bioRxiv preprint doi: https://doi.org/10.1101/034710; this version posted December 17, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A Model of Avian Genome Evolution

LiaoFu LUO

Faculty of Physical Science and Technology, Inner Mongolia University, Hohhot 010021, China

email: lolfcm@mail.imu.edu.cn

Abstract

A model of avian genome evolution is proposed. The classical evolutionary equation is written in terms of nucleotide frequencies of the genome varying in time. The evolution is described by a second-order differential equation. The diversity and the environmental potential play dominant role on the genome evolution. Environmental potential parameters, evolutionary inertial parameter and dissipation parameter are estimated by avian genomic data. To describe the speciation event the quantum evolutionary equation is proposed which is the generalization of the classical equation through correspondence principle. The Schrodinger wave function is the probability amplitude of nucleotide frequencies. The discreteness of quantum state is deduced and the ground-state wave function of avian genome is obtained. New species production is described by the quantum transition between discrete quantum states. The quantum transition rate is calculated from Schrodinger equation which provides a clue to understand the law of the rapid post-Cretaceous radiation of neoavian birds. The present article is the continuation of the author's work "a model on genome evolution" published at arXiv:1411.2205 [q-bio.OT] (2014).

Key words genome evolution, avian genome, evolutionary equation, Schrodinger wave function, quantum state, quantum transition

Introduction

We have proposed a quantitative model of genome evolution.[1] The extinction of the dinosaurs and pterosaurs at the end of Cretaceous (66 millions of years ago) left open a wide range of ecological niches, allowing birds and mammals to diversity. Most of the vacated avian niches were thought to have been filled by a rapid post-Cretaceous radiation of neognathous birds which today include over 99% of all bird species.[2]-[5]. The avian species afford a splendid field for investigation of the genome evolution. Recently, the bird macroevolution was explored by using full genomes from 48 avian species representing major extant clades.[6][7]. Then a more comprehensive phylogeny of 198 avian species was established by using targeted next-generation DNA sequencing.[8] Avian genomes show a remarkably high degree of evolutionary stasis at the level of chromosomal structure, gene syntency and nucleotide sequence. The macrochromosome studies reveals that the ancestral pattern has remained largely unaltered in the majority of avian genomes.[9] At the level of nucleotide sequence it was indicated that the overall pan-genome background substitution rate in birds is about 1.9×10^{3}, lower than in mammals, 2.7×10^{3} substitutions per site per million years.[6] The most interesting examples of evolutionary stasis are Struthio camelus (Common ostrich) and Ophisthocomus hoatzin. The present Struthio camelus

bioRxiv preprint doi: https://doi.org/10.1101/034710; this version posted December 17, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

belongs to palaeognathae，branching out at 50 Mya, while the present Ophisthocomus hoatzin is also a palae-species branching out at 64 Mya.[7][8] Most neoavian birds occurred in a relatively short period of 10 – 15 million years after the mass extinction event about 66 million years ago.

From the full-genome sequence data of 48 avian species it is found the size N (sequence length) of 48 bird genomes takes a value between (1044 – 1258)$\times 10^6$. The informational redundancy

$$R_I = 2 + \sum_i^4 x_i \log x_i \tag{1}$$

(x_i - the frequency of i-th nucleotide in a genome) for 48 genomes takes a value between 0.008 – 0.026 that shows x_i very near to $N/4$.

In the article the classical evolutionary equation of the genome will be written in terms of nucleotide frequencies varying with time. The evolution is described by a second-order differential equation. The diversity and the environmental potential play dominant role on the genome evolution. The measure of diversity was taken from [10] and a simple form of environmental potential will be assumed. Environmental potential parameters, evolutionary inertial parameter and dissipation parameter will be estimated by fitting avian genomic data. To describe the speciation event the quantum evolutionary equation is proposed which is the generalization of the classical equation through Bohr's correspondence principle in atomic physics. Then the discreteness of quantum state can be deduced and the ground-state wave function of avian genome can be obtained. New species production will be described and calculated by the quantum transition between discrete quantum states. The results are expected to provide a clue to understand the law of the rapid post-Cretaceous radiation of neognathous birds.

