Superstatistical and DNA sequence coding of the human genome

M. O. Costa*
Departamento de Física, Universidade Federal do Rio Grande do Norte, Natal - RN, 59072-970, Brasil

R. Silva† and D. H. A. L. Anselmo‡
Departamento de Física, Universidade Federal do Rio Grande do Norte, Natal - RN, 59072-970, Brasil and Programa de Pós-Graduação em Física, Universidade do Estado do Rio Grande do Norte, Mossoró - Rio Grande do Norte, 59610-210, Brasil

(Received 6 July 2022; accepted 16 November 2022; published 12 December 2022)

In this work, by considering superstatistics we investigate the short-range correlations (SRCs) and the fluctuations in the distribution of lengths of strings of nucleotides. To this end, a stochastic model provides the distributions of the size of the exons based on the q-Gamma and inverse q-Gamma distributions. Specifically, we define a time series for exon sizes to investigate the SRC and the fluctuations through the superstatistics distributions. To test the model's viability, we use the Project Ensembl database of genes to extract the time evolution of exon sizes, calculated in terms of the number of base pairs (bp) in these biological databases. Our findings show that, depending on the chromosome, both distributions are suitable for describing the length distribution of human DNA for lengths greater than 10 bp. In addition, we used Bayesian statistics to perform a selection model approach, which revealed weak evidence for the inverse q-Gamma distribution for a considerable number of chromosomes.

DOI: 10.1103/PhysRevE.106.064407

I. INTRODUCTION

The growing amount of genomic data comes from various DNA sequencing projects. In this regard, statistical physics emerges as a tool that allows us to analyze the complexity of the DNA structure. Various approaches were used, e.g., random walk [1–3], Ising model [4], and wavelet transforms [5,6], to name a few. As a result, DNA is associated with an aggregation phenomenon, resulting in a fractal cluster with power-law correlations in space or time. Furthermore, long-range [7] and short-range [8] correlations have been widely discussed, especially in the context of the length distribution of coding and noncoding sequences from many living organisms, including human DNA [8–12].

The exon length distribution analysis has received attention since the 1980s. Hawkins [13] surveyed a broad class of distinct phyla regarding the lengths of introns and exons. Soon after, Höglund *et al.* [14] reported a study of size distributions of 411 mammalian exons which were selected randomly. Long and Deutsch [15] analyzed the distribution of intron-exon structures of eukaryotic genes, and their findings suggested a causal relationship between intron content and genome complexity. Melodelima *et al.* [16] proposed a sum of geometric distributions with equal or different parameters to describe the length distribution of exons and introns of the human genome. The authors have also used hidden Markov models to process the data in order to identify genes. Sakharkar *et al.* [17] reported a comparative analysis of human and mouse genome lengths regarding the differences and similarities of exon-intron distributions. To reconcile various hypotheses on the segmentation of eukaryotic genes, Gudlaugsdottir and colleagues [18] presented mixed statistics of exon lengths' distributions, namely, a sum of pure exponential and Weibull distributions. Finally, Li [19] demonstrated that Menzerath's law [20] could be applied to genes: the more exons in a gene, the shorter the average exon size.

Regarding the dynamics of exon lengths, Wang and Stein [21] proposed a stochastic model about the splitting of exons by introns, and their model predicts that the chance for an exon to obtain an intron is proportional to l_e^3, where l_e is the exon's length. Also, Martignetti and Caselle [22] investigated the (power-law-like) length distribution of the 5' untranslated sections' exons, a specific subclass of DNA sequences, through a Markov chain modeling. Polychronopoulos and colleagues [23] have also obtained a power-law behavior for the size distribution of noncoding elements for phylogenetically distinct datasets.

Based on generalized entropies, some statistical approaches investigate several complex systems [24,25]. From a molecular biology standpoint, the Tsallis and Kaniadakis formalism was used to describe coding and noncoding sections of human DNA [26–31], and even plant DNA [32]. Another statistical approach providing a more general class of distributions and containing the Tsallis one, in particular, is a so-called superstatistics [33]. The core of this formalism is to decompose the dynamics of the system into different scales so that its statistical properties are given by a superposition of statistics that have a Boltzmann factor $e^{-\beta E}$, weighted with a probability density $P(\beta)$ for which it has a fluctuating

*marconeoliveiraa@gmail.com
†raimundosilva@fisica.ufrn.br
‡doryh@fisica.ufrn.br

intensive parameter β. Also, one assumes local equilibrium on each of these scales, which is achieved at distinct β values. Examples of such parameters are dissipation energy and inverse temperature [33,34].

The superstatistics formalism successfully described several systems, e.g., econophysics [35,36], geophysics [37,38], turbulence [39–41], plasmas [42–44], and ultracold gases [45,46] to high-energy scattering processes [47,48], spin systems [49,50], cosmology [51,52] and stellar systems [53]. In biophysics, there are already some special applications, such as a superstatistical model (and its corresponding DNA generation algorithm), which emulates the rules which dictate the (empirical) nucleotide arrangement properties of some DNA sequences [54,55]. Also, more recently, Itto and Beck [56] reported an analysis of DNA-binding proteins that exhibit highly heterogeneous diffusion processes in bacteria [57]. The fractional Brownian motion is used as a possible local model, and this model is, in turn, based on superstatistics with two variables.

In contrast to using the empirical cumulative distribution function (ECDF) to calculate exon length distributions [29–31] and the various approaches discussed above, in this work the main goal is to investigate the short-range correlations (SRCs) and the fluctuations in the distribution of the lengths of strings of nucleotides through the superstatistics framework. As a result, we developed a stochastic model to provide the distributions of the size of the exons. Specifically, the stochastic model leads to the q-Gamma and inverse q-Gamma distributions. Moreover, we performed a selection model approach through Bayesian statistics, which revealed a weak evidence for the inverse q-Gamma distribution, for a considerable number of chromosomes.

This article is organized as follows. The next section details the stochastic model based on superstatistics that are presented. In Sec. III we will apply the stochastic model in the "time" series produced from the data collected from the Ensembl Project. The main conclusions are presented in Sec. IV.

II. THE METHOD

Let us introduce the fluctuations in the distribution of exon size through the superstatistics framework. In this regard we consider the superstatistics from a dynamical viewpoint by using a Langevin equation of the form

$$\dot{x} = -\gamma F(x) + \zeta L(t), \tag{1}$$

where $L(t)$ is a white Gaussian noise, $\gamma > 0$ is a friction constant, ζ describes the noise intensity, and $F(x)$ is a drift force, which is given by $F(x) = -dV(x)/dx$. Generally, we can let the parameters γ and ζ fluctuate so that $\beta = \gamma/\zeta^2$ has the probability density $f(\beta)$ [33]. In this case the conditional probability $p(x|\beta)$ is obtained,

$$p(x|\beta) = \frac{1}{Z(\beta)} \exp[-\beta V(x)], \tag{2}$$

and for the marginal probability $p(x)$,

$$p(x) = \int p(x|\beta) f(\beta) d\beta. \tag{3}$$

The distribution for $f(\beta)$ is a central choice that defines the concept of superstatistics. Initially, $f(\beta)$ can be any normalized probability density. However, as β needs to be positive [33], we use two types of superstatistics that best fit this case: Gamma superstatistics and inverse Gamma superstatistics.

Here, we consider that the evolution of the size distribution of exons obeys the following set of stochastic differential equations:

$$dl = -\gamma(l - \tau)dt + \left(\sqrt{\frac{2}{\phi}\tau\gamma l}\right)dW_t, \tag{4}$$

and

$$dl = -\gamma(l - \tau)dt + \left(\sqrt{2\frac{\gamma}{\alpha}l}\right)dW_t, \tag{5}$$

where W_t is the regular Wiener process that follows a normal distribution with zero mean and unitary variance, $l > 0$, and ϕ, α are adimensional constant characteristics of the system. In Eqs. (4) and (5) the first term on the right-hand side is associated with a deterministic process that aims to keep the size of the exons around a characteristic value τ. However, the second term on the right-hand side of the equations is interpreted as a stochastic term. It represents a memory effect on the evolution of the size of exons. Due to the random signal, W_t increments ($W_t > 0$) or decrements ($W_t < 0$) the size of l. The stochastic equations (4) and (5) are an example of multiplicative noise, known as a Feller process [58]. Heston initially proposed the process represented by Eq. (4) to describe the stochastic volatility in trading price returns [59]. In the same way, Queirós describes sequences of volumes traded in stocks in the financial markets [35]. Finally, Michas and Vallianatos used the same process to describe the time series of earthquakes [37]. However, Eq. (5) appears as a modification of the Heston model proposed by de Sousa *et al.* to consider a parabolic profile in the distribution in active trades [36]. In the following sections we used the above results and those obtained in Refs. [35,36], which were adjustable with our scheme, to introduce fluctuations of exon size distributions.

