Immunogenetics (2002) 54:39–47
DOI 10.1007/s00251-002-0433-0

ORIGINAL PAPER

Viet Khong Nguyen · Chen Su · Serge Muyldermans
Wessel van der Loo

Heavy-chain antibodies in *Camelidae*; a case of evolutionary innovation

Received: 17 September 2001 / Revised: 17 December 2001 / Published online: 26 February 2002

Abstract The emergence in *Camelidae* species of functional antibodies devoid of light chains (referred to as heavy-chain antibodies or HCAbs) is an intriguing evolutionary event. Homodimeric HCAbs have also been documented in spotted ratfish (*Cos5*-Abs) and nurse shark (NAR). To reveal the evolutionary history of HCAbs, we evaluated the phylogenetic and phenotypic relationships among HCAbs and conventional antibodies across taxa and confirmed the current viewpoint that different groups of HCAbs have evolved independently in the three lineages. At least, in the camelids, HCAbs are not the result of resuscitation of dormant genes. They are derived from the conventional antibodies within the *Camelidae* lineage, and are apparently the outcome of more recent adaptive changes occurring in the compartment of heteromeric antibodies. The shared structural properties of HCAbs across taxa are therefore explained by convergent evolution due to similar constraints related to the absence of pairing to the light chain. It appears that innovative evolutionary changes in *Camelidae* have led to a new level of antigen binding repertoire diversification and have allowed acquisition of novel antigen-receptor properties.

Keywords Antibody · Camel · Evolution · Heavy-chain antibody · Phylogeny

V.K. Nguyen (✉) · S. Muyldermans · W. van der Loo
Department of Ultrastructure,
Vlaams Interuniversitair Instituut voor Biotechnologie,
Vrije Universiteit Brussel, Paardenstraat 65,
1640 Sint Genesius Rode, Belgium
e-mail: nguykhon@vub.ac.be
Tel.: +32-235-90227, Fax: +32-235-90289

V.K. Nguyen
Central Vietnam Veterinary Institute,
Agricultural and RD Ministry, Milestone 4, Avenue 2/4,
Dong De, Nha Trang, Vietnam

C. Su
Institute of Molecular Evolutionary Genetics
and Department of Biology, Pennsylvania State University,
University Park, PA 16802, USA

Introduction

Typical functional antibodies of all jawed vertebrates are heteromeric molecules composed of light (L) and heavy (H) chains (herein referred to as conventional antibodies) (Tonegawa 1983; Padlan 1994). The random pairing of an H-chain and an L-chain considerably diversifies the antigen (Ag)-binding repertoire. The discovery in *Camelidae* species of additional classes of functional antibodies composed solely of H-chains was unexpected because of the inherent loss of this combinatorial diversity (Hamers-Casterman et al. 1993). It is, therefore, interesting to understand how and why these heavy-chain antibodies (HCAbs) emerged, and what evolutionary mechanism led to the acquisition of a diverse Ag-binding repertoire in the absence of an L-chain.

The H-chains of HCAbs in *Camelidae* deviate from that of the conventional antibodies in two important aspects. First, the sequences of HCAb variable domains (V_HH) differ significantly from those of the conventional antibodies (V_H) (Muyldermans et al. 1994; Nguyen et al. 1998). Second, the first domain of the constant region (C_H) is absent in HCAbs, and the matured V_HH-DJ is directly joined to the hinge region. Conventional antibodies and HCAbs are encoded in separated sets of the *V* and *C* genes (Nguyen et al. 1999, 2000; Woolven et al. 1999), but appear to the share some *D* genes (Nguyen et al. 2000, 2001) and an identical J_H region (VK Nguyen, unpublished). The coexistence of HCAbs and conventional antibodies, both contributing to the humoral immune response in *Camelidae* species (Hamers-Casterman et al. 1993; Muyldermans and Lauwereys 1999) provides an opportunity to study the evolutionary forces underlying this dichotomic humoral immune system.

Another case of HCAb, the new antigen receptor (NAR), was reported for the nurse shark (Greenberg et al. 1995). The homodimeric NAR H-chain contains one V_{NAR} and five C_H domains. Because nurse shark belongs to a primitive vertebrate lineage, the NAR may represent a candidate for a remnant of the evolutionary intermediate between the putative primordial Ag-receptor and the

heterodimeric antibodies (Roux et al. 1998). HCAbs were also reported for the ratfish, in which the *Cos5*-type Abs (*Cos5*-Ab) also lack L-chains (Rast et al. 1998).

Previous structural analyses strongly suggested that HCAbs in shark and camelids arose by convergence (Roux et al. 1998). In this study, we analyzed this problem in more detail, using compiled phylogenetic and phenotypic data to clarify the origin of *Camelidae* HCAbs during the evolution of immunoglobulins within jawed vertebrates. Alternative hypotheses were envisaged in which HCAbs – or parts of them – were (1) derived from conventional antibodies undergoing an evolutionary convergence, or (2) originated directly from a dormant ancestor gene(s) that has subsisted throughout evolution. We propose that the creation of new opportunities in Ag-binding might have been the driving evolutionary force to develop the HCAbs in *Camelidae*.

Materials and methods

The *C* genes

Differences between the HCAbs and the conventional antibodies reside in both the V_H and the C_H portions. To reveal the evolutionary history of the C_H region of HCAbs, the most conserved portions of their constant region (*C*) were used as a query to search in the EMBL (Stoesser et al. 2001) and IMGT (Lefranc 2001) databases. This search found closely related *C* genes of the conventional antibodies, as documented in previous reports (Greenberg et al. 1996; Rast et al. 1998; Nguyen et al. 1999). The C_H/C_HH sequences used in the analysis (Fig. 1) are: (1) the ratfish *Cos5*-Ab (AF003844–45) and *IgM* (AF003862–63) (Rast et al. 1998); (2) the nurse shark NAR (U18701) and *IgNARC* (U51450) (Greenberg et al. 1995, 1996); (3) the dromedary HCAb *IgG2a* (AJ131945) (Nguyen et al. 1999) and conventional *IgG1a* (AJ421266, this work); (4) the llama conventional antibody *IgG1a* (AF305955) and HCAb clone *B3* (Woolven et al. 1999) [namely, *IgG2c* (Nguyen et al. 2001)]; and (5) the bovine *IgG1* (X16701) (Symons et al. 1989).

The *V* sequences

Representative V_H and V_HH sequences from a wide range of organisms were retrieved either from EMBL and IMGT databases or from the literature. Only germline sequences for which the complete *V* segment is available and that are known to be functional or potentially functional were included. However, for taxa where no germline V_H sequences were available, we used the rearranged sequences (cDNAs), one per each V_H family. *V* segments of ≥80% sequence identity belong to the same family (Brodeur and Riblet 1984). A total of 403 sequences (352 germline + 51 cDNAs) were collected, of which 378 were unique. These 378 sequences (available from the corresponding author upon request) comprise 332 conventional V_Hs: 43 dromedary and 3 llama V_Hs (Vu et al. 1997), and 286 V_Hs (95 V_H families) from 29 non-camelid species, and 46 V_HH sequences: 32 camel and nine llama V_HHs, 4 NARs and 1 ratfish *Cos5*-Ab V_HH.

