

Simulação de divergência genética de populações utilizando algoritmos evolutivos

Gustavo M. Costa¹, Vinicius B. Bastos², Iago A. Carvalho³

¹Departamento de Ciência da Computação
Universidade Federal de Alfenas (UNIFAL)

{gustavo.morais, vinicius.bueno, iago.carvalho}@unifal-mg.edu.br

Abstract. *Despite being two distinct areas, evolutionary computation and population genetics are research fields that have the evolutionary process as a means of studying. However, most studies in evolutionary computation do not seek to draw a parallel between the theory of population genetics and the evolutionary algorithms. Therefore, this article seeks to explore the concepts studied in population genetics in order to observe the genetic divergence in populations through algorithms used in evolutionary computing, where we conducted various computational experiments to analyze the relationship between these two fields of study. At the end of our analysis, we found that the obtained data conformed to the assumptions of population genetics, demonstrating that it is possible to establish a parallel between these two areas.*

Resumo. *Apesar de serem duas áreas distintas, a computação evolutiva e a genética populacional são campos de pesquisa que possuem como meio de estudo o processo evolutivo. Entretanto, a maioria dos estudos na computação evolutiva não buscam fazer um paralelo entre a teoria da genética populacional e os algoritmos evolucionários. Sendo assim, o presente artigo busca explorar os conceitos estudados na genética populacional a fim de observar a divergência genética em populações através de algoritmos utilizados na computação evolutiva, onde realizamos diversos experimentos computacionais para a análise da relação entre esses dois campos de estudo. Ao final de nossa análise, constatamos que os dados obtidos respeitaram os pressupostos da genética populacional, mostrando que é possível estabelecer um paralelo entre essas duas áreas.*

1. Introdução

A computação evolutiva (Mitchell 2014) e a genética populacional (Hartl and Clark 2007) são dois campos de pesquisa correlacionados que estudam o processo evolutivo. A computação evolutiva tenta replicar a dinâmica evolutiva da população em algoritmos, com foco na resolução de problemas de otimização. Por outro lado, a genética populacional busca entender a evolução de populações naturais estudando processos como mutação, seleção, recombinação e migração, entre outros. Embora tenham interesses semelhantes, a computação evolutiva e a genética populacional têm evoluído separadamente há décadas (Paixão et al. 2015).

A genética populacional é um subcampo da genética e um ramo da matemática aplicada. Ela se baseia em modelos matemáticos para estudar a dinâmica populacional e as diferenças genéticas dentro e entre populações. Esses modelos matemáticos são usados para entender como novas espécies evoluem a partir de uma espécie anterior (Girardin and Débarre 2021; Scliar et al. 2014) ou a ancestralidade genômica de uma população (Lima-Costa et al. 2015; Migicovsky et al. 2016), entre outros fenômenos.

A computação evolutiva é um subcampo da inteligência artificial. Algoritmos evolutivos baseados em população, como Algoritmos Genéticos (Mitchell 2014), Evolução Diferencial (Price et al. 2006) e Otimização de Enxame de Partículas (Shi et al. 2001), são os principais objetos de estudo da computação evolutiva. Eles são algoritmos estocásticos que empregam os princípios de recombinação, mutação e evolução natural de Darwin em uma população finita (um conjunto de soluções) para explorar um espaço de estados amplo. Esses princípios, também chamados de operadores genéticos, são aplicados iterativamente na população até que algum critério de parada pré-definido seja alcançado (Mitchell 2014).

Nos últimos anos, os algoritmos de computação evolutiva foram influenciados pela estrutura de populações naturais, levando ao desenvolvimento de Algoritmos Evolutivos Baseados em Ilhas (IBEA, do inglês *Island-based Evolutionary Algorithms*) (Tomassini 2005; Gustafson and Burke 2006). Em vez de ter uma única população panmítica grande, um IBEA possui várias subpopulações pequenas que evoluem separadamente umas das outras. Para espalhar boas soluções, essas subpopulações trocam indivíduos entre si por meio de um operador de migração. A eficiência de um IBEA está diretamente relacionada aos parâmetros desse operador de migração, como o número de indivíduos migrantes (Tomassini 2005), o intervalo entre duas migrações consecutivas (Mambrini and Sudholt 2015) e a política de migração usada (Alba and Troya 2000; Cantú-Paz 2001; Magalhães et al. 2015).

A ambição inicial do IBEA era simplesmente acelerar a convergência de algoritmos evolucionários explorando o hardware paralelo dos computadores (Alba 2005; Da Silveira et al. 2021). No entanto, mesmo sem processamento paralelo, um IBEA pode obter melhores soluções com a mesma quantidade de avaliações de função do que sua versão panmítica (Gordon and Whitley 1993; Alba and Troya 1999). Alguns aspectos teóricos da genética populacional podem

explicar esses resultados experimentais. O tamanho efetivo da população em um IBEA, dividida em várias subpopulações pequenas, é maior do que o número de indivíduos nela (Wakeley 1999; Wakeley 2000). Portanto, a diversidade genética em várias subpopulações pequenas é maior do que a diversidade genética em uma população panmítica com o mesmo número de indivíduos. Ao lidar com algoritmos evolutivos, nota-se que há uma diversidade de soluções maior em um IBEA do que em um algoritmo evolucionário com o mesmo número de soluções. Portanto, um IBEA pode explorar o espaço de soluções de maneira mais eficiente, levando a melhores soluções (Mambrini and Sudholt 2015).

Essa interseção entre computação evolutiva e genética populacional foi a principal motivação deste trabalho. Se a genética populacional pode explicar os resultados obtidos por Gordon and Whitley 1993, ela pode explicar outros aspectos dos algoritmos evolutivos? Além disso, os algoritmos evolutivos se comportam como populações naturais?

Portanto, o objetivo deste trabalho é estudar a relação entre esses campos de pesquisa. Foi feita uma revisão da literatura sobre genética populacional e algoritmos evolutivos. Em seguida, foi feito um paralelo entre aspectos teóricos da genética populacional e algoritmos evolutivos, com foco no IBEA. Várias questões surgem dessa comparação, e essas questões serão respondidas por meio de experimentos computacionais realizados em um problema de otimização de *benchmark*.

O restante deste trabalho está organizado da seguinte forma. Uma descrição biológica e genética do modelo de subpopulações é desenvolvida na Seção 2. Uma descrição computacional e uma breve revisão da literatura sobre o IBEA são apresentadas na Seção 3. A Seção 4 apresenta as funções de *benchmark* usadas nas simulações, e a descrição da heurística evolutiva do IBEA, utilizada em todas as ilhas do algoritmo. Logo após, a Seção 5 define uma série de perguntas correlacionando algoritmos evolucionários e genética de populações e a Seção 6 reporta o resultado de simulações computacionais para verificar a validade de algumas destas correlações. Por fim, as conclusões e trabalhos futuros são apresentados na última seção.

2. Uma descrição genética do modelo de subpopulação

A maioria das populações é agrupada em subpopulações menores. Por exemplo, os seres humanos estão concentrados em cidades, vilas e aldeias, os animais são subdivididos em rebanhos e bandos, as plantas são agregadas em agrupamentos. Esse agrupamento é chamado de subdivisão populacional, ou estrutura populacional, e é quase universal entre os organismos vivos. Dentro dessa subdivisão populacional, é quase inevitável obter alguma diferenciação genética entre as subpopulações (Hartl and Clark 2007; Kardos et al. 2021).

Essa diferenciação genética entre as subpopulações é a variável descritiva mais importante na genética populacional. Ela tem aplicações em muitas áreas da ciência, como biologia evolutiva e conservação, ecologia molecular, imunologia, antropologia, entre outras (Jost 2008).

Um dos aspectos mais importantes que influenciam a diversidade ge-

nética de uma população ou conjunto de subpopulações é a *endogamia*, ou seja, o acasalamento que ocorre entre indivíduos que têm parentesco familiar. A prole resultante do acasalamento entre parentes é chamada de *endogâmica* (Wright 1922; Wright 1932; Keller and Waller 2002). Indivíduos que fazem parte da mesma família têm, pelo menos, um ancestral em comum. Esse ancestral comum contribui desproporcionalmente para o genótipo¹ dos indivíduos endogâmicos. A primeira fórmula matemática que mede a endogamia de uma população foi desenvolvida por Wright 1922 e é chamada de *coeficiente de endogamia*. Essa fórmula foi posteriormente aprimorada por Cotterman 1940 e Malécot 1948 e é comumente denotada pelo símbolo F .

A diversidade genética de uma população pode ser medida por sua *frequência genotípica*, calculada como o número de indivíduos com um determinado genótipo dividido pelo número total de indivíduos na população. Na genética populacional, a frequência genotípica é equivalente à proporção de genótipos em uma população e é representada no intervalo $[0, \dots, 1]$. A endogamia afeta principalmente a diversidade genética de uma população. Quando $F = 0$, a frequência genotípica permanece inalterada durante o processo evolutivo, seguindo o princípio de Hardy-Weinberg² (Stern 1943; Roux 1974; Schaap 1980). Por outro lado, quando $F = 1$, uma população é considerada completamente endogâmica, e não há diversidade genotípica nessa população (Hartl and Clark 2007).

