genotype dependent components of mortality can be recorded at the population genetic levels; 4) the rabbit is a recent world wide distributed species with a documented and interesting history of epizootic diseases (myxomatosis and hemorrhagic viral diseases).

Another unusual feature of the rabbit immune system resides in the fact that the antibody repertoire is diversified in this species after birth. The primary Ab repertoire of rabbit results from the preferential rearrangement of a very limited number of *V* gene segments (i.e. those that encode the allotypic markers). The B cells are developed in the fetal liver and bone marrow and migrate to appendix and other gut-associated lymphoid organs (GALT) where are diversified by somatic hypermutation and/or somatic gene conversion-like mechanism. In the rabbit this process only occurs at 1-2 months of age (Cooper et al., 1968; Weinstein *et al.*, 1994) and is not developmentally regulated like it appears to be in other species, such as chicken, sheep and cattle. It has been shown that the Ab repertoire diversification requires interaction of exogenous factors, such as the intestinal microflora (Lanning *et al.*, 2000a, 2000b; Sehgal *et* al., 2002) (for details see chapter 3).

# Citations

Adkins R, Gelke E, Rowe D and Honeycutt L (2001). Molecular phylogeny and divergence time estimates for major rodent groups: evidence from multiple genes. *Mol Biol Evol* **18:** 777-91.

Agrawal A, Eastman QM, Schtz DG (1998). Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. *Nature* **394:** 744-51.

Akamatsu Y, Tsurushita N, Nagawa F, Matsuoka M, Okazaki K, Imai M and Sakano H (1994). Essential residues in V(D)J recombination signals. *J Immunol* **153**: 4520-529.

Akimenko MA, Mariame B and Rougeon F (1986). Evolution of the immunoglobulin kappa light chain locus in the rabbit: evidence for differential gene conversion events. *Proc Natl Acad Sci* U S A **83**(14): 5180-183.

Almagro JC, Hernández L, Ramirez M de C and Vargas-Madrazo E (1997). The differences between the structural repertoire of VH germ-line gene segments of mice and humans: implications for the molecular mechanism of the immune response. *Mol Immunol* **34**: 1199-214.

Arakawa H, Furusawa S, Ekino S and Yamagishi H (1996). Immunoglobulin gene hyperconversion ongoing in chicken splenic germinal centers. *EMBO J* **15:** 2540-546.

Arakawa H, Kuma K, Yasuda M, Furusawa S, Ekino S and Yamagishi H (1998). Oligoclonal development of B cells bearing discrete Ig chains in chicken single germinal centers. *J Immunol* **160**(9): 4232-241.

Archer OK, Sutherland DER and Good RA (1963). Appendix of the rabbit: a homologue of the bursa in chicken. *Nature* **200**: 337-40.

Becker RS and Knight KL (1990). Somatic diversification of immunoglobulin heavy chain VDJ genes: evidence for somatic gene conversion in rabbits. *Cell* **63**(5): 987-97.

Belessi C, Stamatopoulos K, Stavroyianni N, Zoi K, Papadaki T and Kosmas C (2001). Somatic hypermutation targeting to intrinsic hotspots of immunoglobulin genes in follicular lymphoma and multiple myeloma. *Leukemia* **15**(11): 1772-778.

Benammar A and Cazenave PA (1982). A second rabbit kappa isotype. *J Exp Med* **156(2)**: 585-95.

Betz AG, Neuberger MS, Milstein C (1993). Discriminating intrinsic and antigen-selected mutational hotspots in immunoglobulin V genes. *Immunol Today* **14**(8): 405-11.

Biju-Duval C, Ennafaa H, Dennebouy N, Monnerot M, Mignotte F, Soriguer RC, Gaaied A, Hili A and Mounolou JC (1991). Mitichondrial DNA Evolution in Lagomorphs: Origin of Systematic Heteroplasmy and Organization of Diversity in European Rabbits. *J Mol Evol* **33:** 92-102.

Bolet G, Monnerot M, Arnal C, Arnold J, Bell D, Bergoglio G, Besenfelder U, Bosze ZS, Boucher S, Brun JM, Chanteloup N, Ducourouble MC, Durand-Tardif M, Esteves PJ, Ferrand N, Hewitt G, Joly T, Koehl PF, Laube M, Lechevestrier S, Lopez M, Masoero G, Piccinin R, Queney G, Saleil G, Surridge A, van der Loo W, Vanhommerig, J, Vicente JS, Virag G and Zimmermann JM (1999). A program for the inventory, characterisation, evaluation, conservation and utilisation of european rabbit (*Oryctolagus cuniculus*) genetic resources. *Animal Genetic Resources Information,* **25**: 57-70.

Branco M (2000). *Estrutura genética das populações de coelho europeu (Oryctolagus cuniculus) na Península ibérica. Isolamento, diferenciação de duas unidades evolutivas, expansão geográfica e contacto secundário.* Dissertação de Doutoramento, Universidade do Porto.

Branco M, Ferrand N and Monnerot M (2000). Phylogeography of the European Rabbit (*Oryctolagus cuniculus*) in the Iberian Peninsula inferred from RFLP analysis of the cytochrome b gene. *Heredity*, **4:** 307-17.

Branco M, Monnerot M, Ferrand N and Templeton AR (2002). Postglacial dispersal of the European Rabbit (*Oryctolagus cuniculus*) on the Iberian Peninsula reconstructed from nested clade and mismatch analyses of mitochondrial DNA genetic variation. *Evolution* **56(4):** 792-803.

Burnett RC, Hanley WC, Zhai SK, and Knight KL (1989). The IgA heavy chain gene family in rabbit: cloning and sequence analyses of 13 Cα genes. *EMBO J* **8:** 4041-047.

Cabrera (1914). *Fauna Ibérica.* Museo Nacional de Ciências Naturales de Madrid, *Madrid.*

Callou C (1995). Modifications de l'aire de répartition du Lapin (*Oryctolagus cuniculus*) en France et en Espagne, du Pléistocène à l'époque actuelle. État de la question. *Anthropozoologica* **21**: 95-114.

Cazenave PA, Bennamar A, Sogn JA and Kindt TJ (1987). Immunoglobulin genes in feral populations. In Dubiski S (ed.). *The Rabbit in Contemporary Immunological Research.* Longman Scientific & Technical. pp 148-60.

Chapman J and Flux J (1990). Introduction and overview of the lagomorphs. *In Rabbits, Hares and Pikas, Status survey and conservation action plan.* (Eds) J Chapman and J Flux. UICN/SSC Lagomorph Specialist Group, Switzerland. Gland, pp. 1-6.

Chothia C, Novotny J, Bruccoleri R and Karplus M (1985). Domain association in immunoglobulin molecules. The packing of variable domains. *J Mol Biol* **186**: 651-663.

Claman HN, Chaperon EA and Triplett RF (1966). Thymus-marrow cell combinations. Synergism in antibody production. *Proc Soc Exp Biol Med* **122**: 1167-171.

Coloma MJ, Trinh KR, Wims LA and Morrison SL (1997). The hinge as a spacer contributes to covalent assembly and is required for function of IgG. *J Immunol* **158:** 733-40.

Cooper MD, Perry DY, Gabrielsen AE, Sutherland DE, McKneally MF and Good RA (1968). Production of an antibody deficiency syndrome in rabbit by neonatal removal of organized lymphoid intestinal tissue. *Int Arch Allergy Appl Immunol* **33**: 65-88.

Cooper MD and Lawton AR (1972). The mammalian "bursa equivalent": does lymphid differentiation along plasma cell lines begin in the gut-associated lympho-epithelial tissues (GALT) of mammals? *Contemp Top Immunobiol* **1**: 49-68.

Corbet GB (1994). Taxonomy and origins. In: *The European Rabbit* (eds HV Tompson and CM king), Oxford Science Publications, Oxford. pp 1-6.

Dawson M (1981). Evolution of the modern Lagomorphs. In *Prooceding of the world Lagomorph Conference*. K Myers and CD McInnes (Eds.), Univ. Guelph, Ontario. pp 1-16.

Deisenhofer J (1981). Crystallographic refinement and atomic models of a human Fc fragment and its complex with fragment B of protein A from staphylococcus aureus at 2.9- and 2.8-A resolution. *Biochemistry.* **20:** 2361-370.

Dreyer WJ, Bennet JC (1965). The molecular basis of antibody formation: a paradox. *Proc Natl Acad Sci USA* **54:** 864.

Dubiski S, Dudziak Z, Skalba D and Dubiski A (1959). Serum groups in rabbits. *Immunology* **2**: 84.

Erbajeva M (1981). Late cenozoic lagomorpha of Transbaikalia. In *Proceedings of the world Lagomorph conference*. K Myers and CD McInnes (Eds). Univ. Guelph, Ontario. pp 53-55.

Ferrand N (1995). *Variação genética de proteínas em populações de coelho (Oryctolagus cuniculus). Análise da diferenciação subespecífica, subestruturação, expansão geográfica e domesticação*. Dissertação de Doutoramento, Universidade do Porto.

Ferrand N, Azevedo M and Mougel F (2000). A diallelic short tandem repeat (CCCCG)4 or 5, located in intron 1 of rabbit alpha-globin gene. *Anim Genet* **31**(1): 74-5.

Frazer JC and Capra JD (1999). Immunoglobulins: *Structure and function. In: Fundamental Immunology*, 4th edition, Paul W. E., Lipincot-Raven publishers. pp 37-74.

Fu SM, Chiorazzi N and Kunkel HG (1979). Differentiation capacity and other properties of the leukemic cells of chronic lymphocytic leukaemia. *Immunol Rev* **48**: 23-44.

González-Fernandez A, Gupta SK, Pannell R, Neuberger MS and Milstein C (1994). Somatic mutation of immunoglobulin λ chains: A segment of the major intron hypermutates as much as the complementarity-determining regions. *Proc Natl Acad Sci* USA **91:** 12614-618.

Graur D, Duret L and Gouy M (1996). Phylogenetic position of the order Lagomorpha (rabbits, hares and allies). *Nature* **379**: 333-35.

Greenberg AS, Avila D, Hughes M, Hughes A, McKinney EC and Flajnik MF (1995). A new antigen receptor gene family that undergoes rearrangements and extensive somatic diversification in sharks. *Nature* **374**: 168-73.

Gridley J (1912). The lagomorphs and independent order. *Science* **36:** 285-86.

Hamers R and Hamers-Casterman C (1965). Molecular localization of A chain allotypic specificities in rabbit IgG (7S gamma-globulin). *Mol Biol* **14**(1): 288-89.

Hamers R (1987). Allotypy of immunoglobulins. In Dubiski S (ed.). *The Rabbit in Contemporary Immunological Research*, Longman Scientific & Technical. pp 65-77.

Halanych KM and 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)**: 369-79.

Hardy C, Vigne JD, Casane D, Dennebouy N, Mounoulou JC and Monnerot M (1994). Origin of European Rabbit (*Oryctolagus cuniculus*) in a Mediterranean island: zooarchaeology and ancient DNA examination. *J Biol Evol* **7:** 217-26.

Heidmann O and Rougeon F (1983). Diversity in the rabbit immunoglobulin kappa chain variable regions is amplified by nucleotide deletions and insertions at the V-J junction. *Cell* **34(3)**: 767-77.

Hsu E and Steiner L (1992). Primary structure of immunoglobulins through evolution. *Curr Opin Struct Biology* **2**: 422-31.

Hsu E (1994). The variation in immunoglobulin heavy chain C regions in evolution. *Semin Immunol* **6**: 383-91.

Huchon D, Catzeflis F and Douzery E (1999). Molecular evolution of nuclear von Willebrand factor gene in mammals and the phylogeny of rodents. *Mol Biol Evol* **16:** 557-89.

Idusogie EE, Presta LG, Gazzano-Santoro H, Totpal K, Wong PY, Ultsh M, Meng YG and Mulkerrin MG (2000). Mapping of C1q binding sites on rituxan, a chimeric antibody with a human IgG1 Fc. *J Immunol* **164:** 4178-184.

Kabat EA, Wu TT, Perry HM, Gottesman KS, and Foeller C (1991). *Sequences of proteins of immunological interest*, 5th edn. Public health service, NIH, Besthesda, MD.

Kelus AS and Gell PG (1967). Immunoglobulin allotypes of experimental animals. *Prog Allergy* **11:** 141-84.

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): 255-59.

Knight KL and Crane MA (1994). Generating the antibody repertoire in rabbit. *Adv Immunol* **56:** 179-218.

Knight KL and Winstead CR (1997). B lymphocyte development in rabbit. *Int Rev Immunol* **15**: 129-63.

Kirschbaum T, Pourrajabi S, Zocher I, Schwendinger J, Heim V, Roschethaler F, Kirschbaum V, Zachau H-G (1998). The 3' part of the immunoglobulin k locus of the mouse. *Eur J Immunol* **28:** 1458-466.

Kulics J, Dekegel M, Naessens J, van der Loo W, Hammers R (1983). The quaternary Gs3 and Gs7 allotypes of the rabbit: generation of the determinants by interspecies molecular hybridization. *Mol Immunol* **20**(1): 101-11.

Kumar S and Hedges SB (1998). A molecular time scale for vertebrate evolution. *Nature* **392**: 920-23.

Landau NR, Schatz DG, Rosa M, and Baltimore D (1987). Increased frequency of N-region insertion in a murine pre-B cel line infected with a terminal deoxynucleotidyl transferase retroviral expression vector. *Mol Cell Biol* **7:** 3237-243.

Lanning DK, Knight KL (1997).Somatic hypermutation: mutations 3' of rabbit *VDJ* H-chain genes. *J Immunol* **59**(9)**:** 4403-407.

Lanning D, Zhu X, Zhai SK, Knight KL (2000a). Development of the antibody repertoire in rabbit: gut-associated lymphoid tissue, microbes, and selection. *Immunol Rev*, **175**:214-28.

Lanning D, Sethupathi P, Rhee KJ, Zhai SK and Knight KL (2000b). Intestinal microflora and diversification of the rabbit antibody repertoire. *J Immunol*, **165**(4): 2012-019.

Lefranc, M.-P. 2001. IMGT, the international ImMunoGeneTics database. *Nucleic Acids Research* **29**:207.

Liu F and Miyamoto M (1999). Phylogenetic assessment of molecular and morphological datafor eutherian mammals. *Systematic Biology* **48:** 54-64.

Lopez-Martinez N (1989). *Revision sistematica y biostratigrafica de los lagomorphos (Mammalia) del Terciario y Cuaternario de España.* Memorias del Museo Paleontologico de la Universidad de Zaragoza, nº3. Diputacion General de Aragon.

Lopez-Martinez N (1998). A look to the lagomorph fossil record. In *Abstracts of the Euro-American Mammal Congress*. (Ed.) S Reig. Universidade de Santiago de Compostela, Santiago de Compostela. pp 86.

Lucier MR, Thompson RE, Waire J, Lin AW, Osborne BA and Goldsby RA (1998). Multiple sites of VL diversification in cattle. *J Immunol* **161**: 5438-444.

Mage RG, Young-Cooper GO and Alexander C (1971). Genetic control of variable and constant regions of immunoglobulin heavy chain. *Nature-New Biology* **230**: 63.

Mage RG (1979). A new look at the biological and genetic significance of rabbit heavy chain allotypes. *Ann Immunol* **130**: 105-14.

Mage RG (1987). Molecular biology of rabbit immunoglobulin and T-cell receptor genes. In Dubiski S (ed.). *The Rabbit in Contemporary Immunological Research*, Longman Scientific & Technical. pp106-33.

Malcolm S, Barton P, Murphy C, Ferguson-Smith MA, Bentley DL and Rabbitts TH (1982). Localisation of human immunoglobulin kappa light chain variable region genes to the short arm of chromosome 2 by in situ hybridization. *Proc Natl Acad Sci* USA **79:** 4957-961.

Matsuda F, Ishii K, Bourvagnet P, Kuma K, Hayashida H, Miyata T and Honjo T (1998). The complete nucleotide sequence of the human immunoglobulin heavy chain variable region locus. *J Exp Med* **188**: 2151-162.

Meier JT and Lewis SM (1993). P nucleotides in *V(D)J* recombination: a fine-structure analysis. *Mol Cell Biol* **13:** 1078-092.

Meng J, Wyss A, Dawson M and Zhal R (1994). Primitive fossil rodent from inner Mongolia and its implications for mammalian phylogeney. *Nature* **370:** 134-36.

Monnerot M, Vigne JD, Biju-Duval C, Casane D, Callou C, Hardy C, Mougel F, Soriguer R, Dennebouy N amd Mounolou JC (1994). Rabbit and Man: Genetic and Historic Approach. *Genet Sel Evol* **26 (suppl 1):** 167-82.

Motoyama N, Okada H, Azuma T (1991). Somatic mutation in constant regions of mouse lambda 1 light chains. *Proc Natl Acad Sci* U S A **88**(18): 7933-937.

Mougel F, Mounolou JC and Monnerot M (1997). Nine polymorphic microsatellite loci in the rabbit, *Oryctolagus cuniculus*. *Anim Genet* **28**(1): 58-9.

Murphy W, Eizirik E, O'Brien SJ, Madsen O, Scally M, Douday CJ, Teeling E, Ryder OA, Stanhope MJ, de Jong WW and Springer MS (2001). Resolution of the Early Placental Mammal Radiation Using Bayesian Phylogenetics. *Science* **294:** 2348-351.

Mussmann R, Du Pasquier L and Hsu E (1996). Is Xenopus IgX an analog of IgA? *Eur J Immunol* **26**: 2823-830.

Nei M, Xu P and Glazko G (2001). Estimation of divergence times from multiprotein sequences for a few mammalian species and several distantly related organisms. *Proc Natl Acad Sci* U S A **98(5):** 2497-502.

Nguyen VK, Hamers R, Wyns L and Muyldermans S (2000). Camel heavy-chain antibodies: diverse germline V(H)H and specific mechanisms enlarge the antigen-binding repertoire. *EMBO J* 19(5): 921-30.

Novacek MJ (1992). Fossils, topologies, missing data and the higher level phylogeny of eutherian mammals. *Systematic Biology* **41:** 58-73.

Nuñez C, Nishimoto N, Gartland GL, Bilips LG, Burrows PD, Kubagawa H and Cooper MD (1996). B cells are generated throughout life in humans. *J Immunol* **156**: 866-72.

Oudin J (1956). L'"allotypie" de certains antigènes proteídique du sérum. *Comp. Rend. Acad. Sci. Paris* **242:** 2606-608.

Oudin J (1960). Allotypy of rabbit serum proteins. I. Immunochemical analysis leading to the individualization of seven main allotypes. *J Exp Med* **112:** 107-24.

Padlan EA (1994). Anatomy of the antibody molecule. *Mol Immunol* **31**: 169-217.

Padlan EA (1996). X-ray christallography of antibodies. *Adv Prot Chem* **49**: 57-133.

Pages MV (1980). Èssai de reconstitution de l'histoire du lapin de garenne en Europe. *Bul M Off Nat Chasse*, special number: 13-21.

Parng CL, Hansal S, Goldsby RA and Osborne BA (1996).Gene conversion contributes to Ig light chain diversity in cattle. *J Immunol* **157**(12): 5478-486.

Parodi U (1907). In Dubiski S (ed.). *The Rabbit in Contemporary Immunological Research.* Longman Scientific & Technical.

Pasteur L, Chamberland CE, Roux PPE, Thuillier L (1881). In Dubiski S (ed.). *The Rabbit in Contemporary Immunological Research.* Longman Scientific & Technical.

Prahl JW, Mandy WJ and Todd CW (1969). The molecular determinants of the A11 and A12 allotypic specificities in rabbit immunoglobulin. *Biochem* **8**: 12.

Queney G, Ferrand N, Weiss S, Mougel F and Monnerot M (2001). Stationary distributions of microsatellite loci between divergent population groups of the European Rabbit (*Oryctolagus cuniculus*). *Mol Biol Evol* **18**(12): 2169-178.

Queney G, Vachot AM, Brun JM, Dennebouy N, Mulsant P, Monnerot M (2002).Different levels of human intervention in domestic rabbits: effects on genetic diversity. *J Hered* **93**(3): 205-09.

Ramsden DA, van Gent DC and Gellert M (1997). Specificity in V(D)J recombination: new lessons from biochemistry and genetics. *Curr Opin Immunol* **9:** 114-20.

Reynaud CA, Anquez V, Dahan A and Weill JC (1985). A single rearrangement event generates most of the chicken immunoglobulin light chain diversity. *Cell* **40**(2): 283-91.

Reynaud CA, Mackay CR, Muller RG and Weill JC (1991). Somatic generation of diversity in a mammalian primary lymphoid organ: the sheep ileal Peyer's patches. *Cell* **64**(5): 995-1005.

Reynaud CA, Garcia C, Hein WR and Weill JC (1995). Hypermutation generating the sheep immunoglobulin repertoire is an antigen-independent process. *Cell*, **80**(1): 115-25.

Reynolds JD and Morris B (1983). The evolution and involution of peyer's patches in fetal and post-natal sheep. *Eur J Immunol* **13**: 627-35.

Reynolds JD (1987). Mitotic rate maturation in the peyer's patches of fetal sheep and in the bursa of Fabricius of the chicken embryo. *Eur J Immunol* **17**: 503-07.

Richardson BJ, Rogers PM and Hewitt GM (1980). Ecological genetics of the wild rabbit in Australia. II. Protein variation in British, French and Australian rabbits and the geographical distribution of the variation in Australia. *Aust J Biol Sci* **33:** 371-83.

Robinson R (1984). Rabbit. In: *Evolution of domesticated animals* (ed. IL Mason), Longman.

Sanchez P, Nadel B, and Cazenave P-A (1991). Vλ-Jλ rearrangements are restricted within a V-J-C recombination unit in the mouse. *Eur J Immunol* **21**: 907-11.

Sehgal D, Johnson G, Wu TT and Mage RG (1999). Generation of the primary antibody repertoire in rabbits: expression of diverse set of igk-V genes may compensate for limited combinatorial diversity at the heavy chain locus. *Immunogenetics* **50:** 31-42.

Sehgal D, Obiakor H and Mage RG (2002). Distinct clonal Ig diversification patterns in young appendix compared to antigen-specific splenic clones. *J Immunol*, **168**(11): 5424-433.

Severinson E, Bergstedtlindqvist S, van der Loo W and Fernandez C (1982). Characterization of the IGG response induced by polyclonal B-cell activators. *Immunological Reviews* **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): 4014-018.

Spiegelberg HL (1977). The structure and biology of human IgD. *Immunol Rev* **37:** 3-23.

Springer MS, Murphy WJ, Eizirik E and O'Brien SJ (2003). Placental mammal diversification and the Cretaceous-Tertiary boundary. *Proc Natl Acad Sci* U S A **100**(3): 1056-061.

Stiernholm NBJ, Kuzniar B and Berinstein N (1994). Identifications of a new human Vλ gene family- VλX. *J Immunol* **152**: 4969-975.

Storb U, Arp B and Wilson R (1981). The switch region associated with immunoglobulin C-mu genes is DNAse-I hypersensitive in lymphocytes. *Nature* **294**(5836): 90-92.

Sun J and Butler JE (1996). Molecular characterization of VDJ transcripts from a newborn piglet. *Immunology* **88:** 331-39.

Su C and 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): 9710-715.

Surridge AK, Bell DJ, Rico C and Hewitt GM (1997). Polymorphic microsatellite loci in the European Rabbit (*Oryctolagus cuniculus*) are also amplified in other lagomorph species. *Anim Genet* **28**(4): 302-05.

Tao MH, Smith TR and Morrison SL (1993). Structural features of human immunoglobulin G that determine isotype-specific differences in complement activation. *J Exp Med* **178:** 661-67.

Todd CW (1963). Allotypy in rabbit 19S protein. *Bioch. Biophys. Res. Comm.* **11**: 170-75.

Tonegawa S (1983). Somatic generation of antibody diversity. *Nature* **302**: 575-81.

van der Loo W, Gronowicz ES, Strober S, Herzenberg LA (1979).Cell-differentiation in the presence of cythochalasin-B – Studies on the switch to IGG secretion after polyclonal-B cell activation. *Journal of Immunology* **122**(4): 1203-208.

van der Loo W and Arthur CH (1986). Etude des associations gametiques entre alleles de quatre loci d'immunoglobuline dans de populations naturelles de Lapin de Garenne. In Legay JM (ed.). *Actes du Colloque Biologie des Populations, Lyon*. University Claude Bernard. pp 374-82.

van der Loo W (1987). Studies on the adaptive significance of the immunoglobulin polymorphisms (Ig allotypes) in wild rabbits. In Dubiski S (ed.). *The Rabbit in Contemporary Immunological Research*. Longman Scientific & Technical. pp 164-90.

van der Loo W, Arthur CP, Richardson BJ, Wallage-Drees M and 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): 3075-079.

van der Loo W, Ferrand N and 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): 789-99.

van der Loo W and Verdoodt B (1992). Patterns of interallelic divergence at the rabbit *b*-locus of the immunoglobulin light chain constant region are in agreement with population genetical evidence for overdominant selection. *Genetics* **132**(4): 1105-117.

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): 171-87.

van der Loo W, Bousses P, Arthur CP and Chapuis JL (1996). Compensatory aspects of allele diversity at immunoglobulin loci: gene correlations in rabbit populations devoid of light chain diversity (*Oryctolagus cuniculus* L.; Kerguelen Islands). *Genetics* **144**(3): 1181-194.

van der Loo W, Mougel F, Sanchez MS, Bouton C, Castien E, Fonseca A, Ferrand N, Soriguer R and 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): 629-43.

van Haeringen WA, den Bieman M, van Zutphen LF and van Lith HA (1997). Polymorphic microsatellite DNA markers in the rabbit (*Oryctolagus cuniculus*). *J Exp Anim Sci* **38**(2): 49-57.

Vigne JD (1988). Donnés preliminaries sur l'histoire du peuplement mammalien de l'ilot de Zembra (Tunisie). *Mammalia* **52:** 567-74.

Warr GW (1995). The immunoglobulins gene of fish. *Dev Comp Immunol* **19**: 1-12.

Wasserman A, Neisser A and Bruck C (1906). Eine serodiagnostische Reaktion bei Syphilis. In Dubiski S (ed.). *The Rabbit in Contemporary Immunological Research.* Longman Scientific & Technical.

Webster DM, Pedersen J, Stauton D, Jones A and Rees AR (1994). Antibody-combining sites. Extending the natural limits. *Appl Biochem Biotechnol* **47:** 119-32.

Weinstein PD, Anderson AO and Mage RG (1994). Rabbit IgH sequences in appendix germinal centers: VH diversification by gene conversion-like and hypermutation mechanisms. *Immunity* **1:** 647-59.

# CHAPTER 2

## Genetic variation at the constant region of the IgG antibody in leporids

# Introduction

## Rabbits possess only one IgG class

IgG antibodies are synthesised at a high level and constitute up to 75% of serum immunoglobulins. Four subclasses are produced in the human (IgG1, IgG2, IgG3, IgG4), and in the mouse (IgG1, IgG2a, IgG2b, IgG3). The choice of an IgG subclass is not random and different antigenic stimuli induce the production of different IgG subclasses. For example, in the mouse, protein antigens stimulate predominantly IgG1 responses while carbohydrate antigens induce substantial IgG3 responses (Perlmutter *et al.*, 1978; Slack *et al.*, 1980; Scott and Fleischman, 1982). In addition, the selection of an IgG subclass is strongly influenced by the numerous cytokines produced by Th1 and Th2 cells (helper T cells). It is therefore surprising that in the rabbit no subclasses of IgG antibodies have been identified.

## Indirect evidence for adaptive variation at the rabbit IgG CH2 domain (*e*-locus)

*Structural evidence*

Epitopes at the IgG CH2-CH3 interface play, in conjunction with the Fc receptors, a crucial role in the processing of immunoglobulins including their transplacental transport (Johanson *et al.* 1981, Ghetie and Ward 1997). A major antibody function is the recruitment of the complement pathway, which requires binding of the CH2 domain to C1q, a constituent of the first complement. Mapping by site-directed mutagenesis showed that in human, several amino acids of the CH2 region participate in complement binding (Idusogie *et al.*, 2001). The CH2-CH3 interface is also the target of pathogens (Deisenhofer, 1981; Kim *et al.*, 1999), and different lines of evidence suggest that isotype and allotype variation at this region can be adaptive (Dubiski and Good, 1972; Granoff *et al.*, 1984).

The rabbit *IgGCH2* or *e* locus is localised on the chromosome 17 (Medrano and Duttrillaux, 1984) and exhibits two serological markers, *e14* and *e15* (Hamers and Hamers-Casterman, 1965; Mage, 1981). These alleles are inherited as codominant Mendelian alleles

and are correlated with an Alanine-Threonine replacement at amino acid position 309 (79 in the IMGT numbering) (Appella *et al.*, 1971). This position is adjacent to the canonical histidine at 310, which is known to affect interaction with a variety of Fc receptors (West and Bjorkman, 2000). X-ray crystallography reveals direct contact between the side chains of *IgG*-Leu309 of the rat Fc receptor (Burmeister *et al.*, 1994). The *e15* allotype was found in species of genera *Oryctolagus*, *Lepus*, *Sylvilagus, Romerolagus* and *Ochotona* (Cazenave *et al.* 1977; van der Loo and Hamers-Casterman, 1977). A number of species seem to be fixed for this allele, while the *e14-e15* polymorphism was observed in the Mexican volcano rabbit (*Romerolagus diazi*) (van der Loo and Hamers-Casterman, 1979, 1981) and in cottontail rabbit (*Sylvilagus floridanus*) (Teherani and Mandy, 1976; Cazenave *et al.*, 1977). Both alleles were absent in mountain hare (*Lepus timidus*).

*Population genetic evidence*

In the most recent distribution range of the species, all wild and feral rabbit populations studied were polymorphic, and allele frequencies of *e14* were very similar among major geographical regions ($p_{e14} \approx 20\%$, van der Loo, 1993; van der Loo *et al.*, 1996). However, all Iberian wild rabbit populations studied so far showed the fixation of *e15* (N. Ferrand and W. van der Loo, unpublished data cited in van der Loo and Arthur (1985), and van der Loo (1987)). The absence of the polymorphism in the original distribution area is puzzling. Indeed, the *e14* allele is among the rare genetic variants that are not present in Iberian populations. Strong linkage disequilibria (LD) were observed between alleles of *e* locus and other H chain loci; i.e. *d* locus (H chain Hinge region) and *a* locus ($V_H$ genes). The allotype *d11-e14* has not been reported so far (Mage *et al.*, 1977; van der Loo, 1987), while the *a-e* haplotype *a3-e14* has not been observed in most domestic breeds (see Mage *et al.*, 1977). The *a-e* and *d-e* LD can be explained by the close linkage between these H chain loci and a rather recent origin of the *e14* variant (<5000 years). More surprising was the observation of a highly significant LD between the *e* locus and alleles of the L-chain constant region (*b* locus), because both loci are on different chromosomes (van der Loo *et al* 1987, 1996). This LD was observed within and between populations from all areas where the interacting loci were polymorphic (Australia, Great Britain, Continental Europe North of the Loire river). The observation of LD (or non-random association) between alleles of independently segregating loci in natural populations is considered one of the most reliable indicators of selection (Franklin and Lewontin, 1970; Lewontin, 1974), and, as a matter of fact, had never been documented for protein loci in mammalian populations before.

The population genetic data suggest therefore that the *b* and *e* locus polymorphisms are maintained by epistatic selection pressure. van der Loo and co-workers have presented population genetic evidences for the hypothesis that the extremely large interallelic distances among *b* locus alleles is due to overdominant selection (some 10% heterozygosity excess; van der Loo and Verdoodt 1992; van der Loo 1993). van der Loo (1996) observed furthermore that the heterozygosity excess at one Ig locus tends to be higher if heterozygosity at another Ig locus is low (higher order LD), and introduced the concept of "compensatory overdominance". This author speculated that the emergence of *e14-e15* polymorphism in the recent species range might have to do with the sharp reduction in allele diversity at the *b* locus observed in this region (of more than 16 *b* locus alleles occurring in the original range, only four are present in the recent range).

We have conducted an evolutionary study at the molecular level, which confirms the existing evidence that the *e* locus polymorphism is maintained by selection. Previous molecular studies have been limited to the relationships between serology and protein variation at *IgGCH2* domain (Appella *et al.* 1971; Aggarwal and Mandy, 1976; Cazenave *et al.*, 1977; Teherani *et al.* 1979, 1982). Until the present work, no nucleotide data on lagomorph *IgGCH2* were available, except for three sequences of domestic rabbit (Martens *et al.*, 1982; Bernstein *et al.*, 1983; Martens *et al.*, 1984).

# Article 1

P.J. Esteves, P.C. Alves, N. Ferrand and W. van der Loo (2002).

Restriction fragment alleles of rabbit IGHG genes with reference to the rabbit *IgGCH2* or *e* locus polymorphism. *Animal Genetics* **33**: 309-311.

SHORT COMMUNICATION

# Restriction fragment alleles of the rabbit *IGHG* genes with reference to the rabbit *IGHGCH2* or *e* locus polymorphism

**P. J. Esteves*†, P. C. Alves*, N. Ferrand* and W. van der Loo†**

*Centro de Estudos de Ciência Animal (CECA), ICETA-UP, Campus Agrário de Vairão, Vairão, Portugal, and Departamento de Zoologia-Antroplogia Faculdade de Ciências da Universidade do Porto, Porto, Portugal. †Institute of Molecular Biology and Biotechnology, Vrije Universiteit Brussel, St Genesius-Rode, Belgium

**Summary**

Among domesticated mammals, rabbit (*Oryctolagus cuniculus*) is the only species possessing not more than one subclass of immunoglobulin (IgG) antibodies. The rabbit *IGHGCH2* or *e* locus presents two serologically defined alleles, the *e14* and *e15* allotypes, which are correlated with amino acid variation at the IgG *CH2–CH3* interface. Genetic studies, while revealing the adaptive value of this polymorphism, have relied so far entirely upon allo-antisera. Here we show how these alleles can be distinguished by polymerase chain reaction–restriction fragment length polymorphism (PCR–RFLP) methods. The proposed PCR–RFLP approach allows the monitoring of *IGHG* locus diversity in rabbit.

Abbreviations: IgG, gamma class immunoglobulin; IGHG, IgG immunoglobulin heavy chain gamma constant region; IGHGCH2, second domain of IGHG; bp, base pair; aa, amino acid; PCR, polymerase chain reaction, RFLP, restriction fragment length polymorphism

**Keywords** allotypes, antibody, *e* locus, *IGHGCH2* locus, PCR–RFLP, polymorphism, rabbit.

In conjunction with FcRn receptors, epitopes at the IgG *CH2–CH3* interface play a crucial role in the processing of IgGs and their transplacental transport (Johanson *et al.* 1981; Gethie & Ward 1997). They are also the target of pathogens (Kim *et al.* 1999), and different lines of evidence suggest that genetic variation in this region is adaptive (van der Loo *et al.* 1987, 1996). Unlike other mammals, rabbit possesses only one subclass of IgG antibodies. Two serologically defined alleles, the *e14* and *e15* allotypes, were found to be correlated with an alanine–threonine exchange at amino acid (aa) position 309 (Appella *et al.* 1971; Mage 1981; for IMGT numbering see http://imgt.cines.fr). Position 309 is adjacent to the canonical histidine at 310, which is known to affect interaction with a variety of Fc receptors (West & Bjorkman 2000).

Address for correspondence

Wessel van der Loo, Institute of Molecular Biology and Biotechnology, Vrije Universiteit, Brussels, Paandestr. 65, 1640 St Genesius-Rode, Belgium. E-mail: wvdloo@ben.vub.ac.be

*Accepted for publication 22 February 2002*

Each of more than 6000 specimen of European rabbit showed either the *e14* or the *e15* allotype, or both (van der Loo 1993 unpublished data). In the more recent distribution range of the rabbit species (Continental Europe North of the Loire river, Great Britain, Australia, Kerguelen Islands), all wild and feral rabbit populations were polymorphic at the *IGHGCH2* locus (van der Loo *et al.* 1996). Allele frequencies were very similar among regions ($p_{e14} \approx 20\%$, van der Loo 1993). However, in a survey of gene diversity in seven domestic breeds, the mean *e14* frequency was below 5% (G. Bolet *et al.* unpublished data). Although highly reliable, serological methods depend upon good allo-antisera, which are difficult to obtain. We show here how the *e15* allotype, or its absence, can be detected by restriction fragment length polymorphism (RFLP) analysis of polymerase chain reaction (PCR) products. We also present for the first time the *CH2* nucleotide sequence as obtained with genomic DNA of *e14* rabbits.

Serological analyses was performed by immunodiffusion, in the presence of *e14* or *e15* positive control sera as previously described (van der Loo 1993). Genomic DNA was

© 2002 International Society for Animal Genetics, *Animal Genetics*, **33**, 309–311

obtained by Qiagen extraction of tissues (including blood clots) of specimen that were also studied serologically. The RFLP analysis was performed on PCR products. These were obtained using oligonucleotide primers, constructed according to the sequence of rabbit *IGHGCH2* gene (EMBL databank accession number L29172). The primers were F3: (5'GTGAGTCCCATCTAGCCTCAC3') and R2: (5'GGGCCCAT GGTGTAGACCTT3'). The primers F3 and R2 are homologous to positions (705–724) and (1256–1236) of the L29172 sequence, respectively. The PCR amplification and DNA sequencing protocols were routine, and can be obtained upon request.

The PCR products were 553–557 base pair (bp) in length. They were digested with *Tha*I prior to electrophoresis. All PCR products obtained (19 rabbits studied) showed a *Tha*I site (i.e. GCGC) that comprises a shared Arg301(CGC) codon (T1 in Figs 1 and 2). A second shared *Tha*I site (T3) was found in the 3' intron region, 130 bp downstream of T1. A third *Tha*I site (T2), which cuts the 130 bp fragment into 106 and 24 bp fragments, was exclusively found in PCR products obtained with *e15* rabbits (15 sequences). This site depends upon the allotype-specific $Ala_{309}$(GCG) codon and is absent in *e14* rabbits, which instead feature $Thr_{309}$(ACG). As expected, PCR products of heterozygous *e14/e15* rabbits showed partial digestion of the *Tha*I 130 bp fragment and had either G or A nucleotides at first position of codon 309 (cf. Table 1).

This shows that, although the rabbit allo-antisera remain the safest shortcut for detecting either $Ala_{309}$ or $Thr_{309}$ at the *CH2–CH3* interface of IgG, PCR–RFLP can be helpful in population studies, allowing the monitoring of the *IGHGCH2* diversity in breeding programmes.

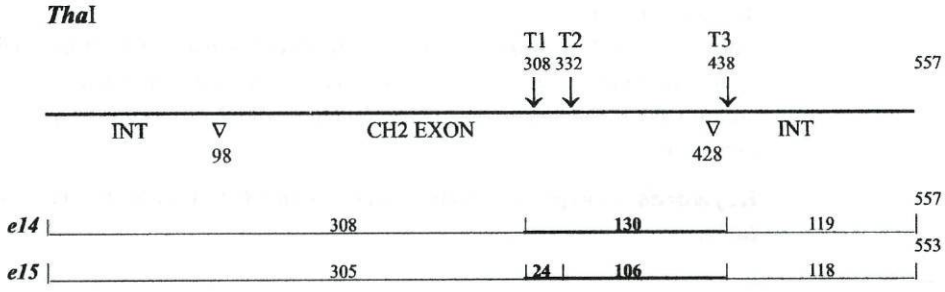



**Figure 1** *Tha*I restriction patterns of the PCR products of the *IGHGCH2* exon. Products include the 5' primer *F3* (20 bp), the 3'primer R2 (20 bp), the entire *CH2* exon (330 bp) and parts of the flanking introns. ∇ indicate the exon–intron boundaries; ↓ indicate the potential restriction sites T1, T2, T3.

```
                    |Int->                                                
```

**Table 1** Correspondence of rabbit *IGHGCH2* allotypes with *Tha*I restriction sites and DNA sequences of PCR fragments.

| Serology | | PCR–RFLP | | | DNA sequencing | | |
|---|---|---|---|---|---|---|---|
| *e14* | *e15* | T1 | T2 | T3 | $G(CGC)_{301}$ | $(GCG)_{309}C$ | GCGC |
| + | – | + | – | + | + | $(ACG)_{309}C$ | + |
| + | + | + | +/– | + | + | $(^{G}/_{A}CG)_{309}C$ | + |
| – | + | + | + | + | + | $(GCG)_{309}C$ | + |

+: Positive or present; –: negative or absent; ±: partial digestion; $^{G}/_{A}$: G and A nucleotides at this site. The *e15* allotype is clearly correlated with the presence of *Tha*I restriction site T2, implying the presence of $Ala_{309}$.

## Acknowledgements

This work was supported by Praxis XXI/BD/16207/98 of the FCT (Foundation for Science and Technology – Portugal) and by Krediet 1.5579.98-FWOKN35 of the FWO-Vlaanderen (Fonds for Scientific Research of the Dutchspeaking community of Belgium).

## References

Appella E., Chersi A., Mage R. & Dubiski S. (1971) Structural basis of the A14 and A15 allotypic specificities in rabbit IgG. *Proceedings of the National Academy Science USA* **68**, 1341–9.

Bernstein K.E., Alexander C.B. & Mage R.G. (1983) Nucleotide sequence of a rabbit IgG heavy chain from recombinant F-I haplotype. *Immunogenetics* **18**, 387–97.

Gethie V. & Ward E.S. (1997) FcRn: the MHC class I-related receptor that is more than an IgG transporter. *Immunology Today* **18**, 592–8.

Johanson R.A., Shaw A.R. & Schlamowitz M. (1981) Evidence that the CH2 domain of IgG contains the recognition unit for binding the fetal rabbit yolk sac membrane receptor. *Journal of Immunology* **126**, 194–9.

Kim J.K., Firan M., Radu C.G., Kim C.H., Ghetie V. & Ward E.S. (1999) Mapping the site on human IgG for binding of the MHC class I-related receptor, FcRn. *European Journal of Immunology* **29**, 2819–25.

van der Loo (1993) Variance analysis of immunoglobulin alleles in natural populations of rabbit (*Oryctolagus cuniculus*). *Genetics* **134**, 171–87.

van der Loo W., Arthur C.P., Drees-Wallage M. & Richardson B. (1987) Nonrandom allele associations between unlinked protein loci: are the polymorphisms of the immunoglobulin constant regions adaptive? *Proceedings of the National Academy Science USA* **84**, 3075–9.

van der Loo W., Boussès P., Arthur C.P. & Chapuis J.-L. (1996) Compensatory aspects of allele diversity at immunoglobulin loci: gene correlations in rabbit populations devoid of light chain diversity (*Oryctolagus cuniculus* L., Kerguelen Islands). *Genetics* **144**, 1181–94.

Mage R. (1981) The phenotypic expression of rabbit immunoglobulins: a model of complex regulated gene expression and cellular differentiation. *Contemporary Topics Molecular Immunology* **8**, 89–112.

Martens C.L., Currier S.J. & Knight K.L. (1984) Molecular genetic analysis of genes encoding the heavy chain of rabbit IgG. *Journal of Immunology* **133**, 1022–7.

West A.P. & Bjorkman P.J. (2000) Crystal structure and immunoglobulin G binding properties of the human major histocompatibility complex-related Fc receptor. *Biochemistry* **39**, 9698–708.

© 2002 International Society for Animal Genetics, *Animal Genetics*, **33**, 309–311

## Genetic diversity at the *IgGCH2* region in *Oryctolagus cuniculus*

Until this work only three sequences of rabbit CH2 domain were available in Genbank. These sequences were obtained from rabbits serologically determined for the genetic markers *d* and *e* locus as *d12-e15* ("*e15*" in Figure 2.1), *d11-e15* ("*e15A*") and *d12-e14* ("*e14*"). Genbank accession numbers are L29172 (Martens *et al.*, 1984), M16426 (Martens *et al.*, 1982), and K00752 (Bernstein *et al.*, 1983), respectively. We compared the published rabbit CH2 sequences to those obtained in this study and with other leporid sequences. The phylogenetic tree shown in Figure 2.1 was constructed by applying the Neighbor-Joining (Saitou and Nei, 1987) and the p-distance methods provided in the software program MEGA 2 (Kumar *et al.*, 2001) to the nucleotide sequences listed in Table 2.1.

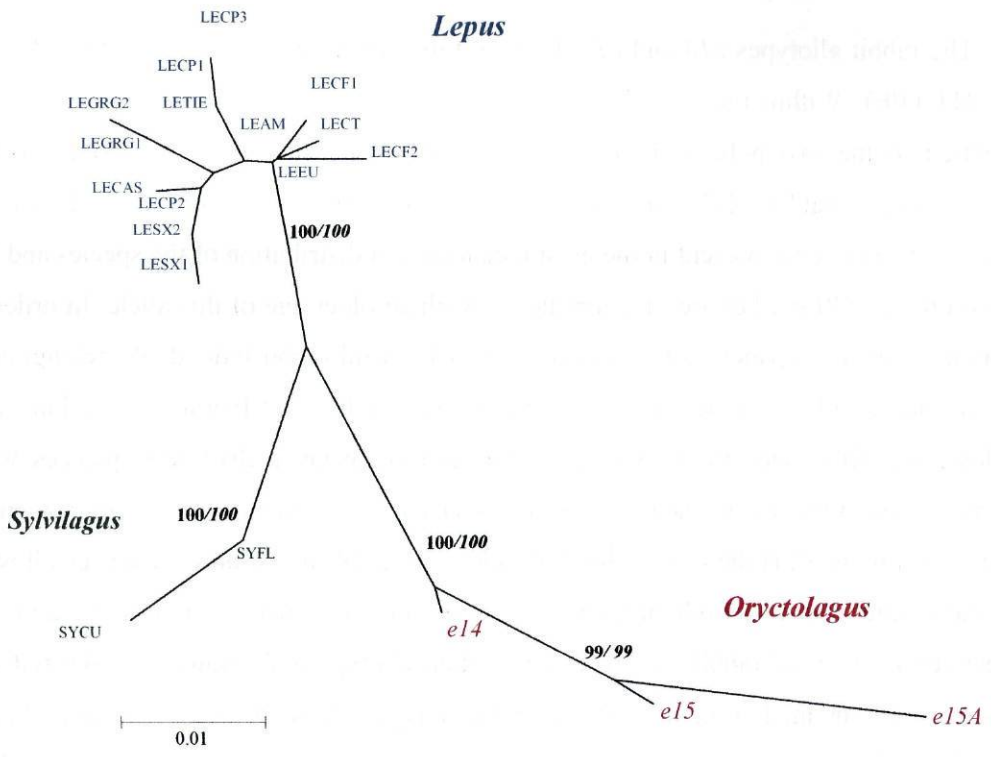



**Figure 2.1**: NJ tree of *IgGCH2* exon of three leporid genera using nucleotide sequences The bootstrap (italic) and the confidence probability values are indicated. Sample codes are shown in Table 2.1.

**Table 2.1**
Sequences obtained for *IgGCH2* in *Lepus* and *Sylvilagus* species

| Species | Sample code | localisation | Genbank (accession number) |
|---|---|---|---|
| ***Lepus*** | | | |
| *L. granatensis* | LEGR1 | Iberian Peninsula | AJ295217 |
| *L. granatensis* | LEGR2 | Iberian Peninsula | |
| *L. timidus* | LETI | Scotland | AJ295216 |
| *L. capensis* | LECP1 | Morocco | AJ295218 |
| *L. capensis* | LECP2 | Morocco | |
| *L. americanus* | LEAM | North America | |
| *L. californicus* | LECF | Mexico | AJ295219 |
| *L. castroviejoi* | LECS | Iberian Peninsula | |
| *L. callotis* | LECT | Mexico | AJ295220 |
| *L. europaeus* | LEEU | France | AJ295221 |
| *L. saxatilis* | LESX1 | South Africa | AJ295222 |
| *L. saxatilis* | LESX2 | South Africa | |
| ***Sylvilagus*** | | | |
| *S. floridanus* | SYFL | North America | |
| *S. cunicularis* | SYCU | Mexico | AJ295223 |

The rabbit allotypes *e14* and *e15* differ in three amino acid positions (268 Q/E, 309 T/A, 311 Q/E). Within the *e15* allotype only synonymous substitutions were observed. Compared to the two published *e15* sequences, the *e14* sequence is the closest to the common node of rabbit CH2, which would suggest that this is the oldest allele. However, the fact that *e14* is only present in the most recent area of distribution of the species and the LD with the *a3 VH* marker are in contradiction with an older age of this allele. In order to understand this discrepancy, we study this marker for wild rabbit individuals belonging to both subspecies (*O .c. algirus* and *O. c. cuniculus*) from different Iberian populations, and for domestic rabbit specimens. A total of 30 *IgGCH2* gene nucleotide sequences were determined and their amino acid sequences inferred. The comparison of the sequences obtained in this work (rabbit and other leporids) with published rabbit sequences allowed the establishment of nucleotide and amino acid consensus sequences (Figure 2.2 and 2.3, respectively). The wild rabbits, as expected by their allotype *e15*, showed Ala309 (GCG). All sequences obtained from domestic rabbit (allotypes *d11-e15*, *d12-e15* and *d12-e14*) were identical, except in the position 309 (data not shown). The study of wild rabbit populations revealed 6 new polymorphic positions (nucleotide positions: 23 C/T, 60 C/T, 78 C/T, 167 C/A, 178 C/T; 204 G/A) of which two were associated with amino acid substitutions (amino acid positions: 240 S/F, and 282 A/D) (see Figures 2.2 and 2.3).

