Detection of Molecular Signatures of Selection at Production Genes for Eggs and Meat Production using dN/dS

Eunice Wainaina, Daniel Kariuki & Caroline Sigei
Department of Biochemistry,
Jomo Kenyatta University of Agriculture and Technology (JKUAT),
P.O. Box 62000-00200, City Square 00200, Nairobi-Kenya;
E-Mails: wainainaeunice1@gmail.com (EW); dkariuki@jkuat.ac.ke (DK);
carolinecsigey@gmail.com (CS)

Simon Maina & Philip Oyier
Department of Information Technology,
Jomo Kenyatta University of Agriculture and Technology (JKUAT),
P.O. Box 62000-00200, City Square 00200, Nairobi-Kenya;
E-Mails: msimog@gmail.com (SM); oyier@itc.jkuat.ac.ke (PO)

Jacqueline Lichoti
State Department of Livestock,
Ministry of Agriculture, Livestock and Fisheries,
P.O. Box 29089-00625, Kangemi-Kenya;
E-Mail: kasiiti.orengo@gmail.com (JL)

Sheila Ommeh
Institute for Biotechnology Research (IBR),
Jomo Kenyatta University of Agriculture and Technology (JKUAT),
P.O. Box 62000, City Square 00200, Nairobi-Kenya;
Correspondence E-Mail: sommeh@jkuat.ac.ke (SO)

Abstract: *Livestock are widely reared in Africa. The livestock are variable phenotypically due to Natural Selection. However, improvement of their productivity is still a challenge. The aim of this study was to perform analysis for signatures of selection at candidate genes for egg and meat production. Genes for egg production were prolactin (prl), vasoactive intestinal peptide 1(vip1) and vasoactive intestinal peptide receptor 1(vipr1) while genes for meat production were growth hormone (gh), growth hormone receptor (ghr), insulin like growth factor I (igf1) and insulin like growth factor I receptor (igf1r). A reciprocal BLASTp using BLOSUM 62 substitution matrix was performed. Thereafter, Multiple Sequence Alignment using MUSCLE was performed and phylogenetic trees inferred using FastME. Finally, Phylogenetic analysis was done by estimating the rate of non-synonymous to synonymous substitution in the lineages and amino acid sites using codon-based substitution models of PAML4. Likelihood Ratio Tests were performed to compare the models followed by Bayes Empirical Bayes analysis to predict signatures of selection. In vipr1 and growth hormone receptor, all the lineages were shown to be under purifying selection while in prolactin, the lineages leading to poultry and other birds were under positive selection. In growth hormone, positive selection was detected in the artiodactyls. Insulin like growth factor I receptor had positive selection on amino acid isoleucine at position*

460 located on Receptor L domain. The positive selection on igf1r may be used as a molecular marker in genetic improvement of growth of livestock.

Keywords: *Adaptive evolution, candidate genes, computational molecular evolution, dN/dS, indigenous poultry, in silico*

1. Introduction

Livestock are largely reared in Africa by the resource-limited majority of who live in the rural areas (Sserugga et al., 2014). Chickens account for largest livestock species reared by man with indigenous chickens forming 70% of the chicken population (Olwande et al., 2010). The livestock fulfill multiple roles like provision of meat, milk, eggs, manure and draught power. However, there are challenges faced by developing countries to improve their production to meet the rising demand without depleting the natural resources. In the developed countries, biotechnology has offered unprecedented opportunities to improve agricultural production. In the contrary, there are few success stories in the application of biotechnology to improve livestock production in Africa. Here, improvement of livestock is not advanced (Dana et al., 2010). For example, in an attempt to improve indigenous poultry production, the government and some Non-Governmental Organization poultry programs have tried cross-breeding with commercial genotypes which have been unsuccessful as this leads to genetic dilution of indigenous genetic resources.

Livestock are highly variable phenotypically in size, skin color, live weights, egg and meat production among other traits (King'ori et al., 2010). This variation is as a result of individuals of a population changing to adapt to their environment through the process of Natural Selection. This variation provides an opportunity for livestock improvement through artificial selection. The availability of sequenced data and statistical tools brings hope to Africa as genetic improvement depends on access to genetic variation and effective methods for analyzing the data. The computational approach that has been used in this study is accurate as it provides evidence of the influence of the genes on the phenotype.

This may save time used in performance of many experiments in the attempt to get the desired phenotypes. A study done by Lwelamira et al (2008) indicated that it would take 5-10 generations of selection corresponding to about 3-6 years of selection to improve body weight at 16 weeks. More than 85% of people in developing countries live in rural areas. Majority of these people keep livestock and thus through breeders coming up with improved livestock, the farmers' economic situation will be improved.

Zhu et al., (2010) used the computational approach to detect positively selected sites in the Mx protein which are important markers that may be used in improving avian anti-viral activity. In this study, we used codon-based maximum likelihood models of PAML to detect signatures of selection at genes for egg and meat production.

2. Materials and Methods

2.1. Mining of sequence data

Cross-database ENTREZ searches in GenBank were used to retrieve either the full coding sequence or mRNA sequences of genes implicated for egg production; Prolactin, vasoactive intestinal peptide, and vasoactive intestinal peptide receptor and genes implicated for growth; Growth hormone, growth hormone receptor, insulin-like growth factor I and insulin-like growth factor I receptor. The chicken genome was used as the reference. This is available at the ENSEMBL annotated database.

2.2. Homology search using Reciprocal BLAST

To infer homology, a reciprocal BLAST was done using BLASTp of the NCBI. The search was done in the nr database using the amino acid sequence for each gene in the fasta format as the query. BLOSUM62 was used as the substitution matrix. The homologues that were selected were species with an E value greater than 1e-10. Only birds were selected as homologues in egg production genes while for growth, animals which are commonly eaten by man were selected. The sequences of the homologues were converted each to its corresponding coding sequence. The stop codons were thereafter removed manually to prevent them from interfering with subsequent analysis.

2.3. Multiple Sequence Alignment

Multiple Sequence Alignment softwares that are progressive in nature; ClustalX2 (Larkin et al., 2007), PRANK and MAFFT (Katoh and Standley, 2013) were used to align the different homologues for each gene. In addition, MUSCLE version 1.3.8.31-1 (Edgar, 2004) which is based on an iterative algorithm was used for the alignment. The alignments were visualized and edited using Jalview version 2.8 (Waterhouse et al., 2009) and Seaview. A comparison of the alignments from the four softwares was done. Alignments from MUSCLE were found to be better and were used for subsequent analysis.

2.4. Phylogeny

MEGA6 (Tamura et al., 2013) was used to model the amino acid substitution and the rate heterogeneity within the various homologues [11]. The model was selected using the Lowest Bayesian Information Criterion. Phylogenetic trees were then inferred using Nearest Neighbour Interchange with subtree pruning and regrafting of FastME2 (Lefort et al., 2015). 1000 bootstraps were used to test for the robustness of the branching topology. The trees were saved in Newick format. Figtree version 1.4.2 (Rambaut, 2014) was used for graphical visualization of the trees.

2.5. Likelihood ratio tests for molecular signatures of selection

Codon-based substitution models which were developed by Nielsen and Yang (1998) and Z Yang et al (2000) which are implemented in CODEML program of PAML were used to detect selection in lineage and in sites. For lineage, 2 codon-based models which are the model 0 and model 1 were used. For the site, the codon-based models used were model 7(beta) and model 8(beta and ώ). The PAML5 package (Ziheng Yang, 2007) was then used for phylogenetic analysis using maximum likelihood. CODEML was run on the command prompt. Analysis was done in two parts, first was the branch or lineage analysis and the second was the amino acid or site analysis. For the lineage analysis, the log likelihood ratio (LRT) was used to compare two

nested models: a null model that does not allow for any codons with ώ ratio of greater than 1(Model 0) against a more general model that does (Model 1) (Z. Yang, 1998). Twice the log likelihood difference between the two models was compared against a X^2 distribution to test the significance of the result obtained. For site analysis, twice the log likelihood difference between model7 (beta) with ω ratios of between 0 and 1 and model8 (beta & ω) that adds a class of sites with dN/dS of >1 were compared against a X^2 distribution to test the significance of the result obtained.

