Immunogenetics (2002) 54:353–364
DOI 10.1007/s00251-002-0458-4

ORIGINAL PAPER

**Bettina Wagner · Irene Greiser-Wilke**
**Anja K. Wege · Andreas Radbruch**
**Wolfgang Leibold**

# Evolution of the six horse *IGHG* genes and corresponding immunoglobulin gamma heavy chains

Received: 14 January 2002 / Accepted: 4 April 2002 / Published online: 4 July 2002


**Abstract** It is generally assumed that the different mammalian IgG isotypes have developed during evolution by duplications of a common ancestor gamma heavy chain constant region gene (*IGHG*). In contrast to other species studied so far, which express between one and four *IGHG* genes, the horse (*Equus caballus*) genome contains six *IGHG* genes, and it has been postulated that they all can be expressed. For determination of the evolutionary history of the six horse *IGHG* genes, genomic DNA and cDNA of the *IGHG* genes were sequenced. The structure of these genes with reference to exons and introns was determined. Comparison of the deduced amino acid sequences of the horse *IGHG* genes revealed the greatest divergences in the hinge regions, and in the proximal CH2 domains. A phylogenetic comparison of the amino acid sequences of the six horse *IGHG* genes to those of other species shows that the horse *IGHG* genes form a distinct cluster. This indicates that the mammalian species included in this study probably share only one common ancestor *IGHG* gene with the horse. The six horse *IGHG* genes probably then evolved by gene duplication after species separation. In addition, various segmental exchanges were found between the horse *IGHG* genes, which might be the result of unequal crossing over and/or gene conversion events during the evolution of the six horse *IGHG* genes.

**Keywords** *Equus caballus IGHG* genes · Horse immunoglobulin gamma heavy chains · Horse *IGHG* evolution · Segmental homology · FcγR binding

The nucleotide sequence data reported in this paper have been assigned the EMBL/GenBank accession numbers: AJ300675, AJ302055, AJ302056, AJ302057, AJ302058, AJ312379, AJ312380 and AJ312381.

B. Wagner (✉) · A.K. Wege · W. Leibold
Immunology Unit, School of Veterinary Medicine,
Bischofsholer Damm 15, 30173 Hannover, Germany

I. Greiser-Wilke
Institute of Virology, School of Veterinary Medicine,
Bünteweg 17, 30559 Hannover, Germany

A. Radbruch
German Rheumatism Research Center, Schumann Strasse 21/22,
10117 Berlin, Germany

*Present address*:
B. Wagner, James A. Baker Institute for Animal Health,
College of Veterinary Medicine, Cornell University,
Hungerford Hill Road, Ithaca, NY 14853, USA
e-mail: bw73@cornell.edu,
Tel: +1-607-2565615, Fax: +1-607-2565608

## Introduction

Antibodies can mediate different effect or functions during an immune response, depending on their antibody class. Some can activate the classical pathway of the complement cascade, when bound to antigen (Duncan and Winter 1988), others interact with Fcγ receptors (FcγR) on various cells of the immune system, mediating responses like phagocytosis, antibody-dependent cytotoxicity or immunoregulatory signaling of proliferation and antibody secretion. These interactions are mediated by the constant region domains (isotypes) of the antibody heavy chains (for review: Ravetch and Kinet 1991; Sensel et al. 1997).

The functional properties of the different IgG isotypes have been intensively studied for the antibody classes of mice and man. In the last 10 years, it also became evident that in other mammalian species the numbers of IgG isotypes vary considerably, depending on the number of functional *IGHG* genes, ranging from one in the rabbit (Knight et al. 1985), three in cattle (Knight et al. 1988), four in human (Flanagan and Rabbitts 1982), mouse (Shimizu et al. 1982) and rat (Brüggemann et al. 1986), five in the pig (Butler and Brown 1994), up to six in the horse (Wagner et al. 1998). It could be speculated that the increase in the number of IgG isotypes might be accompanied by an increase in functional diversity.

It has been postulated that the IgG isotypes had developed during the evolution of mammalian species by several duplication events, resulting in the amplification of one common ancestral *IGHG* gene. The functional diver-

gence of the *IGHG* genes might then have happened after their duplication (Honjo and Matsuda 1995). A strong evolutionary relationship between mouse and rat *IGHG* genes is suggested by the existence of three common ancestor *IGHG* genes in these species (Brüggemann 1988). In addition to the four *IGHG* genes, which were described in BALB/c mice (Shimizu et al. 1982), duplicated *IGHG2* genes were found in other mouse strains (Fukui et al. 1984; Morgado et al. 1989; Martin et al. 1997). Thus, the number of mouse *IGHG* genes varies from four to five. During intraspecies evolution of the human *IGHG* genes, duplication of an *IGHG-IGHG-IGHE-IGHA* fragment must have occurred (Flanagan and Rabbitts 1982). In addition, an open reading frame (*IGHGP*) is located between these two "G-G-E-A" clusters (Bensmana et al. 1988). Thus, the usual number of *IGHG* genes in human is five per haploid genome, but it may vary from one (Lefranc et al. 1982) to nine genes, due to polymorphic multiple gene deletions, duplications and triplications described in the *IGH* locus of healthy individuals (Lefranc and Lefranc 2001). In rabbits, with no duplication of the unique *IGHG* gene, duplications of the ancestral *IGHA* gene created 14 different *IGHA* genes (Knight et al. 1985; Volgina et al. 2000).

The six horse (*Equus caballus*) *IGHG* genes have been isolated and shown to be aligned on the horse antibody heavy chain gene locus in the order: 5′ *IGHM* // *IGHG1* // *IGHG2* // *IGHG3* // *IGHG4* // *IGHG5* // *IGHG6* // *IGHE – IGHA* 3′ (Wagner et al. 1998). Here, we describe the nucleotide and predicted amino acid sequences of the six horse *IGHG* genes and compare their evolutionary relationship to other mammalian species.

The sequence data also clarify whether all horse *IGHG* genes are functional or whether pseudogenes exist. Until now, there was a discrepancy between the genetic approach, clearly indicating the existence of six *IGHG* genes in the horse and the results of serological and biochemical studies, suggesting the existence of only five expressed horse IgG isotypes. Most of the work on horse immunoglobulins was done 30–40 years ago, when horses were frequently used to produce hyperimmune sera for human vaccination. The five IgGs described in these early studies were designated as IgGa, IgGb, IgGc, IgG(T) and IgG(B). IgG(B) was also designated as aggregating immunoglobulin (AI; reviewed in: Montgomery 1973). Evidence for the existence of two different subclasses of IgG(T) was presented in a more recent study (Sheoran and Holmes 1996). Three of the horse IgG isotypes could be linked to their corresponding *IGHG* genes by deletion analysis of heterohybridomas secreting horse IgG: the *IGHG1* genes encodes for IgG1, which is the IgGa according to the old nomenclature, the *IGHG3* gene encodes for IgG3 [IgG(T)], and the *IGHG4* gene encodes for IgG4 [IgGb] (Wagner et al. 1998). The nucleotide sequence data of the remaining *IGHG2*, *IGHG5* and *IGHG6* genes should clarify which of them encoded for complete constant heavy chain regions.

## Materials and methods

### Sequencing strategy

The nucleotide sequences of the horse *IGHG* genes originated from the domestic horse (*Equus caballus*). To identify the exon and intron structure of the horse *IGHG* genes, both genomic DNA and cDNAs of the corresponding mRNAs were analysed. DNA sequences were obtained by automated sequencing (SEQ LAB, Göttingen, Germany).

### Sequencing of the genomic *IGHG* genes

The genomic DNA sequences comprised the cloned DNA fragments γ1–74, γ2–92, γ3–91, γ4–13, γ5–51, γ5–72 and γ6–22 (Overesch et al. 1998), as well as the genomic 4.6 kb *Bam*HI *IGHG6* fragment of the heterohybridoma EqL2 (Wagner et al. 1998). All cloned genomic DNA fragments originated from an Arabian stallion; the heterohybridoma EqL2 was generated using peripheral blood mononuclear cells (PBMC) from an Icelandic horse.

Initially, several combinations of primers, derived from the conserved region of the human *IGHG* genes, were tested for their ability to amplify the horse genomic *IGHG1* gene by PCR. Using the horse 1.8 kb *Nco*I *IGHG1* fragment (Wagner et al. 1998) with a human *IGHG4* sense primer from the CH2 exon and a human *IGHG1* antisense primer from the CH3 exon (5′ AGACAGGGAGAGGCTCTTCTGC 3′) a horse 650 bp *Pfu* PCR fragment was obtained. Nucleotide sequence analysis showed that the human *IGHG1* primer had been extended by PCR from both ends of the 650 bp PCR product. The product consisted of the 3′ part of the CH2 exon, the CH2-CH3 intron and the CH3 exon of the horse *IGHG1* gene. Subsequently, horse primers were selected from this sequence and used for further sequencing of the genomic horse *IGHG* genes as well as the *IGHG* cDNA. Additional sequencing was performed by gene walking (Fig. 1). The horse primers used are listed in Table 1.

