DISCRETE AND CONTINUOUS
DYNAMICAL SYSTEMS
Supplement Volume 2005

Website: http://AIMsciences.org
pp. **91–99**

BIFURCATIONS OF SELF-SIMILAR SOLUTIONS OF THE FOKKER-PLANK EQUATIONS

F. Berezovskaya[1] and G. Karev[2]

[1]Department of Mathematics, Howard University
Washington D.C., 20059,USA
[2]Oak Ridge Institute for Science and Education (ORISE)
8600 Rockville Pike
Bldg. 38A, Rm. 5N511N
Bethesda, MD 20894, USA

Abstract. A class of one-dimensional Fokker-Plank equations having a common stationary solution, which is a power function of the state of the process, was found. We prove that these equations also have generalized self-similar solutions which describe the temporary transition from one stationary state to another. The study was motivated by problems arising in mathematical modeling of genome size evolution.

1. **Motivations and statement of the problem.** A broad variety of phenomena in physics, biology, economics, etc. is described by power law distributions. Recent studies have shown that the distributions of many genome-related quantities could be well described by the so-called Pareto distribution: $P(i) = c(i+a)^{-\gamma}$, where $\gamma > 0$, a are parameters [3, 8, 7]. In our previous work [4, 5], we have shown that well-known birth-and-death processes are a natural source of the power solutions; a special class of such processes (models BDIM, after birth, death and innovation model) shows the power law distribution of its stationary solutions, which are consistent with known data on the sizes of gene families.

The analysis of stochastic BDIMs [6] showed that non-linear versions of such models can well approximate not only the size distribution of gene families but also the dynamics of their formation during genome evolution. The fact that only higher degree BDIMs are compatible with the observed characteristics of genome evolution suggests that the growth of gene families is self-accelerating, which might reflect differential selective pressure acting on different genes.

However, even non-linear BDIMs give unreasonable estimation of the average time of formation of the largest gene families (10^{11} yrs compared to the realistic $\sim 10^9$) and only the minimal time (about $2.5 \cdot 10^9$ yrs) is close to reality. Thus, the problem arises: can we find a different, not purely stochastic approach to model rapid increase of genome size. To examine this problem, we formulate a diffusion approximation of the BDIM (it is well known that a birth-and-death process with discrete space of states is an analog of a diffusion process with a continuous space of states, and vice versa).

In the framework of the diffusion model, we found generalized self-similar solutions (gss) that could be interpreted as the process of deterministic self-accelerating increase of the genome size.

2000 *Mathematics Subject Classification.* Primary: 92B05.
Key words and phrases. Fokker-Plank equations, self-similar solutions, genom evolutions.

This paper is organized as follows. In section 2, we describe the Fokker-Plank equation (FPE) corresponding to the non-linear BDIM; the stationary solutions of this FPE are considered in section 3. An important example of the linear diffusion model is considered in section 4. In section 5, we describe a class of non-linear FPEs which have a given common stationary solution. The core section 6 contains the definition and investigation of the generalized self-similar solutions of the FPE. Section 7 contains a brief discussion of obtained solutions. The proofs of main theorems are given in the Appendix.

2. **Diffusion version of the BDIM and the Fokker-Plank equation.**

Let a population be subdivided into N (finite or infinite) different groups, which we will call "families" and $f(t,x)$ be the number of families of the size x in t time moment. Let us suppose that the "individual" birth and death rates in a family of the size x are $\lambda(x)$ and $\delta(x)$ respectively. Then the equation

$$\frac{\partial f(t,x)}{\partial t} = f(t,x-1)\lambda(x-1) - (\lambda(x)+\delta(x))f(t,x) + \delta(x+1)f(t,x+1), \quad (1)$$

where $x > 0$, (subject to boundary conditions) describes the birth-and-death process with the set of states $\{0, 1, \ldots N\}$.

A formal continuous approximation of equation (1) gives the equation:

$$\frac{\partial f(t,x)}{\partial t} = -\frac{\partial}{\partial x}[f(t,x)\mu(x)] + \frac{1}{2}\frac{\partial^2}{\partial x^2}[f(t,x)\sigma^2(x)] \quad (2)$$

where $\mu(x) = \lambda(x) - \delta(x)$ is the drift and $\sigma^2(x) = \lambda(x) + \delta(x)$ is the diffusion coefficient.

Equation (2) is the *Fokker-Plank equation* (FPE) for the considered process.