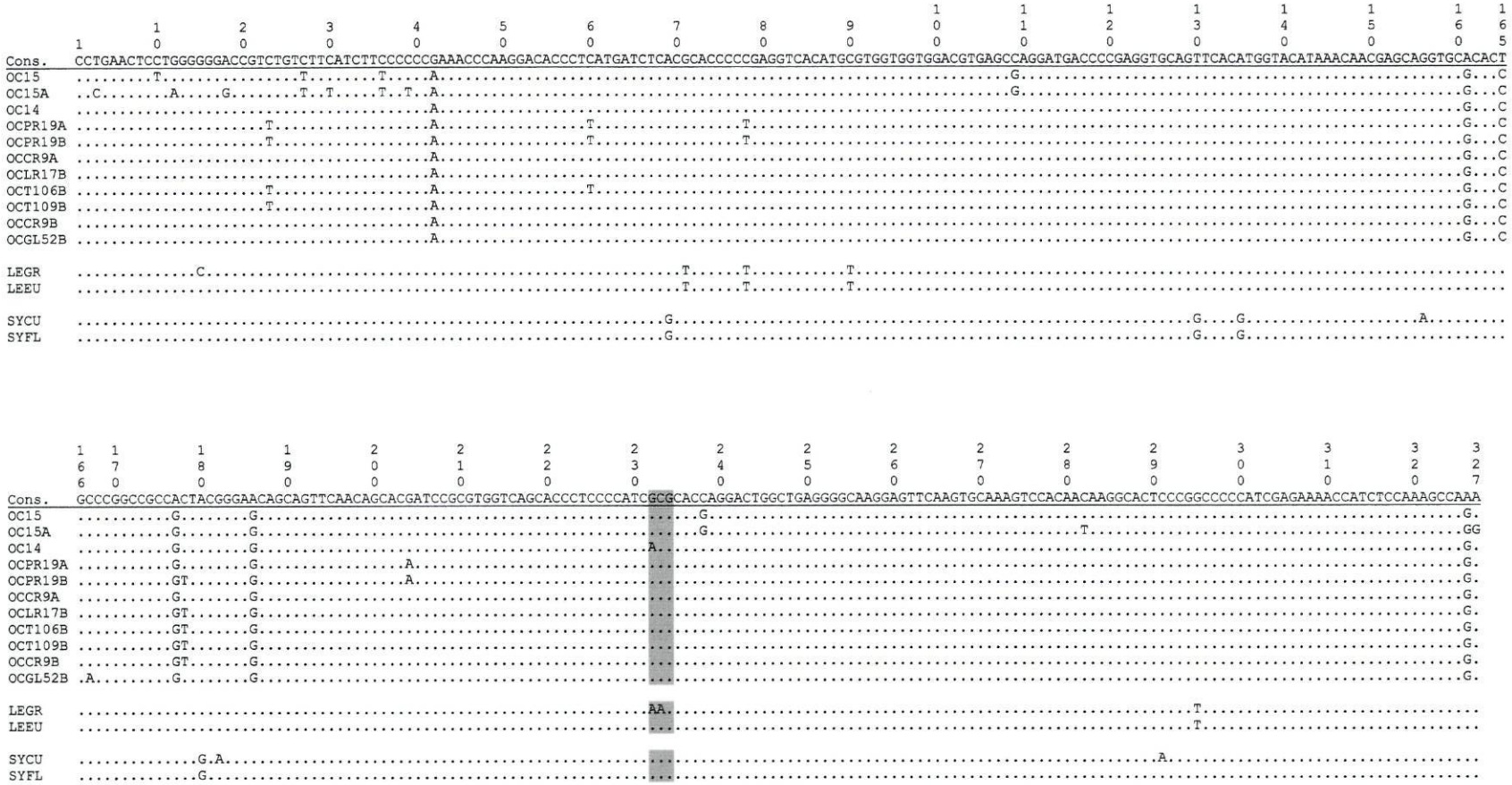
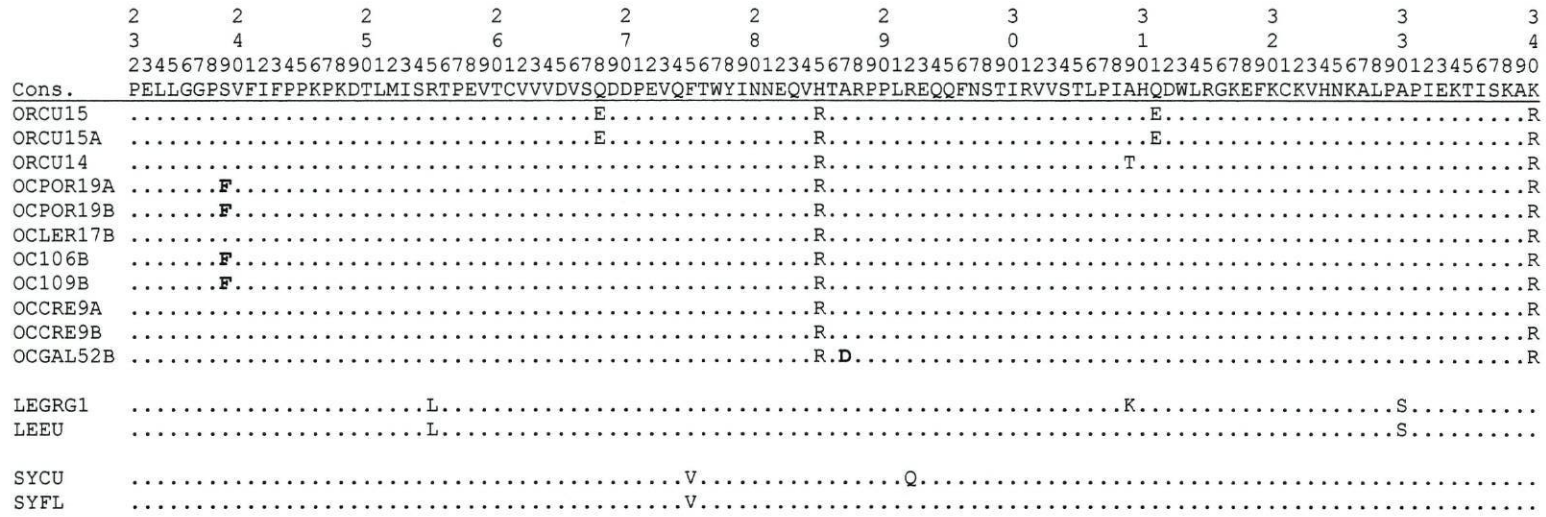



**Figure 2.2.** DNA sequences of *IgGCH2* obtained in wild rabbits from Iberian Peninsula. The comparison of these sequences with those obtained in *Lepus* and *Sylvilagus* allowed the establishment of a consensus sequence (Cons.). The codon that encodes position H309 is grey-shaded.



```
          2       2       2       2       2       2       2       3       3       3       3       3
          3       4       5       6       7       8       9       0       1       2       3       4
          2345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
Cons.     PELLGGPSVFIFPPKPKDTLMISRTPEVTCVVVDVSQDDPEVQFTWYINNEQVHTARPPLREQQFNSTIRVVSTLPIAHQDWLRGKEFKCKVHNKALPAPIEKTISKAK
ORCU15    ....................................E...............R.........................E.............................R
ORCU15A   ....................................E...............R.........................E.............................R
ORCU14    ..........................................................R.....................T.............................R
OCPOR19A  .......F..................................................R.....................................................R
OCPOR19B  .......F..................................................R.....................................................R
OCLER17B  ..........................................................R.....................................................R
OC106B    .......F..................................................R.....................................................R
OC109B    .......F..................................................R.....................................................R
OCCRE9A   ..........................................................R.....................................................R
OCCRE9B   ..........................................................R.....................................................R
OCGAL52B  ..........................................................R.D...................................................R

LEGRG1    ......................L.....................................................K.....................S..........
LEEU      ......................L.....................................................................S..........

SYCU      ........................................V.................Q.....................................
SYFL      ........................................V.....................................................
```

**Figure 2.3.** Amino acid sequences as inferred from the nucleotide sequences displayed in figure 2.2. The consensus sequence is represented. Dots indicate identity with the consensus sequence. Bold indicates the new amino acids detected in wild rabbit specimens.

Compared to the published ORCU*e15* and ORCU*e15a* sequences, the *e15* sequences obtained in the Iberian rabbits showed more similarity to the domestic *e14* sequence. This suggests that *e14* derived from an *e15* lineage present in the Iberian Peninsula. The larger diversity among *e15* sequences could indeed also be explained by the much longer history of this lineage. However, the highly divergent sequences ORCU*e15* and ORCU*e15a*, which were obtained in a laboratory breed, are surprising. It is possible that they represent lineages of *e15* not yet observed in the Iberian Peninsula. Another possibility is that mutant recruitment rate increased in that particular domestic breed. Indeed it has been suggested that positive selection may cause increased rates of synonymous substitution in gene regions that are closely linked to sites under selection (e.g Kreitman and Hudson, 1991). However, the possibility of a cloning mistake cannot be discarded. More data are needed to evaluate these different possibilities.

## The phylogeny of *IgGCH2*

The *IgGCH2* gene phylogram was constructed by the addition of all published *IgCH2* sequences to our data. The accession numbers for all sequences used are shown in Table 2.2. The NJ tree of nucleotide sequences (Figure 2.4) was obtained using the p-distance method in the program software MEGA 2 (Kumar *et al.*, 2001). Five major clearly separated clusters were obtained: 1) the lagomorph clade, which includes *Oryctolagus* (rabbit), *Lepus* (hares) and *Sylvilagus* (cottontail); 2) the primate clade, that encompasses *Homo sapiens* (human), *Macaca* sp. (macaque) and *Pan troglodytes* (chimpanzee); 3) the ferungulate clade, composed by artiodactyls (*Bos taurus* (cattle), *Sus scrofa* (swine), *Ovis aries* (sheep), *Camelus dromedarius* (camel), *Llama glama* (Llama)); carnivores (*Canis familiaris* (dog), *Felis catus* (cat) and *Mustela vison* (vison)) and perissodactyls (*Equus caballus* (horse)); 4) the rodent clade with *Mus musculus* (mouse), *Rattus norvegicus* and *Rattus rattus* (rat); and 5) the outgroup, marsupials: *Monodelphis* (opossum) and *Trichosurus* (possum). The lagomorph group is composed of three well-separated clades: a) *Oryctolagus*; b) *Sylvilagus*; c) *Lepus*.

**Table 2.2**

Accession numbers for each CH2 sequence used. The sample code and the IgG class are indicated.

| species | IgG subclass | Sample code | Accession number |
|---|---|---|---|
| ***Pan troglodytes*** | 1 | PT1 | X61311 |
| ***Canis familiaris*** | 1 | CF1 | AF354264 |
| | 2 | CF2 | AF354265 |
| | 3 | CF3 | AF354266 |
| | 4 | CF4 | AF354267 |
| ***Rattus norvegicus*** | 1 | RANV1 | M28670 |
| | 2A | RANV2A | M28669 |
| | 2B | RANV2B | M28671 |
| ***Rattus rattus*** | 2C | RARA2C | X07189 |
| ***Homo sapiens*** | 1*1 | HS1 1 | J00228 |
| | 1*2 | HS1 2 | Z17370 |
| | 2*1 | HS2 1 | J00230 |
| | 2*2 | HS2 2 | AJ250170 |
| | 3*1 | HS3 1 | M12958 |
| | 3*9 | HS3 9 | AJ390242 |
| | 3*12 | HS3 12 | AJ390252 |
| | 3*14 | HS3 14 | AJ390254 |
| | 3*17 | HS3 17 | AJ390262 |
| | 3*19 | HS3 19 | AJ390276 |
| | 4*1 | HS4 1 | K01316 |
| | 4*2 | HS4 2 | AJ001563 |
| ***Equus caballus*** | 1*1 | EC1 1 | AJ302055 |
| | 1*2 | EC1 2 | AJ300675 |
| | 2 | EC2 | AJ302056 |
| | 3 | EC3 | AJ312379 |
| | 4 | EC4 | AJ302058 |
| | 5 | EC5 | AJ312380 |
| | 6 | EC6 | AJ312381 |
| ***Mus musculus*** | 1*1 | MM1 1 | J00453 |
| | 1*2 | MM1 2 | L35252 |
| | 2A*1 | MM2A 1 | V00825 |
| | 2A*2 | MM2A 2 | X16997 |
| | 2A*3 | MM2A 3 | J00479 |
| | 2B*1 | MM2B 1 | V00763 |
| | 2B*2 | MM2B 2 | J00461 |
| | 2B*3 | MM2B 3 | V00799 |
| | 2C*1 | MM2C 1 | J00479 |
| | 2C*2 | MM2C 2 | X16998 |
| | 3 | MM3 | X00915 |
| ***Bos taurus*** | 1*1 | BT1 1 | X16701 |
| | 1*3 | BT1 2 | S82409 |
| | 2*1 | BT2 1 | M36946 |
| | 2*3 | BT2 3 | S624073 |
| | 3*1 | BT3 1 | U63638 |
| ***Ovis aries*** | 1 | OA1 | X69797 |
| | 2 | OA2 | X70983 |
| ***Felis catus*** | 1*1 | FC1 1 | AB016711 |
| | 1*2 | FC1 2 | AB016710 |
| ***Camelus dromedarius*** | 1A | CD1A | AJ421266 |
| | 2A | CD2A | AJ131945 |
| | 3 | CD3 | Z48947 |
| ***Monodelphis*** | *1 | MD 1 | AF035195 |
| ***Thricosurus*** | *1 | TV 1 | AF157619 |
| | *2 | TV 2 | AF191648 |
| ***Mustela vison*** | | MV | L07788 |
| ***Llama glama*** | | LG | AF305955 |
| ***Sus scrofa*** | 1 | SS1 | U03778 |
| | 2A*1 | SS2A 1 | U03779 |
| | 2A*2 | SS2A 2 | M81769 |
| | 2B | SS2B | U03780 |
| | 3 | SS3 | U03781 |
| | 4 | SS4 | U03782 |
| | 5 | SS5 | M81771 |
| ***Macaca mulata*** | 1 | MAMU1 | AF045537 |
| | 2 | MAMU2 | AF045539 |
| | 3 | MAMU3 | AF045538 |
| ***Macaca fascicularis*** | 1 | MAFA1 | AF045536 |
| ***Cricetulus migratorius*** | | CM | U17166 |

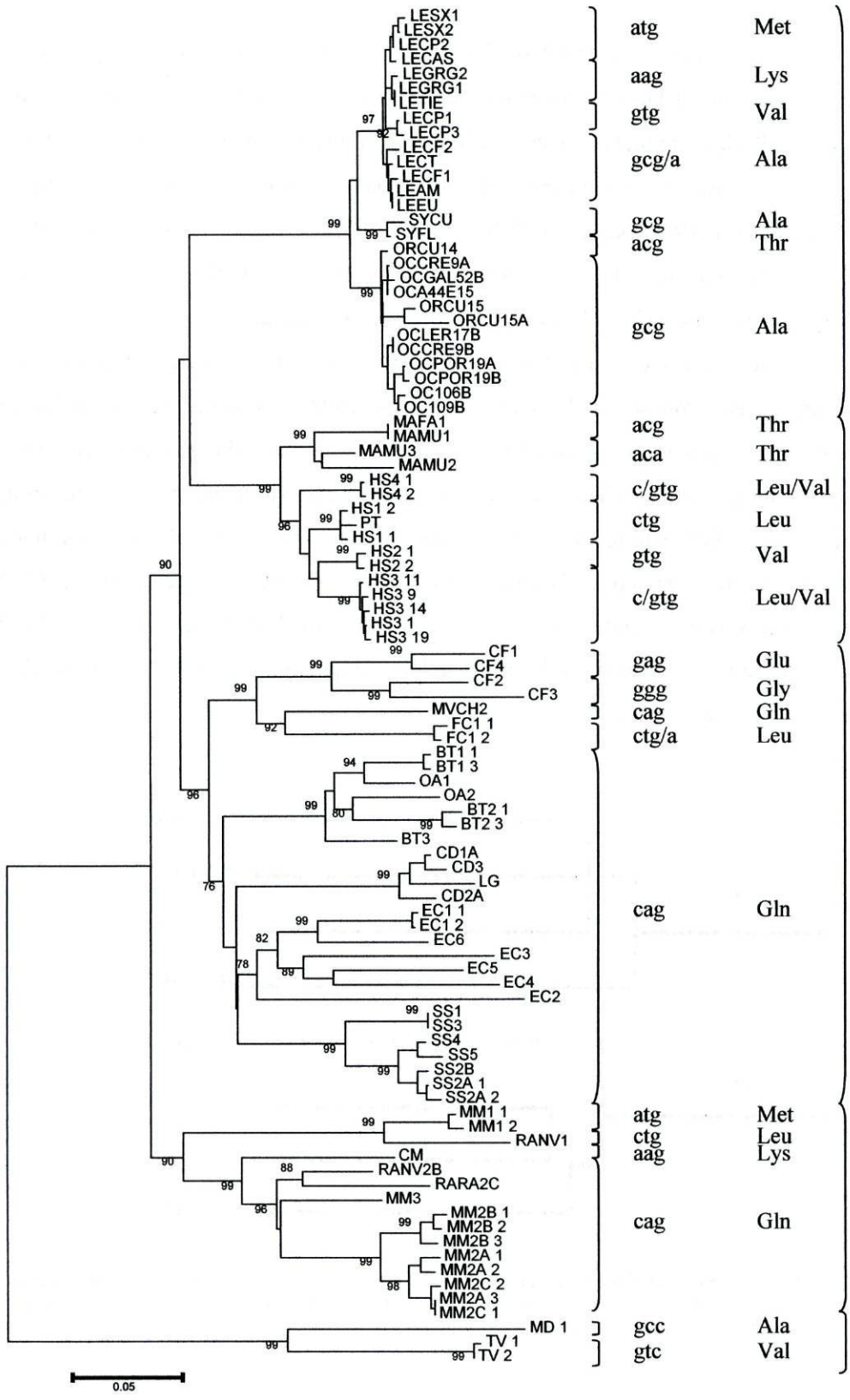



**Figure 2.4.** Neighbor-Joining tree for CH2 genes sequences obtained in species belonging to lagomorphs, primates, ferungulates and rodents groups. Marsupial sequences determine the root of the tree. Numbers represent the bootstrap values when 1000 replicate samplings were done. For each sequence the nucleotide triplet and the corresponding amino acid observed at position H309 are shown.

Analyses of molecular data have lead to two conflicting hypothesis regarding the phylogenetic position of lagomorphs among mammals (Figure 2.5). According to the first hypothesis the Rodents branched apart 110 My ago, before the great mammalian radiation (90 My). In that case, the lagomorphs are more related to the primates and artiodactyls (Kumar and Hedges, 1998; Nei *et al.* 2001) than to rodents. Under the second hypothesis artiodactyls diverged 95 My ago, before the separation of the primates from a rodents/lagomorph branch (Murphy *et al.*, 2001; Springer *et al.*, 2003).

The gene tree of the *IgGCH2* exon is concordant with the classification of the major phylogenetic groups proposed by Kumar and Hedges (1998) and Nei *et al.* (2001). Indeed, in the gene tree, rodents form a well-defined group apart from the other groups (lagomorphs, ferungulates and primates). There are some incongruences within the ferungulate group, since the swine, which belongs to the artiodactyls seems to be more related to the horse, which belongs to the *Perissodactyla* group, than to other *Artiodactyla* species like cattle or sheep. However, it is important to remind that we are only working a small portion of DNA, which does not allow inferring phylogenetic relationships. We only infer the most likely phylogeny for this particular gene.

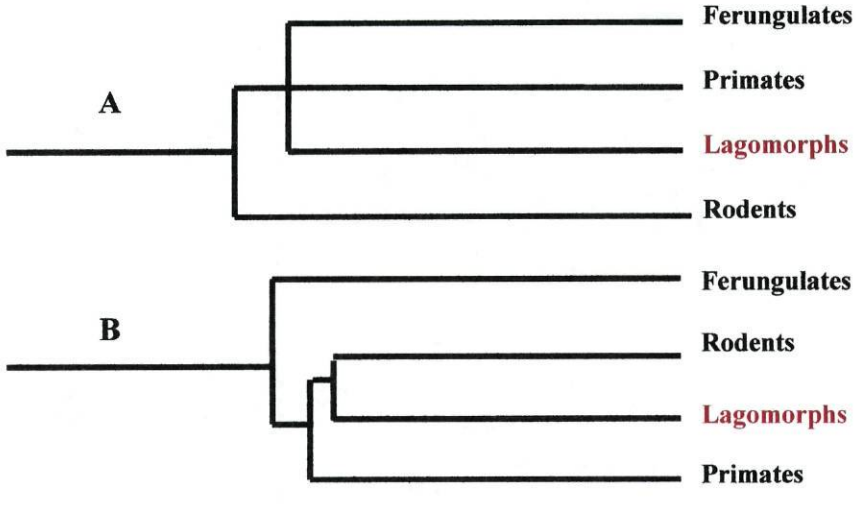



**Figure 2.5:** Two different evolutionary scenarios proposed for the relationship between the major mammal groups. A: Models based in the phylogenetic trees: A - obtained by Kumar and Hedges (1988); B – obtained by Springer *et al.* (2003). The *IgGCH2* gene tree conforms to model A.

# Article 2

P.J. Esteves, P.C. Alves, N. Ferrand and W. van der Loo (2002).

Hotspot variation at the CH2-CH3 interface of Leporid IgG antibodies (*Oryctolagus*, *Sylvilagus* and *Lepus*). *European Journal of Immunogenetics* **29** (6): 529-536.

**Short Communication**

# Hotspot variation at the *CH2–CH3* interface of leporid IgG antibodies (*Oryctolagus, Sylvilagus* and *Lepus*)[1]

**P. J. Esteves,*†‡ P. C. Alves,*† N. Ferrand*† & W. van der Loo‡**

## Summary

The European rabbit (*Oryctolagus cuniculus*) is the only species known to express only one subclass of gamma class immunoglobulin (IgG) antibodies. The rabbit *IGHGCH2* (second domain of the gene region encoding the IgG heavy chain constant region) or *e* locus presents two serologically defined alleles, the *e14* and *e15* allotypes. These are correlated with amino acid variation at position 309, which is located within the target region of the neonatal FcRn receptor. The *e14* and *e15* markers were also observed in other lagomorph species. Population genetic research has indicated that polymorphism at this locus is sustained by selection. We present here the *IGHGCH2* exon sequences for 12 species of rabbit and hare (genera *Oryctolagus, Sylvilagus* and *Lepus*). The inferred amino acid sequences reveal that, despite an overall sequence identity of 97%, five different residues can occur at position 309. As for *Oryctolagus*, the *e15* allotype was always associated with the presence of an Ala309 codon. In all but one case, this codon defined an allotype-specific *Tha*I restriction site. The potential of PCR/*Tha*I restriction fragment length polymorphism (RFLP) analyses for studying *IGHGCH2* variation within and between populations is emphasized.

## Introduction

In conjunction with FcRn receptors, epitopes at the gamma class immunoglobulin (IgG) *CH2–CH3* interface play a crucial role in the processing of IgGs and their transplacental transport (Johanson *et al.*, 1981; Gethie & Ward, 1997). The *CH2–CH3* interface is also the target of pathogens (Deisenhofer, 1981; Kim *et al.*, 1999), and various lines of evidence suggest that isotype and allotype variation in this region can be adaptive (Dubiski & Good, 1972; Granoff *et al.*, 1984). Unlike in other mammals, there are no subclasses of IgG antibodies in the European rabbit (*Oryctolagus cuniculus*). The rabbit IgG *CH2* domain presents two alleles that are distinguishable by serological markers, the *e14* and *e15* allotypes (Hamers & Hamers-Casterman, 1965; Mage, 1981). In the modern distribution range of the European rabbit, highly significant allele associations ('linkage disequilibrium') have been observed between the *IGHGCH2* locus (the second domain of the gene region encoding the IgG heavy-chain constant region) and other *IG* loci, including the *IGKC1* or *b* locus of the light-chain constant region (van der Loo *et al.*, 1987, 1996; van der Loo, 1993). The polymorphism was found to be correlated with an alanine-threonine replacement at amino acid position 309 (Appella *et al.*, 1971). Position 309 (position 79 in the International Immunogenetics Database (IMGT) numbering system) is adjacent to His310, which is known to affect interactions with a variety of Fc receptors (Medesan *et al.*, 1997; West & Bjorkman, 2000).

The *e15* allotype has been found in species of the genera *Oryctolagus*, *Lepus*, *Sylvilagus*, *Romerolagus* and *Ochotona* (Cazenave *et al.*, 1977; van der Loo & Hamers-Casterman, 1979): A number of species seem to be fixed for this allele (P. J. Esteves and W. van der Loo, unpublished data), while the *e14–e15* polymorphism has been found in the Mexican volcano rabbit (*Romerolagus diazi*) and in the cottontail rabbit (*Sylvilagus floridanus*). Both markers were absent in the mountain hare (*Lepus timidus*). The relationship between serology and protein variation at IgG *CH2* domains has been studied by amino acid sequencing of tryptic peptides in various lagomorph species (Appella *et al.*, 1971; Aggarwal & Mandy, 1976; Cazenave *et al.*, 1977; Teherani *et al.*, 1979, 1982). In contrast, the DNA sequence repertoire for the lagomorph IgG *CH2* domain consists of only four sequences, all obtained from domestic rabbits (*Oryctolagus cuniculus*). In this paper we present the *CH2* exon sequences of the *IGHG* locus for 12 leporid species. We also show how restriction fragment length polymorphism (RFLP) analysis of PCR products could replace serology in population genetic studies.

[1] The nucleotide sequence data described in this paper have been submitted to the EMBL nucleotide sequence database and have been assigned the accession numbers AJ295216, AJ295217, AJ295218, AJ295219, AJ295220, AJ295221, AJ295222 and AJ295223.

* Centro de Estudos de Ciência Animal (CECA), ICETA-UP, Vairão, Portugal, † Departamento de Zoologia-Antroplogia Faculdade de Ciências da Universidade do Porto, Porto, Portugal, and ‡ Institute of Molecular Biology and Biotechnology, Vrije Universiteit Brussel, St-Genesius-Rode, Belgium.

Received 8 March 2002; accepted 15 July 2002

Correspondence: W. van der Loo, Institute of Molecular Biology and Biotechnology, Evolutionary Immunogenetics Research Group, Vrije Universiteit Brussel, St-Genesius-Rode 1640, Belgium. E-mail: wvdloo@mail.icav.up.pt or wvdloo@ben.vub.ac.be

**Figure 1.** DNA sequences of F3-R2/PCR products obtained with genomic DNA of, respectively, *Lepus granatensis* (Legr_*eng*), *Lepus americanus* (Leam_*e15*), *Lepus timidus* (Leti_*eng*), *Lepus capensis* (Lecp_*eng*), *Lepus europaeus* (Leeu_*e15*), *Lepus castroviejoi* (Lecs_*eng*), *Lepus saxatilis* (Lesx_*eng*), *Lepus calotis* (Lect_*e15*), *Lepus californicus* (Lecf_*e15*), *Sylvilagus cunicularis* (Sycu_*e15*), *Sylvilagus floridanus* (Syfl_*e15*) and *Oryctolagus cuniculus* (Orcu_*e14* and Orcu_*e15*), where extentions '_*e15*', '_*e14*' and '_*eng*' refer to the serological allotypes *e15*, *e14*, and *e*-negative, respectively. Polymorphisms: Y = C and T; R = A and G; S = G and C. '.' indicates identity with Legr_*eng*; '*' indicates insertions/deletions; '~' indicates no clear data. The primers F3 and R2 are shown in bold italic, and the *Tha*I restriction sites in bold underlined.

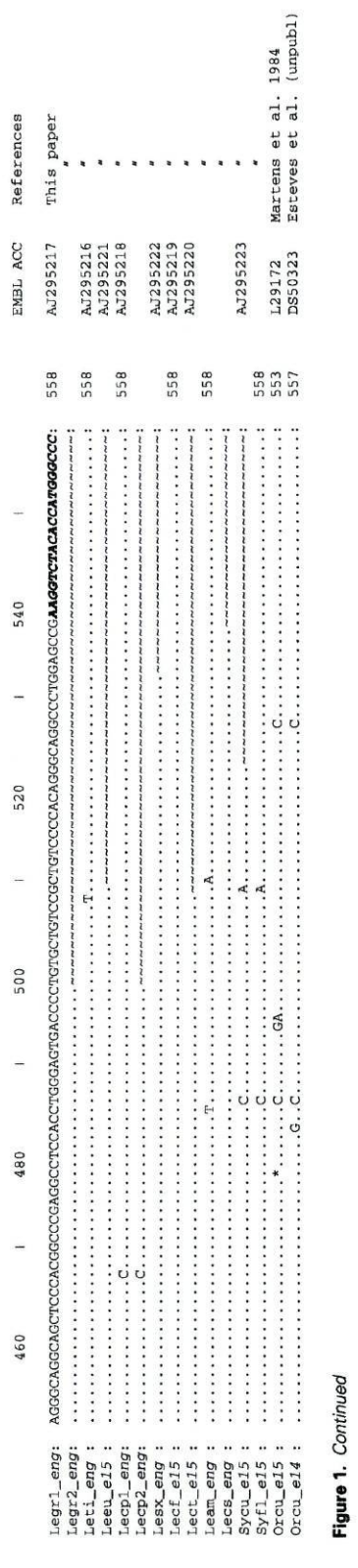



**Figure 1.** *Continued*

## Materials and methods

Tissue and blood samples were taken from wild specimens collected in the Iberian Peninsula (*Lepus granatensis* and *Lepus castroviejoi*), France (*Lepus europaeus*), Scotland (*Lepus timidus*), Morocco (*Lepus capensis*), Mozambique (*Lepus saxatilis*), Mexico (*Sylvilagus cunicularis*, *Lepus calotis* and *Lepus californicus*), and the USA (*Sylvilagus floridanus* and *Lepus americanus*). Allotypes were determined in gel precipitation with allotype-specific alloantisera, as previously described (van der Loo, 1993). The DNA was extracted from liver and/or blood clots using a Qiagen extraction kit (Qiagen, Vienna, Austria). The *CH2* exon regions were amplified by PCR using oligonucleotide primers designed according to the genomic DNA sequence of *Lepus granatensis* (P. J. Esteves, unpublished data). The primers were F3 (5′gtgagtcccatTagcctcac3′) and R2 (5′gggcccatggtgtagacctt3′). The PCR amplification protocol can be obtained upon request. Sequences were determined by automated sequencing following the Big Dye Terminator Cycle Sequencing protocol (Perkin Elmer, Warrington, UK) using the primers F3 and R2. For RFLP analysis, the PCR products were digested with restriction enzyme *Tha*I (New England Bio Laboratory (Beverly, MA)), according to the manufacturer's instructions. Sequence diversity was estimated using MEGA version 2.1 (Kumar *et al.*, 2001) as well as DNASP version 3.51 (Rozas & Rozas, 1999).

## Results and Discussion

As none of the *Sylvilagus* or *Lepus* specimens studied here displayed the *e14* allotype in their serum, the DNA data presented for these species are for two serologic phenotypes: *e15*-positive and *e14*-negative on the one hand ('*e15*'), and *e15*-negative and *e14*-negative ('*eng*') on the other.

Sequences including the entire *CH2* domain exon were determined for a total of 25 specimens. In Fig. 1, the unique sequences are displayed for each species. Results obtained with heterozygous *e14/e15* rabbits confirmed that heterozygous nucleotide positions were very clearly revealed in the sequencing profiles (note: our *e15* rabbits did not show the deletions occurring in the 5′ intron region of the L29172 sequence reported by Martens *et al.*, 1984). In Fig. 2, we show the inferred amino acid sequences, together with data obtained with tryptic digests, as published in Teherani *et al.* (1982). Ignoring the usual Gln-Glu and Asn-Asp ambiguities, there was good agreement between the peptide data and the genomic data. There were, however, conflicting results for the presence of Arg(285) in hares (HA-1 and HA-3). As His285 was found in each of the 48 amplified genomes of *Lepus* and *Sylvilagus*, and occurs most frequently in other mammalian species, the assignment of Arg(285) to *Lepus* and *Ochotona* species may reflect the difficulty of amino acid sequencing. If this is the case, the hare specimen studied by Teherani *et al.* (1982) probably belongs to *Lepus californicus*. HA-1 and HA-3 could indeed be homozygous for one of the two alleles (Lecf_*e15**1 and Lecf_*e15**2) present in the specimen studied here.

| EU numbering | 2 3 2345678901 | 2 4 2345678901 | 22 55 2345678901 | 2 6 2345678901 | 2 7 2345678901 | 2 8 2345678901 | 2 9 2345678901 | 3 0 1 2345678901 | 3 1 2345678901 | 3 2 2345678901 | 3 3 2345678901 | n[a] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *PCR DNA sequences* | | | | | | | | | | | | |
| Leeu_*e15* | PELLGGPSVF | IPPPKPKDTL | **MI**SLTPEVTC | VVVDVSQDDP | EVQFTWYINN | EQVHTARPPL | REQQFNSTIR | VVSTLPIA**HQ** | DWLRGKEFKC | KVHNKALPSP | IEKTISKAK | |
| Leam_*e15* | .......... | .......... | .......... | .......... | .......... | .......... | .......... | .......... | .......... | .......... | .......... | 2 |
| Lect_*e15* | .......... | .......... | .......... | .......... | .......... | .......... | .......... | .......... | .......... | .......... | .......... | 2 |
| Lecf*1_*e15* | .......... | .......... | .......... | .......... | .......... | **K**......... | .......... | ...**I**..... | .......... | .......... | .......... | 1 |
| Lecf*2_*e15* | .......... | .......... | .......... | .......... | .......... | .......... | .......... | .......... | .......K.. | .......... | .......... | 1 |
| Leti_*eng* | .......... | .......... | .......... | .......... | .......... | .......... | .......... | .......... | .......K.. | .......... | .......... | 4 |
| Legr1_*eng* | .......... | .......... | .......... | .......... | .......... | .......... | .......... | .......... | .......K.. | .......... | .......... | 20 |
| Legr2_*eng* | .......... | .......... | .......... | .......... | .......... | .......... | .......S.. | .......... | .......K.. | .......... | .......... | 2 |
| Lecp1_*eng* | .......... | .......... | .......... | .......... | .......... | ...Y...... | .......... | .......... | .......V.. | .......... | .......... | 2 |
| Lecp2*1_*eng* | .......... | .......... | .......... | **P**......... | .......... | ...**Y**..... | .......... | .......... | .......**V**.. | .......... | .......... | 1 |
| Lecp2*2_*eng* | .......... | .......... | .......... | .......... | .......... | .......... | .......... | .......... | .......**M**.. | .......... | .......... | 1 |
| Lecs_*eng* | .......... | .......... | .......... | .......... | .......... | .......... | .......... | .......... | .......M.. | .......... | .......... | 4 |
| Lesx*1_*eng* | .......... | .......... | .......... | .......... | .......... | .......... | ..R....... | .......... | .......M.. | .......... | .Q........ | 1 |
| Lesx*2_*eng* | .......... | .......... | .......... | .......... | .......... | .......... | ..R....... | .......... | .......M.. | .......... | .......... | 1 |
| Sycu_*e15* | .......... | ...R...... | .......... | .......... | ...V...... | .......... | Q......... | .......... | .......... | .......... | .....A.... | 2 |
| Syfl_*e15* | .......... | ...R...... | .......... | .......... | ...V...... | .......... | .......... | .......... | .......... | .......... | .....A.... | 2 |
| Orcu_*e14* | .......... | ...R...... | .......... | .......... | .......... | ...R...... | .......... | .......... | .......T.. | .......... | .....A.... | .....R |
| Orcu_*e15* | .......... | ...R...... | .......... | .......... | .......... | ...R...... | .......... | .......... | .......... | .......... | .....A.... | .....R |
| *Tryptic peptides (Teherani et al. 1982)* | | | | | | | | | | | | |
| CT-12 | | | .V...... | .......... | ...V....D. | | .....D.... | .......... | .......... | ...D...... | .......... | |
| HA-3 | | | .......... | .......E.. | .......... | .....D. | .E.R...... | .....D.... | .......... | .......... | ...D...... | .......... |
| HA-1 | | | .......... | .......... | .......... | .....D. | K..R...... | .......... | ...I...... | ...S...... | ...D...... | .......... |
| LT | | | .......... | .......E.. | ... | | | | ......K... | .... | | |
| PIKA | | | .......S.. | .....E.. | .......... | .....D. | ...R. | .......... | .....L. | ...K..... | ..... | ...R.. |

**Figure 2.** Amino acid sequences, as inferred from nucleotide sequences displayed in Fig. 1. Dots indicate identity with sequence Leeu_*e15*. Data obtained from tryptic digests are from Teherani *et al.* (1982): CT-12: *Sylvilagus floridanus*; HA-1 and HA-3, unspecified *Lepus* spp. from Kansas (USA); LT: *Lepus timidus* (Ireland); PIKA: *Ochotona rufescens*. Italics suggest that data might be incorrect. 'N' indicates that the *cis-trans* configuration between alleles (n*1 and n*2) was not resolved. n[a], number of genes sequenced.

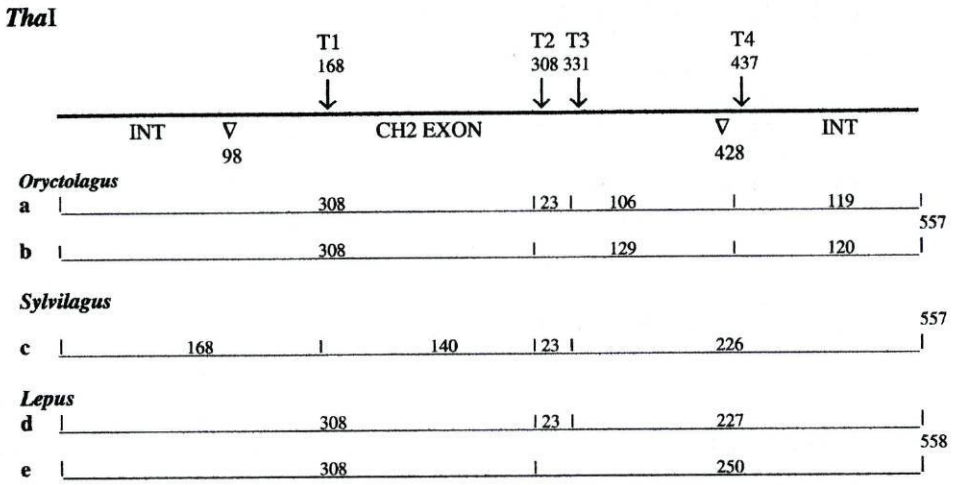



**Figure 3.** *Tha*I haploid restriction patterns inferred from RFLP analysis of F3-R2/PCR products. The full length of the PCR product is 557–558 bp, and includes the 5′ primer (20 bp), the 3′primer (20 bp), the entire *CH2* exon (330 bp), and parts of the flanking introns. '∇' indicates the exon–intron bounderies. '↓Π' indicates the potential restriction sites T1–T4. The nucleotide sites are numbered according to the sequence obtained for *Lepus granatensis* (AJ295217, Fig. 1). The patterns a and b correspond to *Oryctolagus cuniculus* allotypes *e15* and *e14*, respectively. c: *Sylvilagus* spp. showing allotype *e15* (*S. cunicularis* and *S. floridanus*); d: *Lepus* spp. showing allotype *e15* (*L. europaeus, L. calotis, L. californicus* and *L. americanus*); e: *Lepus* spp. that are negative for *e15* and *e14* markers (*L. timidus, L. granatensis, L. castroviejoi, L. capensis* and *L. saxatilis*), and *Lepus californicus* displaying a Ala(GCA) codon at amino acid position 309.

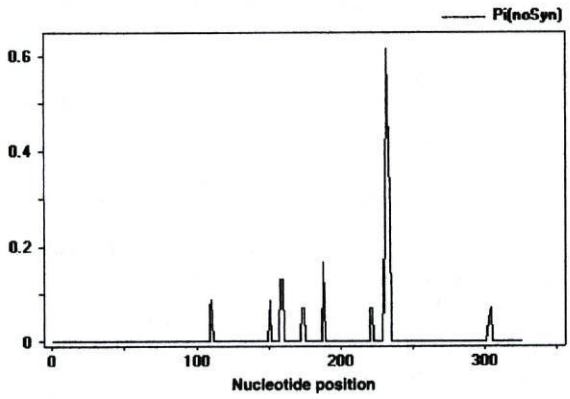



**Figure 4.** Non-synonymous nucleotide diversity among *Lepus IGHGCH2* coding regions, shown in a sliding window plot of non-synonymous diversity among exon sequences of *Lepus* spp. displayed in Fig. 1, using the DnaSP program (Rozas & Rozas, 1999), with window size = 2, step = 1. The value of Pi(noSyn) for the total sequence was 0.01.

Amino acid differences were observed at 13 sites, seven of which showed within-species variation, while five could define species-specific differences. Only position 309 was polymorphic in more than one species. This site can be occupied by five different amino acid residues. The *e15* allotype was always associated with a Ala309 codon. This codon contributes to a *Tha*I restriction site (C[G^CG], '*Tha*I-3' in Fig. 1), except for one allele of *Lepus californicus* (Lecf_*e15*), where Ala(309) was encoded by a GCA. Because all PCR products share the *Tha*I site related to the Arg(301) codon ([CG^C]G; '*Tha*I-2'), the presence of Ala309(GCG) was marked by a 23-bp fragment in RFLP analysis. This fragment was diagnostic for the *e15* allotype in all species studied. In Fig. 3 we show the different haploid *Tha*I restriction patterns as inferred from RFLP and sequence data. Two restriction sites were species-specific: one was specific to *Oryctolagus cuniculus* and was created by a nucleotide deletion within the 3′ intron region ('*Tha*I-4'), and the second was only found in *Sylvilagus* spp. and involved the Ser254Arg255 codons ([TCG]^[CGC], '*Tha*I-1'). Interestingly, Teherani *et al.* (1982) reported Val255 in tryptic peptides of *Sylvilagus floridanus* IgG (CT-12). Position 255 is situated close to the canonic Ile253, which, like His310, is known to interact with FcRn (Kim *et al.*, 1999). As *e14*-positive cottontail rabbits have been found (Cazenave *et al.*, 1977), it appears that, in this species, *Tha*I RFLP may detect amino acid variation at positions 255 and 309.

Although the rabbit allo-antisera provide the safest shortcut for detecting either Ala309 or Thr309 at the IgG *CH2–CH3* interface, they cannot detect new variants. In contrast, PCR-RFLP combined with sequence determination has allowed us to discover new polymorphisms in *Lepus* and *Sylvilagus* species. Whereas the *Tha*I-3 site always implies a GCG codon at position 309, five different 309 codons were associated with its absence (i.e. ACG, GTG, ATG, AAG and GCA). In previous studies we had observed that sera of *L. europaeus* always presented the *e15* allotype, whereas this marker was absent in all 30 specimens of *L. timidus* and *L. granatensis* studied (Schröder *et al.*, 1987; P. J. Esteves and W. van der Loo, unpublished data). As expected, this dichotomy is reflected in the *Tha*I restriction patterns (not shown). PCR-RFLP methods can therefore help to overcome the need for allotype-specific antisera for measuring gene flow between the different *Lepus* species in areas of range

overlap (cf. Schröder *et al.*, 1987). In fact, the spread of *e*-allotype negative genes within *L. europaeus* will be better monitored by RFLP than by serology. An adapted version of the PCR-RFLP method described here could also contribute to species determination for tissue samples and fossil remains deposited in field and museum collections.

Considering the very limited degree of sequence variation among the IgG *CH2* domains of leporids, it is interesting that position 309 was polymorphic within each genus. Despite the 97% average amino acid sequence similarity, five different residues were observed at this particular position, namely, Ala, Thr, Lys, Val, and Met. Amino acid sequence diversity ($\pi$) at 309 was 30 times larger than the mean diversity of the entire sequence ($\pi_{309}$ = 0.742, $\pi_{CH2}$ = 0.024; MEGA2 software: distances > sequence diversity > model > amino acid > p-distances). Figure 4 highlights the 'hotspot' variability among *Lepus* sequences, where position 309 accounts for some 40% of the total amino acid diversity. It is remarkable that human *IGHG3* and *IGHG4* show allotypic variation, which is caused by a Val309Leu mutation (Brusco *et al.*, 1998; Dard *et al.*, 2001). X-ray crystallography has revealed direct contact between the side-chains of *IGHG*-Leu309 of the rat FcRn receptor (Burmeister & Bjorkman, 1994), while population genetic studies have indicated that gene diversity at this site can be adaptive in the rabbit (van der Loo *et al.*, 1987, 1996). Within this context, it seems that the reported 'hotspot' variability at position 309 may have to do with diversity enhancement selection. A better knowledge of the sequence variation, as well as more advanced methods for measuring allele distribution in populations, could contribute to the understanding of the biological meaning of gene diversity in the antibody constant region.

## Acknowledgements

We thank Fernando Cervantes, Brahim Haddane, Juan Martinez, Franz Suchentrunk and Mario Vargas for providing us with samples of different *Lepus* and *Sylvilagus* species. This work was partially supported by Direcçao Geral de Florestas, by a grant of the Foundation for Science and Technology-Portugal (Praxis XXI/BD/16207/98) to P.E., and by a grant of the Belgian Fonds voor Wetenschappelijk Onderzoek Vlaanderen (Krediet 1.5579.98-FWOKN35) to W.vdL.

## References

Aggarwal, S.K. & Mandy, W.J. (1976) Lagomorph IgG hinge region: allotype associated amino acid sequence variations. *Immunochemistry*, **13**, 215.

Appella, E., Chersi, A., Mage, R. & Dubiski, S. (1971) Structural basis of the A14 and A15 allotypic specificities in rabbit IgG. *Proceedings of the National Academy of Sciences of the USA*, **68**, 1341.

Brusco, A., Saviozzi, S., Cinque, F., DeMarchi, M., Boccazzi, C., de Lange, G., van Leeuwen, A.M. & Carbonera, A.O. (1998) Molecular characterisation of immunoglobulin G4 isoallotypes. *European Journal of Immunogenetics*, **25**, 349.

Burmeister, W.P., Huber, A.H. & Bjorkman, P.J. (1994) Crystal structure of the complex of rat neonatal Fc receptor with Fc. *Nature*, **372**, 379.

Cazenave, P.A., Brezin, C., Puget, A. & Mandy, W.J. (1977) Lagomorph constant region IgG allotypes: shared e determinants in *Ochotona* sp. *Immunogenetics*, **4**, 489.

Dard, P., Lefranc, M.P., Osipova, L. & Sanchez-Mazas, A. (2001) DNA sequence variability of IGHG3 alleles associated to the main G3m haplotypes in human populations. *European Journal of Human Genetics*, **9**, 765.

Deisenhofer, J. (1981) Crystallographic refinement and atomic models of a human Fc fragment and its complex with fragment B of protein A from *Staphylococcus aureus* at 2.9- and 2.8-Å resolution. *Biochemistry*, **20**, 2361.

Dubiski, S. & Good, P.W. Jr (1972) Population genetics of the heavy chain immunoglobulin allotypes in the rabbit. *Proceedings of the Society of Experimental Biology and Medicine*, **141**, 486.

Gethie, V. & Ward, E.S. (1997) FcRn: the MHC class I-related receptor that is more than an IgG transporter. *Immunology Today*, **18**, 592.

Granoff, D.M., Boies, E., Squires, J., Pandey, J.P., Suarez, B., Oldfather, J. & Rodey, G.E. (1984) Interactive effects of genes associated with immunoglobulin allotypes and HLA specificities on susceptibility to *Haeophylus influenzae* disease. *Journal of Immunogenetics*, **11**, 181.

Hamers, R. & Hamers-Casterman, C. (1965) Molecular localisation of A chain allotypic specificities in rabbit IgG (7S globin). *Journal of Molecular Biology*, **14**, 288.

Johanson, R.A., Shaw, A.R. & Schlamowitz, M. (1981) Evidence that the *CH2* domain of IgG contains the recognition unit for binding the fetal rabbit yolk sac membrane receptor. *Journal of Immunology*, **126**, 194.

Kim, J.K., Firan, M., Radu, C.G., Kim, C.H., Ghetie, V. & Ward, E.S. (1999) Mapping the site on human IgG for binding of the MHC class I-related receptor, FcRn. *European Journal of Immunology*, **29**, 2819.

Kumar, S., Tamura, K., Jakosen, I. & Nei, M. (2001) *MEGA2: Molecular Evolutionary Genetics Analysis Software.* Arizona State University, Tempe, AZ.

van der Loo, W. (1993) Variance analysis of immunoglobulin alleles in natural populations of rabbit (*Oryctolagus cuniculus*). *Genetics*, **134**, 171.

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.

van der Loo, W., Arthur, C.P., Drees-Wallage, M. & Richardson, B. (1987) Nonrandom allele associations between unlinked protein loci: are the polymorphisms of the immunoglobulin constant regions adaptive? *Proceedings of the National Academy of Sciences of the USA*, **84**, 3075.

van der Loo, W., Boussès, P., Arthur, C.P. & Chapuis, J.-L. (1996) Compensatory aspects of allele diversity at immunoglobulin loci: gene correlations in rabbit populations devoid of light chain diversity (*Oryctolagus cuniculus* L., Kerguelen Islands). *Genetics*, **144**, 1181.

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 chain of rabbit IgG. *Journal of Immunology*, **133**, 1022.

Medesan, C., Matesoi, D., Radu, C., Ghetie, V. & Ward, E.S. (1997) Delineation of the amino acid residues involved in transcytosis and catabolism of mouse IgG1. *Journal of Immunology*, **158**, 2211.

Rozas, J. & Rozas, R. (1999) Dnasp version 3: an integrated program for molecular and population genetics and molecular evolution analysis. *Bioinformatics*, **15**, 174.

Schröder, J., Soveri, T., Suomalainen, H.A., Lindberg, L.A. & van der Loo, W. (1987) Hybrids between *Lepus timidus* and *Lepus europaeus* are rare although fertile. *Hereditas*, **107**, 185.

Teherani, J., Capra, J.D., Aggarwal, S., Mandy, W.J. (1979) Amino acid sequence analysis of group e allotype related peptides derived from lagomorph IgG. *European Journal of Immunology*, **9**, 690.

Teherani, J., Capra, J.D. & Mandy, W.J. (1982) Amino acid sequence of the *CH2* domain from various lagomorph IgGs. *Molecular Immunology*, **19**, 841.