A. Inverse Gamma superstatistics

Initially, to determine the "temporal" evolution, after some discrete time t, given by the probability distribution f, we can write the corresponding Fokker-Planck equation for Eq. (4),

$$\frac{\partial f}{\partial t} = \frac{\partial}{\partial l}[\gamma(l - \tau)f] + \frac{\partial^2}{\partial l^2}\left[l\tau\frac{\gamma}{\phi}f\right], \tag{6}$$

and determine the stationary solution,

$$f(l) = \frac{\phi^\phi}{\tau\Gamma(\phi)}\left(\frac{l}{\tau}\right)^{\phi-1} \exp\left(-\frac{\phi}{\tau}l\right). \tag{7}$$

In order to introduce the fluctuations in the local mean τ of the exon sizes we assume the stationary inverse Gamma distribution:

$$p(\tau) = \frac{\left(\frac{\phi}{\lambda}\right)^\delta}{\Gamma(\delta)}\tau^{-\delta-1} \exp\left(-\frac{\phi}{\tau\lambda}\right). \tag{8}$$

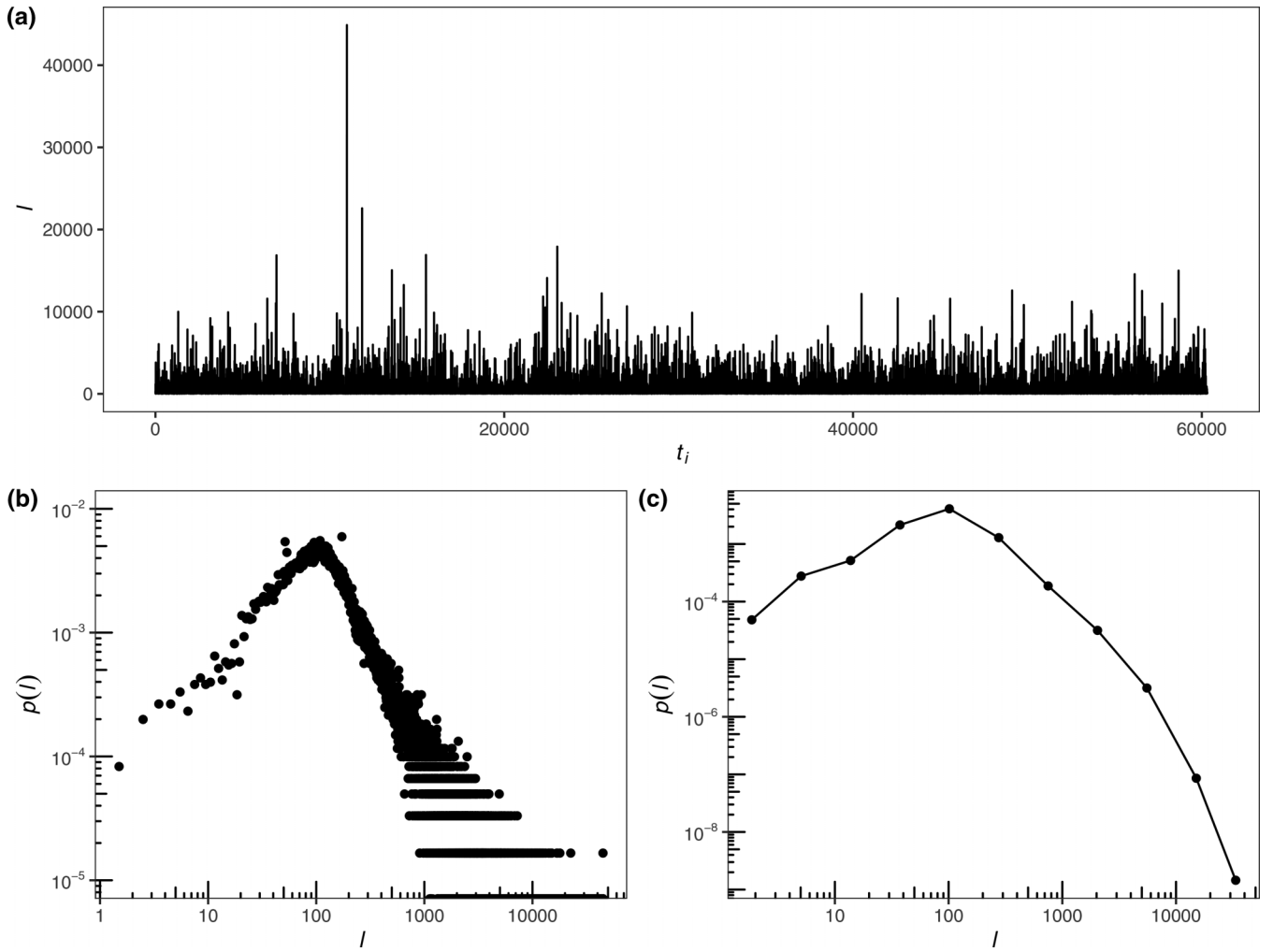

FIG. 1. Time series representation and statistical evaluation method. (a) Time series for chromosome 01 was created using data from the Ensembl Project [60]. The "spatial" series (l) represents the spatial displacement along the DNA sequence at "time" t. The coordinate position (t_i) is also linked to a "temporal" index, with $i = 1, \ldots, n$. (b) The probability density derived from the chromosome 01 time series. (c) The logarithmic box representation of probability density.

In this case Eq. (7) provides the conditional probability of l given τ:

$$f(l) \to p(l|\tau) = \frac{\phi^{\phi}}{\tau\Gamma(\phi)}\left(\frac{l}{\tau}\right)^{\phi-1}\exp\left(-\frac{\phi}{\tau}l\right). \tag{9}$$

Therefore the joint probability of getting certain values of l and τ is $P(l,\tau) = p(l|\tau)P(\tau)$, and the marginal probability of having some value l, independent of τ, is given by

$$p(l) = \int_0^{\infty} p(l|\tau)p(\tau)d\tau. \tag{10}$$

Applying Eqs. (8) and (9) in Eq. (10) and performing the integration, we have

$$p(l) = \frac{\lambda\Gamma(\phi+\delta)}{\Gamma(\phi)\Gamma(\delta)}(\lambda l)^{\phi-1}(1+\lambda l)^{-\phi-\delta}. \tag{11}$$

Using the following variable changes,

$$\lambda = \frac{q-1}{\sigma}, \quad \delta = \frac{1}{q-1} - \phi, \quad \alpha = \phi - 1, \tag{12}$$

we can write Eq. (11) as $p_G(l)$ in the form

$$P_G(l) = \frac{(q-1)^{\alpha+1}\Gamma\left(\frac{1}{q-1}\right)}{\sigma\Gamma\left(\frac{1}{q-1}-\alpha-1\right)\Gamma(\alpha+1)}\left(\frac{l}{\sigma}\right)^{\alpha}\exp_q\left(-\frac{l}{\sigma}\right), \tag{13}$$

which is the q-Gamma probability density function. It is worth noting that

$$\exp_q\left(-\frac{l}{\sigma}\right) = \left[1+(1-q)\left(-\frac{l}{\sigma}\right)\right]^{\frac{1}{1-q}}. \tag{14}$$

B. Gamma superstatistics

By using the process described by Eq. (5), we can construct the discrete-time "temporal" evolution for the size distribution of the exons. Thus we can write the Fokker-Planck equation in the form

$$\frac{\partial f}{\partial t} = \frac{\partial}{\partial l}[\gamma(l-\tau)f] + \frac{\partial^2}{\partial l^2}\left[\frac{\gamma}{\alpha}l^2 f\right], \tag{15}$$