Diffusion model (2) could be considered as a limit of birth-and-death chains in some exact sense under "scaling conditions" (see, e.g., [2] for details). In general, the problem of "equivalence" of models (1) and (2) is non-trivial and the models may have different dynamics. Our aim is to develop the appropriate mathematical tools for modeling of the genome evolution rather then investigation of the equivalence of different mathematical approaches. So, we prefer here to consider the FPE (2) as a separate mathematical model of genome evolution rather then only technical approximation of the initial birth-and-death process with discrete set of states.

Let us denote $J(t,x)$ the *current* of particles through the point x at instant t:

$$J(t,x) = f(t,x)\mu(x) - \frac{\partial}{\partial x}[\frac{1}{2}f(t,x)\sigma^2(x)].$$

Then the Fokker-Plank equation (2) could be written as the equation of continuity (or the equation of mass conservation):

$$\frac{\partial f(t,x)}{\partial t} = -\frac{\partial}{\partial x}J(t,x). \quad (3)$$

To solve this equation, we need an initial condition and boundary conditions at the ends of the interval $[r,\ N]$. For example, if the system is "closed", i.e., a particle cannot leave the interval and there is zero net flow across the ends, then

$$J(t,x) = 0 \quad \text{at } x = r \text{ and } x = N. \quad (4)$$

If the system is "open" and innovation is possible, then the current $J(t,r)$ through the left end, or the rate of innovation $\nu(t) = -\frac{\partial}{\partial x}f(t,x)\lambda(x)|_{x=r}$ could be taken as the boundary condition of Fokker-Plank equation (2) or (3).

In most applications $\sigma^2(x)$ is smaller or of the same order as the drift $\mu(x)$. For example, as follows from formula (5) below, if $\sigma^2(x)$=const and $\mu(x) = cx$, $c < 0$ is a constant, then a stationary solution $f_{st}(x)$ of (2) follows the (truncated) normal distribution; if $\sigma^2(x)$=const and $\mu(x) = c < 0$, then its stationary solution follows the (truncated) exponential distribution.

In this paper, we explore the diffusion approximation (2) of the polynomial and rational BDIMs which have been previously considered for a discrete phase space [4, 5]. We show that the stationary solution $f_{st}(x)$ of model (2) follows the Pareto distribution only when $\sigma^2(x)$ increases faster then $|\mu(x)|$ and the drift is negative as opposed to the usually considered cases.

Next, the problem of estimation of the duplication/deletion rates of genes is hard because the birth and death rates or drift and diffusion coefficients in models of genome size evolution are actually unknown. By contrast, the distributions of gene family sizes are well-established empirical data. Thus, we constructed and explored a class of diffusion models that have a given common stationary solution but various drift and diffusion coefficients. We proved that such models have special "self-similar" solutions that describe the transitions from one stationary solution to another. A speed of the movement of the "front" of this solution depends on the model parameters.

3. **Stationary solution of the model and the power asymptotics.** The stationary solution $f_{st}(x)$ of model (2), for which $\frac{df_{st}(x)}{dt} = 0$, satisfies the equation

$$-\frac{\partial}{\partial x}[f_{st}(x)\mu(x)] + \frac{1}{2}\frac{\partial^2}{\partial x^2}[f_{st}(x)\sigma^2(x)] = \frac{\partial}{\partial x}J_{st}(x) = 0,$$

so the current $J_{st}(x) = \text{const}$ at all x.

If the system is closed and hence $J(t, x) = 0$ at $x = r$ (due to the boundary condition (4)) then

$$f_{st}(x)\mu(x) = \frac{1}{2}\frac{\partial}{\partial x}[f_{st}(x)\sigma^2(x)]$$

for all $x \in [r, N]$.

The solution of this equation is

$$f_{st}(x) = \frac{\sigma^2(r) f_{st}(r)}{\sigma^2(x)} \exp(2\int\limits_r^x \frac{\mu(u)}{\sigma^2(u)} du). \tag{5}$$

If the system is open and $J_{st}(t, x) \neq 0$ at $x = r$, then other stationary solutions can exist (see, e.g., [2], ch.5). The following assertion easily follows from (5).

Theorem 1. *Let* $[\mu(x) + \frac{1}{2}\frac{\partial}{\partial x}\sigma^2(x)]/\sigma^2(x) = -\frac{\gamma}{2x} + S(x)$, *where* γ *is a constant and* $S(x)$ *satisfies the condition:* $\int_r^x S(y)dy \to \text{const}$ *at* $x \to \infty$. *Then* $f_{st}(x) \sim x^{-\gamma}$.

