

1

2 **Dynamic Co-Evolution and Interaction of Avian Leukosis Virus**

3 **Genetic Variants and Host Humoral Immunity**

4 Xuan Dong^{§,1}, Fanfeng Meng^{§,1}, Tao Hu², Sidi Ju¹, Yang Li¹, Peng Sun¹, Yixin Wang¹,
5 Wenqing Chen¹, Fushou Zhang¹, Hongqin Su¹, Sifei Li¹, He Cui¹, Junxia Chen¹, Shuzhen
6 Xu¹, Lichun Fang¹, Huaibiao Luan¹, Zhenjie Zhang², Shuang Chang¹, Jianliang Li¹, Lei
7 Wang¹, Peng Zhao¹, Weifeng Shi^{*,2} and Zhizhong Cui^{*,1}

8

9 ¹College of Veterinary Medicine, Shandong Agricultural University, Taian, Shandong, China

10 ²Institute of Pathogen Biology, Taishan Medical College, Taian, Shandong, China

11 [§]These authors contributed equally to this work.

12 ^{*}**Corresponding author:** Zhizhong Cui, zzcui@sdau.edu.cn; Weifeng Shi, shiwf@ioz.ac.cn.

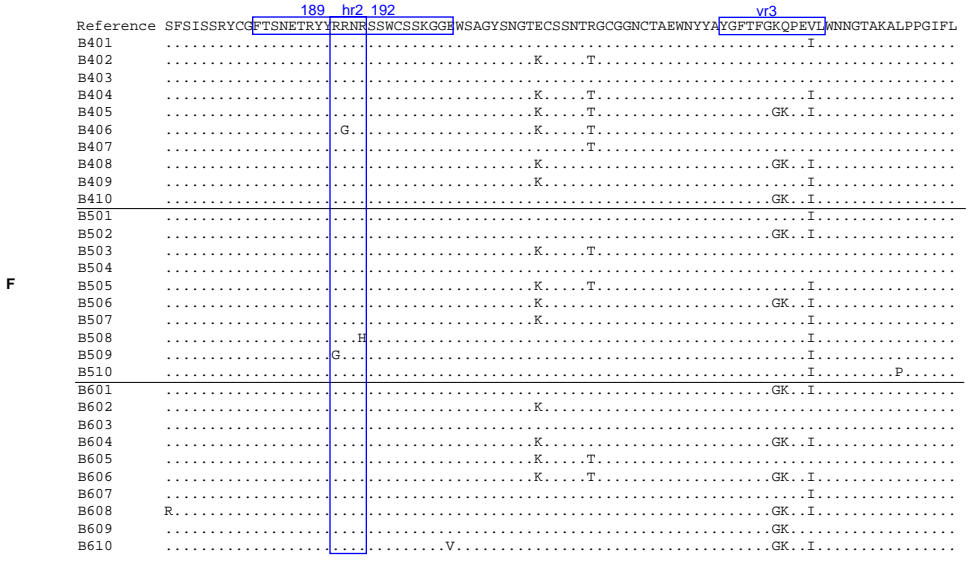
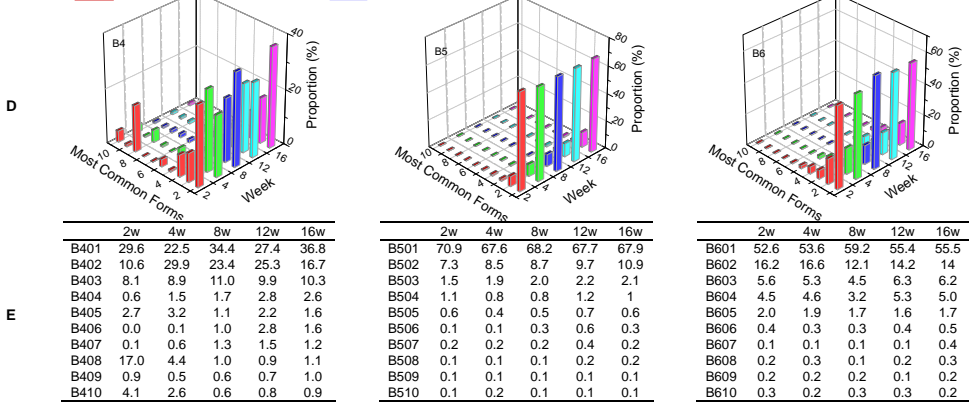
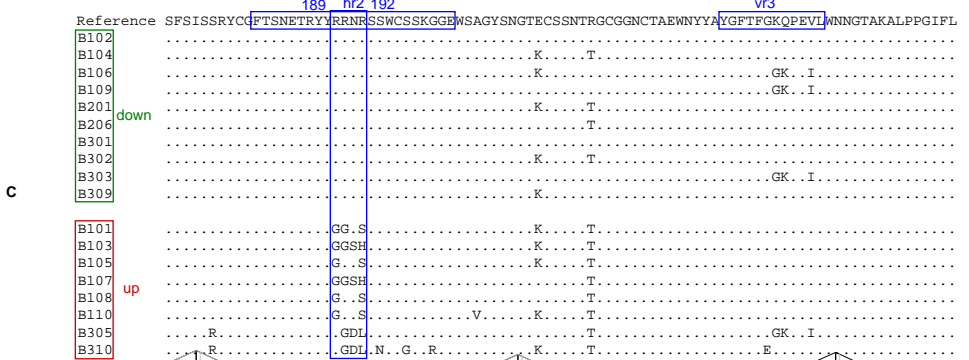
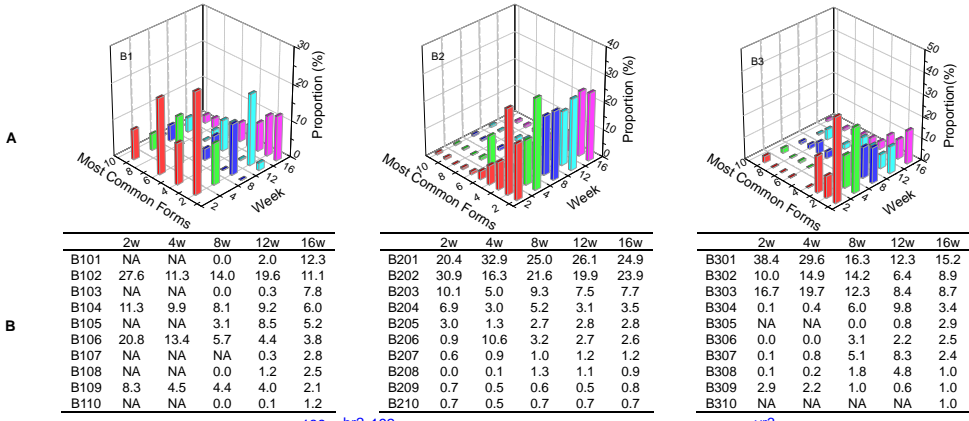