A endogamia causa uma redução na diversidade genética de uma subpopulação. Essa redução, chamada de *depressão endogâmica*, é simbolizada por F_{IS} (Li and Horvitz 1953; Hartl and Clark 2007) e pode ser medida pela Equação (1). A variável H_I denota a probabilidade de um indivíduo endogâmico ser heterozigoto, ou seja, um indivíduo que possui alelos diferentes³ em um determinado gene⁴. A variável H_S denota a proporção esperada de genótipos heterozigotos com acasalamento aleatório em uma subpopulação da qual I é membro.

$$F_{IS} = \frac{H_S - H_I}{H_S} \quad (1)$$

Quanto mais próxima for a relação de parentesco entre dois indivíduos, mais prejudicial é a endogamia. Ela aumenta a homozigosidade⁵ de alelos recessivos raros. Um grande número de anomalias genéticas causadas por alelos recessivos raros é observado em populações humanas devido à endogamia. A maioria dos casos é observada em filhos gerados por acasalamento entre primos de primeiro grau (Devi et al. 1987; Al-Gazali et al. 2006), pois esse é o tipo de acasalamento mais frequente entre parentes na população humana. A pro-

¹Um genótipo é a sequência de genes (o DNA) que constitui uma célula, um organismo ou um indivíduo.

²O princípio de Hardy-Weinberg afirma que o genótipo de uma população permanecerá constante de geração em geração sob certas condições.

³Um alelo é uma forma variante de um determinado gene.

⁴Um gene é uma sequência de ácidos nucleicos que faz parte de um cromossomo. Eles armazenam as características genéticas de um indivíduo.

⁵Um gene é considerado homozigoto se todos os seus alelos forem iguais.

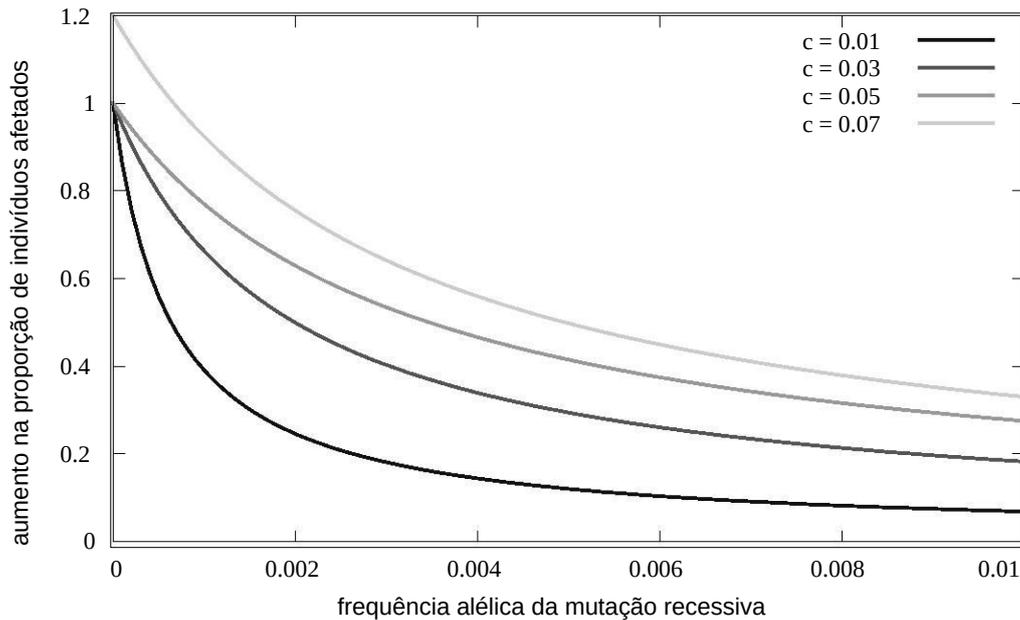


Figura 1. Influência de diferentes proporções de acasalamentos entre primos de primeiro grau (c) na proporção de crianças afetadas por distúrbios genéticos

porção crescente D é dada pela Equação (2), em que c representa a proporção de acasalamentos entre primos de primeiro grau em uma população e q representa a frequência alélica da mutação recessiva.

$$D = \frac{c(1 + 15q)}{c + 16q - cq} \quad (2)$$

A Figura 1 mostra uma representação gráfica da Equação (2). Ela mostra a proporção de crianças afetadas por doenças genéticas resultantes de acasalamentos entre primos de primeiro grau, estabelecendo uma relação entre a frequência de acasalamentos entre primos de primeiro grau em uma população e a frequência alélica de uma mutação recessiva. Pode-se observar que quanto mais rara for a mutação recessiva, maior é a influência dos acasalamentos entre primos de primeiro grau na presença de distúrbios genéticos.

Definimos como *geração* o período de tempo necessário para que os indivíduos se reproduzam e seus descendentes formem a próxima geração. Assim, de uma geração para outra, o valor de F é afetado por duas variáveis principais: (i) a frequência de endogamia; e (ii) o grau de parentesco entre os indivíduos que geraram filhos. Quanto maior for a frequência de endogamia e quanto mais próximo for o grau de parentesco, maior será o aumento do valor de F . Por outro lado, F pode diminuir se essas variáveis também diminuírem.

Outro aspecto importante relacionado à variabilidade genética nas populações é a *deriva genética*, ou seja, a mudança na frequência de uma variante genética em uma população devido aos acasalamentos. Em populações em que a

escolha dos parceiros é feita completamente ao acaso, pode-se dizer que há uma *deriva genética aleatória*. Isso leva a uma mudança aleatória na frequência das variantes genéticas em uma população de uma geração para outra e tende a causar diferenciação na frequência gênica entre as subpopulações. Diferentes subpopulações terão diferentes frequências alélicas como resultado da deriva genética aleatória (Masel 2011).

A diferenciação na frequência alélica entre as subpopulações pode ser facilmente medida pelo *índice de fixação* (F_{ST}), apresentado pela primeira vez por Wright 1922. Ele denota a probabilidade de que os alelos de um indivíduo sejam *idênticos por descendência*, *i.e.* ou seja, a probabilidade de que esses alelos sejam réplicas de outros presentes em uma geração anterior. No restante deste trabalho, F_{ST} será referido apenas como F_t , para acompanhar as mudanças de F_{ST} ao longo do tempo. F_t se refere à probabilidade de um alelo ser idêntico por descendência na geração t .

Em populações diploides, o aumento de F_t de uma geração para outra pode ser medido pela Equação (3), onde $2N$ representa o número de alelos nessa população (Wright 1922; Hartl and Clark 2007). A proporção de heterozigotos diminuirá de geração para geração. Quando $F_t = 1$, não há heterozigotos na população.

$$F_{t+1} = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right) \times F_t \quad (3)$$

A Figura 2 mostra uma representação gráfica da Equação (3) para diferentes tamanhos de população. Pode-se observar que, quanto maior o tamanho da população (N), menor é o aumento de F_t de uma geração para outra.

O *tamanho efetivo da população* em um processo de deriva genética aleatória é maior do que a soma do número de indivíduos. Para uma população dividida em várias subpopulações pequenas de tamanho igual, o tamanho efetivo da população N_e foi avaliado pela primeira vez por Wakeley 1999; Wakeley 2000. Ele pode ser expresso pela Equação (4), onde n representa o número de indivíduos em cada subpopulação, P representa o número de subpopulações e β representa a probabilidade de um alelo escolhido aleatoriamente em uma subpopulação ter sido herdado de uma das $P - 1$ subpopulações restantes. É evidente que o tamanho efetivo da população N_e é maior do que o tamanho da população nP , a menos que 4β seja muito grande. Quando isso ocorre, $\frac{1}{4\beta}$ tende a zero e $N_e \approx nP$.

$$N_e = nP \left(1 + \frac{1}{4\beta}\right) \quad (4)$$

Com essa estrutura populacional, ocorrem simultaneamente dois níveis de deriva genética aleatória. Um primeiro processo de deriva ocorre dentro de cada subpopulação, o qual ocorre relativamente rápido. O segundo processo de deriva ocorre na população como um todo e ocorre mais lentamente do que o primeiro processo de deriva (Wakeley 1999; Wakeley 2000). As subpopulações devem ser pequenas o suficiente para que a deriva genética seja importante e

