Genetic characterization of the chemokine receptor CXCR4 gene in lagomorphs: comparison between the families Ochotonidae and Leporidae

J. Abrantes,*† P. J. Esteves,* C. R. Carmo,* A. Müller,‡ G. Thompson‡ & W. van der Loo*†

Summary
Chemokines receptors are transmembrane proteins that bind chemokines. Chemokines and their receptors are known to play a crucial role in the immune system and in pathogen entry. There is evidence that myxoma virus, the causative agent of myxomatosis, can use the chemokine receptor CXCR4 to infect cells. This virus causes a benign disease in its natural host, Sylvilagus, but in the European rabbit (Oryctolagus cuniculus) it causes a highly fatal and infectious disease known as myxomatosis. We have characterized the chemokine receptor CXCR4 gene in five genera of the order Lagomorpha, Ochotona (Ochotonidae), and Oryctolagus, Lepus, Bunolagus and Sylvilagus (Leporidae). In lagomorphs, the CXCR4 is highly conserved, with most of the protein diversity found at surface regions. Five amino acid replacements were observed, two in the intracellular loops, one in the transmembrane domain and two in the extracellular loops. Oryctolagus features unique amino acid changes at the intracellular domains, putting this genus apart of all other lagomorphs. Furthermore, in the 37 European rabbits analysed, which included healthy rabbits and rabbits with clinical symptoms of myxomatosis, 14 nucleotide substitutions were obtained but no amino acid differences were observed.

Introduction
Chemokines receptors are seven-transmembrane domain G-protein coupled receptors that allow binding of chemokines (chemotactic cytokines), a superfamily of small heparin-binding proteins that share a sequence homology and are characterized by four conserved cysteines residues linked to disulfide bonds. Chemokines and their receptors play a central role in directing the migration of circulating leucocytes during immune surveillance, inflammation and the establishment of immunity, and are involved in organogenesis, hematopoiesis, angiogenesis, tumor growth, metastasis and dendritic cell maturation (reviewed in Rossi & Zlotnik, 2000). Chemokines receptors have also been associated with pathogen entry (e.g. Horuk et al., 1993; Feng et al., 1996). CXCR4, CCR1 and CCR5 are among the candidate receptors used by the myxoma virus to infect cells (Lalani et al., 1999).

The myxoma virus is the causative agent of myxomatosis, which can be a highly fatal disease in the European rabbit (Oryctolagus cuniculus) although benign in its natural host, species of Sylvilagus (Fenner, 1983). This virus belongs to the family poxviridae and like all the members of this family has a large, linear, double-stranded DNA genome that encodes more than 20 host-related proteins (Cameron et al., 1999; Zúñiga, 2002; Seet et al., 2003). A previous study on CCR5 in Oryctolagus revealed the replacement of a specific peptide motif of the second extracellular loop (ECL2) of the CCR5 protein by a motif which in other species characterizes the CCR2 molecules. This replacement was found in all the Oryctolagus specimen analysed while absent in Sylvilagus and Lepus species (Carmo et al., 2006).

Here, we have characterized the genetic diversity of the CXCR4 gene among lagomorphs. The CXCR4 has an open-reading frame that consists of two exons separated by an intron of 2.1kb. The first exon includes the sequence corresponding to the first five amino acids of the CXCR4 protein and a 5’ UTR region and the second exon codes for the last 347 amino acids and a-3’ UTR region (Wegner et al., 1998). We have sequenced the exon 2 of the receptor for Ochotona, Bunolagus, Lepus, Oryctolagus and Sylvilagus. Gene regions will be defined and named in accordance with the structural domains of the CXCR4 protein they encode, which consist, respectively, of the N-terminal domain (N-tail), the C-terminal domain (C-tail), the extracellular loops (ECL1, ECL2, ECL3) and the intracellular loops (ICL1, ICL2, ICL3) and the seven transmembrane domains (TM1-TM7).

The order Lagomorpha includes two families, the leporidae and the ochotonidae. The ochotonidae family differentiated from the leporidae family in the Oligocene (~35 Mya) (Yu et al., 2000) and is represented by one
extant genus, Ochotona (pikas), with approximately 30 species restricted to Asia and North America. The teiropidae family comprises 11 genera, Brachylagus, Bunolagus, Caprolagus, Lepus, Nesolagus, Oryctolagus, Pentalagus, Poelagus, Pronolagus, Romerolagus and Sylvilagus, with near 60 species with a worldwide distribution. The evolutionary relationships between leporid genera are not clearly established (Matthee et al., 2004; Robinson & Matthee, 2005). Oryctolagus is the only leporid genus with a European origin and comprises two subspecies, O. cuniculus cuniculus and O. c. algirus. The subspecies O. c. algirus is restricted to the south-west of Iberian Peninsula and a few Atlantic Islands (reviewed in Ferrand & Branco, 2007) while O. c. cuniculus has a worldwide distribution, man-made mainly, and includes all domestic breeds.