For the construction of a phylogenetic tree (shown in Fig. 2), we generated a subset of data, in which the dromedary was represented by a selection of three bona fide V_Hs and one for each of the seven V_HH subfamilies (Nguyen et al. 2000), and the llama by the 12 cDNA sequences (three V_H and nine V_HH). For the non-camelids each V_H family was represented by one member, except for NARs, where two sequences were included. Thus, we obtained a subset of 118 sequences with accession numbers as follows, where an asterisk indicates the cDNA sequences: Mammalian: bovine U55164; camel AJ245107, AJ245113, AJ245117, AJ245124, AJ245134, AJ245135, AJ245148, AJ245163, AJ245177, AJ245194; dog AF067449*; horse U15150*; human X92209, X62111, M99676, X62112, X92227, X92224, L10057; llama AJ237296*, AJ237312*, AJ237341*, AJ237386*; the V_HH07*, 23*, 41*, 56* and V_H13*, V_H71* (from Vu et al. 1997), AJ421267*, AJ42128* (this work); mice L14548, U53526, K02791, X01437, X01113, X03398*, J00526, L14362, AF064444, M35502*, X53338*, X55935*, X55934, U39293*; pig AF064688; rabbit M93173; rat AF098325; sheep Z49181; opossum AF012113*, AF012124. Birds: chicken M30319; duck X65218*. Reptile: caiman M12769, M12770. Amphibian: axolotl L20243*, AF027254*, AF027255*, AF027257*, AF011571, AF027260*, AF027261*, AF027262*, AF027267*, AF027268*, AF027269*; xenopus M97006, (V2 and V3 from Schwager et al. 1989), X56858, X56859, X56860, X56861, X56862, X56863, X56864, X56865. Fish: catfish AF068137, AF068137, U09721, U09722, U09724; coelacanth X57354, X57353; elops M26182*; fugu AF108421; goldfish J03616, X61312; rainbow M57442, X65262*, L28805*, L28744*, L28745*, L28746*, L28747*, X81501*, X81505*, X81508*, X81511*; ratfish AF003853*, AF003841, AF003946; sturgeon Y13256*, Y13261*, Y13265*; clear-nose skate *Vx2113* U08009; little skate M29672*, X16146, and X15124; bull shark U50614*; horned shark M12195 and Z11776; nurse shark U51450*, L38966, L38968; and sandbar shark U50610*.

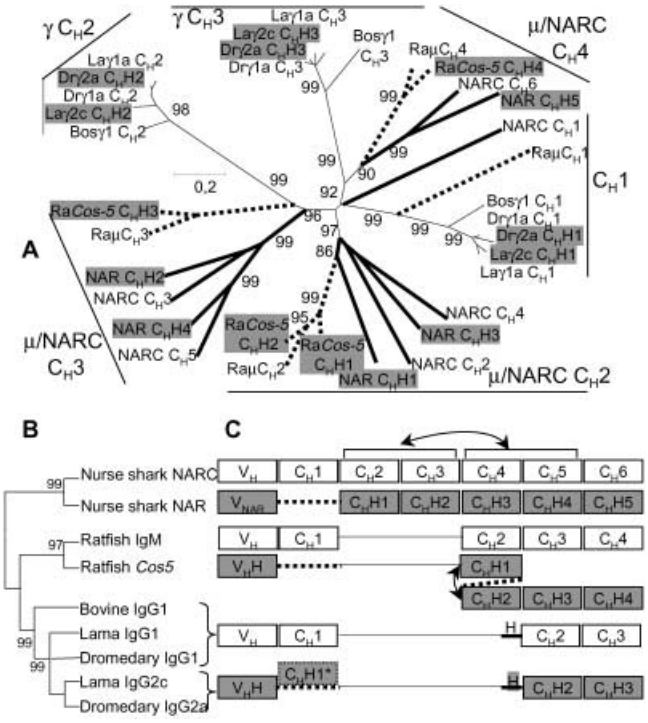
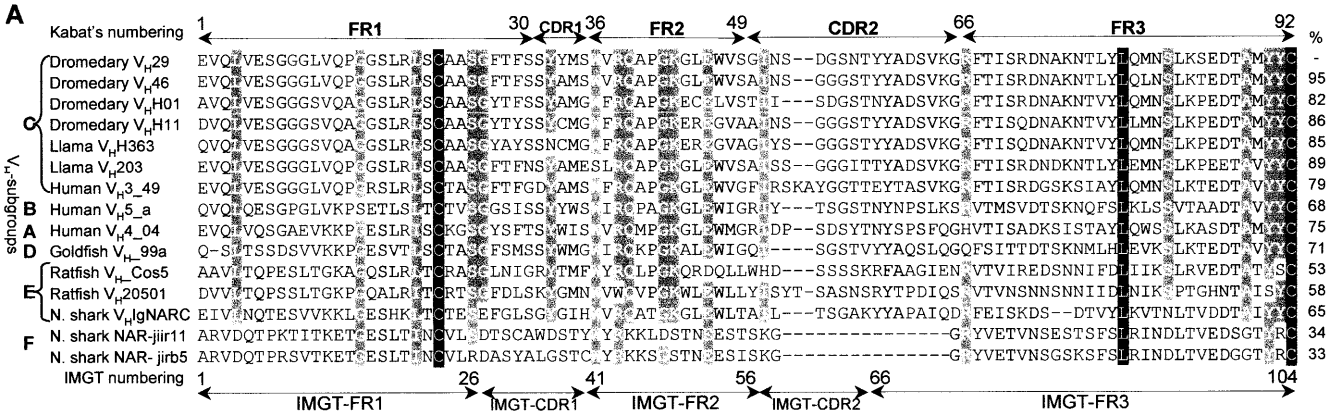


Fig. 1 **A** Phylogenetic (ME) tree of the C domains of C_HH genes (*shaded box*) from four species possessing HCAbs and those of their closest C_H genes, reconstructed with *p* distance (Nei and Kumar 2000). Clusters of C_H domains typical of the *γ* and *μ/NARC* genes (Greenberg et al. 1995; Rast et al. 1998; Nguyen et al. 1999) are indicated. Simplified relationships between C_HH genes and their most closely related C_H genes are illustrated on the ME tree (**B**) and by domain homologies (**C**) as inferred from the tree topology shown in **A**. *Bos* Bovine; *Dr* dromedary; *La* llama; *NAR* nurse shark new antigen receptor; *Ra* ratfish

Evolutionary analyses

The phylogenetic analysis was conducted using the MEGA2 computer program (http://www.megasoftware.net) (Kumar et al.



