

Supporting Information

Table of Content

Supporting Tables

Table S1. Distribution of CDR3 amino acid length of top 11 V germline genes.

Table S2. Substitution rates of VHHs from top 11 V germline genes.

Table S3. Spearman's rank correlation of germline usage, pairing, and substitution preference for the five samples.

Supporting Figures

Fig. S1. Distribution of CDS length of VHHs.

Fig. S2. Phylogenetic tree of V germline genes in the IgH locus of *Vicugna pacos*. The phylogenetic tree is built using MEGA version X. The germline genes, which contain at least two hallmark residues, are shown in the rectangular box.

Fig. S3. Position specific score matrix (PSSM) of sample Naïve-1. The logo plot was built using WebLogo version 3.6. The hallmark residues are labeled with asterisk, corresponding to F37, E44, R45, and G47 in the Kabat numbering system.

Table S1. Distribution of CDR3 amino acid length of top 11 V germline genes.

V-germline a	J-germline	Sequence Data	CDR3 AA length			
			Minimal Length	Maximal Length	Average Length	Medium Length
IGHV3-1*01	IGHJ2*01	1026	0	21	10.88	10
IGHV3-1*01	IGHJ3*01	776	0	30	14.43	15
IGHV3-1*01	IGHJ4*01	6553	0	30	13.11	13
IGHV3-1*01	IGHJ5*01	10	5	22	13.30	11.5
IGHV3-1*01	IGHJ6*01	1216	0	24	12.86	12
IGHV3-1*01	IGHJ7*01	1716	0	26	13.29	13
IGHV3-3*01	IGHJ2*01	7609	0	30	14.75	14
IGHV3-3*01	IGHJ3*01	3284	0	35	14.52	15
IGHV3-3*01	IGHJ4*01	280120	0	30	17.86	22
IGHV3-3*01	IGHJ5*01	378	0	32	14.44	15
IGHV3-3*01	IGHJ6*01	38420	0	33	15.92	16
IGHV3-3*01	IGHJ7*01	31872	0	26	14.43	15
IGHV3S25*01	IGHJ2*01	520	2	25	10.40	11
IGHV3S25*01	IGHJ3*01	628	0	29	13.25	14
IGHV3S25*01	IGHJ4*01	8873	7	22	14.73	17
IGHV3S25*01	IGHJ5*01	11	0	26	12.24	11
IGHV3S25*01	IGHJ6*01	3123	1	26	15.18	15
IGHV3S25*01	IGHJ7*01	1604	0	26	13.61	15
IGHV3S39*01	IGHJ2*01	10253	0	28	12.40	12
IGHV3S39*01	IGHJ3*01	2077	0	30	12.56	12
IGHV3S39*01	IGHJ4*01	68826	0	25	11.49	8
IGHV3S39*01	IGHJ5*01	135	0	27	12.23	11
IGHV3S39*01	IGHJ6*01	24082	0	29	14.72	14
IGHV3S39*01	IGHJ7*01	12983	0	26	13.51	15
IGHV3S41*01	IGHJ2*01	1027	0	26	12.52	13
IGHV3S41*01	IGHJ3*01	392	0	30	11.83	12
IGHV3S41*01	IGHJ4*01	20302	0	22	8.09	7