Corollary 1. *Let* $\sigma^2(x) = x^\rho(a + o(1/x))$, *where* ρ, a *>0 are constants; let* $\frac{\mu(x)}{\sigma^2(x)} = -\eta/(2x) + O(1/x^2)$. *Then* $f_{st}(x) \sim x^{-\eta-\rho}$.

As a representative example, let us consider a *linear* diffusion model with $\lambda(x)$ and $\delta(x)$ being linear functions of x: $\lambda(x) = \lambda_0(x + a)$, $\delta(x) = \delta_0(x + b)$ where λ_0 and δ_0 are positive constants. Then $\mu(x) = (\lambda_0 - \delta_0)x + \lambda_0 a - \delta_0 b$, $\sigma^2(x) = (\lambda_0 + \delta_0)x + \lambda_0 a + \delta_0 b$. Suppose also that the constants a, b are such that $\sigma^2(x) > 0$ in $[r, N]$. The linear discrete-state BDIM [4] has a stable distribution $f_{st}(x)$, which is asymptotically equal to the power-law distribution if and only if $\lambda_0 = \delta_0$; then

$f(x) \sim x^{-\gamma}$, where $\gamma = b - a + 1$. Similar result is valid also for diffusion linear model:

$$\frac{\partial f(x,t)}{\partial t} = -\frac{\partial}{\partial x}[f(x,t)((\lambda_0 - \delta_0)x + \lambda_0 a - \delta_0 b)] + \frac{1}{2}\frac{\partial^2}{\partial x^2}[f(x,t)(\lambda_0 + \delta_0)(x+s)], \quad (6)$$

where $s = (a\lambda_0 + b\delta_0)/(\lambda_0 + \delta_0)b$.

Proposition 1. *For linear diffusion model* (6), $f_{st}(x) = C\exp(-lx)(s+x)^{-\gamma}$ *where* $\gamma = 4(b-a)\frac{\lambda_0\delta_0}{(\lambda_0+\delta_0)^2} + 1$, $C = f_{st}(r)\exp(lr)(s+r)^{\gamma+1}$ *and* $l = 2(\delta_0 - \lambda_0)/(\lambda_0 + \delta_0)$.

Corollary 2. *The stationary distribution of the linear diffusion model is a Pareto distribution if and only if* $\lambda_0 = \delta_0$ *and* $a - b \neq 1$*; under these conditions,* $f_{st}(x) = c_1(s+x)^{-\gamma}$ *where* $\gamma = b - a + 1$, $s = (a+b)/2$ *and* $c_1 = f_{st}(r)(s+r)^{\gamma}$.

Next, let us consider a more general case and suppose that birth and death rates are polynomials on x. Informally, polynomial models could take into account interactions between particles and reflect a feedback between the family size and growth rate (these models with discrete space of states were studied in [4, 5]. Formula (5) yields an explicit stationary distribution for polynomial models because the rational function $\mu(x)/\sigma^2(x)$ is integrable.

The problem of the most interest is the asymptotical behavior of the stationary distribution. This behavior critically depends on the relation between the degrees of the polynomials.

Proposition 2. *Let the birth and death rates be of the form* $\lambda(x) = \lambda_0 x^{\rho} R(x)$, $\delta(x) = \delta_0 x^{\rho} Q(x)$, *where* $\rho > 0$, $R(x)$,$Q(x)$ *are polynomials of the same degree* $m \geq 1$*:*

$$R(x) = \sum_{s=0}^{m} r_s x^{m-s}, \quad Q(x) = \sum_{s=0}^{m} q_s x^{m-s}, \quad r_0 = q_0 = 1.$$

Then $f_{st}(x) \sim \exp(-lx)x^{-\nu-m}$ *where* $l = 2(\delta_0 - \lambda_0)/(\lambda_0 + \delta_0)$, $\nu = 4(q_1 - r_1)$*; in particular,* $f_{st}(x) \sim x^{-\nu-m}$ *if* $\delta_0 = \lambda_0$.

4. **Spatial-temporal dynamics of the linear model; a special class of solutions.** Let us consider the linear model (6) with $\lambda_0 = \delta_0$; in this case $\mu(x) = \lambda_0(a-b)$, $\sigma^2(x) = \lambda_0(2x + a + b)$. After simple algebra, equation (6) reads

$$\frac{\partial f(x,t)}{\lambda \partial t} = (1+\gamma)\frac{\partial f}{\partial x} + (x+s)\frac{\partial^2 f}{\partial x^2}, \quad (7)$$

where $\gamma = b - a + 1, s = (a+b)/2$.

