A case study of genome evolution: from continuous to discrete time model*

Jerzy Tiuryn\textsuperscript{1}, Ryszard Rudnicki\textsuperscript{2}, and Damian Wójcik\textsuperscript{1}

\textsuperscript{1} Institute of Informatics, Warsaw University
\textsuperscript{2} Institute of Mathematics, Polish Academy of Sciences

Abstract. We introduce and analyse a simple model of genome evolution. It is based on two fundamental evolutionary events: gene loss and gene duplication. We are mainly interested in asymptotic distributions of gene families in a genome. This is motivated by previous work which consisted in fitting the available genomic data into, what is called paralog distributions. Two approaches are presented in this paper: continuous and discrete time models. A comparison of them is presented too - the asymptotic distribution for the continuous time model can be seen as a limit of the discrete time distributions, when probabilities of gene loss and gene duplication tend to zero. We view this paper as an intermediate step towards mathematically settling the problem of characterizing the shape of paralog distribution in bacterial genomes.

1 Introduction

Fitting data into various kinds of plots is a common practice of modern biology. A typical case is a study of genome organization and evolution, which can be viewed as a branch of a relatively new area of computational biology, called \textit{comparative genomics} (see \cite{Koc03}). We can view a genome not simply as a set of genes, but rather as a dynamic collection of genes which changes in time. Various biochemical processes (e.g. point mutation, recombination, gene conversion, replication, DNA repair, translocation, horizontal transfer) constantly act on genomes and drive them to evolve dynamically. A problem which has been addressed in late 90's in this framework is an estimate of the distribution of paralogs in a genome. Two genes in a genome are said to be \textit{paralogs} if they have evolved through duplication from a single ancestral gene. We do not discuss here the important issue of deciding which genes are paralogs. We assume that all genes have been clustered into groups of pairwise paralogous genes. The question which was asked in 1998 by P. Slonimski (\cite{Slon98,Slon99}) and independently by M.A. Huynen and E. van Nimwegen \cite{Huyn98} was about the distribution of the numbers of $i$-element clusters of paralogous genes (for consecutive $i$'s) in several microbial genomes which have been sequenced till then. The distribution was

* This research was partially supported by the State Committee for Scientific Research (Poland) Grants No. 2 PO3A 031 25, and 7 T11F 016 21 and by the EC programme \textit{Centres of Excellence for States in phase of pre-accession}, No. ICA1-CT-2000-70024.
estimated by fitting the available data. Since the method of deciding paralogy is only approximate and the size of the genomes was not large, as a consequence both authors came up with different answers: [Slo98] claims that the distribution is logarithmic, while [Huy98] claims that it follows the so called power law distribution. In 2001 Jordan et al. [Jor01] have analysed 21 completely sequenced bacterial genomes and claimed that the logarithmic approximation fits the distribution slightly better than the power law approximation, although the difference between the two fits is not significant.

It should be obvious from the above description that it will be impossible to decide what actually is the observed distribution if we rely merely on the biological data. A decisive answer should come by adopting a certain mathematical model of genome evolution together with a rigorous analysis of the asymptotic distribution within this model. This is the main motivation for the present paper to build and analyse a simple model of genome evolution. The model we study is very simple indeed. It addresses only two evolutionary events: gene loss and gene duplication. Even though it is too simple to settle the problem of distribution of paralog families in genomes it can be used to study various subtleties of the model. We treat this paper as an intermediate step towards analysis of a more complicated model, which we postpone for future publication.

There is a short history of mathematical modeling of genome evolution. In 2000 Yanai et al. [Yan00] designed a simple model of genome evolution based on random gene duplication and point mutations. The paper did not analyse the model. The main result consisted in showing that it is possible for each of the 20 microbial genomes to tune the parameters of the model so that the obtained distribution matches closely the paralog distribution of the genome. Recently Koonin's group has published in a series of papers [Koo02, Kar02, Ka03a, Ka03b] a simple model (called BDIM) of genome evolution which resembles our continuous time model. However, there are two important differences between the two models. In BDIM model in addition to gene loss and duplication there is an external source of new genes, called invention. This source is used to stabilize the asymptotic behavior of the model, i.e. to make sure that the supply of genes does not vanish at some point of evolution. On the contrary, we are interested in asymptotic distributions for the two extreme situations: genome collapse and genome explosion. The reason for this is that if we assume that the two events: gene loss and gene duplication are independent of each other, it follows that we have to assume that their probabilities should not be equal. This leads the model to one of the two extreme situations. The second difference is more important. BDIM model sets an upper bound on the maximal size of gene family in the model. Technically speaking, this assumption implies that the system of differential equations is finite and the theory of finite dimensional matrices is applicable here. In the model which we investigate in this paper we do not impose any bound on the maximal size of a gene family and we end up with an infinite system of differential equations, for which existence of stationary solutions needs a special justification (see Theorem 1).
Continuous time model represents an ideal situation: in one unit of time two or more events can happen to a single gene, even though the probability of this is very low. In discrete time model we assume that in one unit of time every gene of the genome is subject to exactly one of the following events: removal, duplication, idle; each with a fixed probability. So, discrete time model is an approximation (and simplification) of the continuous model. Discrete model is much more suitable for computer simulations. Also, as we will see, the asymptotic distributions for both models are always different. The analysis of the discrete model is apparently more complicated, presumably due to lack of strong analytical tools. Moreover, as it follows from one of our results (see Theorem 7), the distribution for a continuous model can be seen as a limit of the discrete time distribution, when the probabilities of gene loss and duplication tend to zero. Another noteworthy property of the discrete model is a very nice isomorphism (see Theorem 8) between the situation of genome collapse (i.e. when Prob(gene duplication) < Prob(gene loss)) and a genome explosion (when Prob(gene duplication) > Prob(gene loss)). This allows us to reduce the latter situation to the former. It appears that direct analysis of genome explosion is very difficult since the distribution looks more like a uniform distribution on an infinite set. The discrete model presented in this paper is in the same spirit as the model of DNA evolution presented in [Tiu99,Tiu00].

The paper is organized as follows. Section 2 contains a description of results for the continuous time model, together with asymptotic distributions for genome collapse (Theorem 2) and explosion (Theorem 3). Section 3 is devoted to discrete time. In particular we give a characterization of a generating function for the asymptotic distribution for collapse (Theorem 6). All longer proofs are moved to the Appendix.

2 Continuous time model

Before we start a description of the genome evolution, let us introduce all entities used in our model: genes, gene families, class of genes families and genomes. Genes are atomic units, i.e. we do not assume any internal structure of these objects. A genome is a finite set of all genes. A gene family in a genome is a set of genes of that genome which are paralogs. We group families according to their size. Classes are sets of gene families which have the same number of elements, i.e. S is a class i if every family in S has i elements. One gene duplication in a family belonging to class i results in relocation of this family from class i to class i + 1. Conversely, one gene removal relocates the corresponding family to class i − 1 or eliminates this family if i = 1. In this section, we consider time to be continuous.