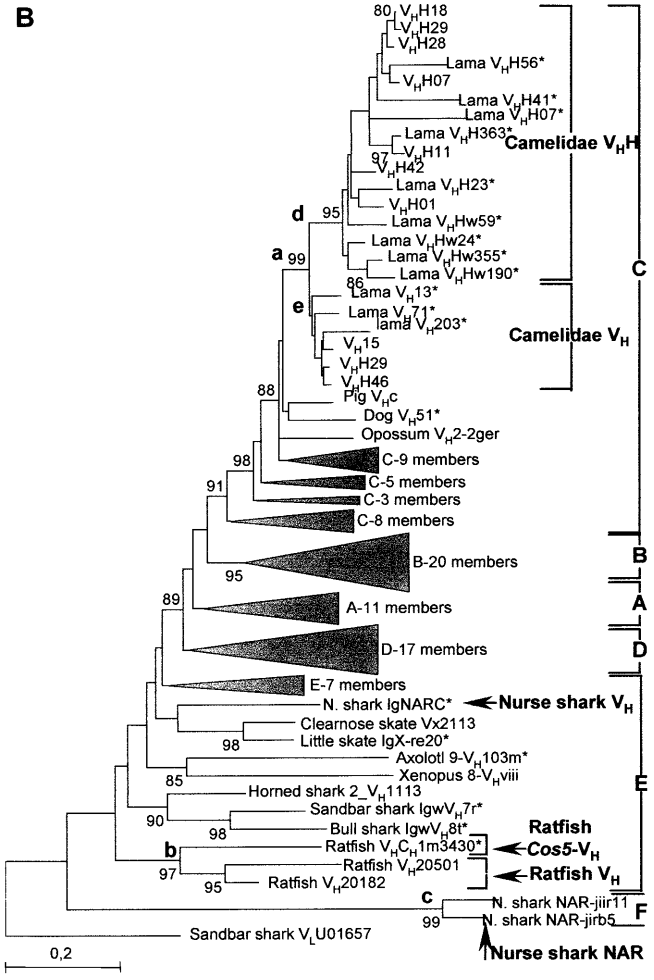


Fig. 2 **A** Deduced amino acid sequence alignments (residues 1 to the second conserved Cys) of representative *V* segments of major V_H subgroups (*left panel*) and V_HH segments of camelids, ratfish and nurse shark. Conserved V_H residues are *shaded*. The Kabat's (*top*) and IMGT (*bottom*) *FR/CDR* definitions are indicated with corresponding numberings. The degree of amino acid sequence similarity (%) to the dromedary V_H29 is shown at the end of each sequence. **B** Phylogenetic (ME) tree of 118 representative V_H and V_HH segments from 31 vertebrate species. For clarity, topologies of numbers (*indicated*) of V_H members are reduced (compress-subtree option in the MEGA2 program). The number immediately following the common name for the species indicates the gene family (where applicable), the gene notations from the original publication, or the EMBL GenBank accession number is also shown. V_H subgroups (A, B, C, D, E and F) are indicated. The clusters of V_H/V_HHs of camelids, the ratfish $V_H/Cos5$-Ab V_HH and the nurse shark *NAR* are indicated and marked by *a*, *b*, and *c*, respectively. The P_B values above 80% are shown on the branches. The sandbar shark V_L (U01657) was used as an outgroup. The same topology was also obtained with other tree making methods, e.g., the UPGMA, NJ, MP and ML

1994) and the GCG programs made available by BEN (http://ben.vub.ac.be/www2gcg). The gene trees of different species were constructed using the minimum evolution (ME) method (Rzhetsky and Nei 1992a) and Poisson-correction distances for pairwise amino acid sequences (see in Rzhetsky and Nei 1992b, 1993 and Nei and Kumar 2000, p.19). Optimal alignment was obtained using the ClustalW program (Thompson et al. 1994) with visual inspections for gaps according to V_H-residue nomenclature (Chothia et al. 1989; Kabat et al. 1991; Lefranc 2001). The reliability of trees was examined using the interior-branch test (Rzhetsky and Nei 1992b; Sitnikova et al. 1995). Other available tree reconstruction methods, including maximum parsimony (MP), neighbor joining (NJ), the unweighted paired-group method using arithmetic averages (UPGMA) and maximum likelihood (ML) were also used.

The rates of the non-synonymous (dN) and synonymous (dS) substitutions were computed using the method of Nei and Gojobori (1986), as implemented in MEGA2 (Nei and Kumar 2000). The null hypothesis of $\overline{d}N = \overline{d}S$ was evaluated using the one-tailed test.

Results

Phylogeny of *HCAb* constant region genes

Previous reports (Greenberg et al. 1996; Rast et al. 1998; Nguyen et al. 1999; Woolven et al. 1999) have independently showed that no known HCAbs have a typical C_H1 domain as compared to their close-related conventional counterparts. To reveal the phylogenetic relationship between all *HCAb* isotypes, we constructed phylogenetic trees of the constant region of these HCAb (C_HH) and of the conventional Ab (C_H) using the exon sequences. Regardless of the phylogenetic methods used, the C_HH genes of ratfish (*Cos5*-C_HH), shark (NAR) and camelids (llama *IgG2c* and camel *IgG2a*) were always part of

three clearly separated branches (Fig. 1B), and closely grouped with their conventional counterparts. A similar result was obtained using a larger number of C_H sequences from different species (not shown). As each domain (exon) might have a different origin, we also analyzed the phylogenetic relationships among different exons (Fig. 1A).

In ratfish, it is clear that C_H4 and C_HH4 are monophyletic, as well as C_H3 and C_HH3. Interestingly, the cluster of C_H2 and C_HH2 groups with C_HH1, whereas C_H1 is monophyletic. This corroborates previous analysis, where sequence comparison indicated that the typical first exon of ratfish *HCAbs* was eliminated and replaced by a duplicated second exon (Rast et al. 1998); Fig. 1C).

As well in the nurse shark, the C_H1 misses a counterpart in the *C* gene of HCAbs (NAR) (Greenberg et al. 1996). In this analysis, both the tree topology and the sequence alignment (Fig. 1) also show that the C_H1 homologue has been eliminated in the nurse shark HCAbs (or, the C_H1 was inserted in the constant gene for conventional antibodies).

For the γ genes in *Camelidae*, the C_H3 and C_HH3 form a tight cluster to which the bovine C_H3 belongs. A similar clustering pattern is found for their second and first exons (Fig. 1A). This clustering pattern and the fact that the C_HH1 is eliminated during the mRNA maturation of llama and dromedary HCAbs (Nguyen et al. 1999; Woolven et al. 1999) strongly suggests that the HCAb-specific *Ig constant* genes in *Camelidae* were derived in a relatively recent past from genes coding for conventional antibodies.