IGHV3S41*01	IGHJ5*01	159	0	27	12.19	12
IGHV3S41*01	IGHJ6*01	5833	0	26	14.22	14
IGHV3S41*01	IGHJ7*01	3201	0	32	13.78	14
IGHV3S53*01	IGHJ2*01	14604	0	28	13.55	14
IGHV3S53*01	IGHJ3*01	3925	0	37	12.34	12
IGHV3S53*01	IGHJ4*01	200241	1	30	16.14	17
IGHV3S53*01	IGHJ5*01	291	0	29	12.00	12
IGHV3S53*01	IGHJ6*01	45693	0	34	14.29	14
IGHV3S53*01	IGHJ7*01	27188	0	28	15.51	16
IGHV3S61*01	IGHJ2*01	3183	0	30	14.70	16
IGHV3S61*01	IGHJ3*01	1535	0	30	14.44	14
IGHV3S61*01	IGHJ4*01	48385	0	22	15.84	21
IGHV3S61*01	IGHJ5*01	104	0	27	14.38	14
IGHV3S61*01	IGHJ6*01	14018	0	28	16.31	16
IGHV3S61*01	IGHJ7*01	10913	0	30	16.67	17
IGHV3S65*01	IGHJ2*01	25355	0	29	15.67	16
IGHV3S65*01	IGHJ3*01	12143	0	30	15.64	16
IGHV3S65*01	IGHJ4*01	515979	0	26	19.48	22
IGHV3S65*01	IGHJ5*01	1241	0	29	15.69	16
IGHV3S65*01	IGHJ6*01	152998	0	30	16.72	17
IGHV3S65*01	IGHJ7*01	126146	0	26	15.31	16
IGHV3S66*01	IGHJ2*01	1706	0	26	14.45	16
IGHV3S66*01	IGHJ3*01	1405	0	30	15.07	15
IGHV3S66*01	IGHJ4*01	41604	0	27	14.80	16
IGHV3S66*01	IGHJ5*01	61	0	27	15.73	16
IGHV3S66*01	IGHJ6*01	9252	0	28	16.03	16
IGHV3S66*01	IGHJ7*01	13043	0	23	13.74	14
IGHV3S9*01	IGHJ2*01	272	0	23	12.90	14
IGHV3S9*01	IGHJ3*01	152	0	30	12.93	13
IGHV3S9*01	IGHJ4*01	5405	7	22	15.89	18
IGHV3S9*01	IGHJ5*01	9	0	28	11.75	12
IGHV3S9*01	IGHJ6*01	3281	0	27	14.79	13

IGHV3S9*01	IGHJ7*01	990	0	27	14.79	13
IGHV3S1*01	IGHJ2*01	515	0	23	11.25	11
IGHV3S1*01	IGHJ3*01	648	0	22	12.90	13
IGHV3S1*01	IGHJ4*01	5222	0	33	12.11	11
IGHV3S1*01	IGHJ5*01	12	0	22	10.33	8.5
IGHV3S1*01	IGHJ6*01	1617	0	27	12.03	11
IGHV3S1*01	IGHJ7*01	1391	0	24	12.61	12

- a. The V germline genes containing hallmark amino acid residues are shown in bold.

Table S2. Substitution rates of VHHs from top 11 V germline genes.

V-germline^a	J-germline	Substitutions	Substitution Rate (%)
IGHV3-1*01	IGHJ2*01	940	13.86
IGHV3-1*01	IGHJ3*01	728	12.75
IGHV3-1*01	IGHJ4*01	5986	13.13
IGHV3-1*01	IGHJ6*01	935	13.13
IGHV3-1*01	IGHJ7*01	1579	12.51
IGHV3-3*01	IGHJ2*01	7088	15.49
IGHV3-3*01	IGHJ3*01	2949	16.40
IGHV3-3*01	IGHJ4*01	271675	15.62
IGHV3-3*01	IGHJ5*01	321	15.84
IGHV3-3*01	IGHJ6*01	36444	16.34
IGHV3-3*01	IGHJ7*01	30220	15.56
IGHV3S1*01	IGHJ2*01	494	22.27
IGHV3S1*01	IGHJ3*01	634	13.56
IGHV3S1*01	IGHJ4*01	5139	17.28
IGHV3S1*01	IGHJ6*01	1581	19.77
IGHV3S1*01	IGHJ7*01	1365	17.02
IGHV3S25*01	IGHJ2*01	502	17.04
IGHV3S25*01	IGHJ3*01	615	15.28
IGHV3S25*01	IGHJ4*01	8780	17.42
IGHV3S25*01	IGHJ5*01	11	15.33
IGHV3S25*01	IGHJ6*01	3089	18.70
IGHV3S25*01	IGHJ7*01	1572	16.92
IGHV3S39*01	IGHJ2*01	10056	17.28
IGHV3S39*01	IGHJ3*01	2024	16.05
IGHV3S39*01	IGHJ4*01	68063	16.30
IGHV3S39*01	IGHJ5*01	121	14.37
IGHV3S39*01	IGHJ6*01	23720	17.00
IGHV3S39*01	IGHJ7*01	12782	16.11