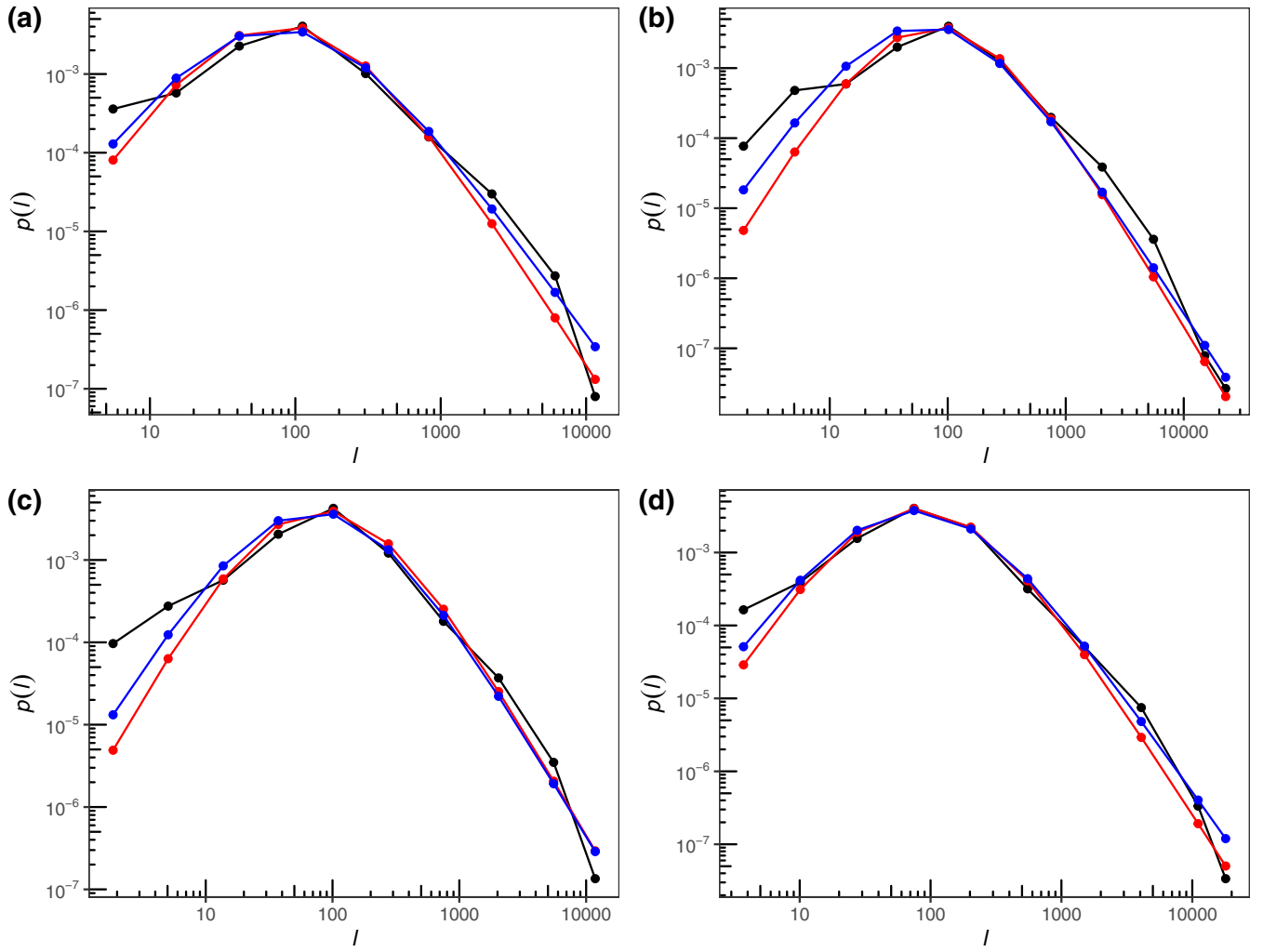

FIG. 2. Superstatistical distribution of sizes for strings of exons from chromosomes: (a) 04, (b) 06, (c) 10, and (d) 15. The black, red, and blue curves represent observational data, the q-Gamma distribution, and the inverse q-Gamma distribution, respectively.

with the stationary solution

$$f(l) \to p(l|\tau) = \frac{\alpha^{\alpha+1}}{\omega\Gamma(\alpha+1)}\left(\frac{l}{\tau}\right)^{-\alpha-2}\exp\left(-\frac{\alpha\tau}{l}\right). \quad (16)$$

Assuming that τ follows the Gamma distribution,

$$p(\tau) = \frac{1}{\lambda\Gamma(\delta)}\left(\frac{\tau}{\delta}\right)^{\delta-1}\exp\left(-\frac{\tau}{\delta}\right). \quad (17)$$

Applying Eqs. (16) and (17) to the conditional probability equation in the form of the Eq. (10), and performing the integration of the conditional probability, we arrive at

$$P_{IG}(l) = A_{IG}\left(\frac{l}{\sigma}\right)^{-\alpha-2}\exp_q\left(-\frac{\sigma}{l}\right), \quad (18)$$

where $P_{IG}(l)$ is the inverse q-Gamma probability density function, with

$$A_{IG} = \frac{\Gamma\left(\frac{1}{q-1}\right)}{\sigma(q-1)\Gamma(\alpha+1)\Gamma\left(\frac{1}{q-1}-\alpha-1\right)}\left(\frac{1}{q-1}\right)^{-\alpha-2}, \quad (19)$$

where

$$\alpha\lambda = \sigma(q-1), \quad \delta = \frac{1}{q-1} - \alpha - 1 \quad (20)$$

are the changes of variables.

III. RESULTS AND DISCUSSIONS

A. Data and time series

In our analysis we used the Project Ensembl [60] database of genes to extract the time evolution of exon sizes and validate our stochastic model. The size of a coding region sequence (exon), l, is given in terms of the number of base pairs (bp) in these biological databases. The core of the whole argument follows from the definition of time series for exon sizes (see Fig. 1). Indeed, the time series is defined by the fact that the x axis represents a discrete-time instant, t_i. For the occurrence of the ith exon, the y axis reflects exon sizes for each event on the x axis. The time series for chromosome 01 is depicted in Fig. 1(a).

Now, let us obtain the time series' statistical information. To do so we compute the probability density, $p(l)$, for the

TABLE I. The model parameters that best fit the chromosomal dataset. A_G, α_G, σ_G, and q_G are free parameters for the q-Gamma distribution as defined by Eq. (13).

