

## ORIGINAL ARTICLE

Evidence for contrasting modes of selection at interacting globin genes in the European rabbit (*Oryctolagus cuniculus*)R Campos<sup>1,3</sup>, JF Storz<sup>2</sup> and N Ferrand<sup>1,3</sup><sup>1</sup>CIBIO-Centro de Investigação em Biodiversidade e Recursos Genéticos, Campus Agrário de Vairão, 4485-661, Vairão, Portugal; <sup>2</sup>School of Biological Sciences, University of Nebraska, Lincoln, NE, USA and <sup>3</sup>Departamento de Zoologia e Antropologia, Faculdade de Ciências da Universidade do Porto, Praça Gomes Teixeira, 4099-002, Porto, Portugal

In hybrid zones between genetically differentiated populations, variation in locus-specific rates of introgression may reflect adaptation to different environments or adaptation to different genetic backgrounds. The European rabbit, *Oryctolagus cuniculus*, is well-suited to studies of such hybrid zone dynamics because it is composed of two genetically divergent subspecies that hybridize in a zone of secondary contact in central Iberia. A species-wide survey of allozyme variation revealed a broad range of locus-specific divergence levels ( $F_{ST}$  ranged from 0 to 0.54, mean  $F_{ST} = 0.16$ ). Interestingly, the two loci that fell at opposite ends of the distribution of  $F_{ST}$  values, haemoglobin  $\alpha$ -chain (HBA) and haemoglobin  $\beta$ -chain

(HBB), encode interacting subunits of the haemoglobin protein. The contrasting patterns of spatial variation at these two loci could not be reconciled under a neutral model of population structure. The HBA gene exhibited higher-than-expected levels of population differentiation, consistent with a history of spatially varying selection. The HBB gene exhibited lower-than-expected levels of population differentiation, consistent with some form of spatially uniform selection. Patterns of linkage disequilibrium and allele frequency variation do not appear to fit any simple model of two-locus epistatic selection. *Heredity* (2008) **100**, 602–609; doi:10.1038/hdy.2008.26; published online 30 April 2008

**Keywords:** adaptation; selection; allozyme; population structure; coalescent; HBA and HBB

## Introduction

Hybrid zones can be regarded as natural filters for the identification of genes that distinguish closely related populations or species (Barton and Bengtsson, 1986; Arnold, 1997; Martinsen *et al.*, 2001; Wu, 2001; Orr *et al.*, 2004; Wu and Ting, 2004). When two hybridizing populations harbour alternative sets of coadapted alleles at different loci, hybrids may suffer reduced fitness because the allelic combinations specific to one population do not function properly when arrayed against the novel genetic background of the other population (Bateson, 1909; Dobzhansky, 1936; Muller, 1942). This type of negative epistasis should result in reduced levels of introgression at the interacting genes relative to a genome-wide average for unlinked marker loci (Payseur *et al.*, 2004; Dopman *et al.*, 2005; Payseur and Hoekstra, 2005). Alternatively, when alleles that are specific to one population are recombined against the genetic background of another population, the novel combinations of alleles at different loci may provide an important source of adaptive genetic variation (Barton and Hewitt, 1985; Woodruff, 1989; Rieseberg *et al.*, 1996; Arnold *et al.*, 1999,

2001; Allendorf *et al.*, 2001). This type of positive epistasis should result in increased levels of introgression at the interacting genes relative to a genome-wide average for unlinked marker loci. Introgression in hybrid zones may also be promoted by other mechanisms, such as additive effects of alleles at loci controlling fitness-related traits (for example, Rieseberg *et al.*, 2003; Martin *et al.*, 2006).

In Europe, most hybrid zones are the result of secondary contact between phylogenetically distinct populations (or 'phylogroups') that were isolated in allopatry during the glacial cycles of the Quaternary (Hewitt, 1996). The European rabbit (*Oryctolagus cuniculus*) originated in the Iberian Peninsula and is composed of two divergent phylogroups that correspond to the subspecies *O. c. algirus* and *O. c. cuniculus*. Analysis of mtDNA variation has revealed two highly divergent lineages with distinct geographic distributions consistent with an allopatric origin: type A (predominant in *O. c. algirus*) in the southwest Iberia and type B (predominant in *O. c. cuniculus*) in the northeast and the rest of colonized areas (Biju-Duval *et al.*, 1991; Monnerot *et al.*, 1994; Branco *et al.*, 2000). Levels of divergence at mitochondrial and X-linked loci suggest that divergence between the two groups dates back to the late Pleistocene, 2 million years BP (Biju-Duval *et al.*, 1991; Geraldès *et al.*, 2006) and spatial patterns of variation in mtDNA, Y-chromosome and two centromeric X-chromosome markers reveal a narrow, sharply delineated zone of secondary contact between the two phylogroups in central Iberia (Branco *et al.*, 2002; Geraldès *et al.*, 2005,

Correspondence: Dr R Campos, CIBIO-UP, Centro de Investigação em Biodiversidade e Recursos Genéticos, Campus Agrário de Vairão, Vairão 4485-661, Portugal.

E-mail: ritacampos@mail.icav.up.pt

Received 25 May 2007; revised 7 February 2008; accepted 17 March 2008; published online 30 April 2008

2006). In contrast to the patterns revealed by uniparentally inherited, haploid markers, multilocus surveys of nuclear-encoded allozyme variation have suggested a more recent divergence time of 275 000–550 000 years BP and also reveal a much broader zone of admixture (Ferrand and Branco, 2007).

In comparisons between *O. c. algirus* and *O. c. cuniculus*, surveys of allozyme variation at 21 polymorphic loci revealed a surprisingly large range of divergence levels (mean  $F_{ST}$  = 0.16, range 0–0.54; Branco, 2000; Campos *et al.*, 2007). This broad range of  $F_{ST}$  values presumably reflects stochastic variation due to drift during the period of allopatric divergence, as well as locus-specific differences in ongoing rates of introgression across the contact zone. Interestingly, the two loci that fell at opposite ends of the distribution of  $F_{ST}$  values, haemoglobin  $\alpha$ -chain (HBA) ( $F_{ST}$  = 0.54) and haemoglobin  $\beta$ -chain (HBB) ( $F_{ST}$  close to zero; Campos *et al.*, 2007), encode interacting subunits of the same multimeric protein. Specifically, the HBA and HBB genes encode the  $\alpha$ - and  $\beta$ -chain subunits of the tetrameric haemoglobin protein. In rabbits, as in all other amniote vertebrates, the  $\alpha$ - and  $\beta$ -globin genes are located on different chromosomes (Xu and Hardison, 1989, 1991). The unusually high level of differentiation at HBA is attributable to the fact that the two main protein electromorphs exhibit pronounced allele frequency differences between the two subspecies: the HBA\*1 allele is present at high frequency in *O. c. cuniculus* to the north, and the alternative HBA\*2 allele is present at high frequency in *O. c. algirus* to the south. Differences in electrophoretic mobility of the two HBA alleles are attributable to a set of three amino acid polymorphisms that are in complete linkage disequilibrium with one another: HBA\*1 is defined by the three-site amino acid haplotype,  $\alpha 29^{Val}/48^{Phe}/49^{Thr}$  and HBA\*2 is defined by the alternative haplotype,  $\alpha 29^{Leu}/48^{Leu}/49^{Ser}$  (Hunter and Munro, 1969; Hardison *et al.*, 1991; N Ferrand, unpublished).