Bayes Empirical Bayes analysis was done to infer the category an amino acid site belongs to. Amino acid sites with a predetermined posterior probability of >0.95 belong to a site class that has a dN/dS ratio of > 1 and thus believed to be under positive selection (Ziheng et al., 2005).

2.6. Prediction of 3D structure of the proteins and visualization of the positively selected sites

The 3D structure of the protein encoded by Insulin Growth Factor1 receptor gene was predicted by sending the selected amino acid sequence to Raptor X structure prediction server (Källberg et al., 2012) which is a template-based modeling server. Thereafter, the positively selected site that was obtained from Bayes Empirical Bayes Analysis was selected on the sequence of the structure. Pymol version 1.7.6 (DeLano, 2002) was used to visualize the positively selected site on the protein structure.

3. Results and Discussion

The multiple sequence alignments showed the variation of sequences between the homologues of the different genes (Figure S1).

3.1. Selection Signatures at Production Genes for Eggs

3.1.1. Prolactin

From the LRT (Table 1) that was performed to compare the two nested models: one ratio model (M0) that does not allow ω >1 in the distribution and the alternative free ratio model (M1) that does allow, the hypothesis that there was heterogeneity in selection pressures was confirmed to be true from the P<0.0005 as shown in Table 1. From the phylogenetic tree in Figure 1, there were different selection pressures acting on the lineages as seen from the dN/dS values. Positive selection was detected in the two branches leading to poultry (dN/dS>1). The branch leading to the ancestor of the wild chicken, quails and common peafowl had a high dN/dS ratio of 3.64. However, there were no sites detected to be under positive selection from the BEB analysis. Prolactin is a pituitary hormone that has many and diverse roles that vary between species. These range from effects in mammalian reproduction to osmoregulation in fish. Although its role in birds is not well understood, it is believed to be responsible for the onset and maintenance of broodiness (Wilkanowska et al., 2014). Broodiness causes egg production to be low. Domestication or artificial selection is likely to cause positive selection in poultry prolactin (Cheng, 2010). The domestication brought about rapid changes and development, growth and behavior to fit the particular environment through adaptation. For example, commercial layers produce more than 300 eggs per year while their ancestor the jungle fowl lays about 4-6 eggs per year. The reason for the difference in production is because jungle fowls show incubation behavior whereas commercial breeds do not go broody as they have been subjected to intensive artificial selection to meet human needs.

Table 1. Likelihood Ratio Test (LRT) to detect selection in lineages and amino acid sites in prolactin gene. In the Lineages, P values of < 0.05 confirm heterogeneity in selection pressures while in Site analysis, this is an indication of positive selection which is confirmed through BEB analysis.

Gene	Model Analysis	2Δ1	X^2 value	Degrees of freedom	Probability value
Prolactin	Lineage Analysis (M0 vs M1)	2(-5519.413936+5449.921133)=138.9866	138.9866	139-71=68	<0.05
	Site Analysis (M7 vs M8)	2(-5469.7794-5466.9746)=5.6096	5.6096	74-72=2	>0.05

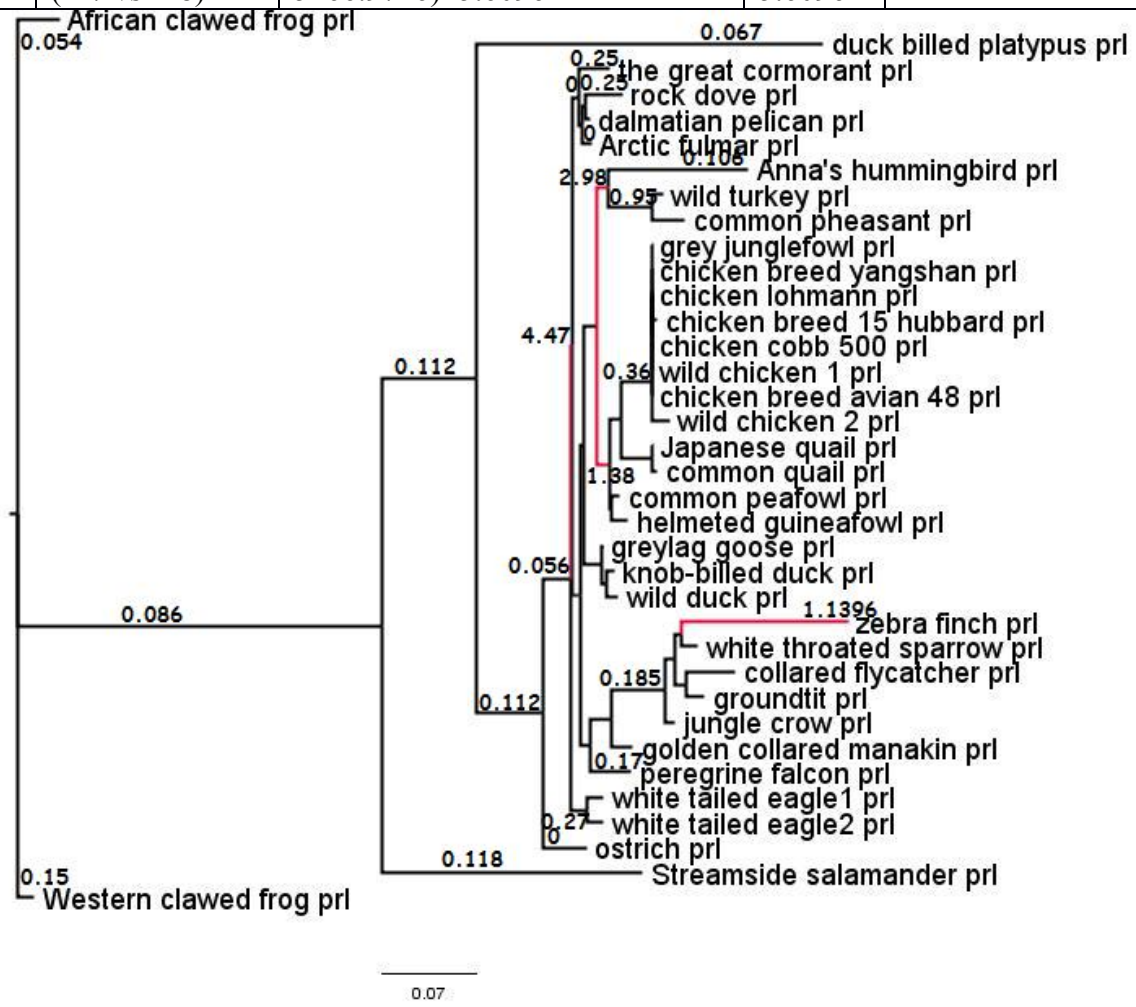

Figure 1. Phylogeny illustrating evolution of prolactin. The phylogram was constructed using distance-based FastME with 1000 bootstrap replicates to test the confidence of the topology. The branch lengths were estimated using maximum likelihood under the free ratio model (M1) which assumes an independent ω for each branch. dN/dS values are shown for each branch. The branches in red have a dN/dS>1 hence indicate positive selection

3.1.2. Vipr1

From the LRT results shown in Table 3, M1 fit significantly better with a P<0.0005, an indication that the lineages were under different selection pressures. The dN/dS values in the phylogenetic tree in Figure 2 indicated that vipr1 was under purifying selection (dN/dS<1). There were no amino acid sites detected to be under positive selection. Vipr1 gene is responsible for regulating broodiness in avian species (Chaiseha et al., 2004). From this study, it is likely that vipr1 gene is under purifying selection to maintain this function.