West, A.P. & Bjorkman, P.J. (2000) Crystal structure and immumoglobulin G binding properties of the human major histocompatibility complex-related Fc receptor. *Biochemistry*, **39**, 9698.

## Variability at position H309 among mammalian species

The nucleotide triplet and the inferred amino acid observed at position H309 (79 IMGT numbering) for each sequence are presented in Figure 2.4. Figure 2.6 shows the non-synonymous nucleotide diversity of the *IgGCH2* coding regions within mammalian groups (*Primata, Rodentia, Lagomorpha, Carnivora* and *Perissodactyla/Artyodactyla*). In the primates and carnivores, like in the lagomorphs, the highest $\pi$ (the average number of nucleotide differences per site) is observed at nucleotides encoding H309. In the lagomorphs six different triplets are observed, corresponding to five different amino acid residues (Met, Lys, Val, Ala and Thr). Four different triplets, coding for amino acids Val, Leu, and Thr, are present in the primates. In the carnivores, there are five different triplets and four amino acids (Glu, Gly, Gln, and Leu) and in the rodents, four different triplets for four amino acids (Met, Leu, Lys and Gln). In the *Perissodactyla/Artiodactyla* group, the Gln triplet CAG is fixed at position 309.

The occurrence of positive selection at the amino acid level was tested by the differences between non-synonymous and synonymous substitutions (Hughes and Nei, 1988, 1989). Table 2.3 shows the values of dS (Synonymous distance) and dN (Non-synonymous distance) at position 309 and the average of these two measures in all remaining positions. The estimation of the distances was performed in the computer program MEGA 2 using Nei-Goijobori method (p-distance). It shows that the dN/dS ratios are higher at postion H309, whereas on the average, in the remaining part of the exon, the opposite situation prevails.

**Table 2.3**

Values of dS (Synonymous distance) and dN (Non-synonymous distance) obtained at position 309 and in the average of all remaining positions for several mammalian groups. The standard error values are in brackets.

| Mammalian groups | all positions | | | H309 | | |
|---|---|---|---|---|---|---|
| | *dS* | *dN* | *dS-dN* | *dS* | *dN* | *dS-dN* |
| Marsupials | 0.197 (0.04) | 0.112 (0.02) | 0.085 (0.05) | ---- | 0.333 | -0.333 |
| Rodents | 0.259 (0.03) | 0.122 (0.01) | 0.137 (0.03) | ---- | 0.324 | -0.324 |
| Artyodactyls/Perissodactyls | 0.297 (0.03) | 0.130 (0.02) | 0.167 (0.03) | ---- | ---- | ---- |
| Carnivores | 0.288 (0.04) | 0.103 (0.02) | 0.185 (0.04) | 0.411 | 0.517 | -0.106 |
| Primates | 0.107 (0.01) | 0.035 (0.01) | 0.072 (0.02) | 0.184 | 0.318 | -0.134 |
| Lagomorphs | 0.068 (0.01) | 0.014 (0.004) | 0.054 (0.02) | 0.090 | 0.394 | -0.304 |

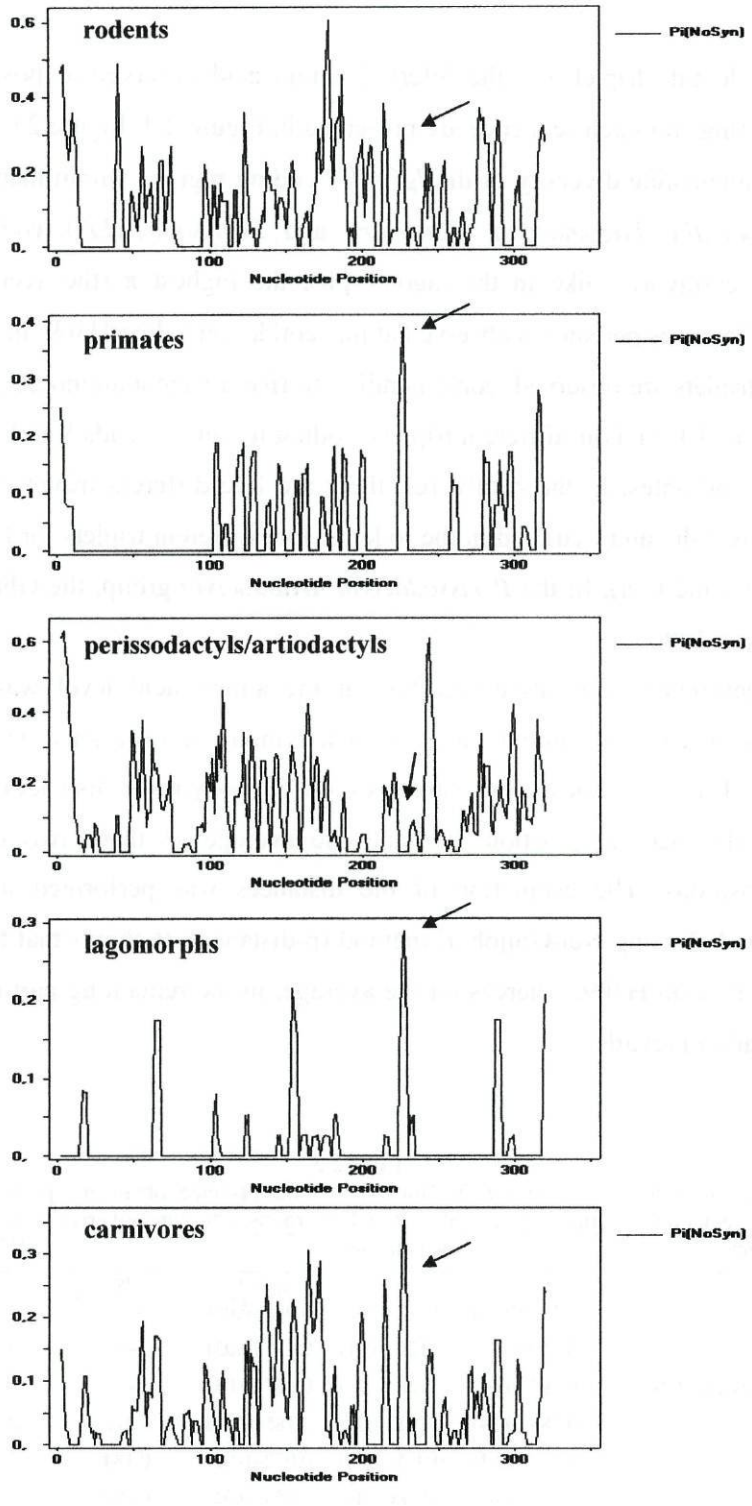



**Figure 2.6.** Non-synonymous nucleotide diversity (Pi) values obtained in *IgGCH2* locus shown in a sliding window (window size = 3, step = 1) for several mammalian groups: rodents – *Rattus*, *Mus and Cricetulus;* primates – *Pan*, *Homo* and *Macaca;* perissodactyls/artiodactyls – *Equus*, *Bos*, *Ovis*, *Camelus*, *Llama* and *Sus;* lagomorphs – *Oryctolagus*, *Sylvilagus* and *Lepus;* carnivores – *Canis*, *Mustela* and *Felis*. The analyses were performed using the program DnaSP (Rozas and Rozas, 1999). The arrows indicate the Pi values obtained at position H309.

These results seem to indicate that the "hotspot" variability at position 309 may have to do with diversity enhancement selection. However, this position does not seem to have the same importance for all groups, since it is fixed in the *Artiodactyla/Perissodactyla* group.

It can also be considered as fixed in the Iberian rabbits, which may be in contradiction with the adaptive value of the polymorphism at this coding region. It is therefore important to emphasize the population genetic evidence that the *e14-e15* polymorphism is only adaptive in situations where diversity at the $V_H$ and/or $C_{K1}$ loci is reduced (compensatory overdominance; van der Loo *et al.*, 1996).

# Citations

Aggarwal SJ and Mandy WJ (1976). Lagomorph IgG hinge region: allotype associated amino acid sequence variations. *Immunochemistry* **13**(3): 215-20.

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**(6): 1341-345.

Bernstein KE, Alexander CB and Mage RG (1983). Nucleotide sequence of a rabbit IgG heavy chain from the recombinant F-I haplotype. *Immunogenetics* **18**(4): 387-97.

Burmeister WP, Huber AH and Bjorkman PJ (1994). Crystal structure of the complex of rat neonatal Fc receptor with Fc. *Nature* **372**(6504): 379-83.

Cazenave PA, Brezin C, Puget A and Mandy WJ (1977). Lagomorph constant region IgG allotypes: shared *e* determinants in *Ochotona* sp. *Immunogenetics* **4**: 489-93.

Deisenhofer J (1981). Crystallographic refinement and atomic models of a human Fc fragment and its complex with fragment B of protein A from *staphylococcus aureus* at 2.9- and 2.8-A resolution. *Biochemistry.* **20:** 2361-370.

Dubiski S and Good PW Jr (1972). Population genetics of the heavy chain immunoglobulin allotypes in the rabbit. *Proc Soc Exp Biol Med* **141**(2): 486-89.

Franklin I and Lewontin RC (1970). Is the gene the unit of selection? *Genetics* **65**(4): 707-34.

Ghetie V and Ward ES (1997). FcRn: the MHC class I-related receptor that is more than an IgG transporter. *Immunol Today* **18**(12): 592-98.

Granoff DM, Boies E, Squires J, Pandey JP, Suarez B, Oldfather J and Rodey GE (1984). Interactive effect of genes associated with immunoglobulin allotypes and HLA specificities on susceptibility to *Haemophilus influenzae* disease. *J Immunogenet* **11**(3-4): 181-88.

Hamers R and Hamers-Casterman C (1965). Molecular localization of A chain allotypic specificities in rabbit IgG (7S gamma-globulin). *J Mol Biol* **14**(1): 288-89.

Hughes AL and Nei M (1988). Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* **335**(6186): 167-70.

Hughes AL and Nei M (1989). Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. *Proc Natl Acad Sci* U S A **86**(3): 958-62.

Idusogie EE, Wong PY, Presta LG, Gazzano-Santoro H, Totpal K, Ultsch M, Mulkerrin MG (2001). Engineered antibodies with increased activity to recruit complement. *J Immunol* **166**(4): 2571-575.

Johanson RA, Shaw AR and Schlamowitz M (1981). Evidence that the *CH2* domain of IgG contains the recognition unit for binding the fetal rabbit yolk sac membrane receptor. *J. Immunol.* **126**: 194-9.

Kim JK, Firan M, Radu CG, Kim CH, Ghetie V and Ward ES (1999). Mapping the site on human IgG for binding of the MHC class I-related receptor, FcRn. *Eur J Immunol* **29**(9): 2819-825.

Kreitman M and Hudson RR (1991). Inferring the evolutionary histories of the Adh and Adh-dup loci in *Drosophila melanogaster* from patterns of polymorphism and divergence. *Genetics* **127**(3): 565-82.

Kumar S and Hedges SB (1998). A molecular time scale for vertebrate evolution. *Nature* **392**: 920-923.

Kumar S, Tamura K, Jakobsen IB and Nei M (2001). MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* **17**(12): 1244-5.

Lewontin RC (1974). *The Genetic Basis of Molecular Change*. Columbia Univ Press, NY and London.

Mage RG, Young-Cooper GO, Rejnek J, Ansari AA, Alexander CB, Appella E, Carta-Sorcini M, Landucci-Tosi S, Tosi RM (1977). Rabbit immunoglobulin allotypes: complexities of their genetics, expression, structural basis, and evolution. *Cold Spring Harb Symp Quant Biol* **41** Pt 2: 677-86.

Mage RG (1981). The phenotypic expression of rabbit immunoglobulins: a model of complex regulated gene expression and cellular differentiation. In: *Contemporary Topics in Molecular Immunology* **8**: 89.

Martens CL, Moore KW, Steinmetz M, Hood L and Knight KL (1982). Heavy chain genes of rabbit IgG: isolation of a cDNA encoding gamma heavy chain and identification of two genomic C gamma genes. *Proc Natl Acad Sci* U S A **79**(19): 6018-022.

Martens CL, Currier SJ and Knight KL (1984). Molecular genetic analysis of genes encoding the heavy chains of rabbit IgG. *J Immunol* **133**(2):1022-027.

Medrano L and Dutrillaux B (1984). Chromosomal location of immunoglobulin genes: partial mapping of these genes in the rabbit and comparison with Ig genes carrying chromosomes of man and mouse. *Adv Cancer Res* **41**: 323-67.

Murphy W, Eizirik E, O'Brien SJ, Madsen O, Scally M, Douday CJ, Teeling E, Ryder OA, Stanhope MJ, de Jong WW and Springer MS (2001). Resolution of the Early Placental Mammal Radiation Using Bayesian Phylogenetics. *Science* **294:** 2348-351.

Nei M, Xu P and Glazko G (2001). Estimation of divergence times from multiprotein sequences for a few mammalian species and several distantly related organisms. *Proc Natl Acad Sci* U SA **98(5):** 2497-502.

Perlmutter RM, Hansburg D, Briles DE, Nicolotti RA and Davie JM (1978). Subclass restriction of murine anti-carbohydrate antibodies. *J Immunol* **121**(2): 566-72.

Rozas J and Rozas R (1999). DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15**(2): 174-5.

Saitou N and Nei M.(1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4(4): 406-25.

Scott MG and Fleischman JB (1982). Preferential idiotype-isotype associations in antibodies to dinitrophenyl antigens. *J Immunol* **128**(6): 2622-628.

Slack J, Der-Balian GP, Nahm M and Davie JM (1980). Subclass restriction of murine antibodies. II. The IgG plaque-forming cell response to thymus-independent type 1 and type 2 antigens in normal mice and mice expressing an X-linked immunodeficiency. *J Exp Med* **151**(4): 853-62.

Springer MS, Murphy WJ, Eizirik E and O'Brien SJ (2003). Placental mammal diversification and the Cretaceous-Tertiary boundary. *Proc. Natl. Acad. Sci.* U S A **100**(3): 1056-061.

Teherani J and Mandy WJ (1976). Constant region IgG allotypes in cottontail rabbits: group E allelic polymorphism. *Immunochemistry* **13**(3): 221-27.

Teherani J, Capra JD, Aggarwal S and Mandy WJ (1979). Amino acid sequence analysis of group *e* allotype-related peptides derived from lagomorph IgG. *Eur J Immunol* **9**(9): 690-95.

Teherani J, Capra JD and Mandy WJ (1982). Amino acid sequence of the CH2 domain from various lagomorph IgGs. *Mol Immunol* **19**(7): 841-46.

van der Loo W and Hamers-Casterman C (1979). Phylogeny of the rabbit immunoglobulin allotypes: rabbit anti e-locus antisera detected two allelic forms of IgG molecules in *Romerolagus diazi. Annals de l'Institut Pasteur, Immunologie C* **130:** 761.

van der Loo W and Hamers-Casterman C (1981). Genetic polymorphisms of the immunoglobulin heavy chain in *Romerolagus diazi*. In Myers and MacInnes CD (eds) Proceedins world Lagomorph Conference, Guelph, Ontario.

van der Loo W and Arthur CH (1985). Geographical immunogenetics of European rabbit. *Mammal Rev* **16**: 199.

van der Loo W (1987). Studies on the adaptive significance of the immunoglobulin polymorphisms (Ig allotypes) in wild rabbits. In Dubiski S (ed.). *The Rabbit in Contemporary Immunological Research*. Longman Scientific & Technical. pp 164-90.

van der Loo W, Arthur CP, Richardson BJ, Wallage-Drees M and 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): 3075-079.

van der Loo W and Verdoodt B (1992). Patterns of interallelic divergence at the rabbit *b*-locus of the immunoglobulin light chain constant region are in agreement with population genetical evidence for overdominant selection. *Genetics* **132**(4): 1105-117.

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): 171-87.

van der Loo W, Bousses P, Arthur CP and Chapuis JL (1996). Compensatory aspects of allele diversity at immunoglobulin loci: gene correlations in rabbit populations devoid of light chain diversity (*Oryctolagus cuniculus* L.; Kerguelen Islands). *Genetics* **144**(3): 1181-94.

West AP Jr and Bjorkman PJ (2000). Crystal structure and immunoglobulin G binding properties of the human major histocompatibility complex-related Fc receptor. *Biochemistry* **39**(32): 9698-708.

# CHAPTER 3

## Studies of gene diversity at the Heavy chain variable region gene locus (*a* locus or *IgVH* locus) in leporids

# Introduction

The VH region diversity is well characterized in the domestic rabbit. Three serologically defined allotypic lineages exist, the so-called $V_Ha$ allotypes *a1, a2* and *a3*. These 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.,* 1984). Except for chicken, $V_H$ allotypic markers have not been found in other species, probably because each individual genome disposes of a large number of $V_H$ gene segments. The genomic mapping of the $V_H$ gene segments showed that the rabbit uses in the *VDJ* rearrangements almost exclusively the *D*-proximal $V_H$ gene segment ($V_H1$) (see Figure 3.1), which can explain the allelic inheritance of the $V_H$ allotypes in this species (Knight and Becker, 1990; Knight, 1992). About 80 to 90% of circulating Ig molecules are derived from the $V_H1$ gene and express the $V_Ha$ allotypic markers (e.g. Kindt, 1975; Margolies *et al.,* 1977). The VH regions of the remaining 10-20% are encoded by the $V_Ha$-negative genes ($V_Hx$ and $V_Hy$) (Kim and Dray, 1972; Kim and Dray, 1973; Roux, 1981), which are localized at least 50Kb upstream of $V_H1$ (Knight and Becker, 1990). These three alleles are highly divergent (+/- 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, 10, 12, 13, 16 and 17) and 3 (positions 65, 67, 70, 71, 74, 75, 84 and 85) (see Figure 3.2) (Tonnele *et al.,* 1983; Mage *et al.,* 1984; Knight and Becker, 1990).

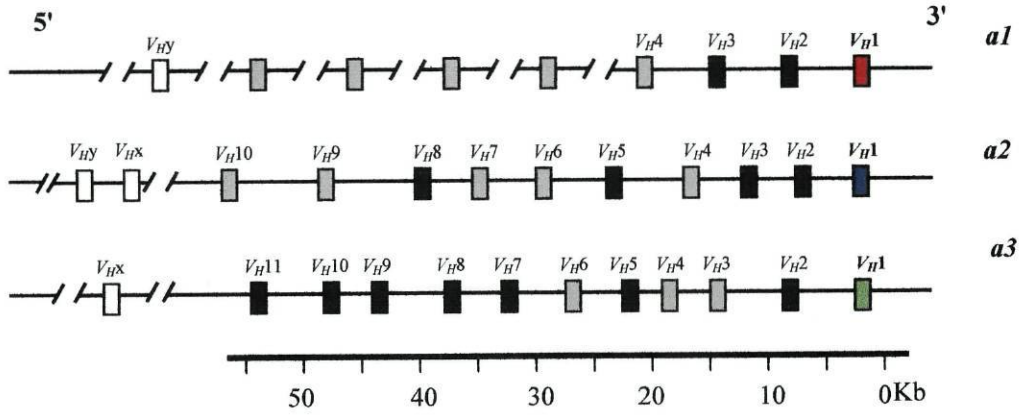



**Figure 3.1.** Genomic organisation of $V_H$ genes in haplotypes: $V_H$-*a1*, -*a2* and -*a3* (Tunyaplin and Knight, 1995). Each haplotype was estimated to have more than 100 $V_H$ genes in a 750-Kb region. Only genes with known genomic locations are shown. $V_H1a1$, $V_H1a2$ and $V_H1a3$ are indicated with red, blue and green boxes, respectively and the remaining functional $V_H$ genes are presented in grey boxes. The putative pseudogenes are represented by black boxes. White boxes represent $V_Hx$ and $V_Hy$. The genomic maps are approximate.

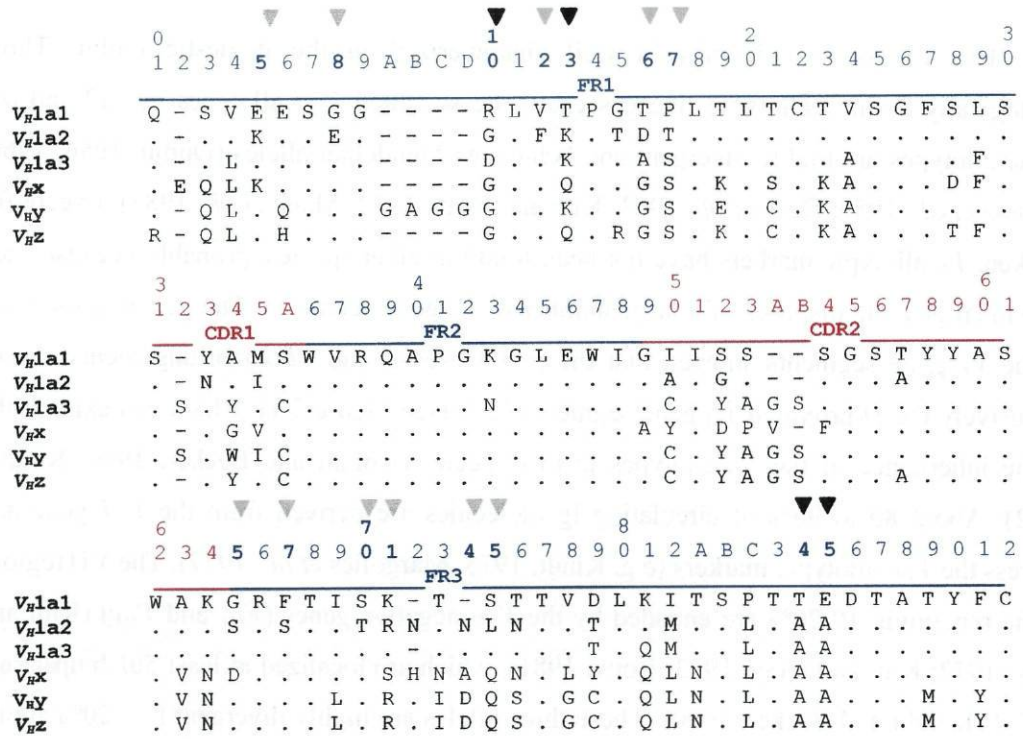



**Figure 3.2.** Amino acid sequences of rabbit germline $V_H$ genes: $V_H1a1$, $V_H1a2$, $V_H1a3$, $V_Hx$ and $V_Hy$. The amino acids that are characteristic of allotypes *a1* and *a2* are marked with black and grey arrows, respectively. The borders of each framework (FR1, FR2 and FR3) and each CDR (CDR1 and CDR2) are defined according to Kabat *et al.*, (1991) numbering. Dots (.) indicate identity with $V_H1a1$ and dashes (-) represents indels (insertions or deletions).

An important contribution to the understanding of the mechanisms that drive the rearrangement process in the rabbit was the study of a rabbit strain, called "Alicia", raised in the Basel Institute, Switzerland. In contrast to normal individuals of the *a2* lineage, the young homozygous *ali/ali* "mutant" Alicia rabbits produced only trace amounts of *a2* molecules and contains mostly Ig encoded by $V_H n$ genes (Kelus and Weiss, 1986; Di Pietro *et al.*, 1990; Chen *et al.*, 1993). The "mutant" rabbits derived from a $V_H 1a2$ parental rabbit in which a 10 Kb segment of genomic DNA, including the $V_H 1$ gene, was deleted (Knight and Becker, 1990). The study of this "mutant" rabbit strain showed that the allelic inheritance of the allotypes in normal rabbit is explained by the preferential usage of only one $V_H$ gene, i.e. the $V_H 1$ gene which is deleted from "ali" genome (Knight and Becker, 1990). The level of B cells that show the allele *a2* specificities increases with age in the "Alicia" rabbit (Kelus and Weiss, 1986; Di Pietro *et al.*, 1990; Alegrucci *et al.*, 1990; Chen *et al.*, 1993; Pospisil *et al.*, 1995). The analysis of nucleotide sequences of the promoter region showed that more than 80% of the *VDJ* rearrangements utilise either the functional $V_H 4$ or $V_H 7$ genes localized upstream of $V_H 1$. The $V_H 4$ and the $V_H 7$ have 7 (out of 11) specific nucleotides associated to the allotype *a2*, while the other nucleotides that characterise *a2* are gained through somatic gene conversion using $V_H 9$ or a $V_H 9$-like germline gene as donor (Shegal *et al.*, 1998; Zhu *et al.*, 1999).

The $V_H$ genes in the rabbit are localized in the chromosome 17 (Medrano and Dutrillaux, 1984). Despite having, like human and mouse, more than 100 $V_H$ genes (approximately 50 % are functional) (Gallarda *et al.*, 1985; Currier *et al.*, 1988), only a small fraction of them have been sequenced (see Figure 3.1).

## Population genetic studies

The study by double immunodiffusion of domestic breeds and wild rabbits of Continental Europe (North of Pyrenean Mountains), Great Britain and overseas showed the presence of allotypes *a1*, *a2* and *a3* only. However, in the French population of Versailles, an allotype with partial reaction against *a3*-specific antiserum was observed, having a gene frequency of less than 5% (van der Loo, 1987, 1993). In all the populations studied the gene frequencies of *a1* and *a3* were similar (≅40%) and higher than *a2* (≅20%). These allelic frequencies reflect the "pecking order" of their relative expression in heterozygote animals (Mage, 1967; Lummus *et al.*, 1967). An excess of homozygotes was detected mainly due to *a2* homozygotes (van der Loo, 1982, 1987; van der Loo and Arthur, 1986).

Radioimmunoassay studies performed in wild rabbit populations from the Iberian Peninsula showed nine variants, *a100* to *a105*, *a107*, *a108* and *a109* (Cazenave *et al.*, 1974; Brézin *et al.*, 1979; Brézin and Cazenave, 1980; Haouas *et al.*, 1987, 1989; Haouas and Benammar-el Gaaied, 1994). The partial amino acid sequence of allotype *a100* showed a close relationship with the allotype *a3* (Tonnelle *et al.*, 1983).

## Indications that the $V_H1$ polymorphism can be trans-specific

The serological analyses of 660 specimens of *Lepus americanus* (snowshoe hare) belonging to 27 populations showed cross-reaction against rabbit anti-*a1*, -*a2* and -*a3* antisera. These allotypes were called *am1*, *am2* and *am3*, respectively (De Poorter, 1984). These names reflect the chronology of their discovery (van der Loo, personal communication). Some analogies with the results that had been described in the rabbit were observed, such as the existence of a less frequent allele (*am3*) with an average gene frequency of 14% and with significant homozygous excess (van der Loo, 1987) (as observed in the rabbit populations for allele *a2*). Su and Nei (1999) considered these data as evidence for the trans-specific nature of the $V_H1$ polymorphism. Furthermore, these authors compared the extent of sequence divergence between the rabbit allotypes *a1*, *a2* and *a3* with that between human and mouse $V_H$ gene sequences and concluded that, assuming a "normal" mutation rate, the $V_H1$ polymorphism has persisted for about 50 My. Since the age of separation between rabbit and *Lepus* is estimated between 13.1 and 29 million years ago (Halanych and Robinson, 1999; Su and Nei, 1999), the allelic lineages present in one species should be more related to some of the alleles expressed in the other species than to their allelic counterparts.

## The diversification of Ab primary repertoire in the rabbit

Among the 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 a number of mammal species (rabbit, sheep, pig and bovine), have less extensive germline repertoires than mice and humans and consequently use strategies of primary repertoire development

that overcome this limitation. In these cases, the primary antibody repertoire is post-rearrangement diversified by two mechanisms: gene conversion and somatic hypermutation.

In chicken, gene conversion is the major mechanism of the Ab binding site diversification. A set of pseudogenes are used as donors and the unique rearranged $V_H/V_L(D)J$ acts as acceptor (Reynaud *et al.*, 1985). In sheep, B cell diversification occurs in early development in the ileal Peyer's patches and results from somatic hypermutation rather than gene conversion (Reynaud *et al.*, 1991). Both processes are used to diversify the Ag-binding repertoire in cattle (Parng *et al.*, 1996; Lucier *et al.*, 1998). Although it is not yet completely proven, gene conversion mechanisms are possibly used in the Ab repertoire diversification in swine (Sun *et al.*, 1996).

Becker and Knight (1990) showed that the rearrangement process in the rabbit involves preferentially a single $V_H$ gene that is diversified mainly through gene conversion using other upstream functional $V_H$ genes (Becker and Knight, 1990; Knight and Crane, 1994). Somatic hypermutation also takes place, distributing several point mutations throughout the entire *VDJ* genes (Short *et al.*, 1991; Weinstein *et al.*, 1994). In this species B cells are generated, proliferate and develop in the bone marrow and in the foetal liver (lymphopoiesis is absent in the adult (Crane *et al.*, 1996)), and migrate to appendix and 0other gut-associated lymphoid tissues (GALT) where they are intensely diversified. The $V_H$ genes upstream of $V_H1$ ($V_H4a1$, $\varphi V_H2a1$, $\varphi V_H3a1$, $V_H4a2$, $\varphi V_H3a2$, $V_H7a2$, $V_H9a2$, etc...) contribute to the expressed $V_H1$ diversification by gene conversion. They evolved to some extent in concert, which resulted in the presence of the allotypic motifs on other $V_H$ gene segments of the chromosome (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 Ab repertoire diversification only happens 1-2 months after birth (Cooper *et al.*, 1968; Weinstein *et al.*, 1994). The experiment of surgically removing the appendix and *sacullus rotundus* on the day of birth and the Peyer's patches at three weeks of age showed that GALT is essential in the generation of the primary antibody repertoire (Vadjy *et al.*, 1998). It has been shown that exogenous factors, such as the intestinal microflora, are also required for the Ab repertoire diversification (Lanning *et al.*, 2000a, 2000b; Sehgal *et al.*, 2002).

## Evolution of $V_H$ gene family

The study of $V_H$ genes from several vertebrate species showed that $V_H$ genes from the same species could belong to different groups. $V_H$ gene sequences have been classified into five different groups (A-E). Those of cartilaginous fishes form a monophyletic group (E). $V_H$ genes of bony-fish form the group D. Groups A, B and C include $V_H$ gene sequences of bony-fish, amphibians, reptiles, birds, and mammal (Tutter and Riblet 1989; Schroeder *et al.*, 1990; Ota and Nei, 1994). The human and mouse possess $V_H$ genes that are very diversified and may cluster with either A, B or C groups (polyphyletic), whereas $V_H$ genes of birds, rabbits, horse, and artiodactyls (cattle, sheep, and swine) are monophyletic. Bird, rabbit and swine each form distinct clusters within group C, whereas horse, cattle and sheep each form distinct clusters within group B. Since $V_H$ genes of different artiodactyl species sequences cluster within different groups (B and C), their common ancestor must have possessed $V_H$ genes from both groups (Sitnikova and Su, 1998) (see Figure 3.3).

Monofunctional multigene families are generally believed to undergo processes of genetic exchange, like gene conversion and unequal crossover, which homogenizes the DNA sequences (Smith *et al.*, 1971; Smith, 1994; Zimmer *et al.*, 1980). This has been called "concerted" evolution. The cconcerted evolution model has been invoked to explain the evolution of $V_H$ genes (Hood *et al.*, 1975; Ohta, 1983). However, the study of $V_H$ gene families in human and mouse (Gojobori and Nei, 1984; Tutter and Ribblet, 1989), and the studies on a much longer evolutionary time scale, showed that the pattern of $V_H$ 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), 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 $V_H$ gene libraries are duplicated and can diverge functionally, while others become pseudogenes or/and 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) (see Figure 3.4).

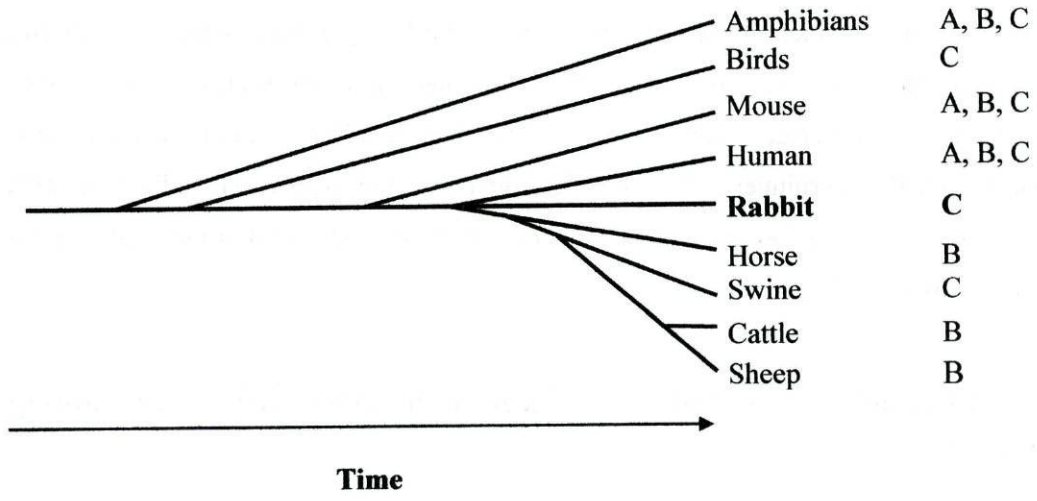



**Figure 3.3.** Schematic representation of $V_H$ gene groups present in several tetrapod species.

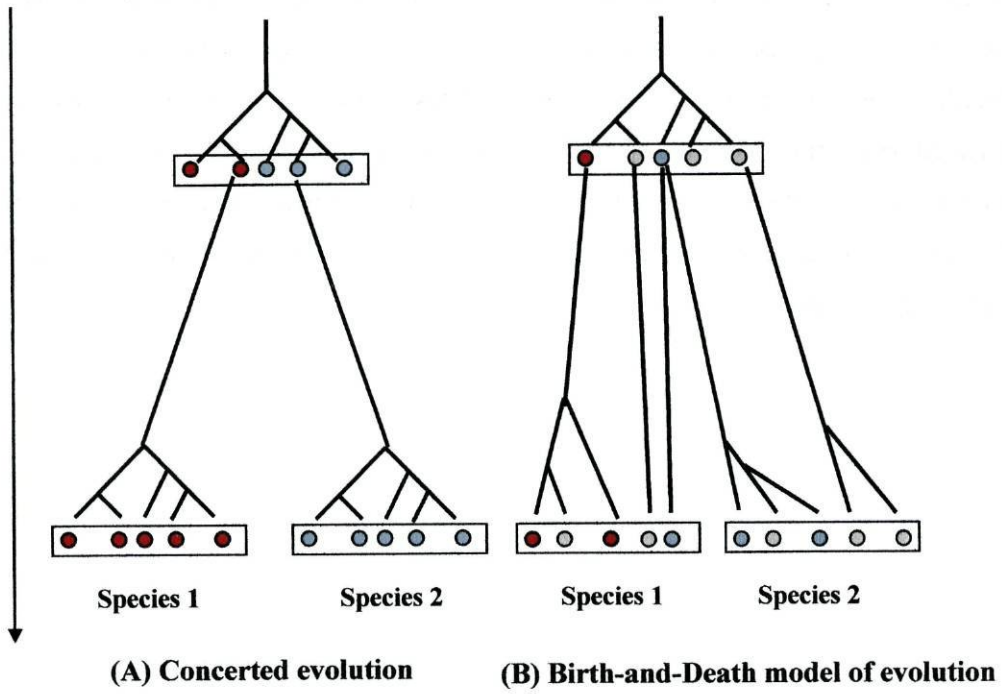



**Figure 3.4.** Schematic representation of two different models of evolution proposed to explain the evolution of monofunctional multigene families. Blue and red circles represent functional genes. Grey circles represent pseudogenes. In the concerted evolution model the genes have high homogeneity within species. In the Birth-and-Death model of evolution genes from one species can be more related to genes present in other species.

According to this model, the tetrapod ancestor possessed three $V_H$ gene groups (A, B, and C) and library contraction events occurred in several lineages independently. In species that have inherited only one group of $V_H$ genes, antibody diversification is mainly due to somatic hypermutation or gene conversion. It seems that the VH gene library contraction is associated with the recruitment of a specific organ for extensive somatic diversification (bursa of Fabricius in chicken; ileal Peyer`s patches in sheep, cattle and probably swine and horse; appendix in rabbit).

## Different hypotheses to explain the large inter-allelic differences among $V_H1$ genes

The large differences between alleles can be explained either by prolonged allele persistence times and/or increased evolutionary rates. If evolutionary rates among the allelic lineages are the same as in reference lineages (rate monotony), the genetic distance reflects the time of divergence between alleles. If on the contrary, evolutionary rates are increased, divergence time between alleles will be overestimated. The knowledge of the diversity of the rabbit $IgV_H$ genes remains fragmentary and concerns only a small number of domestic breeds. These breeds are recent genetic isolates of the subspecies *O. c. cuniculus*, and retained only a fraction of the species diversity. To test the hypothesis of rate monotony we have used two different approaches: 1) a study of the patterns of $V_H$ gene diversity in the Iberian Peninsula (subspecies *O. c. algirus* and *O. c. cuniculus*; 2) sequence determination of $V_H$ genes in other *Leporidae*.

# Material and Methods

## Serological analyses

In this work, 402 specimens from 18 wild rabbit populations from the Iberian Peninsula belonging to the subspecies *Oryctolagus cuniculus algirus* and *O. c. cuniculus* were analysed by double immunodiffusion tests according to van der Loo *et al.*, (1991). The serum of each sample was tested independently against allo-antisera specific for the allotypes *a1*, *a2* and *a3*, by immunodiffusion in 1% agar gel containing 2% polyethylene glycol. In each test a reference was included to monitor the degree of serological similarity of the sample being tested with samples from rabbits expressing the immunizing allotype. The reactions were recorded as: 1) *"1"* (identity reactions) in case of perfect fusion of the precipitation lines of test and reference sera with the antiserum; 2) *"p"* (cross-reactions) in case the reference line spurs over the test line (such spur implies that the tested serum displays only a part of the epitopes recognized by the allo-antiserum on the Ig molecules of the reference allotypes); 3) *"0"* in case of negative test reaction (i.e. absence of precipitation lines). Additionally, using the same test system, the serological analyses were performed in 147 specimens of three different *Lepus* species (*L. granatensis*, *L. europaeus* and *L. capensis*).

The allotype with the negative reaction when tested against allo-antisera specific for the allotypes *a1*, *a2* and *a3* (*000*) was called *"a-blank"*. The gene frequencies of the putative *"a-blank"* allele were calculated assuming Hardy-Weinberg equilibrium and using the formula:

$$(p + q)^2 = p^2 + 2pq + q^2,$$

where $p^2$ is the frequency x of homozygous *"a-blank"*; q corresponds to the sum of gene frequencies of the other alleles; the gene frequency of p (*"a-blank"*) is therefore calculated by $p = \sqrt{x}$, where x represents the relative frequency of *a*-negative individual samples.

## Analysis of genetic diversity

For each population analysed the heterozygosity (Hs) was calculated using the formula:

$$H_S = 1 - \sum_{i=1}^{n} (p_i)^2$$

where $p_i$ is the gene frequency observed at each allele in each population.

To measure the genetic differentiation between groups of populations $F_{ST}$ values were calculated using the formula:

$$F_{ST} = 1 - \frac{\overline{H_S}}{H_T}$$

$H_T$ is the total heterozygosity in the group of populations and is obtained with the mean allele frequencies over localities. $\overline{H_S}$ is the average of $H_S$ values calculated for each population.

Cytonuclear disequilibria (*a* locus vs mtDNA) values were determined according to Asmussen and Arnold (1991).

## Sequencing of *VDJ* gene segments and germline $V_H$ gene segments

Two different types of tissues were used: liver, to obtain genomic DNA, and spleen, to extract total RNA, to be used as template for synthesizing cDNA. Standard DNA and RNA extraction methods were used.

### *Sequencing of VDJ gene segments expressed in* Lepus *specimens*

The samples were collected from wild specimens in Alcochete, South of Portugal (17 specimens of rabbit belonging to the subspecies *O. c. algirus*, and seven specimens of *Lepus granatensis*) and in Austria (two specimens of *Lepus europaeus*). The amplification of the genes encoding the *IgVH* regions (Figure 3.5) was performed using the primers VHRPS and JH-B described by Zhu *et al.*, 1999. All PCR amplifications proceeded for 30 cycles (each cycle: 94°C for 45 s, 60°C for 45 s, and 72°C for 60 s). PCR fragments with approximately 500 bp in length were gel-purified and then cloned and sequenced.

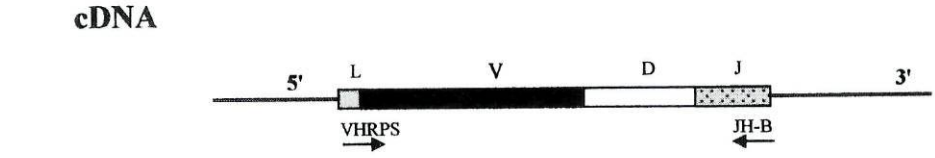



**Figure 3.5.** Schematic representation of the gene fragments that were amplified: Leader (L), Variable (V), Diversity (D) and Joining (J) segments of the Heavy chain; locations of primer homology are indicated (5' primer VHRPS and 3' primer JH-B).

### *Sequencing of germline $V_H$ gene segments*

Sequences from single specimens of *Ochotona princeps* (Pika), *Sylvilagus floridanus*, *Lepus americanus*, *Lepus granatensis*, and *Oryctolagus cuniculus algirus* were obtained. Germline $V_H$ genes were PCR amplified with a 5' primer, VHRPS (localized in the leader segment), and a 3' primer, localized in the end of the V segment (OaFR3: 5`-GAGCTCACCTGAGAGACGGTGACCA-3`) (see Figure 3.6) and proceeded for 30 cycles (each cycle: 94°C for 45 s, 60°C for 45 s, and 72°C for 60s). PCR fragments with approximately 550bp in length were obtained by gel electrophoresis and then cloned and sequenced.

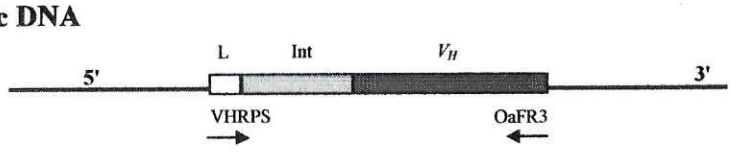



**Figure 3.6.** Schematic representation of the genomic DNA fragments that were amplified. This fragment comprises the complete Leader (L), Intron (Int) and Variable ($V_H$) gene segments of Heavy Chain. Regions of primer homology are indicated by horizontal arrows.

## Phylogenetic analysis

The sequence alignments were made using the pileup program (GCG software), followed by visual adjustments. To construct a phylogenetic tree we used sequences obtained in this work and sequences available in the Genbank (accession numbers are presented in table 3.1): 1) Genbank rabbit germline $V_H$ genes; 2) rabbit cDNA $V_H$ gene sequences representative of each allotype; 3) hares cDNA $V_H$ gene sequences; 4) other functional germline $V_H$ gene sequences obtained in this work; 5) Genbank $V_H$ gene sequences classified in the same group of the rabbit, the group C: human (class 3), mice, camel and domestic duck. As outgroup we used Genbank $V_H$ gene sequences classified in

the group B: human (class 4), mouse and cattle. The reliability of the trees was examined by the bootstrap (Felsenstein, 1985) and the interior-branch (Rzhetsky and Nei, 1992; Sitnikova, 1996) methods, which produced the bootstrap probability (BP) and confidence probability (CP) values, respectively, for each interior branch in the tree. All data analyses were conducted by using the computer program MEGA (Kumar *et al.*, 2001).

**Table 3.1**

Accession numbers for each $V_H$ gene used. The sample code and the major $V_H$ gene group to which each sequence belongs are indicated.

| **Species** | **$V_H$ gene identification** | **Genbank (accession number)** |
|---|---|---|
| **Mammals** | | |
| *Oryctolagus cuniculus* | $V_H$1a1 (group C) | M93171 |
| | $V_H$1a2 (group C) | M93172 |
| | $V_H$1a3 (group C) | M93173 |
| | $V_H$4a1 (group C) | M93181 |
| | $V_H$4a2 (group C) | M93182 |
| | $V_H$4a3 (group C) | M93183 |
| | $V_H$7a2 (group C) | U51027 |
| | $V_H$9a2 (group C) | U51029 |
| | $V_H$x (group C) | L03846 |
| | $V_H$y (group C) | L03890 |
| | $V_H$z (group C) | AF264469 |
| *Homo sapiens* | Ig$V_H$3 (group C) | M99666 |
| | " | M99672 |
| | " | Z18900 |
| | " | M99675 |
| | " | M99679 |
| | " | M99682 |
| | Ig$V_H$4 (group B) | X05713 |
| | " | L10088 |
| | " | M29811 |
| *Mus musculus* | $V_H$12 (group B) | M22439 |
| | $V_H$13 (group C) | X55935 |
| *Camelus dromedarius* | (Group C) | AF000604 |
| *Bos taurus* | 174 (group B) | U55174 |
| | 172 (group B) | U55172 |
| | 170 (group B) | U55170 |
| | 168 (group B) | U55168 |
| | 166 (group B) | U55166 |
| | 164 (group B) | U55164 |
| **Birds** | | |
| *Anas platyrhynchos* | 42M (group C) | X65218 |
| | 31M (group C) | X65219 |

# Results and Discussion

## Genetic diversity in Iberian rabbit populations

### *Population genetic studies*

In the serological analyses six different allotypes were distinguished in wild rabbit populations from the Iberian Peninsula (*a1 (100)*, *a2 (010)*, *a3 (001)*, *a1v (p00)*, *a3v (00p)* and *"a-blank" (000)*). Estimates of the gene frequencies of these alleles and heterozygosity values are shown in Table 3.2. The geographical distribution of alleles is presented in Figure 3.7.

In this work, we divided the Iberian wild rabbit populations into three groups: 1) Southwest, where the mitochondrial lineage *A* is predominant; 2) populations in the hybrid region; 3) Northeast, where the mtDNA lineage *B* is prevalent. The *"a-blank"* allele is present in all Iberian populations (except Alicante). However, the gene frequency of this allotype decreases from southwestern (average gene frequency of 0.56) to northeastern populations (average gene frequency of 0.19). The allele *a1* is present in all Iberian Peninsula but, in contrast to the *"a-blank"* allele, its gene frequency increases from Southwest populations (average gene frequency of 0.27) to Northeast populations (average gene frequency of 0.45). The same pattern of variability is observed in the alleles *a2* and *a3*, hat triplicate the gene frequency from the Southwest (0.04 and 0.06, respectively) to the Northeast (0.12 and 0.18, respectively). The rare alleles *a3v* and *a1v* do not show any particular geographical pattern (note that serological identity between "variant" allotypes among regions has not been tested).

In rabbit populations from the recent area of distribution (France, Great Britain, Australia) the allotype *"a-blank"* is absent, and the alleles *a1*, *a2* and *a3* are present in all populations (Table 3.2).