The unusually low level of differentiation at HBB is attributable to the fact that the two main protein alleles, HBB\*1 and HBB\*2, are present at frequencies of  $\sim 0.10$  and  $\sim 0.90$ , respectively, across the entire range of the species. This uniform pattern of allele frequency variation is all the more remarkable given the pronounced level of genetic subdivision revealed by other unlinked markers (Branco *et al.*, 2000; Geraldes *et al.*, 2005, 2006; Ferrand and Branco, 2007). Similar to the case with HBA, differences in electrophoretic mobility of the two HBB alleles are attributable to a set of four amino acid polymorphisms that are in complete linkage disequilibrium with one another: HBB\*1 is defined by the four-site amino acid haplotype,  $\beta 52^{His}/56^{Ser}/76^{Asn}/112^{Val}$  and HBB\*2 is defined by the alternative haplotype,  $\beta 52^{Asn}/56^{Asn}/76^{Ser}/112^{Ile}$  (Galizzi, 1970; Bricker and Garrick, 1974; Campos *et al.*, 2007). The fact that the two major alleles at both HBA and HBB are distinguished by three and four replacement substitutions, respectively, suggests that both polymorphisms are fairly old and are probably related to allopatric divergence during the Quaternary.

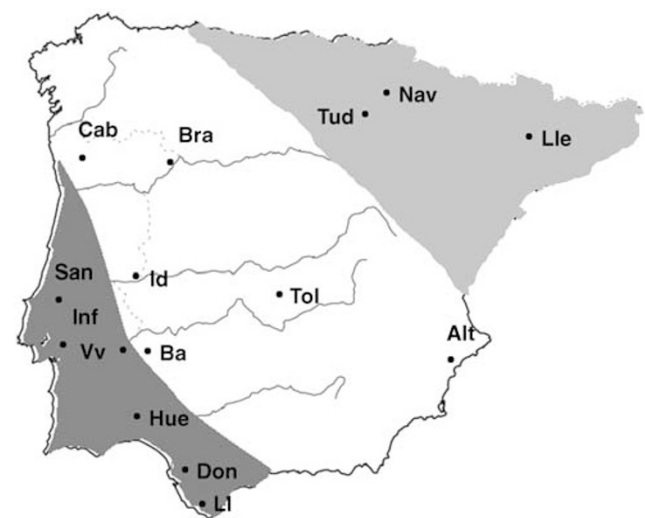
The objective of this study was to assess whether it is necessary to invoke some form of natural selection to account for the contrasting spatial patterns of allele frequency variation at the HBA and HBB genes. We also assessed whether epistasis between the two unlinked

genes may help explain the observed patterns of geographic variation. We conducted a multilocus analysis using a set of 25 polymorphic allozyme loci and a set of 15 population samples from the Iberian Peninsula. Our results suggest that the contrasting levels of spatial differentiation at these two globin genes cannot be reconciled under a neutral model of population structure, and that patterns of variation in the  $\alpha$ - and  $\beta$ -chain subunits of rabbit haemoglobin have been shaped by different modes of selection.

## Materials and methods

### Sampling

The data used in this study resulted from a long-standing project on genetic characterization of wild and domestic European rabbit populations developed in CIBIO laboratories. We analyzed data from a total of 324 wild rabbits that were collected from 15 localities across the Iberian Peninsula (Figure 1). The sampling localities cover the range of both subspecies as well as the zone of secondary contact between them. Previous allozyme surveys demonstrated that these samples cluster into clearly delineated groups that are referable to the subspecies '*algirus*' in southwest Iberia (Doñana, Huelva, Infantado, Las Lomas, Santarém and Vila Viçosa), and the subspecies '*cuniculus*' in northeast Iberia (Lérida, Navarra and Tudela; Ferrand and Branco, 2007). These two subspecies also represent the putative ancestral populations that were isolated during the glacial periods. A third group, '*hybrid*', represents the populations from the zone of secondary contact in central Iberia (Bragança, Idanha, Badajoz, Cabreira, Toledo and Alicante).



**Figure 1** Distribution map of the three population groups used in the multilocus analysis of 25 allozyme loci in the European rabbit. Group *algirus* (dark grey shaded area): Santarém (San), Infantado (Inf), Vila Viçosa (Vv), Huelva (Hue), Doñana (Don) and Las Lomas (Li). Group *cuniculus* (light grey shaded area): Navarra (Nav), Tudela (Tud), Lérida (Lle). Populations from the admixture area (white area): Cabreira (Cab), Bragança (Bra), Idanha (Id), Badajoz (Ba), Toledo (Tol) and Alicante (Alt).

**Table 1** Protein locus, chromosome location and separation systems used in the survey of outlier loci in natural populations of the European rabbit