### Sequencing of *IGHG* cDNA

Horse cDNA was generated from PBMC of three Icelandic horses (*IGHG2*, *IGHG5* and *IGHG6* genes) and from the heterohybridomas EqG3 (*IGHG1*), AFRC CR2 (*IGHG3*) and EqG2 (*IGHG4*; Wagner et al. 1998), obtained by fusion of PBMC of an Icelandic horse (Wagner et al. 1995), except for the AFRC CR2 cell line which originated from a Welsh Mountain pony (Richards et al. 1992). The cDNA was obtained by reverse transcription of mRNA using oligo(dT) as primer. The *IGHG* cDNA was amplified by PCR and cloned into the pCR™II vector (Invitrogen, Groningen, The Netherlands) as described previously (Wagner et al. 2001). Primers for PCR amplification were the same as described above for the sequencing of the genomic *IGHG* genes. The cDNA sequences were used to determine the coding regions of the corresponding genomic *IGHG* genes.

### Sequences used for phylogenetic analyses

For the phylogenetic comparison of immunoglobulin gamma chains from different species, the sequences with the following accession numbers were obtained from the GenBank: ambystoma (X69492); bovine *IGHG1* (X16701), *IGHG2* (X16702) and *IGHG3* (U63638); camel *IGHG2a* (AJ131945); chicken (X07174); chimpanzee *IGHG* (X65284, X65285, X61310, X61311); duck (X65219); human *IGHG1* (J00228), *IGHG2* (J00230), *IGHG3* (D78345) and *IGHG4* (K01316); macaca (rhesus monkey) *IGHG1* (AF045537), *IGHG2* (AF045539) and *IGHG3* (AF045538); mouse *IGHG1* (D78344), *IGHG2a* (V00825), *IGHG2b* (J00461) and *IGHG3* (D78343); sheep *IGHG1* (X69797) and *IGHG2* (X70983); rabbit *IGHG* (K00752); rat *IGHG1* (M28670), *IGHG2a* (M28669),

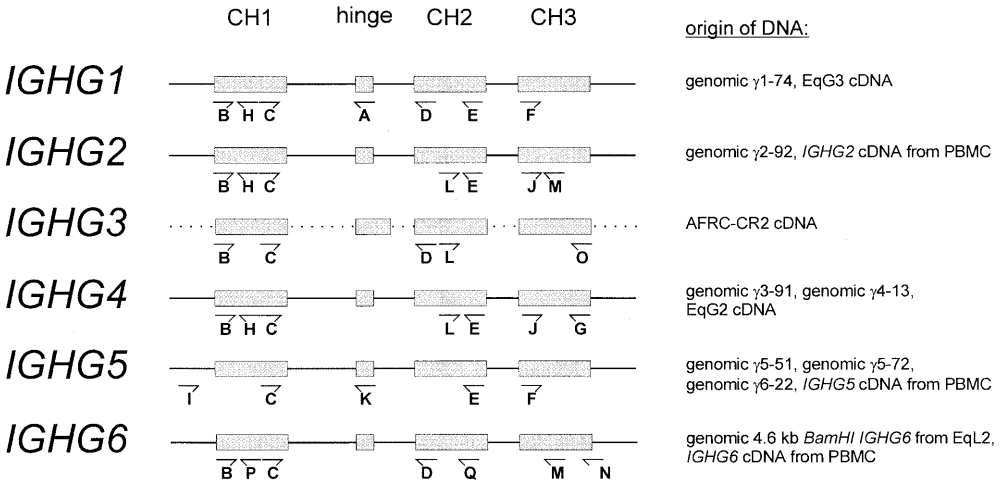


**Fig. 1** Sequencing strategy for the horse *IGHG* genes. Nucleotide sequences were obtained from genomic horse *IGHG* gene clones (Wagner et al. 1998). In addition, cDNAs of the respective *IGHG* genes from horse peripheral blood mononuclear cells (PBMC) and horse-mouse heterohybridoma cell lines (EqG3, AFRC-CR2, EqG2 and EqL2) were sequenced to determine the coding regions. The horse *IGHG* primers A–Q are described in detail in Table 1

**Table 1** Horse *IGHG* primers used for sequencing (*s* sense, *as* antisense)

| Primer | *IGHG* gene[a] | s/as | Primer sequence 5′→3′ |
|---|---|---|---|
| A | $IGHG1_{(607–578)}$ | as | CACTTTTTGGTGGTTGTCGGGAATGGGCTC |
| B | $IGHG1_{(1–21)}$ | s | GCCTCCACCACCGCCCCGAAG |
| C | $IGHG1_{(265–288)}$ | s | GCCAGCAACTTCAAGGTGGACAAG |
| D | $IGHG1_{(811–791)}$ | as | GGGGAAGATGAAGACCGAAGG |
| E | $IGHG1_{(979–956)}$ | as | GTAAGTGCTGTTGAACTGCTCCTC |
| F | $IGHG1_{(1189–1213)}$ | s | GTCCCAGGAGCCGCAAGTGTACGTC |
| G | $IGHG1_{(1513–1494)}$ | as[b] | CGCAAGCTTCATTTACCCGGGTTCTTGG |
| H | $IGHG4_{(82–63)}$ | as | GGCAGCCCAGGGCCACCGT |
| I | $IGHG4_{(-58\ –\ -32)}$ | s[c] | GACRCCCTSCCCGGACCSTCRCAGGG |
| J | $IGHG4_{(1146–1172)}$ | s | GGGCAGCCCCGGGAGCCGCAGGTGTAT |
| K | $IGHG5_{(588–565)}$ | as | GGAGCCTTTTACGACTGCAGAGAG |
| L | $IGHG5_{(856–876)}$ | s | CCCGATGTCCAGTTCACCTGG |
| M | $IGHG5_{(1241–1221)}$ | as | GACCAGGCAGGTCACGCTGAC |
| N | $IGHG2_{(1511–1491)}$ | as[d] | CCGGCGTGTGGCTCATTTACC |
| O | $IGHG5_{(1481–1457)}$ | as[b] | GCGTAAGCTTTACCCGGAGAGTGGGAGACGTTC |
| P | $IGHG6_{(294–272)}$ | as | GATTCTCTTGTCCACCTTGGTGC |
| Q | $IGHG6_{(1066–1045)}$ | as | TTTGGCCTTGGTGATGGTCCTC |

[a] Horse *IGHG* genes and sequence positions from which the primer sequences were selected are given in parentheses according to Fig. 2
[b] Primer contains a *Hin*dIII site
[c] Primer from the 5′UTR
[d] Primer contains 12 bases of the 3′UTR of the *IGHG2* gene

*IGHG2b* (M28671) and *IGHG2c* (X07189); swine *IGHG1* (U03778), *IGHG2a* (U03779), *IGHG2b* (U03780) and *IGHG3* (U03781); xenopus (X15114), and horse *IGHG1* (AJ302055), *IGHG2* (AJ302056), *IGHG3* (AJ312379), *IGHG4* (AJ302057), *IGHG5* (AJ312380), *IGHG6* (AJ312381), *IGHM* (L49414), *IGHE* (AJ305046), *IGKC* (X75612) and *IGLC* (L07563).

### Phylogenetic analysis

Sequence alignments and phylogenetic trees were calculated with the CLUSTAL X (version 1.8) analysis program (Thompson et al. 1997), using the default settings. For protein and DNA sequence alignments, the Blosum protein matrix and the ClustalW 1.6 DNA matrix, respectively, were used. The alignments were redefined in GeneDoc (version 2.6.002; Nicholas and Nicholas 1997) to maximize sequence homology. Phylogenetic trees were calculated using the neighbor joining algorithm of Saitou and Nei (1987). Confidence values were calculated by the bootstrapping method (Felsenstein 1985), as implemented in CLUSTAL X. Visualization of the phylogenetic trees, including the setting of outgroups, was performed using TREEVIEW, version 1.6.5 (Page 1996).