We assume that the probability of a gene duplication to happen during time interval of length \(\Delta t\) is \(d \cdot \Delta t + o(\Delta t)\). Similarly, probability of gene removal in time interval \(\Delta t\) is \(r \cdot \Delta t + o(\Delta t)\). It is assumed that \(\lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} = 0\). Moreover, we assume that all elementary events (gene duplication and removal) are independent of each other.

\(^1\) Obviously, a class may include families that are completely unrelated biologically.
Let $C_i(t)$ be the number of $i$-element families in our model at the time $t$. It follows from the description of our model that we have the following equations which describe an increment $\Delta C_i(t) = C_i(t + \Delta t) - C_i(t)$ during time interval $\Delta t$:

$$\Delta C_1(t) = -(d + r)C_1(t)\Delta t + 2rC_2(t)\Delta t + o(\Delta t)$$

and

$$\Delta C_i(t) = d(i - 1)C_{i-1}(t)\Delta t - (d + r)iC_i(t)\Delta t + r(i + 1)C_{i+1}(t)\Delta t + o(\Delta t),$$

for $i \geq 2$.

Hence, dividing both sides of the above equations by $\Delta t$ and passing with $\Delta t$ to 0, we obtain the following infinite system of differential equations:

$$C'_i(t) = d(i - 1)C_{i-1}(t) - (d + r)iC_i(t) + r(i + 1)C_{i+1}(t), \quad (1)$$

where $i \geq 1$. The above equation for $i = 1$ reduces to $C'_1(t) = -(d + r)C_1(t) + 2rC_2(t)$, independently of the value of $C_0(t)$. We assume that the latter is just $C_0(t) = 0$. Let us also observe that $C_i(t) = 0$, for $i \geq 1$ and $t \in \mathbb{R}$ is a (trivial) solution of $(1)$.

**Theorem 1.** If $r > 0$ and $d > 0$, then for each non-zero and non-negative absolutely summable sequence $(C_i(0))_{i \geq 1}$, equation $(1)$ has a unique solution such that $C_i(t) > 0$ for all $t > 0$ and all positive integers $i$.

The remainder of this section is devoted to the asymptotic behavior of a solution of $(1)$, as $t \to \infty$. It turns out that the behavior of the system is quite different, depending on whether $r > d$, or $r < d$. In the former case all genes are eventually removed, while in the latter case we have an exponential explosion of the number of genes in the genome. We consider each case separately.

### 2.1 Collapse of the genome: $r > d$

The next result characterizes an asymptotic behavior of solutions of $(1)$ when $r > d$.

**Theorem 2.** Let $(C_i(t))_{i \geq 1}$ be non-negative and non-zero solution of $(1)$. If $r > d > 0$, then there exists a constant $c > 0$ such that for all $i \geq 1$,

$$\lim_{t \to \infty} e^{(r-d)t} \cdot C_i(t) = c \cdot \left(\frac{d}{r}\right)^i.$$

Hence for sufficiently large $t$ the number $C_i(t)$ of $i$-element gene families has the following asymptotics

$$C_i(t) \sim c \cdot e^{-(r-d)t} \cdot \left(\frac{d}{r}\right)^i,$$

for all $i \geq 1$. 
Corollary 1. If \( r > d \), then for sufficiently large \( t \), the distribution of gene families in the genome at time \( t \) is close to geometric, i.e. if \( X \) is a random variable which measures the size of a gene family of paralogous genes, then
\[
\Pr(X = i) = c \cdot \theta^i,
\]
where \( \theta = d/r \) and \( c \) is a normalizing constant.

2.2 Exponential explosion of the genome; \( d > r \)

We have the following result

**Theorem 3.** Let \( (C_i(t))_{i \geq 1} \) be non-negative and non-zero solution of (1). If \( d > r > 0 \), then there exist constants \( c_1 \) and \( c_2 \) such that for \( i \geq 1 \) and all \( t \in \mathbb{R} \)
\[
\lim_{t \to \infty} e^{(d-r)t} \cdot C_i(t) = c_1.
\]

In particular
\[
\lim_{t \to \infty} \frac{C_i(t)}{C_1(t)} = 1.
\]

Moreover
\[
e^{r-\theta t} \sum_{i \geq 1} i \cdot C_i(t) = c_2.
\]

The above result shows a peculiar behavior of the system during genome explosion: each \( C_i(t) \) vanishes exponentially fast (with the same speed, independent of \( i \)), while the total mass shows an exponential increase.

3 Discrete time model

In the previous section we introduce formal definition of all entities required in our discrete model. Now, we start with a Markov chain which is going to model the genome evolution with discrete time. States are infinite sequences \((s_i)_{i \geq 1}\) of non-negative integers. A state \((s_i)_{i \geq 1}\) represents a genome which for every \( i \geq 1 \) has \( s_i \) \( i \)-element gene families. As in the continuous time approach we introduce two parameters \( p_R > 0 \) and \( p_D > 0 \) which express the influence of removal and duplication, respectively. Let us assume that \( p_R + p_D \leq 1 \). Each gene of the genome is subject in one step of the Markov chain to:
- removal, with probability \( p_R \);
- duplication, with probability \( p_D \).

The gene remains unchanged with probability \( p_U = 1 - p_R - p_D \).

For \( i, j \geq 1 \) let \( q_{i,j} \) be the probability that a gene family of size \( i \) gets size \( j \), after every gene of the family has been subject to the above events. It can be shown that
\[
q_{i,j} = \sum_{k=0}^{\lfloor j/2 \rfloor} \binom{i}{j - 2k} p_U^{j-2k} p_D^k p_R^{i-j+k}.
\]
It follows that if \((s_i)_{i \geq 1}\) is a current state of the Markov chain, then the expected number of \(j\)-element gene families in the next state is

\[
E_j = \sum_{i=1}^{\infty} q_{i,j} \cdot s_i.
\]

Let \(\hat{Q} = (q_{i,j})_{i,j \geq 1}\). Hence, for \(n \geq 0\)

\[
(C_1^{(n)}, \ldots) = (1, 0, \ldots) \cdot \hat{Q}^n
\]

is an infinite sequence, where \(C_i^{(n)}\) is the expected number of \(i\)-element families of the Markov chain after \(n\) steps, assuming that the chain has started with the initial state consisting of just one gene. Again, we are interested in the observed distribution of gene families as we run the Markov chain a sufficient number of steps, and assuming that we started initially with a sufficient number of one element gene families.\(^2\)

It is useful to extend the concept of a state \((s_i)_{i \geq 1}\) by introducing the quantity \(s_0\) which represents the number of genes which have been removed. Thus an extended state is a sequence \((s_i)_{i \geq 0}\), where \((s_i)_{i \geq 1}\) is the ordinary state and \(s_0\) is the number of genes which have been removed. Then equation (2) makes sense for \(i \geq 1\) and \(j \geq 0\), where \(q_{i,0}\) is the probability that a family of size \(i\) disappears completely from the genome. We also set \(q_{0,0} = 1\) and \(q_{0,j} = 0\), for \(j > 0\), which again satisfies equation (2).