| Locus | Location <sup>a</sup> | Separation system | Reference <sup>b</sup>       | Na             |                  |                  | Heterozygosity | F <sub>ST</sub> |
|-------|-----------------------|-------------------|------------------------------|----------------|------------------|------------------|----------------|-----------------|
|       |                       |                   |                              | <i>algirus</i> | <i>Admixture</i> | <i>cuniculus</i> |                |                 |
| ACP3  | —                     | IF                | Branco and Ferrand, 1998     | 3 (147)        | 3 (93)           | 2 (10)           | 0.100632       | 0.136266        |
| ADA   | —                     | SGE               | Amorim <i>et al.</i> , 1982  | 5 (147)        | 3 (125)          | 4 (64)           | 0.531470       | 0.216774        |
| ALB   | OCU15q23              | IF                | Ferrand and Rocha, 1992      | 3 (147)        | 3 (125)          | 3 (42)           | 0.529559       | 0.126363        |
| AT3   | —                     | HIF               | Branco and Ferrand, 2002     | 5 (86)         | 3 (88)           | 4 (49)           | 0.233733       | 0.274642        |
| BF    | —                     | HIF               | Branco <i>et al.</i> , 1998  | 6 (147)        | 6 (101)          | 5 (52)           | 0.728147       | 0.200965        |
| CAI   | —                     | AGE               | Branco and Ferrand, 2003     | 2 (147)        | 2 (107)          | 2 (36)           | 0.010566       | -0.001490       |
| CAII  | —                     | AGE               | Branco and Ferrand, 2003     | 2 (147)        | 2 (121)          | 2 (35)           | 0.505758       | 0.423347        |
| DIA1  | —                     | SGE               | Unpublished                  | 2 (147)        | 3 (116)          | 3 (51)           | 0.025993       | 0.002150        |
| GALT  | OCU1                  | SGE               | Ferrand, 1995                | 6 (126)        | 6 (118)          | 4 (43)           | 0.448057       | 0.116883        |
| GC    | OCU15q23dist          | IF                | Ferrand, 1995                | 7 (147)        | 6 (124)          | 1 (63)           | 0.250974       | 0.093434        |
| GPI   | —                     | IF                | Branco, 2000                 | 4 (147)        | 2 (125)          | 1 (42)           | 0.041890       | 0.049660        |
| HBA   | OCU6q12               | SGE               | Ferrand, 1990                | 3 (147)        | 3 (119)          | 5 (42)           | 0.660677       | 0.465294        |
| HBB   | OCU1q14-q21           | SGE               | Ferrand, 1989                | 2 (147)        | 2 (125)          | 2 (46)           | 0.162794       | -0.015135       |
| HP    | OCU5 <sup>c</sup>     | HIF               | Branco and Ferrand, 2002     | 3 (45)         | 2 (1)            | 3 (39)           | 0.583895       | 0.295764        |
| HPX   | —                     | IF                | Branco and Ferrand, 2002     | 7 (147)        | 7 (110)          | 4 (41)           | 0.568341       | 0.064845        |
| IDH   | —                     | IF                | Unpublished                  | 2 (126)        | 1 (105)          | 1 (3)            | 0.007326       | 0.012987        |
| MPI   | —                     | SGE               | Vieira and Ferrand, 1995     | 7 (125)        | 4 (81)           | 1 (12)           | 0.350684       | 0.065971        |
| NP    | OCU17                 | SGE               | Amorim <i>et al.</i> , 1982  | 3 (124)        | 3 (125)          | 1 (66)           | 0.122675       | 0.030835        |
| PEPA  | —                     | SGE               | Branco <i>et al.</i> , 1999  | 5 (147)        | 2 (124)          | 1 (59)           | 0.443755       | 0.298558        |
| PEPB  | OCU17                 | SGE               | Branco <i>et al.</i> , 1999  | 4 (147)        | 3 (125)          | 2 (64)           | 0.209843       | 0.094504        |
| PEPC  | —                     | SGE               | Branco <i>et al.</i> , 1999  | 7 (147)        | 7 (125)          | 3 (63)           | 0.557989       | 0.095157        |
| PEPD  | —                     | SGE               | Branco <i>et al.</i> , 1999  | 6 (147)        | 4 (125)          | 2 (64)           | 0.415364       | 0.226517        |
| PGD   | —                     | SGE               | Amorim <i>et al.</i> , 1982  | 4 (147)        | 3 (125)          | 4 (70)           | 0.522833       | 0.199744        |
| PGM2  | —                     | SGE               | Branco, 2000                 | 3 (84)         | 2 (95)           | 2 (33)           | 0.208783       | 0.111088        |
| TF    | OCU14                 | AGE               | Ferrand <i>et al.</i> , 1988 | 5 (147)        | 6 (125)          | 2 (67)           | 0.378303       | 0.138766        |

Abbreviations: Na, number of alleles; IF, isoelectric focusing; HIF, hybrid isoelectric focusing; SGE, starch gel electrophoresis; AGE, agarose gel electrophoresis.

Population diversity patterns are given by the total number of alleles observed per locus per population group (defined in Materials and methods; sample size is in brackets). Heterozygosity and F<sub>ST</sub> are shown for *algirus* and *cuniculus* only (median sample size per locality = 42 individuals).

<sup>a</sup><http://dga.jouy.inra.fr/cgi-bin/lgbc/main.pl?BASE=rabbit>.

<sup>b</sup>The references indicated refer only to methods of electrophoretic separation.

<sup>c</sup>Korstanje *et al.*, 2003.

### Electrophoretic analysis

We collected polymorphism data for a total of 25 loci (Table 1). Thirteen loci were analyzed using starch gel electrophoresis: adenosine deaminase (ADA), diaphorase I, galactose-1-phosphate uridylyltransferase (GALT), haemoglobin  $\alpha$ -chain (HBA), haemoglobin  $\beta$ -chain (HBB), mannose phosphate isomerase (MPI), nucleoside phosphorylase (NP), peptidase A (PEPA), peptidase B (PEPB), peptidase C (PEPC), peptidase D (PEPD), phosphogluconate dehydrogenase (PGD) and phosphoglucomutase 2 (PGM2). Three loci were analyzed using agarose gel electrophoresis: carbonic anhydrase I (CAI) and II (CAII) and transferrin (TF). Six loci were screened using isoelectric focusing: acid phosphatase 3 (ACP3), albumin (ALB), vitamin D-binding protein (GC), glucose phosphate isomerase, haemopexin (HPX) and isocitrate dehydrogenase (IDH). Finally, hybrid isoelectric focusing was used in the analysis of three additional loci: antithrombin III (AT3), properdin factor B (BF) and haptoglobin (HP) (see Table 1 and references therein for details).

Codominant segregation of alleles has been verified for all loci used in this analysis (Ferrand, 1995; unpublished results). Using the Weir and Cockerham (1984) estimator of the inbreeding coefficient  $f$  ( $=F_{IS}$ ), we used a randomization test to determine whether observed genotypic proportions deviated from Hardy–Weinberg expectations. To test for linkage disequilibrium (LD)

between each pairwise combination of loci, we used a contingency table test on diploid genotypes based on the log-likelihood ratio G-statistic. Confidence limits and probability values for  $f$  were calculated by permutating alleles within individual samples 1000 times. Bonferroni corrections were performed on the results as appropriate (Rice, 1989). All analyses were implemented in the Fstat v2.9.3.2 program (Goudet, 2001).

Additionally, we specifically tested the null hypothesis that genotypes at HBA and HBB are independent of one another. Using the program GenePop web version 3.1c (updated from version 1.2; Raymond and Rousset, 1995), we tested for evidence of pairwise LD by performing a Fisher's exact test on contingency tables of diploid genotypes.

