

## Appendix A

**Table S1** The primer sequences used in this study

Name	Sequence(5'-3')	Usage
Not I-d (T) <sub>18</sub>	AAC TGG AAG AAT TCG CGG CCG CAG GAA TTT TTT TTT TTT TTT TTT	First-strand cDNA synthesis
IgHM C1SP	GAC TTC CTG CCT GGC TCC AT	cDNA sequences amplification
IgHM C1AP	TTG GAG TGC TGG ACT TTG CA	cDNA sequences amplification
IgHM VGSP1	TTT GCA AAT CAG CTC TGA	cDNA library
IgHM VGSP2	GGA AGG TCT TGA TGC TCT GG	cDNA library
J <sub>H</sub> SP	GAC CCA GGT CAC CGT CTC CTC AG	Specific primers for 3'RACE
RT-P1	AAC TGG AAG AAT TCG CGG CC	Specific primers for 3'RACE
IgHMCSP	GCA AAG TCC AGC ACT CCA AG	Clone screening
IgHMCAP	GAA ATG CTC AGG CTG TCG TA	Clone screening
IgHGCAP	AAG TAG GTC CCG TCG TTG TCC AG	Clone screening
IgHG1aHSP	GAA CTC GAG ATA TCT GAA CC	Clone screening
IgHG1bHSP	GAA CCG CAT GGA GGA TGC CC	Clone screening
IgHG2aHSP	GAA CCC AAG ATA CCC CAA CC	Clone screening
IgHG2cHSP	GCA CCA CCA CCC CGA AGA CCC	Clone screening
IgHG3HSP	GGA ACG AAT GGA GGA TGC AA	Clone screening
IgHECAP	CAA CAG GGT GGA TGT GGC AGT	C $\epsilon$ amplification
IgHECSP	CAG GCG GTC ACA GAC AGG TTC	C $\epsilon$ amplification
IgHGVGSP1	GTT ACG CTC ATG GTG TCC T	cDNA library
IgHG3VGSP2	TTG CAT CCT CCA TTC GTT CC	cDNA library
IgHG1a/bVGSP2	GGT GGG CTA CGT TGC AGA AG	cDNA library
IgHGC1SP	CCC ATC GGT CTA TCC TCT GA	DNA sequence amplification
IgHG1aHAP	GCA TCC TGG TTG TGG TTG AG	DNA sequence amplification
IgHG1bHAP	GCA TCC TCC ATG CGG TTC	DNA sequence amplification
IgHG2aHAP	ACG TGC ATT CTG GTT CAG GT	DNA sequence amplification
IgHG2cHAP	CAC TGG GAG CTG GGG TCT T	DNA sequence amplification

**Table S2** The *VH* genes and *CH* genes extracted from the NCBI and IMGT database for phylogenetic tree construction

Species	VH family	Accession number	<i>CH</i> gene	Accession number
Human	VH1	L22582	<i>Cμ</i>	X1494040
	VH2	X62111	<i>Cα</i>	J00220
	VH3	M99649	<i>Cε</i>	J00222
	VH4	M29811	<i>Cγ1</i>	Z17370
	VH5	M99686	<i>Cγ2</i>	J00230
	VH6	X92224	<i>Cγ3</i>	X03604
	VH7	L10057	<i>Cγ4</i>	P01861
Mouse	VH1	AC090843	<i>Cμ</i>	V00818
	VH2	AC090887	<i>Cα</i>	D11468
	VH3	AC073563	<i>Cε</i>	X01857
	VH4	AC079273	<i>Cγ1</i>	J00453
	VH5	AC090887	<i>Cγ2a</i>	V00825
	VH6	AC073590	<i>Cγ2b</i>	V00763
	VH7	AC079273	<i>Cγ2c</i>	J00479
	VH8	AC079181	<i>Cγ3</i>	X00915
	VH9	AC073563		
	VH10	AC073561		
	VH11	AC073563		
	VH12	AC073590		
	VH13	AC073589		
	VH14	AC079273		
	VH15	AC090843		
	VH16	AC073563		
Pig	VH1	AF064687	<i>Cμ</i>	AAC48775
			<i>Cα</i>	I47175
			<i>Cε</i>	AAC48776
			<i>Cγ1</i>	I47158
			<i>Cγ2a</i>	I47159
			<i>Cγ2b</i>	I47160
			<i>Cγ3</i>	I47161
		<i>Cγ4</i>	I47162	
Rabbit	VH1	M93171	<i>Cμ</i>	AAA64251
			<i>Cα</i>	S09264
			<i>Cγ</i>	AAB59265
Sheep	VH1	Z49180	<i>Cμ</i>	S25705
			<i>Cα</i>	AA64980
			<i>Cε</i>	AAA51378
			<i>Cγ1</i>	S31459
			<i>Cγ2</i>	X70983

