bioRxiv preprint doi: https://doi.org/10.1101/034710; this version posted April 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity.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

Based on the reconstruction of evolutionary tree for avian genome a model of genome evolution is proposed. The importance of k-mer frequency in determining the character divergence among avian species is demonstrated. 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. As the evolutionary inertia decreasing the classical phase of evolution is transformed into quantum phase. New species production is described by the quantum transition between discrete quantum states. The quantum transition rate is calculated 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 — reconstruction of evolutionary tree of birds

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

bioRxiv preprint doi: https://doi.org/10.1101/034710; this version posted April 30, 2016. 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.

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 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. We shall study how to obtain a more comprehensive picture on the evolutionary history for hundreds and thousands of birds.

The phylogenetic tree provides an appropriate quantitative description on evolution. To reconstruct an evolutionary tree from genome sequence data many methods have been proposed. However, aiming to find the dynamical variable of evolution we shall use k-mer non-alignment algorithm (KNA).[10] Denote the probability of base a (a =A,G,C or T) occurring in a sequence by p_a , and the joint probability of base a and b occurring sequentially in the sequence by p_{ab} . In general, let $\sigma = abc\ldots$ being an oligonucleotide k bases long, we denote the joint probabilities of the bases in σ occurring in the sequence by p_σ . In the calculation of joint probabilities all sequences are assumed to be circular. For any given k we always have $\sum_\sigma p_\sigma = 1$, where the summation over σ is over the set $\{\sigma\}$ of the 4^k oligonucleotides of length k. Given two sequences Σ and Σ' with sets of joint probabilities $\{p_\sigma\}$ and $\{p_\sigma'\}$, respectively, define a distance, called a k-distance, between the two sequences based on the difference of joint probabilities in the two sets as follows

$$D_k(\Sigma,\Sigma') = \sum_\sigma \left| p_\sigma - p_\sigma' \right|$$

$$(k = 1,2,\ldots)$$

There is practical reason for considering k-distances as an appropriate definition of evolutionary distance. The computation time for an k-distances grows linearly with sequence length whereas that for conventional definition based on sequence alignment grows exponentially with sequence length. On the other hand, for long sequences a k-distance is insensitive to minor misalignments between two sequences.

The avian evolutionary tree obtained by Jarvis et al [7] was based on genome sequence data of 48 avian species, while the avian evolutionary tree obtained by Prum et al [8] was based on 198 avian species. In reference [8] only one penguin is listed, which corresponds to two species, namely Adelie penguin and Emperor penguin in reference [7]. So, there are 47 birds which were analyzed commonly in both literature. To compare more easily with [7] and [8] the same 47 avian genomes will be studied in the article. They will be looked as a typical evolutionary system in the following analyses. A data set of 259 genes with the highest quality assemblies was provided in ref [8] which contains about 394000 sites for each avian genome. By using these sequence data and calculating k-distance the distance matrix for 47 species is obtained. Then, by use of Neighbor Joining method the phylogenetic tree is deduced for each k . We find that the quality of tree is gradually improved and the tree topology is stabilized with increasing k and reaches a plateau at k=11 or 12. The tree in Figure 1 [11] is plotted for k equal 12. It remains unchanged with increasing k. The tree is dated using the branching time given in Fig S7 of ref [8]. The Neoaves radiation appears to exhibit successive nested radiations. Four neoavian sister clades, Neoaves 1 to Neoaves 4, are diverged from one palaeognathae/ galloanserae clade. The result is basically consistent with [7] [8] and experimental data.

bioRxiv preprint doi: https://doi.org/10.1101/034710; this version posted April 30, 2016. 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.