Fig S1. Evolutionary dynamics of gp85-B (hr2 and vr3 regions) genetic variants in antibody-positive and -negative chickens during the first 16 weeks.

Chickens 1-3 were antibody-positive, and 4-6 were antibody-negative, as in Figure 1. (A and D) The proportions of the top ten variants of chickens 1-3 and 4-6 at week 16. (B and E) The dynamic proportional changes in top ten variants at week 16 in chickens 1-3 and 4-6, respectively. NA means no available data. (C and F) The amino acid substitutions of the variants of chickens 1-3 and 4-6, respectively. In the two panels, the representative variants were classified into two groups, down and up, in which the proportion of the variants decreases or increases over time, respectively. The most dominant variant when inoculated is used as the reference, and the amino acid positions are numbered from the first codon of the mature gp85 protein. The two previously identified variable regions are highlighted. The variants are ordered according to their rankings at week 16. For example, in B101, “B” refers to region B of the gp85 gene (gp85-B), the first number “1” means chicken 1, and “01” represents the top one variants at week 16; therefore, B101 represents the top one gp85-B variant of chicken 1 at week 16.

Figure S2

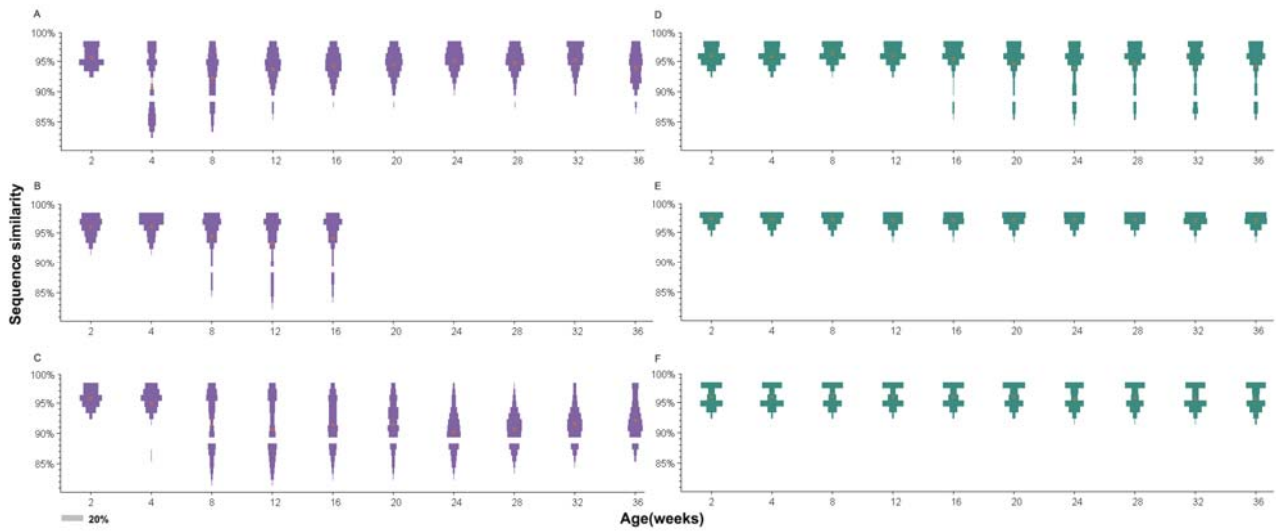


Fig S2. Distribution of sequence similarity of the gp85-B genetic variants during a 36-week period.

(A, B and C) The sequence similarity of chickens 1-3, respectively. (D, E and F) The sequence similarity of chickens 4-6, respectively. The x-axis of each panel represents different sampling time-points. The y-axis represents sequence similarity. The red bar represents 100% identity, and the orange circle represents the mean value of sequence similarity. The width of the grey bar represents the proportion of this sequence similarity value, and the legend is given in the lower left corner.