Table 3. Likelihood Ratio Test (LRT) to detect selection in lineages and amino acid sites in vasoactive intestinal peptide receptor 1 (vipr1) gene. In the Lineages, P values of < 0.05 confirm heterogeneity in selection pressures while in Site analysis, this is an indication of positive selection which is confirmed through BEB analysis.

Gene	Model Analysis	2Δ1	X^2 value	Degrees of freedom	Probability value
Vipr1	Lineage Analysis (M0 vs M1)	2(-9226.699685+9164.329783)=124.7	124.7	115-59=56	<0.05
	Site Analysis (M7 vs M8)	2(-9013.544+9005.761)=15.566	15.566	62-60=2	<0.05

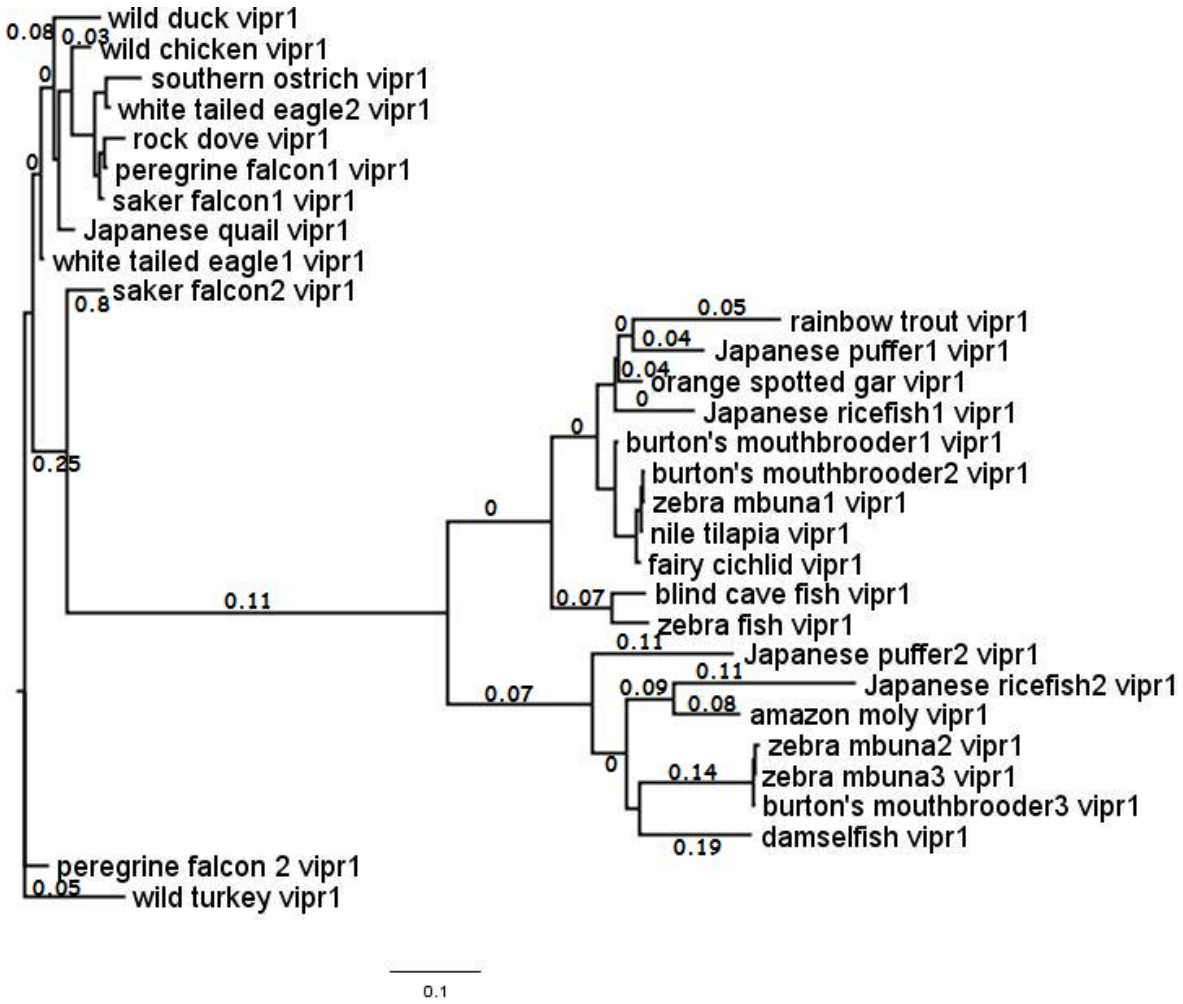

Figure 2. A phylogram showing evolution of vipr1. The branch lengths were estimated by maximum likelihood under the free ratio model (Model 1) which assumes an independent ω for each branch. The branch lengths are drawn in proportion to the expected numbers of nucleotide substitutions per codon. ω values are shown for each branch. Vipr1 is undergoing purifying selection as shown by the dN/dS values of <1.

3.2. Selection Signatures at Production Genes for Meat

3.2.1. Gh

The LRT of the free ratio model (M1) versus the one ratio model (M0) confirmed the hypothesis of heterogeneity in selection pressures in the lineages (Table 4) ($P<0.0005$). Positive selection was detected in the branch leading to cattle, sheep, goat and camel ($dN/dS>1$) (Figure 3). The Gh gene had no sites detected to be under positive selection. Growth hormone is a polypeptide hormone which is present in all vertebrates (Kawauchi et al., 2002). It has a crucial function in growth and promoting differentiation at different target sites. In birds, growth hormone has other secondary functions such as reproduction, egg production and aging (Zhao et al., 2004). The evolution of growth hormone is generally slow because of the important roles it plays and perhaps the constraints imposed by multiple functions. However, there are bursts of rapid change that occurred during the evolution of artiodactyls driven by Natural selection although there lacks a well defined associated functional change.

Studies done by Wallis (1994) also failed to detect any amino acid sites under positive selection despite of the positive selection detected in some lineages. According to him, saturation of sequence substitutions in some regions of the data was possibly the cause.

Table 4. Likelihood Ratio Test (LRT) to detect selection in lineages and amino acid sites in growth hormone (gh) gene. In the Lineages, P values of < 0.05 confirm heterogeneity in selection pressures while in Site analysis, this is an indication of positive selection which is confirmed through BEB analysis.

Gene	Model Analysis	2Δ1	X^2 value	Degrees of freedom	Probability value
Gh	Lineage Analysis (M0 vs M1)	2(-4704.661507+4661.203348)=86.9	86.9	111-57=54	<0.05
	Site Analysis (M7 vs M8)	2(-4634.798616+4634.799225)=0.001218	0.001218	60-58=2	>0.05