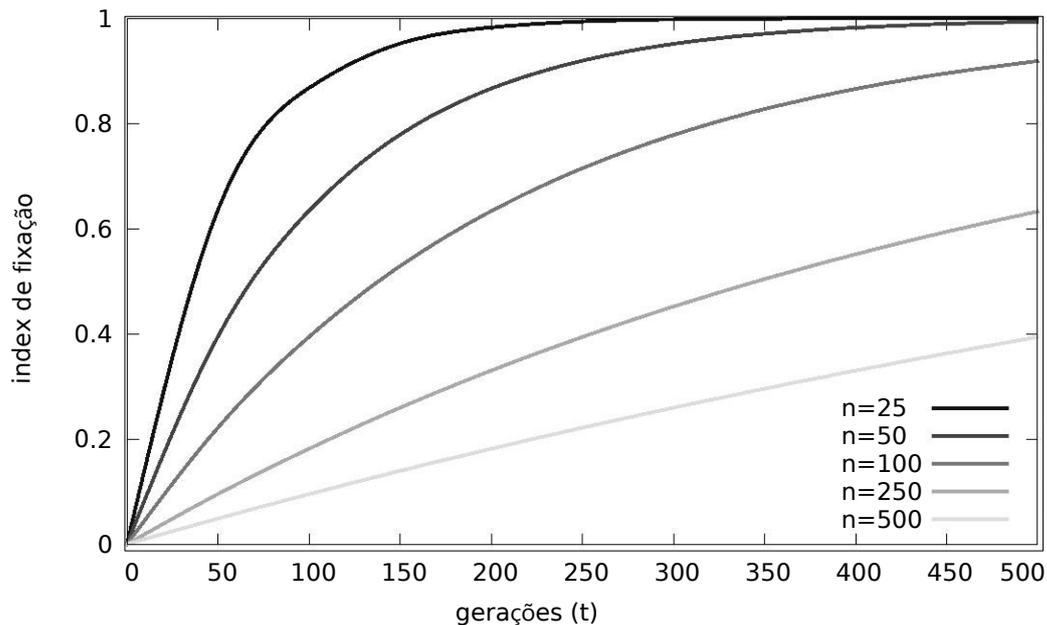


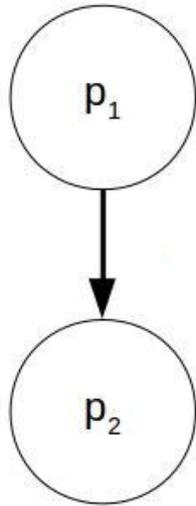
Figura 2. Aumento teórico de F_t em função do tempo para diferentes tamanhos de população

grandes o suficiente para que a seleção massal fixe combinações favoráveis de genes (Hartl and Clark 2007).

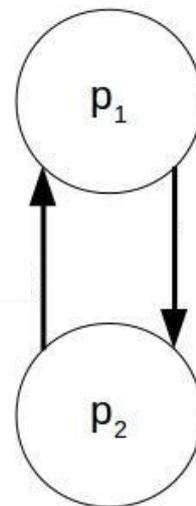
É notável que, em populações pequenas e isoladas, a depressão por endogamia pode levar a um declínio do processo adaptativo dos indivíduos, podendo até mesmo levar essas populações à extinção (Provine 1989). No entanto, raramente uma subpopulação está completamente isolada das outras. O movimento de indivíduos entre subpopulações pode ser entendido como uma *migração* e é um processo fundamental para o desenvolvimento de uma espécie (Lindenmayer and Lacy 1995; Mills and Allendorf 1996). Isso resulta em um *fluxo gênico* entre as subpopulações, reduzindo assim os efeitos da divergência genética entre elas. A migração atua como um sistema regulador evolutivo, mantendo o genótipo de todas as subpopulações mais próximas umas das outras (Provine 1989).

Embora a migração seja um processo essencial, um alto fluxo gênico entre as subpopulações pode levar a resultados indesejáveis. Essas subpopulações precisam estar suficientemente isoladas para que o primeiro nível de deriva genética possa ocorrer. Além disso, elas precisam estar suficientemente conectadas para que combinações favoráveis de genes possam se espalhar por toda a população. Quando há um grande número de subpopulações e um baixo fluxo gênico entre elas, pode levar muito tempo para que quaisquer dois alelos em diferentes subpopulações traçam suas linhagens até encontrarem um alelo comum em uma única subpopulação ancestral (Hartl and Clark 2007).

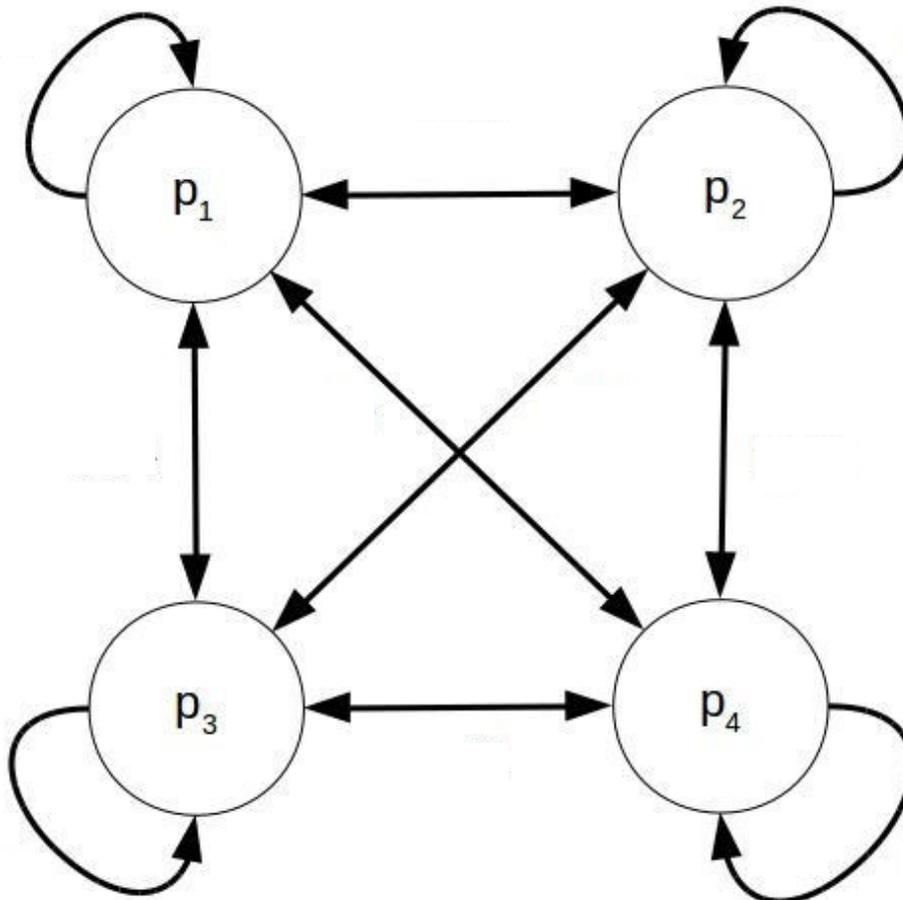
A fim de superar os efeitos prejudiciais do isolamento, os trabalhos de Lewontin et al. 1974 e Spieth 1974 propuseram a regra de um migrante por geração (OMPG, do inglês "One-Migrant-Per-Generation"). Essa regra estabelece que,



(a) Migração unidirecional



(b) Migração bidirecional



(c) Migração de ilhas

Figura 3. Diferentes tipos de migração

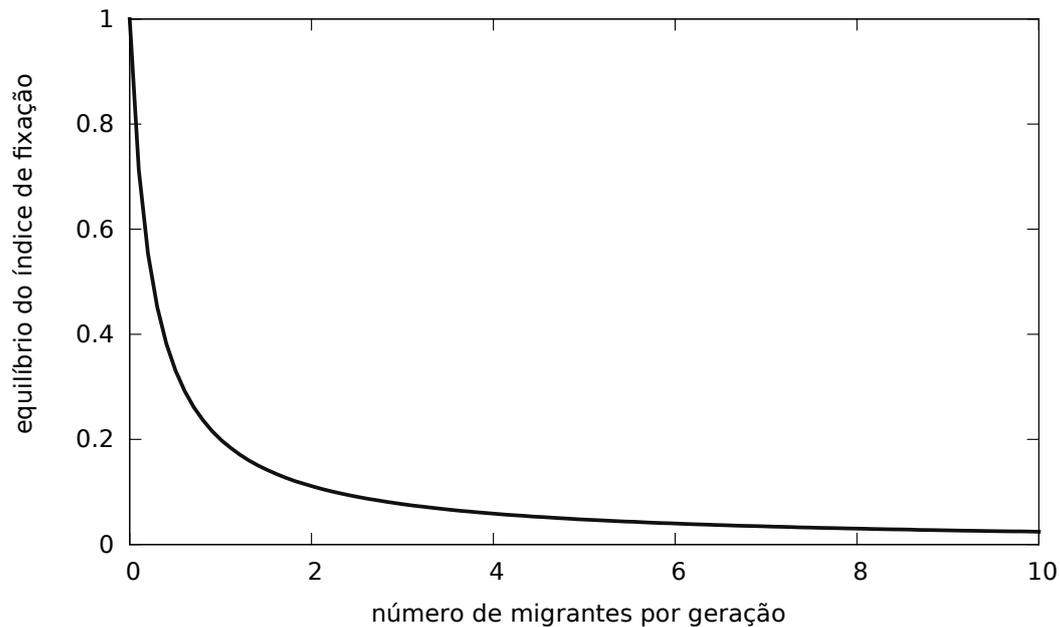


Figura 4. Diminuição do índice de fixação F_t com o aumento indivíduos migratórios por geração

com um único indivíduo migrando entre subpopulações a cada geração, é possível negar os efeitos prejudiciais da depressão por endogamia em populações pequenas e isoladas. No entanto, a OMPG não é uma regra realista. No mundo real, o processo de migração entre subpopulações ocorre em pulsos de vários indivíduos a cada várias gerações, devido a fatores sociais, comportamentais ou logísticos (Mills and Allendorf 1996). O trabalho de Lindenmayer and Lacy 1995 demonstrou, por meio de simulações computacionais, que a OMPG não era capaz de superar os efeitos prejudiciais da depressão por endogamia em populações pequenas e semi-isoladas. Além disso, também mostra que uma taxa de migração alta leva à extinção prematura, limitando a divergência genética entre as subpopulações.