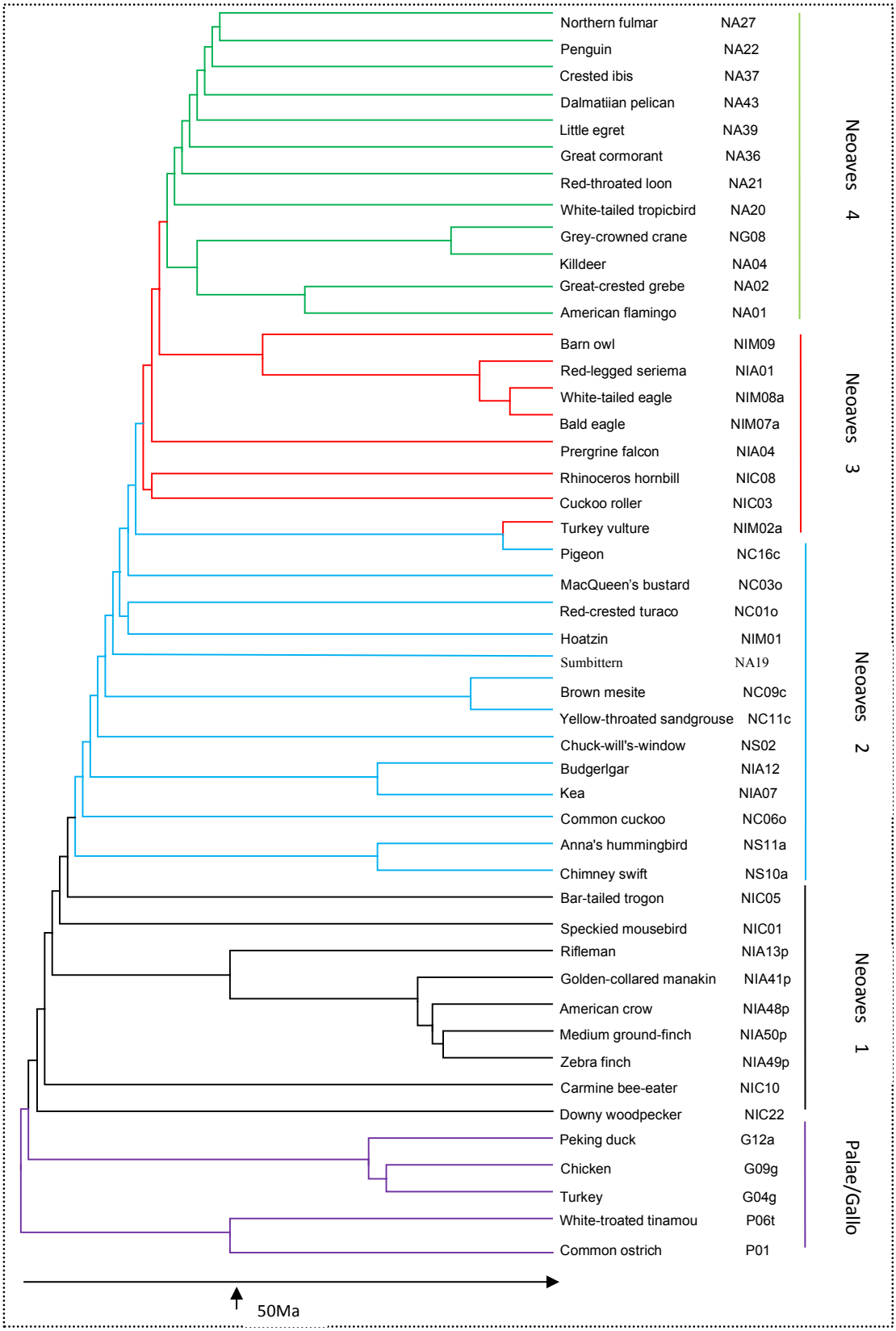

Figure 1 Phylogenetic tree deduced by KNA algorithm

The evolutionary tree of 47 birds was deduced by Zhang et al [11] that can be compared with trees given in [7] [8].

bioRxiv preprint doi: https://doi.org/10.1101/034710; this version posted April 30, 2016. 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.

The notations of 47 birds used in Fig 1 are given by the following rules. Referring to 198 birds in ref [8], several letters are introduced to indicate the taxon of each bird. P01-P09 refer to Palaeognathae, G01-G16 refer to Galloanserae, NS01-NS13 refer to Neoaves- Strisores, NC01-NC16 refer to Neoaves-Columbaves, NG01-NG09 refer to Neoaves-Gruiforms, NA01-NA43 refer to Neoaves-Aequorlitornithes, NIM01-NIM10 refer to Neoaves-Inopinaves-Accipitriformes and related, NIC01-NIC26 refer to Neoaves-Inopinaves-Coraciimorphae, NIA01-NIA56 refer to Neoaves-Inopinaves-Ausralaves. The last lower-case letter indicates a subclass, for exampl;e, the letter p in NIA13p-NIA56p means the subclass passeriformes.