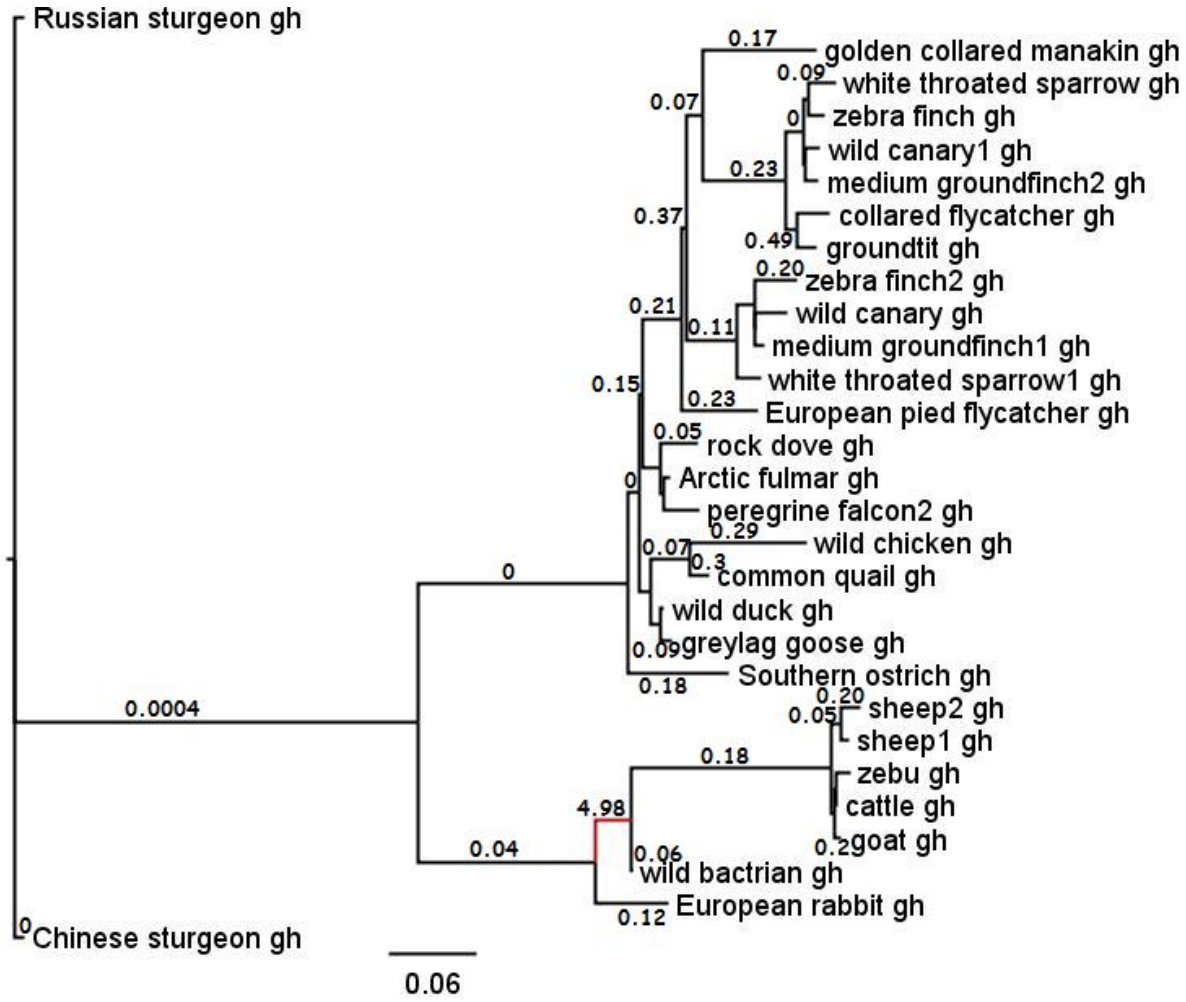

Figure 3. Phylogeny illustrating evolution of gh. The branch lengths were estimated using maximum likelihood under the free ratio model (M1) which assumes an independent ω for each branch. dN/dS values are shown for each branch. The branch in red has a dN/dS>1 hence indicates positive selection.

3.2.2. IGFIR

From the LRT, M1 fit significantly better than M0 (P<0.0005), an indication that there were different selection pressures acting on the lineages in igf1r (Table 7). The hormone was under purifying selection as all the dN/dS values were <1 (Figure 5). An LRT of the null model (M7 beta) which assumes a beta distribution for ω (in the interval 1< ω <1) and the most stringent M8 (beta& ω) which adds an extra class of sites with positive selection (ω>1) confirmed the hypothesis of variable selection pressures among amino acid sites. A Bayes Empirical Bayes analysis (Yang et al., 2005) identified positive selection on igf1r gene on amino acid Isoleucine at position 460 which had a posterior probability of 0.985 which is greater than the predetermined cut-off value of >0.95 for inferring positive selection. Figure 6 shows the position of the site in the Receptor L Domain of igf1r. In almost all proteins where positive selection is operating, only a few sites have been shown to be responsible for the adaptive evolution. Igf1r is a candidate gene for prenatal and postnatal growth, body composition and metabolism, skeletal characteristics and growth of adipose tissue and development (Lei et al., 2007). We postulate that the positive selection is as a result of breeding process in an attempt to increase growth. The L domains are important in binding insulin (Ward et al., 2000) hence the positive selection is a good marker in breeding future generations with improved growth.

There are few studies concentrating on signatures of selection at production genes as compared to disease genes.

Table 7. Likelihood Ratio Test (LRT) to detect selection in lineages and amino acid sites in insulin-like growth factor 1 receptor (igf1r) gene. In the Lineages, P values of < 0.05 confirm heterogeneity in selection pressures while in Site analysis, this is an indication of positive selection which is confirmed through BEB analysis.

Gene	Model Analysis	2Δ1	X^2 value	Degrees of freedom	Probability value
Igf1r	Lineage Analysis (M0 vs M1)	2(-12701.068427-12641.506653)=119.1	119.1	115-57=56	<0.05
	Site Analysis (M7 vs M8)	2(-12582.808851+12564.207151)=37.2	37.2	62-60=2	<0.05

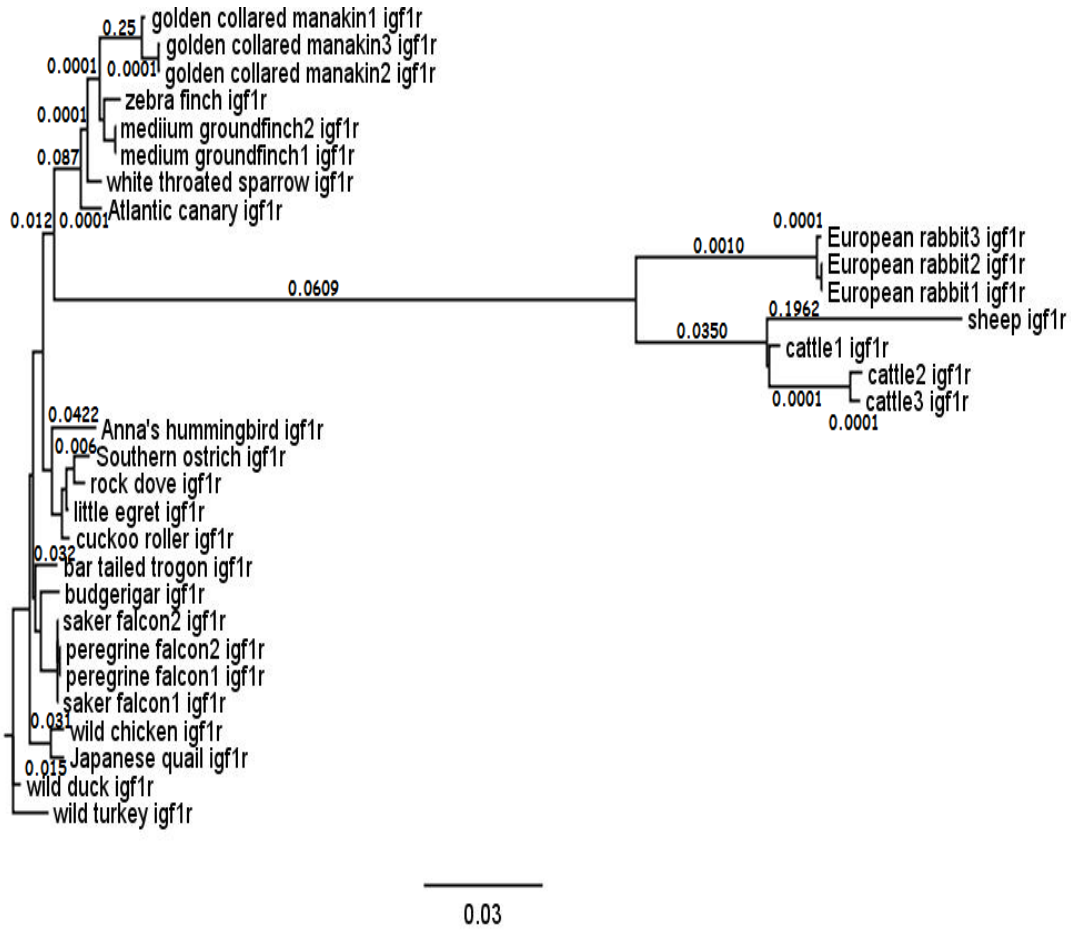

Figure 5. Phylogeny illustrating evolution of igf1r. The branch lengths were estimated using maximum likelihood under the free ratio model (M1) which assumes an independent ω for each branch. dN/dS values are shown for each branch. They indicate purifying selection as their values are < 1. The tree was visualized using Figtree v 1.4.2.