## Results

### Nucleotide sequences of the six horse *IGHG* genes

The nucleotide sequences of the horse *IGHG* genes were obtained from genomic DNA clones, with the exception of the *IGHG3* gene, which was derived from cDNA of the heterohybridoma AFRC CR2, expressing the horse *IGHG3* gene (Wagner et al. 1998). The alignment of the horse *IGHG* gene sequences is shown in Fig. 2. The coding regions and the intervening sequences (IVS) of each gene were determined according to the sequences of the corresponding cDNAs. The length of the coding sequences ranged from 981 bp (*IGHG5*) to 1,065 bp (*IGHG3*). The greatest differences in length were found in the hinge exons (Table 2). The *IGHG4*, *IGHG5* and *IGHG6* genes had short hinge regions of 33 bp or 36 bp; the hinge exons of the *IGHG1* and *IGHG2* genes were of intermediate length (60 bp and 54 bp, respectively), while the hinge exon of the *IGHG3* gene was notably longer (111 bp). Further variations in length were detected at the 5′ end of the CH2 exon. Compared to the

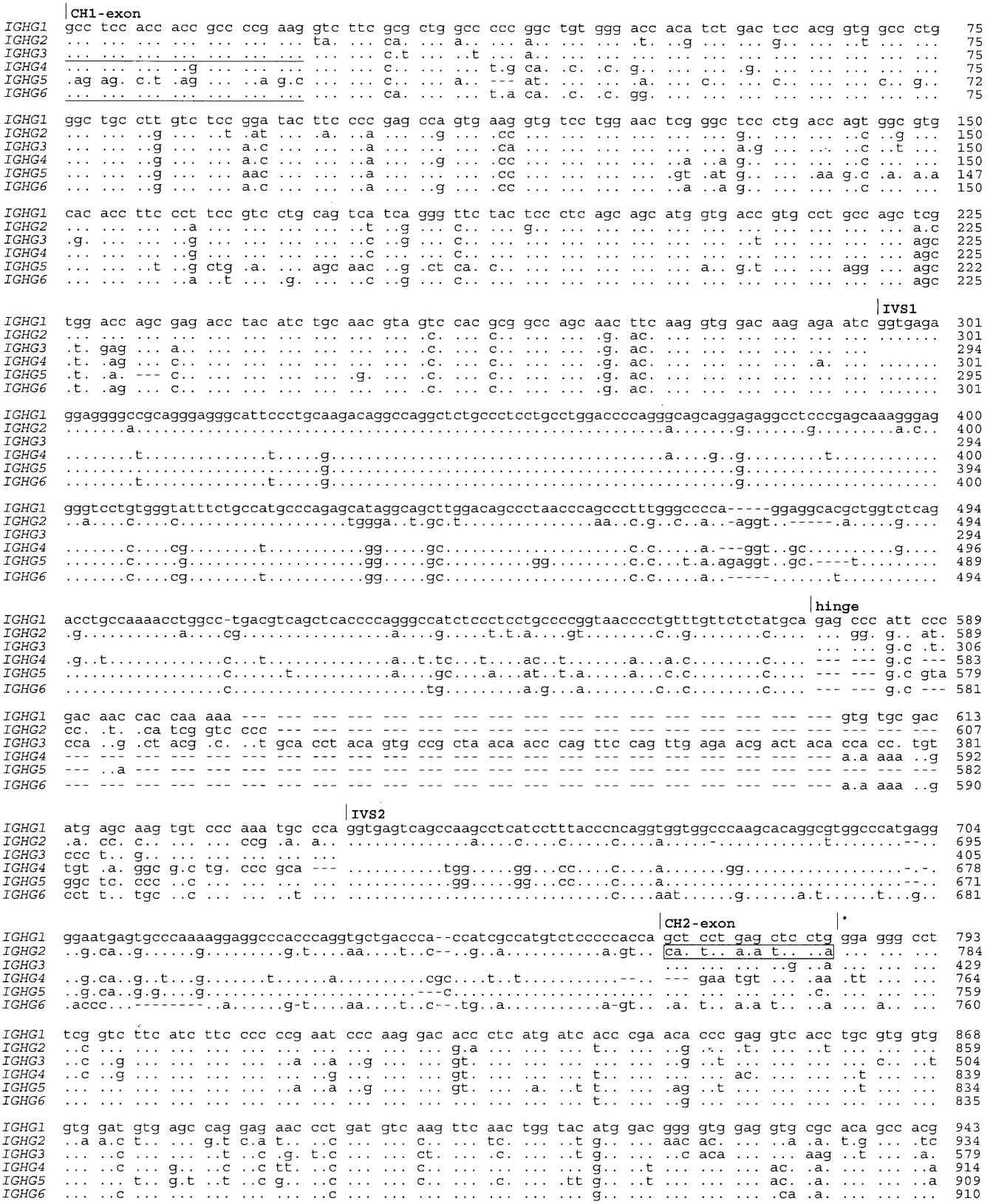


**Fig. 2**

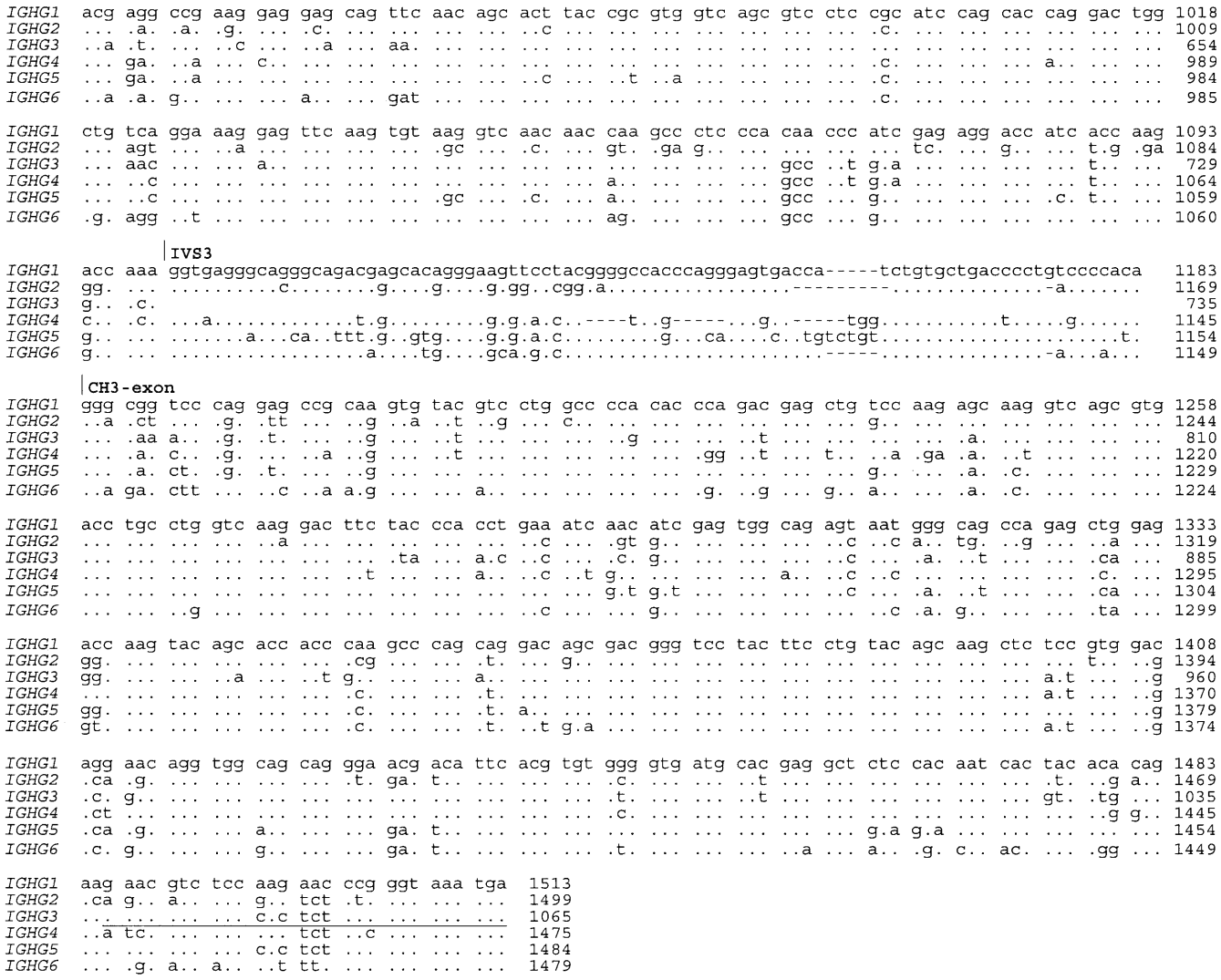


**Fig. 2** Alignment of the *Equus caballus IGHG1 – IGHG6* genes. The nucleotide sequences of the horse *IGHG* genes were obtained by gene walking along genomic clones, except for the *IGHG3* gene sequence, which was obtained from cDNA. Thus, the intervening sequences (IVS) could not be obtained for the *IGHG3* gene. Splicing sites are indicated by *vertical lines*. The beginning of the *IGHG2* CH2 exon is located 15 bp downstream relative to the respective *IGHG6* sequence, which is identical except for the first nucleotide (cytosine instead of guanine). Thus, the splicing site is not conserved in the *IGHG2* sequence and a cryptic splicing site (*asterisk*) is used. The resulting non-translated sequence (c761 to a775) at the beginning of the *IGHG2* CH2 exon is shown in a *box*. Sequences derived from primers are *underlined*