CHR	A_G	α_G	σ_G	q_G
01	$9.9545 \times 10^{-7^{+2.7261\times10^{-7}}_{-2.7261\times10^{-7}}}$	$2.7835^{+0.3603}_{-0.3603}$	$14.0062^{+1.7077}_{-1.7077}$	$1.1841^{+0.0258}_{-0.0258}$
02	$9.7191 \times 10^{-7^{+2.6782\times10^{-7}}_{-2.6782\times10^{-7}}}$	$2.7926^{+0.3594}_{-0.3594}$	$14.1163^{+1.7382}_{-1.7382}$	$1.1765^{+0.0260}_{-0.0260}$
03	$9.9768 \times 10^{-7^{+2.7437\times10^{-7}}_{-2.7437\times10^{-7}}}$	$2.7984^{+0.3654}_{-0.3654}$	$14.5449^{+1.7458}_{-1.7458}$	$1.1684^{+0.0262}_{-0.0262}$
04	$9.7814 \times 10^{-7^{+2.7357\times10^{-7}}_{-2.7357\times10^{-7}}}$	$2.7822^{+0.3610}_{-0.3610}$	$14.5997^{+1.7189}_{-1.7189}$	$1.1764^{+0.0256}_{-0.0256}$
05	$9.9273 \times 10^{-7^{+2.7666\times10^{-7}}_{-2.7666\times10^{-7}}}$	$2.7938^{+0.3666}_{-0.3666}$	$14.0016^{+1.7086}_{-1.7086}$	$1.1796^{+0.0251}_{-0.0251}$
06	$9.6979 \times 10^{-7^{+2.7431\times10^{-7}}_{-2.7431\times10^{-7}}}$	$2.7961^{+0.3649}_{-0.3649}$	$14.0165^{+1.7257}_{-1.7257}$	$1.1775^{+0.0255}_{-0.0255}$
07	$9.8418 \times 10^{-7^{+2.7514\times10^{-7}}_{-2.7514\times10^{-7}}}$	$2.7532^{+0.3597}_{-0.3597}$	$15.0927^{+1.6997}_{-1.6997}$	$1.1745^{+0.0260}_{-0.0260}$
08	$9.6790 \times 10^{-7^{+2.7912\times10^{-7}}_{-2.7912\times10^{-7}}}$	$2.7640^{+0.3550}_{-0.3550}$	$14.4503^{+1.7384}_{-1.7384}$	$1.1795^{+0.0260}_{-0.0260}$
09	$9.9396 \times 10^{-7^{+2.7584\times10^{-7}}_{-2.7584\times10^{-7}}}$	$2.7574^{+0.3522}_{-0.3522}$	$14.4355^{+1.7031}_{-1.7031}$	$1.1813^{+0.0258}_{-0.0258}$
10	$9.9837 \times 10^{-7^{+2.7615\times10^{-7}}_{-2.7615\times10^{-7}}}$	$2.7730^{+0.3535}_{-0.3535}$	$14.2621^{+1.7198}_{-1.7198}$	$1.1850^{+0.0266}_{-0.0266}$
11	$9.8674 \times 10^{-7^{+2.7379\times10^{-7}}_{-2.7379\times10^{-7}}}$	$2.7805^{+0.3591}_{-0.3591}$	$14.4021^{+1.7346}_{-1.7346}$	$1.1793^{+0.0251}_{-0.0251}$
12	$9.9750 \times 10^{-7^{+2.7174\times10^{-7}}_{-2.7174\times10^{-7}}}$	$2.8053^{+0.3638}_{-0.3638}$	$14.1930^{+1.6999}_{-1.6999}$	$1.1709^{+0.0220}_{-0.0220}$
13	$9.9426 \times 10^{-7^{+2.7533\times10^{-7}}_{-2.7533\times10^{-7}}}$	$2.7976^{+0.3610}_{-0.3610}$	$14.2487^{+1.7254}_{-1.7254}$	$1.1761^{+0.0257}_{-0.0257}$
14	$9.7344 \times 10^{-7^{+2.7584\times10^{-7}}_{-2.7584\times10^{-7}}}$	$2.7993^{+0.3627}_{-0.3627}$	$14.1856^{+1.7176}_{-1.7176}$	$1.1774^{+0.0258}_{-0.0258}$
15	$9.5088 \times 10^{-7^{+2.7702\times10^{-7}}_{-2.7702\times10^{-7}}}$	$2.7919^{+0.3642}_{-0.3642}$	$14.2550^{+1.7274}_{-1.7274}$	$1.1790^{+0.0255}_{-0.0255}$
16	$9.9821 \times 10^{-7^{+2.7678\times10^{-7}}_{-2.7678\times10^{-7}}}$	$2.7871^{+0.3584}_{-0.3584}$	$14.0445^{+1.6988}_{-1.6988}$	$1.1820^{+0.0263}_{-0.0263}$
17	$9.8949 \times 10^{-7^{+2.7164\times10^{-7}}_{-2.7164\times10^{-7}}}$	$2.8599^{+0.3657}_{-0.3657}$	$14.0375^{+1.7289}_{-1.7289}$	$1.1616^{+0.0273}_{-0.0273}$
18	$9.8306 \times 10^{-7^{+2.7561\times10^{-7}}_{-2.7561\times10^{-7}}}$	$2.7993^{+0.3554}_{-0.3554}$	$14.1634^{+1.6934}_{-1.6934}$	$1.1765^{+0.0246}_{-0.0246}$
19	$9.9611 \times 10^{-7^{+2.7444\times10^{-7}}_{-2.7444\times10^{-7}}}$	$2.8681^{+0.3641}_{-0.3641}$	$14.6810^{+1.7159}_{-1.7159}$	$1.1521^{+0.0263}_{-0.0263}$
20	$9.9927 \times 10^{-7^{+2.7521\times10^{-7}}_{-2.7521\times10^{-7}}}$	$2.7910^{+0.3571}_{-0.3571}$	$14.1938^{+1.7368}_{-1.7368}$	$1.1723^{+0.0262}_{-0.0262}$
21	$9.7201 \times 10^{-7^{+2.7457\times10^{-7}}_{-2.7457\times10^{-7}}}$	$2.7543^{+0.3651}_{-0.3651}$	$14.6779^{+1.6790}_{-1.6790}$	$1.1798^{+0.0263}_{-0.0263}$
22	$9.8672 \times 10^{-7^{+2.7129\times10^{-7}}_{-2.7129\times10^{-7}}}$	$2.8100^{+0.3645}_{-0.3645}$	$14.1671^{+1.7168}_{-1.7168}$	$1.1702^{+0.0256}_{-0.0256}$
X	$9.7975 \times 10^{-7^{+2.7576\times10^{-7}}_{-2.7576\times10^{-7}}}$	$2.7912^{+0.3648}_{-0.3648}$	$14.1710^{+1.7144}_{-1.7144}$	$1.1756^{+0.0221}_{-0.0221}$
Y	$9.8339 \times 10^{-7^{+2.7263\times10^{-7}}_{-2.7263\times10^{-7}}}$	$2.7578^{+0.3632}_{-0.3632}$	$15.7309^{+1.7380}_{-1.7380}$	$1.1673^{+0.0284}_{-0.0284}$

number of occurrences of exons of size l in Fig. 1(b). We use the logarithmic bin scheme to pinpoint the mean value of the exon sizes and to represent data fluctuations in the tails, as shown in Fig. 1(c). Thus when increasing the bin size with l, we have a better description of the form of the probability density. We categorize the data using b_i, where b_i is the bin size, and $b_i = \exp(i)$, where i is the bin number, $i = 1, 2, \ldots, n$. Let us analyze if the q-Gamma probability density function and inverse q-Gamma probability density function can capture the SRC and fluctuations among the size distribution of exon chains. Therefore the distributions will be fitted to data distribution for the size of the exon chains [see Fig. 1(c)].

In Fig. 2 we show the probability density for a sample of human chromosomes 04, 06, 10, and 15. The remaining chromosomes' behavior is analogous to the behavior shown in Fig. 2. To get the values for $P_G(l)$ and $P_{IG}(l)$, the functions (13) and (18) were utilized to fit the dataset. Tables I and II show the best-fit parameters. Moreover, for both distributions $P_G(l)$ and $P_{IG}(l)$, the fitting procedure was implemented through the LM (Levenberg-Marquardt) numerical algorithm [61] written in *R* [62], which calculates the best-fit parameters for both distributions, adjusted to the chromosomic data. It is worth noting that when l is less than 10 bp for some chromosomes, we observe that the models deviate somewhat from the data. On the other hand, the models can capture the data in the region above 10 bp, which has the most significant number of exons.

B. Bayesian analysis

Before continuing, let us discuss a few points about the Bayesian inference, which is becoming a valuable tool for tackling many problems in physics (for a review, see Ref. [63]), ecology [64–66], and biophysics (for an excellent primer, see Ref. [67]). The rationale behind this technique is updating previous knowledge about some model parameters based on learning new information. Indeed, this is a method of statistical inference in which one uses Bayes' theorem to describe the relationship between models, data, and prior information about model parameters. In a parameter estima-

TABLE II. The same as Table I, but for free parameters A_{IG}, α_{IG}, σ_{IG}, and q_{IG} from the inverse q-Gamma distribution, as defined by Eq. (18).