### Simulation analysis

For the '*algirus*' and the '*cuniculus*' groups, single-locus measures of genetic differentiation were obtained using the  $\theta$  ( $=F_{ST}$ ) estimator of Weir and Cockerham (1984). For each data set, locus-specific departures from neutral expectations were tested by comparing observed  $F_{ST}$  values (conditioned on heterozygosity) to a null distribution generated by a coalescent-based simulation model (Beaumont and Nichols, 1996; Beaumont and Balding, 2004). To generate the null distributions, we used a nonequilibrium model of population structure

that incorporated the history of divergence between the two subspecies as well as internal subdivision within each subspecies. Specifically, we considered a scenario in which two populations (= *O. c. cuniculus* and *O. c. algirus*) diverged from a panmictic ancestral population at a specified time in the past. We used an iterative fitting procedure to generate the expected neutral distribution for the total sample. In each set of iterations, divergence times were varied between  $2.5 \times 10^5$  and  $5 \times 10^6$  years BP to produce a null distribution in which equal numbers of loci fell above and below the median quantile. This range of divergence times for the two subspecies spans the range of empirical estimates based on mtDNA (Biju-Val et al., 1991; Branco et al., 2000), X-linked markers (Geraldès et al., 2006) and nuclear-encoded allozymes (Ferrand and Branco, 2007). In the simulation model, each of the two subspecies was subdivided into a network of equal-sized demes that were interconnected by ongoing migration. Within each of the two subspecies, the rate of migration among demes was set equal to the value that produced the observed level of differentiation under an island model of population structure. This model allowed us to account for the effects of ongoing migration in shaping patterns of allele frequency variation within each subspecies, which was not possible, for example, using the outlier-detection approach of Vitalis et al. (2001), which is based on a drift and divergence model of population structure. We also modelled the effects of bottlenecks following the initial underling of the ancestral population into two refugial isolates. At the time of the initial divergence, each of the two populations underwent a 5- to 10-fold reduction in population size. The duration of the bottleneck was varied between 10 000 to 500 000 years, and was followed by a stepwise increase to the contemporary population size. Coalescent simulations were conducted under the infinite alleles model (IAM) as well as the stepwise mutation model (SMM). In each set of iterations, sample sizes were set equal to the median of actual sample sizes in the specific data set under consideration. Coalescent simulations were used to generate a total of 50 000 paired values of  $F_{ST}$  and  $H$ , which was then used to compute the 0.975, 0.50 and 0.025 quantiles of the conditional distribution (Beaumont and Nichols, 1996; Storz and Nachman, 2003; Storz and Dubach, 2004). Loci with  $F_{ST}$  values that exceeded the 0.975 quantile of the null distribution were considered candidates for spatially varying selection (that is, the observed heterogeneity in allele frequencies exceeded neutral expectations). Conversely, loci with  $F_{ST}$  values that fell below the 0.025 quantile of the distribution were considered candidates for spatially uniform selection (that is, the observed heterogeneity in allele frequencies fell below neutral expectations).

## Results and discussion

### Allozyme polymorphism

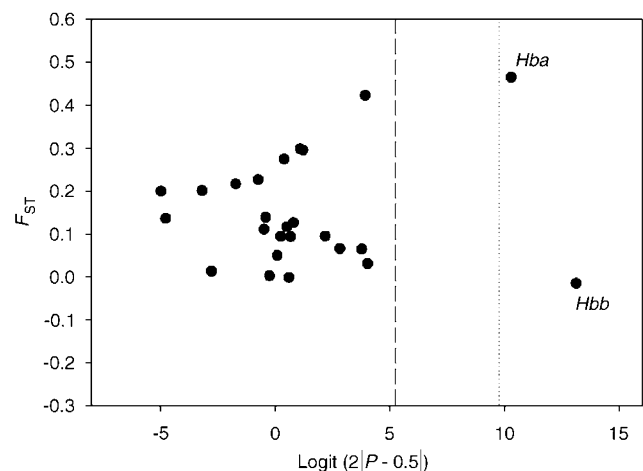
Diversity measured as the number of alleles was consistently higher in *algirus* relative to *cuniculus* (Table 1). This same discrepancy in diversity levels was observed for mtDNA, X-chromosome markers and nuclear-encoded allozymes (Branco et al., 2000; Geraldès et al., 2006; Ferrand and Branco, 2007), and suggests that *algirus* descended from a refugial population that

maintained a larger effective size during the Quaternary glacial cycles.

Samples from all localities conformed to Hardy-Weinberg genotypic proportions, with the exception of the Alicante sample ( $F_{IS} = 0.148$ ;  $P = 0.002$ ). We detected a within-sample deficit of heterozygotes at a total of nine loci (ADA in Cabreira, GALT in Lérida, HBA in Vila Viçosa, HBB in Alicante, HPX in Huelva and Toledo, MPI in Huelva, PEPA in Las Lomas and Idanha, PEPD in Santarém and PGD in Toledo). We detected a within-sample excess of heterozygotes in a single locus  $\times$  locality combination (AT3 in Tudela). A single pair of loci exhibited a statistically significant level of pairwise linkage disequilibrium in the total sample (CAII  $\times$  PEPC,  $P = 0.01$ ). However, significant linkage disequilibrium between these two loci was only detected in two of the individual samples: Alicante and Toledo. We also detected significant linkage disequilibrium in the following locus  $\times$  locality combinations: ALB  $\times$  AT3, BF  $\times$  TF and GALT  $\times$  PEPA in Badajoz, NP  $\times$  PEPD in Infantado, HBA  $\times$  MPI, GC  $\times$  PEPA and ADA  $\times$  MPI in Las Lomas, HBB  $\times$  PEPC in Navarra and BF  $\times$  PEPC and PEPB  $\times$  PEPD in Santarém.

### Levels and patterns of population differentiation

Across the Iberian Peninsula, the weighted mean  $F_{ST}$  value for all 25 loci was 0.200. All loci exhibited a good fit to the neutral model of population structure, with the exception of HBA and HBB (Figure 2). The HBA locus exhibited the highest level of differentiation ( $F_{ST} = 0.465$ ; Table 1) and exceeded the upper 0.975 quantile of the null distribution ( $P = 0.00033$  under the IAM and 0.00028 under the SMM). By contrast, the HBB locus exhibited the lowest level of differentiation ( $F_{ST} \approx 0$ ) and fell below the lower 0.025 quantile of the distribution ( $P < 0.00001$  under both the IAM and SMM). For both loci, observed departures from neutral expectations remained statistically significant under both mutation models even at a Bonferroni-corrected  $\alpha$ -level of 0.002 (= 0.05/25 loci). Locus-specific  $P$ -values were nearly identical for simulations under the IAM and SMM. Hereafter, we restrict the discussion to results obtained under the IAM. HBA and



**Figure 2** Distribution of transformed  $P$ -values for all 25 loci under a neutral model of population structure. The left vertical bar depicts the 0.05 cutoff. The right vertical bar depicts the Bonferroni-adjusted cutoff (= 0.05/25). Locus-specific  $F_{st}$  values are given in Table 1.