**Table 2** Length (bp) of the horse *IGHG* gene exons and intervening sequences (IVS)

| *IGHG* gene | Exon/intron | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | CH1 | IVS1 | Hinge | IVS2 | CH2 | IVS3 | CH3 | Coding sequence[a] |
| *IGHG1* | 294 | 283 | 60 | 132 | 330 | 84 | 330 | 1014 |
| *IGHG2* | 294 | 283 | 54 | 144 | 315 | 79 | 330 | 993 |
| *IGHG3* | 294 | ? | 111 | ? | 330 | ? | 330 | 1065 |
| *IGHG4* | 294 | 286 | 33 | 130 | 327 | 75 | 330 | 984 |
| *IGHG5* | 288 | 285 | 33 | 129 | 330 | 89 | 330 | 981 |
| *IGHG6* | 294 | 284 | 36 | 122 | 330 | 83 | 330 | 990 |

[a] Length (bp) of the coding sequences of the secreted forms of *IGHG* genes; the transmembrane exons have not yet been characterized

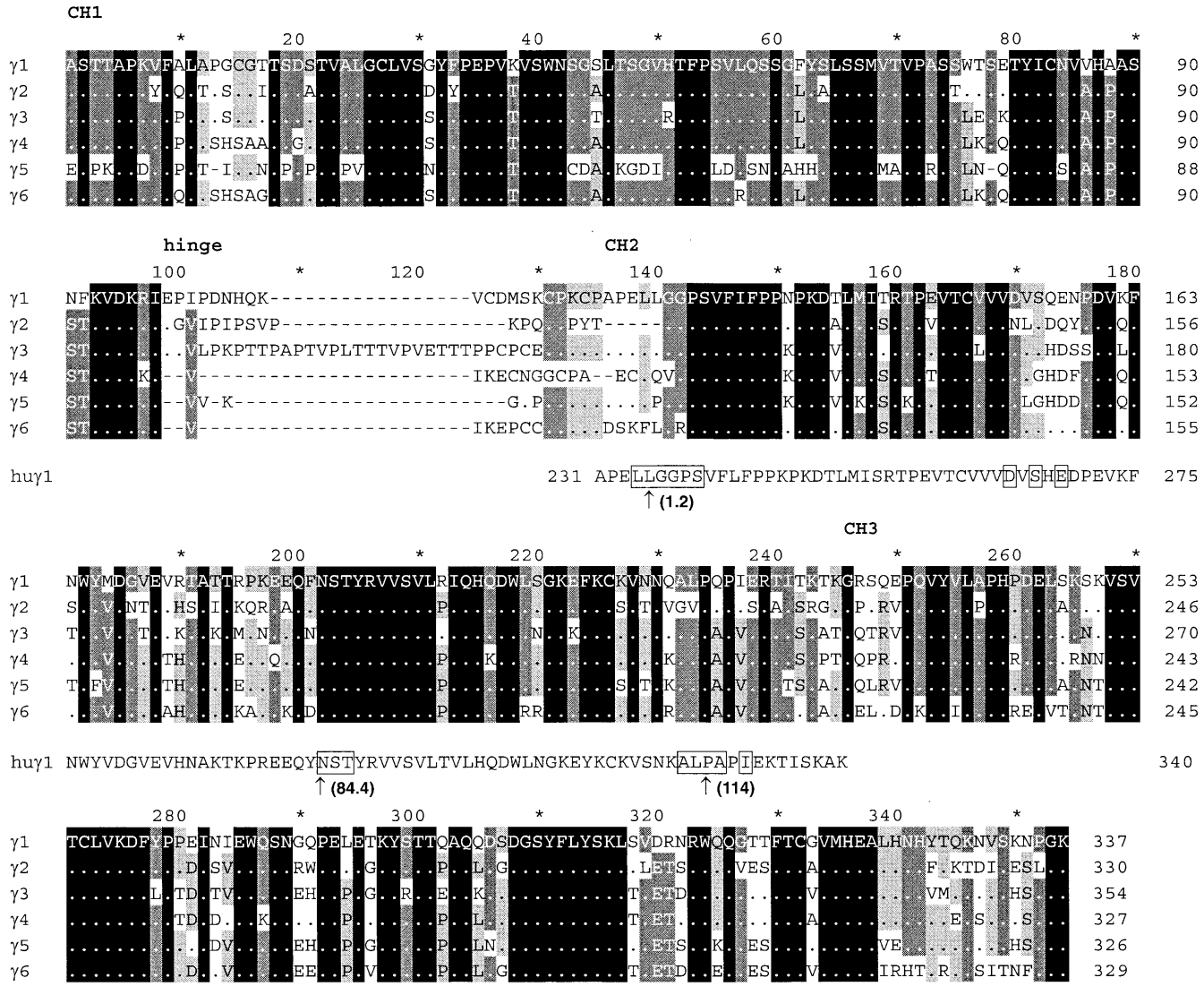


**Fig. 3** Comparison of the deduced amino acid sequences of the six horse *IGHG* genes. The horse immunoglobulin gamma heavy chains (γ1 to γ6) share between 67% and 78% overall homology. The "deletion" of five amino acids (DSKFL) at the beginning of the CH2 domain originated from a cryptic splicing site in the *IGHG2* CH2 exon, which is located 15 bp downstream compared to the CH2 exon of the *IGHG6* gene. In addition, the CH2 domain of the human gamma 1 heavy chain (huγ1) is shown. The amino acids involved in the binding of the CH2 domain of human IgG1 to the FcγRIII (Sondermann et al. 2000) were compared to the CH2 domains of the six horse gamma chains. The residues involved in the FcγRIII binding site of human IgG1 are *boxed*. The *arrows* indicate at the human Leu235 (1.2), the carbohydrate binding site Asn297 (84.4) and the Pro329 (114). The numbers in *parentheses* correspond to the IMGT unique numbering for C domains

*IGHG1*, *IGHG3*, *IGHG5* and *IGHG6* genes, the *IGHG2* gene had a deletion of 15 bp, and a short deletion of three bp was present in the *IGHG4* gene. The last 15 nucleotides of the IVS2 of the *IGHG2* gene corresponded to the beginning of the CH2 exon. The sequence was identical to the first 15 nucleotides of the *IGHG6* CH2 exon, except for the first nucleotide which was cytosine instead of guanine in the *IGHG2* sequence. As mentioned in the IMGT annotations (http://imgt.cines.fr) of AJ302056 (*IGHG2*) the splicing site is not conserved and a cryptic site located 15 bp downstream is used, which explains the "deletion" of five amino acids in the *IGHG2* cDNA. The length of the CH1 exon was conserved except for *IGHG5*, which had a deletion of six bp. The length of the CH3 exon was identical for all six *IGHG* genes.

The intervening sequences between the CH1 and hinge exons (IVS1), hinge and CH2 exons (IVS2) and CH2 and CH3 exons (IVS3) of horse *IGHG* genes decreased in length from the 5′ to the 3′ end of each *IGHG* gene. The transmembrane exons of the horse *IGHG* genes have not yet been characterized.

## Phylogeny of the horse immunoglobulin gamma heavy chains

The amino acid sequences of the six horse immunoglobulin gamma heavy chains were derived from the corre-