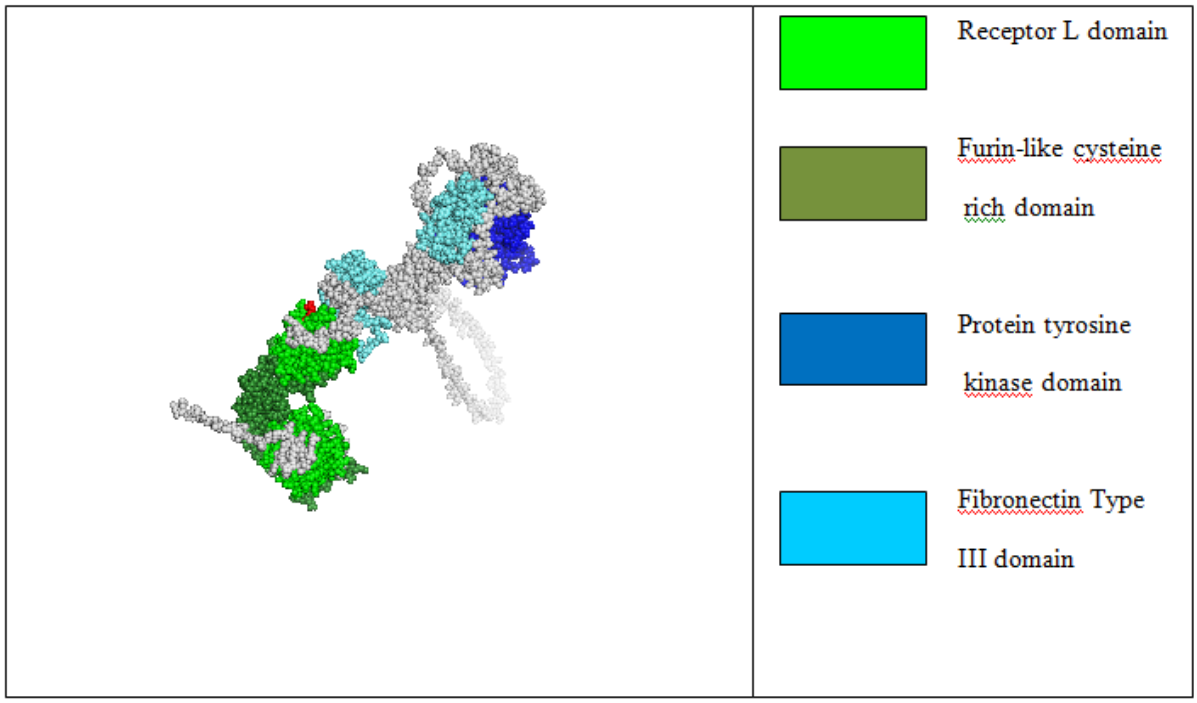

Figure 6. The 3D structure of igf1r as predicted by template-based Raptor X (Källberg et al., 2012). The site shown in red is predicted to be under positive selection by Model 8 (beta&ω). The site is located in the Receptor L domain which functions in binding insulin. Pymol v1.7.6 (DeLano, 2002) was used to dispay the structure.

5. Conclusions

The different proteins are important as they influence traits of economic importance. The computational analysis has shown that the proteins in this study have been influenced greatly by artificial selection during and after domestication. However, they have responded differently. Igf1r is important in growth and has a site under positive selection. Majority of chicken are reared for meat production. This is an important milestone in that *in vitro* and *in vivo* studies could be carried out to confirm the phenotype. This is a fast, accurate and more effective method that may be applied in breeding chickens that mature quickly and have certain meat quality traits that are desirable to farmers and customers.

Acknowledgments: The authors would like to acknowledge Jomo Kenyatta University of Agriculture and Technology, The Department of Veterinary Services, State Department of Livestock, Ministry of Agriculture, Livestock and Fisheries for making this project a success.

Author Contributions: SO conceived the study and designed the method. JL, DK, SM, MM, PO, EN and SO guided the research component. EW, CS, SO, PO and SM performed the analysis. All authors participated in the writing of the manuscript.

Conflicts of Interest: The authors declare that they have no competing interests.

Grant information

The authors wish to acknowledge financial support awarded to Dr. Sheila Ommeh by the Jomo Kenyatta University of Agriculture and Technology research grant no. JKU/2/4/RP/181.

Supplementary Material

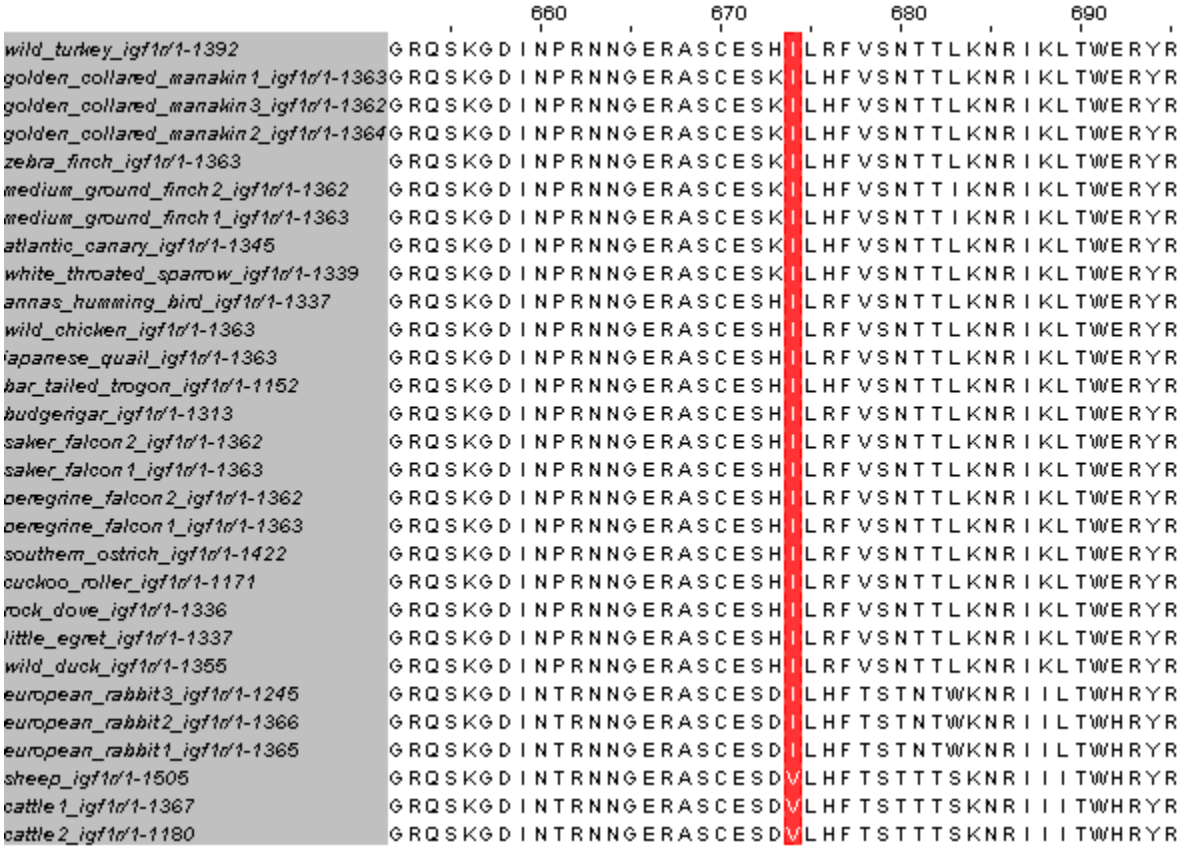

Figure S1: Multiple Sequence Alignment of igf1r sequences using MUSCLE v1.3.8.31-1 (Edgar, 2004). Model 8 (beta&ω) predicted positive selection in igf1r. The amino acid site with Isoleucine is identified to be under positive selection by BEB Analysis. The site is shown in red.