Classical evolutionary equation of avian genome

For any genome there exists a potential to characterize the evolutionary direction [1]

$$V(x_1,...,x_m,t) = D(x_1,...,x_m) + W_{env}(x_1,...,x_m) \tag{2}$$

where x_i means the frequency of the i-th nucleotide (or nucleotide pair) in DNA, W_{env} is a selective potential dependent of environment, and V depends on t through the change of environmental variables. D means the diversity-promoting potential [10][11]

$$D(x_1,...,x_m) = N\log N - \sum_i^m x_i \log x_i, \qquad N = \sum_i^m x_i \tag{3}$$

In the following we shall consider single nucleotide frequency only and take m=4, simultaneously set the logarithm log=$\log_2$.

The classical genome evolution equation reads as [1]

$$\frac{d}{dt}(c^2\frac{dx_i}{dt}) = \frac{\partial V}{\partial x_i} - f\frac{dx_i}{dt}$$

$$\frac{d}{dt}(c^2\frac{dx_i}{dt}) = -\log\frac{x_i}{N} + \frac{\partial W_{env}}{\partial x_i} - f\frac{dx_i}{dt}$$

(4)

bioRxiv preprint doi: https://doi.org/10.1101/034710; this version posted December 17, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

where $f > 0$ is a dissipation coefficient representing the effect of fluctuation force. The parameter c^2 is introduced with the dimension of （time）^{2} which represents the evolutionary inertia of the genome. It means the changing rate of nucleotide frequency $\frac{dx_i}{dt}$ tends to be a constant and as external force is applied the acceleration of frequency change is measured by a dimensionless time t/c.

The environmental potential W_{env} is generally a function of N. For the case of small kinetic energy $(\frac{dx_i}{dt})^2$ the classical trajectory of a mechanical system is always near the bottom of the potential energy. Through the mechanical simulation we can assume the high-order term of N in W_{env} neglected. Thus we have

$$W_{env} = aN - bN^2 \qquad (b > 0 \qquad N = \sum_i x_i)$$

$$\frac{\partial W_{env}}{\partial x_i} = a - 2bN \tag{5}$$

The environmental selection is called positive as the selective potential W_{env} increases with N and called negative as W_{env} decreases with N. As $b > 0$, $\frac{a}{b} < 2N$ the environmental selection is always negative.

Inserting (5) into (4) it leads to the equilibrium occurs at

$$x_i = \frac{N}{4}$$

$$(\frac{dx_i}{dt})_{eq} = \frac{1}{f}(\log\frac{N}{x_i} + a - 2bN) \tag{6}$$

When $N < N_0 = \frac{2+a}{2b}$ the genome size increases with a decreasing rate given by (6) and, as the rate is slow enough, we call it in quasi-stasis phase. When N attains N_0 the genome ceases to increase and the phase is called evolutionary stasis. The genome lengths of Common ostrich and Hoatzin are 1228×10^6 and 1209×10^6 respectively. One may assume both genomes, Common ostrich and Hoatzin, are in the evolutionary stasis and their sizes have attained or approached the threshold value N_0. If the parameter α and β are nearly same for different aves then one estimates $N_0 \approx 1.26\times10^9$. If $|a| << 2$ is assumed then $b = \frac{2+a}{2N_0} \cong \frac{1}{N_0} \approx 0.8\times10^{-9}$ is estimated. The parameter f can be estimated from Eq (6). Considering many birds are in quasi-stasis phase from 50Ma till now we estimate

$$(\frac{dx_i}{dt})_{eq} = \frac{1}{4}(\frac{dN}{dt})_{eq} < \frac{1}{4}\times\frac{(1258-1044)\times10^6}{50\times10^6} \approx 1.1/yr \quad \text{and} \quad f > \frac{2(1-\frac{\langle N\rangle}{N_0})}{1.1} \approx 0.16yr\,.$$

Inserting (5) into (4) and through integration of Eq (4) one obtains

bioRxiv preprint doi: https://doi.org/10.1101/034710; this version posted December 17, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$$\frac{d}{dt}\{\frac{1}{2}c^2(\frac{dx_i}{dt})^2 - D - aN + bN^2\} < 0 \tag{7}$$

due to the damping of friction force. The dissipation effect of friction force is accumulated in time. In the initial evolution of a genome one may neglect f term in Eq (4). By assuming c a constant Eq (4) reads

$$c^2\frac{d^2x_i}{dt^2} = -\log\frac{x_i}{N} + a - 2bN \cong \log\frac{N}{x_i} + a - 2\frac{N}{N_0} \tag{8}$$

