

até a vida evoluir de sua forma mais simplista até o aparecimento de vida complexa. Este cenário sugere que outros planetas possam ter uma trajetória evolutiva similar.

Neste artigo, vamos replicar um método que descreve a evolução de moléculas baseado na ideia de autorreprodução, cópias de si mesmas, sujeitas a uma taxa de mutação (erro ao reproduzir a cadeia que carrega alguma informação)<sup>3</sup> que causa sua eliminação. Este modelo tem sua origem nos trabalhos pioneiros de Manfred Eigen<sup>4</sup> onde princípios da cinética química e conceitos da chamada teoria da informação (entropia da informação)<sup>5</sup> ditam a dinâmica evolutiva de organismos. Nesta abordagem há um limite bem claro que determina quais organismos cujas populações prosperam ou são eliminadas em função do tamanho de sua cadeia e de sua taxa de mutação.

A idade da Terra será utilizada como padrão para determinar quais configurações moleculares obtiveram sucesso no processo reprodutivo aumentando e mantendo suas populações por um período de tempo da ordem de bilhões de anos.

Na próxima seção, demonstramos o modelo de reprodução utilizado em nossas simulações e os principais resultados são apresentados. Discutimos também algumas variantes do modelo de Eigen. Fazemos nossas considerações finais na última seção.

## 2 Evolução da informação genética: O modelo de Eigen

Tão importante quanto entender os mecanismos que deram origem à vida é compreender os processos de evolução e aperfeiçoamento das formas mais primitivas que habitaram nosso planeta.

Seja um organismo primitivo com uma cadeia de informação cujo tamanho é dado por um número inteiro  $L$ . Este organismo é capaz de se

replicar dando origem a novas gerações com um número  $\sigma$  de descendentes. De acordo com Eigen a condição para que esta população sobreviva a longo prazo (critério de sobrevivência) é dada pela condição,

$$pL < \ln \sigma, \quad (1)$$

onde  $p$  é a probabilidade de acontecer um erro em qualquer posição do seu código [2]. Vamos chamar essa variável de *Probabilidade de erro em uma unidade de código ou lacuna*, ou seja, a probabilidade de se inserir o menor erro possível. De modo bem simplista, esse erro, ou mutação, pode ser entendido como uma cópia imperfeita de uma informação, quando a mesma é replicada. Em geral, mutações não são benéficas para o organismo embora que, na prática, uma certa taxa de erro no processo de reprodução é essencial para a evolução via seleção natural. Entretanto taxas muito altas de mutações tendem a levar à extinção, ou *Mutation Meltdown*.

Para ilustrar a relação de Eigen, seja, por exemplo, um organismo extremamente simples que possui sua informação codificada em uma cadeia de tamanho  $L = 2$ . Assim, existem 2 espaços onde informação pode ser depositada. O número de diferentes algarismos/letras que pode ser colocado em cada lacuna não é relevante aqui. No processo de replicação, caso a informação contida seja fielmente copiada para a próxima geração, este organismo dará origem a  $\sigma = 2$  descendentes. A população total agora será  $N = 2$  pois o indivíduo original deixou de existir. O tamanho da população crescerá obedecendo uma progressão geométrica enquanto não houver erro na cópia. A cada nova geração formada a probabilidade de erro  $p$  atua em cada uma das  $L = 2$  lacunas. Ainda neste exemplo, vamos admitir agora que a informação alocada em cada lacuna seja binária (podemos pensar em algo como ativar ou desativar certos processos bioquímicos desse organismo). Um organismo com uma cadeia (0,0) pode ter apenas os seguintes destinos: não sofrer mutação alguma e ficar como (0,0), sofrer 1 mutação na primeira ou na segunda posições e se tornar (1,0) ou (0,1), ou sofrer uma mutação na primeira posição e uma na segunda, transformando-se em (1,1). Se a probabilidade de haver um erro em cada lacuna é  $p$ , então a probabilidade de não haver erro é  $q = 1 - p$ , e assim a chance de o organismo não ser afetado pelo erro é  $qq = q^2$ , de ser afetado por um erro apenas é  $pq + qp = 2pq$