Materials and methods

Blood or tissue samples of 37 European rabbits belonging to both the subspecies O. c. cuniculus and O. c. algirus were obtained from wild specimen collected in Portugal, Spain and France. Eight of the samples presented clinical symptoms of myxomatosis. A sample of a domestic rabbit breed (Fauve de Borgogne) was also included. Ten leporid species and two ochotonid species were analysed: Bunolagus monticularis, Sylvilagus floridanus, Sylvilagus brasiliensis, Lepus americanus, Lepus townsendii, Lepus californicus, Lepus castroviejoi, Lepus europaenus, Lepus granatensis and Lepus saxatilis; Ochotona aucria and Ochotona princeps.

In lagomorphs, the genomic organization of the CXCR4 gene is unknown. Total RNA was isolated from spleen samples from Oryctolagus with TRIzol, following the manufacturer’s instructions (Life Technologies, Grand Island, NY, USA). CDNA was prepared as described (Krug & Berger, 1987) and was polymerase chain reaction (PCR) amplified using the primers CXCR4_F1 and CXCR4_R1 (for description of the primers used in this study see Table 1) designed according to conserved regions of the gene in mammals. PCR products of approximately 1050 bp in length were gel purified and cloned into pGEM-T Easy vector (Promega, Madison, WI, USA).

Total genomic DNA was extracted using a Qiagen extraction kit (Qiagen, Vienna, Austria). The PCRs were carried out in a 25 μl reaction volume containing 50 ng of genomic DNA, 1 x reaction buffer, 2.5 μL of DMSO, 3 mM MgCl₂, 0.5 mM of each dNTP, 0.4 μM of each primer and 1 U of Taq polymerase (EcoTaq, Ecogen, Spain). For Oryctolagus, PCR amplification was performed with the primers CXCR4_F2 and CXCR4_R1. For the other lagomorph genera, a different forward primer had to be used, CXCR4_F3, which was designed based on the rabbit cDNA sequence of CXCR4. The amplification conditions were 5 min at 94 °C followed by 35 cycles at 94 °C (45 s), 54 °C (30 s) and 72 °C (1 min), with a final extension at 72 °C (7 min). Sequencing was performed on an ABI PRISM 310 Genetic Analyser (PE Applied Biosystems, Foster City, CA, USA) following the ABI PRISM BigDye Terminator sequencing protocols. PCR products were sequenced in both directions.

The sequences obtained were aligned with published CXCR4 gene sequences available in NCBI GenBank using ClustalW (Thompson et al., 1994) and adjusted by visual inspection. The nucleotide and inferred protein sequences appear in the tables and figures with the following denominations and GenBank accession numbers: Homo (Homo sapiens: NM_001008540), Macaca (Macaca mulatta: NM_001042645), Sus (Sus scrofa: NM_213773), Bos (Bos taurus: NM_174301), Canis (Canis familiaris: NM_001048026), Felis (Felis catus: NM_001099826), Rattus (Rattus norvegicus: NM_022205), Mus (Mus musculus: NM_009911), Gallus (Gallus gallus: NM_204617) and Xenopus (Xenopus laevis: Y17895). Nucleotide sequence data obtained in this study have been submitted to GenBank and have been assigned the following accession numbers: Oryctolagus cuniculus EU258276 to EU258306 and EU265728 to EU265733; Bunolagus monticularis EU258264; Lepus europaenus EU258265; Lepus townsendii EU258266; Lepus castroviejoi EU258267; Lepus saxatilis EU258268; Lepus americanus EU258269; Lepus granatensis EU258270; Lepus californicus EU258271; Sylvilagus floridanus EU258272; Sylvilagus brasiliensis EU258273; Ochotona aurica EU258274 and Ochotona princeps EU258275. Measures of nucleotide diversity were estimated using the software package DnaSP (Rozas et al., 2003). The number of synonymous substitutions per synonymous site (dS) and the number of nonsynonymous substitutions per nonsynonymous site (dN) were calculated using the MEGA 3.1 software (Kumar et al., 2004). Seven sequences from European rabbit were excluded from the analyses due to the short length. Phylogenetic trees were obtained by neighbour-joining. The reliability of the trees was tested by bootstrap (Felsenstein, 1985).