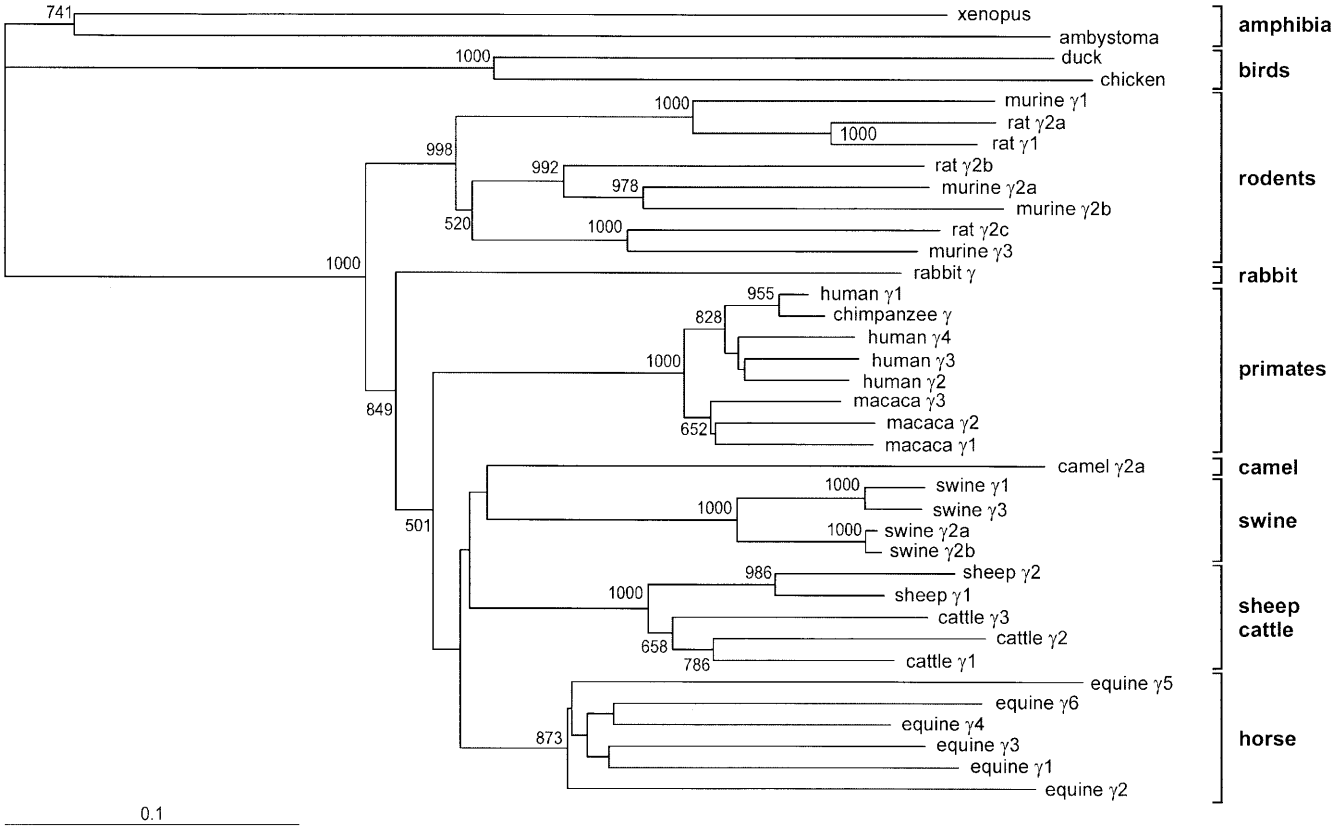


**Fig. 4** Neighbor joining (NJ) phylogenetic tree constructed from the amino acid sequences of the immunoglobulin gamma heavy chains of various mammalian species and the immunoglobulin upsilon chains of two amphibia and two birds. Alignments were calculated with CLUSTAL X (version 1.8) using the BLOSUM series protein matrices, the gap opening penalties for the pairwise and the multiple alignments were set to 10.00, and the gap extension penalties to 0.1 and 0.2 for the pairwise and multiple alignments, respectively. Bootstrapping of the NJ tree was performed with 1,000 replicates. All bootstrapping values ≥500 are shown in the figure. TREEVIEW version 1.6.5 was used to outgroup the tree to the *Ambystoma* and *Xenopus* sequences and for the graphic output. The *bar* indicates the number of substitutions per site

sponding *IGHG* nucleotide sequences (Fig. 3). The horse gamma chains, those of various other mammals and the upsilon heavy chains of some other vertebrates were clustered in relation to their species origin (Fig. 4). As expected, the upsilon heavy chains of amphibia and birds, representing the expected precursors of the mammalian gamma and epsilon heavy chains (for review: Warr et al. 1995), clustered separately from the mammalian immunoglobulin gamma chains. From a common origin, the mammalian gamma heavy chains generated seven distinct clusters. First was a rodent cluster, including the rat and mouse immunoglobulin gamma chains. All other gamma chains branch off into separate clusters, with the rabbit gamma chain (cluster 2), and five more related clusters: cluster 3, the primate cluster, comprising human, chimpanzee and macaca immunoglobulin gamma chains; cluster 4, containing the camel gamma 2a

**Table 3** Amino acid sequence homology between the six horse immunoglobulin gamma chains (γ1 to γ6) and their individual constant domains

| Amino acid sequence homology (%) | | | | | | |
|---|---|---|---|---|---|---|
| | | γ 2 | γ 3 | γ 4 | γ 5 | γ 6 |
| Constant region | γ 1 | 70 | 74 | 75 | 68 | 73 |
| | γ 2 | | 67 | 69 | 67 | 69 |
| | γ 3 | | | 73 | 68 | 70 |
| | γ 4 | | | | 73 | 78 |
| | γ 5 | | | | | 69 |
| CH1 domain | γ 1 | 82 | 85 | 80 | 58 | 81 |
| | γ 2 | | 85 | 81 | 61 | 83 |
| | γ 3 | | | 88 | 64 | 88 |
| | γ 4 | | | | 63 | 94 |
| | γ 5 | | | | | 63 |
| Hinge | γ 1 | 23 | 21 | 20 | 35 | 34 |
| | γ 2 | | 21 | 11 | 22 | 33 |
| | γ 3 | | | 8 | 18 | 21 |
| | γ 4 | | | | 30 | 53 |
| | γ 5 | | | | | 53 |
| CH2 domain | γ 1 | 67 | 79 | 75 | 75 | 81 |
| | γ 2 | | 61 | 65 | 67 | 65 |
| | γ 3 | | | 75 | 76 | 71 |
| | γ 4 | | | | 82 | 75 |
| | γ 5 | | | | | 72 |
| CH3 domain | γ 1 | 72 | 76 | 80 | 76 | 65 |
| | γ 2 | | 72 | 71 | 77 | 66 |
| | γ 3 | | | 79 | 80 | 68 |
| | γ 4 | | | | 77 | 68 |
| | γ 5 | | | | | 73 |