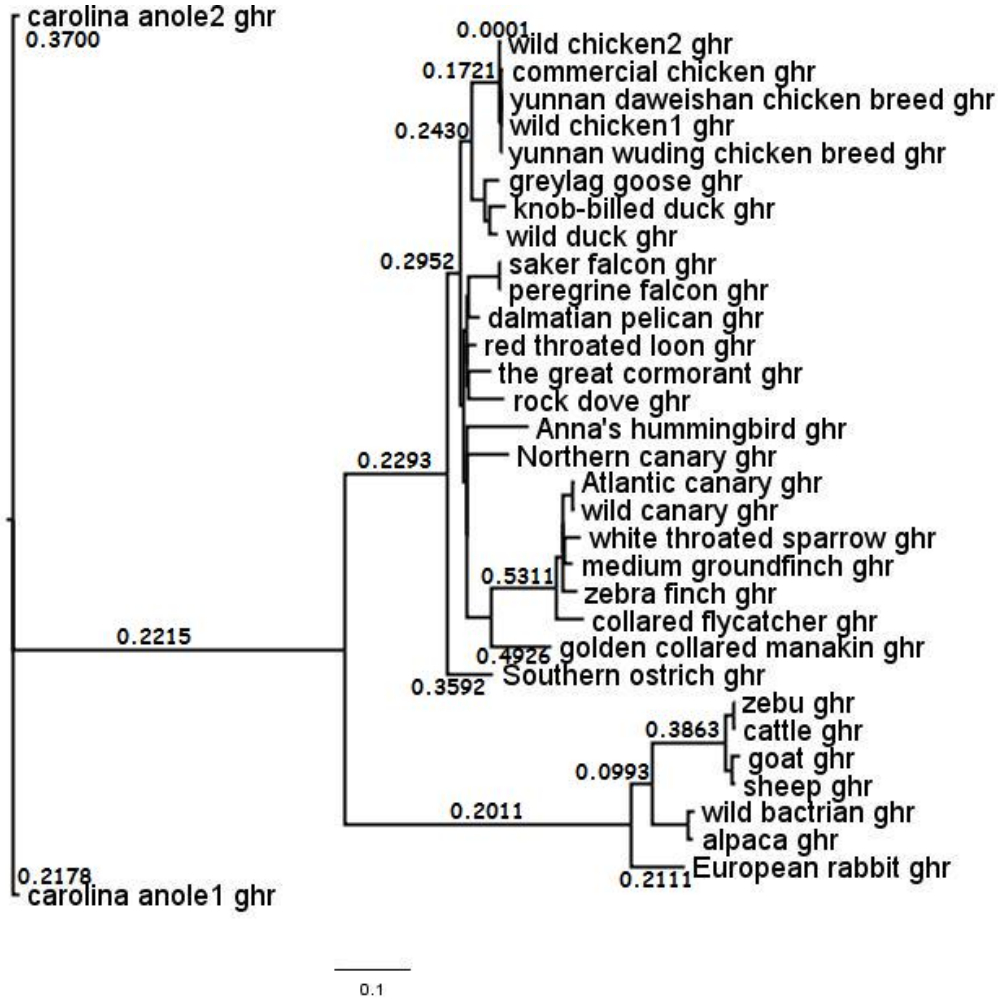

Figure S2: Phylogenetic tree illustrating evolution of growth hormone receptor. The branch lengths were estimated using maximum likelihood under the free ratio model (M1) which assumes an independent ω for each branch. dN/dS values are shown for each branch. They indicate purifying selection as their values are < 1. The tree was visualized using Figtree v 1.4.2.

Table S1: Likelihood Ratio Test (LRT) to detect selection in lineages and amino acid sites in vasoactive intestinal peptide (vip1) gene. In the Lineages, P values of > 0.05 are an indication that all lineages are under the same selective pressure i.e no heterogeneity of selective pressures.

Gene	Model Analysis	2Δ1	X^2 value	Degrees of freedom	Probability value
Vip1	Lineage Analysis (M0 vs M1)	2(-529.499090+514.296366)=30.41	30.41	115-59=56	>0.05

	Site Analysis (M7 vs M8)	2(-517.381707+517.979078)=1.194742	1.19	62-60=2	>0.05

Table S2: Likelihood Ratio Test (LRT) to detect selection in lineages and amino acid sites in growth hormone receptor gene (ghr). In the Lineages, P values of < 0.05 confirm heterogeneity in selection pressures while in Site analysis, this is an indication of positive selection which is confirmed through BEB analysis.

Gene	Model Analysis	2Δ1	X^2 value	Degrees of freedom	Probability value
Ghr	Lineage Analysis (M0 vs M1)	2(-11526.925484+11486.241214)=81.568540	81.568540	127-65=62	<0.05
	Site Analysis (M7 vs M8)	2(-11295.233393+11293.456950)=3.552886	3.552886	68-66=2	>0.05

Table S3: Likelihood Ratio Test (LRT) to detect selection in lineages and amino acid sites in insulin-like growth factor 1 (igf1) gene. In the Lineages, P values of > 0.05 are an indication that all lineages are under the same selective pressure i.e no heterogeneity of selective pressures.

Gene	Model Analysis	2Δ1	X^2 value	Degrees of freedom	Probability value
Igf1	Lineage Analysis (M0 vs M1)	2(-2076.081066+2032.048315)=88.065502	88.065502	151-77=74	>0.05
	Site Analysis (M7 vs M8)	2(-2017.692136+2015.465467)=4.853338	4.853338	80-78=2	>0.05

Data Availability

Dataset 1: Homologues of prolactin gene

Prolactin Gene

Specie	E value	Accession Number
Wild chicken 1	3e-148	BAB18728.1
Chicken breed 15 Hubbard	3e-148	AKQ98504.1
Chicken cobb 500	7e-145	AKQ98502.1
Chicken breed Lohmann	4e-143	AKQ98501.1
Chicken breed Yangshan	5e-162	AAT02223.1

Grey jungle fowl	1e-160	BAJ61716.1
Chicken breed avian 48	1e-160	AKQ98503.1
Wild chicken 2	5e-162	AF288765.1
Common quail	7e-150	BAD10927.1
Wild turkey	5e-147	AAB60604.1
White throated sparrow	3e-148	XP 005481833.1
Gold collared manakin	3e-148	KFW 77597.1
Peregrine falcon	7e-145	XP 005235773.1
Arctic fulmar	5e-163	KFV 94966.1
Greylag goose	5e-145	XP 007653890.0
Ostrich	4e-143	BAF81528.1
Wild chicken	5e-162	AAG01026.1
Common peafowl	1e-160	BAG68293.1
Japanese quail	8e-151	BAJ61717.1
Helmeted guinea fowl	2e-151	BAG68294.1
Dalmatian pelican	2e-143	KFQ 6004.1
Wild duck	3e-144	BAD14942.1
Knob billed duck	1e-143	CAJ55836.1
White tailed eagle 1	2e-139	KFQ 01370.1
Common pheasant	2e-143	BAG68292.1
The great cormorant	2e-143	KFW 89232.1
White tailed eagle 2	1e-139	XP 007867890.0
Rock dove	2e-134	ADK73557.1
Jungle crow	2e-139	BAJ61712.1
Ground tit	4e-139	XP 005525306.1
Collared flycatcher	1e-138	XP 005041658.1
Anna's humming bird	2e-114	XP 908765780.0
Zebra finch	5e-121	XP 004186110.1
Streamside salamander	7e-123	AP93863.1
Duck billed platypus	5e-114	XP 007657220.1
African clawed frog	6e-114	NP 001086486.1
Western clawed frog	5e-114	NP 001093699.1