IGHV3S41*01	IGHJ2*01	972	17.59
IGHV3S41*01	IGHJ3*01	372	17.60
IGHV3S41*01	IGHJ4*01	20068	17.49
IGHV3S41*01	IGHJ5*01	153	12.04
IGHV3S41*01	IGHJ6*01	5591	17.91
IGHV3S41*01	IGHJ7*01	3158	17.79
IGHV3S53*01	IGHJ2*01	14032	16.79
IGHV3S53*01	IGHJ3*01	3672	18.14
IGHV3S53*01	IGHJ4*01	196289	16.45
IGHV3S53*01	IGHJ5*01	268	16.67
IGHV3S53*01	IGHJ6*01	44552	17.10
IGHV3S53*01	IGHJ7*01	26543	16.10
IGHV3S61*01	IGHJ2*01	3094	17.22
IGHV3S61*01	IGHJ3*01	1441	19.08
IGHV3S61*01	IGHJ4*01	47560	16.86
IGHV3S61*01	IGHJ5*01	96	16.68
IGHV3S61*01	IGHJ6*01	13734	17.47
IGHV3S61*01	IGHJ7*01	10735	17.44
IGHV3S65*01	IGHJ2*01	24666	15.13
IGHV3S65*01	IGHJ3*01	11366	17.59
IGHV3S65*01	IGHJ4*01	510159	15.90
IGHV3S65*01	IGHJ6*01	150604	16.66
IGHV3S65*01	IGHJ7*01	124353	15.76
IGHV3S66*01	IGHJ2*01	1640	17.10
IGHV3S66*01	IGHJ3*01	1321	19.79
IGHV3S66*01	IGHJ4*01	41008	17.63
IGHV3S66*01	IGHJ6*01	9055	17.49
IGHV3S66*01	IGHJ7*01	12660	17.73
IGHV3S9*01	IGHJ2*01	262	15.34
IGHV3S9*01	IGHJ3*01	141	17.14
IGHV3S9*01	IGHJ4*01	5328	16.56
IGHV3S9*01	IGHJ6*01	3194	19.02

- a. The V germline genes containing hallmark amino acid residues are shown in bold.

Table S3. Spearman's rank correlation of germline usage, paring, and substitution preference for the five samples.

		Naïve-1		Naïve-2		Naïve-3		Immun-1		Immun-2	
		ρ	P	ρ	P	ρ	P	ρ	P	ρ	P
V usage	Naïve-1	1	0	0.94434647	8.83E-44	0.86700222	4.71E-28	0.9273478	6.65E-39	0.94466337	6.93E-44
	Naïve-2	0.94434647	8.827E-44	1	0	0.8800986	6.93E-30	0.91723198	1.55E-36	0.93624121	2.748E-41
	Naïve-3	0.867002217	4.714E-28	0.8800986	6.93E-30	1	0	0.82767467	1.53E-23	0.84947733	6.954E-26
	Immun-1	0.927347802	6.652E-39	0.917232	1.552E-36	0.8276747	1.53E-23	1	0	0.9332365	1.91E-40
	Immun-2	0.944663373	6.932E-44	0.9362412	2.748E-41	0.8494773	6.954E-26	0.9332365	1.91E-40	1	0
D usage	Naïve-1	1	0	0.96666667	2.155E-05	0.8	0.0096279	0.68333333	0.042442	0.83333333	0.0052657
	Naïve-2	0.96666667	2.155E-05	1	0	0.76666667	0.015944	0.81666667	0.007225	0.9	0.0009431
	Naïve-3	0.8	0.0096279	0.7666667	0.015944	1	0	0.46666667	0.205386	0.8	0.0096279
	Immun-1	0.683333333	0.0424423	0.8166667	0.0072248	0.4666667	0.2053864	1	0	0.83333333	0.0052657
	Immun-2	0.833333333	0.0052657	0.9	0.0009431	0.8	0.0096279	0.8333333	0.005266	1	0
J usage	Naïve-1	1	0	0.97619048	3.314E-05	0.97619048	3.314E-05	0.97619048	3.31E-05	0.97619048	3.314E-05
	Naïve-2	0.976190476	3.314E-05	1	0	0.95238095	0.0002604	0.95238095	0.00026	0.95238095	0.0002604
	Naïve-3	0.976190476	3.314E-05	0.952381	0.0002604	1	0	1	0	1	0
	Immun-1	0.976190476	3.314E-05	0.952381	0.0002604	1	0	1	0	1	0
	Immun-2	0.976190476	3.314E-05	0.952381	0.0002604	1	0	1	0	1	0
VJ pairing	Naïve-1	1	0	0.90947089	7.97E-273	0.81964399	6.08E-174	0.8644189	3.3E-214	0.8947949	7.62E-251
	Naïve-2	0.909470892	7.97E-273	1	0	0.83216526	5.51E-184	0.86203911	1E-211	0.892112	3.53E-247
	Naïve-3	0.819643987	6.08E-174	0.8321653	5.51E-184	1	0	0.80075487	3.5E-160	0.82354836	5.47E-177
	Immun-1	0.864418895	3.26E-214	0.8620391	1E-211	0.8007549	3.48E-160	1	0	0.88430741	4.81E-237
	Immun-2	0.8947949	7.62E-251	0.892112	3.53E-247	0.8235484	5.47E-177	0.8843074	4.8E-237	1	0
DJ pairing	Naïve-1	1	0	0.94775345	1.846E-36	0.95119265	1.805E-37	0.91575864	1.95E-29	0.94033272	1.699E-34
	Naïve-2	0.947753455	1.846E-36	1	0	0.94579776	6.462E-36	0.90583388	8.11E-28	0.94834851	1.249E-36
	Naïve-3	0.95119265	1.805E-37	0.9457978	6.462E-36	1	0	0.92637934	2.09E-31	0.94996302	4.223E-37
	Immun-1	0.91575864	1.951E-29	0.9058339	8.105E-28	0.9263793	2.094E-31	1	0	0.90665672	6.047E-28
	Immun-2	0.94033272	1.699E-34	0.9483485	1.249E-36	0.949963	4.223E-37	0.9066567	6.05E-28	1	0
Amino acid substitution preference	Naïve-1	1	0	0.96238912	0	0.95838737	0	0.9469672	0	0.95198731	0
	Naïve-2	0.96238912	0	1	0	0.95955602	0	0.96277234	0	0.96508444	0
	Naïve-3	0.958387375	0	0.959556	0	1	0	0.94251419	5e-323	0.94829965	0
	Immun-1	0.946967195	0	0.9627723	0	0.9425142	5e-323	1	0	0.9659435	0
	Immun-2	0.951987313	0	0.9650844	0	0.9482997	0	0.9659435	0	1	0