It follows that the extended matrix \(Q = (q_{i,j})_{i,j \geq 0}\) is a stochastic matrix, i.e. \(\sum_{j \geq 0} q_{i,j} = 1\), for all \(i \geq 0\). Moreover, if we define for \(n \geq 0\),

\[
(C_0^{(n)}, C_1^{(n)}, \ldots) = (0, 1, 0, 0, \ldots) Q^n,
\]

then, as before, for \(i \geq 1\), \(C_i^{(n)}\) is the expected number of \(i\)-element gene families after \(n\) steps, and \(C_0^{(n)}\) is the expected number of genes which have been removed after \(n\) steps. The above holds under proviso that the initial state of the Markov chain is \((0, 1, 0, 0, \ldots)\). It also follows that \((C_i^{(n)})_{i \geq 0}\) is a probability distribution, i.e. \(\sum_{i \geq 0} C_i^{(n)} = 1\).

It is easy to show the following conservation property. If \((C_0^{(n)}, \ldots) = (1, 0, \ldots) \hat{Q}^n\) and \((D_0^{(n)}, D_1^{(n)}, \ldots) = (0, 1, 0, 0, \ldots) Q^n\), then \(D_i^{(n)} = C_i^{(n)}\), for all \(i \geq 1\).

It follows from the above remarks that for every \(n \geq 0\), \(Q^n\) is a stochastic matrix, hence each of its rows is a probability distribution. A generating function for probability distribution \((p_i)_{i \geq 0}\) is the function \(f(x) = \sum_{x \geq 0} p_i x^i\). This function is defined at least on the closed interval \(-1 \leq x \leq 1\).

Let

\[
\varphi(x) = p_R + p_U x + p_D x^2.
\]

The next theorem is a standard result in the theory of branching processes (see [Fel61], chapter XII).

\(^2\) If \(p_R > p_D\), then we have to start simulations with sufficient supply of genes to ensure a sufficient number of steps before the genome vanishes.
**Theorem 4.** For all $n, i \geq 0$, the $i$-th row in $Q^n$ has generating function $[\varphi^n(x)]^i$, where $\varphi^n(x)$ is $n$-fold composition of $\varphi$ with itself.

It follows from the above result that

\[ \varphi^n(x) = \sum_{i=0}^{2^n} C_i^{(n)} x^i, \tag{4} \]

and $C_i^{(n)} = 0$, for $i > 2^n$. In particular we have $C_0^{(n)} = \varphi^n(0)$.

Instead of working directly with distribution $(p_R, p_U, p_D)$ it will be sometimes more convenient to represent this distribution as $(p, 1 - p - ap, ap)$, where $0 < p < 1$ is the probability of removal and $0 < a < (1 - p)/p$ is a constant which expresses the ratio $a = \frac{p_D}{p_R}$. We will use both notations interchangeably as it never should cause a confusion.

Let us briefly look at the asymptotic behavior of $C_i^{(n)}$, when $n$ tends to infinity. We have the following easy observation.

\[ \lim_{n \to \infty} C_0^{(n)} = \begin{cases} 1, & \text{if } p_R > p_D, \\ \frac{1}{a}, & \text{if } p_D > p_R. \end{cases} \tag{5} \]

Indeed, since $C_0^{(n)} = \varphi^n(0)$, it follows that the limit must be the smallest fixed point of $\varphi(x)$. Since $\varphi(x) - x = ap(x - 1)(x - a^{-1})$, (5) follows.

Hence we obtain

\[ \lim_{n \to \infty} \sum_{i \geq 1} C_i^{(n)} = \begin{cases} 0, & \text{if } p_R > p_D, \\ \frac{1-a}{a}, & \text{if } p_D > p_R. \end{cases} \tag{6} \]

It follows from the above that when $p_R > p_D$, then for every $i \geq 1$ we have $\lim_{n \to \infty} C_i^{(n)} = 0$.

Let us assume now that $p_D > p_R$. We show that

\[ \lim_{n \to \infty} C_1^{(n)} = 0 \tag{7} \]

Indeed, it follows from (4) that for all $n \geq 0$ we have

\[ C_1^{(n+1)} = p_U + C_1^{(n)} + 2p_DC_0^{(n)}C_1^{(n)}, \]

and $C_1^{(0)} = 0$. Thus, if we let

\[ \alpha_n = p_U + 2p_DC_0^{(n)}, \]

then

\[ C_1^{(n)} = \alpha_0 \alpha_1 \cdots \alpha_{n-1}. \]

Since $(C_1^{(n)})_{n \geq 1}$ is monotone increasing, it follows from (5) that

\[ \alpha_i < p_U + 2p_DC_0^{(n)} = p_U + 2p_R < 1. \]
This proves (7).

It follows from (7) and from Theorem 5 (see below) that for \( i \geq 1, \)
\[
\lim_{n \to \infty} C_i^{(n)} = 0.
\]

Thus we have argued that the sequence of \((C_1^{(n)}, C_2^{(n)}, \ldots)\) converges componentwise to \((0, 0, \ldots)\), when \( n \to \infty \). In order to capture the asymptotic behavior of the distribution of \((C_1^{(n)}, C_2^{(n)}, \ldots)\) we introduce the ratio
\[
\xi_n(i) = \frac{C_i^{(n)}}{C_1^{(n)}}.
\]

We are interested in
\[
\xi(i) = \lim_{n \to \infty} \xi_n(i).
\]

Observe that the information about function \( \xi \) completely determines the asymptotic distribution, provided \( \sum_{i \geq 1} \xi(i) < \infty \). In particular this distribution is geometric iff \( \xi(i) = \theta^{i-1} \), for some \( 0 < \theta < 1 \).

The next result shows that \( \xi \) is well defined.

**Theorem 5.** If \( p_R \neq p_D \), then for every \( i \geq 1 \),
\[
0 < \xi(i) < \infty.
\]

We call \( \xi \) the asymptotic ratio. Clearly it depends merely on the initial probabilities \( p_R \) and \( p_D \).

The main technical result used in the proof of Theorem 5 is the following property. Let
\[
\beta_n = \frac{p_D C_1^{(n)}}{\alpha_n}.
\]

Then it can be proved that for every \( i \geq 2 \) and for all \( n \geq 0 \) we have
\[
\xi_n(i) = \sum_{0 < j < i} \sum_{m < n} \xi_m(j) \cdot \xi_m(i - j) \cdot \beta_m.
\]  
(8)

The above formula can be used to design an algorithm for computing the functions \( \xi_n \). In contrast to Corollary 1 and to the impression one can get from Figure 1(b), it follows from (8) that the asymptotic distribution described by \( \xi \) is never geometric.