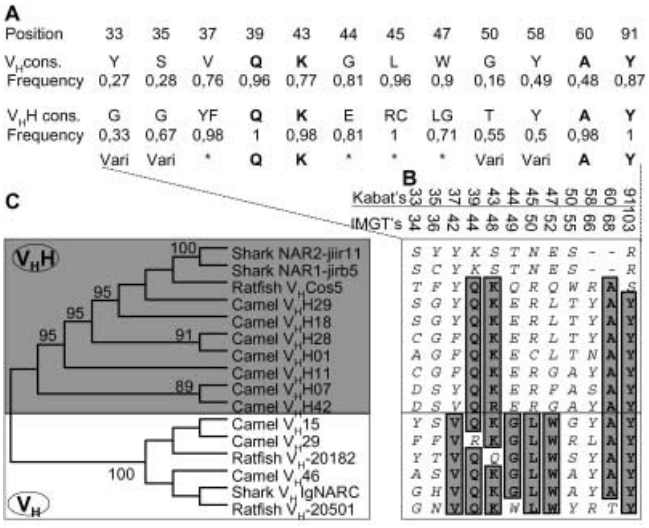


Fig. 3 **A** Amino acids and frequency of the most common amino acid among 332 V_Hs at 12 sites that frequently interact with the V_L, and consensus residues of 42 V_HH sequences at the corresponding sites. *Asterisks* indicate crucial substitutions between V_H and V_HH; *Vari* denotes the variable residues between and among V_H and V_HH at that position. **B** Alignment of the amino acid residues at the 12 sites (Kabat numbering) that frequently interact with V_L in the conventional antibodies, of 16 *V* segments from three species having HCAbs. Conserved V_H residues and those remaining unchanged in the V_HHs are shaded. **C** Maximum parsimony tree of 16 *V* genes (same as in **B**) based on 12 specific sites shown in **A** and **B**

Phylogeny of V_H/V_HH segments

The relationship between V domains of HCAbs and conventional antibodies was assessed in a phylogenetic analysis covering functional V_HH and V_H sequences of all vertebrates studied to date, as illustrated by the alignment of representative sequences (Fig. 2A). In this analysis, only the amino acid sequences of frameworks (FR) (Kabat et al. 1991) were used. Our results (Fig. 2B) show that with the exception of the V_{NARs}, all functional V_H genes of vertebrates belong to one of the currently accepted five major subgroups (A–E) (Ota and Nei 1994), although the subgroup E has less resolution. All mammalian V_Hs genes, inclusive of the V_HH of the camelids, are found within subgroups A–C.

Interestingly, the camel and llama V_HHs and V_Hs form a monophyletic cluster (indicated by "a" in Fig. 2B) within the V_H subgroup C (Fig. 2B, e and d for V_Hs and V_HHs, respectively). These patterns were also observed when 378 unique vertebrate *V* segments were included in the analysis (data not shown). The clustering of V_HHs of llama and dromedary, separated from their V_H counterparts, indicates that the most recent common ancestor of these genes predates the divergence of llama and camel species. The monophyly of camel V_Hs and V_HHs indicates that the V_HH family emerged probably within Tylopoda (the suborder of Artiodactyla to which *Camelidae* belongs).

The ratfish *Cos*5-Ab V_HH segment and its conventional V_H counterpart make a tight cluster within the V_H-subgroup E (b in Fig. 2B). In contrast, the V_{NAR}s (c in Fig. 2B) do not cluster with the nurse shark V_H or with any of the five major subgroups, and were tentatively assigned to a sixth subgroup (F). The large distances among the three branches containing the HCAb *V* segments (a, b and c, Fig. 2B), and the fact that they tend to cluster with the conventional V_Hs suggest that the *V* segments of the HCAbs of camelids, ratfish, and nurse shark emerged independently from each other.

Adaptive changes of V_H and V_HH

In the following sections, we examine evolutionary adaptations in different regions of V_HH genes.

V_H-V_L *interaction region*

The frequency of the most common amino acid residue at 12 V_H positions known to interact with V_L (Padlan 1994) was determined for 332 vertebrate V_H segments. For comparison, the consensus amino acid sequences were calculated for 42 dromedary V_HHs at the corresponding locations (Fig. 3A). The preferred amino acid

residues at four positions (39, 43, 60 and 91; numbering as in Kabat et al. 1991) are the same in V_Hs and V_HHs (for the corresponding V_H-IMGT numbering see Fig. 3B and in IMGTnumberingsTable, http://imgt.cines.fr; Lefranc 2001). At four other sites (33, 35, 50 and 58), neither V_H nor V_HH sequences show a pronounced amino acid preference. At the latter V_H sites, the extent of the contact with the V_L is dependent on the angle between V_H and V_L domains (Padlan 1994), and this may explain the observed amino acid degeneracy. In contrast, crucial differences between V_H and V_HH proteins were found at positions 37, 44, 45 and 47, where the amino acids are highly conserved among V_H genes (i.e., V37, G44, L45, and W47), and strikingly different in the V_HHs [R45 (or C), E44, F37 (or Y) and G47 (or L)]. These results support previous identification of camel V_HH-specific "hallmark" residues (Muyldermans et al. 1994), which presumably arose in response to the absence of the L-chain in HCAbs.

Substitutions of the conserved V_H residues responsible for V_H-V_L dimerization were also reported in the NAR- and *Cos5*-V_HHs (Greenberg et al. 1995; Rast et al. 1998). Some of these confer to the camel V_HH-specific "hallmark" residues (e.g. *Cos5*-V_HH Y37 and R45, and NAR Y37, Fig. 3B), and those residues that deviate from the camel V_HH hallmarks are certainly not of the V_H-type. It appears that the V domains of the camel HCAbs, *Cos5*-Abs, and NARs share phenotypic features that distinguish them from those of conventional antibodies. This is illustrated by the abridged representation of a tree (Fig. 3C) which is based upon these 12 positions for the same 118 *V* segments that were used for constructing the tree shown in Fig. 2B. It is clear that all *V* segments of the V_HH genes cluster apart from the conventional V_H genes.