Dataset2: Homologues of vip1 gene

Vip1 Gene

Species	Accession Number	E value
Wild chicken 1	0.0	NP 990697.2
Swan goose	0.0	XP 008776654.0
Collared flycatcher	0.0	XP 005643344.4
Wild turkey	0.0	XP003204177.1

Southern ostrich 1	0.0	KFV78141.1
Northern fulmar	0.0	AA99877.0
Gold collared manakin	0.0	BAB98877
Rock dove	0.0	BAB99879
Saker falcon	0.0	AA98765
Peregrine falcon	0.0	KFW56765.4
Little egret	0.0	KFV87765.5
Emperor penguin	0.0	XP 004543220.0
Dalmatian pelican	0.0	XP 005465365.1
Wild chicken 2	0.0	NP 7866878.0
Southern ostrich2	0.0	NP 8979898.9
Stinkbird	0.0	NP 64763764.8
Medium groundfinch	0.0	KFW7878787
Anna's hummingbird	0.0	KFQ2334434
White throated sparrow2	0.0	XP 008987487.0
Spotted gar	0.0	XP 006574638.6
Ostrich1	0.0	BAB768567
Ostrich2	0.0	AA6758.0
White tailed eagle	0.0	NP 8976565.0
Zebra fish	0.0	BAB87765
Mexican tetra	0.0	BAB56432
Japanese ricefish	0.0	AA69870.0
Bicolor damselfish	0.0	XP 005644322.3
Olive flounder	0.0	XP 008887766.0
Rainbow smelt	0.0	XP 003476654.1
Blue damsel fish	0.0	XP 007876548.0

Dataset 3: Homologues of vipr1 gene

Vipr1 Gene

Specie	Accession Number	E value
Wild chicken	BAA95164.1	0.0
Japanese quail	AED87510.1	0.0
Wild duck	EOA98591.1	0.0
Rock dove	EMC82014.1	0.0
Southern ostrich	BAA76574.1	0.0
Saker falcon1	XP 005442369.1	0.0
Peregrine falcon1	XP 005229590.1	0.0
White tailed eagle1	XP 007856847.0	0.0
Wild turkey	Q91085.2	0.0
Saker falcon2	XP 005442370.1	0.0

Peregrine falcon2	XP 005229591.1	0.0
White tailed eagle2	AAB67768.0	0.0
Blind cavefish	XP 007249106.1	0.0
Zebra fish	AAI162971	0.0
Nile tilapia	XP 003439239.2	0.0
Fairy cichlid	XP 006802769.1	0.0
Burton's mouthbroooder1	XP 005912343.1	0.0
Zebra mbuna1	XP 005463577.0	0.0
Zebra mbuna2	XP 006756847.8	0.0
Damsel fish	XP 007876487.0	0.0
Orange spotted spinefoot	ACC78770.1	0.0
Japanese rice fish1	AA787879.0	0.0
Burton's mouthbroooder2	XP 005933737.1	0.0
Zebra mbuna3	XP 007878780.0	0.0
Japanese rice fish 2	XP 004081326.1	0.0
Japanese puffer1	CAC82587.1	0.0
Japanese puffer2	XP 003977758.1	0.0
Rainbow trout	AAU29499.1	0.0
Burton's mouth brooder3	XP 007457657.6	0.0
Amazon moly	XP 007548620.1	2e-159

Dataset 4: Homologues of gh gene

Gh Gene		
Specie	Accession Number	E value
Chicken haplotype GH-h20	JN675391.1	2e-27
Chicken haplotype GH-h19	JN675390.1	1e-30
Chicken haplotype GH-h18	JN675389.1	1e-28
Chicken haplotype GH-h17	JN675388.1	1e-28
Chicken haplotype GH-h16	JN675387.1	1e-34
Chicken haplotype GH-h15	JN675386.1	2e-30
Chicken haplotype GH-h14	JN675385.1	3e-34
Chicken haplotype GH-h13	JN675384.1	1e-27
Chicken haplotype GH-h12	JN675383.1	1e-20
Chicken haplotype GH-h11	JN675382.1	3e-25
Wild chicken 1	AY461843.1	3e-30
Wild chicken 2	D10484.1	5e-30
Commercial chicken	M35609.1	6e-30
Domestic chicken	X17618.1	6e-40
Wild duck	EOA99704.1	5e-34
Greylag goose	AAN37412.1	9e-34

Common quail	ACJ73931.1	5e-34
Peregrine falcon	XP 005238874.1	2e-30
Rock dove	EMC85315.1	6e-32
White throated sparrow	XP 005487217.1	2e-28
South ostrich	EOA76764.0	4e-34
European rabbit	XP 007636368.7	2e-28
Wild Bactrian camel	XP 006177464.1	4e-23
Zebu	XP 001122366.7	2e-23
Russian sturgeon	ABK74674.6	9e-24
Sheep1	ABK59498.1	2e-23
Sheep2	ABO21737.1	5e-23
Cattle	ABK67647.0	5e-23
Goat	ADX66303.1	9e-24

Dataset 5: Homologues of ghr gene

Ghr Gene

Specie	Accession Number	E value
Chicken breed Yunnan Daweishan	KC242242.1	0
Chicken breed Yunnan Wuding	KC242241.1	0
Commercial chicken	M74057	0
Wild chicken1	AGG38006.1	0.0
Wild chicken2	NP001001293.1	0.0
Dalmatian pelican	XP 0876532.0	0.0
Greylag goose	ACY38605.1	0.0
Saker falcon	XP 005433804.1	0.0
Peregrine falcon	XP 005242027.1	0.0
Wild duck	ACT 20710.1	0.0
Knob billed duck	ACT 20711.1	0.0
Rock dove	EMC76968.1	0.0
Southern ostrich	EMC9876.0	0.0
Anna's humming bird	ACY2165.0	0.0
Golden collared manakin	ACY3476.0	0.0
Red throated loon	ACY5876.1	0.0
White throated sparrow	XP 005493766.1	0.0
Zebra finch	XP 002193695.2	0.0
Medium groundfinch	XP 005422066.1	0.0
Wild canary	XP 3454267876.0	0.0
The great cormorant	NP6473676787.0	0.0

Collared flycatcher	AA018173.1	0.0
Northern carmine bee-eater	XP 6376387878.0	0.0
Atlantic canary	XP 0087765409.0	0.0
European rabbit	1401239A	0.0
Wild Bactrian camel	AA987430.0	0.0
Alpaca	AA897322.0	0.0
Carolina anole1	XP 008101043.1	0.0
Carolina anole2	XP 0008101044.1	0.0
Goat	XP 0077863233.0	0.0
Zebu	ABM92307.2	0.0
Cattle	AAU94310.1	0.0
Sheep	NP001009323.1	0.0

Dataset 6: Homologues of igf1 gene

Igf1 Gene

Specie	Accession Number	E value
Chicken haplotype h-12	JN609551.1	1e-20
Chicken haplotype h-11	JN609550.1	1e-20
Chicken haplotype h-10	JN609549.1	2e-10
Chicken haplotype h-9	JN609548.1	1e-40
Chicken breed Yunnan Daweishan	KC242240.1	3e-10
Chicken breed Yunnan Wuding	KC2422239.1	3e-10
Chicken haplotype h-8	JN593018.1	3e-25
Chicken haplotype h-7	JN593017.1	3e-20
Chicken haplotype h-6	JN593016.1	1e-40
Chicken haplotype h-5	JN593015.1	1e-27
Chicken haplotype h-4	JN593014.1	1e-35
Chicken haplotype h-3	JN593013.1	1e-35
Chicken haplotype h-2	JN593012.1	1e-35
Chicken haplotype h-1	JN593011.1	1e-64
Commercial chicken	M3279.1	4e-10
Wild chicken	AGG38005.1	2e-94
Japanese quail	AAF67202.1	2e-94
Great cormorant	XP 008766769.2	1e-64
Wild turkey	XP 003202426.1	7e-95
Wild duck	ABS76279.1	2e-94
Zebra finch	XP 006754322.0	2e-62
Collared flycatcher	XP 005040114.1	7e-94
Golden collared manakin	XP 006921111.4	1e-63