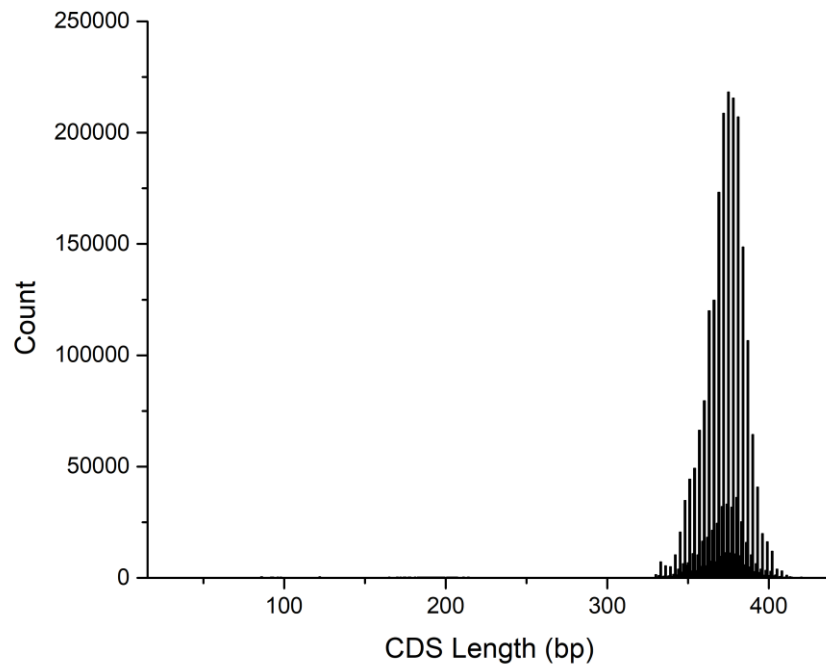


Fig. S1. Distribution of CDS length of VHHs.