CHR	A_{IG}	α_{IG}	σ_{IG}	q_{IG}
1	$7809.203^{+1144.440}_{-1144.440}$	$0.5518123^{+0.05768208}_{-0.05768208}$	$379.7055^{+58.41313}_{-58.41313}$	$1.199526^{+0.02806757}_{-0.02806757}$
2	$7436.514^{+1138.926}_{-1138.926}$	$0.5558285^{+0.05690273}_{-0.05690273}$	$371.7458^{+57.64390}_{-57.64390}$	$1.199960^{+0.03196373}_{-0.03196373}$
3	$7997.262^{+1171.284}_{-1171.284}$	$0.5946445^{+0.05728736}_{-0.05728736}$	$357.6161^{+58.45881}_{-58.45881}$	$1.199874^{+0.02560120}_{-0.02560120}$
4	$7894.546^{+1152.681}_{-1152.681}$	$0.5472515^{+0.05854799}_{-0.05854799}$	$387.5022^{+58.21772}_{-58.21772}$	$1.198991^{+0.03518453}_{-0.03518453}$
5	$7742.633^{+1155.663}_{-1155.663}$	$0.5767669^{+0.05798930}_{-0.05798930}$	$364.1540^{+57.48912}_{-57.48912}$	$1.199902^{+0.02662069}_{-0.02662069}$
6	$7383.890^{+1143.651}_{-1143.651}$	$0.5890860^{+0.05740581}_{-0.05740581}$	$343.4822^{+57.08640}_{-57.08640}$	$1.199539^{+0.02704803}_{-0.02704803}$
7	$7972.993^{+1153.776}_{-1153.776}$	$0.5752212^{+0.05716543}_{-0.05716543}$	$372.3990^{+58.01888}_{-58.01888}$	$1.199994^{+0.02743487}_{-0.02743487}$
8	$7710.580^{+1147.006}_{-1147.006}$	$0.5546483^{+0.05771361}_{-0.05771361}$	$376.7511^{+56.99490}_{-56.99490}$	$1.199383^{+0.03111394}_{-0.03111394}$
9	$7004.210^{+1165.496}_{-1165.496}$	$0.5592837^{+0.05856234}_{-0.05856234}$	$358.7707^{+58.16187}_{-58.16187}$	$1.199847^{+0.03031994}_{-0.03031994}$
10	$7800.751^{+1136.923}_{-1136.923}$	$0.5590480^{+0.05765129}_{-0.05765129}$	$374.9343^{+57.18684}_{-57.18684}$	$1.199946^{+0.02562836}_{-0.02562836}$
11	$7689.615^{+1144.887}_{-1144.887}$	$0.5747293^{+0.05786969}_{-0.05786969}$	$359.8395^{+58.58723}_{-58.58723}$	$1.199871^{+0.03518177}_{-0.03518177}$
12	$7526.128^{+1168.163}_{-1168.163}$	$0.5477628^{+0.05694147}_{-0.05694147}$	$370.6415^{+57.05604}_{-57.05604}$	$1.199685^{+0.03646363}_{-0.03646363}$
13	$6780.360^{+1168.263}_{-1168.263}$	$0.5910883^{+0.05727767}_{-0.05727767}$	$333.5107^{+57.19265}_{-57.19265}$	$1.199689^{+0.02387596}_{-0.02387596}$
14	$6055.631^{+1144.057}_{-1144.057}$	$0.5242399^{+0.05827816}_{-0.05827816}$	$362.4286^{+57.17177}_{-57.17177}$	$1.199968^{+0.02816265}_{-0.02816265}$
15	$7803.435^{+1145.904}_{-1145.904}$	$0.5393364^{+0.05846271}_{-0.05846271}$	$392.1554^{+56.86825}_{-56.86825}$	$1.199411^{+0.03267771}_{-0.03267771}$
16	$7899.175^{+1153.793}_{-1153.793}$	$0.5791303^{+0.05782361}_{-0.05782361}$	$363.9262^{+57.97265}_{-57.97265}$	$1.199695^{+0.03073250}_{-0.03073250}$
17	$6869.810^{+1150.878}_{-1150.878}$	$0.5908990^{+0.05826209}_{-0.05826209}$	$325.5873^{+58.27176}_{-58.27176}$	$1.199341^{+0.03048527}_{-0.03048527}$
18	$7726.757^{+1147.334}_{-1147.334}$	$0.5701726^{+0.05754791}_{-0.05754791}$	$360.7312^{+58.53472}_{-58.53472}$	$1.199895^{+0.02773004}_{-0.02773004}$
19	$5610.925^{+1165.721}_{-1165.721}$	$0.5815002^{+0.05822773}_{-0.05822773}$	$303.6537^{+57.23584}_{-57.23584}$	$1.199576^{+0.03374223}_{-0.03374223}$
20	$7487.947^{+1150.926}_{-1150.926}$	$0.5814504^{+0.05805905}_{-0.05805905}$	$335.1194^{+57.48598}_{-57.48598}$	$1.199572^{+0.02506806}_{-0.02506806}$
21	$7084.840^{+1137.057}_{-1137.057}$	$0.5461979^{+0.05879125}_{-0.05879125}$	$376.4362^{+58.51314}_{-58.51314}$	$1.199018^{+0.02712447}_{-0.02712447}$
22	$7403.418^{+1154.598}_{-1154.598}$	$0.5702446^{+0.05784792}_{-0.05784792}$	$358.1559^{+57.49497}_{-57.49497}$	$1.199587^{+0.02871884}_{-0.02871884}$
X	$7364.109^{+1127.271}_{-1127.271}$	$0.5535639^{+0.05841418}_{-0.05841418}$	$374.7473^{+57.76424}_{-57.76424}$	$1.199424^{+0.02947948}_{-0.02947948}$
Y	$7783.303^{+1157.003}_{-1157.003}$	$0.5262098^{+0.05778872}_{-0.05778872}$	$379.3963^{+58.01137}_{-58.01137}$	$1.198027^{+0.03682755}_{-0.03682755}$

tion problem, the starting point for Bayesian analysis is to calculate the posterior distribution for a set Φ of free parameters given the data D and model M through Bayes' theorem,

$$P(\Phi|D,M) = \frac{\mathcal{L}(D|\Phi,M)P(\Phi|M)}{\epsilon(D|M)}, \tag{21}$$

where $P(\Phi|D,M)$ is the posterior distribution, $\mathcal{L}(D|\Phi,M)$ is the likelihood function, $P(\Phi|M)$ is the prior distribution, and $\epsilon(D|M)$ is the Bayesian evidence.

The Bayesian evidence $\epsilon(D|M)$ is simply a normalization constant regarding parameter constraints. Furthermore, because it is independent of the model parameters, it does not affect the profile of the posterior distribution. Nonetheless, when looking at Bayesian model comparisons, it turns out to be a necessary component. In the continuous parameter space, the Bayesian evidence of a model is given by

$$\epsilon(D|M) = \int \mathcal{L}(D|\Phi,M)P(\Phi|M)d\Phi. \tag{22}$$

An interesting question here is to inspect which length-distribution models of DNA chains should be viable from the point of view of Bayesian inference. Therefore we compare the distributions (13) and (18) through the calculation of Bayesian evidence. As a result, we compute the ratio of the posterior probabilities, which is given by

$$\frac{P(M_1|D)}{P(M_2|D)} = B_{12}\frac{P(M_1)}{P(M_2)}, \tag{23}$$

where B_{12} is known as the Bayes factor, defined as

$$B_{12} = \frac{\epsilon_1(D|M_1)}{\epsilon_2(D|M_2)}, \tag{24}$$

where $\epsilon_1(D|M_1)$ is the evidence of model 01, which in our case is the q-Gamma distribution, and $\epsilon_2(D|M_2)$, model 02, is the evidence of the inverse q-Gamma distribution.

Furthermore, we describe the likelihood function pattern for the entire human genome as follows:

$$\chi^2 = \frac{(p_{\rm obs}(l) - p_{\rm the}(l))^2}{\sigma^2_{\rm obs}}, \tag{25}$$

where $p_{\rm obs}(l)$, $p_{\rm the}(l)$, and $\sigma_{\rm obs}$ are the probability density of the observed nucleotides, the theoretical probability, and the

TABLE III. The Jeffreys scale for interpretation of the Bayes factor. The first column represents the logarithm of the Bayes factor limit values, while the second column is the interpretation of the evidence's strength over the appropriate threshold.

$\ln B_{ij}$	Interpretation
Greater than 5	Strong evidence for model 01
[2.5, 5]	Moderate evidence for model 01
[1, 2.5]	Weak evidence for model 01
[−1,1]	Inconclusive
[−2.5,−1]	Weak evidence for model 02
[−5, 2.5]	Moderate evidence for model 02
Less than −5	Strong evidence for model 02

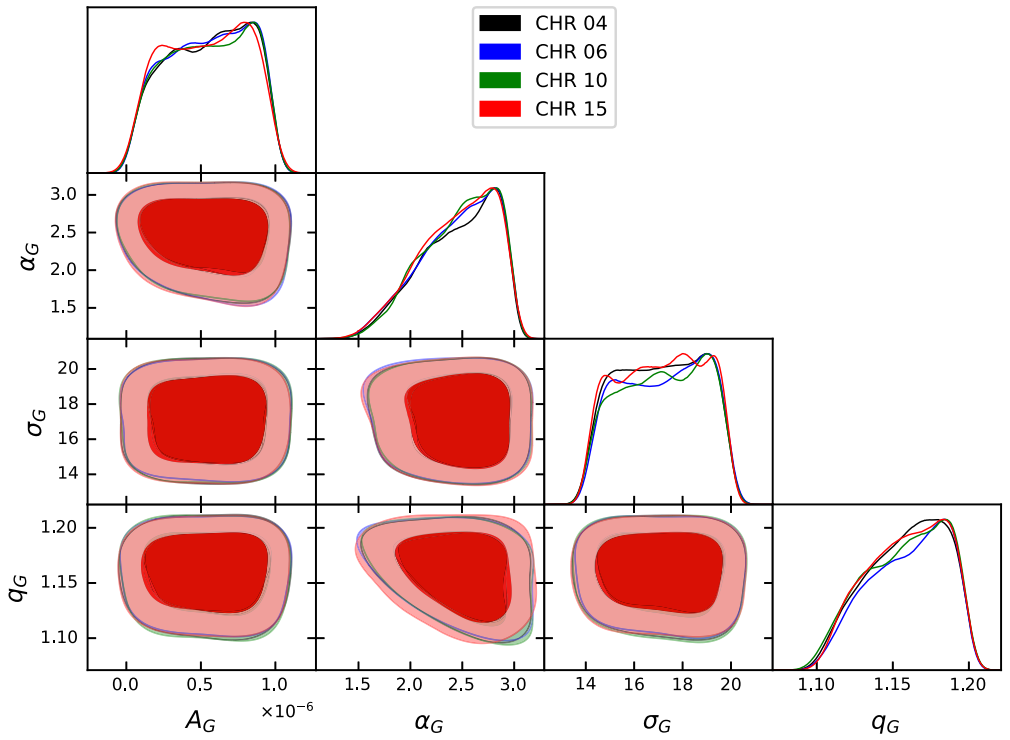

FIG. 3. Results from the Bayesian inference procedure, showing projections of the posterior distributions for the free parameters A_G, α_G, σ_G, q_G according to the q-Gamma distribution for chromosomes 04, 06, 10, and 15.

observed error, respectively. To acquire a clearer picture of whether or not a model contains favorable evidence in comparison to the base model, we make use of the interpretation of the Bayes factor following the Jeffreys scale, Table III [68].