Assim, é fácil perceber que a taxa de migração e o tamanho desempenham um papel importante na estrutura das subpopulações. Taxas de migração altas podem gerar uma população panmítica ampla que não apresenta diferenciação genética significativa entre as subpopulações com até 40 indivíduos cada (Lindenmayer and Lacy 1995). Um estudo desenvolvido por Mills and Allendorf 1996 mostrou que a OMPG é um mínimo aceitável que pode permitir a evolução das subpopulações, mas pode ser inadequado para muitas populações naturais. Por outro lado, mais de 10 migrações por geração podem levar à extinção prematura de muitas subpopulações, limitando a divergência genética entre elas. Assim, um mínimo de 1 e um máximo de 10 indivíduos migrantes por geração podem ser uma regra genérica apropriada para superar a depressão por endogamia e evitar a extinção prematura.

É possível definir vários tipos de processo de migração que ocorre entre as subpopulações, como ilustrado na Figura 3. A *migração unidirecional* (Figura 3(a))

(Servedio and Kirkpatrick 1997) pode ser definida como um processo de migração que ocorre de uma população para outra, sem uma quantidade relevante de migração no sentido oposto. A *migração bidirecional* (Figura 3(b)) também foi estudada por Servedio and Kirkpatrick 1997 e consiste em uma população dividida em duas subpopulações diferentes com uma quantidade equivalente de migração entre elas. Esses dois modelos de migração ocorrem sob condições específicas na natureza. No entanto, um modelo de migração mais realista é o modelo de migração de ilhas (Figura 3(c)) (Prout 1981; Petit et al. 1993; Balloux and Goudet 2002). Ele consiste em uma população de N indivíduos dividida em P subpopulações com uma taxa de migração de m_s indivíduos a cada geração entre elas.

A população humana é um exemplo de uma população dividida em várias subpopulações. Portanto, ela pode ser modelada como um modelo de ilhas. Quando cada subpopulação é grande o suficiente, o efeito do acaso genético aleatório pode ser negligenciado. Cada subpopulação recebe uma proporção igual de migrantes de cada outra subpopulação. Como o número de migrantes a cada geração é m , o número de migrantes de cada subpopulação pode ser calculado como m/P (Hartl and Clark 2007).

Assumindo que m_s é pequeno o suficiente e que N é grande o suficiente, o equilíbrio do índice de fixação F_t entre as subpopulações diminui rapidamente à medida que m_s aumenta. Assim, ele limita a divergência genética entre as subpopulações (Hartl and Clark 2007). Esse efeito é ilustrado na Figura 4, que correlaciona o equilíbrio do índice de fixação e o número de organismos migrantes por geração. Pode-se observar que esse efeito é mais drástico com um maior número de indivíduos que migram a cada geração.

Além dos pontos abordados nesta seção, um estudo da genética de populações pode ser encontrada em Hartl and Clark 2007. Fornecendo uma compreensão mais profunda da teoria.

3. Algoritmos Evolutivos Baseados em Ilhas

Um IBEA é um método simples, popular e eficaz para implementar algoritmos evolutivos em máquinas seriais e paralelas (Whitley et al. 1998). Ele utiliza um vetor P de ilhas, com um número total de N indivíduos. Assim, cada ilha tem $n_i = N/P$ indivíduos. Além dos parâmetros comuns de sua versão não distribuída, um IBEA também possui a taxa de migração (m_r), o tamanho da migração (m_s) e o número de ilhas (P). O m_r representa o intervalo de gerações entre duas migrações, e m_s denota o número total de indivíduos que migram a cada m_r gerações. Tanto o m_r , o m_s quanto o P têm grande influência na convergência do IBEA (Skolicki and De Jong 2005; Tomassini 2005).

A principal diferença entre um IBEA e um algoritmo evolucionário panmítico *PBEA*, *Panmitic Based Evolutionary Algorithm* é a separação dos indivíduos em ilhas, também denominadas subpopulações. As ilhas interagem entre si por meio da migração, espalhando informações genéticas dos indivíduos por toda a população. Sem esse operador de migração, um IBEA não passa de um conjunto de execuções separadas de um PBEA (Skolicki and De Jong 2005).

O Algoritmo 1 mostra um pseudocódigo de um IBEA geral. Primeiro, ele

Algorithm 1 Algoritmo evolutivo baseado em ilha

```
1: for  $i \leftarrow 0$  to  $P$  do
2:   inicializar população  $p_i$ 
3: end for
4: while critério de parada não é atendido do
5:   for  $j \leftarrow 0$  to  $m_r$  do
6:     for  $i \leftarrow 0$  to  $P$  do
7:       evoluir uma geração da ilha  $p_i$ 
8:     end for
9:   end for
10:  aplicar o operador de migração
11: end while
12: return melhor indivíduo de todas as ilhas
13: end
```

inicializa as ilhas $p_i \in P$ (Linhas 1 a 3). Em seguida, cada ilha evolui por m_r gerações (Linhas 5 a 9). Em seguida, o operador de migração é aplicado (Linha 10). O algoritmo é executado até que um critério de parada seja atingido (Linhas 4 a 11). Por fim, a melhor solução encontrada entre todas as ilhas é selecionada e retornada (Linha 12) e o algoritmo é encerrado. Uma versão distribuída do IBEA pode ser facilmente desenvolvida calculando o processo evolutivo (Linhas 5 a 9) em paralelo. Detalhes sobre essa implementação paralela podem ser encontrados em Alba 2005.

3.1. Trabalhos relacionados a IBEA

Vários estudos sobre o IBEA podem ser encontrados na literatura. O trabalho desenvolvido por Lässig and Sudholt 2013 demonstram o impacto dos parâmetros de migração na diversidade de soluções e na convergência do IBEA. Indica que um IBEA com uma taxa de migração alta se comporta como um PBEA. Além disso, uma taxa de migração pequena pode retardar o processo evolutivo, de modo que um maior número de avaliações de função é necessário para obter boas soluções. Além disso, também mostra que um IBEA com uma topologia densamente conectada alcança melhores resultados com intervalos de migração maiores, enquanto um intervalo de migração pequeno é indicado para ser usado com uma topologia esparsamente conectada.

Um conceito de migração por meio de cópias é descrito por Schwehm 1996. De acordo com isso, quando um indivíduo migra de uma subpopulação p_i para uma subpopulação p_j , ele continua a existir na subpopulação p_i . Em contraste, o trabalho desenvolvido por Martin et al. 1997 apresentou um processo no qual, quando um indivíduo migra de uma subpopulação p_i para uma subpopulação p_j , o indivíduo não existe mais na subpopulação p_i . No entanto, não foi encontrado nenhum trabalho na literatura que compare essas duas estratégias de assimilação da migração diferentes.

Um trabalho que trata de diferentes políticas de migração foi desenvolvido por Alba and Troya 2000. Mostrou que as políticas de migração sofrem grande influência das topologias de migração. Outro trabalho que lidou com diferentes políticas de migração e seus efeitos no IBEA foi apresentado por Cantú-Paz 2001.O

estudo apresentou diversas políticas de migração e analisou seu impacto na convergência do IBEA. Mostrou que a escolha de indivíduos aleatórios para a migração não aumenta a pressão de seleção. Portanto, pode ser apropriado para estudar outros parâmetros de migração (Skolicki and De Jong 2005)..

Outro trabalho que compara várias políticas de migração foi desenvolvido por Magalhães et al. 2015. Ele apresentou um IBEA com topologia em anel e mostrou que políticas de migração baseadas em estratégias elitistas são eficientes para resolver problemas simples. Além disso, problemas complexos são melhor tratados por políticas de migração baseadas em similaridades e comparação de aptidão.

Uma comparação entre diferentes topologias de migração foi desenvolvida por Sekaj 2004. Este trabalho analisa as topologias mais frequentemente utilizadas, como anel, matriz, grade, estrela e topologias totalmente conectadas. Mostrou-se que modelos com conexões esparsas têm uma maior diversidade de soluções, pois espalham lentamente as informações genéticas dos indivíduos para toda a população. Por outro lado, topologias densamente conectadas com uma alta taxa de migração se comportam como uma única população.