Results

The exon 2 of CXCR4 was sequenced for the lagomorph species aforementioned. All the sequences obtained cover the three intracellular, the three extracellular loops and the seven transmembrane domains of the CXCR4 protein.

Table 1. Sequences of the primers used for amplification by polymerase chain reaction and schematic representation of their localization. The nucleotide positions are according to the human cDNA sequence of CXCR4 GenBank accession number NM_001008540

<table>
<thead>
<tr>
<th>Primer</th>
<th>Composition (5’-3’)</th>
<th>Positions (according to the human cDNA sequence)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CXCR4_F1</td>
<td>ATGGAAGGGGATCAGTATAC</td>
<td>1–21</td>
</tr>
<tr>
<td>CXCR4_R1</td>
<td>TTAGCTGGAGTGAACATTGA</td>
<td>1039–1059</td>
</tr>
<tr>
<td>CXCR4_F2</td>
<td>ACTTCAAGCACAATCACGGA</td>
<td>22–41</td>
</tr>
<tr>
<td>CXCR4_F3</td>
<td>CATAAAGGACCCCTGCTCC</td>
<td>69–88</td>
</tr>
</tbody>
</table>

The exon 2 of CXCR4 was sequenced for the lagomorph species aforementioned. All the sequences obtained cover the three intracellular, the three extracellular loops and the seven transmembrane domains of the CXCR4 protein.
The data for the carboxyl- and the amino-terminal domains are incomplete since these regions were used as template for the primers.

### Nucleotide polymorphism

Table 2 presents the nucleotide variation at the CXCR4 CDS region among 30 individuals of *Oryctolagus cuniculus*. These included healthy rabbits as well as rabbits with clinical symptoms of myxomatosis. Each of the 14 nucleotide polymorphisms observed were synonymous. Table 3 shows the results of the parameters of gene diversity. All the parameters calculated suggest a low level of nucleotide diversity. The overall average number of nucleotide differences per site between two sequences was lower than 0.001. Nucleotide differences

---

**Table 2.** Nucleotide variation at the CXCR4 CDS region among 30 individuals of *Oryctolagus cuniculus*. The nucleotide positions are numbered according to the putative initiation site of human CXCR4 gene. R = A/G, M = A/C, Y = C/T, represent polymorphic positions; ‘.’ represents identity with the reference sequence (GenBank accession number EU258276); the code of each sample is in brackets; the rabbits with clinical symptoms of myxomatosis are shaded in grey

<table>
<thead>
<tr>
<th>Population/Breed</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2 2 2 4 5 6 7 7 7 8 8 9 9</td>
</tr>
<tr>
<td></td>
<td>3 4 7 7 1 3 9 1 4 8 1 9 0 6</td>
</tr>
</tbody>
</table>

| O.c.algirus             | Sernancelhe (Sern2) |       |       |       |       | C | Y |       |       |       |       | R |       |
| Algete (Alg1)           | . . . . . . . . . . | . . . | . . . | . . . | . . . | . | . | . . . | . . . | . . . | . . . | R | . . .   |
| Algete (Alg2)           | . . . . . . . . . . | . . . | . . . | . . . | . . . | K | . | . . . | . . . | . . . | . . . | R | . . .   |
| Algete (Alg3,4)         | . . . . . . . . . . | . . . | . . . | . . . | . . . | K | . | . . . | . . . | . . . | . . . | R | . . .   |
| Algete (Alg5,16)        | . . . . . . . . . . | . . . | . . . | . . . | . . . | K | . | . . . | . . . | . . . | . . . | R | . . .   |
| Algete (Alg6,13,17)     | . . . . . . . . . . | . . . | . . . | . . . | . . . | K | . | . . . | . . . | . . . | . . . | R | . . .   |
| Algete (Alg10)          | . . . . . . . . . . | . . . | . . . | . . . | . . . | M | . | . . . | . . . | . . . | . . . | R | . . .   |
| Algete (Alg12)          | . . . . . . . . . . | . . . | . . . | . . . | . . . | K | . | . . . | . . . | . . . | . . . | R | . . .   |
| Algete (Alg18)          | . . . . . . . . . . | . . . | . . . | . . . | . . . | A | . | . . . | . . . | . . . | . . . | R | . . .   |
| Alicante (Alic18)       | . . . . . . . . . . | . . . | . . . | . . . | . . . | A | . | . . . | . . . | . . . | . . . | R | . . .   |
| F.Borg. (fba18)         | . . . . . . . . . . | . . . | . . . | . . . | . . . | M | . | . . . | . . . | . . . | . . . | R | . . .   |
| Domestic (CD1,3,4)      | . . . . . . . . . . | . . . | . . . | . . . | . . . | A | . | . . . | . . . | . . . | . . . | R | . . .   |
| Domestic (CD2)          | . . . . . . . . . . | . . . | . . . | . . . | . . . | A | . | . . . | . . . | . . . | . . . | R | . . .   |