The stationary solution of this model is

$$f_{st}(x) = c_1(s+x)^{-\gamma} \quad (8)$$

where $c_1 = f_{st}(r)(s+r)^{\gamma}$; its current is

$$J(t,x) = -\lambda_0\gamma f(t,x) - \lambda_0(x+s)\frac{\partial}{\partial x}f(t,x).$$

The following theorem describes a special case of "self-similar" solutions.

Theorem 2. *Linear model* (7) *with* $N = \infty$ *for any* $t_0 \geq 0$ *has a solution*

$$f(x,t) = (x+s)^{-\gamma}\{C_1 + C_2\Gamma(\gamma, \frac{x+s}{\lambda t})\} \quad (9)$$

where $\gamma = b - a + 1, s = (a+b)/2$ *and* $\Gamma(a,x) = \int_x^{\infty}\exp(-t)t^{a-1}dt$ *is the incomplete Gamma-function.*

Formula (9) describes the conversion of the initial stationary solution $f_{st}(x) = C_1(x+s)^{-\gamma}$ into another stationary solution of the same "Pareto shape", $f_{st}(x) = (C_1 + C_2\Gamma(\gamma))(x+s)^{-\gamma}$. Indeed, according to the properties of incomplete Γ-function, $\Gamma(\gamma, 0) = \Gamma(\gamma)$, $\Gamma(\gamma, \infty) = 0$. Hence, $f(x,0)=C_1(x+s)^{-\gamma}$, and $f(x,t) \to (C_1 + C_2\Gamma(\gamma))(x+s)^{-\gamma}$ at $t \to \infty$.

More general results are proved below (see Theorems 3 and 4).

5. **Transformations of the model.**

The initial point of our investigation was the stationary distribution of sizes of domain families, which was extracted from the empirical data and followed the (truncated) Pareto distribution. Here we construct a class of FPE models with various dynamics but with a common, fixed stationary distribution.

The diffusion model

$$\frac{\partial f(t,x)}{\partial t} = -\frac{\partial}{\partial x}[f(t,x)(\lambda(x) - \delta(x))] + \frac{1}{2}\frac{\partial^2}{\partial x^2}[f(t,x)(\lambda(x) + \delta(x))] \tag{10}$$

has the stationary solution (5) with $\mu(x)=\lambda(x) - \delta(x)$, $\sigma^2(x) = \lambda(x) + \delta(x)$, $r < x < N$.

Let $g(x)$ be a positive smooth function; transform the initial birth and death rates $\lambda(x)$ and $\delta(x)$ using formulas

$$\lambda^*(x) = \lambda(x)g(x) + \frac{1}{4}(\delta(x) + \lambda(x))\frac{dg(x)}{dx}) = \lambda(x)g(x) + \frac{1}{4}\sigma^2(x)\frac{dg(x)}{dx},$$
$$\delta^*(x) = \delta(x)g(x) - \frac{1}{4}(\delta(x) + \lambda(x))\frac{dg(x)}{dx}) = \delta(x)g(x) - \frac{1}{4}\sigma^2(x)\frac{dg(x)}{dx}$$

We suppose that the function $g(x)$ is such that $\lambda^*(x)$ and $\delta^*(x)$ are non-negative for $r < x < N$.

Proposition 3. *The stationary solutions of the initial diffusion model* (10) *and transformed model with*

$$\sigma^{*2}(x) = \lambda^*(x) + \delta^*(x) = \sigma^2(x)g(x),$$
$$\mu^*(x) = \mu(x)g(x) + \frac{1}{2}\sigma^2(x)\frac{dg(x)}{dx}$$

are identical up to the normalizing constant.

The proof easily follows from formula (5).

For the given functions $\mu(x)$ and $\sigma^2(x)$, let us define the operator

$$J[f](t,x) = f(t,x)\mu(x) - \frac{1}{2}\frac{\partial}{\partial x}[f(t,x)\sigma^2(x)].$$

If the function $f(t,x)$ satisfies the FPE (3), then $J[f]$ is the current for diffusion model (2). Denote $J^*[f]$ the operator corresponding to the transformed functions $\mu^*(x)$ and $\sigma^{*2}(x)$.

Lemma 1.

$$J^*[f](t,x) = g(x)J[f](t,x).$$

Let us explore the transformations of the linear diffusion model (7) which has the stationary solution (8). The current for this model is $J[f](t,x) = -\lambda_0[\gamma f(t,x) + (x+s)\frac{\partial}{\partial x}f(t,x)])$, hence the current for the transformed model is $J^*[f](t,x) = -\lambda_0 g(x)\ [\gamma f(t,x) + (x+s)\frac{\partial}{\partial x}f(t,x)]$.