**Table 3.2**

Gene frequencies obtained in rabbit populations at *a* locus. The immunodiffusion reactions obtained against specific antiserum for the allotypes *a1*, *a2* and *a3*, respectively, are shown in brackets. For identification of Australian and British locations see van der Loo (1987).

| | n | *a1* (100) | *a2* (010) | *a3* (001) | *a1v* (p00) | *a2v* (0p0) | *a3v* (00p) | *"a-blank"* (000) | *H* | Reference |
|---|---|---|---|---|---|---|---|---|---|---|
| **Inside Iberian Peninsula** | | | | | | | | | | |
| ***Southwest*** | | | | | | | | | | |
| Portimão | 18 | 0,25 | ---- | 0,06 | 0,03 | ---- | ---- | 0,66 | ***0.497*** | **This work** |
| Santarém | 27 | 0,18 | ---- | 0,10 | 0,08 | ---- | 0,08 | 0,54 | ***0.631*** | " |
| Alcochete | 17 | 0.28 | 0.09 | 0.06 | ---- | ---- | 0.03 | 0.54 | ***0.617*** | " |
| Idanha | 18 | 0,33 | 0,14 | ---- | 0,03 | ---- | ---- | 0,50 | ***0.621*** | " |
| Cabreira | 19 | 0,22 | 0,07 | 0,13 | 0,03 | ---- | ---- | 0,55 | ***0.626*** | " |
| Las Lomas | 40 | 0,38 | 0,03 | 0,09 | 0,13 | ---- | 0,10 | 0,27 | ***0.747*** | " |
| Huelva | 23 | 0,13 | 0,02 | 0,06 | 0,10 | ---- | ---- | 0,69 | ***0.493*** | " |
| Doñana | 19 | 0,26 | 0,05 | 0,08 | ---- | ---- | ---- | 0,61 | ***0.551*** | " |
| Ciudad Real | 20 | 0,32 | 0,03 | ---- | 0,03 | ---- | ---- | 0,62 | ***0.511*** | " |
| Infantado | 17 | 0,29 | 0,03 | ---- | 0,03 | ---- | ---- | 0,65 | ***0.492*** | " |
| Vila Viçosa | 18 | 0.31 | ---- | 0.11 | ---- | ---- | 0.05 | 0.53 | ***0.608*** | " |
| ***average unweighted*** | **236** | ***0,27*** | ***0.04*** | ***0.06*** | ***0.04*** | **—** | ***0,02*** | ***0,56*** | ***0.604*** | |
| ***Populations in hybrid zone*** | | | | | | | | | | |
| Amoladeras | 11 | 0,40 | ---- | ---- | ---- | ---- | ---- | 0,60 | ***0.480*** | " |
| Toledo | 22 | 0,26 | ---- | 0,26 | ---- | ---- | ---- | 0,48 | ***0.634*** | " |
| Bragança | 15 | 0,39 | 0,23 | ---- | 0,13 | ---- | ---- | 0,25 | ***0.716*** | " |
| ***average unweighted*** | **48** | ***0.35*** | ***0,08*** | ***0,09*** | ***0,04*** | **—** | **—** | ***0,44*** | ***0,666*** | |
| ***Northeast*** | | | | | | | | | | |
| Tarragona | 20 | 0,49 | 0,08 | 0,08 | 0,10 | ---- | ---- | 0,31 | ***0.641*** | " |
| Alicante | 19 | 0,58 | 0,08 | 0,34 | ---- | ---- | ---- | ---- | ***0.542*** | " |
| Navarra | 59 | 0,40 | 0,14 | 0,06 | 0,10 | ---- | 0,05 | 0,25 | ***0.742*** | " |
| Lérida | 20 | 0,33 | 0,18 | 0,25 | 0,03 | ---- | ---- | 0,21 | ***0.751*** | " |
| ***average unweighted*** | **118** | ***0,45*** | ***0,12*** | ***0,19*** | ***0,06*** | **—** | ***0,01*** | ***0,20*** | ***0.709*** | |
| ***total (18 populations)*** | ***402*** | ***0.33*** | ***0.07*** | ***0.09*** | ***0.07*** | **—** | ***0,01*** | ***0.45*** | ***0.670*** | |
| **Outside Iberian Peninsula** | | | | | | | | | | |
| *France* | | | | | | | | | | |
| Perpignan | 30 | 0,40 | 0,20 | 0,40 | ---- | ---- | ---- | ---- | ***0,640*** | **van der Loo, 1987** |
| Camargue | 40 | 0,45 | 0,18 | 0,37 | ---- | ---- | ---- | ---- | ***0,628*** | " |
| Versailles | 35 | 0,37 | 0,21 | 0,40 | ---- | ---- | 0,02 | ---- | ***0.659*** | " |
| Ile de France | 147 | 0.39 | 0.21 | 0.40 | ---- | ---- | ---- | ---- | ***0.644*** | " |
| ***average unweighted*** | **252** | ***0,40*** | ***0,20*** | ***0,39*** | **—** | **—** | ***0,005*** | **—** | ***0,644*** | |
| *Australia* | | | | | | | | | | |
| BB | 23 | 0.37 | 0.13 | 0.50 | ---- | ---- | ---- | ---- | ***0596*** | " |
| BW | 39 | 0.26 | 0.29 | 0.45 | ---- | ---- | ---- | ---- | ***0.646*** | " |
| CH | 83 | 0.49 | 0.02 | 0.49 | ---- | ---- | ---- | ---- | ***0.519*** | " |
| CN | 71 | 0.46 | 0.18 | 0.36 | ---- | ---- | ---- | ---- | ***0.626*** | " |
| UN | 21 | 0.44 | 0.16 | 0.40 | ---- | ---- | ---- | ---- | ***0.621*** | " |
| UR | 77 | 0.54 | 0.31 | 0.15 | ---- | ---- | ---- | ---- | ***0.589*** | " |
| GC | 575 | 0.58 | 0.09 | 0.33 | ---- | ---- | ---- | ---- | ***0.547*** | " |
| ***average unweighted*** | **889** | ***0.45*** | ***0.17*** | ***0.38*** | **—** | **—** | **—** | **—** | ***0.624*** | |
| *Great Britain* | | | | | | | | | | |
| DO | 89 | *0.43* | *0.03* | *0.54* | ---- | ---- | ---- | ---- | ***0.523*** | " |
| GL | 22 | *0.61* | *0.05* | *0.34* | ---- | ---- | ---- | ---- | ***0.510*** | " |
| HI | 75 | *0.45* | ---- | *0.55* | ---- | ---- | ---- | ---- | ***0.495*** | " |
| PO | 117 | 0.42 | 0.23 | 0.35 | ---- | ---- | ---- | ---- | ***0,648*** | " |
| *average unweighted* | **303** | ***0.45*** | ***0.10*** | ***0.45*** | **—** | **—** | **—** | **—** | ***0,585*** | |
| ***total (15 populations)*** | ***1444*** | ***0.44*** | ***0.15*** | ***0.40*** | **—** | **—** | ***0.001*** | **—** | ***0.618*** | |
| ***total (33 populations)*** | ***1846*** | ***0.38*** | ***0.11*** | ***0.24*** | ***0.03*** | **—** | ***0.008*** | ***0.25*** | ***0.728*** | |

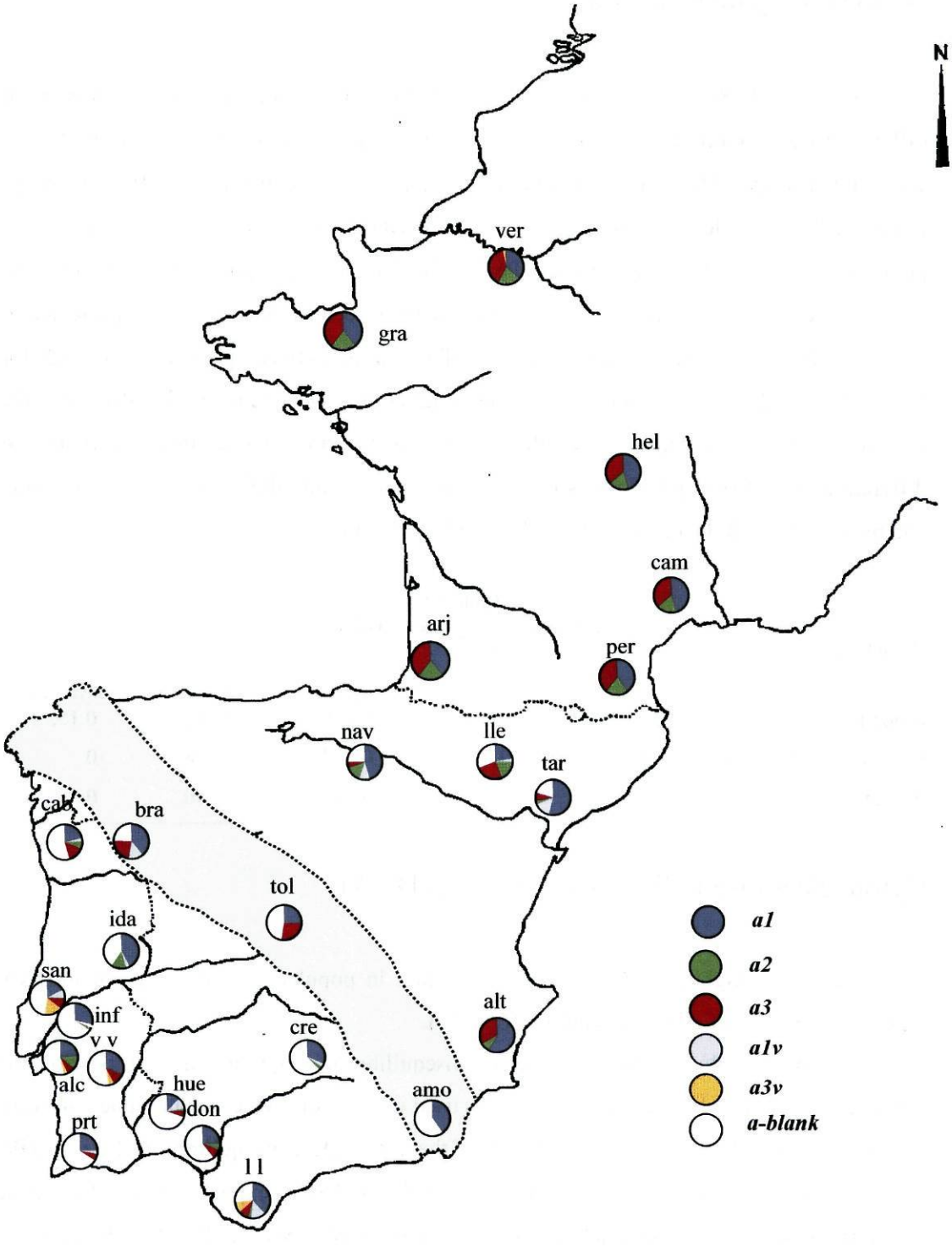



**Figure 3.7.** Geographic distribution of *a* locus alleles in wild rabbit populations of the Iberian Peninsula and France. Iberian Penisula populations: 1) Southwest: prt - Portimão; san - Santarém; alc - Alcochete; ida - Idanha; cab - Cabreira; l l - Las Lomas; hue - Huelva; don - Doñana; cre - Ciudad Real; inf - Infantado; v v - Vila Viçosa; 2) populations in the hybrid zone: amo - Amoladeras; tol - Toledo; bra - Bragança; 3) Northeast: tar-Tarragona; alt- Alicante; nav - Navarra; lle - Lerida. France: per - Perpignan; cam - Camargue; ver - Versailles; arj - Arjuzanx; gra - grax; hel - Helene. The hybrid zone is represented by a grey-shaded area.

## Analysis of genetic diversity

Wright (1978) proposed a qualitative distinction of four classes of population differentiation based on $F_{ST}$ values: low (<0.05), moderate (0.05-0.15) high (0.15-0.25) and very high (>0.25). The values of total $F_{ST}$ obtained vary in the three different groups considered. It is moderate within the Iberian Peninsula populations (0.097), low among the populations from outside the Iberian Peninsula (0.039) and high over all populations (0.176). The $F_{ST}$ values obtained for each allele in the three different groups are presented in table 3.3. Within the Iberian Peninsula the differentiation between populations is high for the *"a-blank"* allele and moderate for all the other alleles. Outside of Iberian Peninsula the $F_{ST}$ values are low, except for the allele *a2* that is moderate. Using all populations, the differentiation between populations is very high for the blank allele, high for the *a3* allele and moderate for all the other alleles (*a1*, *a2*, *a1v* and *a3v*).

**Table 3.3**
$F_{ST}$ values obtained at each allele

| Populations | alleles | | | |
|---|---|---|---|---|
| | *a1* | *a2* | *a3* | *"a-blank"* |
| Inside Iberian Peninsula (18 populations) | 0.053 | 0.078 | 0.110 | 0.152 |
| Outside Iberian Peninsula (15 populations) | 0.029 | 0.067 | 0.038 | 0 |
| All populations (33 populations) | 0.055 | 0.091 | 0.180 | 0.391 |

## Cytonuclear disequilibria (*a* locus *vs* mtDNA)

Figure 3.8 highlights the positive correlation in population frequencies of mtDNA type *B* and the sum of alleles *a1* and *a2* ($R^2$=0.71).

For the calculus of the cytonuclear disequilibrium ($D_T$), it was not possible to calculate the value of $D_S$, because the samples used in this work were not the same that were used in the mtDNA study. Previous studies of the cytonuclear disequilibria in Iberian wild rabbit populations (van der Loo *et al.*, 1999) have shown that, as expected, within localities, equilibrium prevails between nuclear markers and maternal types (i.e. Ds =0). The value of $D_T$ obtained in this work was therefore equivalent to interlocality covariances in gene frequencies.

The frequency of the nuclear allele *"a-blank"* and the mtDNA type are indicated in Table 3.4. The *"a-blank"* allele is significantly correlated to the maternal lineage *A* (cytonuclear disequilibrium $D$ *(mtDNA –type A,"a-blank")* $= 0.06$; $X^2_1 = 17$).

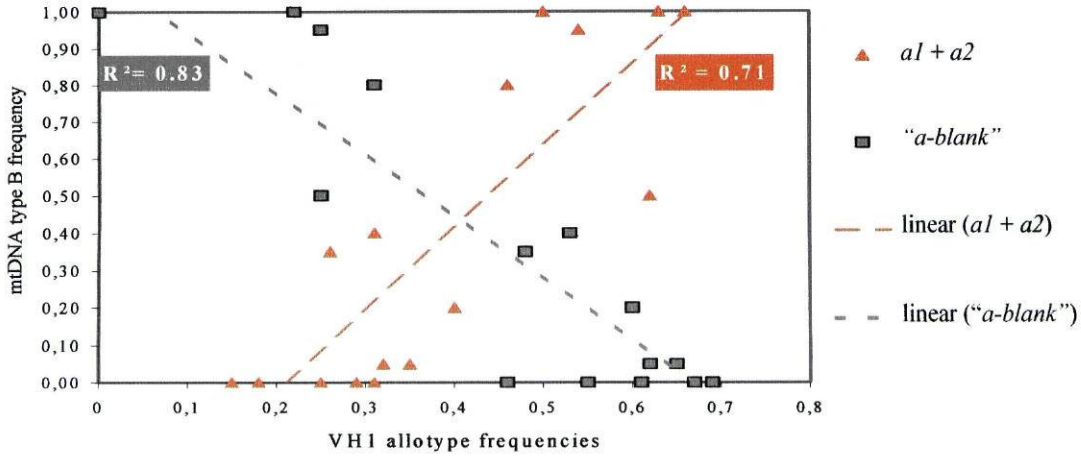



**Figure 3.8.** Frequency correlations between *a* locus alleles and mtDNA type *B*. $R^2$-value measured the correlations *"a-blank"*/mtDNA *B* and (*a1+a2*)/mtDNA *B*.

**Table 3.4**

Gene frequencies of *"a-blank"* allele and mtDNA type *A* in wild rabbit populations from Iberian Peninsula.

| | n | ***"a-blank"* allele** | mtDNA *A* |
|---|---|---|---|
| **Inside Iberian Peninsula** | | | |
| ***Southwest*** | | | |
| Portimão | 18 | **0.66** | 1 |
| Santarém | 27 | **0.54** | 1 |
| Alcochete | 17 | **0.54** | 1 |
| Idanha | 18 | **0.50** | 1 |
| Cabreira | 19 | **0.55** | 1 |
| Las Lomas | 40 | **0.27** | 1 |
| Huelva | 23 | **0.69** | 1 |
| Doñana | 19 | **0.61** | 1 |
| Ciudad Real | 20 | **0.62** | 0.95 |
| Infantado | 17 | **0.65** | 0.95 |
| Vila Viçosa | 18 | **0.53** | 0.60 |
| ***average unweighted*** | **236** | ***0.561*** | ***0.95*** |
| ***Population in hybrid zone*** | | | |
| Amoladeras | 11 | **0.60** | 0.80 |
| Toledo | 22 | **0.48** | 0.65 |
| Bragança | 15 | **0.25** | 0.50 |
| ***average unweighted*** | **48** | ***0.443*** | ***0.65*** |
| ***Northeast*** | | | |
| Tarragona | 20 | **0.31** | 0.20 |
| Alicante | 19 | **0.00** | 0.00 |
| Navarra | 59 | **0.25** | 0.05 |
| Lérida | 20 | **0.21** | 0.00 |
| ***average unweighted*** | **118** | ***0.193*** | ***0.07*** |
| ***average*** | ***402*** | ***0.459*** | **0.70** |

# Article 3

P.J. Esteves, D. Lanning, S.K. Zhai, N. Ferrand, K.L. Knight and W. van der Loo (submitted).

Allelic variation at the $V_Ha$ locus in natural populations of rabbit (*Oryctolagus cuniculus*, L.)

# Allelic variation at the $V_H a$ locus in natural populations of rabbit (*Oryctolagus cuniculus*, L.)

P.J. Esteves*,¶, D. Lanning†, S.K. Zhai†, N. Ferrand*, K.L. Knight† and W. van der Loo*¶

*Centro de Investigação em Biodiversidade e Recursos Genéticos (CIBIO-UP) and Departamento de Zoologia e Antropologia - Faculdade de Ciências da Universidade do Porto, Porto, Portugal.
¶Institute of Molecular Biology and Biotechnology, Vrije Universiteit Brussel, Belgium.
†Department of Microbiology and Immunology, Loyola University Chicago, Maywood, IL, USA.

## Abstract

The large interallelic distances between the three rabbit Ig $V_H a$ lineages, *a1*, *a2*, and *a3*, have suggested that the persistence time of the $V_H a$ polymorphism could amount to 50 million years, which is much longer than that of *MHC* polymorphisms. Rabbit originated in the Iberian Peninsula where two subspecies coexist, one of which is confined to Southwestern Iberia (*O. c. algirus*). We studied the $V_H$ loci in the original species range to obtain a better understanding of the evolutionary history of this unusual polymorphism. Serological surveys revealed that a majority of sera of the subspecies *algirus*, when tested with $V_H a$-locus specific alloantisera, showed either crossreactivity (*"a-positive"* variants) or no reaction at all (*"a-blank"*). Using RT-PCR, we determined 120 sequences of rearranged $V_H$ genes expressed in seven *algirus* rabbits that were typed as either *"a-positive"* or *"a-blank"*. The data show that the $V_H$ genes transcribed in *"a-positive"* rabbits are closely related to one or the other of the $V_H 1$ alleles of domestic rabbits. In contrast, *"a-blank"* rabbits were found to preferentially use $V_H$ genes that, although clearly related to the known $V_H a$ genes, define a new major allotypic lineage, designated *a4*. The *a4* sequences have hallmark rabbit $V_H a$ residues together with a number of unprecedented amino acid changes in FR2 and FR3. The net protein distances between the $V_H$-*a4* and the $V_H$-*a1*, *a2*, and *a3* lineages, was 20%, 29%, and 21% respectively. We conclude that at least four distantly related lineages of the rabbit $V_H a$-locus exist, one of which seems to be endemic in the Iberian range.

# Introduction

In rabbit, the gene locus controlling the variable region of the Ig heavy chain (*IGVH, or* $V_H$) is unusual by showing extensive allelic diversity. Initially, this was discovered by serological methods, which revealed that $V_H$ regions of nearly all rabbit antibodies can be placed, on the basis of the allotypic motifs, into one of three groups, namely, *a*, *x*, or *y* (Kim and Dray, 1972, 1973; Roux *et al.*, 1981). Three "*a*" allotypic lineages have been described, and their Mendelian inheritance was established by extensive breeding studies (Oudin, 1960; Dray *et al.*, 1963; Kim and Dray, 1972). These so called *a1*, *a2*, and *a3* allotypes have characteristic amino acid sequence differences in FR1 (Kabat positions 5, 8, 10, 12, 13, 16, and 17) and FR3 (positions 65, 67, 70, 71, 74, 84 and 85; Mage *et al.*, 1984). These differences reflect allelic variation at the *D*-proximal $V_H$ gene, $V_H1$, which accounts for approximately 80% of *VDJ* gene rearrangements (Knight and Becker, 1990). Unlike for other mammalian species, where $V_H1$, $V_H2$, $V_H3$, *etc.* refer to gene "families" with pronounced sequence similarity, in rabbit, $V_H$ gene segments are numbered according to their relative position on the chromosome of a given allotype (Note: for convenience "$V_H$ gene segments" will be written below "$V_H$ genes"). Thus, $V_H1$-*a1* is the 3'-most $V_H$ gene of the *a1* haplotype; $V_H7$-*a2* is immediately 5' of the $V_H6$-*a2* gene of the *a2* haplotype, *etc.* Only a small number of the more than 100 rabbit germline $V_H$ genes have been sequenced (Gallarda *et al.*, 1985). Besides $V_H1$, other $V_H$ genes have been identified that encode $V_H1$-associated "allotypic" motifs (including $V_H4$-*a1*, $\varphi V_H2$-*a1*, $\varphi V_H3$-*a1*, $V_H4$-*a2*, $\varphi V_H3$-*a2*, $V_H7$-*a2*, $V_H9$-*a2*). They contribute to *VDJ* gene diversification by gene conversion (Becker and Knight, 1990) and evolved to some extent in concert with the neighboring $V_H1$ gene, defining therefore a haplotype polymorphism. The group of these genes will be indicated as "$V_Ha$" or "*a-positive*", to be distinguished from $V_Hn$ genes, which do not encode $V_H1$-associated allotypic markers ($V_Hx$, $V_Hy$ and $V_Hz$, or "*a*-negative"; Kelus and Weiss, 1986). The latter genes are situated more than 50 Kb upstream of the $V_Ha$ genes and contribute to less than 20% of rabbit antibodies (Di Pietro *et al.*, 1990; Short *et al.*, 1991). In this study, the concept of "$V_Ha$-genes", which depends on serology, will be extended to "$V_H$ genes showing significant monophyletic clustering with one or more known $V_Ha$-genes in phylogenetic inference analyses based upon DNA or protein sequences". Newly described $V_H$ genes sharing derived character states with the $V_Ha$ genes can then be classified as such, independent of serology. $V_Ha$ genes of the *aj* haplotype will be designated as $V_Ha$-*aj*, where *j* = 1, 2, 3, 4, etc.

Rabbit Ig allotypes are among the first protein polymorphisms ever described (Kelus and Gell, 1967), but after 50 years of research, we still do not understand its *raison d'être*. The problem posed by the Mendelian inheritance of genetic markers at a multi-gene locus was resolved by the finding that rabbit preferentially uses the $V_H1$ gene (Knight and Becker, 1990), and further, by the mechanism of concerted evolution (Smith, 1976). However, neither preferential expression nor concerted evolution can explain how, or why, such large differences between allelic lineages have evolved (concerted evolution being a mechanism which tends to homogenise the gene pool). Large differences between alleles are believed to be the outcome of diversity enhancement selection, which can favor both prolonged allele persistence times and increased evolutionary rates. According to an evolutionary analysis by Su and Nei (1999), the large genetic distances between the three $V_H1$ alleles suggest allele persistence times of about 50My, which is an order of magnitude larger than average mammalian speciation times. Both prolonged persistence times and increased evolutionary rates imply that this population diversity must fulfill some crucial function (cf. Raman and Knight, 1992; Pospisil and Mage, 1998; Mage and Pospisil, 2000). However, the former hypothesis has specific and profound population genetic implications in regard to founder population sizes, which must always be large enough to contain each of the different alleles. In view of the importance of the questions raised by the rabbit IgH *a*-locus polymorphism for half a century, it might be surprising that data on the rabbit $V_H$ genes remain fragmentary and concern only a small number of domestic breeds. These breeds are recent genetic isolates (<200 years) of the subspecies *O. c. cuniculus*, while the genus originated some 4-6 My ago on the Iberian Peninsula (Lopez-Martinez, 1989; Callou, 1995). Today, this area is inhabited by two subspecies, *O. c. cuniculus* and *O. c. algirus*, which, according to mitochondrial DNA data, could be separated by some 2 My (for details see Branco *et al.*, 2000, 2002).

Serological studies of Cazenave *et al.*, (1987) had already indicated that wild rabbits from the Iberian Peninsula can express $V_Ha$ allotypes that differ from those occurring in domestic breeds. A large number of alleles would not be supportive of extremely long allele persistence times, because it is unlikely that many allelic lineages would be maintained throughout numerous speciation steps. However, the serological characterization suggested that the "wild-type" allotypes could represent more recent variations occurring within one of the 3 lineages of domestic rabbits (for convenience referred to as "domestic" allotypes). This was also supported by partial amino acid sequence of one of these allotypes (*a1*00), which showed a close relationship with the *a3* lineage (i.e. 4 a. a. differences in the FR

regions; Tonelle *et al.*, 1983). When we analysed $V_H$ allotypes in wild rabbits from the Iberian Peninsula, a large fraction of samples from the Southwestern areas did not show any cross-reactivity with $V_Ha$-specific allo-antisera. As illustrated in Figure 1, this phenotype was correlated with genetic markers characteristic of the subspecies *O. c. algirus*, and in particular to the mtDNA-type *A* (Hardy *et al.*, 1995; Monnerot *et al.*, 1994). The cytonuclear disequilibrium between the cytotype *A* and the *"a-blank"* allotype (nucleotype), when estimated according to Asmussen and Arnold (1991), was highly significant ($X^2_1 = 36$; data not shown). Among the possible explanations were (1) these rabbits preferentially use $V_Hx$ or $V_Hy$ genes, possibly because the *D*-proximal $V_H$ genes were reshuffled, damaged or deleted (*cf.* Kelus and Weiss, 1986; Knight and Becker, 1990) or (2) they express a $V_Hl$ allotype that differs markedly from those occurring in domestic breeds.

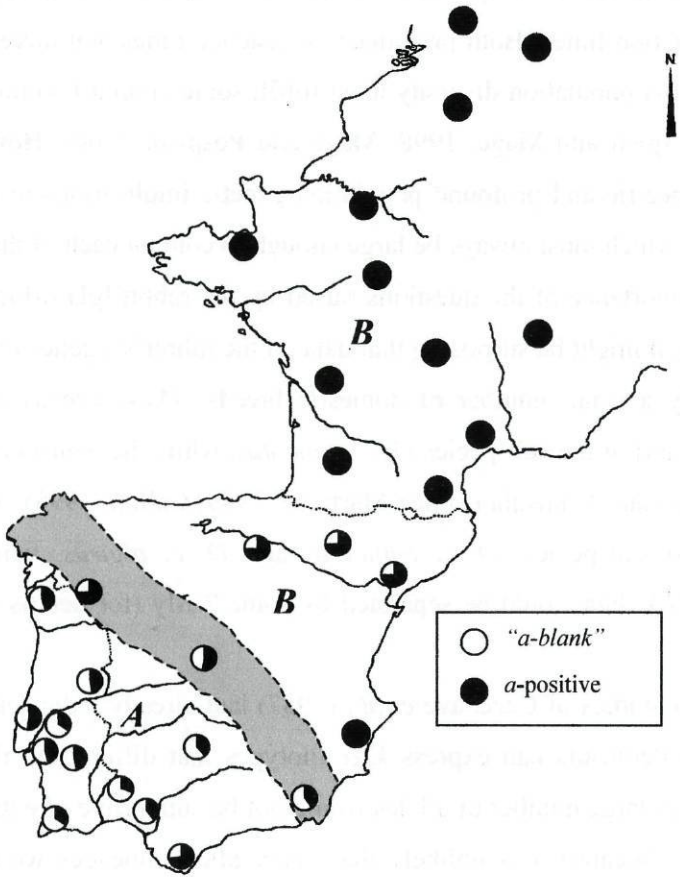



**Figure 1** - Geographic distribution of $V_Ha$ locus allotype frequencies determined for populations of European wild rabbit (*Oryctolagus cuniculus*) from the Iberian Peninsula, France, and Belgium. The subspecies *O. c. algirus* occupies the Southwestern part of Iberia, which is indicated by "*A*". Rabbits of the Northeastern 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 grey. The colouring of the disks reflects the relative allele frequencies per locality, analysed as *a*-allele locus with *"a-blank"* (white) and "*a*-positive" (black colouring) alleles.

We have sequenced the $V_H$ genes expressed in eight wild specimens of the subspecies *O. c. algirus*, four of which were typed *"a-blank"*. The data show that these latter rabbits preferentially express $V_H$ genes, which define a fourth, major lineage of $V_H1$ allotypes (hypothesis 2). The lack of cross-reactivity with allotype-specific alloantisera can be explained by the observed amino acid changes. At the same time, the sequences obtained from *"a-positive" O. c. algirus* rabbits suggest that variations have also occurred within the major $V_H1$ lineages.

## Material and Methods

### *PCR amplification, cloning and nucleotide sequencing of rearranged VDJ genes*

Spleen and serum samples were obtained from adult wild rabbits (*O. c. algirus)* that were collected within the military domain of Alcochete (Ribatejo, Portugal). Serum samples were analysed serologically by double immunodiffusion to determine a1, a2 and a3 *VH* allotypes. Total RNA was isolated from frozen spleen samples with TRIzol, following the manufacturer's instructions (GibcoBRL, Grand Island, NY). First strand cDNA was synthesized from 3µg of RNA using oligo(dT) as a primer (Krug and Berger, 1987). *VDJ* gene rearrangements were PCR-amplified using primers specific for conserved regions in the $V_H$ leader and in the $J_H$ gene segments (Zhu *et al.*, 1999). PCR amplifications proceeded for 30 cycles (each cycle: 94°C for 45 s, 60°C for 45 s, and 72°C for 60 s). PCR products with approximately 500 bp in length were gel-purified and cloned into the pGEM-TEasy vector (Promega, Madison, WI). Nucleotide sequences of the cloned *VDJ* genes were determined using an automated ABI Prism 310 sequencer with Big Dye-labelled terminators (Perkin-Elmer Applied Biosystems, Foster City, CA).

### *Phylogenetic studies*

The $V_H$ gene nucleotide sequences were aligned using the computer program pileup (GCG package, Wilkinson, as provided by the Belgian EMBnet Node (BEN; *http://www.be.embnet.be*) and Clustal W (Thompson *et al.*, 1994) as provided by the freeware DAMBE package (Xia and Xie, 2001; *http://aix1.uottawa.ca/~xxia*). The alignments were corrected upon visual inspection using Genedoc (Nicholas and Nicholas 1997, *http://www.psc.edu/biomed/genedoc*). Phylogenetic analyses were performed using the MEGA 2.1 computer program (Kumar *et al.*, 2001; *http://www.megasoftware.net*).

Evolutionary distances between nucleotide or protein sequences were estimated using uncorrected p-distance. For nucleotide sequences, we also show results using Kimura's 2-parameter option. We prefer the former approach because it makes a minimal number of assumptions, and according to Nei and Kumar (2000), it gives better results when large numbers of relatively short sequences are compared. Phylogenetic trees were constructed using the neighbour-joining (NJ) method (Saitou and Nei, 1987). For the construction of the phylograms we included a sample of published diversified rabbit $V_H$ gene sequences representing the three domestic allotypes (Accession numbers AF264476, AF264478-81, AF264493, AF264496, AF264499, AF264528, AF264544, AF264545, AF264546, AF264571-73, AF264577, AF264583, AF264585, AF264587, AF264589, AF264592, AF264595-96). The reliability of the phylogenetic tree was tested by determining the Bootstrap values (BP; Felsenstein, 1985) and Confidence Probability (CP; Rhezetsky and Nei, 1992; Sitnikova, 1996) values. For sequence pairs under comparison, alignment gaps were excluded from analysis (pairwise deletion option in MEGA). The molecular clock was tested by the relative rate test (Tajima, 1993), as provided in the MEGA package.

*Identification of novel genetic variants*

For discerning the germline $V_H$ genes from groups of diversified $V_H$ sequences, the "*algirus*" sequences were aligned with all known rabbit $V_H$ germline genes. The "*algirus*" sequences often show some systematic variation that differentiates them from sequences obtained from domestic rabbits. For the vast majority of sequences of "*a-blank*" rabbits, this "unprecedented" variation clearly indicates novel variants of $V_H$ genes in *algirus* rabbits. This is less obvious for some sequences of "a-positive" rabbits, where recurrent variation could either indicate "novel" genetic variants of the "domestic" allotypes, or the consistent use, for gene conversion events, of donor $V_H$ genes present in both *algirus* and domestic rabbits. The latter possibility was evaluated by searching the Genbank database for rabbit sequences showing the same unexpected amino acid replacements. This was tested by submitting such variant sequences, or segments of them, to protein BLAST searches using the software provided by NCBI (*http://www.ncbi.nlm.nih.gov/blast*). The alignments for the 1000 most similar rabbit $V_H$ gene sequences were edited under the "flat query-anchored with identities" option, and examined visually.

# Results

## *Serological typing and sequence determination*

cDNA was obtained from spleen and peripheral blood cells of eight adult wild rabbits: Rb07, Rb11, Rb27, Rb32, Rb36, Rb40, Rb43 and Rb60. According to their mtDNA (Monnerot *et al.*, 1994), as well as to other genetic markers (protein and microsatellite DNA; Branco, 2000; Queney *et al.*, 2001), these rabbits belong to the subspecies *O. cuniculus algirus* (data not shown). When tested with conventional allo-antisera, Rb43 typed "*a1/a2*" heterozygous, Rb07 and Rb36 "*a1* only" and Rb11 "*a3* only". Sera of Rb32 showed weak cross-reaction with *a3*-specific antisera (*a3*-variant 32). Sera of rabbits Rb27, Rb40 and Rb60 did not show any cross-reactivity (*"a-blank"*). Because of the prevalence of *"a-blanks"* in the study area (>20%), the "*a-positive*" rabbits (Rb07, Rb36, Rb11, Rb32) have a reasonable chance of being heterozygous for the *"a-blank"* genotype.

A total of 120 $V_H$ sequences were obtained, of which 111 sequences are deposited in Genbank under accession numbers AY207941 to AY208048 and AY299401 to AY299403. Except AY208000 and AY208009, all submitted sequences encode unique amino acid sequences. Forty-four of them were expressed in rabbits typed as homozygous *"a-blank"*, while eleven more were obtained from a rabbit that appeared to be heterozygous for the *"a-blank"* allotype (Rb07). Table 1 lists, for each of the 8 serotypes, the number of sequences leading to unique amino acid sequences (110 sequences, of which 52 were classified *"a-blank"*, 55 "*a-positive*", and 3 "*a-negative*). Unless stated otherwise, we present and discuss the results at the level of the protein sequence.

**Table 1**
Number and allotypic classification of *VDJ* sequences obtained from 8 rabbits

| Rabbit | Serotype | putative allotype | | number of sequences [a] | | | |
|---|---|---|---|---|---|---|---|
| | | | | $V_Ha$ | | $V_Hx$ | $V_Hy$ |
| | | G1 | G2 | G1 | G2 | | |
| Rb07 | *a1 only* | *a1* | *a41* | 1 | 9 | 1 | 0 |
| Rb11 | *a3 only* | *a3* | *a3v11* | 5 | 4 | 0 | 0 |
| Rb27 | *"a-blank"* | *a41* | *a42* | 10 | 4 | 0 | 0 |
| Rb32 | *a3v* | *a3v32* | *a3v32* | 18-n | n | 0 | 0 |
| Rb36 | *a1 only* | *a1v36* | *a1v36* | 8-n | n | 0 | 0 |
| Rb40 | *"a-blank"* | *a41* | *a42* | 8 | 7 | 0 | 0 |
| Rb43 | *a1/a2* | *a1v36* | *a2* | 3 | 16 | 0 | 1 |
| Rb60 | *"a-blank"* | *a42* | *a42* | 14-n | n | 1 | 0 |

[a] underlying unique aa. sequences

## *Inferring $V_H$ genes expressed in O. c. algirus*

The $V_H$ regions of rearranged *VDJ* genes in adults are likely somatically diversified versions of the utilized genomic $V_H$ genes. Sequence comparisons can help to infer the likely

sequences of these germline genes, as well as identify genes contributing to the $V_H$ gene diversification by gene conversion.

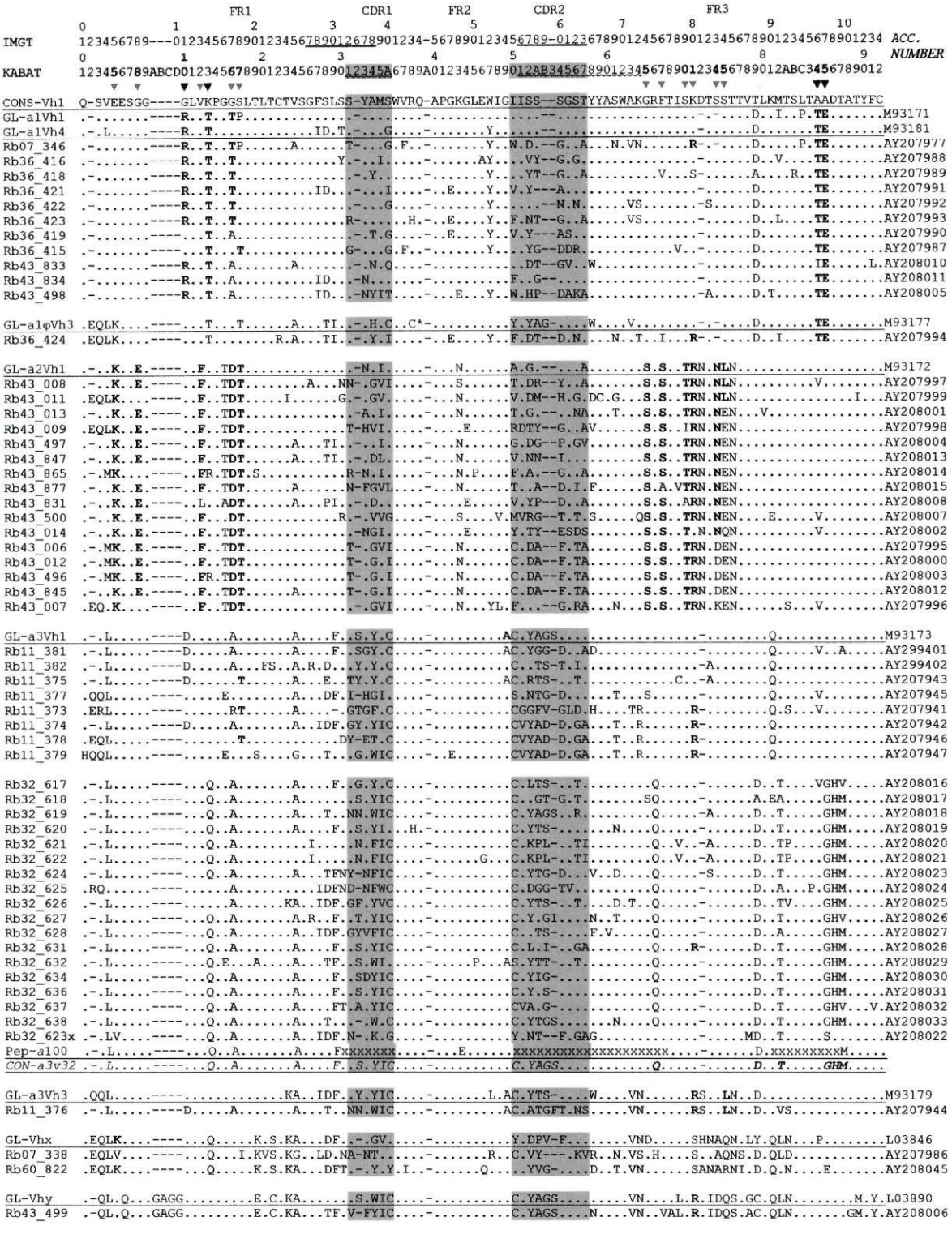
The "*O. c. algirus*" sequences were therefore aligned with all published rabbit genomic $V_H$ sequences. Rabbit $V_H$ sequences belong to the class C (Tutter and Riblet, 1989; Ota and Nei, 1994). For this reason we used, as outgroups in phylogenetic analysis, vertebrate $V_H$ sequences that are separated by a relatively short branch from the root of class C sequences (Human $V_H3$, Camel $V_H$, and Duck $V_H$). Figure 2 displays the unique amino acid sequences inferred from the nucleotide data, grouped according to highest similarity to known rabbit germline genes. They are compared to a consensus sequence of rabbit $V_Ha$ genes, in order to highlight the differences among the $V_Ha$ allotypes. Residues are numbered according to IMGT numberings (*http://imgt.cines.fr*; Lefranc, 2001) or to Kabat *et al.*, (1991), the latter numbering being used throughout the text below. For groups of $V_H$ sequences that show systematic differences from known genes (which is the case for the majority of sequences obtained from *"a-blank"* rabbits), consensus sequences are proposed. These likely represent the underlying germline sequence. The possibility that differences might be due to gene conversion with $V_H$ genes also present in domestic breeds was examined by blast searches outlined in the Material and Methods section.

## $V_H$ genes used by "*a*-positive" *algirus* rabbits

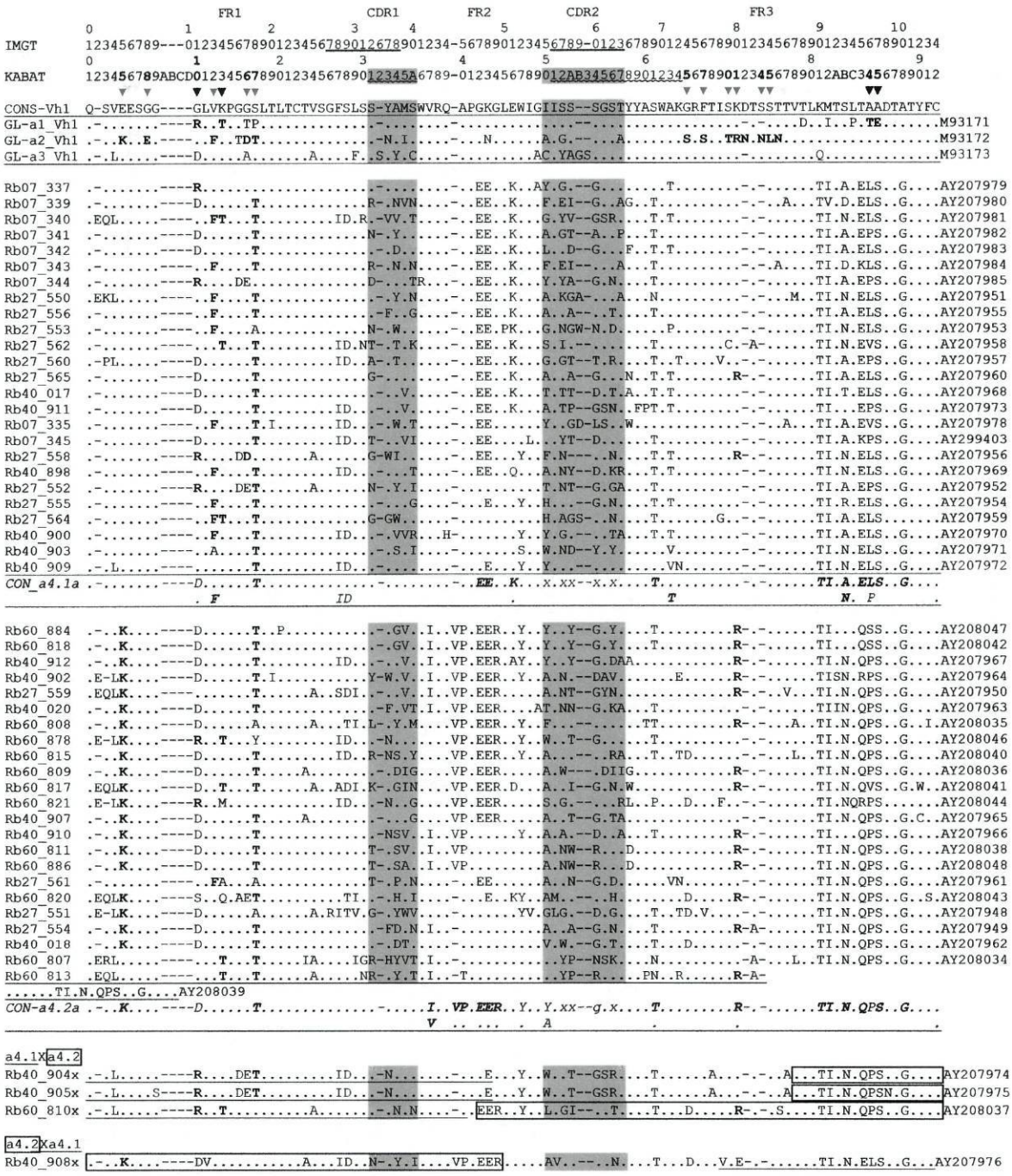
### *$V_Ha$-a1 genes expressed by Rb07, Rb36 and Rb43:*

Rb07 appears heterozygous for the a1 allele and another allele that is also expressed in *"a-blank"* rabbits (see below). Whereas the *a1* sequence of Rb07 conforms to a "domestic" *$V_H1$-a1* gene, those expressed by Rb36 and Rb43 consistently show at positions 17, 74, 82, and 82C the residues that distinguish *$V_H4$-a1* from *$V_H1$-a1*. The second deletion in FR3 (position 74), which is characteristic of *$V_H1$-a1* and absent in *$V_H4$-a1*, is also absent in all *$V_Ha$-a1* sequences of these two rabbits, except for sequence Rb36-424. The latter sequence, however, shows more similarity with the pseudogene *$V_H3$-a1*, and is likely a product of gene conversion (Fig. 2A). One sequence from Rb07 (Rb07_332; Fig. 2A) is clearly derived from a *$V_Hx$*-like gene, while one sequence from Rb43 (Rb_499) is clearly derived from a *$V_Hy$*-like gene.

**A**


FR1 CDR1 FR2 CDR2 FR3
IMGT
KABAT
ACC.
NUMBER
CONS-Vh1 Q-SVEESGG----GLVKPGGSLTLTCTVSGFSLSS-YAMSWVRQ-APGKGLEWIGIISS--SGSTYYASWAKGRFTISKDTSSTTVTLKMTSLTAADTATYFC
GL-a1Vh1 M93171
GL-a1Vh4 M93181
Rb07_346 AY207977
Rb36_416 AY207988
Rb36_418 AY207989
Rb36_421 AY207991
Rb36_422 AY207992
Rb36_423 AY207993
Rb36_419 AY207990
Rb36_415 AY207987
Rb43_833 AY208010
Rb43_834 AY208011
Rb43_498 AY208005
GL-a1φVh3 M93177
Rb36_424 AY207994
GL-a2Vh1 M93172
Rb43_008 AY207997
Rb43_011 AY207999
Rb43_013 AY208001
Rb43_009 AY207998
Rb43_497 AY208004
Rb43_847 AY208013
Rb43_865 AY208014
Rb43_877 AY208015
Rb43_831 AY208008
Rb43_500 AY208007
Rb43_014 AY208002
Rb43_006 AY207995
Rb43_012 AY208000
Rb43_496 AY208003
Rb43_845 AY208012
Rb43_007 AY207996
GL-a3Vh1 M93173
Rb11_381 AY299401
Rb11_382 AY299402
Rb11_375 AY207943
Rb11_377 AY207945
Rb11_373 AY207941
Rb11_374 AY207942
Rb11_378 AY207946
Rb11_379 AY207947
Rb32_617 AY208016
Rb32_618 AY208017
Rb32_619 AY208018
Rb32_620 AY208019
Rb32_621 AY208020
Rb32_622 AY208021
Rb32_624 AY208023
Rb32_625 AY208024
Rb32_626 AY208025
Rb32_627 AY208026
Rb32_628 AY208027
Rb32_631 AY208028
Rb32_632 AY208029
Rb32_634 AY208030
Rb32_636 AY208031
Rb32_637 AY208032
Rb32_638 AY208033
Rb32_623x AY208022
Pep-a100
CON-a3v32
GL-a3Vh3 M93179
Rb11_376 AY207944
GL-Vhx L03846
Rb07_338 AY207986
Rb60_822 AY208045
GL-Vhy L03890
Rb43_499 AY208006

**B**


FR1
CDR1
FR2
CDR2
FR3
IMGT
KABAT
CONS-Vh1 Q-SVEESGG----GLVKPGGSLTLTCTVSGFSLSS-YAMSWVRQ-APGKGLEWIGIISS--SGSTYYASWAKGRFTISKDTSSTTVTLKMTSLTAADTATYFC
GL-a1_Vh1 M93171
GL-a2_Vh1 M93172
GL-a3_Vh1 M93173
Rb07_337 AY207979
Rb07_339 AY207980
Rb07_340 AY207981
Rb07_341 AY207982
Rb07_342 AY207983
Rb07_343 AY207984
Rb07_344 AY207985
Rb27_550 AY207951
Rb27_556 AY207955
Rb27_553 AY207953
Rb27_562 AY207958
Rb27_560 AY207957
Rb27_565 AY207960
Rb40_017 AY207968
Rb40_911 AY207973
Rb07_335 AY207978
Rb07_345 AY299403
Rb27_558 AY207956
Rb40_898 AY207969
Rb27_552 AY207952
Rb27_555 AY207954
Rb27_564 AY207959
Rb40_900 AY207970
Rb40_903 AY207971
Rb40_909 AY207972
CON_a4.1a
Rb60_884 AY208047
Rb60_818 AY208042
Rb40_912 AY207967
Rb40_902 AY207964
Rb27_559 AY207950
Rb40_020 AY207963
Rb60_808 AY208035
Rb60_878 AY208046
Rb60_815 AY208040
Rb60_809 AY208036
Rb60_817 AY208041
Rb60_821 AY208044
Rb40_907 AY207965
Rb40_910 AY207966
Rb60_811 AY208038
Rb60_886 AY208048
Rb27_561 AY207961
Rb60_820 AY208043
Rb27_551 AY207948
Rb27_554 AY207949
Rb40_018 AY207962
Rb60_807 AY208034
Rb60_813
AY208039
CON-a4.2a
a4.1Xa4.2
Rb40_904x AY207974
Rb40_905x AY207975
Rb60_810x AY208037
a4.2Xa4.1
Rb40_908x AY207976

C

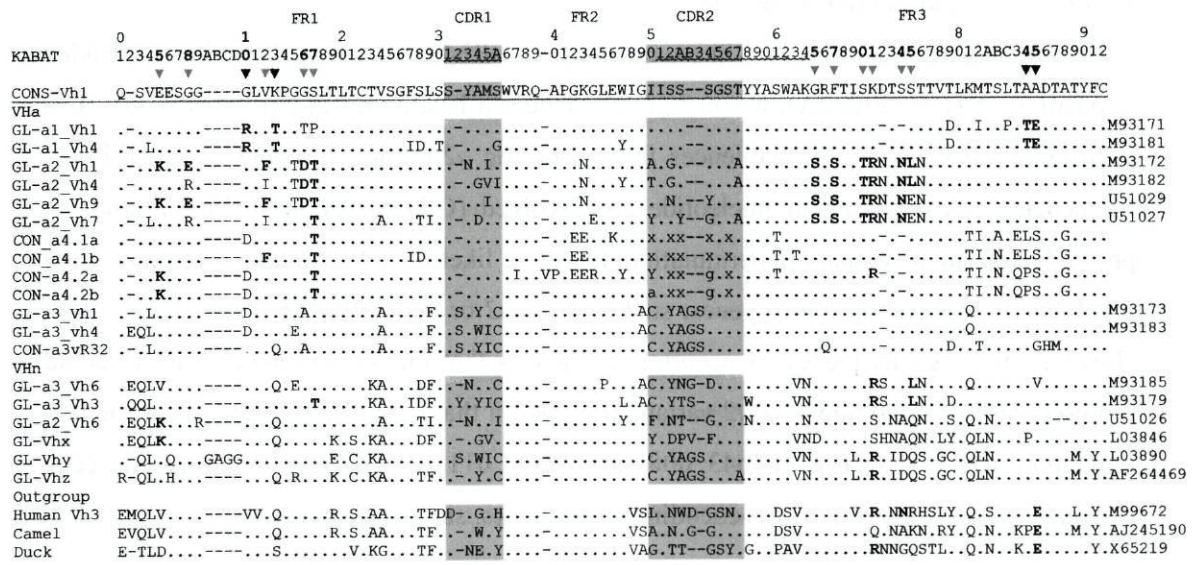



**Figure 2** - *Alignment of unique protein sequences of $V_H$ regions.* $V_H$ regions shown are inferred from *VDJ* genes obtained from 8 wild Iberian rabbits and of published genomic rabbit $V_H$ genes. Sequence names of the former refer to the donor rabbits (Rb07, Rb11, Rb27, Rb32, Rb36, Rb40, Rb43 and Rb60). They are compared to a consensus sequence of rabbit $V_H1$ gene segment (CONS-VH1) and grouped according to phylogenetic relationships, shown in Figure 3. Allotypic marker positions are indicated by▼ and ▼ for the *a1* and *a2* allotypes, respectively, and corresponding matches are highlighted in bold. Genbank accession numbers of underlying nucleotide sequences are included. Each *VDJ* gene cluster is compared to germline $V_H$ gene(s) (marked GL-) of that cluster. For *VDJ* groups showing recurrent differences with known genomic genes, consensus sequences are proposed (indicated by CON-). Potential gene conversion tracts are enclosed in dashed boxes. Sequence names of suspected crossing-over products are marked by "x" (*a4.1* sequences are underlined and *a4.2* sequences are enclosed in boxes). Hypervariable parts of CDR's are shaded. **A:** Sequences showing significant clustering with one or more known genomic sequences. **B:** Sequences clustering apart from known genomic sequences. **C:** Comparison of inferred consensus sequences with relevant genomic sequences of rabbit and with class-C $V_H$ sequences used as outgroups in phylogenetic trees presented in Figures 3 and 4.