HBB also exhibited statistically significant departures from neutral expectations in the simulation models that incorporated population bottlenecks following the initial divergence of northern and southern phylogroups ( $P < 0.0001$  in all cases). In simulation models that spanned the full range of bottleneck durations and population size changes (see Materials and methods), the observed  $F_{ST}$  for HBA consistently exceeded the upper 0.975 quantile of the null distribution and the observed  $F_{ST}$  for HBB consistently fell below the lower 0.025 quantile.

The two globin genes, HBA and HBB, also exhibited strongly contrasting patterns of spatial differentiation within the range of *O. c. cuniculus* in northern Iberia. In the comparisons among the three samples from Lérida, Navarra and Tudela, the weighted mean  $F_{ST}$  for all 25 loci was 0.212. Of these 25 loci, HBA exhibited the highest  $F_{ST}$  value (0.646) and HBB exhibited one of the lowest values ( $\approx 0$ ). By contrast, in the comparison among the six population samples of *O. c. algirus* from southern Iberia, the weighted mean  $F_{ST}$  for all 25 loci was 0.083 and the range was  $-0.010$  to  $0.160$ .  $F_{ST}$  values for HBA and HBB (0.065 and 0.017, respectively) were not especially high or low relative to the multilocus average.

#### Functional significance of haemoglobin polymorphism

The red blood cells of *O. c. algirus* primarily contain (HBA\*2)<sub>2</sub>(HBB\*2)<sub>2</sub> haemoglobin tetramers, whereas the red blood cells of *O. c. cuniculus* primarily contain (HBA\*1)<sub>2</sub>(HBB\*2)<sub>2</sub> tetramers. With reference to the amino acid replacements underlying each HBA and HBB variant,  $\beta 112$  is the only residue located in an inter-subunit contact surface. Specifically, the  $\beta 112$  residue is located within a 6 Å radius of  $\alpha 117^{Phe}$  and  $\alpha 122^{His}$ . Relative to  $\beta 112^{Val}$ , the  $\beta 112^{Ile}$  mutant is slightly less bulky and slightly more hydrophobic. Mutations in these intersubunit contact surfaces may affect the transition in quaternary structure between the oxy- and deoxy-state of the HB tetramer. However, a more detailed functional analysis will be required to determine whether the different  $\beta 112$  mutants have any effect on the equilibrium ratio of oxy- and deoxy-haemoglobin in the rabbit red blood cells.

The fact that HBA and HBB encode polypeptides that are assembled into the same multimeric protein suggests the possibility that the non-neutral patterns of allele frequency variation observed at each of the two genes may reflect some form of epistatic selection. Since the allosteric mechanism of haemoglobin oxygenation and deoxygenation requires coordinated shifts between the two  $\alpha/\beta$  dimers that comprise each tetramer, one possibility is that one allelic type of  $\alpha$ -chain polypeptide only functions well with a particular allelic type of  $\beta$ -chain polypeptide, and vice versa. If this were the case we might expect to see a highly non-random pattern of association between alleles at each of the two genes. In principal, if selection were sufficiently strong, we might observe repulsion-phase LD between the incompatible HBA and HBB alleles, and coupling-phase LD between the compatible HBA and HBB alleles. We would also expect to see high levels of differentiation at both loci, reflecting the fact that *algirus*-specific HBA alleles only function properly in combination with *algirus*-specific HBB alleles and likewise for *cuniculus*-specific alleles.

**Table 2** Probability values for genotypic disequilibrium between HBA and HBB in each population and across all populations

| Pop                           | P-value (s.e.) |
|-------------------------------|----------------|
| Alt                           | 0.262 (0.007)  |
| Ba                            | 1.000 (0.000)  |
| Bra                           | 0.783 (0.003)  |
| Cab                           | 1.000 (0.000)  |
| Don                           | 0.482 (0.002)  |
| Hue                           | 0.066 (0.001)  |
| Id                            | —              |
| Inf                           | 0.458 (0.004)  |
| Ll                            | —              |
| Lle                           | 1.000 (0.000)  |
| Nav                           | 0.252 (0.007)  |
| Tud                           | —              |
| San                           | 0.575 (0.004)  |
| Tol                           | 1.000 (0.000)  |
| Vv                            | 0.092 (0.003)  |
| <i>Across all populations</i> |                |
| $\chi^2$                      | P-value        |
| 20.245                        | 0.683          |

s.e. are given in brackets.

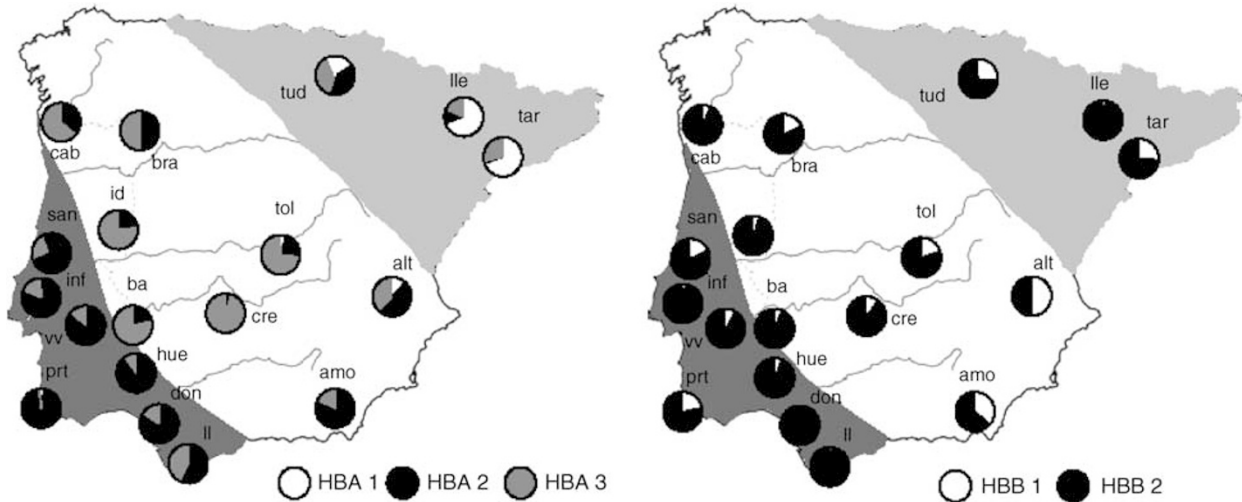
Interestingly, a case of epistatic selection maintaining linkage-disequilibrium between two physically unlinked loci in the European rabbit was already described (van der Loo *et al.*, 1987, 1996; van der Loo, 1993). However, since we observed no statistically significant LD between the HBA and HBB genes (Table 2), and since only HBA exhibits elevated differentiation between the two subspecies, this type of epistasis does not appear to explain the observed patterns of variation. Overall, patterns of variation at the two loci do not appear to conform to any simple model of two-locus epistatic selection.