Thus, the FPE for the transformed model is $\frac{\partial}{\partial t}f(t,x) = -\frac{\partial}{\partial x}J^*[f](t,x)$, or

$$\frac{\partial}{\lambda_0 \partial t}f(t,x) = \frac{\partial}{\partial x}g(x)[\gamma f(t,x) + (x+s)\frac{\partial}{\partial x}f(t,x)].$$

Below we explore the transformation of the linear diffusion model using the function $g(x) = (x+s)^{\rho-1}$, $\rho \geq 1$, $b > (\rho-1)/2$. Then

$$\lambda^*(x) = (x + a + \frac{1}{2}(\rho-1))(x+s)^{\rho-1}, \quad \delta^*(x) = (x + b - \frac{1}{2}(\rho-1))(x+s)^{\rho-1},$$

and $\mu^*(x) = (\rho - \gamma)(x+s)^{\rho-1}$, $\sigma^{*2}(x) = 2(x+s)^{\rho}$. The corresponding FPE is of the form

$$\frac{\partial}{\lambda_0 \partial t}f(t,z) = z^{\rho}\{\frac{\partial^2}{\partial z^2}f(t,z) + \frac{\gamma+\rho}{z}f(t,z) + \frac{\gamma(\rho-1)}{z^2}f(t,z)\}, \tag{11}$$

where $z = x+s$. Equation (11) has a set of stationary solutions:

$$\begin{aligned} f_{st}(t,z) =& z^{-(\gamma+\rho-1)/2}\{C_1 BesselJ[\frac{1}{2}(\gamma+\rho-1), (\gamma(\rho-1))^{1/2}] \\ &+ C_2 BesselY[\frac{1}{2}(\gamma+\rho-1), (\gamma(\rho-1))^{1/2}]\}. \end{aligned}$$

6. **Spatial-temporal dynamics of the transformed models.** A solution $f(t,x)$ of equation (2) is called *generalized self-similar* (gss) if it is of the form $f(x,t) = x^a G(y)$ where $y = x/\phi(t)$ with smooth $\phi(t) \neq 0$ and a is a constant.

The following Theorem 3 describes all possible gss-solutions of model (11) dependently on the "model degree" ρ. Only bounded solutions have a "biological" meaning; corresponding conditions are given in Theorem 4.

Theorem 3. *Let C_1, C_2 be arbitrary constants and $t_0 \geq 0$.*

i) *For $\rho < 2$ equation (11) has a three-parametric (C_1,C_2, t_0)-family of gss-solutions :*

$$f(t,x) = (x+s)^{-\gamma}\{C_1 + C_2\Gamma(1 + \frac{\gamma-1}{2-\rho}, \frac{(x+s)^{2-\rho}}{(2-\rho)^2\lambda_0(t+t_0)})\}. \tag{12}$$

ii) *For $\rho = 2$, equation (11) has a four-parametric (C_1, C_2, t_0, α)-family of gss-solutions*

$$f(t,x) = (x+s)^{-\gamma}\{C_1 + C_2[(x+s)\exp(-\alpha\lambda_0(t+t_0))]^{\gamma-\alpha-1}\}, \tag{13}$$

where α is an arbitrary constant such that $\alpha \neq \gamma - 1$, and three-parametric (C_1, C_2, t_0)-family of gss-solution

$$f(t,x) \; = \; (x+s)^{-\gamma}(C_1 + C_2 \;(\log(x+s) - (\gamma-1)\lambda_0(t+t_0))), \tag{14}$$

if $\alpha \; = \gamma - 1$.

iii) *For $2 < \rho < 1 + \gamma$, equation (11) has (C_1, C_2, t_0)-family of gss-solutions*

$$\begin{aligned} f(t,x) =& (x+s)^{-\gamma}\{C_1 + C_2((\lambda_0(t+t_0))^{1/(\rho-2)}(x+s))^{\gamma-\rho+1} \cdot \\ & E(\frac{\gamma-1}{2-\rho}, \frac{1}{\lambda_0(t+t_0)(s+x)^{\rho-2}(\rho-2)^2})\}, \end{aligned} \tag{15}$$

where $E(u,v) = \int_1^{\infty} \exp(-vt)t^{-u}dt$ is a special function, the Exponential Integral.