The success of deducting phylogenetic tree by use of KNA algorithm means the *k*-mer (namely 12-mer) frequency is an important factor characterizing the avian evolution. Therefore, the nucleotide frequency can be looked as important dynamical variable in genome evolution.

From the full-genome sequence data of 47 (48) avian species it is found the size N (sequence length) of genomes takes a value between $(1044 - 1258)\times 10^6$. The informational redundancy R_I is the first quantity deduced by nucleotide frequency

$$R_I = 2 + \sum_i^4 \frac{x_i}{N} \log_2 \frac{x_i}{N} \tag{1}$$

(x_i - the frequency of nucleotide i in a genome). For 47(48) avian genomes it 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 [12] [13] 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 is deduced and the ground-state wave function of avian genome is obtained. New species production is 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 k-mer) 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 [12][13]

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

We shall first consider single nucleotide frequency and take m=4.

bioRxiv preprint doi: https://doi.org/10.1101/034710; this version posted April 30, 2016. 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.

The classical genome evolution equation reads as [1]

$$\frac{d}{dt}(c^2\frac{dx_i}{dt}) = \frac{\partial V}{\partial x_i} - f\frac{dx_i}{dt} \tag{4}$$

$$\frac{d}{dt}(c^2\frac{dx_i}{dt}) = \log\frac{N}{x_i} - (1-\frac{x_i}{N})\log e + \frac{\partial W_{en}}{\partial x_i} - f\frac{dx_i}{dt} \tag{5}$$

(hereafter the notation log means $\log_2$) 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.

It is easily to prove that

$$\log\frac{N}{x_i} - (1-\frac{x_i}{N})\log e \geq 0 \tag{6}$$

(equality occurs at $N=x_i$ only). This means the diversity force always acts in positive direction. 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

$$W_{env} = \alpha N - \beta N^2 \qquad (\beta > 0 \qquad N = \sum_i x_i)$$

$$\frac{\partial W_{env}}{\partial x_i} = \alpha - 2\beta N \tag{7}$$

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 $\beta > 0$, $\frac{\alpha}{\beta} < 2N$, the environmental selection is always negative.

Inserting (7) into (5) it leads to the equilibrium occurs at

$$(\frac{dx_i}{dt})_{eq} = \frac{1}{f}(\log\frac{N}{x_i} + \frac{x_i}{N}\log e - \log e + \alpha - 2\beta N) \quad \text{for} \quad x_i \approx \frac{N}{4} \tag{8}$$

When $N < N_0 = \frac{2 - 0.75\log e + \alpha}{2\beta}$ the genome size increases with a decreasing rate given by (8) 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

bioRxiv preprint doi: https://doi.org/10.1101/034710; this version posted April 30, 2016. 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.

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 one assumes α a small number between 0 and 0.2 then the range of β

$$\beta = \frac{2-0.75\log e+\alpha}{2N_0} \approx 0.36\times10^{-9}(\alpha=0), \qquad 0.44\times10^{-9}(\alpha=0.2)$$

is estimated. The parameter f can be estimated from Eq (8). Considering many birds are in quasi-stasis phase from 50Ma till now we estimate

$$(\frac{dx_i}{dt})_{eq} = \frac{1}{4}(\frac{dN}{dt})_{eq} < \frac{1}{4}\times\frac{(1258-1044)\times10^6}{50\times10^6} \approx 1.1/\ yr \quad \text{and}$$

$$f = \frac{2\beta N_0(1-\langle N\rangle/N_0)}{(\frac{dx_i}{dt})_{eq}} > 0.073\,yr(\alpha=0), \quad 0.088\,yr(\alpha=0.2)$$

Inserting (7) into (5) and through integration of Eq (5) one obtains

$$\frac{d}{dt}\{\frac{1}{2}c^2(\frac{dx_i}{dt})^2 - D - \alpha N + \beta N^2\} < 0 \tag{9}$$

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 (5) and deduce that x_i increases rapidly with a dimensionless acceleration

$$Ac_i = \log\frac{N(t_0)}{x_i(t_0)} - \log e(1-\frac{x_i(t_0)}{N(t_0)}) + \alpha - 2\beta N(t_0) \tag{10}$$

At $t=t_0$.