*$V_H$a-a2 genes expressed by Rb43:*

The majority (14/16) of *a2*-like sequences from Rb43 encodes a glutamic acid residue at position 75 (E75; Fig. 2A), where *$V_H$1-a2* encodes a leucine residue (L75). E75 has also been observed in *VDJ* gene sequences from domestic *a2* rabbits, although at much lower frequencies. As two tandem nucleotide are involved (GAG vs. CTG), this variation is probably not due to somatic hypermutation, but more likely due to somatic gene conversion involving E75-encoding donor genes, such as *$V_H$7-a2* or *$V_H$9-a2* (cf. Sehgal *et al.*, 1998). However, the high frequency of occurrence suggests that the $V_H$1-a2 gene of Rb43 could represent an E75-encoding *$V_H$1-a2* variant allotype. In that case, the L75 codons of the two remaining Rb43 sequences (Rb43_008 and Rb43_011; Fig. 2A) may represent gene conversion events involving L75-encoding donor genes, such as *$V_H$4-a2*.

*$V_H$a-a3 genes expressed in Rb11*:

The amino acid variation among Rb11 sequences is similar to that of domestic *a3* rabbits (Fig. 2A). However, 4 sequences (Rb11_373, _374, _378 and _379) are atypical in containing arginine (R) residues at positions 64 and 71, where virtually all published *a3* sequences (>99%) contain lysine (K). The sequence of Rb11_376 between positions 63 and 78 could be the product of gene conversion between *$V_H$1-a3* and a gene homologous, or identical, to domestic rabbit pseudogene *$V_H$3-a3*.

*$V_H$a-a3 variant genes expressed in Rb32 ($V_H$1-a3v32)*:

Serological analysis shows that rabbit Rb32 expresses some, but not all, of the *a3* determinants. The Rb32 sequences differ from those of the *$V_H$1-a3* allotype at nine FR positions: two in FR1 (D10, Q13), and seven in FR3 (Q66, D79, K81, T82, G85, H86, M87). The amino acid differences at positions 10 (D/G) and 13 (Q/P) of FR1 have also been observed in domestic *a3* rabbits, although much less frequently, possibly due to gene conversion with *$V_H$6-a3* or a similar donor gene. In contrast, the substitutions in FR3 were never, or extremely rarely, observed in domestic rabbits. Among the 1000 rabbit *VH* sequences retrieved by a Protein Blast search using the FR3 region of Rb32_634 as query sequence (see M&M), none showed either the residues Q66 or H86 that are present in all but two of the R32 sequences. Residues T82, G85 and M87 each occurred at frequencies below 0.5%. D79 and K81 are frequently observed among *a1* and *a2* sequences, but not among *a3* sequences. Rb32 undoubtedly expresses a genetic variant of the *a3* lineage, which

we call "*a3v32*". Interestingly, residues G10, Q13, D79 and M87 have previously been reported in the partial amino acid sequence obtained by Tonelle *et al.*, (1983) from a wild rabbit typed "*a100*" (Fig. 2A). It is possible that *a100* and *a3v32* are different names for the same allotype.

## $V_H$ genes used by *"a-blank" algirus* rabbits define a fourth $V_Ha$ allotypic lineage: $V_Ha$ *-a4*

### *$V_H$ genes expressed in "a-blank" rabbits belong to the $V_Ha$ group*

Except for one single sequence (Rb60-822), which is clearly a product of an "*a-negative*" $V_Hx$-like gene, all of $V_H$ genes expressed in the *"a-blank"* rabbits (Rb27, Rb40, Rb60) appear to be derived from very similar germline genes. Visual inspection of the inferred protein sequences (Fig. 2B) shows that they are clearly more related to the $V_Ha$ than to $V_Hx$ or $V_Hy$ genes. This conclusion is supported by a variety of phylogenetic inference programs conducted at both the nucleotide and protein levels and is independent of the CDR regions.

For the construction of the phylogram shown in Fig 3, we included a random sample of published diversified $V_H$ sequences of domestic rabbit (see Material and Methods). The *"a-blank"* sequences form a monophyletic cluster with the $V_H1$-*a* genes, which clearly branches apart from the $V_Hn$ genes ($V_Hx$, $V_Hy$ and $V_Hz$). At the protein level, the *"a-blank"* sequences share a number of apomorphic (or "derived") characters with one or more of the "domestic" $V_Ha$ genes. Examples are the $V_Ha$ hallmark peptides (16)LTLTCT(23) of FR1, and (61)WAK(65) of FR3. These are unique to rabbit $V_Ha$ genes and do not occur in the $V_Hx$, $V_Hy$ or $V_Hz$ sequences. The *"a-blank"* type shares a larger number of derived characters with the *a1* and *a2* allotypes than with the *a3* allotype. The two deletions between positions 72 and 74, which are characteristic of the $V_H1$-*a1* allotype, are also present in *each* of the 49 *"a-blank"* sequences, where we observe either (71)K-T-ST(76) (as in $V_H1$-*a1*) or (71)<u>R</u>-T-ST(76), confirming a close affiliation with the rabbit $V_H1$ genes.

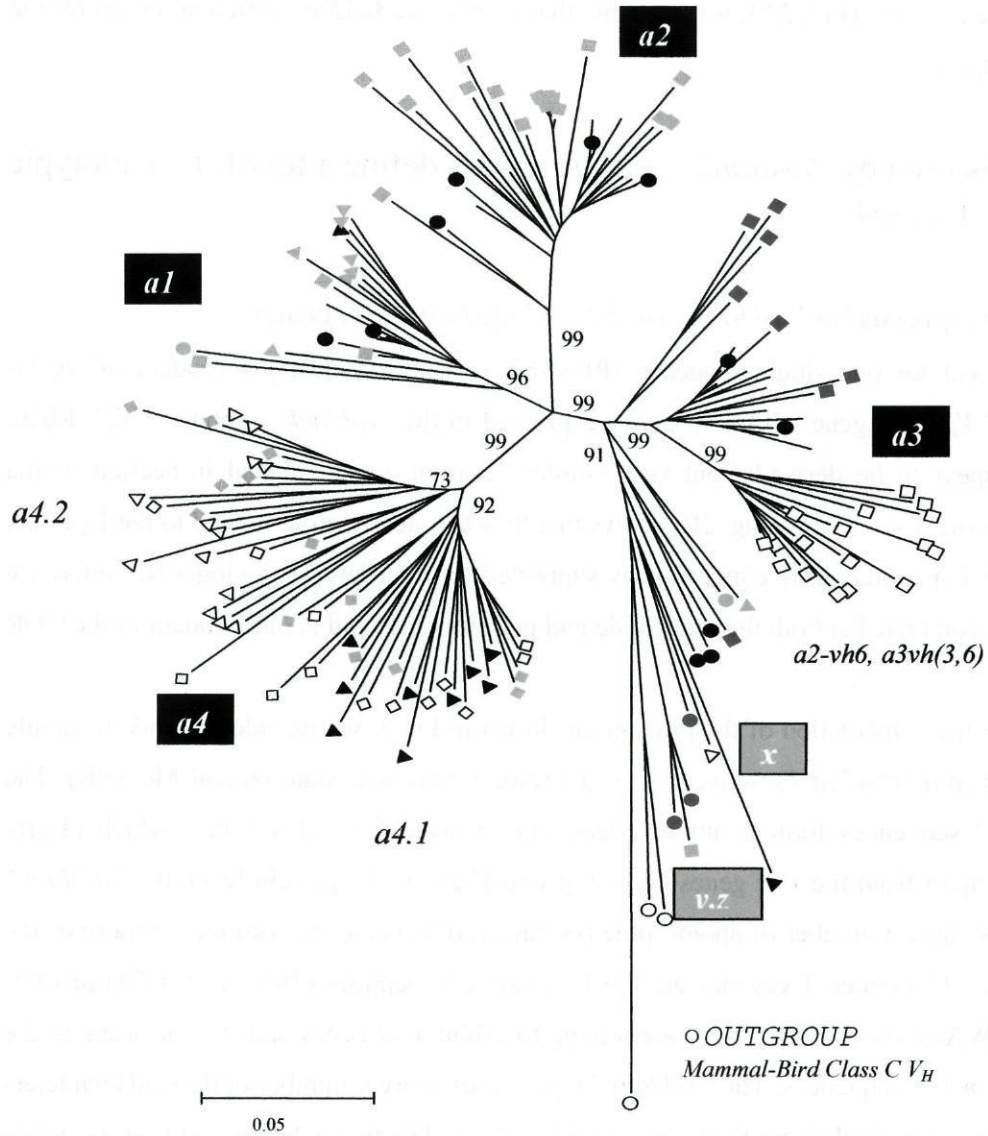



**Figure 3** - *Phylogenetic neighbour-joining tree of rabbit $V_H$ regions.* Based upon nucleotide p-distances. Germline genes are labelled by dots and comprise ●: $V_Ha$ genes (*a1-vh(1,4), a2-vh(1,4,7,9),* and *a3-vh(1,4)*); ●: $V_Hn$ genes (*x, y, z*); ●: $V_Ha$ pseudogenes (*a1-vh(2,3)*); ○: outgroup (*Hum $V_H3$, Camel $V_H$, Duck $V_H$*). Squares, triangles and diamonds indicate $V_H$ regions of RT-PCR-amplified *VDJ* genes obtained from wild rabbits of subspecies *O. c. algirus*: ▲: Rb07; ■: Rb11;□: Rb27; □: Rb32; ▽: Rb36; □: Rb40; ■: Rb43; □: Rb60. *VDJ* gene sequences representing domestic rabbits of the *a1, a2* and *a3* allotypes are not labelled (Accession numbers are displayed in M&M). Sequences reflecting possible crossing-over (marked by "x" in Fig. 2) were excluded from the analysis. CP values (1000 replicates) are shown at branch nodes if larger than 85, except for the *a4.2* clade. The tree illustrates the fact that the vast majority of *VDJ* genes expressed in rabbits Rb07, Rb27, Rb40 and Rb60 form a well-defined cluster that is monophyletic and equidistant from the clades formed by the $V_Ha$ genes expressed in domestic rabbits of the *a1* and *a2* allotypes.

*$V_H$ "a-blank" defines a new $V_H1$ allele*

The *"a-blank"* $V_Ha$ cluster is well separated from the three clusters defined by the sequences expressed in *a1*, *a2* and *a3* rabbits, respectively. BP and CP values are highest when *"a-blank"* sequences, or their consensus sequence, are solely compared to genomic $V_H$ genes (Fig. 4). Visual inspection of the protein sequences (Fig.2B) reveals that the *"a-blank"* sequences, although clearly belonging to the $V_Ha$ genes, have unique features in common that set them apart from the *a1*, *a2* and *a3* alleles. The most consistent marker residues are found in FR3, where, for *each* of 49 *"a-blank"* $V_Ha$ sequences, the $V_Ha$ consensus residue K/Q(81) is replaced by T(81). The $V_Ha$ consensus peptide (82C)TAA(85), which is TTE in the *a1* allotype, is replaced by ELS, or by QPS in most *"a-blank"* sequences.

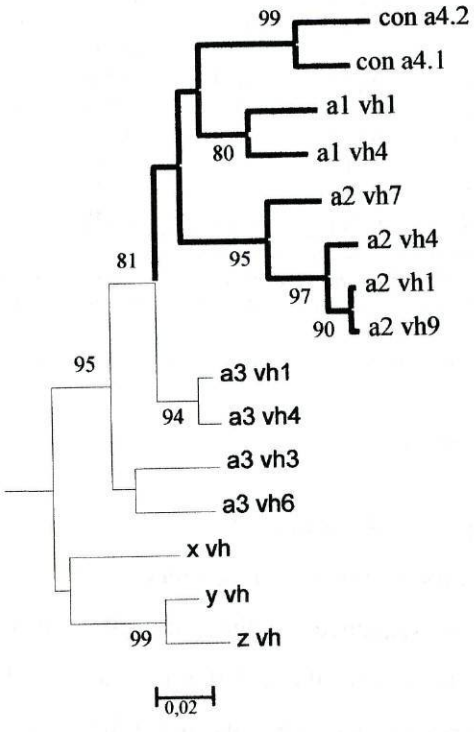



**Figure 4** - *Phylogenetic neighbour-joining tree of rabbit $V_H$ regions.* Based upon amino acid p-distances. The consensus sequences of the *a4* allotype subtypes *a4.1* and *a4.2* were inferred from the data presented in Figure 2B and compared to inferred amino acid sequences of published genomic rabbit $V_H$ genes. BP values were obtained with 1000 bootstrap replicates.

While FR2 is highly conserved among vertebrate class-C $V_H$ genes (including the rabbit $V_H1$ genes), the *"a-blank"* sequences present some unprecedented changes. In the majority of them, the canonical (41)GK/NG(45) is replaced by either EEG or EER. Quite remarkable is the change of A(40) into VP(40), due to a 3 bp insertion before the second

codon position of *A40(gct->g[tg c]ct)*. This was observed in 17 sequences, obtained with 3 different rabbits. Assuming that, like domestic rabbits, *"a-blank"* rabbits preferentially use the *D*-proximal $V_H$ gene, we propose that the majority of the sequences obtained from *"a-blank"* rabbits are derived from a genuine allele of the $V_H1$ gene, which we call $V_H1$-*a4*. The phylograms in Figures 3 and 4 compare the "*a4*" sequences with the known functional genomic $V_H$ sequences of rabbits. The pairwise distances between the *a4* sequence and the *a1*, *a2*, or *a3* sequences were similar to those between the latter three ("domestic") allotypes (Table 2), confirming that the *a4* lineage cannot be considered a subtype of one of the known allotypic lineages.

**Table 2**

Estimation of genetic distances between major $V_H$a allotypic lineages.

| A | B | Net distances between VDJ groups [a] | | | | Pairwise distances between $V_H$ genes [b] | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | a. a. pD[c] | s.e.[d] | nuc K2[e] | s.e.[d] | a. a. pD[c] | s.e.[d] | nuc K2[e] | s.e.[d] |
| *a4.1* | -a1 | 0.137 | *0.030* | 0.094 | *0.018* | 0.203 | *0.045* | 0.124 | *0.024* |
| *a4.2* | *-a1* | 0.158 | *0.032* | 0.103 | *0.019* | 0.241 | *0.048* | 0.129 | *0.025* |
| *a4.1* | *-a2* | 0.227 | *0.042* | 0.146 | *0.024* | 0.291 | *0.051* | 0.171 | *0.030* |
| *a4.2* | *-a2* | 0.226 | *0.041* | 0.143 | *0.024* | 0.278 | *0.050* | 0.154 | *0.028* |
| *a1* | *-a2* | 0.216 | *0.041* | 0.131 | *0.023* | 0.253 | *0.049* | 0.134 | *0.026* |
| *a4.1* | *-a3* | 0.141 | *0.033* | 0.085 | *0.016* | 0.215 | *0.046* | 0.114 | *0.023* |
| *a4.2* | *-a3* | 0.148 | *0.033* | 0.087 | *0.017* | 0.253 | *0.049* | 0.119 | *0.024* |
| *a1* | *-a3* | 0.108 | *0.029* | 0.061 | *0.015* | 0.177 | *0.043* | 0.080 | *0.019* |
| *a2* | *-a3* | 0.187 | *0.038* | 0.114 | *0.021* | 0.250 | *0.048* | 0.132 | *0.025* |

* Calculated using the corresponding option of the MEGA2 package

[a] Net distances were determined for the FR regions between allotypic clusters of *VDJ* sequences formed by the sequences as presented in the tree in Fig. 3 (including *VDJ* sequences of domestic rabbits); they are main distances between groups corrected for variation within groups

[b] Pairwise distances between genomic $V_H$ sequences and between consensus sequences, as presented in the tree in Fig. 4

[c] amino acid p-distances; 93 aa

[d] standard error

[e] Kimura's 2 parameter nucleotide distances; 279 bp.

### *Allelic diversity among $V_H1$-a4 sequences?*

The degree of diversity among *a4* sequences tended to be larger than currently observed among *a-positive* sequences of the same allotype in domestic rabbits (data not shown). This could have to do with the fact that our sample of *a4* sequences was obtained from 7 different haplotypes (1 heterozygous and 3 homozygous *a4* rabbits), *i.e.* cryptic allelic variation could account for some of this diversity. Phylogenetic clustering methods and visual inspection of the amino acid alignments indeed indicate that there might be at least two *a4* variants, *a4.1* and *a4.2*. The consensus sequences of these two variants are presented in Fig. 2B. Rb60 appears to be homozygous for one variant (*a4.1*). Rabbit Rb07, which is heterozygous *a1/a4*, appears to possess the *a4.2* variant, whereas Rb27 and Rb40 appear to be heterozygous *a4.1/a4.2*. The phylogram in Fig. 4 highlights the position of the *a4*.1 and *a4*.2 consensus protein sequences among known genomic $V_H$ genes.

*Gene conversion*

There are strong indications that the putative *a4* variants are diversified by gene conversion events involving $V_H$ donor genes that, in some cases, might be present in both lineages. At a number of positions, amino acid replacements requiring two or more nucleotide substitutions occur more often than expected if due to somatic hypermutation (examples are D/G10, 27ID/FS28, W/Y47, 71AK/VN72, N/A82b). Inspection of the sequence alignments indicates that in some cases variation could possibly be due to PCR crossing-over. Indeed, sequences Rb40_905, Rb40_901, and Rb27_561 show characteristics of *a4.1* and *a4.2* sequences in the 5' and 3' regions, respectively, while the opposite is true for Rb40_908 (Fig. 2). These sequences were not included in the phylogenetic analysis shown in Fig. 3. As these sequences were obtained from rabbits that express both subtypes, PCR crossing-over and/or gene conversion could be an additional source of variation. However, crossing-over was not observed between the *a1* and *a4* genes of Rb07 or between the *a1* and *a2* genes of Rb43. The presence or absence of the valine (position 39A) insertion in FR2 among the *a4.1* sequences indicates that more than one $V_H$ gene contributes to this particular allotype. The sequence of Rb60_810 between position 1 and 17 could be a product of gene conversion of the putative a4.2 gene and a gene resembling the *vh4-a1* gene of domestic rabbit. We conclude that the $V_Ha$-*a4* sequences presented here are likely diversified by gene conversion. In Fig. 2B we have indicated frequent alternates of the consensus sequences.

*Gene usage*

In sharp contrast with our first hypothesis, the *"a-blank"* rabbits do not preferentially use $V_Hn$ genes. In fact, of the 43 sequences obtained from rabbits Rb27, Rb40 and Rb60, 42 had the $V_Ha$ signature, while only 1 was $V_Hx$-like. If indeed the number of PCR clones reflects gene usage, $V_Hn$ genes were used significantly less frequently in *"a-blank"* rabbits than was previously observed in domestic rabbits. This pronounced bias in gene usage is apparently not a characteristic linked to the *"a-blank"* genotype, because it was also observed for the "*a-positive*" rabbits of the same population (55 $V_Ha$ sequences *vs.* zero $V_Hx$ and one $V_Hy$).

*Allelic imbalance*

Another characteristic of the rabbit $V_Ha$ allotypes is "allelic imbalance" in gene expression or "pecking order" (Lummus *et al.*, 1967, Mage, 1967, Akimenko *et al.*, 1986). This refers to the fact that the different $V_Ha$ alleles are not used to the same extent in heterozygous animals. Thus, heterozygous *a1/a2* rabbits consistently use *a1* genes more often than *a2* genes. We note that with Rb43, which typed *a1/a2* heterozygous, the ratio *a1/a2* was on the contrary 3/16. This could possibly have to do with the fact that the $V_H1$-*a1* of this rabbit differs from that in domestic breeds (see above). For rabbit Rb07, which is heterozygous *a1/a4*, the *a1/a4* ratio was 1/7, indicating that allelic imbalance, if occurring at all, would favour the expression of the *a4* allotype in the presence of the *a1* allotype.

*Absence of $V_Ha$ serological markers*

In Fig 2, the $V_Ha$-allotypic residues (*i.e.* residues known to be associated with serological determinants of the allotype) are highlighted. The *a4* sequences present most of the "domestic" allotypic marker residues of FR1. These marker residues however are apparently not presented in combinations required for serological cross-reactivity. For example, while the *a2* marker residues K5, F12 and T17 (FR1 region) are frequently found among *a4* molecules, they never occur together. The same is true for the *a1*-associated FR1 residues R10 and T13, except for two Rb60 sequences (Rb60_810, Rb60_878), where we would expect the phenotypic expression of an *a1* determinant. The fact that rabbit Rb60 was typed *"a-blank"* suggests that the serum level of such molecules is too low for detection or, more likely, that a majority of the antibodies of our *a1*-specific antisera are directed against determinants in the FR3 region. Indeed, none of the marker residues of the *a1, a2,*or *a3* allotypes was observed in the FR3 region of the *a4* sequences.

# Discussion

Proteins encoded by allelic genes show in general not more than 1-3% a. a. differences. Evolutionary theory offers two possible explanations for interallelic distances as large as those observed at the rabbit $V_Ha$ locus: unusually long allele persistence times and increased mutant recruitment rates. The concept of a molecular clock is a powerful tool that evolutionists avoid abandoning. The fact that allo-antisera raised in rabbit distinguish three

$V_H a$ allotypes in populations of *Lepus americanus* suggests that both species have inherited the same three lineages, which could explain the large interallelic distances without invoking acceleration of evolution rates (Su and Nei, 1999).

The main objective of this study was to determine whether the study of natural populations in the original species range could contribute to a more accurate estimate of the genetic variation at this locus. We show here that there are at least four highly divergent lineages (Fig. 2). A correct estimation of the distance between the *a4* lineage and the other $V_H a$ lineages is hampered by the fact that only diversified sequences of the putative $V_H 1$-*a4* gene are available. However, the net distances between diversified $V_H$ sequences of the four lineages are similar (Table 2). If we assume with Su and Nei (1999) that the $V_H a$ genes of different haplotypes evolved at a similar rate as other proteins (*i.e.* $1.4 \times 10^{-9}$ n/y), then not three, but four, lineages have been maintained over numerous speciation steps (40-60 My). However, the present geographical distribution of the four allotypes in the native species range shows that the *a4* allele is associated with the maternal markers of the subspecies *O. c. algirus,* and did not pass the Pyrenean Mountains (Fig 1).

A detailed evolutionary analysis of the present data is beyond the scope of this paper. It is nevertheless interesting to compare the present observations with those made in a study of gene diversity at the *IGKC1*, or $C_{K1}$ *b*-locus in Iberian rabbit populations (van der Loo *et al.,* 1999). The $C_{K1}$ gene, which encodes the constant region of the Ig kappa1 L chain, is used in more than 95% of L chain rearrangements. Allelic differences at this locus are even larger than those at the $V_H a$ locus and contrast sharply with the very limited variation at the quasi-silent $C_{K2}$ *bas* locus. The *b4* and *b5* $C_{K1}$ alleles, which predominate in all populations of *O. c. cuniculus*, were less frequent in populations of *O. c. algirus.* These populations instead express genetic variants serologically related to the *b4* and *b5* alleles but differing from them at up to 15% of the a. a. residues. It was argued that the association of specific alleles with subspecies markers is in contradiction with extremely long allele persistence times.

The present results, showing that the $V_H a$ allotypes that prevail in *O. c. cuniculus* populations, are in part "replaced" in *O. c. algirus* populations by the related *a4* allotype (Fig. 1), mirrors to some extent the situation described for the *b*-locus. In the study mentioned, the molecular clock hypothesis was rejected for the $C_{K1}$ lineages, and the data analysis furthermore indicated that, compared to the *b9* lineage, the *b4* and *b5* lineages showed both higher allele turnover rates and higher evolutionary rates.

Our data suggest that evolutionary modes might also vary among $V_Ha$ lineages. In Fig. 4 we show a phylogenetic tree reduced to the genomic sequences and the *a4* consensus sequences. It strongly suggests that amino acid divergence was markedly increased in the $V_Ha$ *a1*, *a2*, and *a4* lineages in comparison to that in the *a3* and $V_Hn$ lineages. The statistical significance of these differences was confirmed by Tajima's relative rate tests (Table 3) and the Felsenstein ML ratio test (not shown). Because of the increased evolutionary rate in the $V_Ha$ *a1*, *a2*, and *a4* lineages, allele persistence time could be largely overestimated. A considerable fraction of the amino acid differences characterizing the *a4* allele might indeed have accumulated after the separation of the subspecies, explaining its association with subspecies *O. c. algirus*.

**Table 3**

Tajima's relative rate test for rate constancy among rabbit $V_H$ genes

| Seq A $V_H1$- | Seq B $V_H1$- | Outgroup | Chi-sq | P[a] | Chi-sq | P |
|---|---|---|---|---|---|---|
| | | | *279 nucleotides*[b] | | *93 amino acids*[b] | |
| *a1* | *a3* | Duck $V_H$ | 8.07 | **0.005** | 6.4 | **0.011** |
| *a2* | " | " | 8.05 | **0.005** | 7.1 | **0.008** |
| *a4.1* | " | " | 9.00 | **0.003** | 7.1 | **0.008** |
| *a4.2* | " | " | 9.00 | **0.003** | 8.1 | **0.005** |
| *x* | " | " | 0.93 | 0.336 | 0.0 | 1.000 |
| *y* | " | " | 0.50 | 0.480 | 0.1 | 0.815 |
| *a1* | " | Human $V_H$ | 5.40 | **0.020** | 4.5 | 0.034 |
| *a2* | " | " | 7.35 | **0.007** | 4.6 | 0.033 |
| *a4.1* | " | " | 10.71 | **0.001** | 5.3 | 0.021 |
| *a4.2* | " | " | 8.91 | **0.003** | 4.6 | 0.033 |
| *x* | " | " | 1.29 | 0.260 | 1.7 | 0.197 |
| *y* | " | " | 0.12 | 0.730 | 1.1 | 0.285 |
| *a1* | *x* | " | 6.28 | **0.009** | 8.0 | **0.005** |
| *a2* | " | " | 9.76 | **0.002** | 8.0 | **0.005** |
| *a4.1* | " | " | 11.92 | **0.001** | 10.9 | **0.001** |
| *a4.2* | " | " | 11.11 | **0.001** | 8.9 | **0.003** |
| *a3* | " | " | 1.29 | 0.257 | 1.6 | 0.197 |

* The test was performed using the MEGA2 package, applied to the published genomic sequences and the consensus sequences obtained from the *a4.1* and *a4.2 VDJ* sequences

[a] Probability of rate monotony between SeqA and SeqB lineages vs. Outgroup. The hypothesis of identical in evolutionary rates between the $V_H$ lineages *a3*, *x*, *y*, and, *z* can not be rejected (P>0.3). In contrast, evolutionary rates are shown to be significantly differ in the *a1*, *a2* and *a4* lineages compared to that in *a3*, *x* and *y* lineages (P<0.03).

[b] The hypervariable parts of the CDR regions were excluded in the analysis.

In view of the evidence suggesting a role for superantigens in the expansion of B cell lineages that present $V_Ha$ allotypic motifs, (Raman and Knight, 1992, Pospisil and Mage, 1998, Mage and Pospisil, 2000) and of the gut microflora in Ab repertoire diversification (Cooper, 1968, Lanning *et al.*, 2000), another explanation for the correlation between the complex *Ig* allotypes ($V_Ha$ and $C_K1b$) and subspecies markers should be considered. By occupying different habitats, each subspecies might indeed acquire a different intestinal microflora, which could favour habitat-specific frequency distributions of (pre-existing) $V_Ha$ allotypes. Thus, the four $V_Ha$ allotypes were perhaps present in the

common ancestor species, but the expansion of the *a4* allotypes were favoured by selection only in Southwestern regions.

As already mentioned, different authors have commented on the fact that allelic variation concerns residues that are liable to interact with superantigens (Pospisil and Mage 1998). These authors suggest that a relation might exist between the high allotypic variability and the low diversity at the FR3 regions, which results from preferential gene usage. Sehgal *et al.*, (1998) and Zhu *et al.*, (1999) invoked selection factors for explaining the "reconstruction" the *a2* allotypic motives in $V_H$*1-a2* knockout rabbits (i.e. *Alicia* strain). All 52 *a4* sequences here obtained share highly unusual residues at H81-H85 (or H90-H97 in IMGT numbering), suggesting that this region is not affected by gene conversion, or/and, that B cells presenting these motives are preferentially stimulated to expand. X ray crystallography shows that in human $V_H$*3* residues at this position interact with a protein *A* domain of *Staphylococcus aureus* (H81-H82b; Graille *et al.*, 2000).

Unlike FR3, the FR2 region is not expected to interact with superantigens. In fact, the allelic variation at the FR2 of $V_H$*a-a4* proteins affects residues known to interact with the $V_L$ domain, and is highly conserved among vertebrate $V_H$ genes. The replacements of glycine residues at positions H42 and H44 by charged residues are surprising. It is noteworthy that in the camel and lama, amino acid replacements in this molecular region were observed for the $V_HH$ domains of the heavy chain antibodies (incuding a G/E exchange at H44), and have been associated with the loss of L chains (Nguyen *et al.*, 2002).

*In conclusion*, the presented data show that the study of wild specimens of the original species range can contribute significantly to our understanding of the evolution and biological meaning of the rabbit immunoglobulin polymorphisms.

## Citations

Akimenko, MA, B. Mariame, and F. Rougeon. 1986. Evolution of the immunoglobulin kappa light chain locus in the rabbit: evidence for differential gene conversion events. *Proc. Natl. Acad. Sci. U S A 83:5180.*

Asmussen, M.A., and J. Arnold. 1991. The effects of admixture and population subdivision on cytonuclear disequilibria. *Theor. Popul. Biol. 39:273.*

Becker, R.S., and K.L. Knight. 1990. Somatic diversification of immunoglobulin heavy chain VDJ genes: evidence for somatic gene conversion in rabbits. *Cell 63:987.*

Branco M. 2000. *Estrutura genética das populações de coelho europeu (Oryctolagus cuniculus) na Península ibérica. Isolamento, diferenciação de duas unidades evolutivas, expansão geográfica e contacto secundário*. Dissertação de Doutoramento, Universidade do Porto.

Branco, M., M. Monnerot, N. Ferrand, and A.R. Templeton. 2002. Postglacial dispersal of the European rabbit (*Oryctolagus cuniculus*) on the Iberian Peninsula reconstructed from nested clade and mismatch analyses of mitochondrial DNA genetic variation. *Evolution 56:792.*

Branco, M., N. Ferrand, and M. Monnerot. 2000. Phylogeography of the European rabbit (*Oryctolagus cuniculus*) in the Iberian Peninsula inferred from RFLP analysis of the cytochrome b gene. *Heredity 4:307.*

Callou, C. 1995. Modifications de l'aire de répartition du Lapin (*Oryctolagus cuniculus*) en France et en Espagne, du Pléistocène à l'époque actuelle. État de la question. *Anthropozoologica 21:95.*

Cazenave, P.A., A. Bennamar, J.A. Sogn, and T.J. Kindt. 1987. Immunoglobulin genes in feral populations. In *The Rabbit in Contemporary Immunological Research.* Dubiski S ed. Longman Scientific & Technical. p 148.

Cooper, M.D., D.Y. Perey, A.E. Gabrielsen, D.E. Sutherland, M.F. McKneally, R.A Good. 1968. Production of an antibody deficiency syndrome in rabbits by neonatal removal of organized intestinal lymphoid tissues. *Int. Arch. Allergy Appl. Immunol. 33:65.*

Di Pietro, L.A., J.A. Short, S.K. Zhai, A.S. Kelus, D. Meier, and K.L. Knight. 1990. Limited number of immunoglobulin VH regions expressed in the mutant rabbit "Alicia". *Eur. J. Immunol. 20:1401.*

Dray, S.G., O. Young, and A. Nisonoff. 1963. Distribution of allotypic specificities among rabbit immunoglobulin γ-globulin molecules genetically defined at two loci. *Nature 199:52.*

Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution 39:783.*

Gallarda, J.L., K.S. Gleason, and K.L. Knight. 1985. Organization of rabbit immunoglobulin genes. I. Structure and multiplicity of germ-line VH genes. *J. Immunol. 135:4222.*

Graille, M., E.A. Stura, A.L. Corper, B.J. Sutton, M.J. Taussig, J.B. Charbonnier, and G.J. Silverman. 2000. Crystal structure of a Staphylococcus aureus protein A domain complexed with the Fab fragment of a human IgM antibody: structural basis for recognition of B-cell receptors and superantigen activity. *Proc. Natl. Acad. Sci. U S A 97:5399.*

Hardy C, J.D. Vigne, D. Casane, N. Dennebouy, J.C. Mounoulou, and M. Monnerot. 1994. Origin of European rabbit (*Oryctolagus cuniculus*) in a Mediterranean island: zooarchaeology and ancient DNA examination. *J. Biol. Evol. 7: 217-226.*

Kabat, E.A., T.T. Wu, H.M. Perry, K.S. Gottesman, and C. Foeller. 1991. *Sequences of Proteins of Immunologic Interest.* U.S. Department of Health and Human Services, Public Health service, National Institutes of Health, Bethesda, MD.

Kelus, A.S., and P.G. Gell. 1967. Immunoglobulin allotypes of experimental animals. *Prog. Allergy 11:141.*

Kelus, A.S., and S. Weiss. 1986. Mutation affecting the expression of immunoglobulin variable regions in the rabbit. *Proc. Natl. Acad. Sci. U S A 83:4883.*

Kim, B.S., and S. Dray. 1972. Identification and genetic control of allotypic specificities on two variable region subgroups of rabbit immunoglobulin heavy chains. *Eur. J. Immunol. 2:509.*

Kim, B.S., and S. Dray. 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: 750.*

Knight, K.L., and R.S. Becker. 1990. Molecular basis of the allelic inheritance of rabbit immunoglobulin VH allotypes: implications for the generation of antibody diversity. *Cell 60:963.*

Krug, M.S., and S.L. Berger. 1987. First-strand cDNA synthesis primed with oligo(dT). *Methods Enzymol. 152:316.*

Kumar, S., K. Tamura, I.B. Jakobsen, and M. Nei. 2001. MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics 17:1244.*

Lanning, D., X. Zhu, S.K. Zhai, and K.L. Knight. 2000. Development of the antibody repertoire in rabbit: gut-associated lymphoid tissue, microbes, and selection. *Immunol. Rev. 175:214.*

Lefranc, M.-P. 2001. IMGT, the international ImMunoGeneTics database. Nucleic Acids Research 29:207.

Lopez-Martinez N. 1989. *Revision sistematica y biostratigrafica de los lagomorphos (Mammalia) del Terciario y Cuaternario de España.* Memorias del Museo Paleontologico de la Universidad de Zaragoza, nº3. Diputacion General de Aragon.

Lummus, Z., J.J. Cebra, and R.G. Mage. 1967. Correspondence of the relative cellular distribution and serum concentration of allelic allotypic markers in normal and allotype suppressed heterozygous rabbits. *J. Immunol. 99:737.*

Mage, R.G. 1967. Quantitative studies on the regulation of expression of genes for immunoglobulin allotypes in heterozygote rabbits. *Cold Spring Harbor Symp. Quant. Biol. 32:203*

Mage, R.G., K.E. Bernstein, N. McCartney-Francis, C.B. Alexander, G.O. Young-Cooper, E.A. Padlan, and G.H. Cohen. 1984. The structural and genetic basis for expression of normal and latent *VHa* allotypes of the rabbit. *Mol. Immunol. 21:1067.*

Mage, R.G., R. Pospisil. 2000. CD5 and other superantigens may select and maintain rabbit self-renewing B-lymphocytes and human B-CLL cells. *Curr. Top. Microbiol. Immunol. 252:87.*

Monnerot M., J.D. Vigne, C. Biju-Duval, D. Casane, C. Callou, C. Hardy, F. Mougel, R. Soriguer, N. Dennebouy, and J.C. Mounolou (1994). Rabbit and Man: Genetic and Historic Approach. *Genet. Sel. Evol. 26:167.*

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

Nguyen, V.K., C. Su, S. Muyldermans, and W. van der Loo. 2002. Heavy-chain antibodies in Camelidae; a case of evolutionary innovation. *Immunogenetics 54:39.*

Ota, T., and M. Nei. 1994. Divergent evolution and evolution by the birth-and-death process in the immunoglobulin VH gene family. *Mol. Biol. Evol. 11:469.*

Oudin J. 1960. Allotypy of rabbit serum proteins. I. Immunochemical analysis leading to the individualization of seven main allotypes. *J Exp Med* **112:** 107-24.

Pospisil, R., and R.G. Mage. 1998. CD5 and other superantigens as 'ticklers' of the B-cell receptor. *Immunol. Today 19:106.*

Queney G., N. Ferrand, S. Weiss, F. Mougel and M. Monnerot (2001). Stationary distributions of microsatellite loci between divergent population groups of the European rabbit (*Oryctolagus cuniculus*). *Mol Biol Evol* **18**(12): 2169-178.

Raman, C., and K.L. Knight 1992. CD5+ B cells predominate in peripheral tissues of rabbit. *J. Immunol. 149:3858.*

Roux, K.H. 1981. A fourth heavy chain variable region subgroup, w, with 2 variants defined by an induced auto-antiserum in the rabbit. *J. Immunol. 127:626.*

Rzhetsky, A., and M. Nei. 1992. Statistical properties of the ordinary least-squares, generalized least-squares, and minimum-evolution methods of phylogenetic inference. *J. Mol. Evol. 35:367.*

Saitou, N., and M. Nei 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol. 4:406.*

Sehgal D., H. Obiakor, and R.G. Mage. 2002. Distinct clonal Ig diversification patterns in young appendix compared to antigen-specific splenic clones. *J. Immunol. 168:5424.*

Sehgal, D., R.G. Mage, and E. Schiaffella. 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. 160:1246.*

Short, J.A., P. Sethupathi, S.K. Zhai, and K.L. Knight. 1991. *VDJ* genes in *VHa2* allotype-suppressed rabbits. Limited germline VH gene usage and accumulation of somatic mutations in D regions. *J. Immunol. 147: 4014.*

Sitnikova, T. 1996. Bootstrap method of interior-branch test for phylogenetic trees. *Mol. Biol. Evol. 13:605.*

Smith, G.P. 1976. Evolution of repeated DNA sequences by unequal crossover. *Science 191:528.*

Su, C. and M. Nei 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 96:9710.*

Tajima, F. 1993. Simple methods for testing the molecular evolutionary clock hypothesis. *Genetics 135:599.*

Thompson, J.D., D.G. Higgins, and T.J. Gibson. 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.*

Tonnelle, C., P.A. Cazenave, C. Brezin, D. Moinier, and M. Fougereau. 1983. Structural correlates to the rabbit immunoglobulin heavy chain a100 allotype. *Mol. Immunol. 20:753.*

Tutter, A., and R. Riblet. 1989. Evolution of the immunoglobulin heavy chain variable region (Igh-V) locus in the genus *Mus*. *Immunogenetics 30:315.*

Vajdy, M., P. Sethupathi, and K.L. Knight. 1998. Dependence of antibody somatic diversification on gut-associated lymphoid tissue in rabbits. *J. Immunol. 160:2725.*

van der Loo, W., F. Mougel, M.S. Sanchez, C. Bouton, E. Castien, A. Fonseca, N. Ferrand, R. Soriguer, and M. Monnerot. 1999. Cytonuclear disequilibria in wild populations of rabbit (*Oryctolagus cuniculus* L.) suggest unequal allele turnover rates at the *b* locus (*IGKC1*). *Immunogenetics 49:629.*

Xia, X., and Z. Xie. 2001. DAMBE: software package for data analysis in molecular biology and evolution. *J. Hered. 92:371.*

Zhu, X., A. Boonthum, S.-K. Zhai, and K.L. Knight. 1999. B lymphocyte selection and age-related changes in $V_H$ gene usage in mutant Alicia rabbits. *J. Immunol. 163:3313.*

## Genetic diversity in *Lepus* species

In order to evaluate the possibility of a trans-specific polymorphism of locus *a* in lagomorphs, we performed a serological analysis of several *Lepus* populations, using allo-antisera against rabbit *a1*, *a2* and *a3* allotypes.

### *Population distribution*

The gene frequencies obtained are presented in Table 5. The frequency of *"a-blank"* allotype was calculated as for the rabbit populations. In the three different *Lepus* species studied only two phenotypes were observed, *a2v* and *"a-blank"*. This result differs from that obtained with *Lepus americanus*, where cross-reaction with rabbit anti- *a1*, *-a2* and *-a3* antisera and the absence of phenotype *blank* were observed (De Poorter, 1984).

**Table 3.5**

Gene frequencies obtained in populations of *Lepus granatensis*, *L. europaeus* and *L. capensis* at locus *a*. Serotype codes are shown in brackets.

| | n | *a1* *(100)* | *a2* *(010)* | *a3* *(001)* | *a1v* *(p00)* | *a2v* *(0p0)* | *a3v* *(00p)* | *blank* *(000)* |
|---|---|---|---|---|---|---|---|---|
| ***Lepus granatensis*** | | | | | | | | |
| **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 |
| *average unwheighted* | 120 | ---- | ---- | ---- | ---- | *0.29* | ---- | *0.71* |
| ***Lepus europaeus*** | | | | | | | | |
| Lass (France) | 20 | ---- | ---- | ---- | ---- | 0.5 | ---- | 0.5 |
| ***Lepus capensis*** | | | | | | | | |
| Tetouan (Morocco) | 7 | ---- | ---- | ---- | ---- | 0.47 | ---- | 0.53 |

# Article 4

P.J. Esteves, D. Lanning, S.K. Zhai, N. Ferrand, K. L. Knight and W. van der Loo (submitted).

The evolution of the immunoglobulin heavy chain variable region ($IgV_H$) in Leporids: a case of trans-species polymorphism

# The evolution of the immunoglobulin heavy chain variable region ($IgV_H$) in Leporids: a case of trans-species polymorphism

P.J. Esteves*,¶, D. Lanning†, S.K. Zhai†, N. Ferrand*, K. L. Knight†, and W. van der Loo*¶

*, Centro de Investigação em Biodiversidade e Recursos Genéticos (CIBIO-UP) and Departamento de Zoologia e Antropologia - Faculdade de Ciências da Universidade do Porto, Porto, Portugal.
¶Institute of Molecular Biology and Biotechnology, Vrije Universiteit Brussel, Belgium.
†Department of Microbiology and Immunology, Loyola University Chicago, Maywood, IL, USA.

## Abstract

The European rabbit (*Oryctolagus cuniculus*) uses in *VDJ* gene rearrangements preferentially the *D*-proximal Ig heavy chain variable region gene segment, $V_H1$. Three lineages are distinguished at this locus: the a1, a2 and a3 allotypes, which are inherited as codominant Mendelian alleles. The distances between lineages are unusually large and sequence comparisons with human and mouse $V_H$ gene segments have suggested that the $V_H1$ polymorphism has persisted for about 50 Million years (My). By consequence, it should predate the separation between the *Oryctolagus* and *Lepus* genera. To test the trans-generic nature of the polymorphism, we have sequenced $V_H$ gene segments expressed in two European *Lepus* species (*L. granatensis* and *L. europaeus*). Two major lineages were observed, one of which was unrelated to the $V_H1$ alleles expressed in rabbit. In contrast, a second lineage, called *a2L*, was found to be more closely related to the rabbit $V_H1$-*a2* allele than the latter was to its allelic counterparts $V_H1$-*a1* and $V_H1$-*a3*, as would be expected in case the divergence between the $V_H$a alleles preceded the separation of the *Lepus Oryctolagus* split.

# Introduction

The rabbit (*Oryctolagus cuniculus*) immunoglobulin heavy chain variable region ($IgV_H$) possesses three serologically defined allotypes (*a1*, *a2*, *a3* $V_Ha$ allotypes). The *a1* and *a2* allotypic specificities are correlated with several amino acid differences, which are located in framework regions 1 (FR1) and 3 (FR3) (Mage *et al.*, 1984; Knight and Becker, 1990). The $V_H$ region allotypic specificities behave like Mendelian alleles (Oudin, 1956, 1960; Dubiski *et al.*, 1959; Dray *et al.*, 1963; Kim and Dray, 1972, 1973). These lineages differ indeed by some 20% amino acid differences. Both the allelic behaviour and the large genetic distances between the $V_Ha$ allotypes is surprising. The allelic behaviour is explained by the preferential use of the *D* proximal $V_H$ gene segment, $V_H1$, in *VDJ* gene rearrangements (Knight and Becker, 1990; Becker and Knight, 1990): Approximately 80 to 90% of circulating immunoglobulins are derived from this gene segment. The remaining 10 to 20% of serum Ig do not contain the $V_Ha$ allotype-associated amino acids (Friedman *et al.*, 1994; Short *et al.*, 1991). These molecules are called $V_H$ "*a*-negative" or $V_Hn$ and are encoded by separate $V_H$ genes, $V_Hx$, $V_Hy$, and $V_Hz$ (Kim and Dray, 1972; Kim and Dray, 1973; Roux, 1981), which are located at least 50 Kb upstream of the $V_H1$ gene (Knight and Becker, 1990).

There are two mechanisms that could account for the large differences between $V_Ha$ lineages. An acceleration of evolutionary rate and/or very long lineage persistence time. Su and Nei (1999) compared the extent of sequence divergence between the rabbit $V_H1$-*a1*, -*a2* and -*a3* gene segments with that between human and mouse $V_H$ gene sequences. Assuming a normal mutation rate, they concluded that the $V_H1$ polymorphism in the rabbit lineage has persisted for about 50 My, which is much longer than the age of separation between the genera *Oryctolagus* and *Lepus* (Bouton and van der Loo, 1997; Halanych and Robinson, 1999; Su and Nei, 1999). If so, some allelic forms of one species are expected to be more closely related to alleles expressed in related species than to their allelic counterparts ("trans-species", or "trans-generic" polymorphism).

Besides the large interallelic distance, evidence for the trans-species nature of the $V_H1$ polymorphism relies mainly upon the fact that sera of *Lepus* species can show serological cross reactivity with rabbit $V_Ha$-allotype specific antisera (van der Loo *et al.*, 1977, Horng *et*

*al.*, 1980, van der Loo, 1987). While it can allow the distinction of allotypic lineages in *Lepus* populations, it does not warrant phylogenetic clustering of lineages showing cross-reactivity with a same antiserum, nor does the absence of cross-reactivity preclude close relationship with rabbit $V_H1$ genes. Indeed, recently, a fourth allotype (*a4*) was described in wild rabbits of the subspecies *O. c. algirus*, that did not show any cross-reaction with conventional $V_Ha$ specific antisera (*"a-blank"*; Esteves *et al.*, unpublished). These rabbits were found to express preferentially *VH* genes that are clearly monophyletic with $V_H1$ genes of domestic rabbits.

The phylogenetic relationships between $V_H$ gene of rabbits and hares requires therefore the sequence determination in hares. We have determined DNA sequences of $V_H$ genes expressed in specimens of two European *Lepus* species, that were typed either "$V_Ha$-*a2* postive", or "$V_Ha$ blank". The data are in strong support of the trans-generic nature of the $V_Ha$ polymorphism.

# Material and Methods

### *Amplification, cloning and sequencing of rearranged VDJ genes*

Spleen samples were collected from wild specimens in the Iberian Peninsula (*Lepus granatensis*, three specimens), and Austria (*Lepus europaeus*, two specimens). Each sample was serological analysed by double immunodiffusion test as previously described by van der Loo *et al.*, (1974). Total RNA was prepared using TRIzol according to the manufacturer's instructions (Life Technologies, Grand Island, NY). First strand cDNA was prepared from 3μg of RNA and synthesized using oligo (dT) primers (Krug *et al.*, 1987). *VDJ* gene rearrangements were PCR-amplified using primers specific for conserved regions in the $V_H$ leader and in the $J_H$ gene (Zhu *et al.*, 1999). All PCR amplifications proceeded for 30 cycles (each cycle: 94°C for 45 s, 60°C for 45 s, and 72°C for 60 s). PCR products approximately 500 bp in length were gel-purified and cloned into the pGEMT-Easy vector (Promega, Madison, WI). The nucleotide sequences were determined by automated sequencing following the Big Dye Terminator Cycle Sequencing protocol (Perkin-Elmer Applied Biosystems, Foster City, CA).