iv) *For $\rho > 1 + \gamma$, equation (11) has (C_1, C_2, t_0)-family of gss-solutions*

$$f(t,x) = (x+s)^{-\gamma}\{C_1 + C_2\Gamma(1 - \frac{\gamma-1}{2-\rho}, \frac{1}{\lambda_0(t+t_0)(s+x)^{\rho-2}(\rho-2)^2})\}. \tag{16}$$

Theorem 4. *Let* $t_0 = 0$*; then for any fixed value of* x*,*

1) *gss-solutions* (12) *and* (16) *are bounded functions of* t *at* $t \to 0$, $t \to \infty$ *for any constants* C_1, C_2;

2) *gss-solution* (13) *is a bounded function of* t *at* $t \to 0$, $t \to \infty$ *for any value of* α *such that* $0 < \alpha < \gamma - 1$ *or* $0 > \alpha > \gamma - 1$ *and any constants* C_1,C_2*;*

if the constant $C_2 \neq 0$*, then any solution* (13) *with* $\alpha < \gamma - 1 < 0$ *or with* $\alpha > \gamma > 0$ *as well as any solution* (14) *with* $\gamma \neq 1$ *is unbounded at* $t \to \infty$*;*

3) *For any constants* C_1 *and* $C_2 \neq 0$, *gss-solution* (15) *is unbounded at* $t \to \infty$*.*

Proofs of Theorems 3 and 4 are given in the Appendix.

7. **Discussion.** Asymptotic behavior of generalized self-similar solutions of FPE (10), in which diffusion and drift coefficients are rational functions $\sigma^2(x) = x^\rho(a + O(1/x))$ where a >0, and $\mu(x) = x^{\rho-1}(b + O(1/x))$, respectively, is essentially determined by the "degree of non-linearity" $\rho \geq 1$ and the first coefficients a, b of their expansions. If $\mu(x)/\sigma^2(x) = -2\gamma/x + O(1/x^2)$ where γ >0, the FPE has a power stationary solution $f_{st}(x) \sim x^{-\gamma}$. Given the stationary solution, we construct a class of FPE such that every equation from this class has the given stationary solution. For the FPE‘s we found also families of generalized self-similar solutions $f(t, x)$. Let us emphasize that gss-solutions describe the regime of spreading of the "profile", or "front" of the solution along the x axis with time (in terms of the original problem, this solution describes the increase of genome size). The speed and direction of the movement of these solutions can be defined from the relation $f(t, x) = \text{const}$.

In the context of modeling genome size evolution we are interested mainly in bounded solutions of the model. In particular, solution (12) ($1 \leq \rho < 2$) describes the transition from the initial stationary distribution $f(0, x) = C_1(x + s)^{-\gamma}$ to the final stationary distribution $f(\infty, x)$=$(x + s)^{-\gamma}[C_1 + C_2\Gamma(1 + \frac{\gamma-1}{2-\rho})]$ which differs from the initial one only by a constant multiplier. The front of solution (12) moves to the right if C_2 >0 (see Fig.1).

In the case $\rho = 2$ the behavior of gss-solutions can substantially change. Any solution (13) is bounded only if $0 \leq \alpha < \gamma + 1$. In this case, the solution describes the transition from the initial stationary distribution $f(0, x) = C_1(x+s)^{-\gamma}+C_2(x+s)^{-\alpha-1}$ to the final stationary distribution $f(\infty, x) = C_1(x+s)^{-\gamma}$. Let us note that the initial and final solutions now have different shapes; the front of this solution moves to the right if $C_2 > 0$.

For $2 < \rho < 1+\gamma$, all gss-solutions (15) (which are different from the stationary one) tend to infinity in any point x at $t \to \infty$.

Finally, if $\rho > 1 + \gamma$, gss-solution (16) describes the transition from the initial stationary distribution $f(0, x) = C_1(x + s)^{-\gamma}$ to the final stationary distribution $f(\infty, x) = (x + s)^{-\gamma}[C_1+ C_2\Gamma(1 - \frac{\gamma-1}{\rho-2})]$, similarly to the solution (12).

Let us note that the "structure" of the self-similar variable $y = x/t^\beta$ abruptly changes at $\rho = 2$, namely, the exponent β >0 for $1 \leq \rho < 2$ and $\beta < 0$ for $\rho > 2$. We suspect that the case $\rho \leq 2$ is more suitable for modeling genome size evolution.