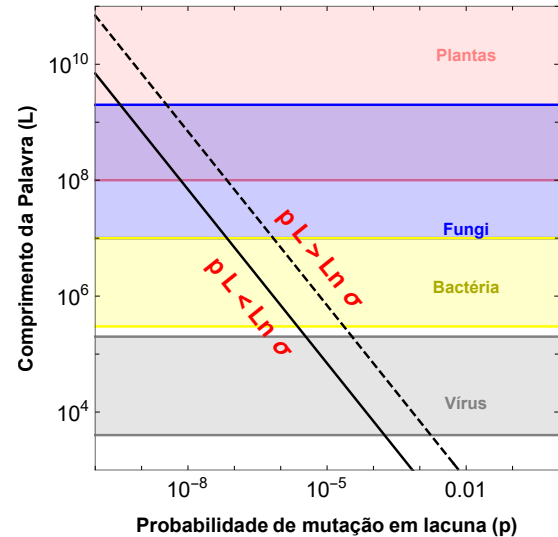
<sup>3</sup>Mesmo que a analogia com o código genético se aplique, evitamos usar tal termo por entender que o gene envolve uma complexidade bem maior do que aqui apresentamos.

<sup>4</sup>Manfred Eigen (9 de maio de 1927 – 6 de fevereiro de 2019), Prêmio Nobel de Química de 1967

<sup>5</sup>Também conhecida como entropia de Shannon. Claude Elwood Shannon (30 de abril de 1916 - 24 de fevereiro de 2001). Grosso modo, a entropia de uma variável é a “quantidade de informação” contida na variável. Nesse caso, a variável em questão é a quantidade de informação que flui no sistema.

e de ser afetado por dois erros é  $pp = p^2$ . Note que a chance total do organismo ser afetado ou não pelo erro é  $p^2 + 2pq + q^2 = (p + q)^2 = 1$  ou 100%. No caso do nosso organismo simples, se  $p = 0.1$ , então  $q = 0.9$ , e a probabilidade de não sofrer mutação é  $q^2 = 0.9^2 = 0.81$ , de sofrer uma mutação apenas é  $2pq = 0.18$  e de sofrer duas mutações é  $p^2 = 0.01$ . Assim sendo, a chance total desse organismo em ter uma cópia errada é  $p^2 + 2pq = 0.19$ . Nesse quadro, como se comportaria um organismo um pouco mais complexo, agora com uma cadeia de tamanho  $L = 5$ ? Para esse caso temos  $(p + q)^5 = p^5 + 5p^4q + 10p^3q^2 + 10p^2q^3 + 5pq^4 + q^5 = 1$ . Note que, com exceção do termo  $q^5$ , que representa a possibilidade desse organismo de não sofrer nenhum erro na sua cadeia quando esta é replicada, todos os outros termos dessa equação são probabilidades de inserção de erro, desde somente uma, até todas as posições.<sup>6</sup> Como exemplo, para esse novo organismo, a probabilidade de ocorrer erro em apenas uma posição é  $5pq^4 = 0.33$ . Note que, para o organismo mais simples descrito acima ( $L = 2$ ) a possibilidade de erro em apenas uma posição era 0.18 ou 18%, enquanto que para esse organismo mais sofisticado ( $L = 5$ ) essa mesma probabilidade é 0.33 ou 33%. Se formos levar em conta todas as mutações, e não apenas aquelas que ocorrem em uma posição, para o caso de  $L = 2$  temos  $p^2 + 2pq = 0.19$  e para  $L = 5$  temos  $p^5 + 5p^4q + 10p^3q^2 + 10p^2q^3 + 5pq^4 = 0.41$ , o que piora ainda mais o quadro dos organismos mais complexos. Isso mostra que, a medida que os organismos se tornam mais complexos, tornam-se também mais suscetíveis à introdução de erros pontuais em suas unidades de código, ou seja, em se tratando de reprodução sexuada, a natureza pode exigir dos organismos um balanço entre o quanto de erros em sua cadeia ele suporta de modo a manter a sua complexidade.