Figure S3

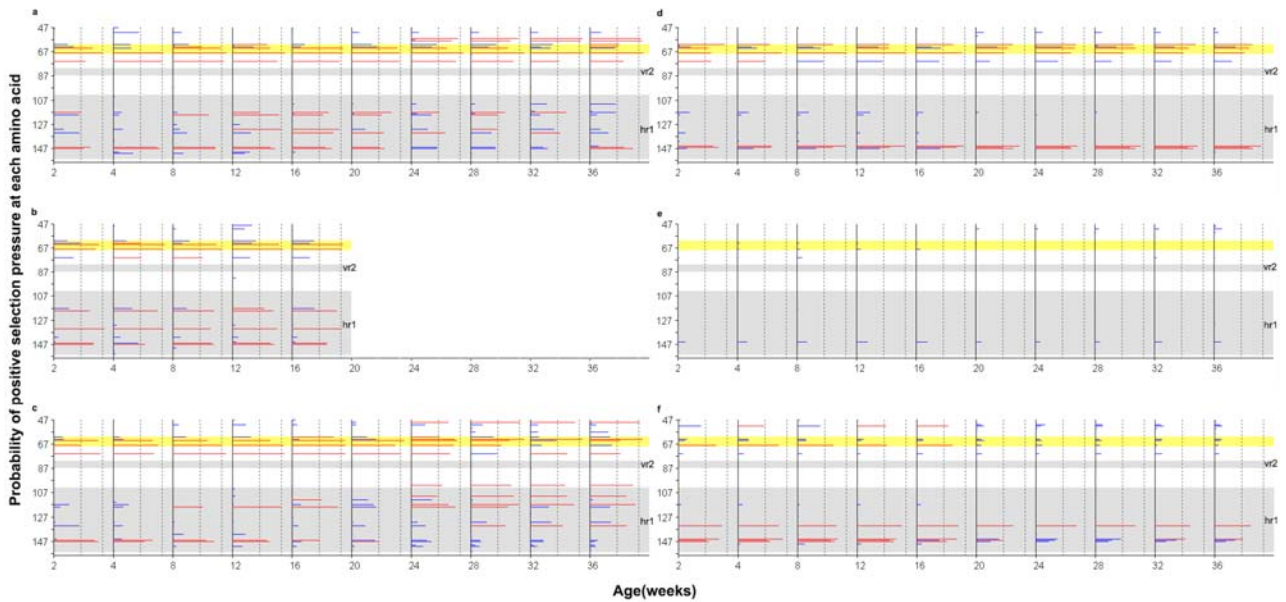


Fig S3. Site-by-site positive selection of the gp85-A gene based on the MEME method.

(A, B and C) The site-by-site positive selection value by the MEME method in chickens 1-3, respectively. (D, E and F) The site-by-site positive selection value by the MEME method in chickens 4-6, respectively. The x-axis of each panel represents different sampling time-points. The y-axis represents the amino acid positions of the gp85 protein. The horizontal lines in each panel represent the positively selected sites, and the length of the lines represents the proportion of the identified positively selected sites in the bootstrap re-sampling replicates. Red lines indicate that these amino acid sites are positively selected in at least 50% of the bootstrap re-sampling replicates, and blue lines indicate positive selection of less than 50%. The two vertical dashed lines represent the 50% and 90% thresholds. The two previously described hypervariable regions, vr2 and hr1, are shown using a grey background. Positions 61-68, a novel hypervariable region, are highlighted using a yellow background.

Figure S4

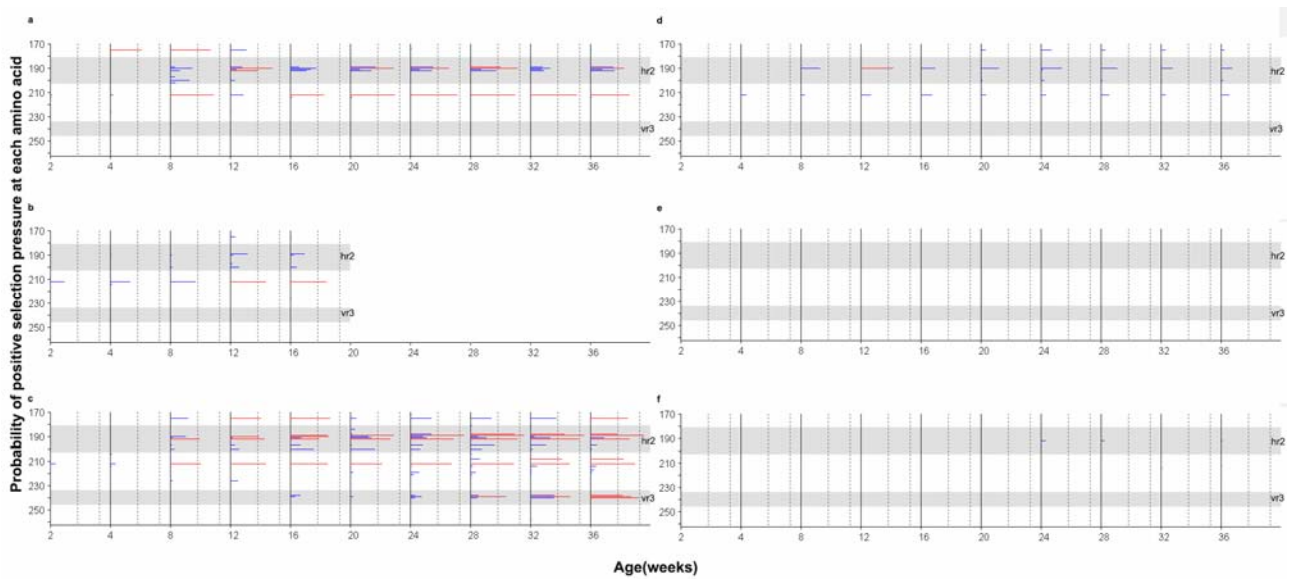


Fig S4. Site-by-site positive selection of the gp85-B gene based on the two rate FEL method.