8. **Appendix.**

Proof of Theorem 3. i) Let $\rho \geq 1$, $\rho \neq 2$. Searching a gss-solution of equation (11) in the form:

$$f(z, \tau) = z^\alpha G(y) = z^\alpha G(z\tau^{-\beta})$$

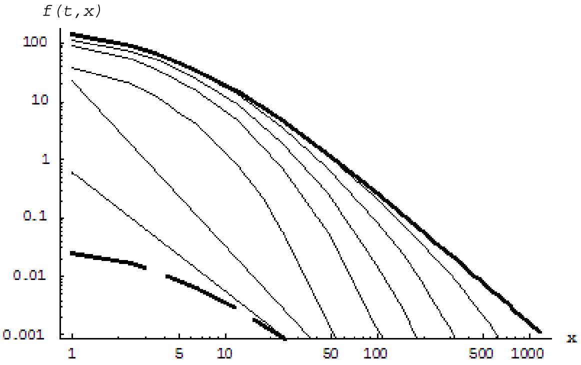

FIGURE 1. Transformation of the initial stationary solution (dash) into the final one (solid); the self-similar solution (12) with $\rho = 1.5$ is shown at different time moments

where $z = x + s$, $\tau = \lambda_0(t + t_0)$ and constants α, β should be determined. One easily gets that $G(y)$ must satisfy the equation

$$y^2 G_{yy}(y) + (y(2\alpha + \gamma + \rho) + \beta y^{\kappa})G_y(y) + (\alpha + \rho - 1)(\alpha + \gamma)G(y) = 0,$$

for $\beta = 1/(2 - \rho)$ and $\kappa = 3 - \rho$. Choosing $a = -\gamma$ one gets

$$yG_{yy}(y) + (\rho - \gamma + \frac{y^{2-\rho}}{2-\rho})G_y(y) = 0. \tag{17}$$

Equation (17) has the following general solution:

$$G(y) = C_1 + C_2 \quad y^{\gamma-\rho+1} \int_1^{\infty} t^{-u} e^{-vt} dt = C_1 + C_2 y^{1+\gamma-\rho} E(u, v), \tag{18}$$

where $v = y^{2-\rho}/(2 - \rho)^2$ and $u = (\gamma - 1)/(\rho - 2)$. Thus, equation (11) has the gss-solution

$$f(t, x) = \frac{1}{(x+s)^{\gamma}}\{C_1 + C_2((\lambda_0(t + t_0))^{\frac{1}{(\rho-2)}}(x + s))^{\gamma-\rho+1}\cdot$$
$$E(\frac{\gamma-1}{2-\rho}, \frac{(s+x)^{2-\rho}}{\lambda_0(t+t_0)(\rho-2)^2})\}.$$

For ρ <2 and $\rho > 1 + \gamma$ (in both cases $1 - u > 0$), solution (18) can be written in the form

$$G(y) = C_1 + C_3\Gamma(1 - u, v) = C_1 + C_3\Gamma\ (1 + \frac{\gamma-1}{2-\rho}, \frac{y^{2-\rho}}{(2-\rho)^2})$$

and corresponding gss-solution read

$$f(t, x) = (x + s)^{-\gamma}\{C_1 + C_2\Gamma(1 + \frac{\gamma-1}{2-\rho}, \frac{(x+s)^{2-\rho}}{(2-\rho)^2\lambda_0(t+t_0)})\}.$$

ii) Let $\rho = 2$; searching a gss-solution of equation (11) in the form

$$f(z, t) = z^{-\gamma}G(y) \quad \text{where } y = z\exp(-\alpha\tau),$$

one can show that $G(y)$ satisfies the equation

$$yG_{yy}(y) + (2 - \gamma + \alpha)G_y(y) = 0.$$

This equation has a general solution $G(y) = C_1 + C_2 y^{-(\alpha+1-\gamma)}$ if $\alpha + 1 - \gamma \neq 0$ and $G(y) = C_1 + C_2 \ln\, y$ if $\alpha = \gamma - 1$. Thus, equation (11) has the gss-solution

$$f(t,x) = C_1(x+s)^{-\gamma} + C_2(x+s)^{-\alpha-1}\exp(\alpha(\alpha-\gamma+1)\lambda_0(t+t_0))$$