Na Figura 1, é possível visualizar a relação de Eigen (1) em um diagrama que conta como parâmetros o comprimento da palavra ( $L$ ) e a Probabilidade de erro em uma unidade de código ( $p$ ). Cada faixa horizontal compreende a região de complexidade de certos seres vivos. Vírus possuem genomas muito simples que podem ser analisados por meio de técnicas modernas de



**Figura 1:** Espaço de fase para o comprimento da palavra ( $L$ ) em função da probabilidade de erro em unidade de código ou lacuna ( $p$ ). Na região abaixo da linha sólida (tracejada) a condição de Eigen prevê a sobrevivência da população com taxa de reprodução  $\sigma = 2$  ( $\sigma = 1000$ ). Cada faixa horizontal indica a complexidade (dada em função do comprimento da palavra) de algumas formas de vida conhecidas.

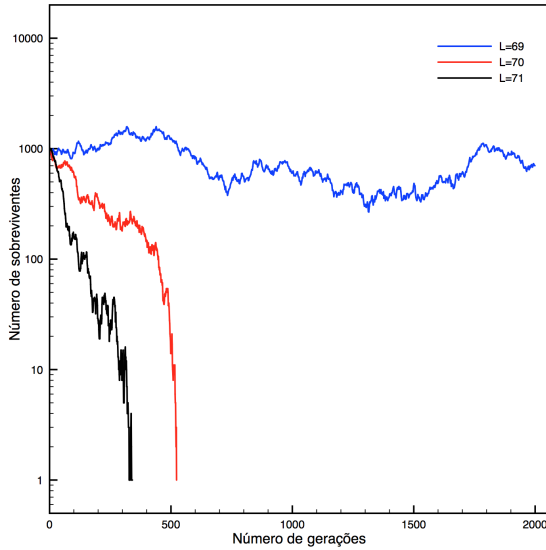
sequenciamento. Seu código genético possui um tamanho de cerca de  $\sim 10^4$ nt.<sup>7</sup> Em organismos mais complexos este número pode ser da ordem de  $10^{12}$ pb, como nos peixes. A relação de Eigen é plotada nas curvas pretas sólidas e tracejadas para organismos que geram  $\sigma = 2$  e  $\sigma = 1000$  descendentes, respectivamente. As regiões neste plano acima destas linhas pretas determinam as configurações genéticas que não sobrevivem no longo prazo de acordo com a relação (1).

A Figura 2 mostra uma simulação de 3 populações tendo cada uma delas uma cadeia de tamanho diferente,  $L = 69$ ,  $L = 70$  e  $L = 71$ , e todas inicialmente com 1000 organismos. Nesta simulação, cada nova geração originará  $\sigma = 2$  descendentes. Podemos ver que as populações com códigos maiores acumulam mais erro, extinguindo-se primeiro, e, de acordo com a Eq. (1), temos um limiar quando  $pL > 0.693$ , que define quando a população se mantém ou não.

Para o caso de um organismo genérico com informações genéticas arquivadas em uma palavra de tamanho  $L$ , podemos calcular a probabilidade  $P_M$  do organismo morrer em função de um número mínimo de erros  $M$ , sendo que esse organismo está sujeito a uma taxa de mu-

<sup>6</sup>De um modo geral, essas probabilidades podem ser calculadas pelo binômio de Newton, onde  $(p + q)^L = \sum_{k=0}^L \binom{L}{k} p^{L-k} q^k$ , e  $\binom{L}{k} = \frac{L!}{k!(L-k)!}$

<sup>7</sup>nt=nucleotídeos, usado para fita simples (RNA); bp=pares de bases, usado para fita dupla (DNA)