Introducing dimensionless $\tau = t/c$ Eq (5) is rewritten as

$$\frac{d^2x_i}{d\tau^2} = \log\frac{N}{x_i} - (1-\frac{x_i}{N})\log e + \alpha - 2\beta N - \frac{f}{c}\frac{dx_i}{d\tau} \tag{11}$$

For $x_i \approx \frac{N}{4}$ the equation describes a damping oscillation with $\omega = \sqrt{8\beta}$. If x_i takes a minimum at the end of Cretaceous (60 Mya) and reaches maximum at present then c is determined by

$$\sqrt{8\beta}\,\frac{6\times10^7}{c} = \pi \tag{12}$$

It leads to c taking a value between 1031 yr and 1138 yr. Of course, this gives only the lower limit of the evolutionary inertia since the genome may evolve not starting from the minimum of the curve $x_i(t)$. The present value of c is larger than the previous estimate in [1] where the form of environmental potential was not given.

In the above discussion the parameters α and β in environmental potential W_{env} are assumed

bioRxiv preprint doi: https://doi.org/10.1101/034710; this version posted April 30, 2016. 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.

to be constant corresponding to the stable environmental condition on 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 (5)) to adapt the environment. However, if the sudden change of the environment, 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.

Now we will discuss the evolution of oligonucleotide frequency of a genome. The Eq (5) still holds for k-mer case (k>1) but the environmental potential W_{env} may take a more general function of $\{x_i\}$. Suppose $W_{env} = \chi(N)$ one obtains a similar set of equations for k-mer frequency, namely

$$\frac{d^2x_i}{d\tau^2} = \log\frac{N}{x_i} - \frac{1}{\ln 2}(1-\frac{x_i}{N}) + \frac{d\chi}{dN} - \frac{f}{c}\frac{dx_i}{d\tau} \qquad (13)$$

$$(i = 1,\ldots,m = 4^k), \qquad (\sum_i x_i = N)$$

Consider k=12 case. For 47(48) avian genomes the12-mer frequency distribution peaks at a small fraction of the total set . The highest frequency of a single 12-mer is 16 in a genome. Nearly for all avian genomes only about 11 to 18 of the total set of $m=4^{12}$ oligomers have frequency equal or larger than 5. These q （q=11,·12·,....., or 18）kinds of 12-mers form a subset of $m=4^{12}$ oligomers for each genome. Set $X=x_1+x_2+\ldots+x_q$ representing the total frequency of the subset. We have

$$\frac{d^2X}{d\tau^2} = \sum_i \log\frac{N}{x_i} + \frac{X}{N\ln 2} - \frac{q}{\ln 2} + q\frac{d\chi}{dN} - \frac{f}{c}\frac{dX}{d\tau} \qquad (14)$$

where

$$\sum_i^q \log\frac{N}{x_i} \cong q\log\frac{qN}{X}$$

and

$$\frac{d^2N}{d\tau^2} = \sum \log\frac{N}{x_i} - \frac{m-1}{\ln 2} + m\frac{d\chi}{dN} - \frac{f}{c}\frac{dN}{d\tau} \qquad (15)$$

where

$$\sum_i^m \log\frac{N}{x_i} \cong q\log\frac{qN}{X} + (m-q)\log\frac{N}{N-X} + (m-q)\log(m-q)$$

bioRxiv preprint doi: https://doi.org/10.1101/034710; this version posted April 30, 2016. 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.

$X(t)$ and $N(t)$ can be solved from Equations（14）and（15）.

Quantum state , quantum transition and avian speciation

The nucleotide frequency moving on classical trajectory and obeying deterministic equation cannot explain the new aves production. The sudden change and stochasticity in speciation event should be understood from a broad view. We have demonstrated that the classic equation can be put in a form of the least action principle and following Feynman's path integral approach, the quantum generalization of the evolutionary equation can be deduced [1]. In fact, there exists another approach from classical to quantum. Bohr's correspondence principle shows a good correspondence between classic-mechanical motion and quantum motion in atomic physics. The correspondence does exist in present case. From both approaches we obtain Schrodinger equation, the quantum-generalization of the deterministic equation (4) (as f=0), satisfied by wave function $\psi(\mathrm{x},t)$