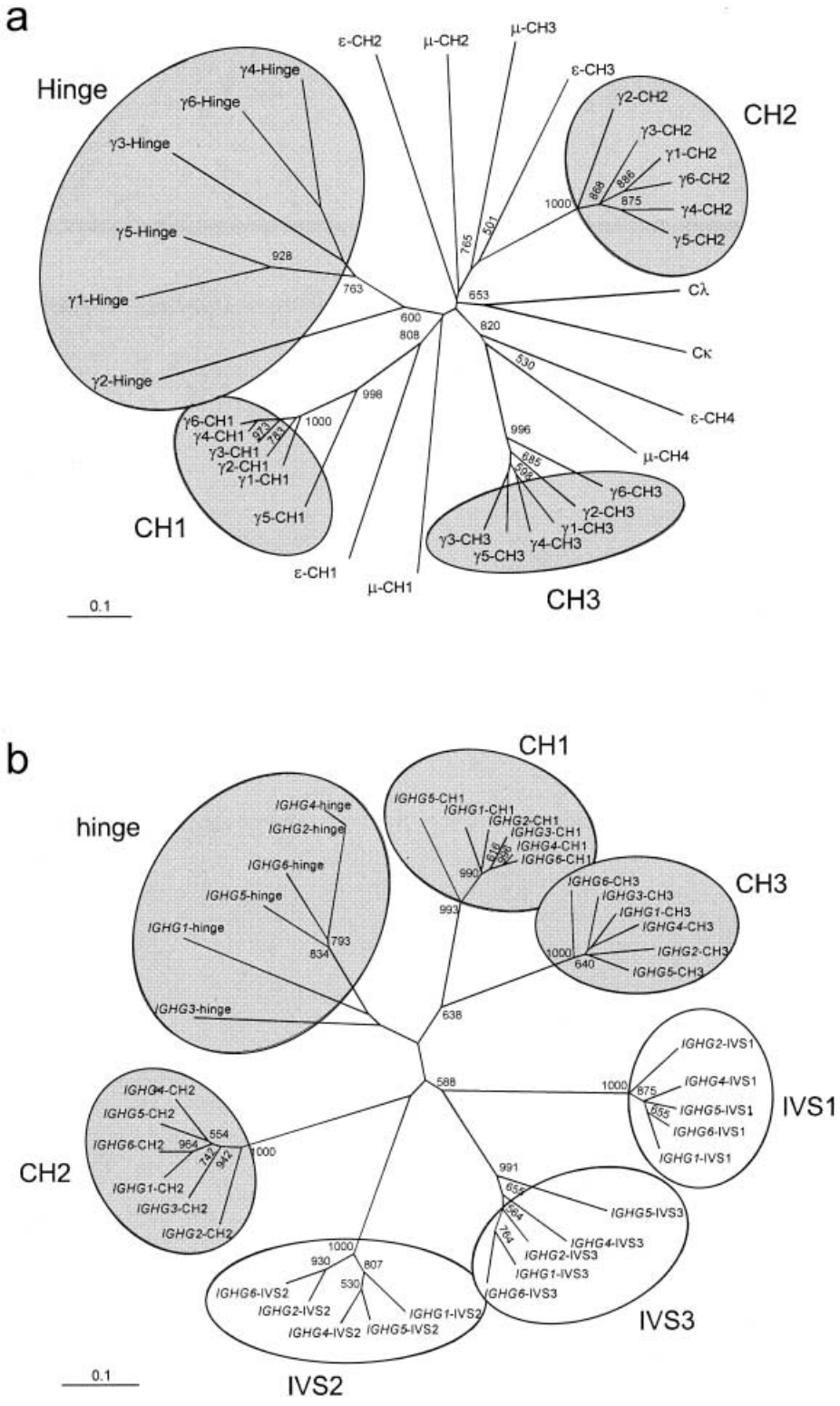


**Fig. 5** Unrooted neighbor joining (NJ) phylogenetic tree constructed with (**a**) the amino acid sequences of the horse immunoglobulin gamma (γ), mu (μ) and epsilon (ε) heavy chain domains and (**b**) the sequences of the *IGHG* exons and intervening sequences (IVS). Calculations of alignments and of the phylogenetic trees were performed using CLUSTAL X (version 1.8). The settings for **a** were as described in Fig. 4. For the alignment of the DNA sequences **b**, the CLUSTAL W (1.6) DNA matrix was used. Gap opening and gap extension penalties were set as described for Fig. 4. The DNA transition weight was set to 0.5. TREEVIEW version 1.6.5 was used for the graphic output. The *bar* indicates the number of substitutions per site

chain, which is part of a heavy chain antibody; cluster 5, containing four porcine gamma chains; cluster 6, a ruminant cluster with sheep and cattle gamma chains; and finally cluster 7, with the six horse immunoglobulin gamma chains.

## Segmental homologies of the horse immunoglobulin gamma chains

The amino acid sequence homologies between the horse immunoglobulin gamma chain constant regions ranged from 67% to 78% (Table 3). In some instances, the homologies between individual domains of two gamma

chains differed from the overall homology between constant regions. For example, the homology between the constant regions of horse gamma 6 and gamma 4 or gamma 5 heavy chains was 78% and 69%, respectively. When individual domains were compared, gamma 4 and gamma 6 chains shared 94% homology in the CH1 domain, 75% in the CH2 domain, but only 68% in the CH3 domain. In contrast, the homology between the respective domains of the gamma 5 and gamma 6 chains was only 63% in the CH1 domain, but 72% and 73% in the CH2 and CH3 domains. Likewise, amino acid sequence homologies between the CH1 domains from gamma 6 and gamma 4 chains were distinctly higher than those between gamma 6 and gamma 5, while for the CH3 domain of the gamma 6 chain, a higher homology was found to the corresponding gamma 5 domain. The lowest homologies were found between the hinge domains (8–53%).

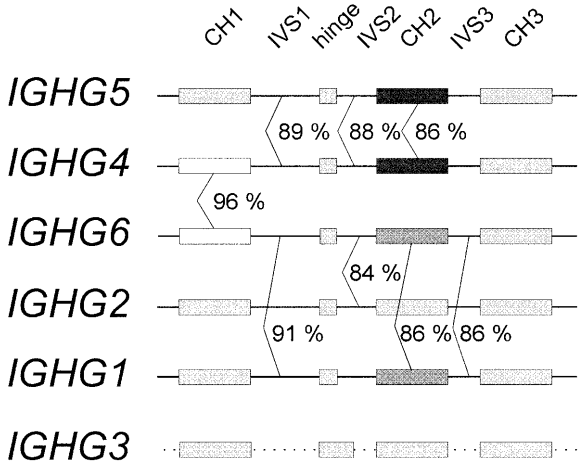


**Fig. 6** Model of segmental homology of horse *IGHG* genes. The related *IGHG* gene segments were determined according to the calculated sequence homologies of the exon and introns, which are shown in this figure for the closely related segments, as well as the CH domains (Table 3) and the phylogenetic comparison at the protein and nucleotide levels shown in Fig. 5. The horse *IGHG* gene segments sharing segmental homology are *joined*. The percentages refer to nucleotide sequence homology

## Phylogenetic relationship between the horse *IGHG* genes

To determine the evolutionary relationship between the individual horse gamma chain domains, an unrooted neighbor-joining phylogenetic tree was constructed, including the constant regions of horse mu and epsilon heavy chains and the lambda and kappa light chains (Fig. 5A). Four major clusters could be distinguished. The CH1 domains of the six horse immunoglobulin gamma chains were closely related to each other, and formed a cluster with the CH1 domains of the horse mu and epsilon chains, respectively. Likewise, the CH2 domain cluster was located on one branch with the second and third CH domains of the mu and epsilon chains. The third cluster combined the CH3 domains of the six gamma chains with the CH4 domains of the mu and epsilon chains. A fourth cluster contained the hinge regions of the gamma chains. The CH domain clusters of the horse gamma chains indicated a higher phylogenetic distance to mu than to epsilon chain for the CH1 and CH3 domains, but for the CH2 and CH4 domains, the mu chain was found to be more closely related than the epsilon chain. The close relationship between the gamma heavy chain domains suggests that the six horse *IGHG* genes developed by duplications from one ancestral *IGHG* gene, and that this ancestral *IGHG* gene most probably originated either from a precursor *IGHM* or *IGHE* gene.

The phylogenetic tree constructed with the nucleotide sequences of the individual horse *IGHG* exons and the corresponding intervening sequences (Fig. 5B) resulted in the same clustering as for the heavy chain domains. The intervening sequences IVS1 to IVS3 each formed three distinct clusters.

The segmental amino acid homologies of the gamma heavy chain domains reflected the phylogenetic relations of both, the CH domains and *IGHG* exons. These results are summarized in a model of horse *IGHG* segmental homology (Fig. 6). We suggest that besides multiple duplications leading to the formation of six horse *IGHG* genes, gene conversion events and/or unequal crossing over must have occurred during the development of constant heavy chain genes. We postulate that sequences were exchanged between the *IGHG4* and *IGHG6* CH1 exons, *IGHG4* and *IGHG5* (IVS1, IVS2 and CH2 exon), *IGHG6* and *IGHG1* (IVS1, CH2 exon and IVS3), as well as *IGHG6* and *IGHG2* (IVS2).

## Hypothetical FcγR-binding sites on horse IgG

The amino acids in the proximal CH2 domain involved in the binding of the human IgG1 to the FcγRIII were recently described (Sondermann et al. 2000). Here, we compared the human IgG1 binding sites to the corresponding regions of the six horse immunoglobulin gamma chains (Fig. 3). The amino acid positions of the human sequence were used as described by Sondermann and coworkers (2000). In addition, the IMGT unique numbering for C domains, which allows a standardized comparison between different species was given after each amino acid position in parentheses: the human Leu234 (1.3) – Ser239 (3) residues are crucial for IgG1-FcγRIII binding. Identical amino acids are found in the horse CH2 domains of the gamma 1 (Leu122 (1.3) – Ser127 (3)) and gamma 3 (Leu139 (1.3) – Ser144 (3)) heavy chains. Other regions of the human IgG1 molecule are also crucial for FcγRIII binding, namely Asp265 (27), Ser267 (29), Glu269 (31), Ala327 (112) – Ala330 (115), Ile332 (117), as well as the N-glycosylation site Asn297 (84.4) – Thr299 (85.3), including the carbohydrate residue NAG1 at this position (Sondermann et al. 2000). Most of these amino acid residues are also conserved in horse gamma 1 and gamma 3 heavy chains, except for the human Ala330 (115) which is Gln in horse

gamma 1, and the human Glu269 (31) and Ile332 (117), which are Asp and Val, respectively, in horse gamma 3 (Fig. 3). The other four horse immunoglobulin gamma chains show at least three (gamma 6), four (gamma 5) or five (gamma 4) amino acid substitutions or, in addition to five substitutions, a deletion of the five proximal amino acids of the CH2 domain (gamma 2), including the amino acids corresponding to human Leu234 (1.3) and Leu235 (1.2).

## Discussion

The evolution of *IGHG* genes is believed to have begun approximately 600 million years ago (Ellison et al. 1982). In mammals, it is assumed that the *IGHM* gene, as the phylogenetically oldest *IGHC* gene, multiplied to generate the ancestral *IGHD*, *IGHG* and *IGHE* genes. These ancestral *IGHC* genes, particularly the *IGHG* genes, then underwent further duplications, resulting in the *IGHG* genes we observe today in the various species (Honjo and Matsuda 1995). Additional duplication events have occurred in individual species, like *Mus musculus* and *Rattus norvegicus*, which are believed to share three ancestral *IGHG* genes, which then evolved into four *IGHG* genes by species-specific duplications of different *IGHG* genes (Brüggemann 1988) or underwent further *IGHG2* gene duplications in some mouse strains, leading to a fifth *IGHG2c* gene (Morgado et al. 1989; Martin et al. 1997). The human species shares only one ancestor *IGHG* gene with the rodent species and the human *IGHG* genes developed in a unique way: after duplication of the ancestral *IGHG* gene, the entire *IGHG-IGHG-IGHE-IGHA* fragment duplicated (Flanagan and Rabbitts 1982; Matsuda and Honjo 1996). Due to the existence of one pseudogene (*IGHGP*) the usual number of *IGHG* genes in human is five (Bensmana et al. 1988).