To compute the Bayesian evidence $\epsilon(D|M)$ with an accompanying error estimate, we use the MULTINEST code [69], a Bayesian inference tool. This approach is implemented for each chromosome and model to conduct the Bayesian

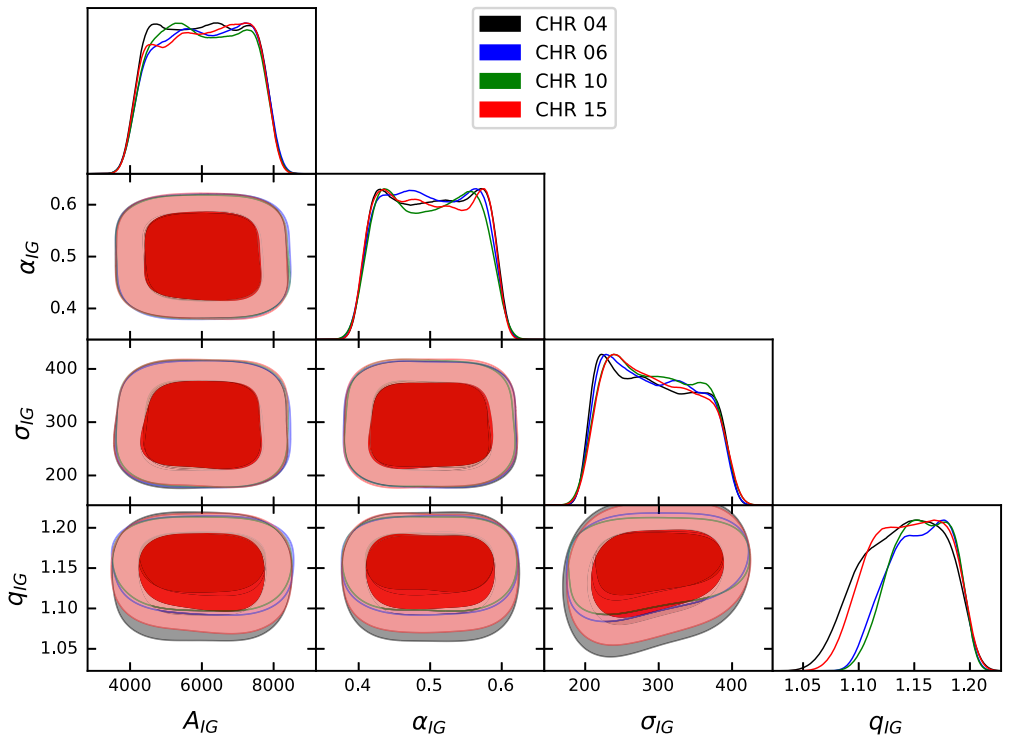

FIG. 4. The same as Fig. 3 but for free parameters A_{IG}, α_{IG}, σ_{IG}, q_{IG} from inverse q-Gamma distribution for chromosomes 04, 06, 10, and 15.

TABLE IV. The table shows the uniform priors on the free parameters of q-Gamma distribution.

CHR	Parameters	Priors
All	A_G	$\mathcal{U}(1\times10^{-8}, 1\times10^{-6})$
	α_G	$\mathcal{U}(1, 3)$
	σ_G	$\mathcal{U}(14, 20)$
	q_G	$\mathcal{U}(1, 1.2)$

analysis. It generates posterior samples from distributions with several modes and numerous curves in a large number of dimensions [70,71]. We provide the triangle plot constructed of confidence zones for the parameters and subsequent distributions in Figs. 3 and 4, respectively, for model 01, q-Gamma distribution, and model 02, inverse q-Gamma distribution. The chromosomes displayed are a random selection from our sample.

We obtained these findings using the priors specified in Tables IV and V. We used the uniform prior for both models because it has the smallest statistical influence on the likelihood [Eq. (25)], resulting in more confidence in the selection analysis. For chromosome 04, the Bayes factor is calculated as $\ln(B_{12}) = -1.3671^{+0.0019}_{-0.0019}$. Using the Jeffreys scale, Table III, we find that model 02 can represent the data set with weak evidence. In the instance of chromosome 06, however, we derived the Bayes factor $\ln(B_{12}) = -0.9298^{+0.0013}_{-0.0013}$. We also saw that the lack of clear evidence for the offered models prevented us from estimating which models best describe the behavior of chromosome 3. The findings for the remaining chromosomes are shown in Table VI. Consequently, chromosomes 3, 6, 10, 13, and 20 provided comprehensive support for the suggested hypotheses, i.e., inconclusive. The remaining 19 chromosomes, however, exhibited minimal support for model 02. Specifically, we found that model 02, the inverse q-Gamma distribution, presented weak evidence to characterize the human genome.

TABLE VI. The results of the Bayesian analysis for each chromosome. The column $\ln(\epsilon_1)$ gives us the Bayesian evidence for each of the models, Eq. (13). The column $\ln(\epsilon_2)$ gives us the Bayesian evidence for each of the models, Eq. (18), and column $\ln(B_{12})$ gives us the Bayes factor.

CHR	$\ln(\epsilon_1)$	$\ln(\epsilon_2)$	$\ln(B_{12})$
01	$-2.1356^{+0.0148}_{-0.0148}$	$-0.9881^{+0.0133}_{-0.0133}$	$-1.1475^{+0.0014}_{-0.0014}$
02	$-2.0501^{+0.0153}_{-0.0153}$	$-0.9918^{+0.0147}_{-0.0147}$	$-1.0582^{+0.0006}_{-0.0006}$
03	$-2.0260^{+0.0152}_{-0.0152}$	$-1.2671^{+0.0154}_{-0.0154}$	$-0.7589^{+0.0002}_{-0.0002}$
04	$-2.1668^{+0.0150}_{-0.0150}$	$-0.7996^{+0.0130}_{-0.0130}$	$-1.3671^{+0.0019}_{-0.0019}$
05	$-2.2890^{+0.0139}_{-0.0139}$	$-1.2290^{+0.0158}_{-0.0158}$	$-1.0599^{+0.0019}_{-0.0019}$
06	$-2.1791^{+0.0148}_{-0.0148}$	$-1.2493^{+0.0162}_{-0.0162}$	$-0.9298^{+0.0013}_{-0.0013}$
07	$-2.2418^{+0.0127}_{-0.0127}$	$-1.1799^{+0.0155}_{-0.0155}$	$-1.0619^{+0.0028}_{-0.0028}$
08	$-2.2871^{+0.0137}_{-0.0137}$	$-0.9765^{+0.0139}_{-0.0139}$	$-1.3105^{+0.0002}_{-0.0002}$
09	$-2.2668^{+0.0133}_{-0.0133}$	$-1.0133^{+0.0145}_{-0.0145}$	$-1.2535^{+0.0011}_{-0.0011}$
10	$-2.2674^{+0.0134}_{-0.0134}$	$-1.2816^{+0.0158}_{-0.0158}$	$-0.9857^{+0.0024}_{-0.0024}$
11	$-2.0635^{+0.0153}_{-0.0153}$	$-0.8889^{+0.0145}_{-0.0145}$	$-1.1746^{+0.0007}_{-0.0007}$
12	$-2.1906^{+0.0158}_{-0.0158}$	$-0.7556^{+0.0118}_{-0.0118}$	$-1.4349^{+0.0040}_{-0.0040}$
13	$-2.1262^{+0.0156}_{-0.0156}$	$-1.2963^{+0.0151}_{-0.0151}$	$-0.8298^{+0.0004}_{-0.0004}$
14	$-2.2396^{+0.0132}_{-0.0132}$	$-1.1546^{+0.0156}_{-0.0156}$	$-1.0849^{+0.0023}_{-0.0023}$
15	$-2.1806^{+0.0135}_{-0.0135}$	$-0.9585^{+0.0151}_{-0.0151}$	$-1.2220^{+0.0016}_{-0.0016}$
16	$-2.1296^{+0.0137}_{-0.0137}$	$-1.0316^{+0.0146}_{-0.0146}$	$-1.0979^{+0.0009}_{-0.0009}$
17	$-2.1629^{+0.0143}_{-0.0143}$	$-1.0899^{+0.0155}_{-0.0155}$	$-1.0730^{+0.0011}_{-0.0011}$
18	$-2.3017^{+0.0119}_{-0.0119}$	$-1.1454^{+0.0157}_{-0.0157}$	$-1.1562^{+0.0037}_{-0.0037}$
19	$-2.2425^{+0.0145}_{-0.0145}$	$-0.9028^{+0.0142}_{-0.0142}$	$-1.3396^{+0.0003}_{-0.0003}$
20	$-2.1533^{+0.0146}_{-0.0146}$	$-1.3560^{+0.0167}_{-0.0167}$	$-0.7973^{+0.0021}_{-0.0021}$
21	$-2.2438^{+0.0132}_{-0.0132}$	$-1.1861^{+0.0156}_{-0.0156}$	$-1.0577^{+0.0024}_{-0.0024}$
22	$-2.0817^{+0.0158}_{-0.0158}$	$-1.0795^{+0.0143}_{-0.0143}$	$-1.0021^{+0.0015}_{-0.0015}$
X	$-2.4011^{+0.0130}_{-0.0130}$	$-1.0556^{+0.0153}_{-0.0153}$	$-1.3454^{+0.0022}_{-0.0022}$
Y	$-1.9682^{+0.0152}_{-0.0152}$	$-0.8268^{+0.0141}_{-0.0141}$	$-1.1414^{+0.0010}_{-0.0010}$