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

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

$$p_i=-iL\frac{\partial}{\partial x_i} \tag{16}$$

To be definite we study mono-nucleotide frequency only and assume $V(\mathrm{x},t)$ given by (2) (3) and (7) and α and β dependent of t in general. H is called Hamiltonian of the genome. L is quantization constant, corresponding to Planck constant in 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 near $3\times10^3\ \tau$ where τ is the average lifetime for one generation of the species (for example, τ ~30yr for human and 1 to several years for bird). The quantum evolutionary equation (16) is the logic generalization of the classical equation. 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. 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

$$H\psi(\mathrm{x})=E\psi(\mathrm{x}) \tag{17}$$

The Hamiltonian given by Eq (16) can be rewritten as

bioRxiv preprint doi: https://doi.org/10.1101/034710; this version posted April 30, 2016. 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.

$$H = H_0 + H'$$

$$H_0 = -\frac{L^2}{2c^2}\sum_i \frac{\partial^2}{\partial x_i^2} - (2+\alpha)N + \beta N^2 \tag{18}$$

$$= -\frac{2L^2}{c^2}\sum_i \frac{\partial^2}{\partial N^2} - (2+\alpha)N + \beta N^2$$

$$H' = NR_I = \frac{2}{N\ln 2}\sum_i x_i^2 - \frac{N}{2\ln 2}$$

where Eqs (1)-(3) and (7) have been used. H' is a small quantity which can be looked as a perturbation. Following perturbation theory, $\psi(\mathrm{x}) = \psi^{(0)}(N) + \psi^{(1)}(\mathrm{x})$, the zero order wave function $\psi^{(0)}$ depends on N only, satisfying

$$\{\frac{-2L^2}{c^2(t_0)}\frac{\partial^2}{\partial N^2} - (2+\alpha(t_0))N + \beta(t_0)N^2\}\psi^{(0)}(N) = E^{(0)}\psi^{(0)}(N) \tag{19}$$

The parameters α, β and c take values at time t_0. The solution of Eq(19) is

$$E^{(0)} = E_m^{(0)} = (m+\frac{1}{2})\sqrt{8\beta}\frac{L}{c} - \frac{(2+\alpha)^2}{4\beta} \qquad (m=0,1,2,\ldots) \tag{20}$$

$$\psi^{(0)}(N) = \psi_m^{(0)}(N) = N_m \exp(-\frac{\xi^2}{2})H_m(\xi)$$

$$H_0(\xi) = 1 \quad H_1(\xi) = 2\xi \qquad H_2(\xi) = 4\xi^2 - 2,\ldots \tag{21}$$

$$\xi = \sqrt{\frac{c}{L}}(\frac{\beta}{2})^{1/4}(N - \frac{2+\alpha}{2\beta})$$

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

$$E_m^{(1)} = H'_{mm} = \int d^4x \psi_m^{(0)} H' \psi_m^{(0)}, \quad d^4x = dx_1 dx_2 dx_3 dx_4 \ .$$

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

Eq (20) shows the Hamiltonian-level is equally spaced by $\sqrt{8\beta}\frac{L}{c}$. The spacing $\sqrt{8\beta}\frac{L}{c}$ is a small quantity if c takes a value of thousand year 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

bioRxiv preprint doi: https://doi.org/10.1101/034710; this version posted April 30, 2016. 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.

state the nucleotide frequency always takes some statistical distribution but not a definite value. Eq (21) shows the statistical distribution of N in ground state (m=0) peaks at $N=\frac{2+\alpha}{2\beta}\sim\frac{1}{\beta}$ with width proportional to $\frac{L}{c\sqrt{\beta}}$.

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 $\sqrt{8\beta}\frac{L}{c}$ in a quantum genome. In fact, the gap of ground state should be deeper than $\sqrt{8\beta}\frac{L}{c}$ if the higher- order terms of N^3 etc have been taken into account in the environmental potential equation (7) of W_{env}. Due to the deep gap occurring in the Hamiltonian ground level 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 $\psi_I(\mathrm{x})$ satisfying Eqs (19)-(21) and the final wave function of the "new" species by $\psi_F(\mathrm{x})$ satisfying the same equation but $c^2(t_0)$, $\beta(t_0)$, $\alpha(t_0)$ replaced by $c^2(t_F)$, $\beta(t_F)$, $\alpha(t_F)$ 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

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

$$=\frac{\partial H}{\partial c}\Delta c+\frac{\partial H}{\partial \beta}\Delta\beta+\frac{\partial H}{\partial \alpha}\Delta\alpha\,. \tag{22}$$