**Table 3.** Estimates of polymorphism in the CXCR4 nucleotide sequences of lagomorphs. *N*, number of individuals; *S*, number of polymorphic sites; *Syn* and *NonSyn*, number of polymorphic sites that are synonymous and nonsynonymous, respectively; *π*, average number of nucleotide differences per site between two sequences; *θ*, proportion of nucleotide sites that are expected to be polymorphic. Number of synonymous substitutions per synonymous site (*dS*) and the number of nonsynonymous substitutions per non-synonymous site (*dN*) were calculated using the formulae of Li, Wu & Luo (1985)

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>S</th>
<th>Syn</th>
<th>NonSyn</th>
<th>dS</th>
<th>dN</th>
<th>dN/dS</th>
<th>π</th>
<th>θ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>45</td>
<td>95</td>
<td>83</td>
<td>12</td>
<td>0.04189 ± 0.00572</td>
<td>0.001221 ± 0.00083</td>
<td>0.05601</td>
<td>0.00984</td>
<td>0.02371</td>
</tr>
<tr>
<td>Bunolagus</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.00000</td>
<td>0.00000</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Oryctolagus</td>
<td>30</td>
<td>14</td>
<td>14</td>
<td>0</td>
<td>0.01251 ± 0.00433</td>
<td>0.00000</td>
<td>—</td>
<td>0.00269</td>
<td>0.00348</td>
</tr>
<tr>
<td>Lepus</td>
<td>7</td>
<td>10</td>
<td>9</td>
<td>1</td>
<td>0.01759 ± 0.00512</td>
<td>0.00056 ± 0.00054</td>
<td>0.03184</td>
<td>0.00303</td>
<td>0.00410</td>
</tr>
<tr>
<td>Sylvilagus</td>
<td>2</td>
<td>6</td>
<td>5</td>
<td>1</td>
<td>0.01724 ± 0.00776</td>
<td>0.00095 ± 0.00091</td>
<td>0.05510</td>
<td>0.00445</td>
<td>0.00445</td>
</tr>
<tr>
<td>Ochotona</td>
<td>2</td>
<td>14</td>
<td>14</td>
<td>0</td>
<td>0.07464 ± 0.02119</td>
<td>0.00000</td>
<td>—</td>
<td>0.01559</td>
<td>0.01559</td>
</tr>
</tbody>
</table>
Figure 1. Differences on the deduced amino acid sequences of CXCR4. The amino acid positions are numbered according to the amino acid sequence of human CXCR4. The different domains of the CXCR4 protein are according to the topology accepted for human CXCR4. Non-identical amino acids were checked visually, and when it was not possible to decide on an amino acid a '?' was used. The Oryctolagus sequence is a consensus based on the 37 sequences of CXCR4 of Oryctolagus cuniculus; Lame, Lepus americanus; Sbra, Sylvilagus brasiliensis.
morphic alteration, I185L. The genus Lepus americanus. Sylvilagus brasiliensis
apomorphic amino acids changes, N 101S, K 110Q, V 177I,
phic.

alteration I 269V and the apomorphic change L 165F in
genus
distance and complete deletion (Fig. 2). The tree was
constructed using MEGA3.1 software, option p-

Amino acid polymorphism
The amino acid sequences of CXCR4 were inferred from the
nucleotide data and were aligned with published
CXCR4 sequences of mammals. The amino acid dif-
fences found among genera are highlighted in Fig. 1.
The consensus sequence is based on the amino acid
CXCR4 sequences of mammals other than lagomorphs
and was constructed using bioedit 7.0.5.3 (Hall, 1999)
with a threshold frequency for inclusion of 80%.