IV. CONCLUSION

The SRCs always present in exon size distributions were investigated in [29–31] through the ECDF with suppressed fluctuations. Certainly, the fluctuations of the exon size distributions should be considered in order to analyze their role in the SRC. In this context we presented models based on the superstatistics formalism discussed in Refs. [33,34,39]. Specifically, we proposed a stochastic model through the Langevin and Fokker-Planck, which provided the superstatistics distributions for the sizes of exon chains, Eqs. (13) and (18), also known as q-Gamma distributions (model 01) and inverse q-Gamma distributions (model 02), respectively. The argument's core followed the construction of a temporal series for exon sizes calculated in terms of the number of base pairs (bp). As a result, we showed that at least for sizes of exon chains over 10 bp, such distributions reasonably adjusted with the data from capturing the SRC and fluctuations in the size distribution of the exon chains. The best fit of free parameters from the superstatistics distributions for the whole genome follows in Tables I and II.

We implemented a selection model approach through Bayesian statistics to compare both superstatistics distributions. In Figs. 3 and 4, we presented the triangle plot constructed of confidence zones for the parameters and subsequent distributions for a selected sample of chromosomes as the main results. Moreover, we presented the result of the Bayesian analysis for the whole genome in Table VI. The analysis was inconclusive for chromosomes 3, 6, 10, 13, and 20. However, the remaining 19 chromosomes exhibited minimal support for inverse q-Gamma distribution (model 02).

The DNA code data that support the findings of this study are available in the Ensembl Project database [60].

TABLE V. The table shows the uniform priors on the free parameters of inverse q-Gamma distribution.

CHR	Parameters	Priors
All	A_{IG}	$\mathcal{U}(4000, 8000)$
	α_{IG}	$\mathcal{U}(0.4, 0.6)$
	σ_{IG}	$\mathcal{U}(200, 400)$
	q_{IG}	$\mathcal{U}(1, 1.2)$

ACKNOWLEDGMENTS

This study was financed in part by CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) and by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES). R.S. thanks CNPq (Grant No. 307620/2019-0) for financial support. D. Anselmo acknowledges CNPq for financial support (Grant No. 317464/2021-3).

[1] C. K. Peng, S. V. Buldyrev, A. L. Goldberger, S. Havlin, F. Sciortino, M. Simons, and H. E. Stanley, Long-range correlations in nucleotide sequences, Nature (London) **356**, 168 (1992).

[2] W. Li and K. Kaneko, Long-range correlation and partial $1/f^{\alpha}$ spectrum in a noncoding DNA sequence, Europhys. Lett. **17**, 655 (1992).

[3] W. Li, The study of correlation structures of DNA sequences: A critical review, Comput. Chem. **21**, 257 (1997).

[4] A. Colliva, R. Pellegrini, A. Testori, and M. Caselle, Ising-model description of long-range correlations in DNA sequences, Phys. Rev. E **91**, 052703 (2015).

[5] A. Arneodo, E. Bacry, P. V. Graves, and J. F. Muzy, Characterizing Long-Range Correlations in DNA Sequences from Wavelet Analysis, Phys. Rev. Lett. **74**, 3293 (1995).

[6] B. Audit, C. Thermes, C. Vaillant, Y. d'Aubenton-Carafa, J. F. Muzy, and A. Arneodo, Long-Range Correlations in Genomic DNA: A Signature of the Nucleosomal Structure, Phys. Rev. Lett. **86**, 2471 (2001).

[7] S. V. Buldyrev, A. L. Goldberger, S. Havlin, R. N. Mantegna, M. E. Matsa, C.-K. Peng, M. Simons, and H. E. Stanley, Long-range correlation properties of coding and noncoding DNA sequences: Genbank analysis, Phys. Rev. E **51**, 5084 (1995).

[8] A. Provata and T. Oikonomou, Power law exponents characterizing human DNA, Phys. Rev. E **75**, 056102 (2007).

[9] Y. Almirantis and A. Provata, Long-and short-range correlations in genome organization, J. Stat. Phys. **97**, 233 (1999).

[10] A. Provata and Y. Almirantis, Fractal cantor patterns in the sequence structure of DNA, Fractals **08**, 15 (2000).

[11] P. Katsaloulis, T. Theoharis, and A. Provata, Statistical distributions of oligonucleotide combinations: Applications in human chromosomes 21 and 22, Physica A **316**, 380 (2002).

[12] P. Katsaloulis, T. Theoharis, W. M. Zheng, B. L. Hao, A. Bountis, Y. Almirantis, and A. Provata, Long-range correlations of RNA polymerase II promoter sequences across organisms, Physica A **366**, 308 (2006).

[13] J. D. Hawkins, A survey on intron and exon lengths, Nucl. Acids Res. **16**, 9893 (1988).

[14] M. Höglund, T. Säll, and D. Röhme, On the origin of coding sequences from random open reading frames, J. Mol. Evol. **30**, 104 (1990).

[15] M. Long and M. Deutsch, Intron-exon structures of eukaryotic model organisms, Nucleic Acids Res. **27**, 3219 (1999).

[16] C. Melodelima, L. Guéguen, D. Piau, and C. Gautier, Modelling the length distribution of exons by sums of geometric laws. analysis of the structure of genes and g+ c influence, in *Proceedings of JOBIM2004, Citeseer, 2004* (unpublished).

[17] M. K. Sakharkar, B. S. Perumal, K. R. Sakharkar, and P. Kangueane, An analysis on gene architecture in human and mouse genomes, In Silico Biol. **5**, 347 (2005).

[18] S. Gudlaugsdottir, D. R. Boswell, G. R. Wood, and J. Ma, Exon size distribution and the origin of introns, Genetica **131**, 299 (2007).

[19] W. Li, Menzerath's law at the gene-exon level in the human genome, Complexity **17**, 49 (2012).

[20] P. Menzerath, Über einige phonetische probleme [about some phonetic problems], in Actes du premier congrès international de linguistes (Leiden, Sijthoff), 1928.

[21] L. Wang and L. D. Stein, Modeling the evolution dynamics of exon-intron structure with a general random fragmentation process, BMC Evol. Biol. **13**, 57 (2013).

[22] L. Martignetti and M. Caselle, Universal power law behaviors in genomic sequences and evolutionary models, Phys. Rev. E **76**, 021902 (2007).