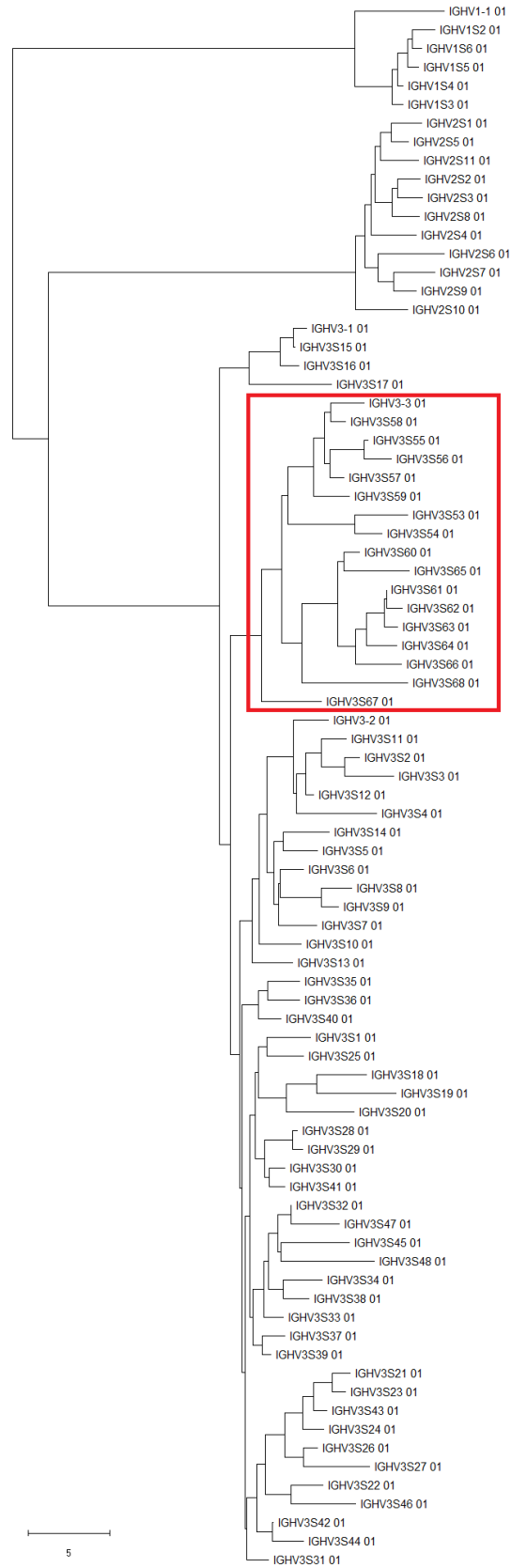


Fig. S2. Phylogenetic tree of V germline genes in the IgH locus of *Vicugna pacos*. The phylogenetic tree is built using MEGA version X. The germline genes, which contain at least two hallmark residues, are shown in the rectangular box.

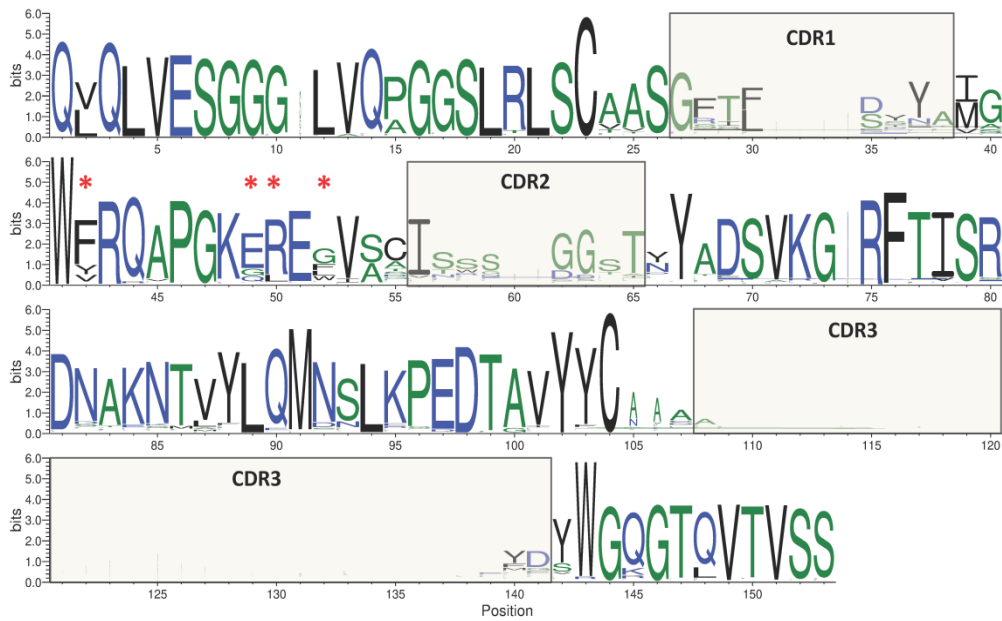


Fig. S3. Position specific score matrix (PSSM) of sample Naïve-1. The logo plot was built using WebLogo version 3.6. The hallmark residues are labeled with asterisk, corresponding to F37, E44, R45, and G47 in the Kabat numbering system.