chaVoc	F	K	G	F	R	G	G	L	D	V	S	H	G	Q	T	G	A	E
egrGar	F	K	G	F	R	G	G	L	D	V	S	H	G	Q	T	G	A	E
pygAde	F	K	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
aptFor	F	K	G	F	R	G	G	L	D	V	S	H	G	Q	T	G	A	E
nipNip	F	K	G	F	R	G	G	L	D	V	S	H	G	Q	T	G	A	E
anaPla	F	K	G	F	R	G	G	L	D	V	S	H	G	Q	T	G	A	E
melGal	F	K	G	F	R	G	G	L	D	V	S	H	G	Q	T	G	A	E
galGal	F	K	G	F	R	G	G	L	D	V	S	H	G	Q	T	G	V	E
strCam	F	K	G	F	R	G	G	L	D	V	S	H	G	Q	T	G	A	E
aptRow	F	K	G	F	R	G	G	L	D	V	S	H	G	Q	T	G	A	E
aptHaa	F	K	G	F	R	G	G	L	D	V	S	H	G	Q	T	G	A	E
aptOwe	F	K	G	F	R	G	G	L	D	V	S	H	G	Q	T	G	A	E
rhePen	F	K	G	F	R	G	G	L	D	V	S	H	G	Q	T	G	A	E
rheAme	F	K	G	F	R	G	G	L	D	V	S	H	G	Q	T	G	A	E
droNov	F	K	G	F	R	G	G	L	D	V	S	H	G	Q	T	G	A	E
casCas	F	K	G	F	R	G	G	L	D	V	S	H	G	Q	T	G	A	E
eudEle	F	K	G	F	R	G	G	L	D	V	S	H	G	Q	T	G	A	E
notPer	F	K	G	F	R	G	G	L	D	V	S	H	G	Q	T	G	A	E
tinGut	F	K	G	W	S	G	G	L	-	-	-	-	-	-	-	-	-	-
cryCin	F	K	G	F	R	G	G	L	D	V	S	H	G	Q	T	G	A	E
aquChr	F	K	G	F	R	G	G	L	D	V	S	H	G	Q	T	G	A	E
halLeu	F	K	G	F	R	G	G	L	D	V	S	H	G	Q	T	G	A	E
lepDis	F	K	G	F	R	G	G	L	D	V	S	H	G	Q	T	G	A	E
colLiv	F	K	G	F	R	G	G	L	D	V	S	H	G	Q	T	G	T	E
cucCan	F	K	G	F	R	G	G	L	D	V	S	H	G	Q	T	G	A	E
picPub	F	K	G	-	P	G	H	L	E	L	-	#	A	A	E	G	A	E
calAnn	F	K	G	F	R	G	G	L	D	V	S	H	G	Q	T	G	A	E
chaPel	F	K	G	F	R	G	G	L	D	V	S	H	G	Q	T	G	A	E
falPer	F	K	G	F	R	G	G	L	D	V	S	H	G	Q	T	G	A	E
melUnd	F	K	G	F	R	G	G	L	D	V	S	H	G	Q	T	G	A	E
corBra	F	K	G	F	R	G	G	L	D	V	S	H	G	Q	T	G	A	E
pseHum	F	K	G	F	R	G	G	L	D	V	S	H	G	Q	T	G	A	E
taeGut	F	K	G	F	R	G	G	L	D	V	S	H	G	Q	T	G	V	E
geoFor	F	K	G	F	R	G	G	L	D	V	S	H	G	Q	T	G	V	E
ficAlb	F	K	-	-	-	-	-	-	-	-	-	-	G	Q	-	-	-	-
serCan	F	K	G	F	R	G	G	L	D	V	S	H	G	Q	T	G	V	E