Skolicki and De Jong 2005 estudaram a influência do tamanho e do intervalo de migração no IBEA. Evidências de que o IBEA é muito sensível a essas duas variáveis foram encontradas. Mostrou-se que um intervalo de migração pequeno deve ser evitado para obter melhores resultados. Além disso, esse trabalho também correlacionou um alto tamanho de migração com a degradação do desempenho do IBEA.

Skakovski and Jędrzejowicz 2019 apresentaram um algoritmo de evolução diferencial baseado em ilhas com populações multi-tamanho. O artigo propôs um novo conceito de modelo de ilha possuindo tamanhos diferentes, bem como algoritmos de evolução diferencial que implementam esse conceito. Essa abordagem multi-tamanho facilita o projeto de algoritmos baseados em ilhas e traz uma variedade de benefícios.

A velocidade com que as boas soluções de uma ilha se espalham pela população foi estudada por Alba and Luque 2005. Experimentos computacionais realizados com diferentes intervalos e tamanhos de migração demonstraram como ajustar esses parâmetros para diferentes topologias.

Várias topologias de migração do IBEA foram estudadas e comparadas por Ruciński et al. 2010. Além das topologias de migração, este trabalho variou o número de ilhas, utilizando dois IBEA para otimizar três funções matemáticas diferentes. Foram apresentados vários parâmetros das topologias de migração, como o grau de conectividade e o coeficiente de agrupamento, demonstrando que esses parâmetros das topologias têm grande influência no IBEA.

Uma política de migração multicultural foi desenvolvida por Araujo and Merelo 2011. Um indivíduo pode migrar da população p_i para a população p_j somente se seu genótipo diferir muito do genótipo médio da população p_j . Assim, é possível aumentar a diversidade de soluções em todas as subpopulações. Essa política mostrou-se mais eficaz do que outras quando se

tem subpopulações de pequeno tamanho.

Um método para analisar o tempo de execução de algoritmos evolucionários paralelos foi apresentado por Lässig and Sudholt 2014. O método apresentado pode ser usado para estimar o limite superior do tempo de execução do IBEA por meio da generalização do tempo de execução da versão panmítica. Esse método foi validado por meio de experimentos computacionais realizados para otimizar duas funções pseudo-booleanas. Foram obtidos limites superiores de tempo de execução para o IBEA com topologia em anel, topologia em hipercubo, topologia em grafo toroidal e topologia totalmente conectada.

Um framework para desenvolvimento do IBEA para otimização multiobjetivo foi apresentado por Vargas et al. 2015. Mostrou que a dinâmica de subpopulações pode melhorar significativamente os resultados de Algoritmos Evolucionários para o caso de problemas multiobjetivo. Além disso, comparou o IBEA com outros modelos distribuídos de PBEA, mostrando que o IBEA supera esses modelos para otimização multiobjetivo.

Mambrini and Sudholt 2015 propôs duas estratégias adaptativas de migração para o IBEA. Essas estratégias ajustam o intervalo de migração analisando se as ilhas conseguiram encontrar soluções melhores desde a última migração. Se uma ilha encontrou uma solução melhorada, então seu intervalo de migração aumenta para espalhar essa solução para toda a população. Por outro lado, se uma ilha não encontrou uma solução melhorada desde a última migração, o intervalo de migração dessa ilha diminui, a fim de evitar um grande esforço de comunicação (quando se lida com hardware paralelo). Experimentos computacionais mostraram que ambos os frameworks foram capazes de reduzir o esforço de comunicação em comparação com um esquema de migração tradicional. Um trabalho similar foi desenvolvido por Osorio et al. 2011; Osorio et al. 2013. No entanto, seus algoritmos foram executados apenas por um número fixo de gerações.

Algumas decisões de implementação também podem impactar a eficiência de um IBEA. As mais importantes são a política de migração, a topologia de migração, o processo de comunicação e a assimilação da migração. Uma política de migração descreve os critérios usados para escolher os indivíduos a serem migrados. Existem inúmeras políticas de migração descritas na literatura. Por exemplo, o melhor indivíduo de cada população pode ser selecionado para migrar (Tomassini 2005), ou o indivíduo com o genótipo mais diferente pode ser escolhido (Araujo and Merelo 2011). As políticas de migração têm influência na pressão de seleção quando se trata do processo de migração (Alba and Troya 2000; Cantú-Paz 2001; Magalhães et al. 2015).

Uma topologia de migração descreve como os indivíduos migrantes se movem entre as ilhas. Muitas topologias de migração foram relatadas na literatura. As mais proeminentes são a topologia de anel, a topologia de grade e a topologia totalmente conectada (Sekaj 2004; Tomassini 2005). Outra escolha importante é usar uma topologia dinâmica (Lin et al. 1994), em vez de uma fixa. Isso pode ser usado para aumentar a exploração no IBEA, de modo que um indivíduo

migre com base em algum critério, como a distância genotípica entre as ilhas, por exemplo.

Dois processos de comunicação diferentes podem ser usados ao desenvolver um IBEA paralelo. Uma implementação *síncrona* evolui todas as ilhas juntas, e o operador de migração é executado a cada m_r gerações. Enquanto isso, uma implementação *assíncrona* também é possível. Nesse modelo de comunicação, indivíduos de diferentes ilhas podem migrar em diferentes gerações. Outra implementação assíncrona possível é quando uma ilha não espera que as outras ilhas terminem suas m_r gerações para migrar seus indivíduos. Implementações assíncronas podem resultar em ilhas com um número flutuante de indivíduos, em vez de um número fixo de indivíduos de um modelo de comunicação síncrono (Cantú-Paz 1998; Alba 2005; Tomassini 2005).

Por fim, é possível discutir duas estratégias diferentes de assimilação da migração, a saber, as estratégias baseadas em cópia e baseadas em transferência. Na estratégia baseada em cópia, quando um indivíduo migra, ele continua a existir em sua ilha original. Em contraste, quando um indivíduo migra na estratégia baseada em transferência, o indivíduo é removido de sua ilha original e inserido em sua nova ilha (Schwehm 1996; Martin et al. 1997).

Apesar do grande volume de trabalhos que lidam com algoritmo evolutivos, até onde sabemos, não há nenhum trabalho na literatura que se preocupe com a teoria genética por trás do IBEA. Portanto, o objetivo deste trabalho é apresentar uma relação entre a genética populacional e essa classe de algoritmos. Em seguida, apresentamos vários experimentos computacionais para simular o comportamento de populações naturais como algoritmos evolucionários. Esses resultados podem ser usados como uma referência quando se procura desenvolver um algoritmo evolucionário, ajudando na definição dos parâmetros do algoritmo.

4. Implementação computacional

Nesta seção apresentamos detalhes da implementação computacional utilizada neste trabalho. Primeiramente, a Seção 4.1 apresenta as funções de *benchmark* utilizadas em nossas simulações computacionais. Logo após, a Seção 4.2 apresenta a heurística evolutiva utilizada no IBEA. Vale a pena ressaltar que todas as ilhas de nosso IBEA utilizam o mesmo algoritmo apresentado nesta seção.

4.1. As funções objetivo de *benchmark* e sua representação computacional

Nós utilizamos 14 problemas diferentes de otimização global, aqui denominados como funções objetivo de *benchmark*. Essas funções objetivo de *benchmark* foram previamente empregadas na competição IEEE CEC de 2015 sobre otimização de parâmetro real de objetivo único baseado em aprendizado (Liang et al. 2014).

A Tabela 1 apresenta as funções empregadas. A primeira e segunda colunas apresentam, respectivamente, o número e o nome da função. A terceira coluna fornece o número de dimensões d da função objetivo de *benchmark*. Finalmente, a quarta coluna denota o valor ótimo da função objetivo. Pode-se observar que o valor de d varia no conjunto $\{10, 30\}$ para todas as funções objetivo de *ben-*

chmark. Quanto maior o valor de d , mais difícil se torna a resolução da função objetivo de *benchmark*.

Tabela 1. Funções objetivo de *benchmark* avaliadas neste trabalho

Número	Nome da função	d	f_{min}
F1	Rotated Bent Cigar Function	10, 30	100
F2	Rotated Discus Function	10, 30	200
F3	Shifted and Rotated Weierstrass Function	10, 30	300
F4	Shifted and Rotated Schwefel's Function	10, 30	400
F5	Shifted and Rotated Katsuura Function	10, 30	500
F6	Shifted and Rotated HappyCat Function	10, 30	600
F7	Shifted and Rotated HGBat Function	10, 30	700
F8	Shifted and Rotated Expanded	10, 30	800
F9	Shifted and Rotated Expanded Scaer's F6 Function	10, 30	900
F10	Hybrid Function 1 (N = 3) Rastrigin	10, 30	1000
F11	Hybrid Function 2 (N = 4) Rastrigin	10, 30	1100
F12	Hybrid Function 3 (N = 5)	10, 30	1200
F13	Composition Function 1 (N = 5)	10, 30	1300
F14	Composition Function 2 (N = 3)	10, 30	1400

Denotamos uma solução para uma função objetivo de *benchmark* como um indivíduo p_m^i de um IBEA. Em nosso problema, uma solução é representada como um vetor d -dimensional de números reais no intervalo $[-100, 100]$. A Figura 5 representa uma solução para $d = 10$. Denotamos cada posição deste vetor como um *gene*. Além disso, denotamos o valor obtido ao se aplicar a solução p_m^i na resolução de uma função objetivo de *benchmark* proposta como seu *valor de aptidão* $f(p_m^i)$.