[23] D. Polychronopoulos, D. Sellis, and Y. Almirantis, Conserved noncoding elements follow power-law-like distributions in several genomes as a result of genome dynamics, PLoS One **9**, e95437 (2014).

[24] M. Gell-Mann and C. Tsallis, *Nonextensive Entropy: Interdisciplinary Applications* (Oxford University Press, Oxford, England, 2004).

[25] G. Kaniadakis, Maximum entropy principle and power-law tailed distributions, Eur. Phys. J. B **70**, 3 (2009).

[26] T. Oikonomou and A. Provata, Non-extensive trends in the size distribution of coding and non-coding DNA sequences in the human genome, Eur. Phys. J. B **50**, 259 (2006).

[27] T. Oikonomou, A. Provata, and U. Tirnakli, Nonextensive statistical approach to non-coding human DNA, Physica A **387**, 2653 (2008).

[28] N. T. C. M. Souza, D. H. A. L. Anselmo, R. Silva, M. S. Vasconcelos, and V. D. Mello, A κ-statistical analysis of the y-chromosome, Europhys. Lett. **108**, 38004 (2014).

[29] M. O. Costa, R. Silva, D. H. A. L. Anselmo, and J. R. P. Silva, Analysis of human DNA through power-law statistics, Phys. Rev. E **99**, 022112 (2019).

[30] R. Silva, J. R. P. Silva, D. H. A. L. Anselmo, J. S. Alcaniz, W. J. C. da Silva, and M. O. Costa, An alternative description of power law correlations in DNA sequences, Physica A: Stat. Mech. Appl. **545**, 123735 (2020).

[31] J. P. Correia, R. Silva, D. H. A. L. Anselmo, and J. R. P. da Silva, Bayesian inference of length distributions of human DNA, Chaos, Solitons Fractals **160**, 112244 (2022).

[32] M. M. F. de Lima, R. Silva, U. L. Fulco, V. D. Mello, and D. H. A. L. Anselmo, Bayesian analysis of plant DNA size distribution via non-additive statistics, Eur. Phys. J. Plus **137**, 495 (2022).

[33] C. Beck and E. G. D. Cohen, Superstatistics, Physica A **322**, 267 (2003).

[34] C. Beck, Dynamical Foundations of Nonextensive Statistical Mechanics, Phys. Rev. Lett. **87**, 180601 (2001).

[35] S. M. D. Queirós, On the emergence of a generalised gamma distribution, Application to traded volume in financial markets, Europhys. Lett. **71**, 339 (2005).
[36] J. de Souza, L. G. Moyano, and S. M. Duarte Queirós, On statistical properties of traded volume in financial markets, Eur. Phys. J. B **50**, 165 (2006).
[37] G. Michas and F. Vallianatos, Stochastic modeling of nonstationary earthquake time series with long-term clustering effects, Phys. Rev. E **98**, 042107 (2018).
[38] A. Iliopoulos, D. Chorozoglou, C. Kourouklas, O. Mangira, and E. Papadimitriou, Superstatistics, complexity and earthquakes: A brief review and application on Hellenic seismicity, B. Geofis. Teor. Appl. **60**, 531 (2019).
[39] C. Beck, Lagrangian acceleration statistics in turbulent flows, Europhys. Lett. **64**, 151 (2003).
[40] A. M. Reynolds, Superstatistical Mechanics of Tracer-Particle Motions in Turbulence, Phys. Rev. Lett. **91**, 084503 (2003).
[41] S. Jung and H. L. Swinney, Velocity difference statistics in turbulence, Phys. Rev. E **72**, 026304 (2005).
[42] K. Ourabah, L. Ait Gougam, and M. Tribeche, Nonthermal and suprathermal distributions as a consequence of superstatistics, Phys. Rev. E **91**, 012133 (2015).
[43] S. Davis, G. Avaria, B. Bora, J. Jain, J. Moreno, C. Pavez, and L. Soto, Single-particle velocity distributions of collisionless, steady-state plasmas must follow superstatistics, Phys. Rev. E **100**, 023205 (2019).
[44] K. Ourabah, Demystifying the success of empirical distributions in space plasmas, Phys. Rev. Res. **2**, 023121 (2020).
[45] I. Rouse and S. Willitsch, Superstatistical Energy Distributions of an Ion in an Ultracold Buffer Gas, Phys. Rev. Lett. **118**, 143401 (2017).
[46] K. Ourabah, Fingerprints of nonequilibrium stationary distributions in dispersion relations, Sci. Rep. **11**, 12103 (2021).
[47] P. Jizba and H. Kleinert, Superstatistics approach to path integral for a relativistic particle, Phys. Rev. D **82**, 085016 (2010).
[48] A. Ayala, M. Hentschinski, L. A. Hernández, M. Loewe, and R. Zamora, Superstatistics and the effective QCD phase diagram, Phys. Rev. D **98**, 114002 (2018).
[49] K. Ourabah and M. Tribeche, Quantum entanglement and temperature fluctuations, Phys. Rev. E **95**, 042111 (2017).
[50] J. Cheraghalizadeh, M. Seifi, Z. Ebadi, H. Mohammadzadeh, and M. N. Najafi, Superstatistical two-temperature Ising model, Phys. Rev. E **103**, 032104 (2021).
[51] P. Jizba and F. Scardigli, Special relativity induced by granular space, Eur. Phys. J. C **73**, 2491 (2013).
[52] K. Ourabah, E. M. Barboza, E. M. C. Abreu, and J. A. Neto, Superstatistics: Consequences on gravitation and cosmology, Phys. Rev. D **100**, 103516 (2019).
[53] K. Ourabah, Generalized statistical mechanics of stellar systems, Phys. Rev. E **105**, 064108 (2022).
[54] M. I. Bogachev, O. A. Markelov, A. R. Kayumov, and A. Bunde, Superstatistical model of bacterial DNA architecture, Sci. Rep. **7**, 43034 (2017).
[55] M. I. Bogachev, O. A. Markelov, A. R. Kayumov, and A. Bunde, Correction: Corrigendum: Superstatistical model of bacterial DNA architecture, Sci. Rep. **7**, 46917 (2017).
[56] Y. Itto and C. Beck, Superstatistical modelling of protein diffusion dynamics in bacteria, J. R. Soc. Interface **18**, rsif.2020.0927 (2021).
[57] A. A. Sadoon and Y. Wang, Anomalous, non-Gaussian, viscoelastic, and age-dependent dynamics of histonelike nucleoid-structuring proteins in live *Escherichia coli*, Phys. Rev. E **98**, 042411 (2018).
[58] W. Feller, Two singular diffusion problems, Ann. Math. **54**, 173 (1951).
[59] F. D. Rouah, *The Heston Model and its Extensions in VBA* (John Wiley & Sons, New York, 2015).
[60] K. L. Howe, P. Achuthan, J. Allen, J. Allen, J. Alvarez-Jarreta, M. R. Amode, I. M. Armean, A. G. Azov, R. Bennett, J. Bhai *et al.*, Ensembl 2021, Nucleic Acids. Res. **49**, D884 (2021).
[61] J. J. Moré, The Levenberg-Marquardt algorithm: Implementation and theory, in *Numerical Analysis*, edited by G. A. Watson (Springer Berlin, Heidelberg, 1978), pp. 105–116.
[62] R. Core Team, *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria, 2022).
[63] U. von Toussaint, Bayesian inference in physics, Rev. Mod. Phys. **83**, 943 (2011).
[64] A. M. Ellison, Bayesian inference in ecology, Ecol. Lett. **7**, 509 (2004).
[65] K. H. Reckhow, Bayesian inference in non-replicated ecological studies, Ecology **71**, 2053 (1990).
[66] P. H. Boersch-Supan, S. J. Ryan, and L. R. Johnson, deBlnfer: Bayesian inference for dynamical models of biological systems in R, Methods Ecol. Evol. **8**, 511 (2017).
[67] K. E. Hines, A primer on Bayesian inference for biophysical systems, Biophys. J. **108**, 2103 (2015).
[68] H. Jeffreys, *The Theory of Probability* (Oxford University Press, Oxford, 1998).
[69] https://github.com/JohannesBuchner/MultiNest.
[70] F. Feroz, M. P. Hobson, and M. Bridges, MultiNest: An efficient and robust Bayesian inference tool for cosmology and particle physics, Mon. Not. R. Astron. Soc. **398**, 1601 (2009).
[71] F. Feroz, M. P. Hobson, E. Cameron, and A. N. Pettitt, Importance nested sampling and the multinest algorithm, Open J. Astrophys. **2**, 1 (2019).