The initial evolution of a genome is x_i increasing rapidly with an acceleration

$$a_c = \log\frac{N(t_0)}{x_i(t_0)} + a - 2\frac{N(t_0)}{N_0}$$

$$x_i = x_{i0} + (\frac{dx_i}{dt})_0(t-t_0) + \frac{a_c(t-t_0)^2}{2c^2} \tag{9}$$

Taking $t - t_0 < 10^7 yr$, $x_i - x_{i0} = \frac{1}{4}(1258-1044)\times 10^6 = 53.5\times 10^6$, $0 < (\frac{dx_i}{dt})_0 < 1.1 yr^{-1}$, and $a_c = 2(1-\frac{\langle N(t_0)\rangle}{N_0}) = 0.17$ we estimate the inertial parameter $c < 450 yr$ for avian genome in the classical phase of evolution. The present c value is larger than the previous estimate in [1] where the form of environmental potential was not given.

In the above discussion on avian genome the parameters α and β in environmental potential W_{env} are assumed to be constant since the environmental condition is stable on the evolutionary trajectory. In general, in the changing environment, α and β varying with t , the genome will select new evolutionary trajectory (new solution of Eq (4)) to adapt the environment. However, if the environmental sudden change, for example the food deficiency, makes $\beta(t)$ increasing too rapidly, the genome will be shrunken and degenerate and the species will be close to extinction.

Quantum state , quantum transition and avian speciation

Following Bohr's correspondence principle in atomic physics there exists a good correspondence between classic-mechanical motion and quantum motion. The quantum-generalization of the deterministic equation (4) （as f=0）is Schrodinger equation satisfied by wave function $y(\mathrm{x},t)$ [1]

$$iL\frac{\partial}{\partial t}y(\mathrm{x},t) = Hy(\mathrm{x},t) \qquad (\mathrm{x} = (x_1, x_2, x_3, x_4)$$

$$H = \frac{1}{2c^2(t)}\sum_i p_i^2 - V(\mathrm{x},t)$$

$$p_i = -iL\frac{\partial}{\partial x_i} \tag{10}$$

$V(\mathrm{x},t)$ is given by (2) (3) and (5) and α and β are dependent of t in general. H is called Hamiltonian of the genome. L is quantization constant , corresponding to Planck constant in

bioRxiv preprint doi: https://doi.org/10.1101/034710; this version posted December 17, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

atomic quantum mechanics. Different from Planck constant, L is in the dimension of time and describes the span of speciation event. It was roughly estimated $L = 3\times10^3\tau$ where τ is the average lifetime for one generation of the species[1] (τ ~30yr for human and 1 yr for bird). The quantum evolutionary equation (10) is the logic generalization of the classical equation through path integral approach. The generalization is valid not only for the evolution in stable environment but also for the evolution in varying environment where the evolutionary potential V and inertia c^2 are time-dependent. The quantum evolutionary theory is applicable in studying the new species formation where the classical theory is incapable of action. In the following we shall discuss the speciation event from the view of quantum transition.

The quantum state described by the eigenstate of Hamiltonian is generally not continuously but discrete. The eigenstate of H satisfies

$$Hy(\mathrm{x}) = Ey(\mathrm{x}) \tag{11}$$

The Hamiltonian given by Eq (10) can be rewritten as

$$\begin{aligned} H &= H_0 + H' \\ H_0 &= \frac{-L^2}{2c^2}\sum_i \frac{\partial^2}{\partial x_i^2} - (2+a)N + bN^2 \\ &= \frac{-2L^2}{c^2}\frac{\partial^2}{\partial N^2} - (2+a)N + bN^2 \\ H' &= N(2-R_I) = \frac{2}{N\ln 2}\sum_i x_i^2 - \frac{N}{2\ln 2} \end{aligned} \tag{12}$$

where Eqs (1)-(3) and (5) has been used. H' is a small quantity which can be looked as a perturbation. Following perturbation theory, $y(\mathrm{x}) = y^{(0)}(N) + y^{(1)}(\mathrm{x})$, the zero order wave function $y^{(0)}$ depends on N only, satisfying

$$\{\frac{-2L^2}{c^2(t_0)}\frac{\partial^2}{\partial N^2} - (2+a(t_0))N + b(t_0)N^2\}y^{(0)}(N) = E^{(0)}y^{(0)}(N) \tag{13}$$