3.2	7.4	9.1	0.9	1.5	0.5	8.8	3.2	5.8	4.6
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

Figura 5. Representação de solução para uma função de otimização de *benchmark* com $d = 10$

4.2. Algoritmo Genético

(Holland et al. 1992) propôs o Algoritmo Genético (GA, do inglês *Genetic Algorithm*), inspirado pela teoria da evolução apresentada por Charles Darwin. O GA atualiza sua população com base nos operadores de cruzamento, mutação e evolução. Ele foi inicialmente proposto para representação binária dos indivíduos. No entanto, diversas variações apresentadas na literatura estenderam sua aplicação para lidar com variáveis de valor real (Reeves 2010), como neste trabalho.

O Algoritmo 2 mostra o pseudo-código da nossa implementação de GA. Primeiro, a população é inicializada na linha 1 e avaliada na linha 2. Em seguida, o loop das linhas 3-8 é executado até que um critério de parada seja atendido. Cada iteração desse loop é denominada uma *geração* do GA. Em cada iteração,

os operadores de seleção, cruzamento, mutação e evolução são aplicados sequencialmente. Finalmente, o melhor indivíduo encontrado pelo GA é retornado na linha 9.

Algorithm 2 Algoritmo Genético (GA)

- 1: Inicialização da população
 - 2: Avaliação da população
 - 3: **while** critério de parada não é atendido **do**
 - 4: Seleção
 - 5: Cruzamento
 - 6: Mutaç o
 - 7: Evoluç o
 - 8: **end while**
 - 9: **return** O melhor indiv duo encontrado
-

4.2.1. Operador de Seleç o

O operador de seleç o obt m um ou dois indiv duos aleat rios para passar pelo processo de evoluç o. Primeiro, ele seleciona um indiv duo aleat rio da populaç o p_m usando um algoritmo de roleta (Reeves 2010). Em seguida, gera um n mero pseudo-aleat rio ι usando uma distribuiç o uniforme real $U(0, 1)$. Se ι for menor ou igual ao par metro da taxa de cruzamento, ent o o indiv duo selecionado anteriormente   usado como entrada para o operador de mutaç o, n o aplicando assim o operador de cruzamento. Caso contr rio, seleciona um indiv duo diferente usando o mesmo algoritmo de roleta e emprega ambos os indiv duos como entrada para o operador de cruzamento.

4.2.2. Operador de Cruzamento

O GA emprega o cl ssico operador de *cruzamento de 2 pontos* (Reeves 2010). Este operador mistura os genes dos indiv duos selecionados e cria um novo indiv duo. Este novo indiv duo ser  usado como entrada para o operador de mutaç o.

4.2.3. Operador de Mutaç o

O operador de mutaç o altera o valor de um  nico gene do indiv duo gerado anteriormente. O gene a ser mutado   escolhido usando uma distribuiç o uniforme inteira $U(1, d)$ e seu novo valor   gerado da mesma forma que na inicializaç o da populaç o, isto  , usando uma distribuiç o uniforme cont nua $\mathcal{U}(-100, 100)$. Esta operaç o   aplicada com uma certa probabilidade regulada pela taxa de mutaç o, um par metro do algoritmo.

4.2.4. Operador de Evolução

Os operadores anteriores são aplicados até que o número de indivíduos gerados seja igual ao tamanho original da população. O operador de seleção constrói um vetor temporário de soluções contendo a população original e todos os indivíduos gerados. Em seguida, seleciona metade deles para passar para a próxima geração.

Este processo de seleção simplesmente ordena o vetor temporário de soluções de acordo com seu *fitness*. Então, os indivíduos com os melhores valores de *fitness* são selecionados e os outros são descartados. Pode-se observar que este operador preserva o tamanho original da população.

5. Planejamento experimental

Os experimentos computacionais foram realizados em um computador pessoal utilizando o sistema operacional Ubuntu 22.04.1. Os experimentos foram executados em um núcleo de um processador Intel i5-12500 com clock de 4,6 GHz e 16 GB de RAM. Todos os algoritmos foram desenvolvidos do zero em C e compilados usando o GNU GCC 11.3.0. Além disso, todos os experimentos foram repetidos 30 vezes com diferentes sementes para o gerador de números pseudo-aleatórios Mersenne-Twister (Matsumoto and Nishimura 1998) para a obtenção dos resultados. Definimos três critérios de parada diferentes: (i) um tempo máximo de execução de 3600 segundos; (ii) um total de 500 gerações sem melhoria no valor da função objetivo; e (iii) um valor inferior ou igual a 100 para o cálculo da divergência genética entre as ilhas.

5.1. Correlações entre computação evolucionária e genética de populações

Há mais de quarenta anos, John Holland foi inspirado pela teoria da seleção natural de Darwin e desenvolveu a base do que viria a ser a Computação Evolucionária. Foi a primeira tentativa de imitar o processo evolutivo com um algoritmo. Em seguida, foram desenvolvidos e aplicados muitos algoritmos inspirados na natureza em quase todas as áreas do conhecimento. Apesar de suas origens na teoria de Darwin, os algoritmos evolucionários não foram influenciados em grande medida pela genética evolutiva nos anos seguintes.

Utilizando a teoria apresentada na Seção 2, é possível estabelecer uma relação entre alguns resultados experimentais encontrados na literatura de algoritmos evolucionários e a teoria da genética populacional. O trabalho Lässig and Sudholt 2013, por exemplo, explora essa conexão em relação à migração de pequena e grande escala. Além disso, existem outros trabalhos que também exemplificam essa relação.

Também é possível estabelecer um paralelo entre a computação evolucionária e a genética populacional. Os algoritmos evolucionários são principalmente utilizados em problemas de otimização. Tanto a genética populacional quanto a teoria da otimização compartilham um aspecto comum: a *paisagem de aptidão*, um espaço subjacente de combinações genéticas possíveis dos indivíduos em um espaço n -dimensional, no qual as combinações entre genes geralmente resultam em

topografias adaptativas complexas com muitos picos e vales. Uma população inteira pode ser mapeada nessa estrutura. O processo evolucionário busca gerar os indivíduos mais adaptados, que são encontrados no ponto máximo dessa paisagem de aptidão (Martin et al. 1997; Hartl and Clark 2007; Reid et al. 2021). Da mesma forma, um problema de otimização (maximização) busca a solução mais adequada, que é encontrada no ponto mais alto da paisagem de aptidão. Um algoritmo evolucionário emprega uma série de operações estocásticas entre indivíduos de sua população para buscar essa solução mais adequada (Talbi 2009).

No entanto, quase toda a teoria da genética populacional é baseada na ideia de indivíduos diploides, e a maioria dos algoritmos evolucionários emprega indivíduos que possuem apenas um alelo por gene, sendo assim indivíduos haploides. Portanto, a teoria descrita na Seção 2 não pode ser aplicada diretamente aos algoritmos evolucionários. No entanto, como mencionado acima, alguns resultados experimentais da computação evolucionária podem ser explicados pela genética populacional, de modo que os algoritmos evolucionários se comportam como populações do mundo real em alguns aspectos. Assim, este trabalho tem como objetivo preencher a lacuna entre essas áreas de pesquisa. Argumentamos que os algoritmos evolucionários, especialmente os Algoritmos Evolucionários Baseados em Ilhas, se comportam como populações naturais em outros aspectos, apesar do mencionado acima. Com base nisso, as seguintes hipóteses são levantadas e vários experimentos computacionais foram realizados para investigar essa relação.

1. Quando existe uma IBEA com um pequeno número de indivíduos em cada ilha, a migração de um único indivíduo por geração nega os efeitos deletérios da depressão endogâmica, como afirmam Lewontin et al. 1974 e Spieth 1974?
2. As ilhas pequenas (que têm 4-10 indivíduos) são gravemente afetadas pelas elevadas taxas de migração entre elas, como afirma Lindenmayer and Lacy 1995?
3. Com uma população dividida em várias ilhas de 40 indivíduos cada, será que uma taxa de migração elevada induz uma população panmítica, como afirma Lindenmayer and Lacy 1995? Além disso, quantas gerações são necessárias para atingir esta população panmítica?
4. Será que várias ilhas isoladas têm uma divergência genética maior do que uma população panmítica, como afirmado em Hartl and Clark 2007?
5. Como referido por Lindenmayer and Lacy 1995; Mills and Allendorf 1996; Hartl and Clark 2007, a migração reduz a divergência genética entre ilhas, de tal forma que o número de migrantes por geração é diretamente inverso à divergência genética das ilhas. Em que proporção o número de migrantes por geração reduz a divergência genética entre as ilhas?
6. A regra OMPG (Lewontin et al. 1974; Spieth 1974) pode levar a uma extinção prematura da população, como afirma Lindenmayer and Lacy 1995. Além disso, mais de dez migrantes por geração também podem ter um impacto negativo no processo evolutivo, tal como referido por Mills and Allendorf 1996. Serão estes limites também aplicáveis aos algoritmos evolutivos?