if $\alpha + 1 - \gamma \neq 0$ and

$$f(t,x) = (x+s)^{-\gamma-1}[C_1 + C_2(\ln(x+s) - \alpha\lambda_0(t+t_0))]$$

for $\alpha + 1 = \gamma$. □

Proof of Theorem 4. Recall that $\Gamma(u,v) \to 0$ at $v \to \infty$ and $\Gamma(u,v) \to \Gamma(u)$ at $v \to 0$ for positive u. Hence, for $\rho < 2$ and for $\rho > 1+\gamma$, $\Gamma(1+\frac{\gamma-1}{2-\rho}, \frac{(x+s)^{2-\rho}}{(2-\rho)^2\lambda_0 t}) \to 0$ at $t \to 0$, and $\Gamma(1+\frac{\gamma-1}{2-\rho}, \frac{(x+s)^{2-\rho}}{(2-\rho)^2\lambda_0 t}) \to \Gamma(1+\frac{\gamma-1}{2-\rho}) > 0$ at $t \to \infty$. Hence, $f(0,x) = C_1(x+s)^{-\gamma}$ and $f(t,x)$ tends to $f(\infty,x) = (x+s)^{-\gamma}\{C_1 + C_2\Gamma(1+\frac{\gamma-1}{2-\rho})\}$ at $t \to \infty$. So, solution (12) describes the transition from the initial stationary distribution $f(0,x)$ to the final stationary distribution $f(\infty,x)$, which differs from the initial one only by a constant multiplier.

For $\rho = 2$, gss-solutions depend on a free parameter α. Solution (13) is bounded for all t only if $0 \leq \alpha < \gamma - 1$ or $0 > \alpha > \gamma - 1$. It is easy to see that at $\alpha \to \gamma - 1$ solution (13) transforms to (14) which is unbounded at $t \to \infty$.

It is known that $e^v E(u,v) = O(1/v)$ ([1], 5.1.19), hence $v^s E(u,v) \to 0$ at $v \to \infty$ for any $s > 0$. When this relation is applied to solution (15) for $2 < \rho < 1+\gamma$, we see that term $C_2((\lambda t)^{\frac{1}{\rho-2}}(s+x))^{\gamma-\rho+1}E(\frac{\gamma-1}{2-\rho}, \frac{1}{(\lambda_0 t)(s+x)^{\rho-2}(\rho-2)^2})$ tends to 0 with $t \to 0$. Hence, solution (15) tends to $C_1(x+s)^{-\gamma}$ at $t \to 0$. Next, for $u > 1$, $E(u,v) \to \frac{1}{u-1}$ at $v \to 0$. If $2 < \rho < 1+\gamma$ then $E(\frac{\gamma-1}{2-\rho}, \frac{1}{(\lambda_0 t)(s+x)^{\rho-2}(\rho-2)^2}) >$ const >0 at $t \to \infty$. Hence, the solution (15) is unbounded because this exponential integral is multiplied by the factor $(\lambda_0 t)^{\frac{(\gamma-\rho+1)}{\rho-2}}$ that tends to infinity. □

Acknowledgements. Author thanks Dr. E. Koonin for valuable discussions and help in preparation of the manuscript. The work of F. Berezovskaya was supported by NSF Grant #634156.

REFERENCES

[1] M. Abramovitz and I.A. Stegun, "Handbook of Mathematical Functions", Dover Publ. Inc., New York, 1970.

[2] C.W. Gardiner, "Handbook fo Stochastic Models for Physics, Chemistry and the Natural Sciences", Springer, Berlin, 3rd edition, 2004.

[3] T. Gisiger, *Scale invariance in biology: coincidence or footprint of a universal mechanism?* Biol. Rev. Camb. Phil. Soc., **76** (2001), 161–209.

[4] G.P. Karev, Yu.I. Wolf, A.Y. Rzhetsky, F.S. Berezovskaya, E.V. Koonin, *Birth and death of protein domains: A simple model of evolution explains power law behavior.* BMC. Evol. Biol., **2** (2002), 18.

[5] G.P. Karev, Yu.I. Wolf, E.V. Koonin, *Simple stochastic birth and death models of genome evolution: was there enough time for us to evolve?* Bioinformatics, **19** (2003), 1889–1900.

[6] G.P Karev, Y.I. Wolf, F.S. Berezovskaya, E.V. Koonin, *Gene family evolution: an in-depth theoretical and simulation analysis of non-linear birth-death-innovation models.* BMC Evol. Biol., **4** (2004), 32.

[7] E.V. Koonin, Y.I. Wolf, G.P. Karev, *The structure of the protein universe and genome evolution*, Nature, **420** (2002), 218–223.

[8] J. Qian, N. Luscombe, M. Gerstein, *Protein family and fold occurrence in genomes: power-law behaviour and evolutionary model*, J. Mol. Biol., **313** (2001), 673–681.

Received September, 2004; revised May, 2005.

E-mail address: `fsberezo@hotmail.com, karev@ncbi.nlm.nih.gov`