The overall degree of diversity among lagomorphs was
low with the largest distance found between ochotonid
and leporid species and did not exceed 2% of amino acid
differences. The genus Oryctolagus harboured two amino
acid changes, Q69L and K146R, that are clearly apomor-
differences found among lagomorph species were predominantly
synonymous.

Phylogenetic relationships
A neighbour-joining tree for CXCR4 nucleotide sequences
was constructed using MEGA3.1 software, option p-
distance and complete deletion (Fig. 2). The tree was

Discussion
CXCR4 was reported as the most evolutionary conserved
of all CXCRs (Murphy, 1993; Huising et al., 2003; Lio &
Vannucci, 2003) and despite the redundancy within the
chemokine system, it has only one ligand. Its wide range
of functions and wide variety of tissues where it is
expressed seem to be implicated in this specificity and
conservative status. For example, mice lacking CXCR4
or SDF-1 die perinatally and show profound defects in
the haematopoietic and nervous systems (Ma et al.,
1998).

As expected, the characterization of CXCR4 in
lagomorphs showed that genetic diversity at this locus
is very limited. More in particular, for the seven trans-
membrane domains (157 amino acids) no variation was
found among lineages that have evolved independently
for more than 35 myr. Only one polymorphism was
recorded for these regions in an individual sample of Lepus
americanus (L165F). Pikas appear as the most divergent
group compared to the mammalian consensus, with two
substitutions in ECL1 and eight substitutions in ECL2.
Also for the other species, most changes, when occurring,
affect ECL regions. These observations suggest that the
internal regions (ICL and TM) of the CXCR4 protein are
under strong purifying selection. Oryctolagus appears as
a notable exception, where both ICL1 and ICL2 show
amino acids replacements, respectively, L69Q and R146K.
Among lagomorphs, these replacements are unique to
rabbit. They are apparently apomorphic mutations,
although the former is also present in chicken and the
latter in rhesus monkey and in cattle (Fig. 1).

In a recent study on CCR5 in rabbit, a remarkable gene
conversion was described between this gene and CCR2. A
specific peptide motif of the second extracellular loop
of the CCR5 protein is replaced by a motif that in other
species characterizes the CCR2 protein (Carmo et al.,
2006). Here we report two amino acid changes in intracellu-
lar domains of the CXCR4 protein. The intracellular
regions are known to be involved in signal transduction
that influences patterns of cellular immune response and
causes cell death (Vlahakis et al., 2002; Roland et al.,
2003).

The clinical symptoms that follow the myxoma infection
observed among susceptible lagomorph hosts are not the
same. In the members of Sylvilagus species the infection
is restricted to the inoculation site and results in the forma-
tion of skin tumours but no mortality has been observed.
In European rabbits the infection is much more severe and
reaches high mortality rates. Following the infection,
myxoma virus replicates in class II MHC+ cells with a
dendritic shape and then in T lymphocytes of the draining lymph node that probably disseminate the virus through the body (Zúñiga, 2002).

The different outcome between Sylvilagus and Oryctolagus strongly suggests the existence of species-specific differences at genes involved in the acquired and innate immune response. The use of chemokine receptors by myxoma virus may also interfere in the type of immune response since chemokines and chemokine receptors can regulate leucocyte trafficking, angiogenesis and inflammation. There is clear evidence of a different cellular immune response in rabbits that died from myxomatosis compared to those that overcome the infection (Best & Kerr, 2000a) and apoptosis only occurs in cells with no detectable viral antigen but adjacent to infected cells (Best et al., 2000b). The intracellular pathways and the cellular messengers activated after the infection are still unknown but the possibility exists that the unique features found in the intracellular domains of the Oryctolagus CXCR4 protein (Fig. 3) may somehow be involved in the different course of the disease. Functional studies should evaluate the possible effect of these alterations.

Acknowledgements

We thank Alison Surridge and Paulo Célio Alves for providing us the samples of Bunolagus, Sylvilagus, Ochotona and Lepus. We thank also Nuno Ferrand for all the support for the execution of this work. This work was supported by a grant of the Foundation for Science and Technology-Portugal (SFRH/BD/31048/2006 and SFRH/BPD/27021/2006) to J.A and P.J.E., respectively, and by a project of the Foundation for Science and Technology-Portugal (POCI/BIA-BDE/61553/2004).

References


analyses and a supermatrix analysis. Mammal Review, 35, 231.