According to the nucleotide sequences of the six horse *IGHG* genes provided here, the deduced amino acid sequences of the horse immunoglobulin gamma chains were obtained. The discrepancy between the five horse IgG subclasses (IgGa, IgGb, IgGc, IgG(T) and IgG(B)) described earlier and the six horse *IGHG* genes leaves us with the question of how many of the *IGHG* genes are expressed. The *IGHG1*, *IGHG3* and *IGHG4* genes could be linked to their corresponding proteins IgG1 (IgGa), IgG3 [IgG(T)] and IgG4 (IgGb), respectively (Wagner et al. 1998). However, all six *IGHG* genes have switch regions (Overesch et al. 1998; Wagner et al. 1998) and, as described in this study, all *IGHG* genes encode for complete constant regions, making it unlikely that any are pseudogenes. Additional evidence for the expression of all six horse *IGHG* genes is provided by comparison of the C-terminal amino acid sequence of IgG(T), described earlier (Weir et al. 1966) with the corresponding C termini of the predicted gamma 1 to gamma 6 heavy chains, indicating that the gamma 5 sequence His308 – Gly325 has 100% homology to the 18 C-terminal amino acids described for IgG(T). Horse IgG(T) could be previously linked to the *IGHG3* gene by heterohybridoma expression (Wagner et al. 1998) and could be separated by affinity chromatography into two different fractions (Sheoran and Holmes 1996). Thus, we suggest that the originally described IgG(T) includes two different horse isotypes, namely IgG3 and IgG5, and that the isotypes designated as IgGc and IgG(B) are expressed by the remaining *IGHG2* and *IGHG6* genes.

The six horse immunoglobulin gamma heavy chains were compared with those of other species. The phylogenetic analysis clearly indicates that horses share one common ancestor with the other mammalian species investigated here. After the horse had separated as an individual species, the six horse *IGHG* genes apparently have developed from this common ancestor *IGHG* gene by multiple duplications.

*IGHG* gene duplication is also apparent for many of the other species included in the phylogenetic tree, except for the mouse and rat *IGHG* genes, as described above. Further evidence for different species sharing common ancestor *IGHG* genes is found in the example of humans and chimpanzees. Only one *IGHG* is described, showing a close phylogenetic relationship to the human *IGHG1* gene. In accordance, all rhesus monkey and human *IGHG* genes share one common ancestor. *IGHG* gene duplications in ruminants, represented here by sheep and cattle, also have happened after species separation. To clarify the phylogenetic relationship of other perissodactyla to the horse, it remains to be shown whether the *IGHG* genes of those closely related species such as donkey, zebra, rhinoceros and tapir share one or several *IGHG* precursor genes with horses.

For the mammalian species compared here, the evolutionary clustering of the individual immunoglobulin gamma chains is obvious and clear, as reflected by the high bootstrap values of >800 (Fig. 4). Between the individual horse gamma heavy chains, however, bootstrap values are mostly low (<500), suggesting that, apart from the duplication events leading to the formation of six horse *IGHG* genes, segmental exchanges have occurred between them. We suggest further that horse *IGHG* gene segments have been subject to gene conversion events and/or unequal crossing over, as has been postulated earlier for the human and mouse *IGHG* and *IGHA* genes, which also share segmental gene homology (Takahaski et al. 1982; Flanagan et al. 1984; Hayashida et al. 1984; Lefranc et al. 1986; Huck et al. 1989). In particular, the segmental homologies between the individual horse *IGHG* genes (Fig. 6) suggest that gene conversion events and/or unequal crossing over occurred between horse *IGHG4* and *IGHG6*, *IGHG4* and *IGHG5*, *IGHG6* and *IGHG1* as well as *IGHG6* and *IGHG2* genes. Additional point mutations have added diversification later on and the exact sizes of the exchanged fragments can no longer be determined.

The horse *IGHG* nucleotide sequences show a high degree of divergence in the hinge region and the 5′ end of the CH2 exon. The hinge exon of horse *IGHG3* is dis-

tinctively longer than those of the other five *IGHG* genes. In that aspect, horse *IGHG3* resembles human *IGHG3*. The number of hinge exons in the human *IGHG3* may vary from two to five (Huck et al. 1989; Dard et al. 1997, 2001), but the hinge region is usually encoded by four exons (Takahaski et al. 1982). Due to the 11 interchain Cys disulfide bridges of its four exon hinge region, human IgG3 is very rigid, while the horse IgG3 hinge region, containing only four Cys residues, allows a high degree of flexibility, like the shorter horse IgG hinge regions.

By analogy to mouse and human IgGs, the amino acid divergences and deletions in the proximal CH2 domains of the horse gamma chains may be most relevant for transmission of effector functions. In human and mouse IgG, this part of the CH2 domain has recently been shown to be crucial for IgG1-FcγRIII binding (Sondermann et al. 2000). Out of 17 amino acids of the human IgG1 CH2 domain, which are crucial for interaction with the receptor, 16 are also present in the horse gamma 1 and 15 in the gamma 3 heavy chain, while the other four horse gamma chains differ by at least three amino acid residues (Fig. 3). On the basis of the conserved amino acid sequences of the human FcRs and different IgGs, Sondermann and coworkers (2000) proposed a model describing the principal interactions within the various IgG-FcR complexes, in which the "proline sandwich" appears to be the principal binding motif. This proline sandwich characterizes the interaction of Pro329 (corresponding to position 114 in the IMGT unique numbering for C domains) of one CH2 domain of the IgG1 with Trp87 and Trp110 of the FcγRIII, both interacting with the carbohydrate residues NAG1 of the Fc fragment and the interchain disulfide bridge of the Cys229 (14) residues of the IgG1 hinge region. The residues corresponding to Pro329 (114) as well as the Asn297 (84.4) carbohydrate binding site of human IgG1 are conserved in all human and mouse IgG and IgE isotypes, as well as in all horse gamma (Fig. 3) and epsilon heavy chains (Pro301 and Asn269) as described by Wagner and coworkers (2001). For the mouse IgG-FcγRII interaction, the same contact areas as in the human IgG1-FcγRIII complex have been identified in the proximal sequence of the CH2 domain of mouse IgG2b (Kato et al. 2000). The significantly higher affinity of FcγRI is thought to be mediated by its third domain, because the two N-terminal domains show an affinity to IgG comparable to that of FcγRII/III (Hulett and Hogarth 1998). The N-terminal domains of the FcγRI will probably bind the IgG Fc fragment as in the FcγRII/III complex, because the mutation of the Glu235 to Leu increases the affinity of mouse IgG2b by more than 100-fold (Duncan et al. 1988). This underlines the central importance of the Leu235 (1.2) residue for the FcγRI binding. The comparison of the amino acid residues of horse immunoglobulin gamma chains corresponding to position Leu235 shows that only the gamma 1, gamma 3 and gamma 6 chains contain a Leu at that position. Provided that the crucial binding sites of the respective FcγRs are conserved between man and horse, horse IgG1, IgG3 and probably IgG6 but not the other IgG isotypes should have the ability to bind to the horse FcγRs. This will have to be verified experimentally.

In summary, our analysis of the evolutionary relationship of the six horse *IGHG* genes and the *IGHG* genes of other species shows that the horse *IGHG* genes are not individually linked to their human or mouse counterparts by evolutionary descent. The diversification and duplication of the horse *IGHG* genes is unique for the horse and maybe its close relatives, but different from primates and rodents. It remains to be shown to what extent co-evolution of genes involved in distinct types of immune response has generated *IGHC* genes with analogous effector functions in different mammalian species. In this context, it would be of interest to discover to what extent the apparently conserved mechanisms of Th1/Th2 dichotomy in the control of immune effector functions are connected to the diverse *IGHG* genes. In man and mouse, class switch recombination is targeted to distinct isotypes by cytokines characteristic for Th1 or Th2 cells, which then recruit distinct IgG classes for immune reactions controlled by those cytokines (Siebenkotten and Radbruch 1995; Lorenz and Radbruch 1996). Targeting of switch recombination by T cell cytokines operates via cytokine-induced switch transcripts of the recombinogenic DNA switch regions in front of the respective *IGHG* genes. Analysis of the phylogeny of horse promotors of switch transcripts will clarify which horse *IGHG* genes are recruited into Th1 or Th2 immune responses.