#### Contrasting patterns of differentiation in HBA and HBB

The high level of differentiation at HBA reflects the fact that each allele is either restricted or more common within each population group (HBA\*1 in *cuniculus* and HBA\*2 in *algirus* populations, respectively; Figure 3; see also Campos *et al.*, 2007; Ferrand and Branco, 2007). HBA\*3 is the most common allele in the contact zone and has no association with mtDNA haplotype background (Campos *et al.*, 2007). The prevalence of this allele in the zone of admixture suggests that it is a novel recombinant that originated by crossing-over between the parental alleles, HBA\*1 and HBA\*2. The creation of novel alleles in hybrid zones has been documented in several species (for example, Woodruff, 1989; Bradley *et al.*, 1993; Godinho *et al.*, 2006; Sequeira, 2006) which makes the hybrid origin of these alleles a plausible hypothesis, given its geographical distribution.

HBB is characterized by lower-than-expected  $F_{ST}$  values, which are nearly zero in subspecies comparisons (Campos *et al.*, 2007). This low level of differentiation reflects a surprisingly uniform distribution of allele frequencies across the species range (Figure 3), especially considering that unlinked markers exhibit high levels of genetic subdivision (Branco *et al.*, 2000; Geraldès *et al.*, 2005).

Both globin genes exhibit patterns of variation that are consistent with completely different modes of natural



**Figure 3** Geographical distribution of the alleles of HBA (left map) and HBB (right map) detected by electrophoresis in wild rabbit populations from the Iberian Peninsula. Populations and the three geographic groups are as described in Figure 1 except for Portimão (prt) from the *algius* group (populations from the dark grey shaded area), Ciudad Real (cre) and Amoladeras (amo) from the admixture area (white area) and Tarragona (tar) from the *cuniculus* group (populations from the light grey shaded area). In the HBA map, only the three common alleles are represented (see text for details on the other alleles). Adapted from Campos *et al.*, 2007.

selection. However, in each case it is possible that the true target of selection is actually a different, closely linked locus. HBA and HBB are members of the  $\alpha$ - and  $\beta$ -globin gene families, respectively. Each family consists of multiple genes that control haemoglobin synthesis during different stages of development (Hardison, 2001). In rabbits, the  $\alpha$ -globin gene family spans approximately 60 kb on chromosome 6 (Hardison *et al.*, 1991; Xu and Hardison, 1991) and the  $\beta$ -globin gene family spans 45 kb on chromosome 1 (Margot *et al.*, 1989; Xu and Hardison, 1989). If the observed patterns of variation at the HBA and HBB genes are attributable to hitchhiking associated with selection at linked loci, in each case the true target of selection is likely to be another closely linked globin gene.

In mammals and other vertebrates, haemoglobin polymorphism plays a well-documented role in adaptation to hypoxic environments (Poyart *et al.*, 1992; Weber and Fago, 2004; Storz, 2007; Storz *et al.*, 2007), and in humans, amino acid and deletion polymorphisms in the  $\alpha$ - and  $\beta$ -globin genes have been implicated in resistance to malaria (for example, Agarwal *et al.*, 2000; Ohashi *et al.*, 2004; Kwiatkowski, 2005; Williams *et al.*, 2005). In the case of the European rabbit, patterns of HBA and HBB variation across the Iberian Peninsula do not correlate with altitude or any obvious environmental factors. It thus seems unlikely that the HB polymorphism is involved in some form of adaptation to the abiotic environment. Another possibility is that the observed patterns of  $\alpha$ - and  $\beta$ -globin polymorphism reflect selection for resistance to blood-borne pathogens, as allelic variation in haemoglobin function is known to play an important role in modulating the reduction–oxidation status of red blood cells. In humans, despite the long term knowledge of the selective agent for the most common haemoglobinopathies in sub-Saharan Africa, the existence of negative epistasis between both conditions was reported only recently (Williams *et al.*, 2005). It may therefore prove difficult to unravel the causes of the observed patterns of allele frequency variation at both

globin loci in the European rabbit. Nevertheless, while the mechanisms of selection acting on this system remain to be discovered and experimental physiological and expression assays are yet to be done, the intriguing patterns of variation seen at the rabbit globin loci warrant further study to elucidate the possible causes of fitness variation.

## Acknowledgements

This work was partially financed by project Grant POCTI/BSE/34151/2000 from Fundação para a Ciência e a Tecnologia (FCT). R Campos was supported by PhD Grant SFRH/BD/1409/2000 from FCT. We thank H Moriyama for help in the 3D structural analysis of the amino-acid replacements described for HBA and HBB and W van der Loo, R Nichols and three anonymous referees for helpful comments and suggestion on the article.

## References

- Agarwal A, Guindo A, Cissoko Y, Taylor JG, Coulibaly D, Kone A *et al.* (2000). Hemoglobin C associated with protection from severe malaria in the Dogon of Mali, a West African population with a low prevalence of hemoglobin S. *Blood* **96**: 2358–2363.
- Allendorf FW, Leary RF, Spruell P, Wenburg JK (2001). The problems with hybrids: setting conservation guidelines. *Trends Ecol Evol* **16**: 613–622.
- Amorim A, Kömpf J, Schunter F, Ritter H (1982). Aminolevulinic dehydratase (E.C.4.2.1.24): linkage analysis. *Hum Genet* **61**: 48–49.
- Arnold ML (1997). *Natural Hybridisation and Evolution*. Oxford University Press: New York.
- Arnold ML, Bulger MR, Burke JM, Hempel AL, Williams JH (1999). Natural hybridisation: how low can you go and still be important? *Ecology* **80**: 371–381.
- Arnold ML, Kentner EK, Johnston JA, Cornman S, Bouck AC (2001). Natural hybridisation and fitness. *Taxon* **50**: 93–104.