MALMA_AF344483	A	R	E	G	N	E	I	I	R	E	A	S	K	W	S	P	E	L	A	A	A	C	E	V	W	K	E	I	K	F
MALMA_AF344519	A	R	E	G	N	E	I	I	R	E	A	S	K	W	S	P	E	L	A	A	A	C	E	V	W	K	E	I	K	F
MALMA_AF344467	A	R	E	G	N	E	I	I	R	E	A	S	K	W	S	P	E	L	A	A	A	C	E	V	W	K	E	I	K	F
MALMA_AF344470	A	R	E	G	N	E	I	I	R	E	A	A	K	W	S	P	E	L	A	A	A	C	E	V	W	K	E	I	K	F
MALMA_AF344507	A	R	E	G	N	E	I	I	R	E	A	S	K	W	S	P	E	L	A	A	A	C	E	V	W	K	E	I	K	F
BBOBB_AJ235773	A	R	E	G	N	E	I	I	R	E	A	S	N	W	S	P	E	L	A	A	A	C	A	V	W	K	E	I	K	F
HANCO_AF323190	A	R	E	G	N	E	I	I	R	E	A	S	K	W	S	P	E	L	A	A	A	C	E	V	W	K	E	I	K	F
HANCO_AF323188	A	R	E	G	N	E	I	#	R	E	A	S	K	W	S	P	E	L	A	A	A	C	E	V	W	K	E	I	K	F
CRNCO_AY530929	A	R	E	G	N	E	I	I	R	E	A	S	K	W	S	P	E	L	A	A	A	C	E	V	W	K	E	I	K	F
CRNCO_L14395	A	#	E	G	N	E	I	I	R	E	A	S	K	W	S	P	E	L	A	A	A	C	E	V	W	K	E	I	K	F
CRNCO_AY530927	A	R	E	G	N	E	I	I	R	E	A	S	K	W	S	P	E	L	A	A	A	C	E	V	W	K	E	I	K	#
CRNCO_AF190433	A	S	E	G	N	E	I	I	R	E	A	S	K	W	S	P	E	L	A	A	A	C	E	V	W	K	E	I	K	F
CRNCO_L11217	A	R	E	G	N	E	I	I	R	E	A	S	K	W	S	P	E	L	A	A	A	C	E	I	W	K	E	I	K	F
BERRA_AF203487	A	A	E	G	N	T	I	I	R	E	A	S	K	W	S	P	E	L	A	A	A	C	E	V	W	K	E	I	R	F
COMMY_AF281478	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
PCIER_AJ428893	A	R	E	G	N	E	I	I	R	E	A	S	K	W	S	P	E	L	A	A	A	C	E	V	W	K	A	I	K	F
MGVER_Z80200	A	R	E	G	N	E	I	I	R	E	A	S	K	W	S	P	E	L	A	A	A	C	E	I	W	K	E	I	K	F
SLAER_AF421109	A	R	E	G	N	E	I	I	R	E	A	C	K	W	S	P	E	L	A	A	A	C	A	V	W	K	E	I	K	F
TSTER_Z83149	A	R	#	G	N	E	I	I	R	E	A	S	K	W	S	P	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SYMER_Z80192	A	R	E	G	N	E	I	I	P	A	A	S	K	W	S	P	E	L	A	A	A	C	E	I	W	K	E	I	K	F