**Proposition 1.** For no \( p_R \) and \( p_D \) with \( p_R \neq p_D \) the asymptotic distribution described by \( \xi \) is geometric.

**Proof.** Using equation (8), it is easy to show that \( \xi(2)^2 < \xi(3) \). Hence the distribution cannot be geometric. \( \square \)

Next, we will study the shape of the asymptotic distribution for case \( p_R > p_D \) and \( p_D > p_R \), separately. It will turn out that the nature of the process is completely different in each of the cases, yet, as we will see, there is a nice isomorphism between distributions which arise in each of the cases.
3.1 Collapse of the genome: \( p_R > p_D \)

Let us look at the shape of the graph of function \( \xi \) depicted in Figure 1. We notice that for small values of \( p_R \) the graph fits very well geometric distribution (see Fig. 1(b)). However, by Proposition 1, for no value \( p_R > 0 \) it coincides with the geometric distribution. The difference can be better seen for larger values of \( p_R \) (see Fig. 1(a)).

Recall that it follows from Theorem 4 that \( \varphi^n(x) \) is the generating function for the row number 1 in \( \mathcal{Q}^n \). Hence it is a generating function for the probability distribution \( (C_i^{(n)})_{i \geq 0} \). It follows that the generating function for the distribution which is obtained from \( (C_i^{(n)})_{i \geq 0} \) by restricting \( i \) to range over positive integers is

\[
f_n(x) = \frac{\varphi^n(x) - \varphi^n(0)}{1 - \varphi^n(0)}.
\]

Hence the generating function for the asymptotic distribution is the limit of \( \{f_n\}_{n \geq 0} \). We have the following result.

**Theorem 6.** Let \( p_R > p_D \). The limit

\[
f(x) = \lim_{n \to \infty} f_n(x)
\]

exists for all \( x \in (\frac{1 - p_D}{p_R} a^{-1}, a^{-1}) \). The above convergence is almost uniform in complex plane \( C \) on the open disk \( |x| < a^{-1} \), i.e., it is uniform on every closed subdisk contained in \( \{x \in C : |x| < a^{-1} \} \). Hence \( f \) is an analytic function with the radius of convergence \( a^{-1} \).

Moreover \( f \) satisfies the following equation

\[
f(\varphi(x)) = \mu \cdot f(x) + 1 - \mu.
\]  

(9)

where \( \mu = \varphi'(1) = p_U + 2p_D \).

It follows from the theory of analytic functions\(^3\) that \( f \) is a unique analytic function which satisfies (9) and the two constraints: \( f(0) = 0 \) and \( f(1) = 1 \). In this sense the above theorem gives a complete characterization of the generating function for the distribution of interest.

Relationship between function \( \xi \) and the generating function \( f \) is given in the next result.

**Proposition 2.** Let \( p_R > p_D > 0 \). Then \( f'(0) \neq 0 \) and

\[
\sum_{i \geq 1} \xi(i) \cdot x^i = \frac{f(x)}{f'(0)}
\]

holds for all \( |x| < a^{-1} \). In particular we have

\[
\sum_{i \geq 1} \xi(i) = \frac{1}{f'(0)}
\]

\(^3\) Two analytic functions which coincide on set containing a point of convergence are equal.
Fig. 1. Plots of function $\xi$ for the case $p_R > p_D$. In (a) probability $p_R$ is relatively large. The continuous line in this plot shows the best fit of the geometric distribution to the observed data. In (b) $p_R$ is very small. The observed data fits very well in this case a geometric distribution.

We conclude this subsection with a result which explains the observed shape of $\xi$ during computer simulations for small values of $p_R$. In order to stress de-
dependence on \( p = p_R \) of functions under consideration we make this dependence explicit. Thus instead of writing \( \varphi(x) \) we shall write \( \varphi(p, x) \). The same applies to \( f \) and \( f_n \). The main result of this section is the following theorem.

**Theorem 7.** Let \( a < 1 \) be fixed. We have

\[
\lim_{p \to 0^+} f(p, z) = \frac{(1 - a)z}{1 - a z},
\]

for all \( z \) from open disk \( |z| < a^{-1} \) in complex numbers.

Hence, by Proposition 2, for all \( i \geq 1 \), we have

\[
\lim_{p \to 0^+} \xi(p, i) = a^{i-1}.
\]

### 3.2 Exponential explosion of the genome: \( p_D > p_R \)

In this subsection we discuss the case \( p_D > p_R \) and we shall explain the observed shape of \( \xi \). Let us first look at computer simulations depicted in Figure 2. The striking feature is a completely different shape of the graph of \( \xi \) with respect to the previous case \( p_R > p_D \). It looks now more like a uniform distribution for small values of \( p_D \), but the domain is infinite. Corollary 2 explains exactly this behaviour.

Fortunately we do not have to repeat all the work done in the previous subsection since there a nice isomorphism which enables us to reduce the present situation to the former case. This isomorphism is given in the next result.

**Theorem 8.** Let \( p, q > 0 \) be reals such that \( p + q < 1 \). Let \( Q_{p,q} \) be the extended stochastic matrix for the initial probabilities \( p_R = p \), \( p_D = q \). Then

\[
Q_{p,q} = \Delta_a \cdot Q_{p,q} \cdot \Delta_a^{-1},
\]

where \( a = q/p \) and \( \Delta_a \) is a diagonal matrix with \( \delta_{i,i} = a^i \), for \( i \geq 0 \).

**Proof.** Since the proof is very short we give it here. Let

\[
\varphi_{p,q}(x) = p + (1 - p - q)x + qx^2
\]

and let \( a = q/p \). It is easy to check that we have

\[
\varphi_{p,q}(x) = a \cdot \varphi_{p,q}(a^{-1} \cdot x).
\]

By Theorem 4 the \( i \)-th row in \( Q_{p,q} \) has the generating function

\[
[\varphi_{p,q}(x)]^i = a^i \cdot [\varphi_{p,q}(a^{-1} \cdot x)]^i.
\]

Hence

\[
Q_{p,q} = \Delta_a \cdot Q_{p,q} \cdot \Delta_a^{-1},
\]

as required. \( \Box \)
Let $\xi_{p,q}$ be the asymptotic ratio for the initial probabilities $p_R = p, p_D = q$. It follows immediately from Theorem 8 that for all $i \geq 1$,

$$\xi_{p,q}(i) = a^{i-1} \xi_{p,q}(i),$$

(10)
where $a = q/p$. Hence, if we apply Theorem 7 to (10), we easily obtain the following result:

**Corollary 2.** Let $a > 1$ be fixed and let for all $i \geq 1$, $\xi(p, i)$ be the asymptotic ratio for the initial probabilities $p_R = p$ and $p_D = a \cdot p$. Then

$$\lim_{p \to 0^+} \xi(p, i) = 1.$$

4 Conclusions

Here we presented a complete mathematical description of the frequency distribution of paralog families in genomes for a simple model of evolution. Our model includes two types of events: gene duplication and removal. Moreover, we analysed and compared two approaches to the time in the model: continuous and discrete.