The solution of Eq(13) is

$$E^{(0)} = E_m^{(0)} = (m+\frac{1}{2})\sqrt{8b}\,\frac{L}{c} - \frac{(2+a)^2}{4b} \qquad (m=0,1,2,\ldots) \tag{14}$$

$$\begin{aligned} & y^{(0)}(N) = y_m^{(0)}(N) = N_m \exp(-\frac{x^2}{2})H_m(x) \\ & H_0(x) = 1 \quad H_1(x) = 2x \qquad H_2(x) = 4x^2 - 2, \ldots \\ & x = \sqrt{\frac{c}{L}}(\frac{b}{2})^{1/4}(N - \frac{2+a}{2b}) \end{aligned} \tag{15}$$

The first-order correction to $E_m^{(0)}$ is

$$E_m^{(1)} = H'_{mm} = \int d^4x y_m^{(0)} H' y_m^{(0)},. \quad d^4x = dx_1dx_2dx_3dx_4 \ .$$

The first-order correction to wave function $y^{(0)}(N)$ can also be calculated from perturbation

bioRxiv preprint doi: https://doi.org/10.1101/034710; this version posted December 17, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

theory.

Eq (14) shows the Hamiltonian-level is equally spaced by $\sqrt{8b}\,\frac{L}{c}$. The spacing $\sqrt{8b}\,\frac{L}{c}$ is a small quantity if c takes a value of several hundreds of years for avian genome as estimated in previous section. So the eigenstates are basically continuous in classical phase of the evolution. However, during speciation the changing environmental pressure makes the evolutionary inertia of new species dropping to a lower value as if in this time all evolutionary events happened more rapidly. If c as a parameter of time dimension decreases to $10^{-2}-10^{-3}$ of the classical value, then the picture of definite trajectory, namely $x_i(t)$ as a function of t, ceases to be correct and the state will be switched to quantum. A series of discrete eigenstates occur in quantum phase. In the quantum state the nucleotide frequency always takes some statistical distribution but not a definite value. Eq (15) shows the statistical distribution of N in ground state (m=0) peaks at $N=\frac{2+a}{2b}$: $\frac{1}{b}$ with width proportional to $\frac{L}{c\sqrt{b}}$.

Therefore, the quantum theory regards that the speciation event is essentially a quantum transition between initial "old" species and final "new" species. There always exists a Hamiltonian-level gap between low-lying ground state (m=0) and excited state (m=1,2,…), that can be seen from the large spacing $\sqrt{8b}\,\frac{L}{c}$ in a quantum genome. In fact, the gap of ground state should be deeper than $\sqrt{8b}\,\frac{L}{c}$ if the higher- order terms of N^3 etc have been taken into account in the environmental potential equation (5) of W_{env}. Due to the deep gap occurring in the Hamiltonian level of ground state one can consider only the transition between ground states in studying speciation event.

Based on Schrodinger equation the speciation rate can be calculated. Suppose the initial wave function of the "old" species denoted by $y_I(\mathrm{x})$ satisfying Eqs (13)-(15) and the final wave function of the "new" species by $y_F(\mathrm{x})$ satisfying the same equation but $c^2(t_0)$, $b(t_0)$, $a(t_0)$ replaced by $c^2(t_F)$, $b(t_F)$, $a(t_F)$ respectively. The transition from "old" to "new" is caused by a time-dependent interaction $H_{\mathrm{int}}(t)$ in the framework of quantum mechanics. One may assume H_{int} comes from the time variation of evolutionary inertia and environmental potential , namely

$$H_{\mathrm{int}}(t)=H(c=c(t),b=b(t),a=a(t))-H(c=c(t_0),b=b(t_0),a=a(t_0))$$

$$=\frac{\partial H}{\partial c}\Delta c+\frac{\partial H}{\partial b}\Delta b+\frac{\partial H}{\partial a}\Delta a\ . \qquad (16)$$

Thus the transitional probability amplitude is expressed by

bioRxiv preprint doi: https://doi.org/10.1101/034710; this version posted December 17, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$$
\begin{aligned}
T_{fi} &= \frac{-i}{L}\int y_F^*(\mathrm{x})\int_{t_0}^{t_F} dt H_{\mathrm{int}}(t) y_I(\mathrm{x}) dx_1 dx_2 dx_3 dx_4 \\
&= \frac{-i}{L}\int y_F^*(\mathrm{x})\int_{t_0}^{t_F} dt\{(\frac{\partial H}{\partial c}\frac{dc}{dt}+\frac{\partial H}{\partial b}\frac{db}{dt}+\frac{\partial H}{\partial a}\frac{da}{dt})(t-t_0)\} y_I(\mathrm{x}) dx_1 dx_2 dx_3 dx_4 \\
&= \frac{-i}{L}\int y_F^*(\mathrm{x})\{H(t_F)(t_F-t_0)-\int_{t_0}^{t_F} H(t)dt\} y_I(\mathrm{x}) dx_1 dx_2 dx_3 dx_4 \\
&= \frac{-i}{L}(E_F-E_I)(t_F-t_0)\int y_F^*(\mathrm{x}) y_I(\mathrm{x}) dx_1 dx_2 dx_3 dx_4
\end{aligned}
\tag{17}
$$