### *Evolutionary analyses*

The phylogenetic tree shown in Figure 1, was constructed using the following sequences a) rabbit germline $V_H$ genes ($V_H1a1$, $V_H1a2$, $V_H1a3$, $V_Hx$, $V_Hy$ and $V_Hz$); b) rabbit

cDNA $V_H$ gene sequences representative of each allotype (*a1*, *a2*, *a3* and *a4*); c) cDNA $V_H$ genes obtained in this study in *Lepus granatensis* (*Lg*) and *Lepus europaeus* (*Le*) specimens. As outgroups, we used human $V_H$ genes of the *VH3* family (accession number: M99672; M99679; M99682; M99666; X92206). These genes belong to the same phylogenetic group as the rabbit $V_H$ genes, i.e. group C (Ota and Nei, 1994; Nei, 1997; Sitnikova and Su, 1998; Su and Nei, 1999). The alignments were made using the computer program "pileup" of the GCG© package (Wilkinson), as provided by the Belgian EMBnet Node (BEN; *http://www.be.embnet.be*) and Clustal W (Thompson *et al.*, 1994) as provided by the freeware DAMBE package (Xia and Xie, 2001), followed by visual adjustments. The reliability of the trees was examined by the bootstrap (Felsenstein, 1985) and the interior-branch (Rzhetsky and Nei, 1992; Sitnikova, 1996) methods, which produced for each interior branch in the tree the bootstrap probability (BP) and confidence probability (CP) values, respectively. All data analyses shown were conducted by using the computer program MEGA2 (Kumar *et al.*, 2001). We used the "pairwise deletion" option to compute evolutionary distances, but essentially the same results were obtained when the alignment gaps were eliminated. The $V_H$ region of a rearranged *VDJ* gene is somatically diversified by gene conversion and somatic hypermutation. For this reason genetic distances between grouped $V_H$ gene were estimated using the option "compute net average distance between groups" of the MEGA2 package, which corrects for distance components due to variance within groups.

# Results

Serum samples Lg7 and Lg15 *Lepus granatensis* failed to show any cross-reactivity with conventional $V_Ha$-allotype specific alloantisera ($V_Ha$ blank). In contrast, the serum of *Lepus granatensis* Lg19 and of *Lepus europaeus* Le1 and Le2 showed clear cross-reaction with anti-*a2* alloantisera raised in rabbits that were homozygous for the *a3* allele. These reactions were very similar to those obtained with hare sera that were described previously in van der Loo *et al.*, (1977).

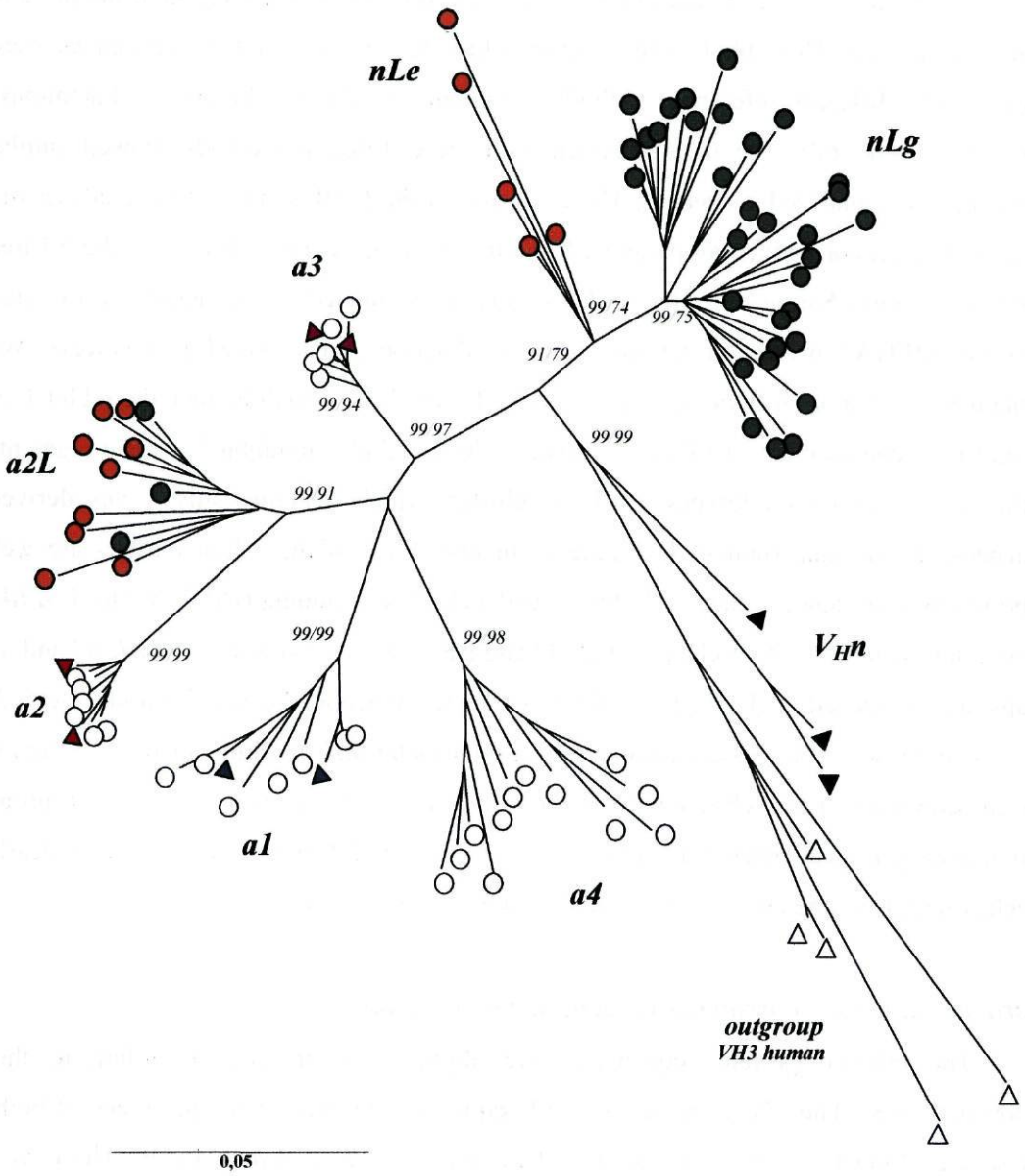



**Figure 1:** Neighbor-Joining tree for *Oryctolagus cuniculus*, *Lepus europeus* and *Lepus granatensis VH* genes. VH genes from human class 3 determine the root of the tree. Numbers represent the Confidence Probability (CP, in bold) and the bootstrap (BP, italic) values when 1000 replicate samplings were done. Expressed $V_H$ genes (cDNA): ● *Lepus europaeus*; ● *Lepus granatensis*; ○ *Oryctolagus cuniculus*; Germline $V_H$ gene sequences: *Oryctolagus cuniculus* allotypes *a1*:▲, *a2*: ▲, *a3*:▲ and $V_Hn$ genes:▲; $V_H3$ human (outgroup):△.

*Phylogenetic tree*

A total of 57 $V_H$ gene nucleotide sequences were determined and their amino acid sequences inferred. Their phylogenetic relationships to known rabbit $V_H$ sequences were studied using different inference methods (Minimum Evolution, Maximum Parsimony, Maximum Likelihood). The trees obtained with these different methods showed similar clustering of the analysed sequences. The exclusion of the CDR regions in the analyses did not affect the general branching patterns. In Figure 1 we present therefore only the NJ tree (Neighbor-Joining; Saitou and Nei, 1987) of the entire nucleotide sequences constructed using the MEGA2 package (Options: pairwise deletions, uncorrected p-distances). We distinguish four major clusters: a) $V_H$ *a*-negative cluster ($V_Hn$), which include the rabbit $V_Hx$, $V_Hy$ and $V_Hz$ gene segments, b) $V_Ha$*-like* cluster, which include the rabbit $V_H1$ gene segments of the *a1, a2,* and *a3* allotypes, c) $V_HnL$ cluster, which contains only *Lepus* derived sequences; d) outgroup (human $V_H3$ gene segments). Each of these four clusters are well supported by confidence values (CP=99) as well as by bootstrapping (BP>74). The $V_Ha$*-like* group in turn consists of four clusters (CP=99 and BP>91). Three of these, the *a1*, *a3* and *a4* groups, are composed exclusively of rabbit sequences, whereas the fourth group, the "*a2-like*", is composed of sequences obtained from both rabbit and hare specimens. The *Lepus*-derived sequences of this cluster will be referred to as "*a2L*" sequences. The $V_HnL$ group encompasses gene sequences from both *Lepus europaeus* and *Lepus granatensis,* and clearly branches apart from the clusters including the rabbit $V_H1$ sequences.

*Characterisation of rearranged $V_H$ gene segments in hare*

The inferred protein sequences were aligned and grouped according to the phylogenetic tree. The *a2L* group consists of sequences obtained from specimens of both hare species studied, *Lepus europaeus* (Le1, Le2) and *Lepus granatensis* (Lg19). The $V_HnL$-like group can be divided into two subgroups, of which one *(nLg)* is composed of sequences obtained from *Lepus granatensis* (Lg7, Lg15, Lg 19), and one *(nLe)* from *Lepus europaeus* (Le1, Le2), respectively. We derived a consensus sequence of the amino acid sequences inside each group, i.e. Cons-*a2L*, Cons-*nLg* and Cons-*nLe* (Figure 2A). In Figure 2B we compare these consensus sequences with the inferred amino acid sequences of the $V_H1a1$, $V_H1a2$, $V_H1a3$, $V_Hx$, $V_Hy$ and $V_Hz$ gene segments, and with Cons-*a4.1* and Cons-*a4.2*, which are the consensus sequences of two variant forms of the allotype *a4* (Esteves *et al.*, unpublished).

**Figure 2.** Alignment of inferred amino acid sequences of $V_H$ genes. **A**: $V_H$ gene sequences of *Lepus europaeus* and *Lepus granatensis* that form the *a2L*, n*Le* and *nLg group*. CONS_*a2L*, CONS_*nLe* and CONS_*nLg* are the established consensus sequences. LE1_*: *Lepus europaeus* 1; LE2_*: *Lepus europaeus* 2; Lg19_*: *Lepus granatensis* 19. **B:** Sequences of rabbit and *Lepus* $V_H$ genes. CONS-vh1 is the established consensus sequence. GL-a1_Vh1, GL-a2_Vh1, GL-a3_Vh1: rabbit germline VH1gene of allotypes a1, a2 and a3, respectively; CONS a4.1, CONS a4.2: consensus sequence established for allotypes *a4.1* and *a4.2*; GL_Vhx, GL_Vhy, GL_Vhz: rabbit germline Vhx, Vhy and Vhz, respectively. The hypervariable region of CDRs are shaded and were excluded from the analyses. Dots (.) indicate identity with the consensus sequence; Dashes (-) represent indels.

The *a2L* sequences have the signature of $V_Ha$ alleles. They share a number of apomorphic characters with the $V_Ha$ genes, such as the $V_H$a hallmark peptides 18LTLTCT23 of FR1, and 62WAK64 of FR3 (amino acid numbering of Kabat 1987). Furthermore, they possess five out of eleven amino acid residues that characterize the allotype $V_Ha$-*a2* (Mage *et al.*, 1984), *i.e.* two in the FR1: K5 and T17, and three in the FR3: S65, R71, and N74. Another feature that they have in common with the $V_H$*1-a2* sequence is the absence of deletions in the FR3 (positions 72 and 74). These shared features might explain the cross-reactivity observed in immunodiffusion tests. Among the leporid sequences shown, the *a2L* sequences are characterized by amino acids residues I12 and E16 (FR1), and C67 (FR3). I12 is also observed in the rabbit *VH4-a2* and *VH7-a2* gene segments. The lineages *nLg* and *nLe* differ at six positions: Q/K13, T/S21, E/K45, G/S80b, T/V84, T/E85. It is interesting that the two *nL* lineages differ at positions 84 and 85, because these positions vary among rabbit allotypes *a1*, *a2/a3*, *a4.1*, and *a4.2* and are known to define allotypic determinants (Mage *et al.*, 1984).

### *Net-distances between rabbit and hare $V_H$ sequences*

The amino acid p-distance between the different clusters of diversified $V_H$ sequences, are presented in Table 1. Other distance methods essentially provided the same result. The net-distances between the (diversified) rabbit *a2* allotype sequences and *a1, a3, a4.1* and *a4.2* allotypes vary between 0.199-0.220, which is significantly larger than the net-distance observed between the rabbit *a2* sequences and the *Lepus a2L* sequences (0.121 +/- 0.028). This supports the phylograms suggesting that the *a2* and *a2L* lineages diverged after the separation of the major $V_Ha$ lineages (Figure 1). The distance between the *a2L* sequences and the *VH* gene groups *nLg* and *nLe* was 0.206 and 0.179 respectively, which is of the same order as that observed between allotypic $V_Ha$ lineages of rabbit.

**Table 1**

Distances between rabbit and hare $V_H$ genes. Net distances were determined using the "uncorrected p-distance" of the MEGA2 package, applied to diversified rabbit and hare VH sequences representative of different clusters defined in Figure 1. The genetic distances are represented below and the standard deviations above the diagonal.

| | *a1* | *a2* | *a3* | *a4.1* | *a4.2* | *a2L* | *nLg* | *nLe* | *n* |
|---|---|---|---|---|---|---|---|---|---|
| *a1* | ---- | 0.039 | 0.035 | 0.031 | 0.030 | 0.033 | 0.041 | 0.038 | 0.040 |
| *a2* | 0.220 | ---- | 0.036 | 0.036 | 0.034 | 0.028 | 0.039 | 0.037 | 0.039 |
| *a3* | 0.153 | 0.208 | ---- | 0.031 | 0.028 | 0.030 | 0.033 | 0.032 | 0.030 |
| *a4.1* | 0.151 | 0.206 | 0.151 | ---- | 0.015 | 0.031 | 0.035 | 0.034 | 0.035 |
| *a4.2* | 0.163 | 0.199 | 0.149 | 0.050 | ---- | 0.028 | 0.032 | 0.031 | 0.031 |
| *a2L* | 0.177 | 0.121 | 0.167 | 0.162 | 0.148 | ---- | 0.032 | 0.031 | 0.022 |
| *nLg* | 0.264 | 0.307 | 0.183 | 0.214 | 0.190 | 0.206 | ---- | 0.014 | 0.014 |
| *nLe* | 0.243 | 0.271 | 0.169 | 0.219 | 0.194 | 0.179 | 0.039 | ---- | 0.016 |
| *n* | 0.249 | 0.272 | 0.143 | 0.243 | 0.214 | 0.140 | 0.186 | 0.163 | ---- |

# Discussion

There are different ways to explain the unusually high allelic divergence observed in rabbit $V_H1$ polymorphism: very long lineage persistence time and/or an acceleration of the evolutionary rate within lineages. If the latter hypothesis is correct, the mutation rate is higher than that estimated under neutrality, and the time of divergence between the alleles will be overestimated. If we assume that evolutionary rates were normal, the allelic lineages must have existed before the separation of the *Lepus* and *Oryctolagus* genera took place. In that case, allelic forms in distinct species can be more closely related than are allelic forms in the same species. *MHC* allelic lineages are prime examples of this phenomenon, having persisted for more than 25 MY in humans and in the old world monkeys (Lawlor *et al.*, 1988; Fan *et al,* 1989; Mayer *et al.*, 1992; Klein *et al.*, 1993).

*Oryctolagus* originated in the Iberian Peninsula, where two subspecies (*O. c. cuniculus* and *O.c. algirus*) have coexisted with *Lepus* species for more than 6 My (Lopez-Martinez 1989, Callou 1995). In this study, we found two $V_H$ lineages in *Lepus* species: one that is clearly part of the cluster formed by the rabbit $V_H$a-like genes (*a2L*), and another that clearly is not (*nL*). Comparison of these lineages with the rabbit alleles (*a1*, *a2*, *a3*, *a4.1* and *a4.2*) revealed that the *a2L* lineage is more similar to the *a2* lineage than the latter is to its allelic counterparts *a1* and *a3* (see Figure 1 and Table 1). These results strongly suggest that the high allelic divergence of rabbit $V_Ha$ alleles is at least in part due to prolonged lineage persistence times.

It seems that in $V_Ha$-positive typing hares, genes encoding *a2L* and *nL* proteins are equally expressed (14 *nL* sequences *vs.* 13 *a2L* sequences for *Lg19, Le1*, and *Le2*). $V_HnL$ genes were expressed in all *Lepus* specimens, and were the only $V_H$ genes detected in *Lg7* and *Lg15* forming a thigh cluster in phylogenetic analysis. In the absence of breeding studies, it would be highly presumptions to speculate about the allelic nature of the *a2L* and *nL* coding genes. Genomic mapping will furthermore be necessary for the understanding of the mechanism that regulates the expression of these $V_H$ gene lineages. This does however not diminish the evidence for the hypothesis that the common ancestors of $V_H1$ alleles of modern rabbits have started their divergence long before the rabbit-hare split.

The origin and maintenance of $V_H$a polymorphism is unknown. Su and Nei (1999) suggested that the presence of only three highly divergent alleles indicates that each might have some specific adaptive significance and proposed that the three known lineages (*a1*, *a2*

and *a3*) have been maintained by overdominant selection. 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), diversification of the primary antibody repertoire in rabbits is not developmentally programmed. It has been shown that Ab repertoire diversification in the rabbit requires interaction with exogenous factors, such as the intestinal microflora (Lanning *et al.*, 2000a; 2000b, Sehgal *et al.*, 2002). It is possible that the mechanism driving diversification of the primary antibody repertoire might impose constraints on the evolution of the $V_H a$ polymorphism.

This study shows how extending the studies on the rabbit $V_H a$ allotypes to other lagomorph species, can contribute to a better comprehension of the evolution and the biological meaning of this unusual polymorphism.

# Citations

Becker RS and Knight KL (1990) Somatic diversification of immunoglobulin heavy chain VDJ genes: evidence for somatic gene conversion in rabbits. *Cell* **63**(5): 987-97.

Bouton C and van der Loo W (1997) The trans-species nature of rabbit b locus polymorphism is supported by studies on the snowshoe hare. *Immunogenetics* **45**(6): 444-46.

Callou C (1995). Modifications de l'aire de répartition du Lapin (*Oryctolagus cuniculus*) en France et en Espagne, du Pléistocène à l'époque actuelle. État de la question. *Anthropozoologica* **21**: 95-114.

Dray SG, Young O and Nisonoff A (1963). Distribution of allotypic specificities among rabbit immunoglobulin γ-globulin molecules genetically defined at two loci. *Nature* **199**: 52.

Dubiski S, Dudziak Z, Skalba D and Dubiski A (1959). Serum groups in rabbits. *Immunology* **2**: 84.

Fan WM, Kasahara M, Gutknecht J, Klein D, Mayer WE, Jonker M and Klein J (1989). Shared class II MHC polymorphisms between humans and chimpanzees. *Hum Immunol* **26**(2): 107-21.

Felsenstein J (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, **39**: 783-91.

Friedman ML, Tunyaplin C, Zhai SK and Knight KL (1994). Neonatal $V_H$, D, and JH gene usage in rabbit B lineage cells. *J Immunol* **152**(2): 632-41.

Halanych KM and 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): 369-79.

Horng WJ, Papagiannes E, Dray S and 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):111-17.

Kabat EA, Wu TT, Perry HM, Gottsman KS and Foeller C (1991). *Sequences of Proteins of Immunologic interest*, 5th Ed. U.S. Departament of Health and Human Services, Public Health Service, National institutes of Health, Bethesda, MD.

Kim BS and 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): 509-14.

Kim BS and 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): 750-60.

Klein J, Satta Y, Takahata N and O'hUigin C (1993). Trans-specific Mhc polymorphism and the origin of species in primates. *J Med Primatol* **22**(1): 57-64.

Knight KL and Becker RS. (1990). Molecular basis of the allelic inheritance of rabbit immunoglobulin $V_H$ allotypes: implications for the generation of antibody diversity. *Cell* **23**: 60(6): 963-70.

Krug MS and Berger S.L. (1987). First strand cDNA synthesis primed with oligo(dT). *Methods Enzymol* **152**: 316.

Kumar S, Tamura K, Jakobsen IB and Nei M. (2001). MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* **17**(12): 1244-245.

Lanning D, Zhu X, Zhai SK and Knight KL (2000a). Development of the antibody repertoire in rabbit: gut-associated lymphoid tissue, microbes, and selection. *Immunol Rev* **175**: 214-28.

Lanning D, Sethupathi P, Rhee KJ, Zhai SK and Knight KL (2000b). Intestinal microflora and diversification of the rabbit antibody repertoire. *J Immunol* **165**(4): 2012-019.

Lawlor DA, Ward FE, Ennis PD, Jackson AP and Parham P (1988). HLA-A and B polymorphisms predate the divergence of humans and chimpanzees. *Nature* **335**(6187): 268-71.

Lopez-Martinez N (1989). *Revision sistematica y biostratigrafica de los lagomorphos (Mammalia) del Terciario y Cuaternario de España.* Memorias del Museo Paleontologico de la Universidad de Zaragoza, nº3. Diputacion General de Aragon.

Lucier MR, Thompson RE, Waire J, Lin AW, Osborne BA and Goldsby RA (1998). Multiple sites of V lambda diversification in cattle. *J. Immunol* **161**(10): 5438-444.

Mage RG, Bernstein KE, McCartney-Francis N, Alexander CB, Young-Cooper GO, Padlan EA and Cohen GH. (1984). The structural and genetic basis for expression of normal and latent $V_Ha$ allotypes of the rabbit. *Mol Immunol*, **21**(11): 1067-081.

Mayer WE, O'hUigin C, Zaleska-Rutczynska Z and Klein J.(1992). Trans-species origin of Mhc-DRB polymorphism in the chimpanzee. Immunogenetics, 37(1): 12-23.

Nei M, Gu X and 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): 7799-806.

Ota T and Nei M (1994). Divergent evolution and evolution by the birth-and-death process in the immunoglobulin $V_H$ gene family. *Mol Biol Evol*,**11**(3): 469-82.

Oudin J (1956). L'"allotypie" de certains antigènes proteídique du sérum. *Comp. Rend. Acad. Sci. Paris* **242**: 2606-608.

Oudin J (1960). Allotypy of rabbit serum proteins. I. Immunochemical analysis leading to the individualization of seven main allotypes. *J Exp Med* **112**: 107-24.

Reynaud CA, Dahan A, Anquez V and Weill JC (1989). Somatic hyperconversion diversifies the single Vh gene of the chicken with a high incidence in the D region. *Cell*, **59**(1): 171-83.

Reynaud CA, Garcia C, Hein WR and Weill JC (1995). Hypermutation generating the sheep immunoglobulin repertoire is an antigen-independent process. *Cell*, **80**(1): 115-25.

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): 626-32.

Rzhetsky A and Nei M (1992). Statistical properties of the ordinary least-squares, generalized least-squares, and minimum-evolution methods of phylogenetic inference. *J Mol Evol*, **35**(4): 367-75.

Saitou N and Nei M (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*, **4**(4): 406-25.

Sehgal D, Obiakor H and Mage RG (2002). Distinct clonal Ig diversification patterns in young appendix compared to antigen-specific splenic clones. *J Immunol*, **168**(11): 5424-433.

Sitnikova T (1996). Bootstrap method of interior-branch test for phylogenetic trees. *Mol Biol Evol*, **13**(4):605-11.

Sitnikova T and Su C (1998). Coevolution of immunoglobulin heavy- and light-chain variable-region gene families. *Mol Biol Evol*, **15**(6): 617-25.

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): 4014-018.

Su C and 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): 9710-5.

Thompson JD, Higgins DG and 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): 4673-680.

van der Loo W, De Baetselier P, Hamers-Casterman C and 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 (1987). In *The rabbit in Contemporary Immunological Research*, ed. Dubiski, S. (Longman, Harlow, UK), pp. 164-190.

Xia X and Xie Z (2001). DAMBE: software package for data analysis in molecular biology and evolution. *J Hered* **92**(4): 371-73.

Zhu X, Boonthum A, Zhai SK and Knight KL (1999). B lymphocyte selection and age-related changes in VH gene usage in mutant Alicia rabbits. *J Immunol* **163**(6): 3313-120.

## Understanding the large inter-allelic differences at the $V_H1$ locus.

Evolution theory offers two possible explanations for interallelic distances as large as those observed at the rabbit $IgV_H1$ locus: unusually long allele turnover times, or increased mutant recruitment rates. The fact that allo-antisera raised in rabbit allowed to distinguish three $V_H$ allotypes in populations of *Lepus americanus* has suggested that both species have inherited the same three lineages, which could explain the large interallelic distances without invoking acceleration of evolution rates (Su and Nei, 1999). In this case allelic forms in distinct species will be more related than are allelic forms in the same species.

In Article 3 we show that there are for the rabbit *IgVH1* locus not three but at least four highly divergent lineages in Iberian wild rabbit populations. The fourth lineage (*a4*) is associated with the maternal markers of subspecies *algirus* and did not cross the Pyrenean Mountains, which is in contradiction with extremely long allele persistence times. It suggests that the large differences between *a4* and other $V_H1$ alleles may have accumulated after the separation of the subspecies, estimated in 2 million years ago.

On the other hand, in Article 4 we observe the existence of a $V_H$ lineage in *Lepus* species (*a2L*) that clusters with sequences of the rabbit *a2* lineage, which supports the hypothesis of very long lineage persistence time. However, the serological analyses indicate that some allelic lineages present in *L. americanus* could be absent in *L. granatensis*, *L. europaeus* and *L. capensis*, suggesting that in the *Lepus* speciation process some of these lineages were lost.

These two different lines of evidence suggest, on one hand, that allelic lineages can be lost stochastically and, on the other hand, that different mutation rates can be associated to each allelic lineage. The study of $V_H$ genes of *Lepus americanus* could help to clarify this situation.

## Evolution of $V_H$ genes in the lagomorph group

### *Phylogenetic analyses*

The lagomorph $V_H$ gene sequences used in this analysis were obtained by "random" PCR sampling among more than 100 $V_H$ gene segments. To construct the NJ tree present in Figure 3.9 only functional $V_H$ genes were considered. The major vertebrate $V_H$ genes classes B and C are well differentiated. Inside of class C three clusters are well supported: 1) lagomorph $V_H$ genes (rabbits, hares, cottontails and pikas; 2) bird $V_H$ genes; 3) other

mammals (camel, human and mice). Four clusters compose the lagomorph group. These are: a) the $V_Ha$ group that encompasses *a1*, *a3*, *a4* and *a2-like* groups; b) $V_HnL$ group that includes both expressed and germline $V_H$ gene sequences of *Lepus granatensis*; c) $V_Hn$ group, that contains germline $V_H$ genes of *Oryctolagus*, *Sylvilagus*, *Ochotona*, *Lepus granatensis* and *L. americanus*; d) $V_HP$ group, represented only by pika germline sequences. The $V_Ha$ group contains three clusters (*a1*, *a3* and *a4*), which are composed exclusively by rabbit sequences. In contrast, the *a2-like* cluster includes sequences of three different Leporid genera, i. e. the rabbit *a2* allele, the lineage *a2L* of *Lepus granatensis* (cDNA) and genomic $V_H$ gene sequences obtained in *Lepus americanus* and *Sylvilagus floridanus*.

*Characterization of lagomorph germline* $V_H$ *gene sequences*

The amino acid sequences of $V_H$ germline genes obtained in this work were compared to other rabbit germline sequences available in Genbank and with the consensus sequences of *a4*, *a2L* and *nL* allelic lineages (see article 3 and 4) in the Figure 3.10.

These results suggest that only $V_H$ sequences belonging to group C are present in the genome of lagomorph species. Probably, the loss of $V_H$ genes from group A and B happened in an ancestral of the lagomorphs at least 50 My years ago, which is the time of divergence estimated between pika and rabbit (Springer *et al.*, 2003). The lagomorph $V_H$ gene sequences define an independent group. Within this cluster the $V_H$ gene sequences are separated into four major groups: $V_Ha$, $V_Hn$, $V_HnL$ and $V_HP$. The $V_H$ genes sequences from different species can be clustered in the same group and $V_H$ gene sequences of the same species are sometimes clustered in different groups. It seems likely that in the lagomorphs, the duplication of $V_H$ genes is in accordance with the model of birth-and-death evolution.

It has been shown that in the rabbit the diversification process is not developmentally regulated like it appears to be in other species, such as chicken, sheep and cattle, and requires interaction of exogenous factors, such as the intestinal microflora (Lanning *et al.*, 2000a; 2000b, Sehgal *et al.*, 2002). We speculate that the intestinal microflora is probably the mechanism that is responsible for the occurrence of highly divergent $V_H$ gene sequences within the lagomorph group.

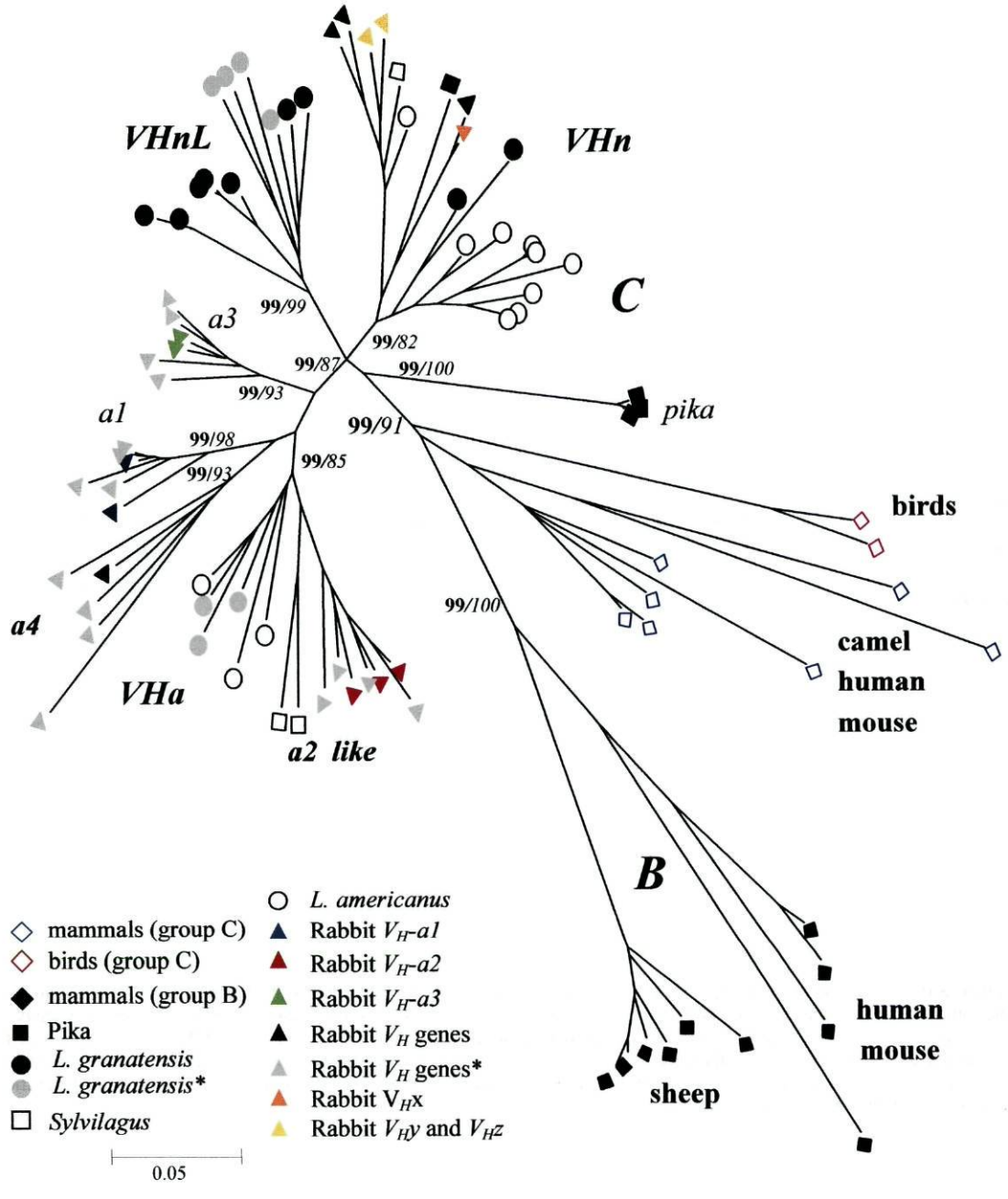



**Figure 3.9.** Neighbor-Joining tree for *Oryctolagus cuniculus*, *Lepus granatensis*, *Lepus americanus*, *Sylvilagus*, *Ochotona* and other mammals and birds $V_H$ genes clustered in group C. $V_H$ genes from mammals belonging to group B determine the root of the tree. Numbers represent the Confidence Probability (CP, in bold) and the bootstrap (BP, italic) values when 1000 replicate samplings were done. The expressed *VH* gene sequences (cDNA) are indicated by *.

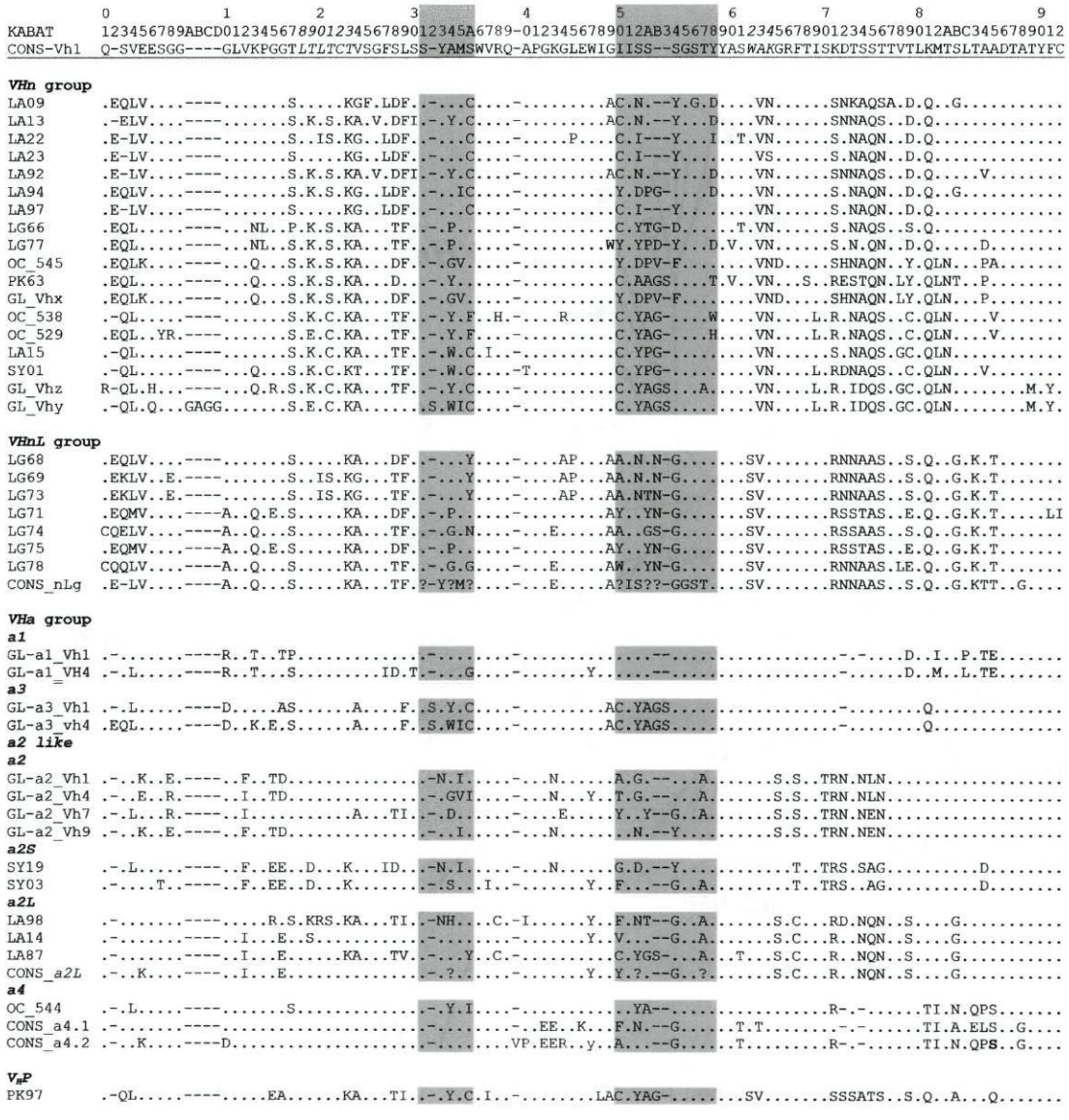


**Figure 3.10.** Alignment of lagomorph $V_H$ gene sequences. The amino acid $V_H$ regions are inferred from: 1) genomic $V_H$ nucleotide sequences obtained in this work for *Oryctolagus* (OC), *Sylvilagus* (SY), *Lepus americanus* (LA), *Lepus granatensis* (LG) and *Ochotona* (PK); 2) genomic rabbit $V_H$ nucleotide sequences (GL) published in Genbank. CONS_*a4.1*, *a4.2*, *nLg* and *a2L* are the consensus sequences of the lineages *a4.1*, *a4.2*, *nLg* and *a2L*, obtained from the $V_H$ gene expressed (cDNA). They are compared to a consensus sequence of rabbit $V_H1$ gene segment (CONS-VH1) and grouped according to phylogenetic relationships, shown in Figure 3.9.

# Citations

Allegrucci M, Newman BA, Young-Cooper GO, Alexander CB, Meier D, Kelus AS and 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): 5444-8.

Asmussen MA, Arnold J (1991). The effects of admixture and population subdivision on cytonuclear disequilibria. *Theor Popul Biol* **39**(3): 273-300.

Becker RS and Knight KL (1990). Somatic diversification of immunoglobulin heavy chain *VDJ* genes: evidence for somatic gene conversion in rabbits. *Cell* **63(5):** 987-97.

Brézin C, Cazenave PA (1980). Allotypy of rabbit immunoglobulins: a fifth allele at the a locus. *J Immunol* **125**(1): 59-62.

Brézin C, Benammar A, Roland J, Cazenave PA (1979). A and b allotypy in *Oryctolagus* and *Lepus* species. *Ann Immunol (Paris)* **130**(2):167-78.

Cazenave PA, Brézin C and Roland J (1974). An allotypic specificity presumably of the *a* series in rabbit immunoglobulins, different from *a1*, *a2*, and *a3*. *Biochem Biophys Res Commun* **61**(2): 664-70.

Chen HT, Alexander CB, Young-Cooper GO and Mage RG (1993). VH gene expression and regulation in the mutant Alicia rabbit. Rescue of VHa2 allotype expression. *J Immunol* **150**(7): 2783-93.

Cooper MD, Perey DY, Gabrielsen AE, Sutherland DE, McKneally MF, Good RA Production of an antibody deficiency syndrome in rabbits by neonatal removal of organized intestinal lymphoid tissues. *Int Arch Allergy Appl Immunol* 1968; **33**(1): 65-88.

Crane MA, Kingzette M and Knight KL (1996). Evidence for limited B-lymphopoiesis in adult rabbits. *J Exp Med* **183**(5): 2119-21.

Currier SJ, Gallarda JL and Knight KL (1988). Partial molecular genetic map of the rabbit VH chromosomal region. *J Immunol* **140**(5): 1651-9.

De Poorter M (1984). *An experimental test of predictions from different hypothesis of self regulation in the snowshoe hare* (*Lepus americanus* Erxleben, 1777). PhD Thesis, Vrije Universiteit Brussel.

Di Pietro LA, Short JA, Zhai SK, Kelus AS, Meier D and Knight KL (1990). Limited number of immunoglobulin VH regions expressed in the mutant rabbit "Alicia". *Eur J Immunol* **20**(6): 1401-4.

Dray SG, Young O, Nisonoff A (1963). Distribution of allotypic specificities among rabbit immunoglobulin γ-globulin molecules genetically defined at two loci. *Nature*, **199**:52.

Dubiski S, Dudziak Z, Skalba D and Dubiski A (1959). Serum groups in rabbits. *Immunology* **2**: 84.

Felsenstein J (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, **39**: 783-791.

Gallarda JL, Gleason KS and Knight KL (1985). Organization of rabbit immunoglobulin genes. I. Structure and multiplicity of germ-line VH genes. *J Immunol* **135**(6): 4222-8.

Gojobori T and Nei M (1984). Concerted evolution of the immunoglobulin VH gene family. *Mol Biol Evol* **1**(2): 195-212.

Halanych KM and 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): 369-79.

Haouas H, Benammar-El Gaaied A, Cazenave PA (1987). A "new allotype" of the *a* series in rabbit immunoglobulins (*a*107). *Mol Immunol* **24**(3): 247-51.

Haouas H, el Gaaied A and Cazenave PA (1989). Genetic polymorphism of rabbit VHa region: a new allotype, *a*108. *Res Immunol* **140**(3): 265-73.

Haouas H, Benammar-el Gaaied A (1994). A genetic marker of rabbit immunoglobulin VHa region: *a*109. *Mol Immunol* **31**(16): 1233-238.

Hood L, Campbell JH and Elgin SC (1975). The organization, expression, and evolution of antibody genes and other multigene families. *Annu Rev Genet* **9**:305-53.

Kabat EA, Wu TT, Perry HM, Gottesman KS, and Foeller C (1991). *Sequences of proteins of immunological interest*, 5th edn. Public health service, NIH, Besthesda, MD.

Kelus AS and Weiss S (1986). Mutation affecting the expression of immunoglobulin variable regions in the rabbit. *Proc Natl Acad Sci U S A* **83**(13):4883-886.

Kim BS and 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): 509-14.

Kim BS and 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): 750-60.

Kindt TJ (1975). Rabbit immunoglobulin allotypes: structure, immunology, and genetics. *Adv Immunol* **21**:35-86.

Klein J, Ono H, Klein D and O'hUigin C (1993). The accordion model of MHC evolution. *Prog. Immunol* **8**: 137-143.

Knight KL and Becker RS (1990). Molecular basis of the allelic inheritance of rabbit immunoglobulin VH allotypes: implications for the generation of antibody diversity. *Cell* **60**(6): 963-70.

Knight KL (1992). Restricted VH gene usage and generation of antibody diversity in rabbit. *Annu Rev Immunol* **10**: 593-616.

Knight KL and Crane MA (1994). Generating the antibody repertoire in rabbit. *Adv Immunol* **56**: 179-218.

Kumar S, Tamura K, Jakobsen IB and Nei M (2001). MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* **17**(12): 1244-245.

Lanning D, Zhu X, Zhai SK, Knight KL (2000a). Development of the antibody repertoire in rabbit: gut-associated lymphoid tissue, microbes, and selection. *Immunol Rev*, **175**:214-28.

Lanning D, Sethupathi P, Rhee KJ, Zhai SK and Knight KL (2000b). Intestinal microflora and diversification of the rabbit antibody repertoire. *J Immunol*, 165(4): 2012-019.

Lucier MR, Thompson RE, Waire J, Lin AW, Osborne BA and Goldsby RA (1998). Multiple sites of VL diversification in cattle. *J Immunol* **161**: 5438-444.

Lummus Z, Cebra JJ and 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): 737-43.

Mage RG (1967). Quantitative studies on the regulation of expression of genes for immunoglobulin allotypes in heterozygote rabbits. *Cold Spring Harbor Symp Quant Biol* **32**: 203-210.

Mage RG, Bernstein KE, McCartney-Francis N, Alexander CB, Young-Cooper GO, Padlan EA and Cohen GH. (1984). The structural and genetic basis for expression of normal and latent *VHa* allotypes of the rabbit. *Mol Immunol*, **21**(11): 1067-81.

Margolies MN, Cannon LE, Kindt TJ and 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): 287-94.

Medrano L and Dutrillaux B (1984). Chromosomal location of immunoglobulin genes: partial mapping of these genes in the rabbit and comparison with Ig genes carrying chromosomes of man and mouse. *Adv Cancer Res* **41**: 323-67.

Nei M, Gu X and 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): 7799-806.

Ohta T (1983). On the evolution of multigene families. *Theor Popul Biol* **23**(2): 216-40.

Ota T and Nei M (1994). Divergent evolution and evolution by the birth-and-death process in the immunoglobulin VH gene family. *Mol Biol Evol* **11**(3):469-82.

Oudin J (1956). L'allotopie de certains antigènes proteidique du sérum. *C. R. Acad. Sci.* (Paris) **242**: 2606.

Oudin J (1960). Allotypy of rabbit serum proteins. I. Immunochemical analysis leading to the individualization of seven main allotypes. *J Exp Med* **112:** 107-24.

Parng CL, Hansal S, Goldsby RA and Osborne BA (1996). Gene conversion contributes to Ig light chain diversity in cattle. *J Immunol* **157(12):** 5478-486.

Pospisil R, Young-Cooper GO and 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): 6961-965.

Reynaud CA, Anquez V, Dahan A and Weill JC (1985). A single rearrangement event generates most of the chicken immunoglobulin light chain diversity. *Cell* **40(2):** 283-91.

Reynaud CA, Dahan A, Anquez V and Weill JC (1989). Somatic hyperconversion diversifies the single Vh gene of the chicken with a high incidence in the D region. *Cell* **59**(1): 171-83.

Reynaud CA, Mackay CR, Muller RG and Weill JC (1991). Somatic generation of diversity in a mammalian primary lymphoid organ: the sheep ileal Peyer's patches. *Cell* **64(5):** 995-1005.

Reynaud CA, Garcia C, Hein WR and Weill JC (1995). Hypermutation generating the sheep immunoglobulin repertoire is an antigen-independent process. *Cell*, **80(1):** 115-25.

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): 626-32.

Rzhetsky A and Nei M (1992). Statistical properties of the ordinary least-squares, generalized least-squares, and minimum-evolution methods of phylogenetic inference. *J Mol Evol*, **35**(4): 367-75.

Schroeder HW Jr, Hillson JL and Perlmutter RM (1990). Structure and evolution of mammalian VH families. *Int Immunol* **2**(1): 41-50.

Sehgal D, Mage RG and 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): 1246-255.

Sehgal D, Obiakor H and Mage RG (2002). Distinct clonal Ig diversification patterns in young appendix compared to antigen-specific splenic clones. *J Immunol*, **168**(11):5424-33.

Short JA, Sethupathi P, Zhai SK and 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): 4014-8.

Sitnikova T (1996). Bootstrap method of interior-branch test for phylogenetic trees. *Mol Biol Evol*, **13**(4):605-11.

Sitnikova T and Su C (1998). Coevolution of immunoglobulin heavy- and light-chain variable-region gene families. *Mol Biol Evol* **15**(6): 617-25.

Smith GP, Hood L and Fitch WM (1971). Antibody diversity. *Annu Rev Biochem* **40:** 969-1012.

Smith GP (1974). Unequal crossover and the evolution of multigene families. *Cold Spring Harb Symp Quant Biol* **38**: 507-13.

Springer MS, Murphy WJ, Eizirik E and O'Brien SJ (2003). Placental mammal diversification and the Cretaceous-Tertiary boundary. *Proc. Natl. Acad. Sci.* U S A **100**(3): 1056-61.

Su C and 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): 9710-5.

Sun J and Butler JE (1996). Molecular characterization of *VDJ* transcripts from a newborn piglet. *Immunology* **88:** 331-39.

Tonnelle C, Cazenave PA, Brezin C, Moinier D and Fougereau M (1983). Structural correlates to the rabbit immunoglobulin heavy chain a100 allotype. *Mol Immunol* **20**(7): 753-61.

Tunyaplin C and Knight KL (1995). Fetal VDJ gene repertoire in rabbit: evidence for preferential rearrangement of VH1. *Eur J Immunol* **25**(9): 2583-7.

Tutter A, Riblet R (1989). Evolution of the immunoglobulin heavy chain variable region (Igh-V) locus in the genus *Mus*. *Immunogenetics* **30**(5): 315-29.

Vajdy M, Sethupathi P and Knight KL (1998). Dependence of antibody somatic diversification on gut-associated lymphoid tissue in rabbits. *J Immunol* **160**(6): 2725-9.

van der Loo W (1982). Study of frequency variances of immunoglobulin alleles in wild rabbit populations. Population Genetic Group Meeting, Cambridge, UK.

van der Loo W and Arthur CH (1986). Etude des associations gametiques entre alleles de quatre loci d'immunoglobuline dans de populations naturelles de Lapin de Garenne. In Legay JM (ed.). *Actes du Colloque Biologie des Populations, Lyon*. University Claude Bernard. pp 374-382.

van der Loo W (1987). Studies on the adaptive significance of the immunoglobulin polymorphisms (Ig allotypes) in wild rabbits. In Dubiski S (ed.). *The Rabbit in Contemporary Immunological Research*. Longman Scientific & Technical. pp 164-190.

van der Loo W, Ferrand N and 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): 789-99.

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): 171-87.

Weinstein PD, Anderson AO and Mage RG (1994). Rabbit IgH sequences in appendix germinal centers: VH diversification by gene conversion-like and hypermutation mechanisms. *Immunity* **1:** 647-659.

Wright S (1978). *Evolution and Genetics of Populations* Vol. 4. University of Chicago Press. pp 79-103.

Zimmer EA, Martin SL, Beverley SM, Kan YW and 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): 2158-62.

Zhu X, Boonthum A, Zhai SK and Knight KL (1999). B lymphocyte selection and age-related changes in VH gene usage in mutant Alicia rabbits. *J Immunol* **163**(6): 3313-20.

# CHAPTER 4

## Evaluation of the bottleneck effect in the Porto Santo wild rabbit population

# Introduction

The presence of rabbits in more than 800 islands all over the world points out the success of this species in the colonization of new environments (Flux and Fullagar, 1992). The reasons for this success are particular features such as high growth and reproduction rate, an efficient food utilisation (associated with coprophagy, i.e. the capacity to eat the faeces increasing nutritional efficiency), and the ability to shift between growth strategies ("r" or "K - selection") according to the carrying capacity of the environment. Other aspects are also fundamental in its success in islands, like the absence of competitors, few predators and inexistence of diseases.