- Barton NH, Bengtsson BO (1986). The barrier to genetic exchange between hybridizing populations. *Heredity* **56**: 357–376.
- Barton NH, Hewitt GM (1985). Analysis of hybrid zones. *Annu Rev Ecol Syst* **16**: 113–148.
- Bateson W (1909). Heredity and variation in modern lights. In: Seward AC (ed). *Darwin and Modern Science*. Cambridge University Press: Cambridge, UK, pp 85–101.
- Beaumont MA, Balding DJ (2004). Identifying adaptive genetic divergence among populations from genome scans. *Mol Ecol* **13**: 969–980.
- Beaumont MA, Nichols RA (1996). Evaluating loci for use in the genetic analysis of population structure. *Proc Roy Soc London Ser B* **263**: 1619–1626.
- Biju-Duval C, Ennafaa H, Dennebouy N, Monnerot M, Mignotte F, Soriguer RC *et al.* (1991). Mitochondrial DNA evolution in lagomorphs: origin of systematic heteroplasmy and organization of diversity in European rabbits. *J Mol Evol* **33**: 92–102.
- Bradley RD, Bull JJ, Johnson AD, Hillis DM (1993). Origin of a novel allele in a mammalian hybrid zone. *Proc Natl Acad Sci USA* **90**: 8939–8941.
- Branco M (2000). *Genetic Structure of the European Rabbit (Oryctolagus Cuniculus) Populations in the Iberian Peninsula. Isolation, Differentiation of two Evolutionary Lineages, Geographic Expansion and Secondary Contact*. PhD Thesis. University of Porto: Portugal.
- Branco M, Ferrand N (1998). Genetic polymorphism of rabbit (*Oryctolagus cuniculus*) tissue acid phosphatases (ACP2 and ACP3). *Comp Biochem Physiol B, Biochem Mol Biol* **120**: 405–409.
- Branco M, Ferrand N (2002). Genetic polymorphism of antithrombin III, haptoglobin, and haemopexin in wild and domestic European rabbits. *Biochem Genet* **40**: 387–393.
- Branco M, Ferrand N (2003). Biochemical and population genetics of the rabbit, *Oryctolagus cuniculus*, carbonic anhydrases I and II, from the Iberian Peninsula and France. *Biochem Genet* **41**: 391–404.
- Branco M, Ferrand N, Monnerot M (2000). Phylogeography of the European rabbit (*Oryctolagus cuniculus*) in the Iberian Peninsula inferred from RFLP analysis of the cytochrome b gene. *Heredity* **85**: 307–317.
- Branco M, Ferrand N, Monnerot M, 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**: 792–803.
- Branco M, Lopes G, Ferrand N (1998). Genetic polymorphism of properdin factor B (BF) in domestic rabbit. *Anim Genet* **29**: 135–137.
- Branco M, Machado JC, 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**: 237–249.
- Bricker J, Garrick M (1974). An isoleucine-valine substitution in the  $\beta$ -chain of rabbit hemoglobin. *Biochim Biophys Acta* **351**: 437–441.
- Campos R, Branco M, Weiss S, Ferrand N (2007). Patterns of hemoglobin polymorphism [ $\alpha$ -globin (HBA) and  $\beta$ -globin (HBB)] across the contact zone of two distinct phylogeographic lineages of the European rabbit (*Oryctolagus cuniculus*). In: Weiss S, Ferrand N (eds). *Phylogeography of European Refugia*. Springer: Netherlands, pp 237–255.
- Dobzhansky TH (1936). Studies on hybrid sterility. II. Localization of sterility factors in *Drosophila pseudo-obscura* hybrids. *Genetics* **21**: 113–135.
- Dopman EB, Perez L, Bogdanowicz SM, Harrison RG (2005). Consequences of reproductive barriers for genealogical discordance in the European corn borer. *Proc Natl Acad Sci USA* **102**: 14706–14711.
- Ferrand N (1989). Biochemical and genetic studies on rabbit hemoglobin. I. Electrophoretic polymorphism of the  $\beta$ -chain. *Biochem Genet* **27**: 673–678.
- Ferrand N (1990). Biochemical and genetic studies on rabbit hemoglobin. II. Electrophoretic polymorphism of the  $\alpha$ -chain. *Biochem Genet* **28**: 117–122.
- Ferrand N (1995). *Variaco gentica de proteinas em populaoes de coelho (Oryctolagus cuniculus). Anlise da diferenciao subespecfica, subestruturaco, expanso geogrfica e domesticao* PhD Thesis. University of Porto: Porto, Portugal.
- Ferrand N, Branco M (2007). The evolutionary history of the European rabbit (*Oryctolagus cuniculus*): Major patterns of population differentiation and geographic expansion inferred from protein polymorphism. In: Weiss S, Ferrand N (eds). *Phylogeography of European Refugia*. Springer: Netherlands, pp 207–235.
- Ferrand N, Carvalho G, Amorim A (1988). Transferrin (Tf) polymorphism in wild rabbit, *Oryctolagus cuniculus*. *Anim Genet* **19**: 295–300.
- Ferrand N, Rocha J (1992). Demonstration of serum Albumin (ALB) polymorphism in wild rabbits, *Oryctolagus cuniculus*, by means of isoelectric focusing. *Anim Genet* **23**: 275–278.
- Galizzi A (1970). The amino acid sequence of the  $\beta$ -chains of rabbit hemoglobin. *Eur J Biochem* **17**: 49–55.
- Geraldes A, Ferrand N, Nachman MW (2006). Contrasting patterns of introgression at X-linked loci across the hybrid zone between subspecies of the European rabbit (*Oryctolagus cuniculus*). *Genetics* **173**: 919–933.
- Geraldes A, Rogel-Gaillard C, Ferrand N (2005). High levels of nucleotide diversity in the European rabbit (*Oryctolagus cuniculus*) SRY gene. *Anim Genet* **36**: 349–351.
- Godinho R, Mendonca B, Crespo EG, Ferrand N (2006). Genealogy of the nuclear beta-fibrinogen locus in a highly structured lizard species: comparison with mtDNA and evidence for intragenic recombination in the hybrid zone. *Heredity* **96**: 4454–4463.
- Goudet J (2001). FSTAT, a Program to Estimate and Test Gene Diversities and Fixation Indices Version 2.9.3. Available from <http://www.unil.ch/izea/software/fstat.html>.
- Hardison R (2001). Organization, evolution, and regulation of the globin genes. In: Steinberg MH, Forget BG, Higgs DR, Nagel RL (eds). *Disorders of Hemoglobin: Genetics, Pathophysiology, and Clinical Management*. Cambridge University Press: Cambridge, UK, pp 95–116.
- Hardison R, Krane D, Vandenberg D, Cheng JF, Mansberger J, Taddie J *et al.* (1991). Sequence and comparative analysis of the rabbit  $\alpha$ -like globin gene cluster reveals a rapid mode of evolution in a G+C-rich region of mammalian genomes. *J Mol Evol* **222**: 233–249.
- Hewitt GM (1996). Some genetic consequences of ice ages, and their role in divergence and speciation. *Biol J Linn Soc Lond* **58**: 247–276.
- Hunter T, Munro A (1969). Allelic variants in the amino acid sequence of the  $\alpha$ -chain of rabbit hemoglobin. *Nature* **223**: 1270–1272.
- Kwiatkowski DP (2005). How malaria has affected the human genome and what human genetics can teach us about malaria. *Am J Hum Genet* **77**: 171–192.
- Korstanje R, Gillissen GF, Versteeg SA, van Oost BA, Bosma AA, Rogel-Gaillard C *et al.* (2003). Mapping of rabbit microsatellite markers using chromosome-specific libraries. *J Hered* **94**: 161–169.
- Margot JB, Demers GW, Hardison RC (1989). Complete nucleotide sequence of the rabbit beta-like globin gene cluster. Analysis of intergenic sequences and comparison with the human beta-like globin gene cluster. *J Mol Evol* **205**: 15–40.
- Martin NH, Bouck AC, Arnold ML (2006). Detecting adaptive trait introgression between *Iris fulva* and *I. brevicaulis* in highly selective field conditions. *Genetics* **172**: 2481–2489.