Medium ground finch	XP 005421104.1	1e-93
Dalmatian pelican	AAF98765.0	2e-94
Southern ostrich	AAF34222.0	3e-75
Wild Bactrian camel1	XP 006186100.1	1e-64
European rabbit1	XP 008254938.1	2e-62
Cattle1	AAF56222.1	2e-62
Goat	BAB77524.1	3e-75
Sheep	ACG49835.1	1e-72
Wild Bactrian camel2	XP 006186101.1	5e-52
Wild Bactrian camel3	XP 006186102.1	7e-66
Wild Bactrian camel4	XP 006186103.1	1e-64
Wild Bactrian camel5	XP 006186104.1	6e-79
Cattle2	AAF22156.2	1e-63
European rabbit2	XP 008254939.1	7e-50
Cattle3	AAF42111.0	7e-50
Cattle4	AAF11114.2	3e-75
European rabbit3	XP 008254940.1	2e-62
European rabbit4	XP 008254941.1	5e-77
Cattle5	AAF75333.2	3e-75
Cattle6	AAF73432.0	3e-75
Cattle7	AAF23407.1	2e-62

Dataset 7: Homologues for igf1r gene

Igf1r Gene

Specie	Accession Number	E value
Wild chicken	AGG38009.1	0.0
Japanese quail	BAF73401.1	0.0
Saker falcon1	XP 005436689.1	0.0
Peregrine falcon1	XP 005242493.1	0.0
Saker falcon2	XP 005436690.1	0.0
Peregrine falcon2	XP 005242494.1	0.0
Medium ground finch1	XP 005424278.1	0.0
Zebra finch	XP 002199843.1	0.0
Medium ground finch2	XP 005424279.1	0.0
Golden collared manakin1	XP 004687532.2	0.0
Golden collared manakin2	XP 000997654.8	0.0
Golden collared manakin3	EMC 77848.3	0.0
Wild duck	EOB07472.1	0.0
Little egret	EOB23699.0	0.0

Rock dove	EMC 77329.1	0.0
Southern ostrich	EMC 76589.5	0.0
Wild turkey	XP 0032009598.1	0.0
Atlantic canary	XP 0034509876.0	0.0
White throated sparrow	XP 0012567876.5	0.0
Anna's humming bird	XP 0035779654.0	0.0
Budgerigar	XP 0045885434.3	0.0
Cuckoo roller	XP 0056328797.1	0.0
Bar tailed trogon	XP 0011187072.0	0.0
Cattle1	XP 0078656766.5	0.0
European rabbit1	XP 0066666988.2	0.0
European rabbit2	XP 0011765445.3	0.0
European rabbit3	XP 0044498885.2	0.0
Sheep	XP 0040085983.1	0.0
Cattle2	XP 0067333333.3	0.0
Cattle3	XP 0078899906.2	0.0

References

Cheng, H.-W. Breeding of tomorrow's chickens to improve well-being. *Poult. Sci.* **89,** 805–813 (2010).

Chaiseha, Y., Youngren, O. M. & Halawani, M. E. E. Expression of Vasoactive Intestinal Peptide Receptor Messenger RNA in the Hypothalamus and Pituitary Throughout the Turkey Reproductive Cycle. *Biol. Reprod.* **70,** 593–599 (2004).

Dana, N., van der Waaij, L. H., Dessie, T. & van Arendonk, J. A. M. Production objectives and trait preferences of village poultry producers of Ethiopia: implications for designing breeding schemes utilizing indigenous chicken genetic resources. *Trop. Anim. Health Prod.* **42,** 1519–1529 (2010).

DeLano, W. L. The PyMOL Molecular Graphics System (2002) DeLano Scientific, Palo Alto, CA, USA. http://www.pymol.org. *ResearchGate* (2002).

Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32,** 1792–1797 (2004).

Källberg, M. *et al.* Template-based protein structure modeling using the RaptorX web server. *Nat. Protoc.* **7,** 1511–1522 (2012).

Källberg, M. *et al.* Template-based protein structure modeling using the RaptorX web server. *Nat. Protoc.* **7,** 1511–1522 (2012).

Kawauchi, H. *et al.* Identification of growth hormone in the sea lamprey, an extant representative of a group of the most ancient vertebrates. *Endocrinology* **143,** 4916–4921

Kingori, A. M., Wachira, A. M. & Tuitoek, J. K. Indigenous Chicken Production in Kenya: A Review. *Int. J. Poult. Sci.* **9,** 309–316 (2010).

Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30,** 772–780 (2013).

Lei, M. *et al.* Polymorphism of Growth-Correlated Genes Associated with Fatness and Muscle Fiber Traits in Chickens. *Poult. Sci.* **86,** 835–842 (2007).

Lwelamira, J., Kifaro, G. C. & Gwakisa, P. S. Genetic parameters for body weights, egg traits and antibody response against Newcastle Disease Virus (NDV) vaccine among two Tanzania chicken ecotypes. *Trop. Anim. Health Prod.* **41,** 51–59 (2009).

Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinforma. Oxf. Engl.* **23,** 2947–2948 (2007).

Lefort, V., Desper, R. & Gascuel, O. FastME 2.0: a comprehensive, accurate, and fast distance-based phylogeny inference program. *Mol. Biol. Evol.* msv150 (2015). doi:10.1093/molbev/msv150

Nielsen, R. & Yang, Z. Likelihood Models for Detecting Positively Selected Amino Acid Sites and Applications to the HIV-1 Envelope Gene. *Genetics* **148,** 929–936 (1998).

Olwande, P. O. *et al.* Assessing the productivity of indigenous chickens in an extensive management system in southern Nyanza, Kenya. *Trop. Anim. Health Prod.* **42,** 283–288 (2010).

Sserugga, J. *et al.* Investing in the Livestock Sector : Why Good Numbers Matter, A Sourcebook for Decision Makers on How to Improve Livestock Data. (2014).

Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* **30,** 2725–2729 (2013).

Ward, C. W. *et al.* The Structure of the Type 1 Insulin-Like Growth Factor Receptor. (2000).

Wallis, M. Variable evolutionary rates in the molecular evolution of mammalian growth hormones. *J. Mol. Evol.* **38,** 619–627 (1994).
(2002).

Wilkanowska, A., Mazurowski, A., Mroczkowski, S. & Kokoszyński, D. Prolactin (PRL) and prolactin receptor (PRLR) genes and their role in poultry production traits. *Folia Biol. (Praha)* **62,** 1–8 (2014).

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinforma. Oxf. Engl.* **25,** 1189–1191 (2009).

Yang, Z., Nielsen, R., Goldman, N. & Pedersen, A. M. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* **155,** 431–449 (2000).

Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24,** 1586–1591 (2007).

Yang, Z. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol. Biol. Evol.* **15,** 568–573 (1998).

Yang, Z., Wong, W. S. W. & Nielsen, R. Bayes Empirical Bayes Inference of Amino Acid Sites Under Positive Selection. *Mol. Biol. Evol.* **22,** 1107–1118 (2005).

Yang, Z., Wong, W. S. W. & Nielsen, R. Bayes Empirical Bayes Inference of Amino Acid Sites Under Positive Selection. *Mol. Biol. Evol.* **22,** 1107–1118 (2005).

Zhao, R., Muehlbauer, E., Decuypere, E. & Grossmann, R. Effect of genotype-nutrition interaction on growth and somatotropic gene expression in the chicken. *Gen. Comp. Endocrinol.* **136,** 2–11 (2004).

Zhu, Y. F. *et al.* Detecting Adaptive Evolution of Galliform and Anseriform Avians Mx Genes. *J. Anim. Vet. Adv.* **9,** 1811–1815 (2010).