This work is our initial step toward modeling of genome evolution using most important evolutionary processes/events. Actually, we should extend our model by some concept of e.g. gene mutation. Nevertheless, even this simple and crude model seems to reveal several interesting aspects of genome evolution.

Acknowledgements

JT thanks Piotr Słonimski for many discussions on various Markov chain models for genome evolution and Stanisław Kwapien and Krzysztof Oleśkiwicz for fruitful discussions on mathematical properties of the model.

References


5 Appendix: proofs of some statements

5.1 Proof of Theorem 1

We change variables in equation (1) to obtain a system which will be easier in analysis. Let

\[ x_i(t) = e^{(r - d)y_i} C_i(t). \] (11)

Then, we obtain

\[ x_i'(t) = (r - d)x_i(t) + e^{(r - d)yi} C_i'(t) \]
\[ = (r - d)x_i(t) + e^{(r - d)yi} [d(i - 1)C_{i-1}(t) - (d + r)iC_i(t) + r(i + 1)C_{i+1}(t)] \]

and therefore

\[ x_i'(t) = i dx_{i-1}(t) - [(i + 1)d + (i - 1)r]x_i(t) + ir x_{i+1}(t), \quad i = 1, 2, \ldots , \] (12)
where we formally assume that $x_0(t) = 0$ for all $t \geq 0$.

Let $l^1$ denote the space of absolutely summable sequences. We check that system (12) generates a Markov semigroup on $l^1$. Recall that a linear mapping $P : l^1 \to l^1$ is called a Markov or stochastic operator if $P(D) \subset D$, where

$$D = \{ x \in l^1 : x_n \geq 0 \text{ for all } n \geq 1 \text{ and } \sum_{n=1}^{\infty} x_n = 1 \}.$$ 

A family $\{ P(t) \}_{t \geq 0}$ of Markov operators which satisfies conditions:

(a) $P(0) = \text{Id}$,
(b) $P(t+s) = P(t)P(s)$ for $s, t \geq 0$,
(c) for each $x \in l^1$ the function $t \mapsto P(t)x$ is continuous with respect to the $l^1$ norm

is called a Markov or stochastic semigroup.

We recall that an abstract equation $x'(t) = Ax(t)$, $x(0) = x_0$ generates a Markov semigroup $\{ P(t) \}_{t \geq 0}$ if $P(t)x_0 = x(t)$.

Let $a_{i,j} = -(i+1)d - (i-1)r$, $a_{i+1,i} = (i+1)d$ for $i \geq 1$, $a_{i-1,i} = (i-1)r$ for $i \geq 2$, and $a_{i,j} = 0$ in other cases. System (12) can be written in the following way:

$$x'(t) = Ax(t), \quad x(0) = x_0,$$  

where $A = (a_{i,j})_{i,j \geq 1}$. Let $A$ be the $A^\ast = (A^\ast_{i,j})_{i,j \geq 1}$ where $a^\ast_{i,j} = a_{j,i}$ for $i, j \geq 1$.

The matrix $A = (a_{i,j})_{i,j \geq 1}$ is a Kolmogorov matrix, i.e. it satisfies conditions: $a_{i,j} \geq 0$ for $i \neq j$ and $\sum_{j=1}^{\infty} a_{i,j} = 0$ for all $j \geq 1$. We need the following result

**Theorem 9.** Let $A$ be a Kolmogorov matrix. Then system (13) generates a Markov semigroup on $l^1$ if and only if for some positive $\theta$ there is no non-zero solution of the equation $A^\ast x = \theta x$, where $x \in l^\infty$.

Recall that $l^\infty$ is the space of bounded sequences. The proof of Theorem 9 can be found in [Nor97]. Consider our case. If $A^\ast x = \theta x$, then

$$(i-1)rx_{i-1} - [(i+1)d + (i-1)r]x_i + (i+1)dx_{i+1} = \theta x_i, \quad i = 1, 2, \ldots, \quad (14)$$

where $x_0 = 0$. Then

$$x_{i+1} = \left(1 + \frac{(i-1)r + \theta}{(i+1)d}\right)x_i - \frac{(i-1)r}{(i+1)d}x_{i-1}. \quad (15)$$

We can assume that $x_1 > 0$. Using simple induction arguments one can check that the sequence $(x_i)$ is increasing and

$$x_{i+1} = \left(1 + \frac{\theta}{(i+1)d}\right)x_i. \quad (16)$$

Thus

$$x_n > x_1 \prod_{i=2}^{n} \left(1 + \frac{\theta}{id}\right) \quad (17)$$
and therefore
\[
\lim_{n \to \infty} x_n \geq x_1 \prod_{i=2}^{\infty} \left(1 + \frac{\theta}{id}\right) = \infty.
\]  
Consequently, \( x \notin l^\infty \). Thus we have proved the following result:

**Corollary 3.** For each \( x \in l^1 \) and \( t \geq 0 \) we denote by \( P(t)x \) the solution \( x(t) \) of equation (13) with the initial condition \( x(0) = x \). Then \( \{P(t)\}_{t \geq 0} \) is a Markov semigroup on \( l^1 \).

Let \( x \in l^1 \) be a non-zero and non-negative sequence. Since \( a_{i,i+1} > 0 \) and \( a_{i+1,i} > 0 \) for all \( i \geq 1 \) we have \( (P(t)x)_i > 0 \) for all \( t > 0 \), \( i = 1, 2, \ldots \). Now, we can return to equation (1). Let \( \{C_i\} \) be a non-zero and non-negative sequence such that \( \sum_{i=1}^{\infty} iC_i < \infty \). Let \( x = (x_i) \) be a sequence defined by \( x_i = iC_i \). Then \( x \in l^1 \). From Corollary 3 it follows that equation 13 has a unique solution for every \( x_0 \in l^1 \) which is given by the formula \( x(t) = P(t)x_0 \). The unique solution of equation (1) such that \( C_i(0) = C_i \) is given by the formula
\[
C_i(t) = e^{(d-r)t}i^{-1}(P(t)x)_i,
\]
where \( x_i = iC_i \). This completes the proof of Theorem 1.