**Acknowledgements** This work was supported by the Stiftung Gestüt Fährhof. The experiments described in this study were performed following the recommendations of the German Gentechnik Gesetz.

## References

Bensmana M, Huck S, Lefranc G, Lefranc MP (1988) The human immunoglobulin pseudo-gamma *IGHGP* gene shows no major structural defect. Nucleic Acids Res 16:3108

Brüggemann M (1988) Evolution of the rat immunoglobulin gamma heavy-chain gene family. Gene 74:473–482

Brüggemann M, Free J, Diamond A, Howard J, Cobbold S, Waldmann H (1986) Immunoglobulin heavy chain locus of the rat: striking homology to mouse antibody genes. Proc Natl Acad Sci U S A 83:6075–6079

Butler JE, Brown WR (1994) The immunoglobulins and immunoglobulin genes of swine. Vet Immunol Immunopathol 43:5–12

Dard P, Huck S, Frippiat JP, Lefranc G, Langaney A, Lefranc MP, Sanchez-Mazas A (1997) The *IGHG3* gene shows a structural polymorphism characterized by different hinge lengths: sequence of a new 2-exon hinge gene. Hum Genet 99:138–141

Dard P, Lefranc MP, Osipova L, Sanchez-Mazas A (2001) DNA sequence variability of *IGHG3* alleles associated to the main *G3m* haplotypes in human populations. Eur J Hum Genet 9:765–772

Duncan AR, Winter G (1988) The binding site for the C1q on IgG. Nature 332:738–740

Duncan AR, Woof JM, Partrigde LJ, Burton DR, Winter G (1988) Localization of the binding site for the human high affinity Fc receptor on IgG. Nature 332:563–564

Ellison J, Berson BJ, Hood L (1982) The nucleotide sequence of a human immunoglobulin C gamma 1 gene. Nucleic Acid Res 10:4071–4079

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

Flanagan JG, Rabbitts TH (1982) Arrangement of human immunoglobulin heavy chain constant region genes implies evolutionary duplication of a segment containing γ, ε and α genes. Nature 300:709–713

Flanagan JG, Lefranc MP, Rabbitts TH (1984) Mechanisms of divergence and convergence of the human immunoglobulin alpha 1 and alpha 2 constant region gene sequences. Cell 36:681–688

Fukui K, Hamaguchi Y, Shimizu A, Nakai S, Moriwaki K, Wang CH, Honjo T (1984) Duplicated immunoglobulin gamma 2a genes in wild mice. J Mol Cell Immunol 1:321–330

Hayashida H, Miyata T, Yamawaki-Kataoka Y, Honjo T, Wels J, Blattner F (1984) Concerted evolution of the mouse immunoglobulin gamma chain genes. EMBO J 3:2047–2053

Honjo T, Matsuda F (1995) Immunoglobulin heavy chain loci of mouse and human. In: Honjo T, Alt FW (eds) Immunglobulin genes. Academic Press 2nd ed. pp 145–172

Huck S, Lefranc G, Lefranc MP (1989) A human immunoglobulin *IGHG3* allele (Gmb0,b1,c3,c5,u) with an IGHG4 converted region and three hinge exons. Immunogenetics 30:250–257

Hulett MD, Hogarth PM (1998) The second and third extracellular domains of the FcγRI (CD64) confer the unique high affinity binding of IgG2a. Mol Immunol 35:989–996

Kato K, Sautès-Fridman C, Yamada W, Kobayashi K, Kim H, Enokizono J, Galinha A, Kobayashi Y, Fridman WH, Arata Y, Shimada I (2000) Structural basis of the interaction between IgG and Fc gamma receptors. J Mol Biol 295:213–224

Knight KL, Burnett RC, McNicholas JM (1985) Organization and polymorphism of rabbit immunoglobulin heavy chain genes. J Immunol 134:1245–1250

Knight KL, Suter M, Becker RS (1988) Genetic engineering of bovine Ig. Construction and characterization of hapten-binding bovine/mouse chimeric IgE, IgA, IgG1, IgG2 and IgG3 molecules. J Immunol 140:3654–3659

Lefranc MP, Lefranc G (eds) (2001) The immunoglobulin facts book. Academic, New York, pp 46–47, 65–68

Lefranc MP, Lefranc G, Rabbitts TH (1982) Inherited deletion of immunoglobulin heavy chain constant region genes in normal human individuals. Nature 300:760–762

Lefranc MP, Helal AN, de Lange G, Chaabani H, van Loghem E, Lefranc G (1986) Gene conversion in human immunoglobulin gamma locus shown by unusual location of IgG allotypes. FEBS Lett 196:96–102

Lorenz M, Radbruch A (1996) Immunoglobulin class switching. In: Snapper CM (ed) Cytokine regulation of humoral immunity: basic and clinical aspects. Wiley, New York, pp 109–135

Martin RM, Silva A, Lew AM (1997) The Igh-1 sequence of the non-obese diabetic (NOD) mouse assigns it to the IgG2c isotype. Immunogenetics 46:167–168

Matsuda F, Honjo T (1996) Organization of the human immunoglobulin heavy-chain locus. Adv Immunol 62:1–29

Montgomery PC (1973) Molecular aspects of horse antibodies. In: Bryans JT, Gerber H (eds) Proceedings of the 3rd International Conference on Horse Infectious Diseases. Karger, Basel, pp 343–363

Morgado MG, Cam P, Gris-Liebe C, Cazenave PA, Jouvin-Marche E (1989) Further evidence that BALB/c and C57BL/6 gamma 2a genes originate from two distinct isotypes. EMBO J 8:3245–3251

Nicholas KB, Nicholas HB (1997) GeneDoc: a tool for editing and annotating multiple sequence alignments. Distributed by the authors: http://www.psc.edu/biomed/genedoc

Overesch G, Wagner B, Radbruch A, Leibold W (1998) Organisation of the horse immunoglobulin constant heavy chain genes. II. Horse *Cγ* genes. Vet Immunol Immunopathol 66:273–287

Page RDM (1996) Treeview: an application to display phylogenetic trees on personal computers. Comput Appl Biosci 12:357–358

Ravetch JV, Kinet JP (1991) Fc receptors. Annu Rev Immunol 9:457–492

Richards CM, Aucken HA, Tucker EM, Hannant D, Mumford JA, Powell JR (1992) The production of horse monoclonal immunoglobulins by horse-mouse heterohybridomas. Vet Immunol Immunopathol 33:129–143

Saitou N, Nei M (1987) The neighbour-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Sensel MG, Kane LM, Morrison SL (1997) Amino acid differences in the N-terminus of the CH2 influence the relative abilities of IgG2 and IgG3 to activate complement. Mol Immunol 34:1019–1029

Sheoran AS, Holmes MA (1996) Separation of horse IgG subclasses (IgGa, IgGb and IgG(T)) using their different binding characteristics for staphylococcal protein A and streptococcal protein G. Vet Immunol Immunopathol 55:33–43

Shimizu A, Takahashi N, Yaoita Y, Honjo T (1982) Organization of the constant region gene family of the mouse immunoglobulin heavy chain. Cell 28:499–506

Siebenkotten G, Radbruch A (1995) Towards a molecular understanding of immunoglobulin class switching. Immunologist 4:141–145

Sondermann P, Huber R, Oosthuizen V, Jacob U (2000) The 3.2-Å crystal structure of the human IgG1 Fc fragment-FcγRIII complex. Nature 406:267–273

Takahaski N, Ueda S, Obata M, Nikaido T, Nakai S, Honjo T (1982) Structure of human immunoglobulin gamma genes: implications for evolution of a gene family. Cell 29:671–679

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 24:4876–4882

Volgina VV, Kingzette M, Zhai SK, Knight KL (2000) A single 3′ alpha hs1,2 enhancer in the rabbit *IgH* locus. J Immunol 165:6400–6405

Wagner B, Radbruch A, Richards C, Leibold W (1995) Monoclonal horse IgM and IgG immunoglobulins. Vet Immunol Immunopathol 47:1–12

Wagner B, Overesch G, Sheoran AS, Holmes MA, Richards C, Leibold W, Radbruch A (1998) Organization of the horse immunoglobulin heavy chain constant region genes. III. Alignment of *Cμ*, *Cγ*, *Cε* and *Cα* genes. Immunobiol 199:105–119

Wagner B, Siebenkotten G, Radbruch A, Leibold W (2001) Nucleotide sequence and restriction fragment length polymorphisms of the horse *Cε* gene. Vet Immunol Immunopathol 82:193–202

Warr GW, Magor KE, Higgins DA (1995) IgY: clues to the origins of modern antibodies. Immunol Today 16:392–398

Weir RC, Porter RR, Givol D (1966) Comparison of the C-terminal amino-acid sequence of two horse immunoglobulins, IgG and IgG(T). Nature 212:205–206