This cross-taxa similarity of the V_HH "hallmark" residues correlates with the shared absence of the V_L in their HCAbs, reflecting a convergent evolution among these species. If acquired independently, these hallmarks must have been selected under similar functional constraints. Conversely, a V_HH subregion containing these hallmarks might be derived from a remnant of the putative primordial Ab *V* gene, which might somehow have persisted in the different vertebrate lineages. We therefore constructed phylogenetic trees for subdomains of 83 *V* segments defined either by the FR/CDR or β-sheet scaffold/H-loop definitions (Chothia et al. 1989; Kabat et al. 1991). Interestingly, all V_HH-types form a cluster, which is separated from the V_H-types, when subdomains were limited to (1) the Kabat's *FR2* (residues 36–49) or the *IMGT-FR2* (41–56, see Fig. 2A), harboring the V_HH hallmark residues; (2) the B-C strands of the Ig protein fold (residues 32–52), encoded by the region bordered by the palindromic nucleotide sequences reported by Nguyen and coworkers (2000); or (3) the region containing residues 27–52 (the IMGT residues 28–57, Fig. 4A). The last subregion encompasses the V_HH hallmarks and a region that is known to interact with the antigen in V_HHs but not in V_Hs (Padlan 1994; Decanniere et al. 1999). In contrast,

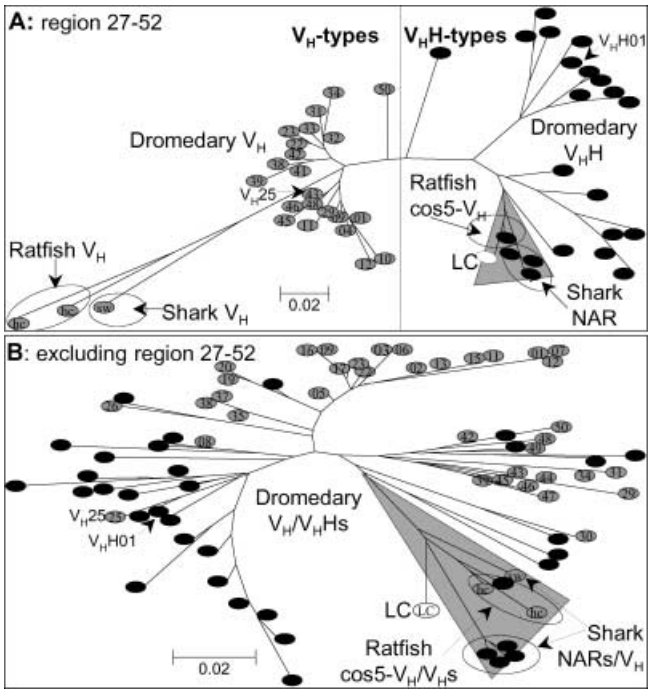
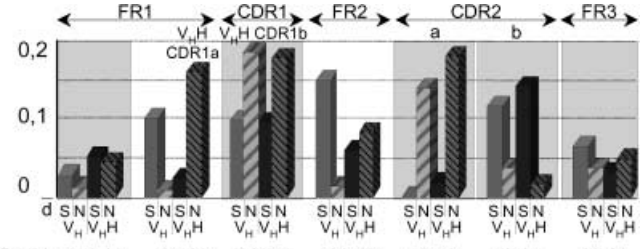


Fig. 4 Phylogenetic (ME) trees of *V* subregions containing amino acid residues 27–52 (IMGT residues 28–57, Lefranc 2001) (**A**) and of subregions excluding residues 27–52 (**B**) for 83 *V* segments of three species that have HCAb-types. Identical subregion sequences were removed prior to the tree reconstruction. The number of unique subdomains is 46 in (**A**) and 76 in (**B**). The V_HH and V_H genes are shown with *black* and *gray ovals*, respectively. Branch lengths are proportional to Poisson corrected distances (PC pairwise distances), estimated by MEGA2 (option: pairwise deletion), and within the *shaded triangle* they are scaled down tenfold for clarity reasons. All sequence accession numbers are shown in the text except for V_{NAR}*2g11* (L38965) and V_{NAR}*2 jiir5* (L38967)

this V_H-V_HH segregation was lost, and the V subdomains became intermingled if this V_HH-specific region was omitted (Fig. 4B).

These evolutionary patterns are compatible with the hypothesis that multiple dromedary V_H genes evolved into V_HH genes by introducing an internal segment comprising codons 27–52. Several pairs of highly divergent dromedary V_Hs and V_HHs can exhibit a close homology when the V_HH-specific region is excluded (Fig. 4). As an example, the 27–52 region of V_HH01 and V_H25 (separated in Fig. 4A) differs by nine amino acid replacements, whereas only one replacement was found for the remaining 69 sites (paired in Fig. 4B). This indicates that the monophyletic clustering of all camelid V_HHs in Fig. 2B (cluster d) is merely due to the *FR2* (within the V_HH-specific region). It appears, therefore, that all dromedary V_HHs were not the result of amplification from a single V_HH ancestor, but rather several distinct V_Hs might have evolved separately into V_HHs. However, we did not find independent evidence for the hypothesis that this subregion could originate from a dormant primordial V_HH-like gene segment. There are no V_H segments bearing an Arg/Cys codon at position 45 in 312 reported pseudogenes from the EMBL and IMGT databases.



Residues		1-26	27-30	31-35	36-49	50-58	59-65	66-92
IMGT N°		1-27	28-31	32-36	41-54	55-66	67-74	75-104
$\overline{dN}/\overline{dS}$	V_H	0.448	0.096	1.873	0.098*	-*	0.328	0.576
	V_HH	0.870	6.56*	1.848	1.387	9.736*	0.109*	1.515
	V_HH/V_H	1.942	**68.22**	0.987	**14.15**	-	0.332	2.632

Fig. 5. Non-synonymous ($\overline{dN}$) and synonymous ($\overline{dS}$) substitutions for *FR* and *CDR* regions (Kabat's numbering and the corresponding IMGT-FR/CDR numbering are indicated) of 43 V_H and 32 V_HH dromedary *V*-segments. The ratios of the $\overline{dN}$ over $\overline{dS}$ for all regions of V_H and V_HH is given below the graph. The last line gives the ratio of dN/dS of the V_H divided by that of the V_HH, of which the number in bold reflect a large difference in substitution pattern between V_H and V_HH. The *asterisk* indicates that $\overline{dN} > \overline{dS}$ or $\overline{dS} > \overline{dN}$ are statistically significant ($P \leq 5\%$).

Interestingly, internal palindromic sequences were found in the V_HH sequences at a high frequency at the boundaries of the 27–52 region (Nguyen et al. 2000). They may possibly act as recombination hotspots and thereby have contributed to the conversion of V_H into V_HH genes and their expansion. However, such a possible gene conversion mechanism did not result in the homogenization of the V_HH genes. On the contrary, the specific V_HH regions are diverse (Fig. 4A), indicating the presence of diversifying forces.

Ag-binding loop region

In the dromedary, indicators of different selective pressures operating on the V_HH or V_H subregions can be assessed by calculating the average synonymous ($\overline{dS}$) and non-synonymous ($\overline{dN}$) substitution rates. $\overline{dS}$ values that are significantly larger than $\overline{dN}$ values are evidence for purifying selection at this region, whereas the opposite situation argues for positive selection; i.e., for forces that favor genetic change or variation (Tanaka and Nei 1989).