A particularly successful introduction was documented for the Porto Santo, a small island of the Portuguese Madeira archipelago. Several historical documents suggest that a single pregnant female was released in this island by a Portuguese navigator, Bartolomeu Perestrelo, in 1419 (Cordeiro, 1717; Barros, 1778; Azurara, 1841; Frutuoso, 1873). This event was followed by a sudden increase in population numbers (Cadamosto, 1812; Frutuoso, 1873). Charles Darwin, in his book "*The variation of animals and plants under domestication*" (1868), described seven specimens brought from Porto Santo Island as remarkably different in appearance from the British rabbit. For Darwin, the rabbit of Porto Santo was a striking example of divergence in form since its introduction to the island. This opinion was also shared by Hæckel (1874) who considered the Porto Santo rabbit a distinct species named as *Lepus huxley*. Later studies, however, demonstrated by comparison with rabbits from the Iberian Peninsula that Porto Santo rabbits were clearly *Oryctolagus cuniculus* (Miller, 1912; França, 1913), although its domestic or wild origin was not clarified (Zeuner, 1963; Hemmer, 1980).

The derivation of populations from one single pregnant female has only been described in *Drosophila* (Prakashi, 1972; Nei *et al*, 1975). If the history of the Porto Santo rabbit is true, this would be the first time that such successful introduction is reported for a mammalian species. Porto Santo could thus be seen as a natural population in nearly "laboratory" conditions, which can be very useful in conservational genetics issues, because it represents the ideal situation where consequences of an extreme bottleneck event can be studied in mammals. Indeed, the study of founder effects has become increasingly important in population genetics, speciation theory and conservation biology. One of the issues raised here is the likelihood that four or more allelic lineages can persist over millions of years, as was proposed for *IgVH1* locus (Su and Nei, 1999).

Population bottlenecks are particularly important because they can increase the demographic stochasticity, the rate of inbreeding and the fixation of deleterious alleles, therefore reducing adaptive potential and increasing the probability of population extinction (e.g. Lande, 1994; O`Brien *et al.*, 1994; Newman and Pilson, 1997; Saccheri *et al.*, 1998)). In natural populations, for selectively neutral loci, the number of alleles and heterozigosity values result from equilibrium between mutation and genetic drift. This equilibrium depends therefore on two parameters: mutation rate and effective population size. When a strong bottleneck occurs there is a high decrease in the number of alleles and heterozigosity values. A potential test for detecting recent bottlenecks is to compare the expected gene diversity (in fact, the variance in gene frequencies) observed in the population (Ho) with that expected (He) from the observed number of alleles for each locus under a particular mutation model. Note that this heterozygosity excess should not be confused with the excess of heterozygotes. The former compares the observed heterozygosities, whereas the latter compares the number of heterozygotes with Hardy-Weinberg expectation. Theoretically, in non-bottlenecked populations a similar number of loci with slight excess or with a slight deficiency of Ho are expected. On the other hand, in a bottlenecked population the majority of loci show an excess of Ho (Nei *et al*, 1975; Chakraborty and Nei, 1977; Chakraborty et al., 1980; Maruyama and Fuerst, 1985; Cornuet and Luikart, 1996; Luikart and Cornuet, 1998).

# Material and Methods

## Assessment of genetic diversity

In this work, the Porto Santo rabbit population was studied by means of a battery of loci that had previously been shown to be polymorphic in Iberian rabbit populations. We used: 1) 20 protein loci (allozymes and plasmatic proteins), 2) three immunological markers; 3) nine microsatellite loci and 4) mtDNA markers, a 1120 bp-long fragment of cytochrome *b* and a fragment of approximately 370 bp length of the control region.

The samples (tissues, red cells and serum) utilized in this work were obtained in November 1995 and stored at -20ºC.

The different alleles obtained in a battery of 20 protein loci were separated by starch gel electrophoresis, agarose gel electrophoresis or isoelectric focusing and detected by general staining of proteins, specific enzymatic activity staining or by immunoblotting. In the study of immunological markers the serum of each sample was tested against antisera specific for particular allotype markers: (*b* locus: *b4*, *b5*, *b6*, *b9*; *a* locus: *a1*, *a2*, *a3*; *e* locus: *e14, e15*). The phenotypes were determined by immunodiffusion in 1% agar gel containing 2% polyethylene glycol. In each test a reference was included to monitor the degree of serological similarity of the sample being tested with samples from rabbits expressing the immunizing allotype. Typing of the microsatellites was achieved with multiple PCR, fluorescently labeled primers, and analyzed using an ABI 310 (Applied Biosystems) automated sequencer. The cytochrome *b* was amplified according to the methodology described by Branco *et al.* (2000) and was then digested with eight restriction enzymes (*Alu*I, *Bsa*I, *Dde*I, *Fok*I, *Hae*III, *Mse*I, *Taq*I, and *Tsp*509I). The fragments patterns were

obtained after electrophoresis on 3% Nusieve agarose gels. A control region fragment with approximately 400 bp was amplified according to the conditions given in Branco *et al.* (2002) and then sequenced.

The loci used and the methodologies followed are indicated in the Table 4.1.

**Table 4.1**
Protein, immunological, microsatellite and mtDNA loci analysed in this work.

| Loci | Separation method | Reference |
|---|---|---|
| **Nuclear** | | |
| **Protein markers** | | |
| Acid phosphatase 3 (ACP3) | IEF | Branco and Ferrand (1998) |
| Adenosine deaminase (ADA) | SGE | Ferrand (1995) |
| Albumin (ALB) | IEF | Ferrand and Rocha (1992) |
| Anti-Thrombine III (ATIII) | IEF | Branco and Ferrand (2002) |
| Carbonic anhydrase I (CAI) | AGE | Branco and Ferrand (2003) |
| Carbonic anhydrase II (CAII) | AGE | Branco and Ferrand (2003) |
| Galactose-1-phosphate uridyl transferase (GALT) | SGE | Ferrand (1995) |
| Vitamin D binding region (GC) | IEF | " " |
| Hemoglobin alpha chain (HBA) | SGE | Campos *et al.* (2003) |
| Hemoglobin beta chain (HBB) | SGE | Campos *et al.* (2003) |
| Hemopexin (HPX) | IEF | Branco and Ferrand (2002) |
| Nucleoside Phosphorylase (NP) | SGE | Ferrand (1995) |
| Properdin factor B (BF) | IEF | Branco *et al.* (1998) |
| Peptidase A (PEPA) | SGE | Branco *et al.*. (1999) |
| Peptidase B (PEPB) | SGE | " " |
| Peptidase C (PEPC) | SGE | " " |
| Peptidase D (PEPD) | SGE | " " |
| Phosphogluconate dehydrogenase (PGD) | SGE | Ferrand (1995) |
| Superoxide Dismutase (SOD) | SGE | Ferrand (1995) |
| Transferrin (TF) | AGE/IEF | Ferrand *et al.* (1988) |
| **Immunogenetic markers** | | |
| IGGHCH2 | IDT | Kellus and Gell (1967) |
| IGVH1 | IDT | " " |
| IGCK1 | IDT | " " |
| **Microssatellites** | | |
| Sat 2 | CE | Queney *et al.* (2001) |
| Sat 3 | CE | " " |
| Sat 4 | CE | " " |
| Sat 5 | CE | " " |
| Sat 7 | CE | " " |
| Sat 8 | CE | " " |
| Sat 12 | CE | " " |
| Sat 13 | CE | " " |
| Sat 16 | CE | " " |
| **Mitochondrial (mtDNA)** | | |
| cytochrome *b* | PCR/RFLP | Branco *et al.* (2000) |
| Control region | S | Branco *et al.* (2002) |

IEF- Isoelectric Focusing
SGE- Starch Gel Electrophoresis
AGE- Agarose Gel electrophoresis
IDT- Immunodifusion Test
PCR/RFLP- Polymerase Chain Reaction/Restriction Fragment of Length Polymorphism
CE- Capilar Electrophoresis
S- Sequencing

## Statistical tests to detect bottleneck situations

To test the occurrence of a bottleneck effect we followed the methodology described by Cornuet and Luikart (1996).

*Assumptions and Principles*

In order to use simulation tests to detect bottleneck events in natural populations it is necessary to have a representative sample of a well-defined population, under no immigration, population substructure and where loci are selectively neutral. All these situations can be generally detected through Hardy-Weinberg equilibrium. Loci that are not in accordance to Hardy-Weinberg equilibrium should be excluded or used only with caution. In our study, all loci (microsatellites and the protein markers) used for testing the bottleneck effect are in Hardy-Weinberg equilibrium.

The method described by Cornuet and Luikart (1996) calculates for each population sample and for each locus the distribution of the gene diversity expected from the observed number of alleles (k), given the sample size (n) under the assumption of mutation-drift equilibrium. This distribution is obtained through simulating the coalescent process of n genes, assuming mutation/drift equilibrium under different mutation models, the Infinite Allele Model (IAM), the Two-Phase Mutation Model (TPM) and the Stepwise Mutation Model (SMM). Under the IAM, a single mutation is allocated at a time and the resulting number of alleles is computed. The process is repeated until the latter reaches the observed number of alleles. Under the SMM, a Bayesian approach is used as explained in Cornuet and Luikart (1996). Briefly, the likelihood distribution of the parameter theta (= $4N_e\mu$) given the number of alleles (k) and the sample size (n) is evaluated as the proportion of iterations (in the simulation process) producing exactly k alleles for a varying set of thetas. Only gene diversities found in iterations producing exactly k alleles are considered. The way in which the coalescent process is simulated is unconventional due to the conditioning by the observed number of alleles. The phylogeny of the n genes is simulated according to Hudson (1990).

This enables the computation of the average (He) that is compared to the observed gene diversity (Ho) (see Nei, 1987) to establish whether there is a gene diversity excess or deficit at each locus. For the protein markers we choose to use the IAM, because the allozyme allele frequency distribution seems to fit this model (Nei *et al.* 1976; Chakraborty

*et al*, 1980; Luikart and Cornuet, 1998), but for the microsatellite loci we used the Two-Phase Model (TPM) (90% of Stepwise Mutation Model (SMM) and 10% of IAM). In IAM each mutation produces a new allele that is different from all existing ones (Kimura and Crow, 1964), while in SMM mutations change the state of an allele by one step forward or backward with equal probability (Ohta and Kimura, 1973). The TPM is intermediate to the SMM and IAM. Most microsatellite data sets fit the TPM better than the SMM or IAM (Di Rienzo *et al.*, 1994). The TPM recommendation for microsatellite consists of mostly one-step mutations, but a small percentage (5-10%) of multi-step changes (Luikart and Cornuet, 1998).

We used two statistical methods: the Standardized Differences Test (SDT) and the Wilcoxon Test (WT). The SDT is more powerful than the WT, but is parametric and requires at least 20 polimorphic loci in order to become valid. The Wilcoxon test is nonparametric, provides relatively high power and it can be used with as few as four polymorphic loci and any number of individuals (15-40 individuals and 10-15 polymorphic loci is recommended to achieve high power). Nevertheless, the Wilcoxon Test provides relatively high power and it can be used with as few as four or five polymorphic loci and any number of individuals. In these statistical tests, probability values lower than 0.05 reject the null hypothesis (mutation/drift equilibrium) in favor of the hypothesis of overall heterozygosity excess and a recent bottleneck effect. Additionally, we used a qualitative descriptor of the allele frequency distribution ("mode-shift" indicator), which discriminates bottlenecked populations from stable populations. However, this qualitative method is not a proper statistical test and cannot be used with confidence with samples of fewer than 30 individuals.

# Results and Discussion

## Assessment of genetic diversity

The gene frequencies obtained for each marker studied for the Porto Santo rabbit population are presented in Table 4.2.

**Table 4.2**
Allele frequencies at nuclear and mitochondrial loci for Porto Santo rabbit population. n is the sample size.

| Markers | n | alleles | gene frequency |
|---|---|---|---|
| **Nuclear** | | | |
| **Protein markers** | | | |
| ACP3 | 59 | 1 | 1.00 |
| ADA | 64 | 1 | 0.64 |
| | | 2 | 0.36 |
| ALB | 22 | 1 | 1.00 |
| ATIII | 20 | A | 1.00 |
| CAI | 22 | 1 | 1.00 |
| CAII | 22 | 2 | 1.00 |
| GALT | 15 | 1 | 0.57 |
| | | 2 | 0.43 |
| GC | 20 | 1 | 1.00 |
| HBA | 59 | 2 | 0.48 |
| | | 3 | 0.52 |
| HBB | 60 | 1 | 0.01 |
| | | 2 | 0.99 |
| HPX | 20 | F1 | 0.50 |
| | | F3 | 0.50 |
| NP | 66 | 1 | 0.50 |
| | | 2 | 0.25 |
| | | 3 | 0.25 |
| BF | 42 | A | 0.54 |
| | | B | 0.06 |
| | | D | 0.40 |
| PEPA | 61 | 1 | 0.21 |
| | | 2 | 0.78 |
| | | 3 | 0.01 |
| PEPB | 68 | 1 | 0.91 |
| | | 2 | 0.09 |
| PEPC | 68 | 1 | 0.18 |
| | | 2 | 0.28 |
| | | 4 | 0.27 |
| | | 8 | 0.28 |
| PEPD | 55 | 1 | 0.74 |
| | | 2 | 0.26 |
| PGD | 64 | A | 0.45 |
| | | C | 0.55 |
| SOD | 68 | 1 | 1.00 |
| TF | 20 | 1 | 1.00 |

*Continued on next page*

*Continued from previous page*

| Markers | n | alleles | gene frequency |
|---|---|---|---|
| **Immunologenetic markers** | | | |
| IGGCH2 | 20 | e15 | 1.00 |
| IGVH1 | 20 | a2 | 0.13 |
| | | a3 | 0.70 |
| | | a1v | 0.17 |
| IGCK1 | 20 | b4wa | 0.30 |
| | | b5wf | 0.25 |
| | | b5wb | 0.45 |
| **Microsatellites** | | | |
| Sat2 | 32 | 237 | 0.23 |
| | | 239 | 0.66 |
| | | 241 | 0.11 |
| Sat3 | 30 | 146 | 0.21 |
| | | 154 | 0.79 |
| Sat4 | 32 | 227 | 0.14 |
| | | 231 | 0.21 |
| | | 235 | 0.65 |
| Sat5 | 32 | 206 | 0.14 |
| | | 210 | 0.38 |
| | | 214 | 0.20 |
| | | 216 | 0.03 |
| | | 222 | 0.25 |
| Sat7 | 33 | 185 | 0.02 |
| | | 187 | 0.54 |
| | | 191 | 0.21 |
| | | 193 | 0.23 |
| Sat8 | 28 | 152 | 0.23 |
| | | 154 | 0.20 |
| | | 166 | 0.22 |
| | | 168 | 0.35 |
| Sat12 | 33 | 126 | 0.11 |
| | | 130 | 0.21 |
| | | 134 | 0.59 |
| | | 138 | 0.09 |
| Sat13 | 31 | 112 | 0.59 |
| | | 114 | 0.05 |
| | | 126 | 0.07 |
| | | 128 | 0.29 |
| Sat16 | 34 | 109 | 0.91 |
| | | 115 | 0.09 |
| **Mitochondrial** | | | |
| Cytochrome b | 27 | Arb4 | 1.00 |
| Control region | 7 | Pst | 1.00 |

# Phylogeographic origin of Porto Santo rabbit population

## Nuclear markers

### *Protein markers*

The NJ tree (Saitou and Nei, 1987) (see Figure 4.1) was obtained using the genetic distance of Nei (Nei, 1978) calculated from the gene frequencies obtained for the rabbit population of Porto Santo as well as for a set of wild rabbit populations of different origins (Ferrand, 1995; Branco, 2000) in 20 protein markers. All analyses were conducted by using the computer program PHYLIP version 3.5 (Felsenstein, 1993). The tree clearly separates two groups (BP=100): one that is composed by Southwest Iberian populations, where Porto Santo is included, clustering with the Portuguese population of Idanha, and other that includes the populations from the Northeast. Previous studies using only sixteen polymorphic markers clustered the French (FR) and the domestic rabbit populations with the Northeast populations (Ferrand, 1995).

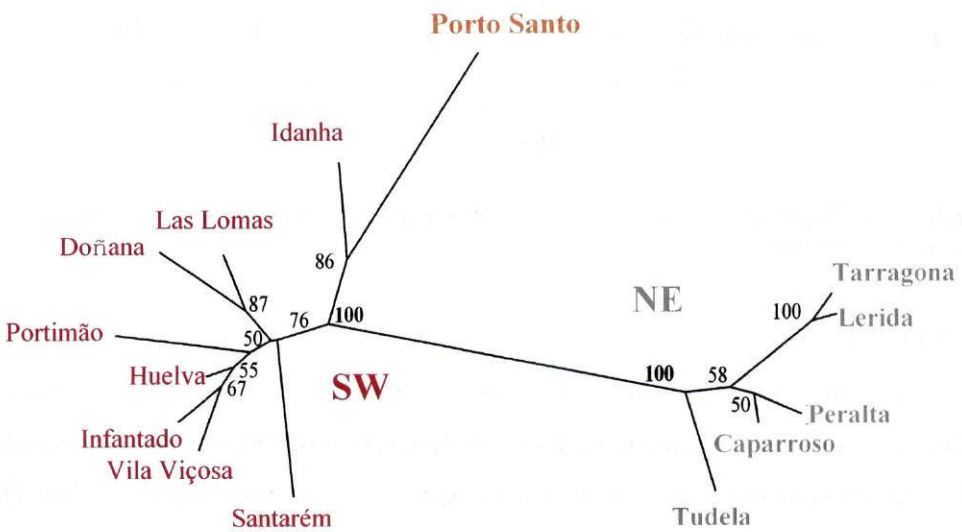



**Figure 4.1.** NJ tree obtained with Nei's genetic distances calculated using the gene frequencies of 20 polymorphic protein loci studied in Porto Santo and in Iberian Peninsula rabbit populations. The bootstrap values are indicated. SW- Southwest and NE- Northeast.

*Immunogenetic markers*

The gene frequencies obtained for *a*, *b* and *e* loci in Porto Santo were compared to the obtained in Southwest and Northeast of Iberian Peninsula and in French rabbit populations (van der Loo *et al.*, 1999). The alleles observed in Porto Santo for *b* locus are typical of Southwestern populations; for example the allele *b5wb* is the most frequent both in Southwestern populations and Porto Santo. On the other hand, allele *b4* is far out the most frequent in the Northeastern Iberian, in the French and in the domestic rabbit populations but is absent from Porto Santo (see Figure 4.2). The *a* and *e* loci are not informative in determining the origin of Porto Santo, since the *a1* and *a-blank*, highly represented in the wild rabbit populations, are both absent from Porto Santo and the *e* locus shows fixation for allele *e15*.

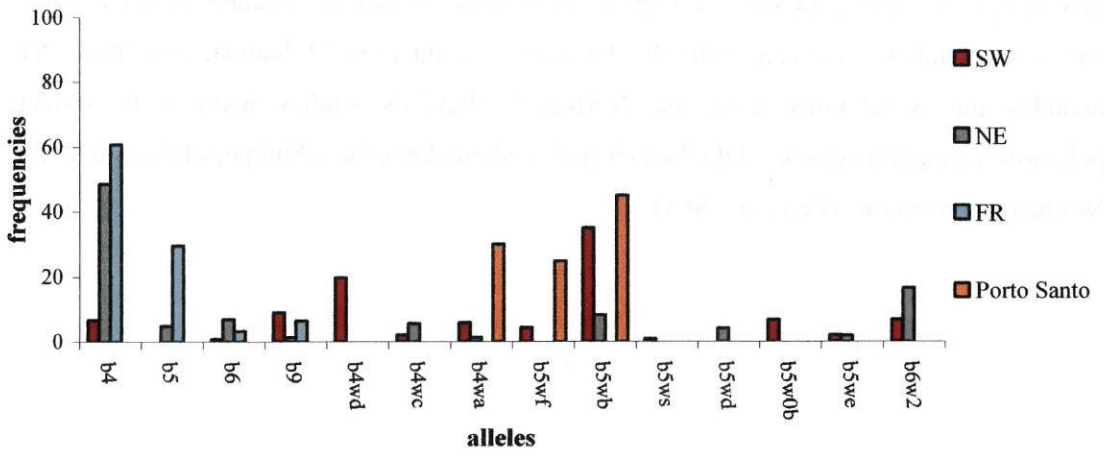



**Figure 4.2.** Gene frequencies obtained at *b* locus in rabbit populations from SW and NE of Iberian Peninsula, France and in Porto Santo.

*Microsatellites*

The gene frequencies obtained in Porto Santo as well as in a set of wild rabbit populations of different origins in 9 microsatellite markers (Queney *et al.*, 2001) were used to construct a NJ tree (see Figure 4.3) using the genetic distance of Nei (Nei, 1978). The tree clearly separates three groups: one that is composed by Southwest Iberian populations, where Porto Santo is included (clustering again with the Portuguese population of Idanha), other that includes the Iberian populations from the Northeast and a third group composed exclusively by French populations. Previous studies with this set of microsatellites showed that the French (FR) populations are closely related with the domestic breeds (Queney *et al.*, 2002).

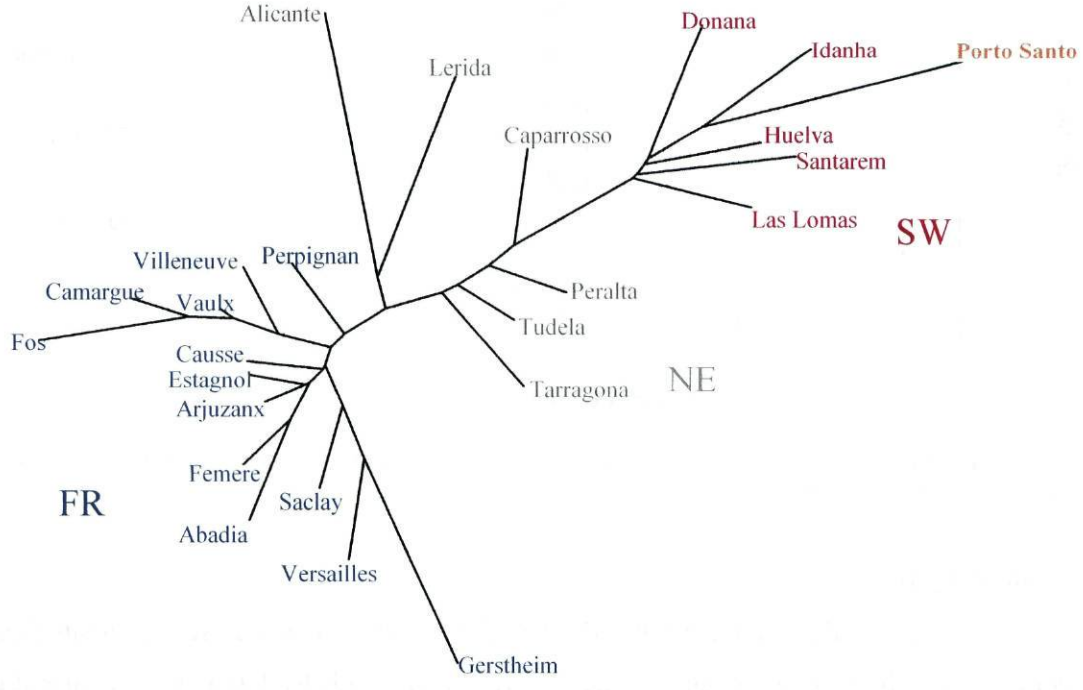



**Figure 4.3.** NJ tree obtained with Nei's genetic distances of calculated using the gene frequencies of 9 microsatellites studied in Porto Santo, the Iberian Peninsula and in French rabbit populations. SW - Southwest of Iberian Peninsula; NE - Northeast of Iberian Peninsula; FR - France.

## Mitochondrial DNA

### *Cytochrome b*

The restriction fragment profiles obtained in the Porto Santo population correspond to the mtDNA haplotype ARb4 (Branco, 2000; Branco *et al.*, 2000), which was only found in Portuguese populations, where it is the most common. The frequencies of each haplotype of mtDNA lineage *A* in Portuguese, Spanish and in Porto Santo populations are shown in Figure 4.4. In all wild rabbit populations from outside of Iberian Peninsula and in all domestic breeds studied so far only the mtDNA lineage *B* was detected. Consequently, this result confirms that the Porto Santo rabbit population does show closest relationships with the Portuguese populations and is clearly not derived from a domestic population.

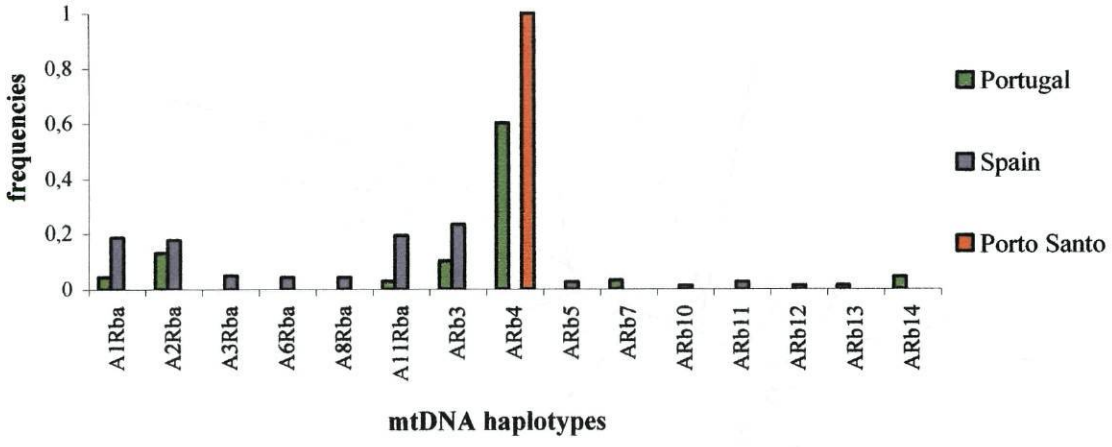



**Figure 4.4.** mtDNA haplotypes of lineage *A* obtained in rabbit populations from Portugal and Spain and in the island population of Porto Santo.

### *Control region*

We sequenced a fragment with 373 bp of the control region in seven rabbits from Porto Santo. All of these sequences were identical. Part of this haplotype was compared to 53 different Iberian haplotypes defined by 56 variable sites in a fragment of 179-181 bp (Branco *et al.*, 2002). The NJ tree representing the relationships between these haplotypes is presented in figure 4.5. Although representing a haplotype that was not detected in the previously studied populations, the sequence found in Porto Santo, called here Pst, clusters within a group that is clearly of Portuguese origin.

Taken together, these results clearly show that founders of the Porto Santo rabbit population were drawn from Southwestern Iberian populations belonging to the subspecies *Oryctolagus cuniculus algirus*. The clustering with the Portuguese population of Idanha both for protein and microsatellite trees, the fixation of a mtDNA *A* type only present in the Portuguese populations and the fact that the control region haplotype found in Porto Santo is more related to the described in the Portuguese populations strongly suggest a Portuguese origin.

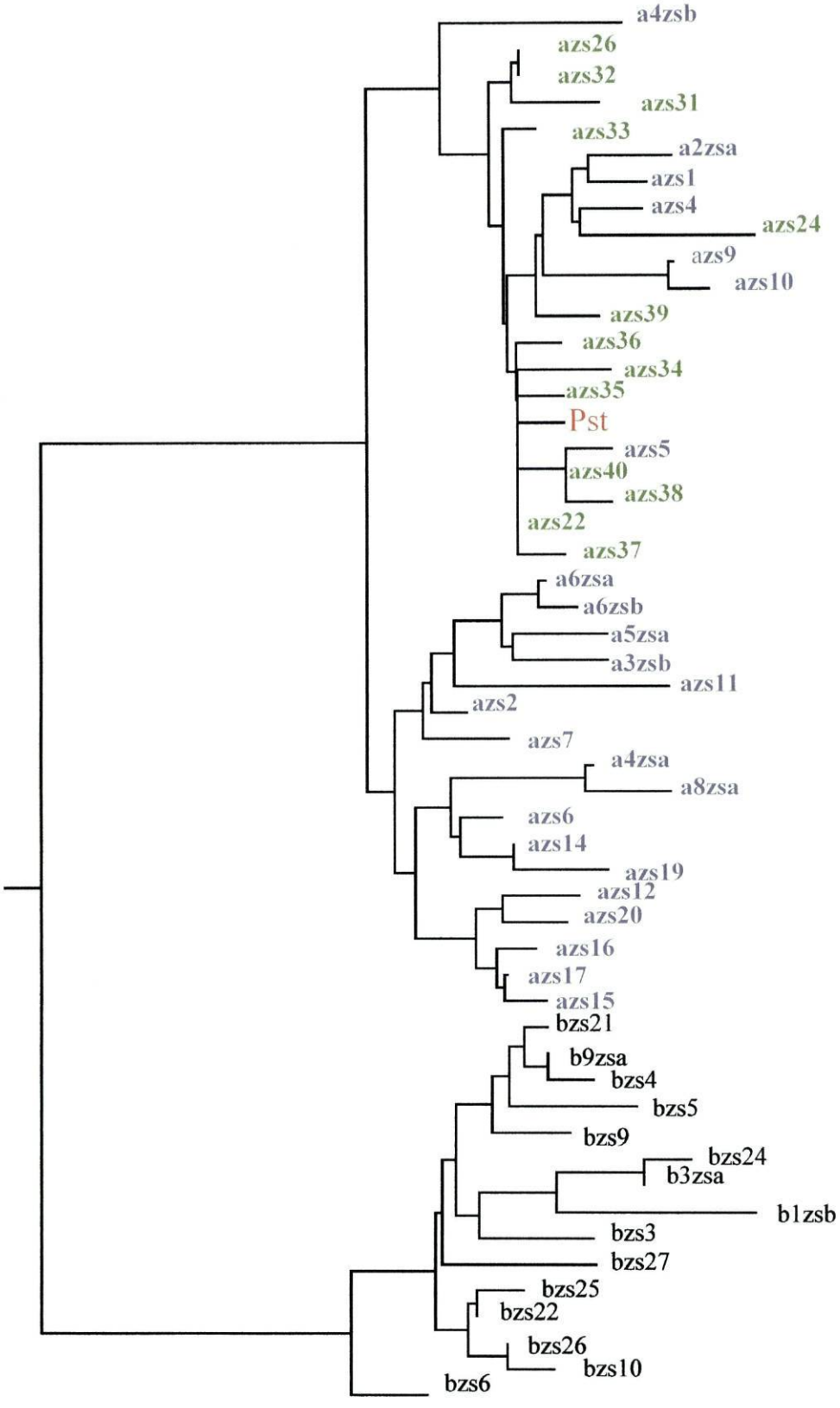



**Figure 4.5.** NJ tree for control region haplotypes obtained in wild rabbit populations. ■ haplotypes A obtained in Western Spanish populations; ■ haplotypes A obtained in Portuguese populations; ■ haplotypes B obtained in Eastern Spanish populations; ■ haplotype obtained in Porto Santo.

# Article 5

P.J. Esteves, S. Weiss, J. Rocha, G. Queney, M. Branco, W. van der Loo and N. Ferrand (Submitted).

The "Rabbit-Eve" of Porto Santo: A 500 year-old mammalian population derived from a single female

# The "Rabbit-Eve" of Porto Santo: A 500 year-old mammalian population derived from a single female.

**Pedro J. Esteves*§, Steven Weiss†, Jorge Rocha‡§, Guillaume Queney‖, Madalena Branco*, Wessel van der Loo*ϒ, Nuno Ferrand*§**

*CIBIO - Centro de Investigação em Biodiversidade e Recursos Genéticos, Campus Agrário de Vairão, 4485-661 Vairão

§Departamento de Zoologia e Antropologia, Faculdade de Ciências da Universidade do Porto, 4099-002 Porto, Portugal. e-mail: nferrand@mail.icav.up.pt

†Department of Zoology, University of Graz, A-8010 Graz, Austria

‡IPATIMUP, Rua Dr. Roberto Frias s/n, 4200-465 Porto, Portugal

‖Antagène, Immeuble Le Meltem, 2 allée des Séquoias, 69760 Limonest, France

ϒInstitute of Molecular Biology and Biotechnology, Vrije Universiteit Brussel, St Genesius-Rode, Belgium

How many individuals does it take for establishing a viable mammal population? An early answer to this key question in evolutionary sciences, was proposed by Noah, who allowed for «one of each gender» on his ark (Genesis 6-9). Closer to us, Charles Darwin[1] reported how in 1419 AD Bartolomeu Perestrelo, by introducing a single pregnant rabbit doe into the island of Porto Santo (Madeira Archipelago, Portugal), founded a population, which within a few years was considered a plague[2,3]. Here we report genetic data that are in strong support of the Single Female Ancestor (SFA) hypothesis of the contemporary Porto Santo rabbits.

Of 38 cyt*b* haplotypes described in Iberian populations[4], one single haplotype occurred in Porto Santo, which was furthermore unique to Portugal. In addition, the sequencing of 350 base-pairs in the mtDNA hypervariable region from seven Porto Santo rabbits confirmed the occurrence of a single haplotype, in sharp contrast with the extreme degree of polymorphism resolved in continental populations[5]. The distinct allelic variation at nuclear loci confirmed Southwestern Iberia as the origin of Porto Santo rabbits[6]. We have estimated the allele frequencies at 30 loci that were polymorphic in this area (22 protein loci and eight microsatellites). In Porto Santo, all loci revealed four or fewer alleles, except one microsatellite locus, which had a fifth allele. Estimating 500 generations since the introduction, and a mean mutation rate of $5 \times 10^{-4}$ per locus/per generation, the emergence of a new microsatellite allele at Porto Santo is expected and therefore concordant with four sets of founder chromosomes.

Of more relevance is the fact that our knowledge of the allele frequencies in the source population allows estimates of the expected number of loci that would retain a given number of distinct alleles, if either four (SFA), or more than four, chromosomes were drawn randomly from it. It appears that the distribution observed in Porto Santo is a perfect match of expectations under SFA. For microsatellites, where most loci display more than ten alleles in the founder region[7], the alternative hypothesis can be rejected ($P<0.01$). Remarkably, two of these microsatellites are closely linked in the mammalian casein gene cluster[8,9] and occur in complete linkage disequilibrium defining four haplotypes, while a sample of four southwestern rabbit populations exhibited a minimum of 80 haplotypes. For protein loci, however, where gene diversity is more limited, observations are in accordance with four to eight founding chromosomes.

We have therefore used an additional criterion. Because the founding event was immediately followed by exponential expansion[2,3], the initial allele frequencies might, to some extent, have been conserved. Under the SFA hypothesis, these frequencies should be multiples of 25% or quartiles. The observed distribution of allele frequency classes across loci most convincingly supports a single female founder (Figure 1).

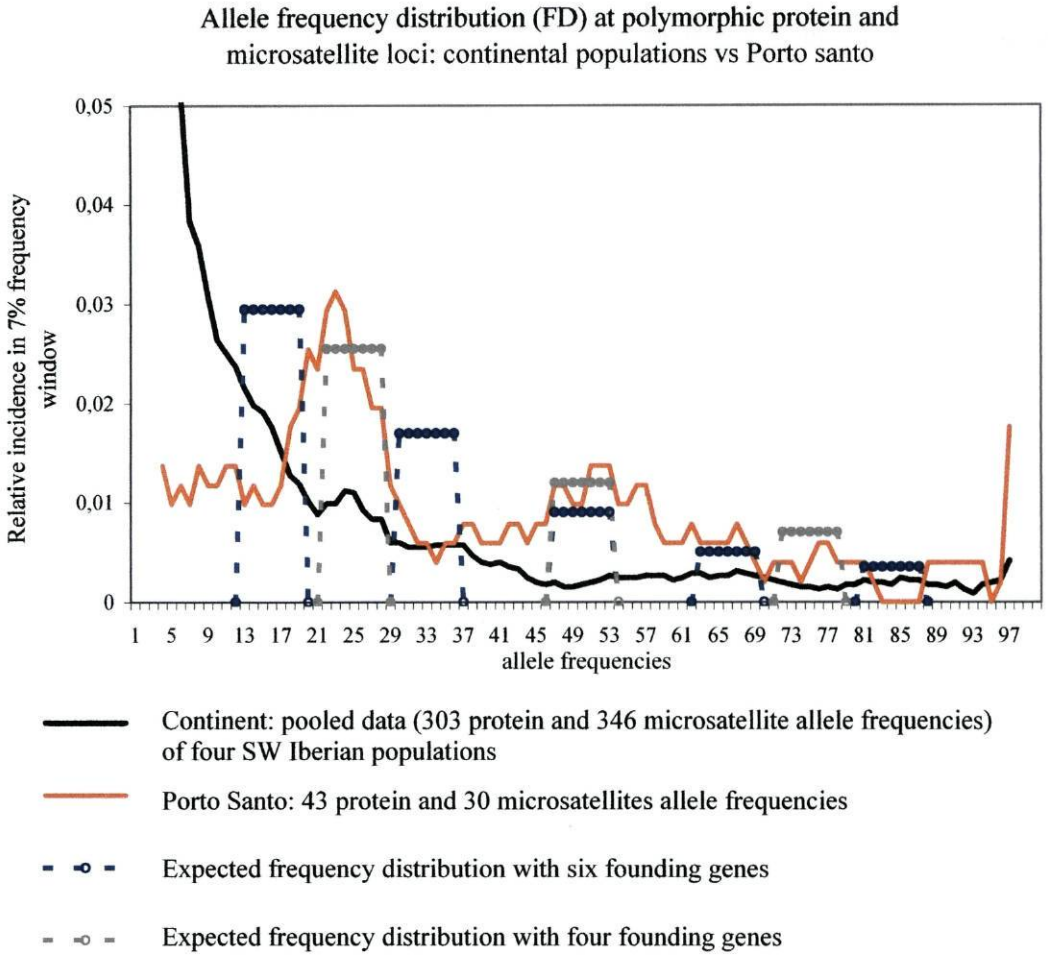



**Figure 1.** Distribution of allelic frequency classes for 8 microsatellite and 22 protein loci. Frequency classes are centered on the expected quartiles among four founding chromosomes, with the distribution tail split between rare (<12.5%) and frequent (>12.5%) alleles. Normal populations display a L-shaped frequency distribution whereas extreme bottlenecks are marked by intermediate frequencies. Several markers in our data reveal allele frequencies matching perfectly or nearly so expected quartiles (e.g. NP, n=66, 3 alleles at 0.50, 0.25, 0.25, HBA, n=59, 2 alleles at 0.52, 0.48, PGD, n=64, 2 alleles at 0.55, 0.45, PEPC, n=68, 4 alleles at 0.28, 0.28, 0.18, 0.27, PEPD, n=55, 2 alleles at 0.45, 0.55, GALT, n=15, 2 alleles, 0.57, 0.43, HPX, n=16, 2 alleles, 0.50, 0.50, sat3, n=31, 2 alleles 0.79, 0.21, sat8, n=30, 4 alleles 0.23, 0.22, 0.20, 0.35).

We provide therefore strong empirical support for the viewpoint that the rabbit population of Porto Santo was indeed founded by a single pregnant female, to our knowledge representing the only such wild mammalian population. As subsequent bottlenecks or introductions apparently did not occur, it provides a superb natural laboratory for testing theoretical expectations of allelic drift, evaluating mammalian mutation rates or studying natural selection. This report also suggests that gene diversity conveyed by one female can sustain long-term population viability.

# Citations

1. Darwin, C. *The Variation of Animals and Plants under Domestication* (New York: Orange Judd, 1868).

2. Barros, J. *Asia. Primeira Década, Livro primeiro* (Lisboa, 1778).

3. Crosby, A.W. *Ecological imperialism. The biological expansion of Europe, 900-1900* (Cambridge University Press, New York, 1986).

4. Branco, M., Ferrand, N. & Monnerot, M. *Heredity* **85,** 307-317 (2000).

5. Branco, M., Ferrand, N., Monnerot, M. & Templeton, A. *Evolution* **56,** 792-803. (2002).

6. Ferrand, N. & Branco M. in *Phylogeography of European Refugia* (eds Weiss, S. & Ferrand, N.) (Kluwer Academic Publishers, Amsterdam, *in press*).

7. Queney, G., Ferrand, N., Weiss, S., Mougel, F & Monnerot, M. *Mol. Biol. Evol.* **18,** 2169-2178 (2001).

8. Rijnkels, M., Wheeler, D.A., de Boer, H.A. & Pieper, F.R. *Mammalian Genome* **8,** 9-15 (1997).

9. Mougel, F., Mounolou, J.-C. & Monnerot M. *Animal Genetics* **28,** 68-71 (1997).

# Evaluating bottlenecks in wild rabbit populations

*Genetic diversity*

Different measures of genetic diversity (average number of alleles per locus ($n_a$), percentage of polymorphic loci (P) and the average of expected heterozigosity (He) were obtained using BIOSYS program version 1.7 (Swofford and Selander, 1989). The average values of the different measures of genetic diversity obtained in Iberian populations for protein loci (Ferrand, 1995; Branco, 2000), microsatellites (Queney, 2000; Queney *et al.*, 2001) and immunoglobulin markers (van der Loo *et al.*, 1999) and the values of genetic diversity obtained in Porto Santo studied in this work are presented in Table 4.3.

**Table 4.3**

Genetic diversity obtained in the SW, NE, FR, domestic and Porto Santo rabbit populations, for protein, microsatellites and two Immunoglobulin markers (*a* and *b* locus). $n_a$ is the average number of alleles per locus; He is the average expected heterozygosity; P is percentage of polymorphic loci.

| | **SW** | | | **NE** | | | **FR** | | | **Porto Santo** | | | **Domestic** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Markers** | $n_a$ | He | P | $n_a$ | He | P | $n_a$ | He | P | $n_a$ | He | P | $n_a$ | He | P |
| Microsatellite | 9.3 | 0.84 | 100 | 7.8 | 0.78 | 100 | 5.2 | 0.60 | 100 | 3.44 | 0.53 | 100 | 3.2 | 0.49 | 100 |
| Protein | 3.0 | 0.35 | 77.8 | 2.1 | 0.23 | 65 | 1.7 | 0.20 | 57 | 1.9 | 0.26 | 63 | 1.6 | 0.14 | 43 |
| Immunogenetic | 7.2 | 0.60 | 100 | 7 | 0.67 | 100 | 4 | 0.59 | 100 | 3 | 0.50 | 100 | 2.5 | 0.40 | 100 |

As expected, gene diversity was lower in Porto Santo than in most Iberian mainland populations. However, the diversity of protein markers in Porto Santo population is higher than the observed in wild rabbit populations from France or in the domestic breeds. In microsatellite markers the levels of diversity observed in Porto Santo are slightly lower than those observed in French populations, but higher that those obtained in the domestic breeds. Also in loci under selection, such as immunoglobulin markers, the level of diversity observed in Porto Santo is similar or higher than the obtained in wild populations outside the Iberian Peninsula and in the domestic populations (van der Loo *et al.*, 1991, 1999). These results show that the genetic diversity conveyed in this population is generally higher than most of the natural French populations and may eventually explain long-term population viability.

*Statistical tests to detect bottleneck situations*

Table 4.4 shows the probability values and the model-shift descriptor outcomes obtained in Portuguese island populations (Porto Santo, São Jorge and Flores), in five Southwest Iberian Peninsula populations (Las Lomas, Huelva, Doñana, Idanha and

Santarém), in three Northeast Iberian populations (Lerida, Alicante and Tudela) and in three French populations (Camargue, Perpignan and Versailles). Independent of the type of data, all statistical tests showed a situation of mutation-drift equilibrium for all continental populations, i.e. approximately 50% of the loci sampled are a slight excess (Ho>He) or deficiency (Ho<He) heterozygosity, and a bottleneck signature for island populations ($p<0.05$), the majority of loci have a heterozygosity excess (Ho>He) (see Figures 4.4 and 4.5). The "mode-shift" descriptor showed for all continental populations a normal L-shaped distribution while in all island populations a mode-shift distribution was observed.

**Table 4.4**

Probability values obtained in the bottleneck test, and the result obtained in the mode-shift descriptor for five Southwest, three Northeast and three island populations. IAM- Infinite Allele Model; TPM- Two-phase Mutation Model; WT- Wilcoxon test; SDT- Standardized difference test.

| | **Microsatellite** | | | **Proteins** | | |
|---|---|---|---|---|---|---|
| | **TPM (90%)** | **Mode-shift** | **Reference** | **IAM** | **Mode-shift** | **Reference** |
| **Continental populations** | | | | | | |
| **Southwest** | | | | | | |
| Donana | 0.08 (WT) | Normal | 1 | 0.07 (SDT) | Normal | 2, 3 |
| Huelva | 0.75 (WT) | Normal | 1 | 0.36 (SDT) | Normal | 2, 3 |
| Santarém | 0.08 (WT) | Normal | 1 | 0.06 (SDT) | Normal | 2, 3 |
| Las Lomas | 0.15 (WT) | Normal | 1 | 0.37 (SDT) | Normal | 2, 3 |
| Idanha | 0.18 (WT) | Normal | 1 | 0.74 (WT) | Normal | 2, 3 |
| **Northeast** | | | | | | |
| Alicante | 0.32 (WT) | Normal | 1 | 0.41 (WT) | Normal | 2, 3 |
| Lerida | 0.75 (WT) | Normal | 1 | 0.34 (WT) | Normal | 2, 3 |
| Tudela | 0.41 (WT) | Normal | 1 | 0.40 (WT) | Normal | 2, 3 |
| **France** | | | | | | |
| Camargue | 0.25 (WT) | Normal | 1 | 0.10 (WT) | Normal | Normal |
| Versailles | 0.15 (WT) | Normal | 1 | 0.07 (WT) | Normal | Normal |
| Perpignan | 0.59 (WT) | Normal | 1 | n. a. | n. a. | n. a. |
| **Island populations** | | | | | | |
| Porto Santo | 0.006*(WT) | Shifted | This work | 0.0004*(WT) | Shifted | This work |
| São Jorge | n. a. | n. a. | n. a. | 0.0005*(WT) | Shifted | 2 |
| Flores | n. a. | n. a. | n. a. | 0.0009*(WT) | Shifted | 2 |

1) Queney *et al.* (2001);
2) Ferrand (1995);
3) Branco (2000).
n. a. - There is no data available

In the SW and NE rabbit populations the most frequent allelic classes are the rare alleles (0.00-0.10 and 0.10-0.20). In contrast, for Porto Santo the most frequent are classes with intermediate frequencies (0.20-0.30, 0.40-0.50 and 0.50-0.60, see Figures 4.4 and 4.5. These results revealed, as expected, that all insular rabbit populations showed the signature of a recent bottleneck and a "shifted mode" allelic distribution. In contrast, the continental rabbit populations are clearly under mutation/drift equilibrium and have a normal L-shaped allelic distribution. This indicates that natural rabbit populations have maintained the same level of genetic diversity even after multiple epizootic episodes of myxomatosis and haemorragic viral diseases.

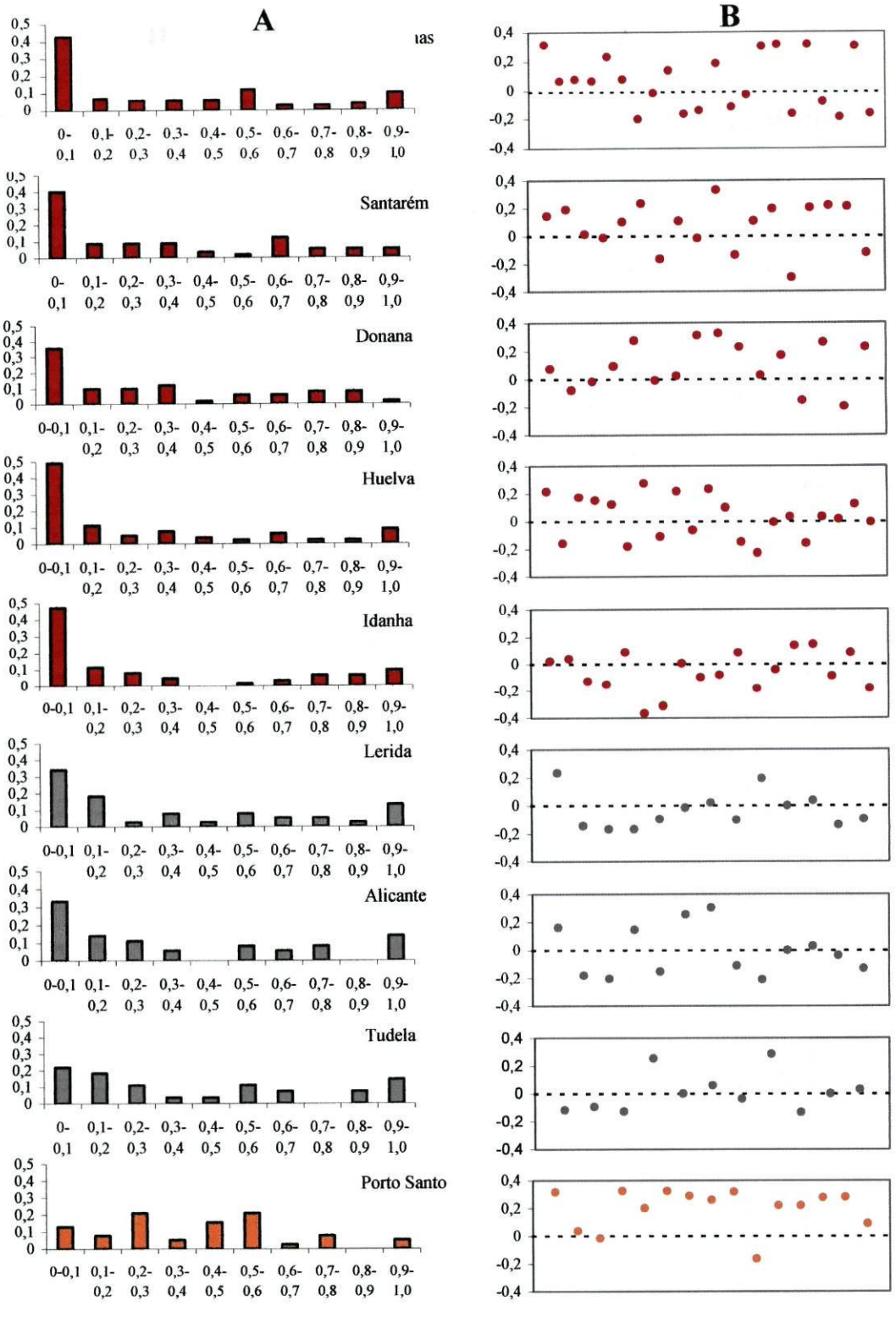



**Figure 4.4. A-** Distribution of allelic frequencies at selected protein loci in wild rabbit populations from Southwestern and Northeastern Iberia and the island of Porto Santo. **B-** Magnitude of heterozygosity excess observed at each polymorphic protein locus. Dots above the dashed line represent loci with heterozygosity excess; dots below are loci with heterozygosity deficiency.

