Journal of Animal and Veterinary Advances 9 (13): 1811-1815, 2010
ISSN: 1680-5593

Detecting Adaptive Evolution of Galliform and Anseriform Avians Mx Genes

Y.F. Zhu, H.F. Li, W. Han, J.T. Shu, W.T. Song, X.Y. Zhang and K.W. Chen
Institute of Poultry Science, Chinese Academy of Agricultural Science,
Yangzhou, 225003, Jiangsu, R.P. China

Abstract: The objective of this study was to trace the route of avian Mx (myxo-virus resistance) gene evolution and knowledge of its structure and function variations. Mx gene cDNA sequences of six galliform and anseriform avians were used in this study. Datamonkey and DAMBE were used to detect recombination events and test nucleotide substitution saturation. PAML4b were used to test the selective pressure on amino acid sites. Saturation test did not indicate any sign of substitution saturation. Single breakpoint scanning and genetic algorithm scanning by Datamonkey found two breakpoints, located, respectively in 162 and 999 bp and divided the sequences into three nonrecombination split partitions. The selection test of site-specific mode showed that avian Mx gene sequences had suffered positive selection pressure. Likelihood Ratio Test (LTR) suggested M2a and M8 be more advantageous models,and in a total of eight positive sites with >95% posterior probability were identified in three nonrecombination split partitions. These detected positive sites distributed in N-ternimal and GTP-binding domain of Mx protein and might be important candidate marks for improving avian antiviral activity.

Key words: Mx gene, adaptive evolution, recombination scanning, Codon substitution models, adaptive evolution, China

INTRODUCTION

Comparison of Synonymous (dS) and Nonsynonymous (dN) substitution rates (ω = dN dS^{-1}) in protein encoding genes provided an important means for understanding molecular evolution. The value of ω ratio <1, = 1 and >1, respectively indicated purifying selection, neutral selection and positive selection (Yang *et al*., 2000a). It had been mainly used to identify adaptive molecular evolution of genes correlated with reproduction and antivirus (Yang *et al*., 2000b; Sainudiin *et al*., 2005). Mx proteins, part of the dynamin family of large GTPases induced by IFN-I (Accola *et al*., 2002) expressed antiviral activity in many species, e.g., human (Aebi *et al*., 1989), murine (Jin *et al*., 1999), sheep (Charleston and Stewart, 1993), bovine (Ellinwood *et al*., 1998) as well as fish (Ooi *et al*. 2006) viral activity had some difference between species. MxA protein of humans had antiviral activity to Vesicular Stomatitis (VSV) and avian influenza virus but MxB protein had no antiviral activity. Similarly, Mx1 protein of rats conferred antiviral activity to avian influenza virus *in vivo* or *in vitro*, Mx2 and Mx3 conferred no activity. In poultry, Ko *et al*. (2002, 2004) found that the amino acid at position 631 of chicken Mx could determine the antiviral activity to Vesicular Stomatitis (VSV) and avian influenza virus and that cell lines expressing chicken Mx proteins with asparagine at position 631 had higher antiviral activity than those with serine. It was not clear if the antiviral activity of avian Mx has evolved to inhibit a set of species-specific pathogens.

In the present study, Maximum-likelihood models of codon substitution were used to determine the evolutionary pattern of six galliform and anseriform avians Mx genes and identify amino acid sites under diversifying selection. This could help to trace the route of chicken Mx evolution and knowledge of its structure and functions. The identified positive sites could contribute to chicken resistant breedings.

MATERIALS AND METHODS

Sequence data: The full Mx cDNA sequences of six species of anseriform and galliform birds were collected from GenBank. Of which, *G. gallus* Mx gene sequence sets were composed of RIR (Accession No: EF575635), TB (Accession No: EF575608), GSP line (Accession No: AB088536), BM line (Accession No: AB088535), WLF (Accession No: AB088534), Comtaier (Accession No: AY695797), SK (Accession No: EF575611), BY (Accession No: EF575614), WL (Accession

Corresponding Author: H.F. Li, Institute of Poultry Science, Chinese Academy of Agricultural Science, Yangzhou, 225003, Jiangsu, R.P. China

No: EF575622) and WLN (Accession No: AB088533). The other five species were *C. coturnix* (Accession No: EF575605), *M. gallopavo* (Accession No: EF575607), *P. colchicus*(Accession No: EF575606), *L. lagopus* (Accession No: EF575604)and *A. platyrhynchos* (Accession No: Z215549).