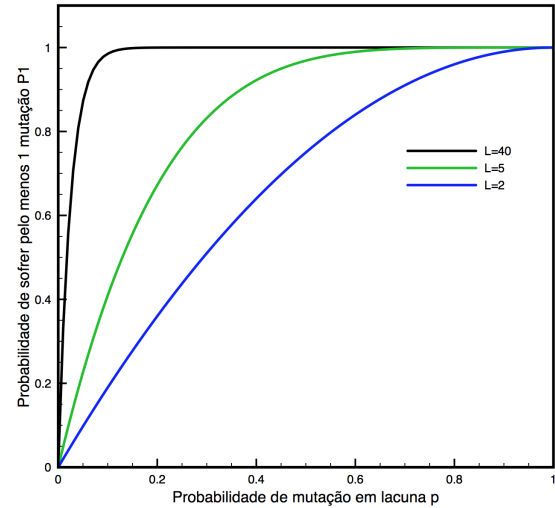
**Figura 2:** Simulação da evolução temporal para 3 populações com diferentes tamanhos de cadeias  $L = 69$ ,  $L = 70$  e  $L = 71$ , com uma taxa de mutação  $p = 0.01$ , número de filhotes  $\sigma = 2$  e população inicial de 1000 organismos. Com esses parâmetros, de acordo com (1) temos  $\ln(2) = 0.693$ , e, para  $p = 0.01$  o limiar é atingido em  $69 < L < 70$ . Simulação baseada na discussão proposta em [4].

tação  $p$  em cada unidade de informação toda vez que ele se divide. Por exemplo, se o organismo morrer com apenas um erro em sua sequência  $P_1 = 1 - (1 - p)^L$ . Caso ele suporte um erro, mas morra com dois ou mais erros temos  $P_2 = 1 - (1 - p)^L - Lp(1 - p)^{L-1}$ , e assim sucessivamente.<sup>8</sup>

A relação entre  $P_1$  e  $p$  pode ser visualizada para 3 organismos com tamanhos de cadeias diferentes através da Figura 3. Podemos notar que, apesar de  $P_1$  sempre aumentar com  $p$ , em cadeias grandes  $P_1 \rightarrow 1$  muito mais rapidamente, evidenciando o fato que organismos mais complexos são muito mais sensíveis a uma dada taxa de erro em sua cadeia. O aumento de complexidade dos organismos só foi possível com o aparecimento de estratégias de minimização e remoção de erros quando há replicação. Assim sendo, esses organismos simples devem ter durado tempo suficiente para que essas estratégias aparecessem, e é esse tempo que esse estudo tenta inferir.

A condição (1) surge como limite assintótico

<sup>8</sup>Uma técnica útil no cálculo de probabilidades é: quando parecer difícil calcular a probabilidade de um evento ocorrer, tente calcular a probabilidade dele não ocorrer, e subtraia de 1. Aqui a probabilidade de não haver erro é  $q^L = (1 - p)^L$ , e assim a probabilidade de haver erro é  $1 - q^L = 1 - (1 - p)^L$ .



**Figura 3:** Probabilidade de haver pelo menos uma mutação  $P_1$  em função de probabilidade de erro em uma unidade de código ou lacuna  $p$  para três organismos com  $L = 40$  (curva preta),  $L = 5$  (curva verde) e  $L = 2$  (curva azul).

de uma equação diferencial, construída através de técnicas de cinética química, cuja dedução completa vamos omitir nesse texto. No entanto, podemos chegar à relação de Eigen (1), a partir de uma equação de dinâmica populacional, que nada mais é que uma relação de recorrência, prevendo a variação  $\Delta N$  da população entre um instante  $t$  e um instante posterior  $t + \Delta t$ . Seja  $N(t)$  a população no instante  $t$ , e seja  $\Delta t$  o tempo necessário para ocorrer um ciclo de replicação. Se cada indivíduo dá origem a outros  $\sigma$ , indivíduos, neste intervalo ocorrem, portanto,  $\sigma N(t)$  nascimentos. Se houvessem apenas nascimentos, teríamos uma variação  $\Delta N = \sigma N(t)$  entre os instantes  $t$  e  $t + \Delta t$ . Mas é preciso lembrar que na reprodução assexuada um indivíduo dá lugar a  $\sigma$  indivíduos ( $N(t)$  mortes). Além disso, é preciso descontar as  $p_1 \sigma N(t)$  mortes por erros de replicação. O índice  $M$  em  $P_M$  indica que o organismo resiste a apenas  $M$  mutações a cada reprodução. Dito isso, chegamos a taxa de variação de indivíduos:

$$\Delta N = \sigma N(t) - N(t) - P_1 \sigma N(t). \quad (2)$$

E como  $\Delta N = N(t + \Delta t) - N(t)$  e a probabilidade de ocorrer ao menos um erro de cópia é  $P_1 = 1 - (1 - p)^L$ , chegamos à relação:

$$N(t + \Delta t) = (1 - p)^L \sigma N(t). \quad (3)$$

A Eq. (3) é o mapa iterativo de nossa dinâmica. Partindo de uma população inicial  $N(0)$

e aplicando essa relação iterativamente é possível obter o valor previsto para a população nos instantes discretos subsequentes. Na Eq. (3), é possível ver como o crescimento ou declínio da população dependem do número  $L$  de lacunas de informação e da probabilidade  $p$  de um erro de cópia ocorrer. A condição para a população crescer é que  $(1 - p)^L \sigma > 1$ , ou *mutatis mutandis*,

$$L \ln(1 - p) > -\ln \sigma. \quad (4)$$

Multiplicando-se ambos os lados da Eq.(4) por  $-1$  e lembrando que para  $p \ll 1$ ,  $\ln(1 - p) \approx -p$  chegamos à condição (1) para que a população sobreviva  $pL < \ln \sigma$ .

### 3 Mecanismo de correção de erro de replicação

A partir da análise da Figura 1 percebemos que a sobrevivência de organismos com  $L \sim 10^8$  só é possível se sua reprodução ocorre com uma taxa de erro da ordem  $p \sim 10^{-8}$  ou menor. Cadeias mais simples,  $L \sim 10^4$ , são capazes de sustentar um erro maior. A partir desta observação, nota-se que organismos modernos evoluíram aprimorando o processo reprodutivo por meio do desenvolvimento de métodos apurados de correção de erros e diminuição de sua taxa de ocorrência. Mas como isso pode ser introduzido no modelo de Eigen?

A Eq. (3) define nosso modelo de dinâmica populacional para o caso em que a ocorrência de um erro de cópia defina a não viabilidade de um descendente. Podemos reescrever esta relação de uma forma mais simples:

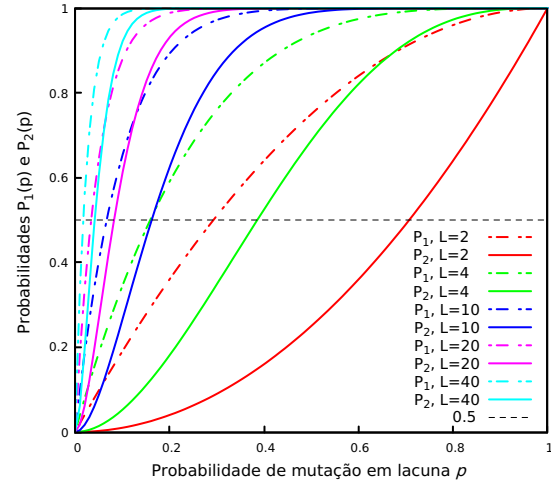
$$N(t + \Delta t) = (1 - P_1)\sigma N(t). \quad (5)$$

O que muda quando surge um mecanismo capaz de corrigir esse erro de cópia? Imaginemos que esse mecanismo fosse capaz de corrigir apenas um erro que surgisse em alguma lacuna. Isso ainda levaria à inviabilidade daquelas cópias onde ocorrem dois erros ou mais, o que ocorre com probabilidade  $P_2 = 1 - (1 - p)^L - Lp(1 - p)^{L-1}$ . Neste cenário, na Eq. (5) deve-se trocar a probabilidade  $P_1$  pela probabilidade  $P_2$ .

A Eq. (5) mostra que uma população só é viável quando

$$P_M < 1 - \frac{1}{\sigma}, \quad (6)$$

onde o subscrito  $M$  pode ser:



**Figura 4:** Comparação das curvas de  $P_1$  (linhas tracejadas) e  $P_2$  (linhas contínuas), como função de  $p$ . Cores iguais mostram  $P_1$  e  $P_2$  para organismos com um mesmo valor de  $L$ . Caminhando sobre a linha tracejada 0,5, entre as curvas azul escuro descontínua e magenta contínua, estamos numa faixa de valores de  $p$  onde não é possível existir estruturas com  $L = 10$ , se não houver mecanismo de correção ( $P_1 > 0.5$ ); mas se este mecanismo estiver presente é possível a existência de populações com  $L = 20$  ( $P_2 < 0.5$ ). Abaixo da linha horizontal tracejada existe viabilidade para  $\sigma = 2$ .