### 5.2 Proof of Theorem 2

Since the semigroup \( \{P(t)\}_{t \geq 0} \) (see Appendix 5.1) is defined on the space \( l^1 \), it is an integral or kernel semigroup. Moreover, \( \{P(t)\}_{t \geq 0} \) is an irreducible semigroup because \( (P(t)x)_i > 0 \) for all \( t > 0 \) if \( x \in l^1 \) is a non-zero and non-negative sequence. If \( \{P(t)\}_{t \geq 0} \) is an integral and irreducible Markov semigroup and there exists \( x^* \in D \) such that \( P(t)x^* = x^* \) for all \( t > 0 \), then \( P(t)x \to x^* \) for every \( x \in D \) as \( t \to \infty \) (see e.g. [Rud95]). If \( r > d > 0 \), then system (12) has a stationary solution \( x_i^* = ei(d/r)^i \) for \( i = 1, 2, \ldots \). We can choose the constant \( c \) such that \( x^* \in D \). Then \( P(t)x \to x^* \) for every \( x \in D \) as \( t \to \infty \). Therefore
\[
\lim_{t \to \infty} e^{(r-d)t}C_i(t) = \frac{x_i^*}{1} = c \cdot \left(\frac{d}{r}\right)^i.
\]
This completes the proof of Theorem 2.

### 5.3 Proof of Theorem 3

First observe that since \( \{P(t)\}_{t \geq 0} \) (see Appendix 5.1) is a Markov semigroup we have \( \sum_{i=1}^{\infty} x_i(t) = \text{const} \) for any solution of system (12). From equation (11) it follows that
\[
e^{(r-d)t} \sum_{i=1}^{\infty} iC_i(t) = \text{const}.
\]
In order to prove the main part of Theorem 3 we substitute \( z_i(t) = e^{(d-r)t} \left(\frac{d}{r}\right)^i iC_i(t) \).

Then
\[
z_i' = irz_{i-1} - [(i+1)r + (i-1)d]z_i + idz_{i+1}.
\]
Equation (20) has a form of equation (12) with $r$ and $d$ interchanged. This implies that the Markov semigroup generated by (20) is asymptotically stable and the sequence $z^* = (z_t^*)_{t \geq 1}$ with $z_t^* = i \left( \frac{r}{d} \right)^t$ is stationary. It means that if $z(0) \in l^1$ then there is a constant $c_1$ such that and $\lim_{t \to \infty} z(t) = c_1i \left( \frac{r}{d} \right)^t$. Therefore $\lim_{t \to \infty} e^{(d-r)t}|C_i(t)| = c_1$.

### 5.4 Proof of Theorem 5

We start by proving (8). It follows immediately from (4) that

$$C_i^{(n+1)} = p_U C_i^{(n)} + p_D \sum_{j=0}^{i} C_j^{(n)} C_{i-j}^{(n)}, \text{ for } i \geq 1. \tag{21}$$

We can rewrite (21) as follows for $i \geq 1$

$$C_i^{(n+1)} = \alpha_n C_i^{(n)} + \sum_{j=0}^{i-1} p_D C_j^{(n)} C_{i-j}^{(n)}.$$

Hence, for $i \geq 2$ we have

$$C_i^{(n)} = \sum_{j=1}^{i-1} \sum_{m=0}^{n-1} \sum_{j=1}^{n-1} p_D \alpha_{m+1} \cdots \alpha_{n-1} C_j^{(m)} C_{i-j}^{(m)}$$

$$= \sum_{j=1}^{i-1} \sum_{m=0}^{n-1} p_D \alpha_{m+1} \cdots \alpha_{n-1} \frac{C_j^{(m)} C_{i-j}^{(m)}}{\alpha_m}$$

$$= \sum_{j=1}^{i-1} \sum_{m=0}^{n-1} p_D C_j^{(m)} C_{i-j}^{(m)}$$

$$= \sum_{j=1}^{i-1} \sum_{m=0}^{n-1} p_D \frac{C_j^{(m)}}{\beta_m \alpha_m} C_{i-j}^{(m)}$$

$$= C_i^{(n)} \sum_{j=1}^{i-1} \sum_{m=0}^{n-1} p_D C_j^{(m)} C_{i-j}^{(m)}$$

$$= C_i^{(n)} \sum_{j=1}^{i-1} \sum_{m=0}^{n-1} \xi_m(j) \xi_m(i-j) \beta_m.$$

This yields (8).

It follows from (8) that for every $i \geq 2$ and for all $n \geq 0$,

$$\xi_n(i) \leq \xi_{n+1}(i).$$

Hence it suffices to show that for every $i \geq 2$, the sequence $(\xi_n(i))_{n=0,1,2,...}$ is bounded from above by a certain constant $B_i$. We prove this statement by
induction on $i$. In order to prove the statement for $i = 2$ let
\[
\gamma = \begin{cases} 
p_U + 2p_D & \text{if } p_D \leq p_R, \\
p_U + 2p_R & \text{if } p_R < p_D.
\end{cases}
\]
Clearly we have $\gamma < 1$ iff $p_D \neq p_R$. Moreover, it follows from (5) that $\alpha_n \to \gamma$. Since $\alpha_n \leq \alpha_{n+1}$, we have $\alpha_n \leq \gamma$, for all $n$. Hence
\[
\xi(2) = \sum_{m < \infty} \beta_m \\
\leq \frac{p_D}{\alpha_0} \sum_{m < \infty} \gamma^m \\
= \frac{p_D}{p_U(1 - \gamma)}.
\]

Assume the statement for all $i' < i$. Without loss of generality we may assume that $B_\rho \leq B_{i'}$ holds for all $i' < i^\theta < i$. By (8) we have
\[
\xi_n(i) = \sum_{j=1}^{i-1} \sum_{m=0}^{n-1} \xi_{m}(j)\xi_{m}(i-j)\beta_m \\
\leq (i - 1)(B_{i-1})^2 \sum_{m=0}^{n-1} \beta_m \\
\leq (i - 1)(B_{i-1})^2 \xi(2).
\]
Hence, we have $\xi(i) < \infty$ for all $i \geq 1$.

Since $\beta_k > \beta_{k+1}$ for all $k \in \mathbb{N}$, thus we have
\[
\xi(i) \geq \xi(i) \geq (\beta_i)^{i-1} > 0,
\]
for $i \geq 1$.

This completes the proof of Theorem 5.

5.5 Proof of Theorem 6

We first need the following two lemmas, where the former is a well known result in dynamical systems [Dev85]. For the sake of completeness we give the proof of this result too.

**Lemma 1.** Let $\varphi : (a, \beta) \to \mathbb{R}$ be a $C^2$-function and $x_0 \in (a, \beta)$. Assume that $\varphi(x_0) = x_0$, $|\varphi'(x_0)| \equiv \mu$, $0 \leq \mu < 1$, $|\varphi(x) - x_0| < |x - x_0|$ and $\varphi(x) \neq x_0$ for $x \neq x_0$. Then for every $x \in (a, \beta)$ there exists the limit
\[
\lim_{n \to \infty} \frac{|\varphi^n(x) - x_0|}{\mu^n} = g(x)
\]
and $0 < g(x) < \infty$.  