Statistical analysis: The cDNA sequences were aligned using CLUSTAL W program implemented in MEGA 3.1 (Kumar *et al*., 2004). The whole level of saturation was determined using the index of substitution saturation, implemented in the DAMBE (http://dambe.bio.uottawa.ca/dambe.asp). Recombination analysis and identification of non-recombinant regions in multiple sequence alignments was done using single breakpoint and Genetic Algorithms (GARD) implemented in Datamonkey (http://www.datamonkey.org/GARD/). Several site-specific models that allow for various dN/dS ratios among sites implemented in Codeml program from PAML4b (http://abacus.gene.ucl.ac.uk/software/paml.html) were used to detect positive selection. The fitness of models was evaluated through Likelihood Ratio Test (LRT). When the likelihood ratio test was significant, the empirical Bayes method (Yang *et al*., 2005) was used to calculate posterior probability for sites classes.

RESULTS AND DISCUSSION

Substitution saturation test: Before carrying out tests of positive selection, the substitution saturation of Mx gene data sets was examined, as saturation could lead to the underestimation of dS and an inflation of the dN/dS ratio. The results showed that the Index Score (ISS) was significantly lower than the critical score (ISS.C) for the Mx gene data sets (ISS = 0.4659, ISS.C = 0.7850; $p < 0.001$). Therefore, Mx gene data sets of six anseriform and galliform birds did not indicated any sign of substitution saturation.

Recombination analysis and Positive site identification: Datamonkey implemented a pragmatic approach, Genetic Algorithm Recombination Detection (GARD for short) to find all the recombination breakpoints. Given the maximum number of break points (B this number can also be inferred), the method will search the space of all possible locations for B or fewer break points in the alignment, inferring phylogenies for each putative nonrecombinant fragment and assessed goodness of fit by small sample Akaike Information Criterion (AIC). For Mx gene data sets of six anseriform and galliform birds, two breakpoints, located, respectively in 162 bp and 999 bp were found (Table 1) and the sequences were divided into three nonrecombination split partitions. Figure 1 showed the breakpoint placement support using c-AIC. The topology of inferred trees based on different split partitions was treated as user-tree in Codeml program to test positive sites.

Table 1: Screen of recombination breakpoints using GARD method

No. of BPs	c-AIC	Δc-AIC	Site of breakpoints (bp)
0	14987.7	-	-
1	14735.3	252.421	162
2	14686.8	48.4837	162, 999
3	14686.8	0	162, 999

The locus of breakpoint was accorded to sequences of chicken in alignments (EF575614)

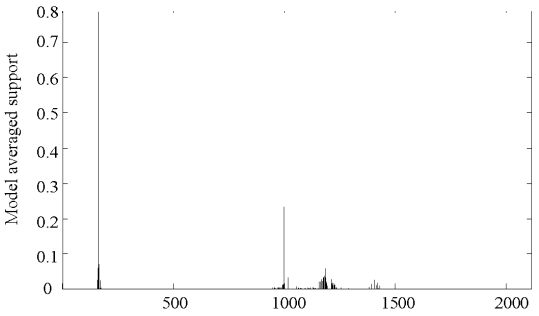

Fig. 1: Breakpoint placement support of Mx gene sequences using c-AIC

Positive site identification: A number of models were implemented in codeml program. Model 0 (one ratio) assumeed one ω for all sites. The neutral Model (M1) assumeed a proportion p_0 of conserved sites with $\omega_0 = 0$ and a proportion p1 of neutral sites with $\omega_1 = 1$. The selection Model (M2) added a third class of sites with ω_2 estimated from data. The discrete Model (M3) used a discrete distribution with three site classes, estimating the proportions (p_0, p_1 and p_2) and the ω ratios (ω_0, ω_1 and ω_2) from the data. Model M7 (beta) used a beta distribution B (p, q) which depending on parameters p and q can take various shapes in the interval (0, 1). Model M8 (β and ω) added an extra class of sites with ω>1 to the beta (M7) model. Comparison of two different models constituted an Likelihood Ratio Rest (LRT) which was done by comparing the log-likelihood values. If the null hypothesis was correct, twice the log-likelihood difference between the two models (2ΔL) asymptotically had a χ^2 distribution. Parameter estimation of the above models for three nonrecombination fragment sequences of Mx gene and results of three LRTs were showed in Table 2-4. About 4, 1 and 3 positive sites were respectively identified in three nonrecombination fragment sequences. Early adaptive evolution studies calculated Synonymous (dS) and Nonsynonymous (dN) substitution rates by averaging over all sites in the protein. However,