- $M = 1$ , no caso  $P_1$ , quando não pode ocorrer nenhum erro de cópia; ou
- $M = 2$ , no caso  $P_2$ , quando há mecanismo para correção de um erro.

No caso específico em que  $\sigma = 2$ , essa condição se resume a  $P_M < 1/2$ .

A Figura 4 mostra as probabilidades  $P_1$  e  $P_2$  como função de  $p$ , para alguns valores de  $L$ . A linha horizontal nos mostra a probabilidade de corte  $1/2$  do caso  $\sigma = 2$ . Comparando essas curvas para um valor específico de  $p$ , é possível notar como o mecanismo de correção de um único erro aumenta a viabilidade de populações com  $L$  significativamente maiores. Este mecanismo de correção, por exemplo, garante o surgimento de organismos de maior complexidade.

### 4 Modelo de Eigen e exoplanetologia

Nesta seção, buscamos discutir como se daria a utilização dos conceitos discutidos nas seções anteriores ao estudo de exoplanetas. Ficará claro que trata-se de um tópico ainda muito especulativo na ciência. Muito provavelmente trataremos



mais perguntas do que respostas, mas nossa intenção é conduzir o leitor à reflexão sobre alguns aspectos da astrobiologia.

É importante lembrar que o planeta Terra possui características muito especiais. Ele orbita sua estrela hospedeira, o nosso Sol, a uma distância ideal, dentro da chamada zona de habitabilidade (ZH) do sistema solar, que é definida como a região onde é permitida a existência de água em sua fase líquida. A água, em sua fase líquida, é considerada um solvente universal e meio necessário para que ocorram as reações químicas necessárias para o desenvolvimento das formas primitivas de vida. O planeta Marte, por exemplo, está no limite desta zona. Saturno e demais planetas exteriores encontram-se fora dessa região, onde seria “muito frio”. Por outro lado, Mercúrio e Vênus encontram-se na região “muito quente” e também fora da ZH. A definição da ZH de um sistema planetário é um dos principais aspectos da astrobiologia.

Outro fator importante, e muito relevante para discussão deste artigo, é a idade do sistema planetário. O nosso Sol tem uma idade aproximada de 4,5 bilhões de anos. A formação dos planetas ocorreu praticamente na mesma escala de tempo. Assim, considera-se que sistemas planetários possuem a mesma idade que sua estrela hospedeira. A Terra era um ambiente hostil para o desenvolvimento de vida em seus primeiros instantes. Estima-se que a Terra foi fortemente bombardeada por meteoritos desde sua formação até cerca de 3,9 bilhões de anos atrás. Ou seja, considera-se que em seus primeiros 600 milhões de anos o planeta Terra era estéril. E quando surgiram os primeiros organismos no nosso planeta? A análise de fósseis de estromatólitos (uma rocha fóssil formada por atividades de microrganismos em ambientes aquáticos) revela que estes organismos, que seriam os primeiros organismos terrestres, surgiram há cerca 3,7 bilhões de anos.

O objetivo da discussão feita no parágrafo anterior é revelar que o surgimento dos primeiros organismos em um planeta pode necessitar de apenas algumas centenas de milhões de anos. Uma estimativa razoável que podemos fazer é que, caso o sistema planetário necessite de cerca de 100 milhões de anos para se formar e que, após esse período, necessite de outros 100 milhões de anos de incubação até o surgimento das primeiras formas de vida, pode-se descartar sistemas planetários

extremamente jovens (estrelas hospedeiras com idades na casa de dezenas de milhões de anos) como alvos na busca de vida fora da Terra. No entanto, estrelas muito jovens são muito raras e, quando encontradas, residem em ambientes de alta formação estelar.

De fato, a vida necessita de tempo para se desenvolver. Em astrofísica existe uma relação entre a massa e idade de estrelas. Quanto mais massiva, menor seu tempo de vida. Portanto, estrelas com massas  $\sim M_{\odot}$  possuem um tempo de vida da ordem de bilhões de anos tornando-se os candidatas preferenciais.