Proof. Without loss of generality we can assume that \( x_0 = 0 \) and \( \varphi(x) > 0 \) for \( x \neq 0 \). From the assumption \( |\varphi(x)| < x \) it follows immediately that \( \lim_{n \to \infty} \varphi^n(x) = 0 \) for every \( x \in (\alpha, \beta) \) and therefore it is sufficient to check (22) for sufficiently small \( x \). We consider the case \( x > 0 \) (the case \( x < 0 \) is analogous). We have

\[
\log \frac{\varphi^n(x)}{\mu^n x} = \log \frac{\varphi^n(x)}{\mu^n \varphi^{n-1}(x)} + \cdots + \log \frac{\varphi(x)}{\mu x}.
\]

Since \( \varphi(x) = \mu x + \frac{1}{2} \varphi''(0)x^2 + \alpha x^2 \) we have \( \varphi(x)/x = 1 + cx + \alpha(x) \), where \( c = \frac{1}{2} \varphi''(0)/\mu \). Therefore \( \log \frac{\varphi(x)}{\mu x} = cx + \alpha(x) \). Observe that for sufficiently small \( x \) we have

\[
(c - 1) \left( \frac{\varphi}{x} \right)^k \leq (c - 1) \varphi^k(x) \leq \log \frac{\varphi^{k+1}(x)}{\mu \varphi^k(x)} \leq (c + 1) \varphi^k(x) \leq (c + 1) \left( \frac{1 + \mu x}{x} \right)^k.
\]

and this implies that the series \( \sum_{n=1}^{\infty} \log \frac{\varphi^n(x)}{\mu^n x} \) is convergent. Thus the limit

\[
\lim_{n \to \infty} \log \frac{\varphi^n(x)}{\mu^n x}
\]

exists. It follows that for sufficiently small \( x \) the limit (22) exists and it is a positive real number. This completes the proof.

\[\square\]

Lemma 2. If \( p_D < p_R \), then for every \( x \) satisfying \( -(p_U + p_R)/p_D < x < p_R/p_D \) we have \( \lim_{n \to \infty} \varphi^n(x) = 1 \).

Proof. We have the following identities.

\[
\varphi(x) - p_R/p_D = p_D(x - p_R/p_D)(x + (p_U + p_R)/p_D) \tag{23}
\]

\[
\varphi(x) - 1 = p_D(x - 1)(x + (p_U + p_D)/p_D) \tag{24}
\]

\[
\varphi(x) - x = p_D(x - 1)(x - p_R/p_D). \tag{25}
\]

Hence if \( -(p_U + p_D)/p_D < x < 1 \), then by (24) and (25) we have \( x < \varphi(x) < 1 \). In this case \( \varphi^n(x) \) has a limit which is the closest fixed point of \( \varphi, i.e. 1 \).

If \( 1 < x < p_R/p_D \) then again by (24) and (25) we have \( 1 < \varphi(x) < x \) and by the same argument \( \lim_{n \to \infty} \varphi^n(x) = 1 \).

Finally, if \( -(p_U + p_R)/p_D < x < -(p_U + p_D)/p_D \), then by (23) and (24) we have \( 1 < \varphi(x) < p_R/p_D \). Thus again \( \lim_{n \to \infty} \varphi^n(x) = 1 \). This completes the proof of Lemma 2.

\[\square\]

Now we start proving Theorem 6.

We have

\[
\frac{\varphi^n(x) - \varphi^n(0)}{1 - \varphi^n(0)} = \frac{(\varphi^n(x) - 1)/\mu^n + (1 - \varphi^n(0))/\mu^n}{(1 - \varphi^n(0))/\mu^n}
\]

Hence by Lemma 1 and Lemma 2 the above limit exists and is finite whenever \( x \) satisfies \( -(p_U + p_R)/p_D < x < p_R/p_D \). On the other hand it is easy to show that the limit is \( +\infty \) for all other \( x \)’s.
To show that the convergence is almost uniform in the open disk $|z| < a^{-1}$ in complex numbers take any $0 < r < a^{-1}$. We have for all $|z| < r$ the following inequalities

$$|f_n(z)| \leq f_n(|z|) \leq f_n(r) \quad (26)$$

The first inequality follows from non-negative coefficients of $f_n$ and the second from its monotonicity. Since the sequence $f_n(r)$ is converging, the sequence of functions $f_n(z)$ is uniformly bounded on the disk $|z| < r$. According to the Vitali theorem a uniformly bounded sequence of analytic functions which is converging in some set with a limit point in this disk, converges uniformly to an analytic function $f(z)$ on each compact subsets of this disk. This implies that the sequence $f_n(z)$ covers uniformly to an analytic function $f(z)$ on each compact subsets of the disk $|z| < a^{-1}$.

Now we prove (9).

$$\frac{\varphi^n(x) - \varphi^n(0)}{1 - \varphi^n(0)} \leq \frac{\varphi^{n-1}(\varphi(x)) - \varphi^{n-1}(0) + \varphi^{n-1}(0) - \varphi^n(0)}{1 - \varphi^n(0)}$$

$$= \frac{\varphi^{n-1}(\varphi(x)) - \varphi^{n-1}(0)}{1 - \varphi^{n-1}(0)} \frac{1 - \varphi^{n-1}(0)}{1 - \varphi^n(0)} + \frac{\varphi^{n-1}(0) - \varphi^n(0)}{1 - \varphi^n(0)}$$

It follows from Lemma 1 that

$$\lim_{n \to \infty} \frac{1 - \varphi^{n-1}(0)}{1 - \varphi^n(0)} = \frac{1}{\mu}$$

and

$$\lim_{n \to \infty} \frac{\varphi^{n-1}(0) - \varphi^n(0)}{1 - \varphi^n(0)} = 1 - \frac{1}{\mu}$$

Hence, by taking limit on both sides we obtain the equality

$$f(x) = f(\varphi(x))/\mu + 1 - 1/\mu,$$

which is equivalent to (9).

### 5.6 Proof of Proposition 2

It is obvious that for a fixed $n$ the sequence $(\xi_n(i))_{i \geq 1}$ uniquely determines (up to a constant factor) the sequence $(C_{i}^{(n)})_{i \geq 1}$. Thus if we normalize these sequences we obtain the same probability distribution.

It follows from Theorem 6 that the limit of the generating functions $f_n(x)$ for the distributions determined by the sequences $(C_{i}^{(n)})_{i \geq 1}$ exists. We have the same for the sequence $(\xi_n(i))_{i \geq 1}$. Moreover, we have to obtain the same limit. Hence we have

$$f(x) = \sum_{i \geq 1} \xi(i)x^i / \sum_{i \geq 1} \xi(i).$$

In particular we also have

$$f'(0) = 1 / \sum_{i \geq 1} \xi(i).$$

This completes the proof of Proposition 2.
5.7 Proof of Theorem 7

We prove the first part of the Theorem. The second part follows easily from the first part. First we show that the limit exists.