Table 2: Parameters estimation of first nonrecombination sequences and test of positive sites

Model	LnL	Estimation values	2ΔlnL	Positive sites
M0	-956.324306	$\omega_0 = 1.82977$	1.1789	Not allowed
M3	-955.734847	$p_0 = 0.04975$, $p_1 = 0.02438$, $p_2 = 0.92588$ $\omega_0 = 0.33123$, $\omega_1 = 1.97938$, $\omega_2 = 1.97938$	p>0.05	
M1a	-960.070076	$p_0 = 0.00742$, $p_1 = 0.99258$	8.6706	5 R, 21 Q, 24 M2a
	-955.734847	$p_0 = 0.04975$, $p_1 = 0.0000$, $p_2 = 0.95025$ $\omega_0 = 0.33123$, $\omega_1 = 1.0000$, $\omega_2 = 1.97938$		N, 37 G,
M7	-960.174890	$p = 99.0000$, $q = 0.0050$	8.8779	2 N, 5 R, 9
M8	-955.735925	$p_0 = 0.05003$, $p_1 = 0.94997$ $p = 50.05739$, $q = 99.0000$ $\omega = 1.97985$	p<0.05	S, 10 S, 13 G, 15 p, 21 Q, 24 N, 32 p, 37 G, 50 F, 51 C

Table 3:Parameters estimation of second nonrecombination sequences and test of positive sites

Model	LnL	Estimation values	2Δ lnL	Positivesites
M0	-2631.623303	$\omega_0 = 0.67471$	39.0118	Not
M3	-2612.117429	$p_0 = 0.25845$, $p_1 = 0.39244$,$P_2 = 0.34911$ $\omega_0 = 0.23634$, $\omega_1 = 0.23634$, ω_2=1.77710	p<0.01	allowed
M1a	-2615.132400	$p_0 = 0.40005$, $p_1 = 0.59995$	6.0299	298 E
M2a	-2612.117429	$p_0 = 0.65089$, $p_1 = 0.0000$,$P_2 = 0.34911$ $\omega_0 = 0.23634$, $\omega_1 = 1.0000$, $\omega_2 = 1.77710$	p<0.05	
M7	-2615.320076	$p = 0.02812$, $q = 0.01605$	6.3861	298 E
M8	-2612.127027	$p_0 = 0.65242$, $p_1 = 0.34758$ $p = 31.05056$, $q = 99.0000$ $\omega = 1.78008$	p<0.05	

Table 4: Parameters estimation of third nonrecombination sequences and test of positive sites

Model	LnL	Estimation values	2lnL	Positive sites
M0	-3595.692088	$\omega_0 = 1.04593$	81.7453	Not allowed
M3	-3554.819433	$p_0 = 0.69388$, $p_1 = 0.20083$,$P_2 = 0.10529$ $\omega_0 = 0.45395$, $\omega_1 = 2.61013$, $\omega_2 = 4.79240$	p<0.01	
M1a	-3577.556981	$p_0 = 0.29765$, $p_1 = 0.70235$	45.3592	413 S,
M2a	-3554.877384	$p_0 = 0.58473$, $p_1 = 0.17573$,$P_2 = 0.23954$ $\omega_0 = 0.41398$, $\omega_1 = 1.0000$, $\omega_2 = 3.84470$	p<0.01	417I,
M7	-3577.650474	$p = 0.03681$, $q = 0.01369$	45.5161	413 S,
M8	-3554.892425	$p_0 = 0.74554$, $p_1 = 0.25446$ $p = 5.03995$, $q = 4.74389$ $\omega = 3.74868$	p<0.01	4171, 425 A, 474 C, 477 L, 653 Y, 681 Q

General font sites with posterior probability >95% level, bold font sites with posterior probability >99% level

due to stronger structural and functional constraints in a functional protein, adaptive selection pressure often only acted on a few sites at a few time points (Yang, 1998). Thus, such analysis rarely find ω ratio >1 sites or detect positive selection.