5.2. Métricas de avaliação

Para a avaliação da diversidade genética nas populações foram obtidas duas medidas a cada geração, sendo elas a Distância Populacional e a Distância Mundial. O cálculo dessas duas medidas depende do cálculo da distância euclidiana, que pode ser descrita como

$$distEuclidiana(x, y) = \sqrt{\sum_{i=1}^d (x_i - y_i)^2},$$

onde x e y são dois indivíduos, i representa o i -ésimo gene do indivíduo e d a dimensão do problema.

Sendo assim, a Distância Populacional é obtida através do cálculo da distância euclidiana entre todos os indivíduos de uma mesma população. Esse cálculo é realizado para todas as populações do IBEA e, ao fim, é calculada a média pelo número total de ilhas.

Já para a Distância Mundial, seja p_1 e p_2 duas populações. Além disso, seja c_i os indivíduos pertencentes a p_1 e k_j os indivíduos pertencentes a p_2 . Podemos calcular a distância entre duas populações p_1 e p_2 como

$$distancia(p_1, p_2) = \sqrt{\sum_{i \in P_1} \sum_{j \in P_2} distEuclidiana(c_i, k_j)}.$$

Deste modo, a Distância Mundial é obtida calculando-se a soma da distância entre todo par de populações do IBEA.

6. Simulações computacionais

Nesta seção nós apresentamos os experimentos computacionais de nosso trabalho. Primeiro, a Seção 6.1 apresenta um experimento inicial onde é realizado a otimização de parâmetros de nosso GA. Logo após, as seções subsequentes apresentam os resultados de simulações computacionais que tentam responder algumas das perguntas levantadas na Seção 5.1.

6.1. Otimização de Hiperparâmetros

Neste primeiro experimento, otimizamos os parâmetros para o algoritmo genético descrito na Seção 4.2. Este experimento foi realizado para uma única população, ou seja, sua implementação PBEA. O objetivo foi encontrar o melhor conjunto de parâmetros para o algoritmo e utilizá-lo nos experimentos subsequentes.

A otimização dos hiperparâmetros foi realizada usando as funções objetivo de *benchmark* $F3$, $F5$ e $F12$. Essas três funções foram selecionadas aleatoriamente. O pacote iRace (López-Ibáñez et al. 2016) foi utilizado para este experimento. Este pacote oferece um conjunto conhecido para configuração automática de algoritmos em tarefas de otimização e aprendizado. Utilizamos o método *Iterated Race* disponível no pacote iRace desenvolvido na linguagem de programação R.

A Tabela 2 mostra os resultados deste experimento. A primeira coluna apresenta o nome do algoritmo, enquanto a segunda coluna indica o nome do parâmetro que está sendo otimizado. A terceira coluna mostra o intervalo considerado, ou seja, o valor mínimo e máximo avaliado para cada parâmetro. Finalmente, a quarta coluna apresenta os resultados obtidos através da aplicação do procedimento Iterated Race. Como estes são os melhores parâmetros, eles serão utilizados nos experimentos subsequentes.

Tabela 2. Experimento de otimização de hiperparâmetros

Algoritmo	Parametro	Intervalo de teste	Melhor valor
GA	Tamanho da População	$[30, 10000] \in \mathbb{N}$	3229
	Taxa de Mutação	$[0, 1] \in \mathbb{R}$	0.13
	Taxa de Cruzamento	$[0, 1] \in \mathbb{R}$	0.08

6.2. Aplicação da regra OMPG em ilhas pequenas

Para esse experimento foi realizado um comparativo entre IBEA's com $n \in \{2, 5, 10, 25, 50\}$ ilhas e $p \in \{4, 10, 20\}$ indivíduos por ilha em dois cenários. No primeiro não realizamos migrações entre as ilhas. No segundo aplicamos a regra OMPG. Desse modo, o objetivo desse experimento consiste em verificar se a regra OMPG é capaz de evitar a depressão por endogamia em populações pequenas. Com base nos resultados obtidos, concluímos que um único migrante por geração não é capaz de evitar a depressão endogâmica em populações pequenas. Para os gráficos considere M_i como sendo o número de migrantes por geração e P_j o número de indivíduos em cada ilha.

Analisando a Figura 6 vemos que, de início, as ilhas isoladas e as ilhas OMPG possuem uma distância genética similar devido a inicialização randômica dos indivíduos em ambos os cenários. Com o passar das gerações as ilhas no cenário OMPG tendem a uma rápida aproximação genética, caminhando para uma divergência genética igual a zero. As ilhas isoladas também tendem a se aproximar geneticamente, mas de uma maneira muito mais lenta quando comparado as ilhas no cenário OMPG, tendendo a uma estagnação. Isso ocorre porque no cenário onde as ilhas estão isoladas, com o passar das gerações, cada ilha tende a convergir para um ótimo local diferente. Durante as primeiras gerações elas tendem a ficar próximas geneticamente de forma célere. Entretanto, após esse intervalo a aproximação genética começa a ficar cada vez mais lenta devido a depressão endogâmica sofrida localmente por cada ilha isolada, que pode ser observada na Figura 7. Nesse ritmo, as ilhas isoladas ficariam completamente estagnadas, não fosse pelo operador de mutação empregado no Algoritmo Genético, que sempre atribui uma variação genética às ilhas, permitindo que elas continuem a explorar o espaço de busca, mesmo que de maneira extremamente lenta.

Por outro lado, observando o cenário OMPG, devido ao fluxo gênico entre as ilhas, a variabilidade genética observada em cada ilha localmente (Figura 7) é maior comparado ao cenário sem migrações, impedindo momentaneamente a depressão endogâmica. Entretanto, terminada a troca inicial de genes, as ilhas

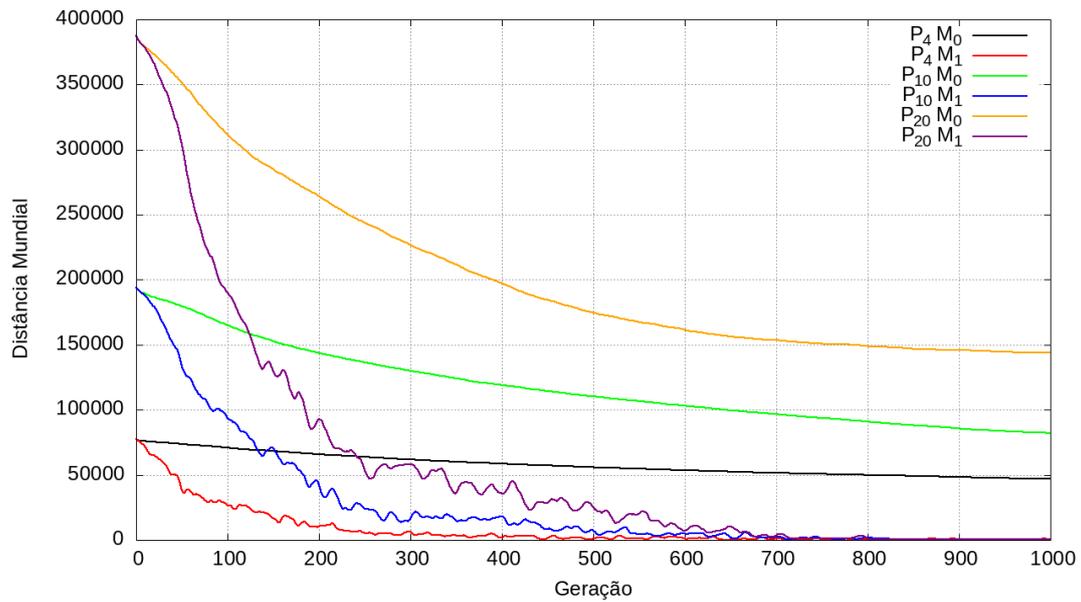


Figura 6. Comparação entre a Distância Mundial de IBEA's com 50 Ilhas para a Função 4