In this analysis, we first identified subregions in FRs and CDRs according to their contribution to antigen specificity. Residues 27–30 of V_HH, but not of V_H, are involved in antigen interaction (Decanniere et al. 1999): they were therefore defined as the *V_HH-CDR1a* subregion. The *V_HH-CDR1b* (31–35) conforms to the canonical *CDR1* of the conventional V_Hs. Because only residues 50–58 of this region interact with antigen (Padlan 1996), the *CDR2* was by analogy divided into *CDR2a* (50–58) and *CDR2b* (59–65) (Tomlinson et al. 1992; Rothenfluh et al. 1995). The $\overline{dN}$ and $\overline{dS}$ values for these different subregions of V_H and V_HH are shown in Fig. 5.

The $\overline{dN}/\overline{dS}$ ratios of the V_HH and V_H were similar, except at the *V_HH-CDR1a* and *FR2* regions, where the $\overline{dN}/\overline{dS}$ ratios were much higher among V_HH genes (5.0 and 1.2) than among V_H genes (0.1 and 0.1). The increased $\overline{dN}$ at the *CDR1a* of the V_HH is obviously related to its Ag-binding function and provides strong evidence for adaptations at the germline level. The excess of $\overline{dN}$ over $\overline{dS}$ observed for the *V_HH-FR2*, suggests that this region is also subjected to positive selection. Clearly, evolutionary changes at these V_HH-specific regions are allowed and possibly selected for. We also noted that the *V_HH-CDR1* has one residue less than the *IMGT-CDR1*, and the *CDR2a* shifts from the *IMGT-FR2* by one residue (Lefranc 2001; Fig. 5). Similar studies are yet not feasible for the NARs and the *Cos5*-Abs due to the limited data on germline sequences.

Discussion

Camelidae HCAbs are derived from the conventional antibodies

Standard phylogenetic inference methods applied to vertebrate V_H and C_H regions allow a more precise evaluation of the hypothesis that HCAbs of *Camelidae* species are – entirely or in part – derived from remnants of a putative primordial HCAb form. Our results do not support this hypothesis. On the contrary, they suggest that HCAbs are a relatively recent adaptation of the conventional heteromeric antibodies.

Constant region

The tree topologies of the HCAb *C* genes and their domains revealed three monophyletic clusters consisting respectively of the NAR/NARC, the *Cos5*-Ab/IgM and the *Camelidae* HCAbs/mammalian IgGs. The fact that the *C_HH* genes, dromedary *γ2a* and llama *γ2c*, show more similarity to conventional C_H genes of camelids than to that of bovines, constitutes the first evidence that these *C_HH*-specific genes emerged within the suborder of Tylopoda, after they diverged from other Artiodactyla.

Previous reports (Greenberg et al. 1996; Rast et al. 1998; Nguyen et al. 1999; Woolven et al. 1999) have shown that all currently known HCAbs lack a typical C_H1 domain. However, the genetic bases for its absence in HCAbs differ fundamentally among species (Fig. 1). In the ratfish *Cos5*-Abs, *C_H1* is substituted by a C_H2-like domain, while in HCAbs of camelids, it is spliced out during mRNA maturation. The fact that different mechanisms were used for omitting the equivalent of a typical C_H1 domain strongly suggests that at least the constant regions of HCAbs evolved independently in different taxa. The universal absence of the typical C_H1 domain among the HCAbs furthermore points towards the obligate avoidance in HCAbs of this particular C_H domain.

Variable region

The large distances between the three V_HH branches among the vertebrate V_Hs (Fig. 2B, nodes a–c) sufficiently show that the *V* genes of the HCAbs have also evolved independently within the different evolutionary lineages.

A relatively recent origin of the camel V_HH is implied by the monophyly of V_HHs and V_Hs of camelids (Fig. 2B, node a). Therefore, evolutionary patterns of both variable and the constant regions favor the hypothesis that the HCAbs in *Camelidae* species originated only relatively recently from the conventional antibodies. Their origin most likely occurred after the divergence of Tylopoda from other Artiodactyla suborders and prior to the C*amelus* and *Lama* split. Similarly, the ratfish *Cos5-*V_HH type also arose from a conventional V_H ancestral gene (Fig. 2B, node d). However, the same could not be said for the NAR-V_H (Fig. 2B, node d) that misses conventional V_H counterpart from currently available data. Nevertheless, the monophyly NAR-V_Hs could be interpreted as representing the most ancient and divergent evolutionary V_H genes (Rast and Litman 1998; Richards and Nelson 2000).

Emergence of HCAbs from conventional antibodies

Although homologues of conventional antibodies are still undocumented in the jawless vertebrates, the presence of heterodimeric antibodies in all jawed vertebrates demonstrates that they were established very early in vertebrate evolution (Litman et al. 1999). The homodimeric isotypes have been documented so far in three unrelated lineages, camelids, ratfish, and nurse shark. In each lineage, HCAbs and conventional antibodies form a dichotomy Ab/HCAb system, which is apparently the outcome of more recent adaptive changes occurring in the compartment of the heteromeric antibodies. Although in the primitive vertebrates, the chances that their HCAbs originated from the primordial homodimeric Ig prototype are greater (Roux et al. 1998; Litman et al. 1999), the evolutionary inference reveals that also in these species at least the *C* regions of these HCAbs were derived from genes encoding conventional antibodies. Furthermore, the fact that the *Camelidae* HCAbs did emerge from conventional antibodies makes it appear more likely that the cartilaginous HCAbs are also derived from the conventional Ab genes

The emergence of HCAbs gave rise to a new generation of Ag receptors, which shows that diversification of the Ag-binding repertoire is an ongoing evolutionary process. This new level of diversification requires the co-evolutionary changes at both C_H and V_H regions. The removal or the reshaping of the typical C_H1 domain has likely played a key role in converting a classical Ab into a HCAb. Indeed, the C_H1 domain of the former binds to a chaperon protein (BIP), which leads to the retention of the H-chain in the endoplasmic reticulum (Hendershot et al. 1987). As a result, the replacement of the BIP by the L-chain is subsequently required to finalize the folding of the H polypeptides and to proceed towards their secretion as immunoglobulins. The absence of the C_H1 domain enables the H-chain to bypass the BIP capturing, making the rescue by the L chain unnecessary.