In the past twenty years, many stochastic models for studying evolution due to point mutations were reported (Muse and Gaut, 1994; Felsenstein and Churchill, 1996; Savill *et al*., 2001). These advances led to rapid development of popular phylogeny-based inference methods for detecting adaptive evolution at the level of sequence codons (Nielsen and Yang, 1998; Suzuki and Gojobori, 1999; Kosakovsky and Frost, 2005). Of which, Maximum Likelihood (ML) methods developed by Nielsen and Yang (1998) and Yang *et al*. (2000a), based on explicit models of codon substitution assuming variable ω ratios among sites had been found to be powerful in detecting adaptive evolution and identifing positive sites (Suzuki and Gojobori, 1999).

Some researchers pointed that the Maximum likelihood methods may suffer high rates of false positives when the analyzed sequences had undergone recombination (Anisimova *et al*., 2003; Shriner *et al*., 2003). So, an essential step before analysis was to screen for and quantify evidence of recombination. Many algorithms and software tools analyzing sequence recombination (Archibald and Roger, 2002) rely on a sliding window approach, however, the length of the sliding window could strongly influence recombination inference. Although, several methods based on Markov Chain Monte Carlo were not limited by sliding window, they were too time expensive and could only be used to

test small or medium data sets. Sergei *et al.* (2006) proposed a pragmatic approach-Genetic Algorithm Recombination Detection (GARD) implemented in datamonkey to rapidly screen multiple-sequence alignments for recombination, quantified the level of support for their locations and identified sequences or clades involved in putative recombination events. Based on reanalysis of previously published biological sequence alignments (Posada, 2002; Chare *et al.*, 2003) and several simulation scenarios, it showed that GARD had good power and accuracy to detect recombination and low rates of false positives.

In this study, GARD approach identified two breakpoints in aligments of Mx gene sequences from six species of anseriform and galliform bird with a total of 300.9047 improvement in c-AIC. The alignments of Mx gene were divided into three nonrecombination split partitions. Then site-special models, allowed the ω ratio to vary among sites (Nielsen and Yang, 1998; Yang *et al.*, 2000b) were used to test diversifying selection for each nonrecombination fragments. The results suggested extreme variation in selective pressure among amino acid sites (M0 VS M3, $p<0.05$) and showed significant positive selection signatures (M1a VS M2a, $p<0.05$; M7 VS M8, $p<0.05$, $p<0.01$) in Mx gene sequences of six anseriform and galliform birds. Hou *et al.* (2007) used sequence comparisons among several mammals, chicken and duck to detect positively selected sites by conducting phylogenetic analysis, also suggested that Mx gene had strong positive selection in chicken and duck lineages. This was consistent with our results.

In this study, the number of positive sites identified by M2a was less than that of M8 and all were contained in the results of M8. In case of false positive, the positive sites identified by M2a were considered. Especially in the first nonrecombination partition sequences, >90% sites suffered positive selection with $\omega = 1.97985$. In the 2nd and 3rd partitions, the rates were also >30 and 20% with higher ω values. The detected positive sites located in the avian specific N-terminal end of the Mx gene and GTPase domain that was essential for activiral activity of Mx protein.

CONCLUSION

In this study, association had analyzed between those detected positive sites with immune traits in Baier chicken (a Chinese indigenous chicken breeds) and found several to be important candidate marks for improving antiviral activity.

ACKNOWLEDGEMENTS

This research was supported by the Program of National Technological Basis from Ministry of Science and Technology of China (No.2008BADB2B08), the high Technological programme of Jiangsu province, China (No. BG2007323) and National High Technology Research and Development Program of China (No. 2008 AA101009).

REFERENCES

Accola, M.A., B. Huang, A. Al-Masri and M.A. McNiven, 2002. The antiviral dynamin family member, MxA, tubulates lipids and localizes to the smooth endoplasmic reticulum. J. Biol. Chem., 277: 21829-21835.

Aebi, M., J. Fah and N. Hurt, 1989. cDNA structures and regulation of two interferon-induced human Mx proteins. J. Mol. Cell Biol., 9: 5062-5072.

Anisimova, M., R. Nielsen and Z.H. Yang, 2003. Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. Genetics, 164: 1229-1236.

Archibald, J.M. and A.J. Roger, 2002. Gene conversion and the evolution of euryarchael chaperonins: A maximum likelihood-based method for detecting conflicting phylogenetic signals. J. Mol. Evol., 55: 232-245.

Chare, E.R., E.A. Gould and E.C. Holmes, 2003. Phylogenetic analysis reveals a low rate of homologous recombination in negative sense RNA viruses. J. General Virol., 84: 2691-2703.