SLAER_AF421108	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
PPLER_AF419239	A	R	E	G	N	E	I	I	R	E	A	S	K	W	S	P	E	L	A	A	A	C	E	V	W	K	E	I	Q	F
SPTER_Z83135	A	R	E	G	N	E	I	I	R	E	A	T	K	W	S	P	E	L	A	A	A	C	E	V	W	K	E	I	K	F
LECER_AF077656	A	R	E	G	N	E	I	I	-	-	-	-	-	-	-	-	-	-	-	-	-	-	R	E	#	S	K	W	S	
LECER_Z80183	A	R	E	G	N	E	I	I	R	E	A	S	K	W	S	P	E	L	A	A	A	C	#	V	W	K	E	I	K	F
LECER_Z80202	A	R	E	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
LECER_Z80180	A	R	Q	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
LECER_AF077652	A	R	E	G	N	E	I	I	R	E	A	S	K	W	S	P	E	L	A	A	A	C	E	V	W	K	E	I	K	F
LECER_AF077653	A	R	E	G	N	E	I	I	R	E	A	S	K	W	S	P	E	L	A	A	A	C	E	V	W	K	E	I	K	F
LECER_Z80177	A	R	E	G	N	E	I	I	R	#	A	S	K	W	S	P	E	L	A	A	V	C	E	V	W	K	E	I	K	F
LECER_Z80181	A	R	E	G	N	E	I	I	R	A	A	S	K	W	S	P	E	L	A	A	A	C	E	V	W	K	E	I	K	F
LECER_AF022134	A	R	E	G	N	E	I	I	R	A	A	S	K	W	S	P	E	L	A	A	A	C	E	V	W	K	E	I	K	F
PMOER_Z83144	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
THPER_AF213818	A	R	E	G	N	E	I	I	R	E	A	S	K	W	S	P	E	L	A	A	A	C	E	V	W	K	E	I	K	F
THPER_AF213819	A	R	E	G	N	E	I	I	R	E	A	C	K	W	S	P	E	L	A	A	A	C	E	V	W	K	E	I	N	L
THPER_AF213817	A	R	E	G	N	E	I	I	R	E	A	S	K	W	S	P	E	L	A	A	A	C	E	V	W	K	E	I	K	F
THPER_AF213815	A	R	E	G	N	E	I	I	R	A	A	S	K	W	S	P	E	L	A	A	A	C	E	V	W	#	E	I	K	F
PRMER_AF395005	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
PRMER_AF394965	A	R	E	G	N	E	V	I	R	E	A	C	K	W	S	P	E	L	A	A	A	C	E	V	W	K	E	I	K	F
PRMER_AF395006	A	R	E	G	N	E	V	I	R	E	A	C	K	W	S	P	E	L	A	A	A	C	E	V	W	K	E	I	K	F
PRMER_AF213794	A	R	E	G	N	E	V	I	R	E	A	C	K	W	S	P	G	L	A	A	A	C	E	V	W	K	E	S	K	F