(A, B and C) The site-by-site positive selection value by the two rate FEL method in chickens 1-3, respectively. (D, E and F) The site-by-site positive selection value by the two rate FEL method in chickens 4-6, respectively. The x-axis of each panel represents different sampling time-points. The y-axis represents the amino acid positions of the gp85 protein. The horizontal lines in each panel represent the positively selected sites, and the length of the lines represents the proportion of the identified positively selected sites in the bootstrap re-sampling replicates. Red lines indicate that these amino acid sites are positively selected in at least 50% of the bootstrap re-sampling replicates, and blue lines indicate positive selection of less than 50%. The two vertical dashed lines represent the 50% and 90% thresholds. The two previously described hypervariable regions, vr3 and hr2, are shown using a grey background.

Figure S5

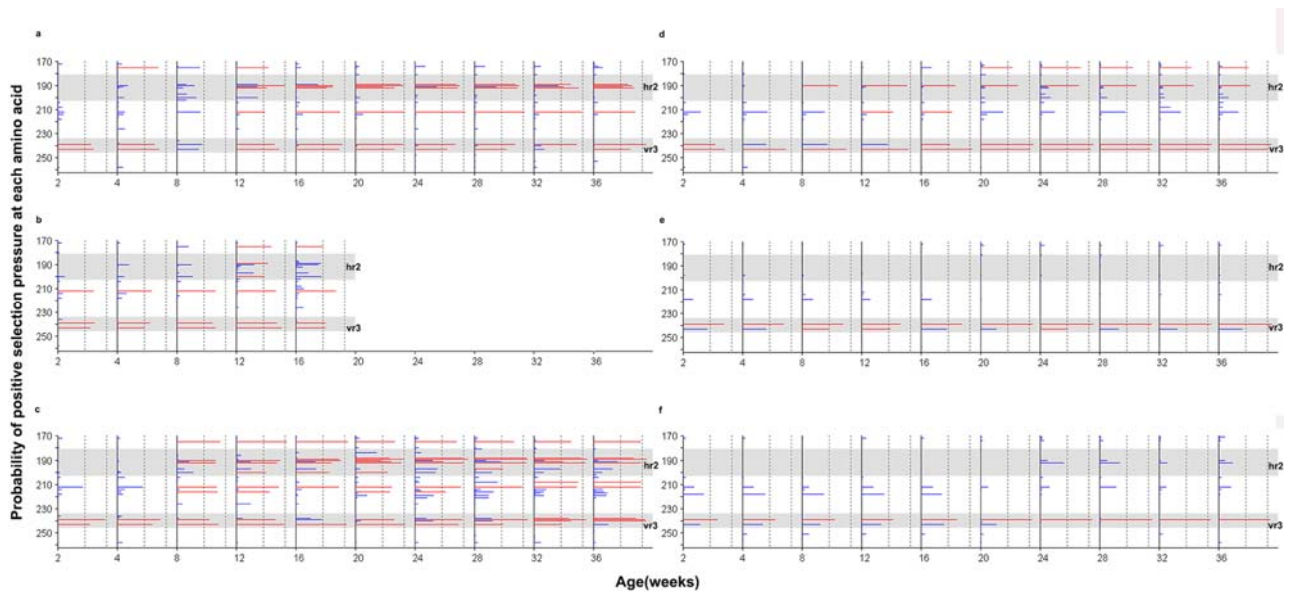


Fig S5. Site-by-site positive selection of the gp85-B gene based on the MEME method.