Charleston, B. and H.J. Stewart, 1993. An interferon-induced Mx protein: cDNA sequence and high-level expression in the endometrium of pregnant sheep. Gene, 137: 327-331.

Ellinwood, N.M., J.M. McCue and P.W. Gordy, 1998. Cloning and characterization of cDNAs for a bovine (*Bos taurus*) Mx protein. J. Interferon Cytokine Res., 18: 745-755.

Felsenstein, J. and G.A. Churchill, 1996. A Hidden Markov Model approach to variation among sites in rates of evolution. Mol. Biol. Evol., 13: 93-104.

Hou, Z.C., G.Y. Xu, Z. Su and N. Yang, 2007. Purifying selection and positive selection on the myxovirus resistance gene in mammals and chickens. Gene, 396: 188-195.

Jin, H.K., A. Takada and Y. Kon, 1999. Identification of the murine Mx2 gene: Interferon-induced expression of the Mx2 protein from the feral mouse gene confers resistance to vesicular stomatitis virus. J. Virol., 73: 4925-4930.

Ko, J.H., A. Takada, T. Mitsuhashi, T. Agui and T. Watanabe, 2004. Native antiviral specificity of chicken Mx protein depends on amino acid variation at position 631. Anim. Genet., 35: 119-122.

Ko, J.H., H.K. Jin, A. Asano, A. Takada and A. Ninomiya *et al*., 2002. Polymorphisms and the differential antiviral activity of the chicken Mx gene. Genome Res., 12: 595-601.

Kosakovsky, P.S.L. and S.D. Frost, 2005. Not so different after all: A comparison of methods for detecting amino-acid sites under selection. Mol. Biol. Evol., 22: 1208-1222.

Kumar, S., K. Tamura and M. Nei, 2004. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief. Bioinform., 5: 150-163.

Muse, S.V. and B.S. Gaut, 1994. A likelihood approach for comparingsynonymous and nonsynonymous nucleotide substitutionrates, with application to the chloroplast genome. Mol. Biol. Evol., 11: 715-724.

Nielsen, R. and Z.H. Yang, 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics, 148: 929-936.

Ooi, E.L., I. Hirono and T. Aoki, 2006. Functional characterisation of the Japanese flounder, *Paralichthys olivaceus*, Mx promoter. Fish Shellfish Immunol., 21: 293-304.

Posada, D., 2002. Evaluation of methods for detecting recombination from DNA sequences: Empirical data. Mol. Biol. Evol., 19: 708-717.

Sainudiin, R., W.S. Wong, K. Yogeeswaran, J.B. Nasrallah, Z. Yang and R. Nielsen, 2005. Detecting site-specific physicochemical selective pressures: Applications to the class I HLA of the human major histocompatibility complex and the SRK of the plant sporophytic self-incompatibility system. J. Mol. Evol., 60: 315-326.

Savill, N.J., D.C. Hoyle and P.G. Higgs, 2001. RNA sequence evolution with secondary structure constraints: comparison of substitution rate models using maximum-likelihood methods. Genetics, 157: 399-411.

Sergei, L.K.P., P. David, B.G. Michael, H.W. Christopher and D.W.F. Simon, 2006. Automated phylogenetic detection of recombination using a genetic algorithm. Mol. Biol. Evol., 23: 1891-1901.

Shriner, D., D.C. Nickle, M.A. Jensen and J. Mullins, 2003. Potential impact of recombination on sitewise approaches for detectingpositive natural selection. Genet. Res., 81: 115-121.

Suzuki, Y. and T. Gojobori, 1999. A method for detecting positive selection at single amino acid sites. Mol. Biol. Evol., 16: 1315-1328.

Yang, Z.H., 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol. Biol. Evol., 15: 568-573.

Yang, Z.H., J.S. Willie and D.V. Victor, 2000a. Maximum-likelihood analysis of molecular adaptation in abalone sperm lysin reveals variable selective pressures among lineages and sites. Mol. Biol. Evol., 17: 1446-1455.

Yang, Z.H., N. Rasmus, G. Nick and K.P. Anne-Mette, 2000b. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics, 155: 431-449.

Yang, Z.H., S.W.W. Wendy and N. Rasmus, 2005. Bayes empirical bayes inference of amino acid sites under positive selection. Mol. Biol. Evol., 22: 1107-1118.