Lemma 3. Let $a = p_0/p_R$ and let $0 < p < 1/(3 - a)$. For all $n \geq 1$ we have

$$\frac{\partial f_n}{\partial p}(p, x) \leq 0, \text{ if } 0 \leq x < a^{-1},$$

and

$$\frac{\partial f_n}{\partial p}(p, x) \geq 0, \text{ if } -a^{-1} < x \leq 0.$$

Proof. Consider the polynomial

$$\psi(p, x) = 1 - \varphi(p, 1 - x).$$

Clearly we have

$$\psi^n(p, x) = 1 - \varphi^n(p, 1 - x).$$

Thus

$$f_n(p, x) = 1 - \frac{\psi^n(p, 1 - x)}{\psi^n(p, 1)}.$$

Let

$$F_n(p, x) = \frac{\partial \psi^n}{\partial p}(p, x) \cdot \frac{1}{\psi^n(p, x)}.$$ 

We prove the following property, for all $n \geq 1$, $|x| \leq a^{-1}$, and $0 < p < 1/(3 - a)$,

$$\frac{\partial F_n}{\partial x}(p, x) \leq 0 \quad (27)$$

Let $\mu = 1 - p + ap$. We have $\psi(p, x) = px - apx^2$. Thus $\psi^{n+1} = \mu \psi^n - ap|\psi^n|^2$.

Hence we have

$$\frac{\partial \psi^{n+1}}{\partial p}(p, x) = [(\mu - 2ap\psi^n(p, x)) \cdot F_n(p, x) - 1 + a + 2ap\psi^n(p, x)] \cdot \psi^n(p, x).$$

Hence

$$F_{n+1}(p, x) = \frac{\mu - 2ap \cdot \psi^n(p, x)}{\mu - a p \cdot \psi^n(p, x)} \cdot F_n(p, x) + \frac{1 + a - a \cdot \psi^n(p, x)}{\mu - a p \cdot \psi^n(p, x)}.$$ 

Since for all $p > 0$,

$$\frac{\partial}{\partial x}\left(\frac{\mu - 2apx}{\mu - apx}\right) < 0$$

and since $\psi(p, x)$ is monotone increasing with respect to $x$ for $|x| < a^{-1}$ and $0 < p \leq 1/(3 - a)$, we have for all $n \geq 1$

$$\frac{\partial \psi^n}{\partial x}(p, x) \geq 0.$$
It follows that
\[ \frac{\partial}{\partial x} \left( \frac{\mu - 2ap \cdot \psi^n(p, x)}{\mu - ap \cdot \psi^n(p, x)} \right) \leq 0. \]

In a similar way we show that
\[ \frac{\partial}{\partial x} \left( \frac{-1 + a - a \cdot \psi^n(p, x)}{\mu - ap \cdot \psi^n(p, x)} \right) \leq 0. \]

It is easy to show that
\[ F_{n+1} = A_n \cdot A_{n-1} \cdots A_1 \cdot B_0 + A_n \cdot A_{n-1} \cdots A_2 \cdot B_1 + \cdots + A_n \cdot B_{n-1} + B_n, \]
where
\[ A_i = \frac{\mu - 2ap \cdot \psi^n(p, x)}{\mu - ap \cdot \psi^n(p, x)} \]
and
\[ B_i = \frac{-1 + a - a \cdot \psi^n(p, x)}{\mu - ap \cdot \psi^n(p, x)}. \]

Since each $A_i$ and $B_i$ is non-negative and both have non-positive derivatives with respect to $x$, it follows that for all $n \geq 1$, for all $|x| \leq a^{-1}$, and $0 < p < 1/(3-a)$,
\[ \frac{\partial F_{n+1}}{\partial x}(p, x) \leq 0. \]

This proves (27).

Now we complete the proof of the lemma.
\[ \frac{\partial f_n}{\partial p}(p, x) = -\frac{\frac{\partial \psi^n}{\partial p}(p, 1 - x) \cdot \psi^n(1) - \frac{\partial \psi^n}{\partial p}(p, 1) \cdot \psi^n(p, 1 - x)}{[\psi^n(1)]^2}. \]

It follows from (27) that for all $0 < p < 1/(3-a)$ we have $F_n(p, 1 - x) \geq F_n(p, 1)$, whenever $0 \leq x < a^{-1}$; and $F_n(p, 1 - x) \leq F_n(p, 1)$, whenever $-a^{-1} < x \leq 0$.

This completes the proof of Lemma 3. \( \square \)

It follows from Lemma 3 that the limit in Theorem 7 exists for $x \in (-a^{-1}, a^{-1})$.

Using once more inequalities (26) and the Vitali theorem we prove that if $p \to 0$ then $f(p, z)$ converges uniformly to an analytic function $g(z)$ on each compact subset of the disk $|z| < a^{-1}$. Denote this limit by $g(z)$. Since $f(p, z)$ are analytic functions we have also
\[ \lim_{p \to 0} \frac{\partial f}{\partial z}(p, z) = g'(z) \]
uniformly on compact sets in $|z| < a^{-1}$.

Consider equation (9), subtract from both sides $f(x)$ and divide both sides by $\phi(x) - x$. Writing explicitly the parameter $p$ we obtain the equation
\[ \frac{f(p, \phi(p, x)) - f(p, x)}{\phi(p, x) - x} = \frac{(1 - \mu) \cdot (1 - f(p, x))}{\phi(p, x) - x}. \]
Transforming the right hand side of the above equation leads to the following equation:

\[
\frac{f(p, \varphi(p, x)) - f(p, x)}{\varphi(p, x) - x} = \frac{(1 - a) \cdot (1 - f(p, x))}{(1 - x) \cdot (1 - ax)}. \quad (29)
\]

We claim that

\[
\lim_{p \to 0^+} \frac{f(p, \varphi(p, x)) - f(p, x)}{\varphi(p, x) - x} = g'(x). \quad (30)
\]

Indeed, by the mean value theorem we know that there exists \( x < \rho_p < \varphi(p, x) \) such that

\[
\frac{f(p, \varphi(p, x)) - f(p, x)}{\varphi(p, x) - x} = f'(p, \rho_p).
\]

Since \( \lim_{p \to 0} \varphi(p, x) = x \), it follows that \( \lim_{p \to 0} \rho_p = x \). From (28) it follows that \( \lim_{p \to 0} f'(p, \rho_p) = f'(0, x) = g'(x) \). Thus we obtain from (29) and (30) the following differential equation

\[
g'(x) = \frac{(1 - a) \cdot (1 - g(x))}{(1 - x) \cdot (1 - ax)}. \quad (31)
\]

This is an equation with split variables and therefore it has a unique solution which satisfies the boundary condition \( g(1) = 1 \). It is easy to check that the function \( g(x) = (1 - a)x / (1 - ax) \) satisfies (31). This completes the proof of Theorem 7.