- Martinsen GD, Whitham TG, Turek RJ, Keim P (2001). Hybrid populations selectively filter gene introgression between species. *Evolution* **55**: 1325–1335.
- Monnerot M, Vigne J-D, Biju-Duval C, Casane D, Callou C, Hardy C *et al.* (1994). Rabbit and man: genetic and historical approach. *Genet Sel Evol* **26** (suppl 1): 167s–182s.
- Muller HJ (1942). Isolating mechanisms, evolution and temperature. In: Dobzhansky T (ed). *Biological Symposia*. The Jaques Catell Press: Lancaster, PA, pp 71–125.
- Ohashi J, Naka I, Patarapotikul J, Hananantachai H, Brittenham G, Looareesuwan S *et al.* (2004). Extended linkage disequilibrium surrounding the hemoglobin E variant due to malarial selection. *Am J Hum Genet* **74**: 1198–1208.
- Orr HA, Masly JP, Presgraves DC (2004). Speciation genes. *Curr Opin Genet Dev* **14**: 675–679.
- Payseur BA, Hoekstra HE (2005). Signatures of reproductive isolation in patterns of single nucleotide diversity across inbred strains of mice. *Genetics* **171**: 1905–1916.
- Payseur BA, Krenz JG, Nachman MW (2004). Differential patterns of introgression across the X chromosome in a hybrid zone between two species of house mice. *Evolution* **58**: 2064–2078.
- Poyart C, Wajzman H, Kister J (1992). Molecular adaptation of hemoglobin function in mammals. *Respir Physiol* **90**: 3–17.
- Raymond M, Rousset F (1995). GENEPOP (version 1.2): Population genetic software for exact tests and ecumenicism. *J Hered* **80**: 248–249.
- Rice WR (1989). Analysing tables of statistical tests. *Evolution* **43**: 223–225.
- Rieseberg LH, Raymond O, Rosenthal DM, Lai Z, Livingstone K, Nakazato T *et al.* (2003). Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* **301**: 1211–1216.
- Rieseberg LH, Sinervo B, Linder R, Ungerer MC, Arias DM (1996). Role of gene interactions in hybrid speciation: Evidence from ancient and experimental hybrids. *Science* **272**: 741–745.
- Sequeira F (2006). *História evolutiva da salamandra-lusitânica, Chioglossal lusitanica. Filogenia, reconstrução dos padrões filogeográficos e análise dos processos de miscigenação na zona híbrida através do uso de múltiplos marcadores moleculares PhD Thesis.*, University of Porto: Porto, Portugal.
- Storz JF (2007). Hemoglobin function and physiological adaptation to hypoxia in high-altitude mammals. *J Mammal* **88**: 24–31.
- Storz JF, Dubach JM (2004). Natural selection drives altitudinal divergence at the Albumin locus in deer mice, *Peromyscus maniculatus*. *Evolution* **58**: 1342–1352.
- Storz JF, Nachman MW (2003). Natural selection on protein polymorphism in the rodent genus *Peromyscus*: evidence from interlocus contrasts. *Evolution* **57**: 2628–2635.
- Storz JF, Sabatino SJ, Hoffmann FG, Gering EJ, Moriyama H, Ferrand N *et al.* (2007). The molecular basis of high-altitude adaptation in deer mice. *PLoS Genet* **3**: e45.
- 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**: 171–187.
- van der Loo W, Arthur CP, Richardson BJ, Wallage-Drees M, Hamers R (1987). Nonrandom allele associations between unlinked protein loci: are the polymorphisms of the immunoglobulin constant regions adaptive? *Proc Natl Acad Sci USA* **84**: 3075–3079.
- van der Loo W, Bousses P, Arthur CP, 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**: 1181–1194.
- Vieira J, Ferrand N (1995). Four new alleles at the mannose-6-phosphate isomerase locus in rabbit. *Anim Genet* **26**: 37–38.
- Vitalis R, Dawson K, Boursot P (2001). Interpretation of variation across marker loci as evidence of selection. *Genetics* **158**: 1811–1823.
- Weber RE, Fago A (2004). Functional adaptation and its molecular basis in vertebrate hemoglobins, neuroglobins and cytoglobins. *Respir Physiol Neurobiol* **144**: 141–159.
- Weir BS, Cockerham CC (1984). Estimating F-statistics from the analysis of population structure. *Evolution* **38**: 1358–1370.
- Williams TN, Mwangi TW, Wambua S, Peto TEA, Weatherall DJ, Gupta S *et al.* (2005). Negative epistasis between the malaria-protective effects of alpha+-thalassemia and the sickle cell trait. *Nat Genet* **37**: 1253–1257.
- Woodruff DS (1989). Genetic anomalies associated with *Cerion* hybrid zones: the origin and maintenance of new electrophoretic variants called hybridzymes. *Biol J Linn Soc Lond* **36**: 281–294.
- Wu C-I (2001). The genetic view of the process of speciation. *J Evol Biol* **14**: 851–865.
- Wu C-I, Ting C-T (2004). Genes and speciation. *Nat Rev Genet* **5**: 114–122.
- Xu J, Hardison R (1989). Localization of the  $\beta$ -like globin gene cluster and the genes for parathyroid hormone and c-Harvey-ras 1 to region q14-q21 of rabbit chromosome 1 by *in situ* hybridization. *Cytogenet Cell Genet* **52**: 157–161.
- Xu J, Hardison R (1991). Localization of the  $\alpha$ -like globin gene cluster to region q12 of rabbit chromosome 6 by *in situ* hybridization. *Genomics* **9**: 362–365.