A simulação apresentada na Figura 2 utiliza o número de gerações como variável temporal. Mas quanto tempo leva para uma população se multiplicar? Obviamente a resposta para esta pergunta dependerá do tipo de organismo. Se tomarmos como exemplo organismos mais simples, como bactérias, existem experimentos conduzidos com culturas de *Escherichia coli* cujas populações são cultivadas por décadas em laboratórios. Alguns estudos reportam taxas de reprodução destas bactérias da ordem de 60 mil gerações em 25 anos [5] e 40 mil gerações em 20 anos [6]. Assim, como valor médio, podemos inferir que a taxa de reprodução destes organismos é da ordem de 6 gerações por dia. Assumindo este resultado, e de volta à Figura 2, podemos concluir que as populações representadas pelas curvas pretas ( $L = 71$ ) e vermelhas ( $L = 70$ ) sobreviveram por períodos de semanas a poucos meses, enquanto que a população com  $L = 69$ , assim como previsto pelo critério de Eigen, terá sua sobrevivência garantida no longo prazo.

Neste artigo, discutimos os aspectos mais básicos do modelo de Eigen. Nosso objetivo é aplicar estes conceitos em alguns problemas de astrobiologia. Existem muitas perguntas ainda sem resposta. Uma proposta futura é aperfeiçoar o próprio modelo. Por exemplo, podemos pensar em micro-organismos cuja probabilidade de erro em lacuna evolui com o número de gerações e isto de fato ocorreu na história da evolução da vida na Terra [7]. Outro ponto interessante é o questionamento se este modelo pode ser aplicado em outros planetas. Em caso afirmativo, quais são as condições necessárias para um exoplaneta que simule a evolução da vida assim como ocorreu na Terra? Ainda, pretendemos simular o modelo de Eigen para valores de  $L$  e  $\sigma$  mais próximos dos

valores que conhecemos para alguns organismos como vírus e bactérias, por exemplo.

### Agradecimentos

Os autores agradecem o apoio de PROPP/UFOP, CNPq, FAPEMIG e FAPES.

### Sobre os autores

Hermano Velten ([hermano.velten@ufop.edu.br](mailto:hermano.velten@ufop.edu.br)) é doutor em física pela UFES e tem foco de suas pesquisas nas áreas de astrofísica e cosmologia.

Alcides Castro e Silva ([alcides@ufop.edu.br](mailto:alcides@ufop.edu.br)) é doutor em física pela UFMG e pesquisa temas ligados à sistemas complexos, mais precisamente ecologia e dinâmicas urbanas.

Carlos Felipe S. Pinheiro ([felipe@ufop.edu.br](mailto:felipe@ufop.edu.br)) é doutor em ciência de materiais pela UFOP e mestre em física aplicada à biologia pela USP, sempre atuando com simulação de sistemas complexos.

### Referências

- [1] B. Pearce, A. Tupper, R. Pudritz, P. Higgs, *Constraining the Time Interval for*

*the Origin of Life on Earth*, *Astrobiology* **3**, 343–364 (2018).

- [2] M. Eigen, *Selforganization of matter and the evolution of biological macromolecules* *Naturwissenschaften* **58**, 465–523 (1971).
- [3] J. Coolidge, *The story of the binomial theorem*, *The American Mathematical Monthly*, **56**, 3, 147–157 (1949).
- [4] A. Brandenburg, *Nonlinear aspects of astrobiological research*, in *Encyclopedia of Complexity and System Science*, editado por R. A. Meyers (Springer, 2009), 3284–3300.
- [5] R. Lenski, *Experimental evolution and the dynamics of adaptation and genome evolution in microbial populations*, *ISME J* **11**, 2181–2194 (2017).
- [6] J. Barrick, D. Yu, S. Yoon, et al., *Genome evolution and adaptation in a long-term experiment with Escherichia coli*, *Nature* **461**, 1243–1247 (2009).
- [7] F. J. Dyson, *Origins of life* (Cambridge University Press, Cambridge, 1999).