The conservation of the V_H*-FR2* region throughout vertebrate evolution is likely related to the critical role of this region in the V_H-V_L interaction (Hsu and Steiner 1992; Padlan 1994; Nuttall et al. 2000). In the conventional Ab system, the failure of V_H-V_L dimerization in H-chain producing B cells can prevent further B cell maturation (Kohler 1980; Keyna et al. 1995). Of the twelve amino acids that are the key in V_H-V_L interaction (Padlan 1994), four differ consistently between V_HH and V_H genes and provide the signature of the HCAbs. These substitutions have three effects. First, they reshape the surface structurally and chemically, impairing the association with a V_L and rendering the isolated domain soluble. Second, they presumably abrogate the anchoring site of a V_H domain by BIP in the endoplasmic reticulum (Vu et al. 1997). Third, they provide a landing place for the long *CDR3* (Desmyter et al. 1996; Decanniere et al. 1999) and therefore can dictate the *CDR3* organization (Muyldermans et al. 2001).

New level of Ab diversification and creation of novel Ag receptor properties

The need for the immune system to establish a highly specific system to discriminate self and non-self has resulted in the evolution of several mechanisms aimed at amplifying and diversifying the Ag-binding repertoire. In addition to combinatorial V_H–V_L variation, they include more advanced mechanisms such as V(D)J recombination, gene conversion, and somatic hypermutation (Tonegawa 1983; Knight and Becker 1990; Reynaud et al. 1994). The emergence of HCAbs, where Ag-binding activity is conveyed by a single domain V_HH, represents another way of Ab diversification. Once HCAbs were established in the dromedary, it seems that their Ag-binding repertoire was increased in the first place by the expansion of *V* gene segments having acquired V_HH characteristics. The number and the high degree of sequence homology of the V_HH genes (more than forty are identified so far) indicate that they have undergone recurrent duplications while “new” V_HH genes could be generated by gene exchange among V_HH and V_H genes. In addition to gene expansion and conversion, more sophisticated mechanisms tend to enlarge the Ag-binding repertoire of the HCAbs. These involve hypermutation hotspots (Milstein et al. 1998) embedded in the V_HH-specific *CDR1a* region (Nguyen et al. 2000) and the acquisition of a non-canonical cysteine in the CDRs and *FR2*. These non-canonical cysteines most likely evolved posterior to the V_HH emergence in the camelids because they occupy different positions in llama and camel V_HHs (Vu et al. 1997; Nguyen et al. 2000).

The co-existence of conventional antibodies and HCAbs in *Camelidae* suggests that they could exert different functions. Indeed, experimental evidence (Lauwereys et al. 1998) showed that the camel HCAbs recognize epitopes that are barely antigenic for conventional antibodies, such as active sites of enzymes. In addition, studies on recombinant Ab fragments revealed that V_HHs are more stable than V_H-V_L pairs (Chothia 1991), probably aided by the additional disulfide bond tethering the Ag-binding loops (Muyldermans et al. 1994; Desmyter et al. 1996; Vu et al. 1997; Nuttall et al. 2000Nguyen et al. 2000). These observations indicate that the camel HCAbs might be selected over conventional antibodies to fulfill a complementary function.

The ongoing pressure for change at the Ag-binding site can therefore explain the high dN/dS ratios observed at *CDR1a*. By analogy, the nucleotide substitution patterns at *V_HH-FR2* regions may also reflect their contribution to Ag-binding. Indeed, the crystal structure of recombinant V_HHs in complex with their antigen proved that the FR2 residues could determine the shape of the CDR3 (Muyldermans et al. 2001). In conclusion, it appears that more residues are involved in Ag-binding in V_HHs than in classical V_Hs, and this may represent a novel way of diversifying the V_HH repertoire. Rather than relying upon the contribution of the V_L polypeptide, the strategy consists in recruitment of more residues (larger CDR1 and CDR3) of the same peptide chain for Ag-binding, and this includes residues that are committed to the V_L interaction in conventional antibodies. The forces favoring repertoire diversification have apparently utilized the opportunities created by the loss of the L-chain. The fact is that this loss has initiated the innovative developments at the Ag-binding site. It appears that, at least in *Camelidae*, the HCAbs are not the reactualization of a primitive form of antibodies. On the contrary, in some aspects, these V_HHs show a higher structural complexity than the V_Hs from which they were derived.

Acknowledgements We are grateful to Dr. Masatoshi Nei for supervision of this work in his laboratory. We thank Dr. K. Decanniere (VUB) and members of IMEG (Pennsylvania State University) for fruitful discussions. This work was supported by research grants from the National Institute of Health to M.N., and the Fonds voor Wetenschappelijk Onderzoek (Vlaanderen).

References

Brodeur PH, Riblet R (1984) The immunoglobulin heavy chain variable region (*Igh-V*) locus in the mouse. I. One hundred *Igh-V* genes comprise seven families of homologous genes. Eur J Immunol 14:922–930

Chothia C (1991) Antigen recognition. Curr Opin Struct Biol 1:53–59

Chothia C, Lesk AM, Tramontano A, Levitt M, Smith-Gill SJ, Air G, Sheriff S, Padlan EA, Davies D, Tulip WR (1989) Conformations of immunoglobulin hypervariable regions. Nature 342:877–883

Decanniere K, Desmyter A, Lauwereys M, Ghahroudi MA, Muyldermans S, Wyns L (1999) A single-domain antibody fragment in complex with RNase A: non- canonical loop structures and nanomolar affinity using two CDR loops. Structure 7:361–370

Desmyter A, Transue TR, Ghahroudi MA, Thi MH, Poortmans F, Hamers R, Muyldermans S, Wyns L (1996) Crystal structure of a camel single-domain V_H antibody fragment in complex with lysozyme. Nat Struct Biol 3:803–811

Greenberg AS, Avila D, Hughes M, Hughes A, McKinney EC, Flajnik MF (1995) A new antigen receptor gene family that undergoes rearrangement and extensive somatic diversification in sharks. Nature 374:168–173

Greenberg AS, Hughes AL, Guo J, Avila D, McKinney EC, Flajnik MF (1996) A novel "chimeric" antibody class in cartilaginous fish: IgM may not be the primordial immunoglobulin. Eur J Immunol 26:1123–1129

Hamers-Casterman C, Atarhouch T, Muyldermans S, Robinson G, Hamers C, Songa EB, Bendahman N, Hamers R (1993) Naturally occurring antibodies devoid of light chains. Nature 363:446–448

Hendershot L, Bole D, Kohler G, Kearney JF (1987) Assembly and secretion of heavy chains that do not associate post-translationally with immunoglobulin heavy chain-binding protein. J Cell Biol 104:761–767

Hsu E, Steiner L (1992) Primary structure of immunoglobulins through evolution. Curr Opin Struct Biol 2:422–431

Kabat EA, Wu TT, Perry HM, Gottesman KS, Foeller C (1991) Sequences of proteins of immunological interest. US Public Health Service, NIH, Besthesda, Md.