E_I and E_F - the eigenvalue of H, given by Eq(14) with m=0 where parameters c, β, α take values at t_0 and t_F respectively . From Eq（15）, taking m=0 one has

$$
y_I(N) = (p a_I)^{1/4} \exp(-\frac{(N-N_I)^2}{2a_I})
$$

$$
y_F(N) = (p a_F)^{1/4} \exp(-\frac{(N-N_F)^2}{2a_F}) \tag{18}
$$

($a_{I,F}$ - the frequency distribution width, and $N_{I,F}$ the frequency distribution centers for two genomes respectively). Inserting (18) into the transitional probability amplitude one obtains the overlap integral

$$
\begin{aligned}
\int y_F^*(N) y_I(N) dN &= p a_I \exp(-\frac{(N_I-N_F)^2}{4a_I}) \quad (\text{as } a_I = a_F) \\
&\sqrt{2p^3} a_I^{3/4} a_F^{1/4} \quad (\text{as } a_I << a_F) \\
&\sqrt{2p^3} a_F^{3/4} a_I^{1/4} \quad (\text{as } a_I >> a_F)
\end{aligned}
\tag{19}
$$

The transition probability equals $|T_{fi}|^2$, proportional to $\left|\int y_F^*(N) y_I(N) dN\right|^2$ and $(E_F-E_I)^2$. During speciation, corresponding to one ancestry genome there are many candidates for the posterity with different probabilities. Those with larger $(E_F-E_I)^2$ will be more probable. Eq (19) shows that as a_F near a_F the transition probability is large only for small distance $|N_I-N_F|$, since it rapidly tends to zero with increasing $\frac{(N_I-N_F)^2}{2a_I}$. The transition probability is irrespective of the symbol of N_I-N_F if the difference between posterity E_F's can

bioRxiv preprint doi: https://doi.org/10.1101/034710; this version posted December 17, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

be neglected. The next two lines of Eq (19) show that for a given initial state the larger the frequency distribution width a_F the greater the transition probability. It means the most probable posterity genome has stronger uncertainty of frequency in its first formation stage. Evidently, it is a quantum peculiarity of genome evolution. The rapid post-Cretaceous radiation of neoavian birds provides a vast amount of experimental data to test the above quantum theory on new species production.

Acknowlegement The author is indebted to Lirong Zhang, Yongfeng Zhang and Xun Zhou for numerous helpful discussions and Dr Yulai Bao for his kind help in literature searching.

References

1. Luo LF. 2014. A Model on Genome Evolution. arXiv:1411.2205 [q-bio. OT].
2. Feduccia A. 1995. Explosive evolution in tertiary birds and mammals. Science. 267: 637-638.
3. Hedges SB, Parker PH, Sibley CG, Kumer S. 1996. Continental breakup and the ordinal diversification of birds and mammals. Nature. 381:226-229.
4. Cooper A, Penny D. 1997. Mass survival of birds across the Cretaceous-Tertiary boundary: molecular evidence. Science.275:1109-1113.
5. Simon YW H, Matthew J P. 2009. Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. Sym. Biol. 58(3):367-380.
6. Zhang GJ et al. 2014. Comparative genomics reveals insights into avian genome evolution and adaption. Science. 346:1311-1319.
7. Jarvis ED et al. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. Science. 346:1320-1331.
8. Prum RO et al. 2015. A comprehensive phylogeny of birds (aves) using targeted next-generation DNA sequencing. Nature. 526:569-573.
9. Romanov MN et al. 2014. Reconstruction of gross avian genome structure, organization and evolution suggests that the chicken lineage most closely resembles the dinosaur avian ancestor BMC Genomics 15:1060.
10. Laxton RR. 1978. The measure of diversity. J Theor Biol. 70(1):51-67.
11. Zhang LR, Luo LF. 2003. Splice site prediction with quadratic discriminant analysis using diversity measure. Nucleic Acids Res. 31(21):6214-6220.