2019

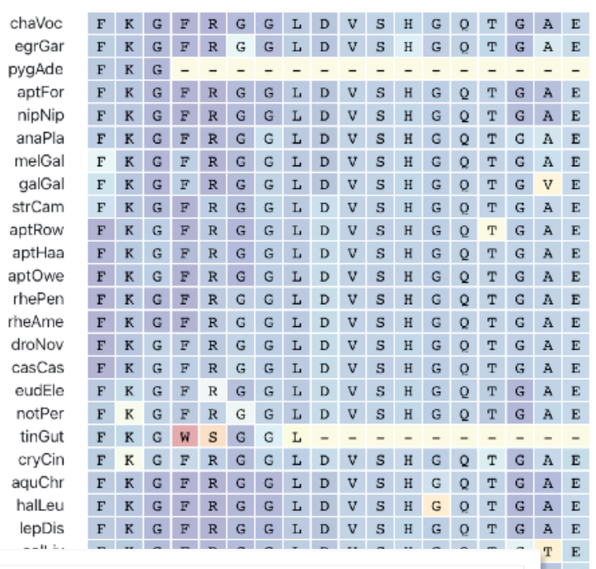
2021

Sample alignments from recently published papers

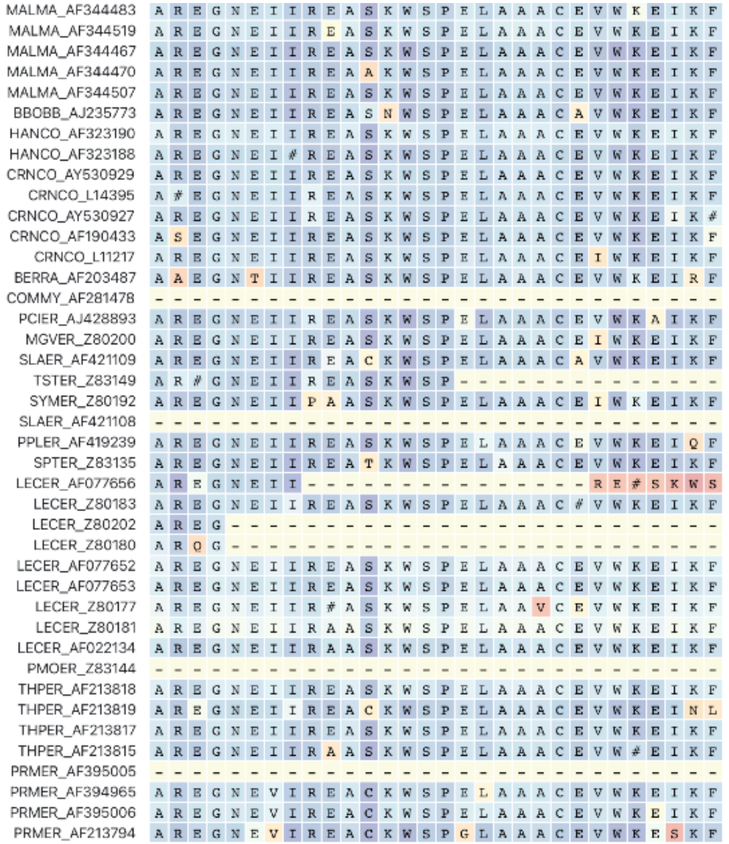
These 45,367 hierarchical orthologous groups, or HOGs, were filtered to retain 16,151 HOGs with sequences for at least four species. Protein sequences were aligned with MAFFT v. 7.245 (Katoh and Standley, 2013), and filtered in three steps. First, entire columns were excluded if missing in more than 30% of species, had sequence in fewer than 10 taxa, or was missing in two of the three of the main taxonomic groups (paleognaths, neognaths, or non-avian outgroups). Second, poorly aligned regions were masked according to Jarvis et al. (2014) using a sliding-window similarity approach. Third, columns were removed using the same criteria as the first round. Next, entire sequences were removed from each alignment if they were over 50% shorter than their pre-filtered length or contained excess gaps. Finally, entire HOGs were removed if they contained more than three sequences for any species, did not have more than 1.5x sequences for the given number of species present in the alignment, or were less than 100 base pairs long. Nucleotide sequences for all remaining HOGs were aligned with the codon model in Prank v. 150803 (Löytynoja and Goldman, 2008). In total, 11,247 HOGs remained after all alignment and filtering steps.

After the application of “due diligence” alignment masking and filtering techniques

Sample alignments from recently published papers



These 45,367 hierarchical orthologous groups, or HOGs, were filtered to retain 16,151 HOGs with sequences for at least four species. Protein sequences were aligned with MAFFT v. 7.245 (Katoh and Standley, 2013), and filtered in three steps. First, entire columns were excluded if missing in more than 30% of species, had sequence in fewer than 10 taxa, or was missing in two of the three of the main taxonomic groups (paleognaths, neognaths, or non-avian outgroups). Second, poorly aligned regions were masked according to Jarvis et al. (2014) using a sliding-window similarity approach. Third, columns were removed using the same criteria as the first round. Next, entire sequences were removed from each alignment if they were over 50% shorter than their pre-filtered length or contained excess gaps. Finally, entire HOGs were removed if they contained more than three sequences for any species, did not have more than 1.5x sequences for the given number of species present in the alignment, or were less than 100 base pairs long. Nucleotide sequences for all remaining HOGs were aligned with the codon model in Prank v. 150803 (Löytynoja and Goldman, 2008). In total, 11,247 HOGs remained after all alignment and filtering steps.



2021

After the application of “due diligence” alignment masking and filtering techniques

WDR72

Research Article
Evolutionary Biology, Genetics and Genomics

Immune genes are hotspots of shared positive selection across birds and mammals

Allison J Shultz, Timothy B Sackton

Harvard University, United States

Jan 8, 2019 · https://doi.org/10.7554/eLife.41815