Keyna U, Beck-Engeser GB, Jongstra J, Applequist SE, Jack HM (1995) Surrogate light chain-dependent selection of Ig heavy chain V regions. J Immunol 155:5536–5542

Knight KL, Becker RS (1990) Molecular basis of the allelic inheritance of rabbit immunoglobulin V_H allotypes: implications for the generation of antibody diversity. Cell 60:963–970

Kohler G (1980) Immunoglobulin chain loss in hybridoma lines. Proc Natl Acad Sci U S A 77:2197–2199

Kumar S, Tamura K, Nei M (1994) MEGA: Molecular evolutionary genetics analysis software for microcomputers. Comput Appl Biosci 10:189–191

Lauwereys M, Ghahroudi MA, Desmyter A, Kinne J, Holzer W, De Genst E, Wyns L, Muyldermans S (1998) Potent enzyme inhibitors derived from dromedary heavy-chain antibodies. EMBO J 17:3512–3520

Lefranc MP (2001) IMGT, the international ImMunoGeneTics database. Nucleic Acids Res 29:207–209

Litman GW, Anderson MK, Rast JP (1999) Evolution of antigen binding receptors. Annu Rev Immunol 17:109–147

Milstein C, Neuberger MS, Staden R (1998) Both DNA strands of antibody genes are hypermutation targets. Proc Natl Acad Sci USA 95:8791–8794

Muyldermans S, Lauwereys M (1999) Unique single-domain antigen binding fragments derived from naturally occurring camel heavy-chain antibodies. J Mol Recognit 12:131–140

Muyldermans S, Atarhouch T, Saldanha J, Barbosa JA, Hamers R (1994) Sequence and structure of V_H domain from naturally occurring camel heavy chain immunoglobulins lacking light chains. Protein Eng 7:1129–1135

Muyldermans S, Cambillau C, Wyns L (2001) Recognition of antigens by single-domain antibody fragments: the superfluous luxury of paired domains. Trends Biochem Sci 26:230–235

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and non-synonymous nucleotide substitutions. Mol Biol Evol 3:418–426

Nei M, Kumar S (eds) (2000) Evolutionary change of amino acid / DNA sequences. In: Molecular evolution and phylogenetics. Oxford University Press, New York, p 17

Nguyen VK, Muyldermans S, Hamers R (1998) The specific variable domain of camel heavy-chain antibodies is encoded in the germline. J Mol Biol 275:413–418

Nguyen VK, Hamers R, Wyns L, Muyldermans S (1999) Loss of splice consensus signal is responsible for the removal of the entire C_H1 domain of the functional camel IgG2 A heavy-chain antibodies. Mol Immunol 36:515–524

Nguyen VK, Hamers R, Wyns L, Muyldermans S (2000) Camel heavy-chain antibodies: diverse germline V_HH and specific mechanisms enlarge the antigen-binding repertoire. EMBO J 19:921–930

Nguyen VK, Desmyter A, Muyldermans S (2001) Functional heavy-chain antibodies in *Camelidae*. Adv Immunol 79:261–296

Nuttall SD, Irving RA, Hudson PJ (2000) Immunoglobulin V_H domains and beyond: design and selection of single-domain binding and targeting reagents. Curr Pharm Biotechnol 1:253–263

Ota T, Nei M (1994) Divergent evolution and evolution by the birth-and-death process in the immunoglobulin V_H gene family. Mol Biol Evol 11:469–482

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

Padlan EA (1996) X-ray crystallography of antibodies. Adv Protein Chem 49:57–133

Rast JP, Litman GW (1998) Towards understanding the evolutionary origins and early diversification of rearranging antigen receptors. Immunol Rev 166:79–86

Rast JP, Amemiya CT, Litman RT, Strong SJ, Litman GW (1998) Distinct patterns of IgH structure and organization in a divergent lineage of chrondrichthyan fishes. Immunogenetics 47: 234–245

Reynaud CA, Bertocci B, Dahan A, Weill JC (1994) Formation of the chicken B-cell repertoire: ontogenesis, regulation of Ig gene rearrangement, and diversification by gene conversion. Adv Immunol 57:353–378

Richards MH, Nelson JL (2000) The evolution of vertebrate antigen receptors: a phylogenetic approach. Mol Biol Evol 17:146–155

Rothenfluh HS, Blanden RV, Steele EJ (1995) Evolution of *V* genes: DNA sequence structure of functional germline genes and pseudogenes. Immunogenetics 42:159–171

Roux KH, Greenberg AS, Greene L, Strelets L, Avila D, McKinney EC, Flajnik MF (1998) Structural analysis of the nurse shark (new) antigen receptor (NAR): molecular convergence of NAR and unusual mammalian immunoglobulins. Proc Natl Acad Sci U S A 95:11804–11809

Rzhetsky A, Nei M (1992a) A simple method for estimating and testing minimum-evolution trees. Mol Biol Evol 9:945–967

Rzhetsky A, Nei M (1992b) Statistical properties of the ordinary least-squares, generalized least-squares, and minimum-evolution methods of phylogenetic inference. J Mol Evol 35:367–375

Rzhetsky A, Nei M (1993) Theoretical foundation of the minimum-evolution method of phylogenetic inference. Mol Biol Evol 10:1073–1095

Schwager J, Burckert N, Courtet M, Du PL (1989) Genetic basis of the antibody repertoire in *Xenopus*: analysis of the V_H diversity. EMBO J 8:2989–3001

Sitnikova T, Rzhetsky A, Nei M (1995) Interior-branch and bootstrap tests of phylogenetic trees. Mol Biol Evol 12:319–333

Stoesser G, Baker W, van den Broek A, Camon E, Garcia-Pastor M, Kanz C, Kulikova T, Lombard V, Lopez R, Parkinson H, Redaschi N, Sterk P, Stoehr P, Tuli MA (2001) The EMBL nucleotide sequence database. Nucleic Acids Res 29:17–21

Symons DB, Clarkson CA, Beale D (1989) Structure of bovine immunoglobulin constant region heavy chain *gamma1* and *gamma2* genes. Mol Immunol 26:841–850

Tanaka T, Nei M (1989) Positive Darwinian selection observed at the variable-region genes of immunoglobulins. Mol Biol Evol 6:447–459

Thompson JD, Higgins DG, Gibson TJ (1994) Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Tomlinson IM, Walter G, Marks JD, Llewelyn MB, Winter G (1992) The repertoire of human germline V_H sequences reveals about fifty groups of V_H segments with different hypervariable loops. J Mol Biol 227:776–798

Tonegawa S (1983) Somatic generation of antibody diversity. Nature 302:575–581

Vu KB, Ghahroudi MA, Wyns L, Muyldermans S (1997) Comparison of llama V_H sequences from conventional and heavy chain antibodies. Mol Immunol 34:1121–1131

Woolven BP, Frenken LG, van der Logt P, Nicholls PJ (1999) The structure of the llama heavy chain constant genes reveals a mechanism for heavy-chain antibody formation. Immunogenetics 50:98–101