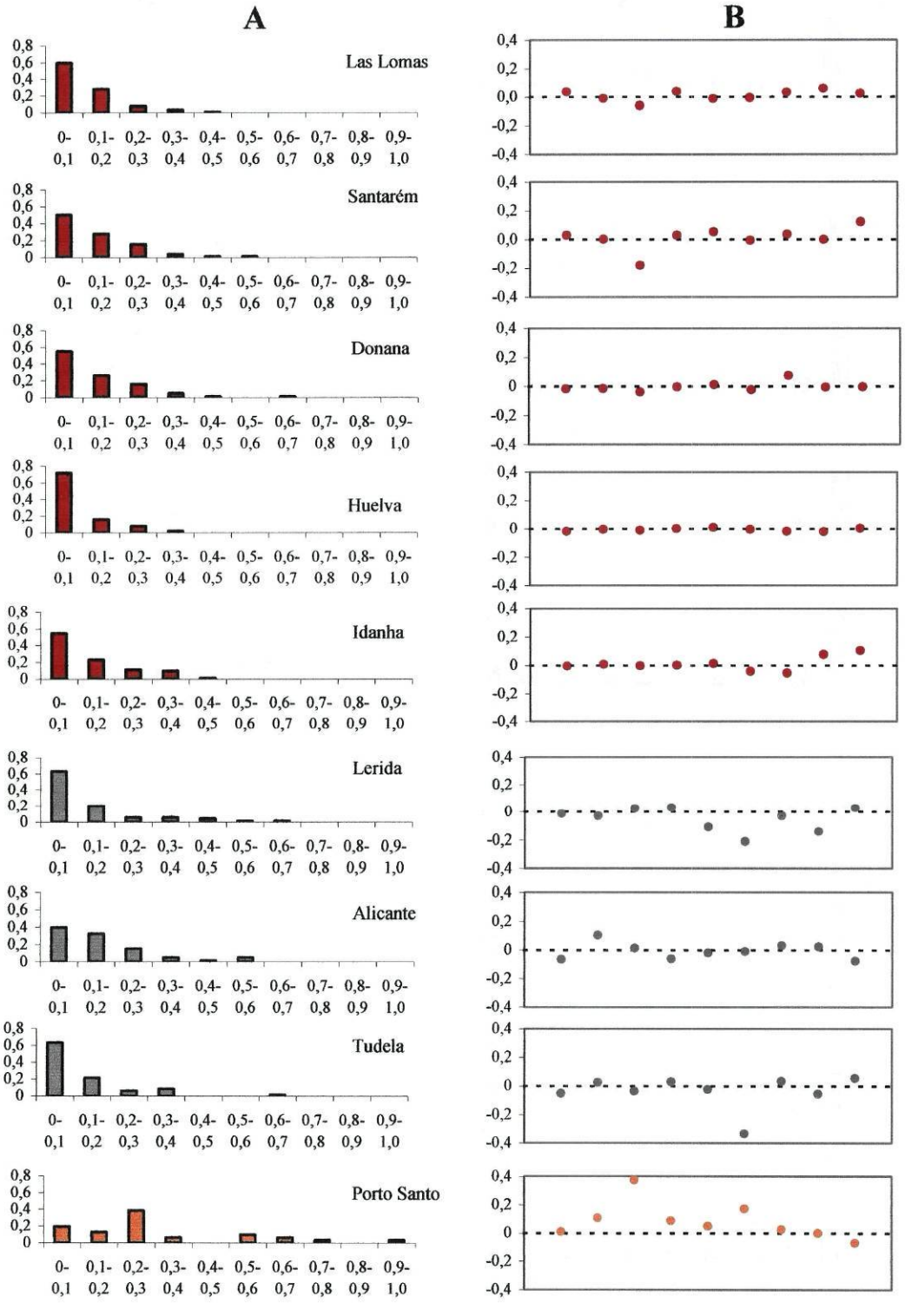



**Figure 4.5. A-** Distribution of allelic frequencies at nine microsatellite loci in wild rabbit populations from Southwestern and Northeastern Iberia and the island of Porto Santo. **B-** Magnitude of heterozygosity excess observed at each polymorphic microsatellite loci. Dots above the dashed line represent loci with heterozygosity excess; dots below are loci with heterozygosity deficiency.

## The likelihood that one single pregnant female originated the Porto Santo rabbit population

In article 5 we provided strong empirical support for the viewpoint that the rabbit population of Porto Santo was indeed founded by a single pregnant female. The methodology used to estimate the expected number of loci that would retain a given number of distinct alleles, if either four, or more than four, chromosomes were drawn randomly from the original species range is given in the appendix 1. The likely allele frequencies in the population from which the founders of the Porto Santo population originated were inferred from the allele frequencies in four rabbit populations (Las Lomas, Doñana, Santarém and Huelva; SW of Iberian Peninsula). Data for protein markers were from Ferrand (1995) and Branco (2000), immunoglobulin markers from van der Loo *et al.* (1999) and microsatellites from Queney *et al.* (2001). The expected probability values under the hypotheses of four or six chromosomes founding the population for microsatellite and protein data are presented in appendix 2 (tables A1 and A2, respectively). Additionally, for microsatellites three different scenarios were considered: 1) no mutations; 2) assuming mutation at Sat5 and 3) mutations at Sat 5 and Sat 7 (appendix 2 see figure A1).

## Implications for the trans-generic origin of the complex Ig allotypes

The explanation of the large inter-allelic distances at the $V_H1$ and the $C_K1$ loci by the trans-generic hypothesis has specific and profound implications with regard to sizes of founder populations, which must always be large enough to contain each of the different alleles. These numbers are at least four, for the $V_H1$ locus, and larger than six, for the $C_K1$ locus, suggesting average founder populations sizes of some 10 to 20 individuals. The demonstration that populations can be founded by just two individuals (four chromosomes) is not supportive for the trans-generic origin of the complex Ig allotypes, and it suggests, on the contrary, that lineages have a high chance of being lost during speciation steps.

# Citations

Azurara GE (1841). *Chronica do descobrimento e conquista da Guiné*. J. P. Aillaud, Paris.

Barros, J (1778). *Asia. Primeira Década, Livro primeiro*, Lisboa.

Branco M, Lopes G and Ferrand N (1998). Genetic polymorphism of properdin factor B (BF) in domestic rabbit. *Anim Genet* **29**(2): 135-7.

Branco M and Ferrand N (1998). Genetic polymorphism of rabbit (*Oryctolagus cuniculus*) tissue acid phosphatases (ACP2 and ACP3). *Comp Biochem Physiol B Biochem Mol Biol* **120**(2): 405-9.

Branco M, Machado JC and Ferrand N (1999). Extensive genetic polymorphism of peptidases A, B, C, and D, in wild rabbit (*Oryctolagus cuniculus*) populations from the Iberian Peninsula. *Biochem Genet* **37**(7-8): 237-49.

Branco M (2000). *Estrutura genética das populações de coelho europeu (Oryctolagus cuniculus) na Península ibérica. Isolamento, diferenciação de duas unidades evolutivas, expansão geográfica e contacto secundário*. Dissertação de Doutoramento, Universidade do Porto.

Branco M, Ferrand N and Monnerot M (2000). Phylogeography of the European rabbit (*Oryctolagus cuniculus*) in the Iberian Peninsula inferred from RFLP analysis of the cytochrome *b* gene. *Heredity* **4**: 307-17.

Branco M, Monnerot M, Ferrand N and Templeton AR (2002). Postglacial dispersal of the European rabbit (*Oryctolagus cuniculus*) on the Iberian Peninsula reconstructed from nested clade and mismatch analyses of mitochondrial DNA genetic variation. *Evolution* **56**(4): 792-803.

Branco M and Ferrand N (2002). Genetic polymorphism of antithrombin III, haptoglobin, and haemopexin in wild and domestic European rabbits. *Biochem Genet* **40**(11-12): 387-93.

Branco M and Ferrand N (2003). Biochemical and population genetics of rabbit (*Oryctolagus cuniculus*) carbonic anhydrase I and II in the Iberian Peninsula and France. *Biochemical Genetics* in press.

Cadamosto L (1812). Collecção de noticias para a historia e geografia das Nações Ultramarinas que vivem nos dominios Portugueses ou lhes são vizinhas. In França C.. Contribution à L'Étude du Lapin de Porto Santo. *Bulletin de la Société Portugaise de Sciences Naturelle*, Tome VI, Fasc. 2.

Campos R, Branco M and Ferrand N (2003) in *Phylogeography of European Refugia* (eds Weiss, S. & Ferrand, N.) (Kluwer Academic Publishers, Amsterdam, *in press*).

Chakraborty R and Nei M (1977). Bottleneck effects on average heterozygosity and genetic distance with the stepwise mutational model. *Evolution* **31**: 347-56.

Chakraborty R, Fuerst PA and Nei M (1980). Statistical studies on protein polymorphism in natural populations. III. Distributions of allele frequencies and the number of alleles per locus. *Genetics* **94**: 1039-063.

Cordeiro A (1717). Historia Insulana. In França C.. Contribution à L'Étude du Lapin de Porto Santo. *Bulletin de la Société Portugaise de Sciences Naturelle*, Tome VI, Fasc. 2.

Cornuet JM and Luikart G (1996). Description and evaluation of two tests for detecting recent bottlenecks. *Genetics* **144**: 2001-014.

Darwin, C. (1868) *The Variation of Animals and Plants under Domestication* (New York: Orange Judd)

Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M and Freimer NB (1994). Mutational processes of simple-sequence repeat loci in human populations. *Proc Natl Acad Sci* U S A **91**(8): 3166-70.

Felsenstein J (1993). PHYLIP (Phylogeny Inference Package) version 3.5. Distributed by the author. Seattle, University of Washington.

Ferrand N, Carvalho G and Amorim A (1988). Transferrin (Tf) polymorphism in wild rabbit, *Oryctolagus cuniculus*. *Anim Genet* **19**(3): 295-300.

Ferrand N and Rocha J (1992). Demonstration of serum albumin (ALB) polymorphism in wild rabbits, *Oryctolagus cuniculus*, by means of isoelectric focusing. *Anim Genet* **23**(3): 275-78.

Ferrand N (1995). *Variação genética de proteínas em populações de coelho (Oryctolagus cuniculus). Análise da diferenciação subespecífica, subestruturação, expansão geográfica e domesticação.* Dissertação de Doutoramento, Universidade do Porto.

Flux JEC and Fullagar PJ (1992). World distribution of the Rabbit (*Oryctolagus cuniculus*) on islands. *Mammal Review* **22**: 151-202.

França C (1913). Contribution à L'Étude du Lapin de Porto Santo. *Bulletin de la Société Portugaise de Sciences Naturelle*, Tome VI, Fasc. 2.

Frutuoso G (1873). As saudades da Terra. *In* França, C.. Contribution à L'Étude du Lapin de Porto Santo. *Bulletin de la Société Portugaise de Sciences Naturelle*, Tome VI, Fasc. 2.

Hæckel E (1874). *História da criação dos seres organizados segundo as leis naturais*. Lello & Irmão Editores, Porto.

Hemmer H (1990). *Domestication. The decline of environmental appreciation.* Cambridge University Press.

Hudson RR (1990) Gene genealogies and the coalescent process, pp. 1-42 in *Oxford Survey in Evolutionary Biology*, Vol. 7, edited by D. Futuyama and J. Antonovics. Oxford University Press, Oxford.

Kelus AS and Gell PG (1967). Immunoglobulin allotypes of experimental animals. *Prog Allergy* **11:** 141-84.

Kimura M and Crow JF (1964). The number of alleles that can be maintained in a finite population. *Genetics*, **49**: 725-38.

Lande R (1994). Risk of population extinction from fixation of new deleterious mutations. *Evolution* **48**: 1460-1469.

Luikart G and Cornuet JM (1998). Empirical Evaluation of a Test for Identifying Recently Bottlenecked Populations from Allele frequency Data. *Conservation Biology* **12**: 228-237.

Miller GS (1912). Catalogue of the Mammals of Western Europe in the collection of the British Museum. *In* França C. Contribution à L'Étude du Lapin de Porto Santo. *Bulletin de la Société Portugaise de Sciences Naturelle*, Tome VI, Fasc. 2.

Maruyama T and Fuerst P (1985). Population bottlenecks and nonequilibrium models in population genetics. I. Number of alleles in a small population that was formed by a recent bottleneck. *Genetics* **111**: 675-689.

Nei M and Maruyama T, Chakraborty R (1975). The bottleneck effect and genetic variability in populations. *Evolution* **29:** 1-10.

Nei M, Chakraborty R and Fuerst PA (1976). Infinite allele model with varying mutation rate. *Proc Natl Acad Sci* U S A. **73**(11): 4164-168.

Nei M (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **23:** 341-69.

Nei M (1987). *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Newman D and Pilson D (1997). Increased probability of extinction due to decreased genetic effective population size: experimental populations of *Clarkia pulchella. Evolution* **51**: 345-62.

O'Brien E, Kerber RA, Jorde LB and Rogers AR (1994). Founder Effect: Assessment of variation in genetic contributions among founders. *Human Biology* **66**: 185-204.

Ohta T and Kimura M (1973). A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genetical Research Cambridge* **22**: 201-04.

Prakashi S (1972). Origin of reproductive isolation in the absence of apparent genic differentiation in a geographic isolate of *Drosophila pseudoobscura. Genetics* **72**: 143-55.

Queney G (2000). *Histoire des populations et organisation sociale du lapin européen (Oryctolagus cuniculus) à travers l'étude de marqueurs microsatellites*. These Doctoural, Paris.

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*). *Mol Biol Evol* **18**(12): 2169-78.

Queney G, Vachot AM, Brun JM, Dennebouy N, Mulsant P, Monnerot M (2002). Different levels of human intervention in domestic rabbits: effects on genetic diversity. *J Hered* 93(3): 205-9.

Saccheri I, Kuussaari M, Vikman W, Fortellius W and Hanski (1998). Inbrreeding and extinction in a metapopulation. *Nature* **392**: 491-94.

Saitou N and Nei M (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*, **4**(4): 406-25.

Su C and 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): 9710-715.

Swofford DL and Selander RB (1989). BYOSIS-1. A computer program for the analysis of allelic variation in population genetics and biochemicals systematic. Release 1.7 Illinois Natural History Survey.

van der Loo W, Mougel F, Sanchez MS, Bouton C, Castien E, Fonseca A, Ferrand N, Soriguer R and 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): 629-43.

Zeuner FE (1963). *A history of domesticated animals*. Hutchinson.

# CHAPTER 5

## Conclusions

In this work we studied two rabbit immunogenetic loci, the *IgGCH2* or *e* locus and the *IgVH1* or *a* locus. These loci both contribute to biosynthesis of an immunoglobulin H chain. However, they differ in complexity, in the level of diversity and in the phylogeographic patterns of distribution. The *IgGCH2* is a single gene locus whereas the *IgVH* gene locus is a multigene locus. The *IgCH2* polymorphism is restricted to a single amino acid substitution whereas large inter-allelic distances characterise the *IgVH1* polymorphism. The *IgGCH2* polymorphism has a recent age whereas the *IgVH1* is very ancient and persisted in the populations for long period of time. For the complex rabbit allotypes (*IgVH* or *a*-locus and *IgCK1* or *b*-locus) the initial evidence for the adaptiveness of variation is based upon large inter-allelic distances. In contrast, in the case of *IgGCH2* it is based upon population genetic evidence (see chapter 2), which indicated that the polymorphism at the *e*-locus is adaptive only through the lack of diversity at *a*- and *b*-loci.

## Gene diversity at the *e* locus (*IgGCH2*) in *Oryctolagus cuniculus*

The geographical variation at this locus differs from that observed at other loci studied until now. Indeed, in the original species range, this polymorphism, defined by one amino acid substitution at position 309, is fixed for *e15* allele whereas it is polymorphic in all wild and feral rabbit populations studied ($p_{e14} \approx 20\%$) in the more recent area of the species distribution. The linkage disequilibrium with $V_H$ locus implies that *e14* is a recent polymorphism (see chapter 2). However, the sequence data available prior to this work indicated that the *e14* allele was more ancestral than *e15*. Our data for Iberian wild rabbit specimens show that the allele *e14* can indeed be derived by a single nucleotide substitution from an *e15* lineage present in the Iberian Peninsula. The sequences obtained here also revealed six new polymorphic nucleotide positions, of which two were associated with amino acid substitutions. The occurrence of this polymorphism can be explained either by compensatory overdominance (i.e. it emerged in order to compensate the reduction of diversity at *a*- and *b*- loci in the populations from outside of Iberian Peninsula) or by a selective advantage of *e14* allele when exposed to a particular microorganism.

The sequences that we produced from other leporids and the analyses of CH2 sequences of other mammals, showed in general an excess of dN (non-synonymous substitutions) at this position, which confirms the population genetic evidence for adaptive variation.

## Gene diversity at the *IgVH* in leporids

### *Genetic diversity in Iberian rabbit populations*

In domestic breeds and in wild and feral populations from outside the Iberian Peninsula in essence three alleles (*a1*, *a2* and *a3*) contribute effectively to the *a* locus diversity. We analysed serologically 18 wild rabbit Iberian Peninsula populations which allowed us to distinguish different endemic alleles, one of which ("*a*-blank") was particular by failing to react with specific *a1*, *a2* and *a3* antisera. The frequency distribution suggests that the "*a-blank*" phenotype is associated with the subspecies *O. c. algirus*. This pattern is similar to that previously reported in the Iberian Peninsula at *IgCK1* or *b* locus, where the most common alleles (*b4* and *b5*) of *O. c. cuniculus* populations were rare or absent in *O. c. algirus* populations. These associations of Ig allotypes with subspecies could be explained by the fact that subspecies occupy different habitats, where there are exposed to different environmental factors. A bottleneck effect seems to be the most likely explanation for the fact that the most common alleles of the Iberian populations did not cross the Pyrenean Mountains.

Our data show that the rabbits serologically homozygous for the "*a-blank*" allele use preferentially $V_H$ genes that, although clearly related to the known $V_H1$ genes, define a new major allotypic lineage, called here *a4*. We conclude that there exist at least four distantly related lineages at the rabbit *IgVH1* or *a* locus.

### *Genetic diversity in* Lepus *species*

*VH* gene sequences of leporids (other than *Oryctolagus*) had never been obtained before this work. Here we sequenced *VH* genes that were expressed in specimens of two *Lepus*, *L. europaeus* and *L. granatensis* species. The results obtained show two *VH* lineages: one with the signature of $V_Ha$ alleles (*a2L*), and another unrelated to these (*nL*). The comparison of these lineages with the rabbit alleles revealed that *a2L* is more similar to the rabbit *a2* than *a2* is to its rabbit allelic counterparts. This supports a trans-species polymorphism hypothesis, i.e. that very long lineage persistence times explain this high allelic divergence.

### *Are the allotypes described in* Lepus *(*a2L *and* nL*) truly alleles?*

The $V_H$ gene segments that encode the allotypes *a1*, *a2* and *a3* have been mapped in the most *D*-proximal region of the $V_H$ cluster (i.e. $V_H1$). However, the genomic location of

these newly described lineages (*a4*, *a2L* and *nL*) is unknown. It seems likely that the lineages *a4* and *a2L*, which cluster with the rabbit allotypes *a1*, *a2* and *a3* in the $V_H$*a*-like group, are encoded by allelic variants of the $V_H$*1* locus. However, the *nL* sequences show more sequence similarity with the $V_H$*x*, and $V_H$*y* gene segments of rabbit, which, in this species, are located more than 50 kb upstream of the $V_H$*1* locus. The preferential expression of the *nL* allotype in *Lepus* specimens could indicate: a) they are indeed encoded by the *D*-proximal $V_H$ locus (i.e. the homologue of rabbit $V_H$*1*). Such situation would suggest that $V_H$ gene segments can be reshuffled; b) unlike in the rabbit, in hare species the preferential expression of $V_H$ genes is not imposed by its *D*-proximal position. To evaluate the likelihood of these scenarios a genomic library of $V_H$ gene segments must be constructed with amplification and sequencing of the most *D*-proximal $V_H$ genes.

### *The role of intestinal microflora in the* $V_H1$ *polymorphism*

The origin of the selection forces that regulate the $V_H$*1* polymorphism is unknown. However, the requirements of exogenous factors, such as the intestinal microflora, for the diversification of the primary rabbit antibody repertoire in the rabbits, suggest that this mechanism might impose constraints on the evolution of the $V_H$*1* alleles. The rabbit is a species with altricial young (naked and blind at birth). Other lagomorph species, such as hares (genus *Lepus*) have precocial young. Considering the differences in reproductive biology between the genera *Oryctolagus* and *Lepus*, it might be interesting to investigate if in *Lepus* the Ab repertoire is diversified neonatally or during foetal life (rather than after two or three weeks like in *Oryctolagus*). The unique situation of post-natal diversification in rabbit could be related to the shortening of gestation times (4 weeks in rabbit and six weeks in hares).

### *The* $V_H1$ *large inter-allelic differences*

Regarding the two different hypotheses invoked to explain the very large interallelic distances observed at the *IgVH1 locus*, the obtained results seem to be contradictory. On one hand, the association of specific alleles with subspecific markers suggests that the large differences between *a4* and domestic alleles (*a1*, *a2* and *a3*) 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 $V_H$*1* polymorphism between *Oryctolagus* and *Lepus* supports the hypothesis that very long lineage persistence times have contributed to allelic divergence. The different evidences obtained point 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.

For persistence of distinct lineages it is necessary that, during the several episodes of population expansion and retraction, the effective population size remains large enough to maintain these lineages. The study of Porto Santo showed that one population might survive with a gene pool reduced to four chromosomes. This makes it unlikely that four or more allelic lineages can be maintained for millions of years.

*Evolution of* $V_H$ *gene in the lagomorph group*

Under the concerted evolution model, the multigene family genes show high homogenisation and all the genes of the same species are clustered together. In contrast, the multigene family genes evolving according to the birth-and-death model, when duplicated, can diverge functionally. The end result of this process is an admixture of both divergent and highly homologous groups of genes. The analyses of $V_H$ gene segments obtained in several vertebrate species showed different evolutionary patterns. Indeed, the human and mouse possess $V_H$ gene sequences that are very diversified and clustered within all groups, whereas $V_H$ gene sequences from birds, rabbit, swine, sheep and cattle, are restricted to one $V_H$ group. To understand the mechanisms that regulate the evolution of these genes, more species must be studied, in particular those that live in natural conditions and are submitted to changing environmental conditions.

We extended the knowledge of the $V_H$ gene segment to other lagomorph species (genera: *Sylvilagus* and *Ochotona*). The phylogenetic analysis of these sequences suggests that only $V_H$ genes belonging to group C are present in the genome of lagomorph species. Probably, the loss of $V_H$ genes from groups A and B happened in a lagomorph ancestor at least 50 My ago, which is the estimated time of divergence between *Ochotonidae* and *Leporidae*. We speculate that in the lagomorph ancestor a particular family of $V_H$ gene (class C) became more used in *VDJ* rearrangements and compensated at the same time the reduction of diversity by somatic diversification. In this way, the loss of all other $V_H$ gene families from the genome became possible.

The lagomorph $V_H$ gene group is divided into four clusters, $V_Ha$, $V_Hn$, $V_HnL$ and $V_HP$. The $V_H$ gene sequences from different lagomorph species can cluster within the same group and $V_H$ gene sequences of the same species are sometimes clustered in different

groups (for example in both $V_Ha$ and $V_Hn$ group are present sequences of *Lepus*, *Oryctolagus* and S*ylvilagus)*. These results suggest that the lagomorph $V_H$ gene segments follow the model of birth-and-death evolution, i.e. during the $V_H$ gene duplication process several genes diverged functionally while others maintained high homology. The intestinal microflora is probably a factor contributing to the occurrence of highly divergent $V_H$ genes sequences within the lagomorph group.

## Evaluating the bottleneck effect in the Porto Santo wild rabbit population

The data reported in this work support historical evidence that only one pregnant rabbit female was introduced to the Porto Santo Island. Introductions with a reduced effective size generally provoke a reduction in the number of alleles and in the heterozygosity values. This is more dramatic when only four chromosomes founded the population. The different parameters of genetic variability (He, P and $n_a$) obtained for protein, microsatellite and immunogenetic markers in Porto Santo show that, although the level of diversity in this island is lower than the genetic diversity detected in the Iberian populations, it is still higher than the natural French populations and domestic breeds. This relatively high level of diversity may explain why a population founded by only one pregnant female became established and survived until today and give some indications about the minimum level of diversity that is necessary for sustaining long-term population viability. It highlights the fact that it is not the size of the founder population that is important, but the amount of genetic variation it contains.

# Appendix 1

## Question: What is the number of possible rearrangements with p alleles taking out n chromosomes from one population?

The total number of possible rearrangements for each possibility is estimated by:

$$\alpha_n{}^{p} = p^n,$$

where p is the number of distinct categories (alleles) and n is the number of classes (chromosomes).

a) The probability of obtaining one allele is given by the formula:

$$P(X) = p^n,$$

where p is the frequency of allele X and n is the number of classes.

b) The probability of obtaining two alleles is given by the formula of the distribution of binomial probability:

$$P(X) = \frac{n!}{n\text{-}x!\times x!} \times \left[\left(q^{n-x}\right) \times p^x\right],$$

where, n is the number of classes, x is the number of copies of each distinct category; p and q are the frequencies of each category.

c) The probability to obtain three or more alleles is giving by the formula of distribution of multinomial probability

$$P(X_1, X_2, \ldots\ldots, X_k) = \frac{n!}{X_1! \times X_2! \ldots X_k} \times \left[\left(p_1{}^{X1} \times p_2{}^{X2} \times \ldots\ldots \times p_k{}^{Xk}\right)\right]$$

where, n is the number of classes, x is the number of copies of each distinct categories, p is the frequency of each category.

Hypothesis 1: The Porto Santo rabbit population was originated from only one pregnant female (four genes)

*1. To take out four genes and to obtain only one allele (a)*

Total number of rearrangements:

$$\alpha_4{}^1 = 1^4 = 1$$

*a) fixation of one allele* (all copies belong to the same category)*

$[\ a^4\ ]\quad 1\ *$

*2. To take out four genes and to obtain two alleles (a, b):*

Total number of possible rearrangements

$$\alpha_4{}^2 = 2^4 = 16$$

*a) fixation of one allele* (all copies belong to the same category)

$$\begin{pmatrix} a^4 \\ b^4 \end{pmatrix} \quad 2$$

*b) to obtain only two alleles* (categories *a* and *b*) *

$$\begin{pmatrix} 4a^3b \\ 6a^2b^2 \\ 4ab^3 \end{pmatrix} \quad 14\ *$$

$$\text{e.g. } P(3a,1b) = \frac{4!}{3!\times 1!} \times \left[\left(p_a{}^3\right) \times p_b\right] = 4ab^3$$

The sum of rearrangements obtained in the point a) and b) (2+14) correspond as expected to the total number of possible rearrangements 16.

*3. To take out four genes and to obtain three distinct alleles (a, b, and c):*

Total number of possible rearrangements

$$\alpha_4^3 = 3^4 = 81$$

*a) fixation of one allele* (all copies belong to the same category)

$$\begin{pmatrix} a^4 \\ b^4 \\ c^4 \end{pmatrix} \quad 3$$

*b) to obtain only two alleles* (categories *a* and *b; a* and *c; b* and *c*)

$$\begin{pmatrix} 4a^3b \\ 6a^2b^2 \\ 4ab^3 \end{pmatrix} \begin{pmatrix} 4a^3c \\ 6a^2c^2 \\ 4ac^3 \end{pmatrix} \begin{pmatrix} 4b^3c \\ 6b^2c^2 \\ 4bc^3 \end{pmatrix} \quad 3\times14 = 42$$

*c) to obtain three distinct alleles* (distinct categories *a, b,* and *c*) *

$$\begin{pmatrix} 12a^2bc \\ 12ab^2c \\ 12abc^2 \end{pmatrix} \quad 3\times12 = 36\ *$$

$$\text{e.g. } P(2a,1b,1c) = \frac{4!}{2!\times1!\times1!}\times\left[\left(p_a{}^2\right)\times p_b\times p_c\right] = 12a^2bc$$

The sum of rearrangements obtained in the point a), b), and c) (3+42+36) correspond as expected to the total number of possible rearrangements 81.

*4. To take out four genes and to obtain four distinct alleles (a, b, c, and d):*

Total number of possible rearrangements

$$\alpha_4^4 = 4^4 = 256$$

*a) fixation of one allele (all copies* belong to the same category)

$$\begin{pmatrix} a^4 \\ b^4 \\ c^4 \\ d^4 \end{pmatrix} \quad 4$$

*b) to obtain only two alleles* (categories: *a* and *b; a* and *c; a* and *d; b* and *c; b* and *d; c* and *d*).

$$\begin{pmatrix} 4a^3b \\ 6a^2b^2 \\ 4ab^3 \end{pmatrix} \begin{pmatrix} 4a^3c \\ 6a^2c^2 \\ 4ac^3 \end{pmatrix} \begin{pmatrix} 4a^3d \\ 6a^2d^2 \\ 4ad^3 \end{pmatrix} \begin{pmatrix} 4b^3c \\ 6b^2c^2 \\ 4bc^3 \end{pmatrix} \begin{pmatrix} 4b^3d \\ 6b^2d^2 \\ 4bd^3 \end{pmatrix} \begin{pmatrix} 4c^3d \\ 6c^2d^2 \\ 4cd^3 \end{pmatrix}$$

$$6 \times 14 = 84$$

*c) to obtain three distinct alleles* (categories: *a, b,* and *c; a, b, and d; a, c, and d; b, c, and d*).

$$\begin{pmatrix} 12a^2bc \\ 12ab^2c \\ 12abc^2 \end{pmatrix} \begin{pmatrix} 12a^2bd \\ 12ab^2d \\ 12abd^2 \end{pmatrix} \begin{pmatrix} 12a^2cd \\ 12ac^2d \\ 12acd^2 \end{pmatrix} \begin{pmatrix} 12b^2cd \\ 12bc^2d \\ 12bcd^2 \end{pmatrix}$$

$$4 \times 36 = 144$$

*d) to obtain four distinct alleles* (distinct categories: a, b, c, and d) *.

$$\begin{pmatrix} 24\ abcd \end{pmatrix} \quad 24\ *$$

$$\text{e.g. } P(1a,1b,1c,1d) = \frac{4!}{1! \times 1! \times 1! \times 1!} \times (p_a \times p_b \times p_c \times p_d) = 24abcd$$

The sum of rearrangements obtained in the point a), b), c), and d) (4+84+144+24) correspond as expected to the total number of possible rearrangements 256.

Hypothesis 2: The Porto Santo rabbit population was originated from three rabbits (six genes)

*1. To take out six genes and to obtain only one allele (a):*

Total number of possible rearrangements

$$\alpha_6^{1} = 1^6 = 1$$

*a) fixation of one allele* (all copies belong to the same category) *

$[\ a^6\ ]\ 1$*

*2. To take out six genes and to obtain two alleles (a,b):*

**Total number of possible rearrangements**

$$\alpha_6^{2} = 2^6 = 64$$

*a) fixation of one allele* **(all copies belong to the same category)**

$$\begin{bmatrix} a^6 \\ b^6 \end{bmatrix} \quad 2$$

*b) to obtain only two alleles* (distinct categories a and b) *

$$\begin{pmatrix} 6a^5b \\ 15a^4b^2 \\ 20a^3b^3 \\ 15a^2b^4 \\ 6ab^5 \end{pmatrix} \quad 62\ *$$

$$\text{e.g. } P(5a,1b) = \frac{6!}{5! \times 1!} \times \left[\left(p_a{}^5\right) \times p_b\right] = 6a^5b$$

The sum of rearrangements obtained in the point a) and b) (2+62) corresponds as expected to the total number of possible rearrangements 16.

*3. To take out six genes and to obtain three distinct alleles (a, b, and c):*

Total number of possible rearrangements

$\alpha_6^3 = 3^6 = 729$

*a) fixation of one allele* (all copies belong to the same category)

$$\begin{pmatrix} a^6 \\ b^6 \\ c^6 \end{pmatrix} \quad 3$$

*b) to obtain only two alleles* (distinct categories: *a* and *b; a* and *c; b* and c).

$$\begin{pmatrix} 6a^5b \\ 15a^4b^2 \\ 20a^3b^3 \\ 15a^2b^4 \\ 6ab^5 \end{pmatrix} \quad \begin{pmatrix} 6a^5c \\ 15a^4c^2 \\ 20a^3c^3 \\ 15a^2c^4 \\ 6ac^5 \end{pmatrix} \quad \begin{pmatrix} 6b^5c \\ 15b^4c^2 \\ 20b^3c^3 \\ 15b^2c^4 \\ 6bc^5 \end{pmatrix} \quad 62\times3=186$$

*c) to obtain three distinct alleles* (distinct categories: *a, b*, and *c*) *

$$\begin{pmatrix} 30a^4bc \\ 60a^3b^2c \\ 60a^3bc^2 \\ 60a^2bc^3 \\ 60a^2b^3c \\ 90a^2b^2c^2 \\ 60ab^3c^2 \\ 60ab^2c^3 \\ 30ab^4c \\ 30abc^4 \end{pmatrix} \quad 540\ *$$

$$\text{e.g. } P(4a,1b,1c) = \frac{6!}{4!\times1!\times1!} \times \left[\left(p_a{}^4\right) \times p_b \times p_c\right] = 30a^4bc$$

The sum of rearrangements obtained in the point a), b), and c) (3+186+540) correspond as expected to the total number of possible rearrangements 729.

*4. To take out six genes and to obtain four distinct alleles (a, b, c, and d):*

Total number of possible rearrangements

$$\alpha_6{}^4 = 6^4 = 4096$$

*a) fixation of one allele* (all copies belong to the same category)

$$\begin{pmatrix} a^6 \\ b^6 \\ c^6 \\ d^6 \end{pmatrix} \quad 4$$

*b) to obtain two alleles* (categories a and b; a and c; a and d; b and c; b and d; c and d).

$$\begin{pmatrix} 6a^5b \\ 15a^4b^2 \\ 20a^3b^3 \\ 15a^2b^4 \\ 6ab^5 \end{pmatrix} \begin{pmatrix} 6a^5c \\ 15a^4c^2 \\ 20a^3c^3 \\ 15a^2c^4 \\ 6ac^5 \end{pmatrix} \begin{pmatrix} 6a^5d \\ 15a^4d^2 \\ 20a^3d^3 \\ 15a^2d^4 \\ 6ad^5 \end{pmatrix} \begin{pmatrix} 6b^5c \\ 15b^4c^2 \\ 20b^3c^3 \\ 15b^2c^4 \\ 6bc^5 \end{pmatrix} \begin{pmatrix} 6b^5d \\ 15b^4d^2 \\ 20b^3d^3 \\ 15b^2d^4 \\ 6bd^5 \end{pmatrix} \begin{pmatrix} 6c^5d \\ 15c^4d^2 \\ 20c^3d^3 \\ 15c^2d^4 \\ 6cd^5 \end{pmatrix}$$

$$6 \times 62 = 372$$

*c) to obtain three alleles* (categories: a, b, and c; a, b, and d; a, c, and d; b, c, and d).

$$\begin{pmatrix} 30a^4bc \\ 60a^3b^2c \\ 60a^3bc^2 \\ 60a^2bc^3 \\ 60a^2b^3c \\ 90a^2b^2c^2 \\ 60ab^3c^2 \\ 60ab^2c^3 \\ 30ab^4c \\ 30abc^4 \end{pmatrix} \begin{pmatrix} 30a^4bd \\ 60a^3b^2d \\ 60a^3bd^2 \\ 60a^2bd^3 \\ 60a^2b^3d \\ 90a^2b^2d^2 \\ 60ab^3d^2 \\ 60ab^2d^3 \\ 30ab^4d \\ 30abd^4 \end{pmatrix} \begin{pmatrix} 30a^4cd \\ 60a^3c^2d \\ 60a^3cd^2 \\ 60a^2cd^3 \\ 60a^2c^3d \\ 90a^2c^2d^2 \\ 60ac^3d^2 \\ 60ac^2d^3 \\ 30ac^4d \\ 30acd^4 \end{pmatrix} \begin{pmatrix} 30b^4cd \\ 60b^3c^2d \\ 60b^3cd^2 \\ 60b^2cd^3 \\ 60b^2c^3d \\ 90b^2c^2d^2 \\ 60bc^3d^2 \\ 60bc^2d^3 \\ 30bc^4d \\ 30bcd^4 \end{pmatrix}$$

$$4 \times 540 = 2160$$

*d) to obtain four distinct categories: a, b, c, and d**

$$\begin{pmatrix} 120a^3bcd \\ 120ab^3cd \\ 120abc^3d \\ 120abcd^3 \\ 180a^2b^2cd \\ 180a^2bc^2d \\ 180a^2bcd^2 \\ 180ab^2c^2d \\ 180ab^2cd^2 \\ 180abc^2d^2 \end{pmatrix} \quad (4x120)+(6x180)=1560\ *$$

$$\text{e.g. } P(3a,1b,1c,1d) = \frac{6!}{3!\times 1!\times 1!\times 1!} \times (p_a \times p_b \times p_c \times p_d) = 120abcd$$

The sum of rearrangements obtained in the point a), b), c), and d) (4+372+2160+1560) correspond as expected to the total number of possible rearrangements 4096.

**Note**: The formulas used to calculate the probabilities expected for Porto Santo under the two hypotheses considered above are marked with *.

# Appendix 2

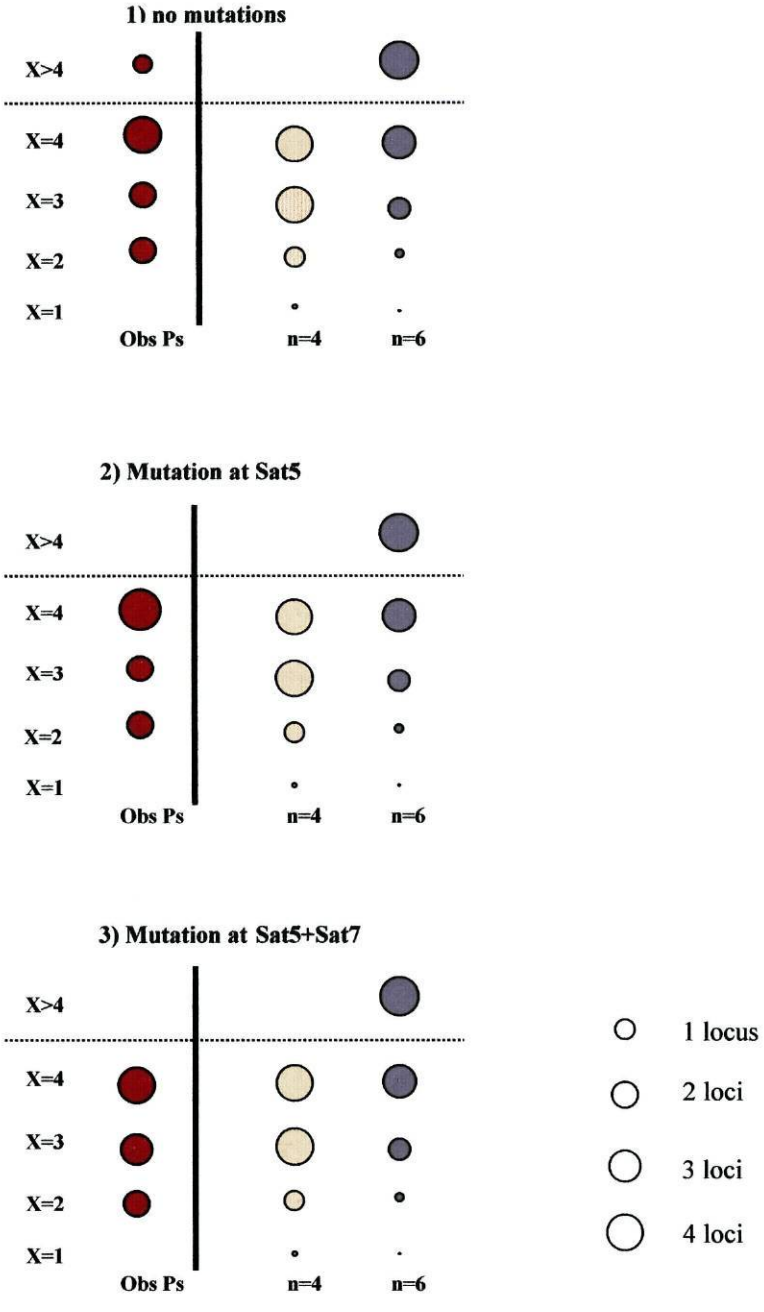



**Figure A1:** Three different scenarios considered to explain the results obtained at microsatellite markers in the Porto Santo rabbit population. 1) no mutations; 2) mutation at Sat5; 3) mutations at Sat5 and Sat7. The criteria used for considering the possibility of mutation at microsatellite loci combined a low allelic frequency ($p_i < 0.05$) with the close vicinity of putative mutants to major alleles (gene frequencies obtained in Sat5: 206-0.14; 210-0.38; 214-0.20; **216-0.03**; 222-0.25 and in Sat7: **185-0,02**; 187-0,54; 191-0,21; 193-0,23). X=1, 2, 3, 4, >4 are the number of alleles per locus observed in Porto Santo (Obs Ps) and expected under the hypothesis four (n=4) or six (n=6) founding chromosomes. The size of each circle corresponds to the number of loci observed and expected with 1, 2, 3, 4 or more than 4 alleles.

**Table A1:** Distribution values obtained and expected from the gene frequencies of 9 microsatellite for each class of alleles per locus (X=1, 2, 3, 4 or >4) under the hypotheses of 4 or 6 chromosomes founding the Porto Santo rabbit population. X is the number of alleles expected per locus.

| | N=4 | | | | N=6 | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Microsatellite | X=1 | X=2 | X=3 | X=4 | X=1 | X=2 | X=3 | X=4 | X>4 |
| | expected | | | | | | | | |
| Sat2 | 0,004 | 0,079 | 0,3917 | 0,5251 | 0,0002 | 0,008 | 0,0784 | 0,2838 | 0,6294 |
| Sat3 | 0,0015 | 0,065 | 0,4110 | 0,5224 | 3,11E-05 | 0,004 | 0,0665 | 0,2979 | 0,6314 |
| Sat4 | 0,01766 | 0,187 | 0,4739 | 0,3214 | 0,0022 | 0,045 | 0,2196 | 0,3910 | 0,3420 |
| Sat5 | 0,0012 | 0,054 | 0,3813 | 0,5634 | 2,42E-05 | 0,003 | 0,0508 | 0,2504 | 0,6956 |
| Sat7 | 0,0052 | 0,142 | 0,5018 | 0,3510 | 0,0002 | 0,019 | 0,1848 | 0,4339 | 0,3619 |
| Sat8 | 0,0005 | 0,03 | 0,3079 | 0,6615 | 3,75E-06 | 0,001 | 0,0230 | 0,1747 | 0,8011 |
| Sat12 | 0,0169 | 0,286 | 0,5291 | 0,1680 | 0,0015 | 0,075 | 0,3654 | 0,4206 | 0,1373 |
| Sat13 | 0,0050 | 0,134 | 0,5193 | 0,3416 | 0,0002 | 0,017 | 0,1671 | 0,4270 | 0,3886 |
| Sat16 | 0,0066 | 0,1689 | 0,5322 | 0,2923 | 0,0066 | 0,0263 | 0,2309 | 0,4609 | 0,2750 |
| total | 0,0587 | 1,1458 | 4,0485 | 3,7469 | 0,0111 | 0,1983 | 1,3869 | 3,1408 | 4,2627 |
| | Observed Ps | | | | | | | | |
| Mutation in Sat5 | 0 | 2 | 2 | 5 | 0 | 2 | 2 | 5 | 0 |
| $\chi^2$ | 0,0587 | 0,6367 | 1,0365 | 0,4190 | 0,0111 | 16,361 | 0,271 | 1,1005 | 4,2627 |
| $\chi^2$ total | | | | 2,1511 | | | | | 22,007 |
| | Observed Ps | | | | | | | | |
| Mutations sat5+Sat7 | 0 | 2 | 3 | 4 | 0 | 2 | 3 | 4 | 0 |
| $\chi^2$ | 0,0587 | 0,6367 | 0,2715 | 0,0170 | 0,0111 | 16,361 | 1,8761 | 0,2350 | 4,2627 |
| $\chi^2$ total | | | | 0,9841 | | | | | 22,746 |
| | Observed Ps | | | | | | | | |
| no Mutations | ---- | ---- | ---- | ---- | 0 | 2 | 2 | 4 | 1 |
| | ---- | ---- | ---- | ---- | 0,0111 | 16,361 | 0,271 | 0,2350 | 2.497 |
| | | | | | | | | | 19.104 |

**Table A2:** Distribution values obtained and expected from the gene frequencies of 20 protein and 2 immunological markers for each class of alleles per locus (X=1, 2, 3, 4 or >4) under the hypotheses of 4 or 6 chromosomes founding the Porto Santo rabbit population. X is the number of alleles expected per locus.

| Proteins | N=4 X=1 | X=2 | X=3 | X=4 | N=6 X=1 | X=2 | X=3 | X=4 | X>4 |
|---|---|---|---|---|---|---|---|---|---|
| | expected | | | | | | | | |
| NP | 0,76 | 0,2311 | 0,0054 | -------- | 0,6671 | 0,3202 | 0,0127 | -------- | -------- |
| PEPA | 0,1087 | 0,7902 | 0,100 | 0,0008 | 0,0255 | 0,7902 | 0,1809 | 0,0034 | 0,00001 |
| PEPB | 0,6395 | 0,3291 | 0,031 | 0,0005 | 0,5114 | 0,4186 | 0,0679 | 0,0021 | -------- |
| PEPC | 0,0685 | 0,4880 | 0,3980 | 0,0453 | 0,0155 | 0,2992 | 0,4735 | 0,1910 | 0,02092 |
| PEPD | 0,0916 | 0,6727 | 0,225 | 0,0107 | 0,0209 | 0,5734 | 0,3605 | 0,0438 | 0,00139 |
| PGD | 0,1226 | 0,8388 | 0,038 | -------- | 0,0314 | 0,8969 | 0,0717 | -------- | -------- |
| ADA | 0,2105 | 0,6943 | 0,093 | 0,0022 | 0,0925 | 0,7273 | 0,1723 | 0,0079 | 0,00004 |
| GALT | 0,3882 | 0,4971 | 0,112 | 0,0026 | 0,2417 | 0,5261 | 0,2205 | 0,01149 | 0,00014 |
| HBA | 0,3332 | 0,6367 | 0,030 | -------- | 0,1902 | 0,7488 | 0,0610 | -------- | -------- |
| HBB | 0,7651 | 0,2349 | -------- | -------- | 0,6692 | 0,3308 | -------- | -------- | -------- |
| ACP3 | 0,7642 | 0,2324 | 0,003 | -------- | 0,6681 | 0,3240 | 0,0079 | -------- | -------- |
| ALB | 0,1243 | 0,8682 | 0,007 | -------- | 0,0311 | 0,9549 | 0,0140 | -------- | -------- |
| TF | 0,18 | 0,7211 | 0,093 | 0,0015 | 0,0737 | 0,7496 | 0,1662 | 0,01018 | 0,00018 |
| CAI | 0,9683 | 0,0316 | -------- | -------- | 0,9529 | 0,0471 | -------- | -------- | -------- |
| CAII | 0,2986 | 0,6474 | 0,054 | 0,0003 | 0,1608 | 0,7316 | 0,1065 | 0,00118 | -------- |
| GC | 0,4211 | 0,4646 | 0,108 | 0,0067 | 0,2731 | 0,4966 | 0,2039 | 0,0255 | 0,00093 |
| BF | 0,0471 | 0,5012 | 0,416 | 0,0352 | 0,0071 | 0,2800 | 0,5769 | 0,12868 | 0,00723 |
| HPX | 0,0652 | 0,4492 | 0,420 | 0,0657 | 0,0153 | 0,2432 | 0,5077 | 0,20701 | 0,02676 |
| SOD | 0,8274 | 0,1726 | -------- | -------- | 0,7527 | 0,2473 | -------- | -------- | -------- |
| AT3 | 0,7032 | 0,2708 | 0,0253 | 0,0006 | 0,5897 | 0,3513 | 0,0559 | 0,00296 | 0,00004 |
| IGHV1 | 0,0696 | 0,4213 | 0,3662 | 0,1428 | 0,0173 | 0,2281 | 0,4453 | 0,24263 | 0,06666 |
| IGKC1 | 0,0167 | 0,2028 | 0,4805 | 0,2999 | 0,0019 | 0,0501 | 0,2489 | 0,4022 | 0,29697 |
| Total | 7,98 | 10,40 | 3,01 | 0,61 | 6,0091 | 10,3354 | 3,9542 | 1,2800 | 0,4213 |
| | Observed Pst | | | | | | | | |
| | 8 | 8 | 5 | 1 | 8 | 8 | 5 | 1 | 0 |
| $\chi^2$ | 4,04029E-05 | 0,5525 | 1,3216 | 0,2416 | 0,6595 | 0,5276 | 0,2766 | 0,0612 | 0,42131 |
| $\chi^2$ total | | | | 2,1158 | | | | | 1,9464 |