Thus the transitional probability amplitude is expressed by

$$T_{fi}=\frac{-i}{L}\int\psi_F^*(\mathrm{x})\int_{t_0}^{t_F}dtH_{\mathrm{int}}(t)\psi_I(\mathrm{x})dx_1dx_2dx_3dx_4$$

$$=\frac{-i}{L}\int\psi_F^*(\mathrm{x})\int_{t_0}^{t_F}dt\{(\frac{\partial H}{\partial c}\frac{dc}{dt}+\frac{\partial H}{\partial \beta}\frac{d\beta}{dt}+\frac{\partial H}{\partial \alpha}\frac{d\alpha}{dt})(t-t_0\}\psi_I(\mathrm{x})dx_1dx_2dx_3dx_4$$

$$=\frac{-i}{L}\int\psi_F^*(\mathrm{x})\{H(t_F)(t_F-t_0)-\int_{t_0}^{t_F}dtH(t)\}\psi_I(\mathrm{x})dx_1dx_2dx_3dx_4$$

where the partial integration has been used. On the other hand,

bioRxiv preprint doi: https://doi.org/10.1101/034710; this version posted April 30, 2016. 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.

$$T_{fi} = \frac{-i}{L}\int \psi_F^*(\mathrm{x})\int_{t_0}^{t_F} dtH_{\mathrm{int}}(t)\psi_I(\mathrm{x})dx_1dx_2dx_3dx_4$$

$$= \frac{-i}{L}\int \psi_F^*(\mathrm{x})\int_{t_0}^{t_F} dt\{H(t)-H(t_0)\}\psi_I(\mathrm{x})dx_1dx_2dx_3dx_4$$

$$= \frac{-i}{L}\int \psi_F^*(\mathrm{x})\{\int_{t_0}^{t_F} dtH(t)-E_I(t_F-t_0)\}\psi_I(\mathrm{x})dx_1dx_2dx_3dx_4$$

So,

$$T_{fi} = \frac{-i}{2L}(E_F-E_I)(t_F-t_0)\int \psi_F^*(\mathrm{x})\psi_I(\mathrm{x})dx_1dx_2dx_3dx_4 \qquad (23)$$

E_I and E_F - the eigenvalue of H, given by Eq(20) with m=0 where parameters c, β, α take values at t_0 and t_F respectively . From Eq（21）, taking m=0 one has

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

$$\psi_F(N) = (\pi a_F)^{1/4}\exp(-\frac{(N-N_F)^2}{2a_F}) \qquad (24)$$

($a_{I,F}$ - the frequency distribution width, and $N_{I,F}$ the frequency distribution centers for two genomes respectively). Inserting (24) into the transitional probability amplitude one obtains the overlap integral

$$\int \psi_F^*(N)\psi_I(N)dN = \pi a_I\exp(-\frac{(N_I-N_F)^2}{4a_I}) \quad (\text{as } a_I = a_F)$$

$$\sqrt{2\pi^3}a_I^{3/4}a_F^{1/4} \quad (\text{as } a_I << a_F)$$

$$\sqrt{2\pi^3}a_F^{3/4}a_I^{1/4} \quad (\text{as } a_I >> a_F) \qquad (25)$$

The transition probability equals $|T_{fi}|^2$, proportional to $\left|\int \psi_F^*(N)\psi_I(N)dN\right|^2$ 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 (25) shows that as a_I 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. The sequence length of posterity genome

bioRxiv preprint doi: https://doi.org/10.1101/034710; this version posted April 30, 2016. 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.

can be longer, or shorter as well, than ancestry. The point is consistent with avian data. The next two lines of Eq (25) 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. Luo LF. 2004. *Theoretic-Physical Approach to Molecular Biology*. Shanghai Scientific & Technical Publishers, Shanghai.
11. Zhang YF, Zhou X, Luo LF,Zhang LR. 2016. Reconstruction of evolutionary tree for avian genomes. International Symposium on Big Data of Science, Baotou.
12. Laxton RR. 1978. The measure of diversity. J Theor Biol. 70(1):51-67.
13. Zhang LR, Luo LF. 2003. Splice site prediction with quadratic discriminant analysis using diversity measure. Nucleic Acids Res. 31(21):6214-6220.