Alpaca	VH1	AM939699	$C\mu$	CAO79568
	VH2	AM939705	$C\alpha$	CAO79580
	VH3	AM939730	$C\varepsilon$	CAO79578
	VHH3	AM939754	$C\gamma1a$	CAO79572
			$C\gamma1b$	CAO79575
			$C\gamma2b$	CAO79570
			$C\gamma2c$	CAO79576
	Bovine	VH	U55164	$C\mu$
$C\alpha$				AF109167
$C\varepsilon$				U63640
$C\gamma1$				S82409
$C\gamma2$				X16702
$C\gamma3$				U63639
Dromedary	VH3	HE653078	$C\mu$	BAD00196
	VH4	HE653061	$C\gamma1a$	CAD13185
	VHH3	AJ245126	$C\gamma2a$	CAB64864
			$C\gamma3$	Z48947
Llama	VH1	AF305949	$C\gamma1b$	AAG42243
	VHH3	AF305946	$C\gamma2b$	AAX73259
			$C\gamma2c$	AAX73260
			$C\gamma3$	ADI87159
Nurse shark	VH1	M92851	$C\mu$	I50731
Horse			$C\mu$	AAU09792
			$C\alpha$	AAP80145
			$C\varepsilon$	AAA85662
			$C\gamma1$	CAC44760
			$C\gamma2$	CAC44761
			$C\gamma3$	CAC86339
			$C\gamma4$	AAS18415
			$C\gamma5$	CAC86340
			$C\gamma6$	CAC86341
	Panda			$C\mu$
			$C\alpha$	AAX73304
			$C\varepsilon$	AAX73306
Platypus			$C\mu$	AAO37747
			$C\alpha$	AAL17700
			$C\varepsilon$	AAL17702
Dog			$C\alpha$	AAA56796
			$C\varepsilon$	AAA56797
			$C\gamma1$	AAL35301
			$C\gamma2$	AAL35302
		$C\gamma3$	AAL35303	
		$C\gamma4$	AAL35304	

---

**Table S3** Statistics of the MELG and MRLG leader sequences from three variable region cDNA libraries

	IgM			IgG1a/b			IgG3		
	MELG	MRLG	others	MELG	MRLG	others	MELG	MRLG	others
VH3 family	159	0	10	125	0	14	167	0	25
VH4 family	0	43	0	0	44	5	0	0	0
Total No. of sequences		212			188			192	

**Table S4** Analysis of Kozak sequences and the first four downstream amino acids of the cDNA leader sequences

Kozak sequences	First four residues of the leader sequences			
	MELG	MRLG	Others	Total
T/ACACC	450	0	22	472
GGAAG	1	86	0	87
Total	451	86	22	559

**Table S5** Two-way ANOVA analysis of the variability index in VH3 sequences derived from IgM and IgG1a/b and VHH3 sequences derived from IgG3

pairwise comparison	<i>P</i> -value
(VH3 IgM)-(VH3 IgG1a/b)	0.7882127
(VHH3 IgG3)-(VH3 IgG1a/b)**	0.0000510
(VHH3 IgG3)-(VH3 IgM)**	0.0000028

Note: \*,  $P < 0.05$ , \*\*,  $P < 0.01$ .

**Table S6** Chi-square tests for the relationship between CDR3 length and the presence of an inter-loop disulfide bond

	Value	df	Sig. (2-sided)
Person Chi-Square	20.544	22	0.549
Likelihood Ratio	26.780	22	0.220
Linear-by-Linear Association	6.767	1	0.009
N of Valid Cases	192		