(A, B and C) The site-by-site positive selection value by the MEME method in chickens 1-3, respectively. (D, E and F) The site-by-site positive selection value by the MEME method in chickens 4-6, respectively. The x-axis of each panel represents different sampling time-points. The y-axis represents the amino acid positions of the gp85 protein. The horizontal lines in each panel represent the positively selected sites, and the length of the lines represents the proportion of the identified positively selected sites in the bootstrap re-sampling replicates. Red lines indicate that these amino acid sites are positively selected in at least 50% of the bootstrap re-sampling replicates, and blue lines indicate positive selection of less than 50%. The two vertical dashed lines represent the 50% and 90% thresholds. The two previously described hypervariable regions, vr3 and hr2, are shown using a grey background.

Table S1. The re-sampling strategy used in the positive selection analysis.

Number	Number of replicates	Number of sequences of each replicate
$N \geq 50000$	750	200
$40000 \leq N < 50000$	600	200
$30000 \leq N < 40000$	450	200
$N < 30000$	300	200

This table describes the re-sampling strategy used in the positive selection analysis. For example, if the number of clean reads in the dataset is no less than 50000, we performed 750 re-sampling replicates, with 200 sequences in each replicate. If the number of clean reads in the dataset is no less than 40000, but less than 50000, we performed 600 re-sampling replicates, with 200 sequences in each replicate. For each replicate, we performed the positive selection analysis and calculated the global ω and the average ω of all the replicates from a single dataset was estimated to represent the selection level for the dataset.

Table S2. The number of raw reads and clean reads of each sample.

Number	Segment A					Segment B					
	Raw reads	Clean reads	Number	Raw reads	Clean reads	Number	Raw reads	Clean reads	Number	Raw reads	Clean reads
A102	46394	45268	A402	57995	56422	B102	44259	43187	B402	28159	27387
A104	38959	37592	A404	44065	42501	B104	44224	42644	B404	50424	48580
A108	52667	44317	A408	21401	20943	B108	24619	23873	B408	34521	33585
A112	34309	33615	A412	7553	7415	B112	48628	47374	B412	50438	49300
A116	16187	15747	A416	22046	21669	B116	43859	42668	B416	29575	28988
A120	37490	34644	A420	28016	26945	B120	17157	16701	B420	7522	7324
A124	32238	31183	A424	47797	46394	B124	17440	17082	B424	31096	30372
A128	32940	31689	A428	29394	28571	B128	12306	12038	B428	19583	19211
A132	46835	45086	A432	24685	23910	B132	11849	11481	B432	10549	10311
A136	25226	23308	A436	41012	39680	B136	7584	7342	B436	17120	16648
A202	42711	41637	A502	15064	14773	B202	40370	39302	B502	18992	18653
A204	10951	10614	A504	44068	42924	B204	36341	35139	B504	47633	46368
A208	36719	36073	A508	14791	14477	B208	37063	36232	B508	31911	31267
A212	3081	3028	A512	38120	37549	B212	48486	47389	B512	48168	47391
A216	33254	32604	A516	38589	37993	B216	41619	40651	B516	38648	38006
			A520	48578	47284				B520	13781	13486
			A524	43742	42877				B524	11856	11662
			A528	24393	23891				B528	8529	8388
			A532	43181	42147				B532	14338	14069
			A536	23005	22173				B536	12329	12073
A302	44164	43092	A602	37799	37055	B302	52977	51698	B602	41360	40420
A304	20522	19921	A604	29273	28781	B304	33163	32031	B604	34250	33629
A308	42011	35350	A608	16051	15781	B308	45770	44615	B608	18006	17671
A312	41748	39956	A612	35969	33077	B312	41373	40410	B612	35631	35038
A316	10735	10534	A616	19509	19213	B316	10988	10767	B616	20914	20545
A320	20369	19422	A620	26176	25427	B320	9027	8755	B620	6287	6150
A324	22155	21222	A624	21189	20686	B324	7034	6794	B624	12401	12162
A328	22436	21525	A628	45032	43819	B328	5212	4991	B628	9949	9772
A332	15692	14936	A632	41373	40109	B332	12203	11617	B632	25171	24629
A336	34528	33016	A636	36206	34758	B336	12845	12264	B636	13323	12893

In this table, the variants is named as follows. For example, A102 means region A of the gp85 gene (gp85-A), in which the first number “1” means chicken 1 and 02 represents that this sample is collected at week 2. Therefore, B424 represents the number of reads of gp85-B at week 24 of chicken 4.