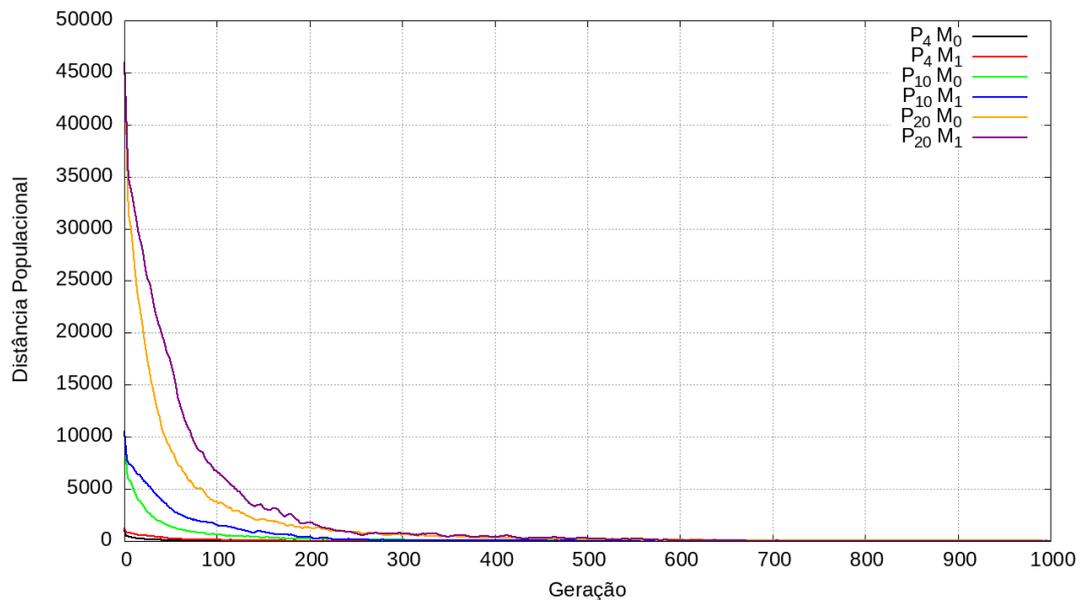


Figura 7. Comparação entre a Distância Populacional de IBEA's com 50 Ilhas para a Função 4

ficam geneticamente próximas. Desse modo, a população sofre rapidamente com a depressão endogâmica. Como várias ilhas agora estão varrendo o mesmo local no espaço de busca juntas, a tendência é que haja uma convergência mais rápida para um mesmo ótimo local entre as ilhas. Podemos observar nas Figuras 8 e 9 que quanto maior o número de ilhas, mais rápida é a convergência da população para um ótimo local quando comparado ao cenário sem migrações."

Sendo assim, concluímos que a OMPG é capaz de postergar a depressão

endogâmica em ilhas pequenas, proporcionando um fluxo gênico entre as ilhas, assim impedindo a rápida estagnação das mesmas. Entretanto, esse processo as torna geneticamente próximas, e devido à baixa variabilidade genética presente nas ilhas pequenas elas acabam sofrendo com a depressão endogâmica.

Em geral, a OMPG não se mostra eficiente para evitar a depressão por endogamia em populações pequenas porque desde o início essas populações já possuem uma baixa variabilidade genética e a migração acaba por acelerar o processo de depressão endogâmica.

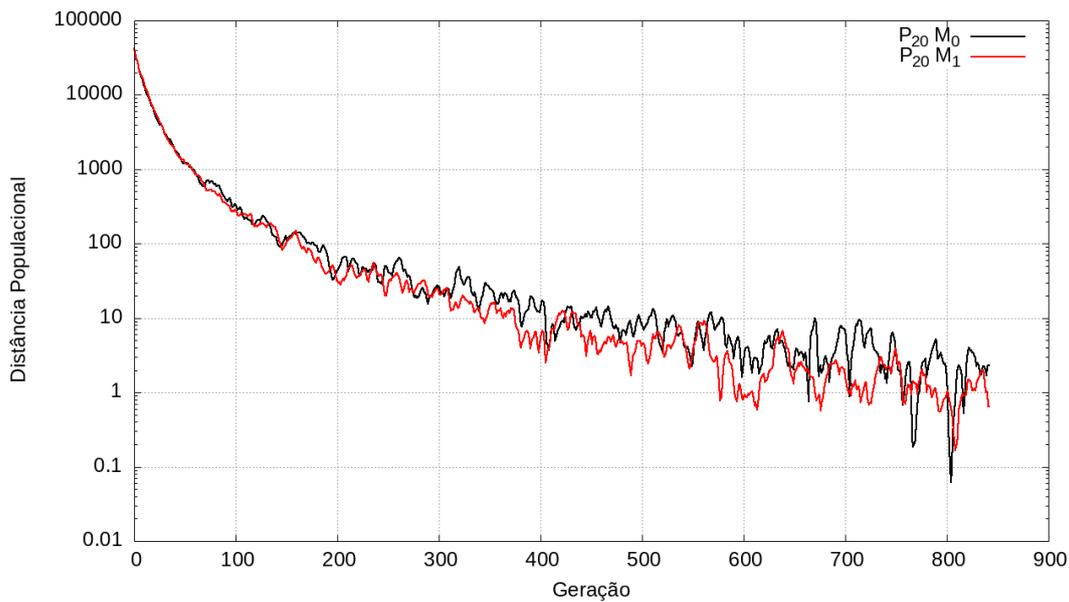


Figura 8. Comparação entre a Distância Populacional de IBEA's com 2 Ilhas para a Função 1

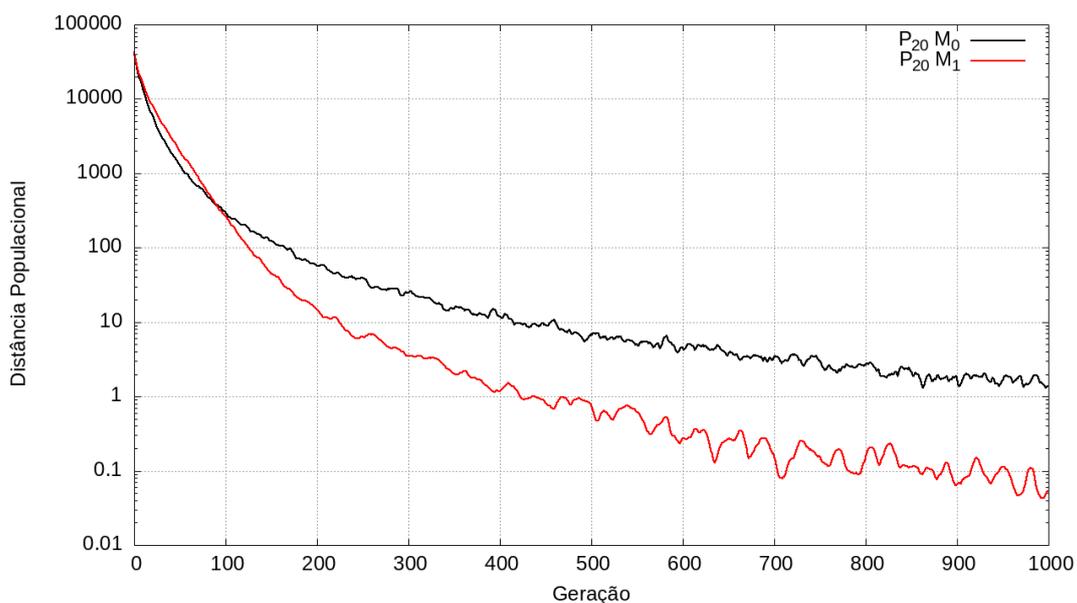


Figura 9. Comparação entre a Distância Populacional de IBEA's com 50 Ilhas para a Função 1

6.3. Ilhas pequenas com altas taxas de migração

Nesse experimento realizamos um comparativo entre IBEA's com diferentes taxas de migração a cada geração. Foram utilizados IBEA's com $n \in \{2, 5, 10, 25, 50\}$ ilhas, $p \in \{4, 10, 20\}$ indivíduos por ilha e uma proporção de $m \in \{0\%, 5\%, 15\%, 25\%, 50\%\}$ migrantes a cada geração. O objetivo consiste em observar o impacto das altas taxas de migração em populações pequenas. Com a análise dos dados obtidos para esse experimento, concluímos que as ilhas pequenas são gravemente afetadas pelas altas taxas de migração. Para os gráficos considere M_i como sendo o número de migrantes por geração e P_j o número de indivíduos em cada ilha.

Ao observarmos a Distância Populacional (Figura 10) para verificar a diversidade genética nas ilhas, percebe-se que de início quanto maior o número de migrantes, maior é a diversidade genética entre os indivíduos. Isso ocorre devido a troca inicial de genes. Entretanto, após algumas gerações, observa-se que quanto maior for a taxa de migração, maior é o efeito da depressão endogâmica na população.

Já para a Distância Mundial (Figura 11) percebemos que o grande fluxo gênico entra as ilhas causado pela alta taxa de migração acaba por acelerar a convergência do Algoritmo Genético, acelerando assim o processo de depressão endogâmica na população.

Nas primeiras gerações essas tendências são claramente visíveis. No entanto, após esse período, o comportamento das diferentes taxas de migração tende a se tornar semelhante, pois as ilhas já se encontram em um estado avançado de depressão endogâmica. Esse comportamento é devido ao fato de que as ilhas possuem poucos indivíduos e, portanto, desde o início, apresentam pouca diversidade genética.

Concluímos então que uma alta taxa de migração em populações pequenas acaba por acelerar o processo de convergência do algoritmo, acelerando também o processo de depressão endogâmica sofrido pela população.

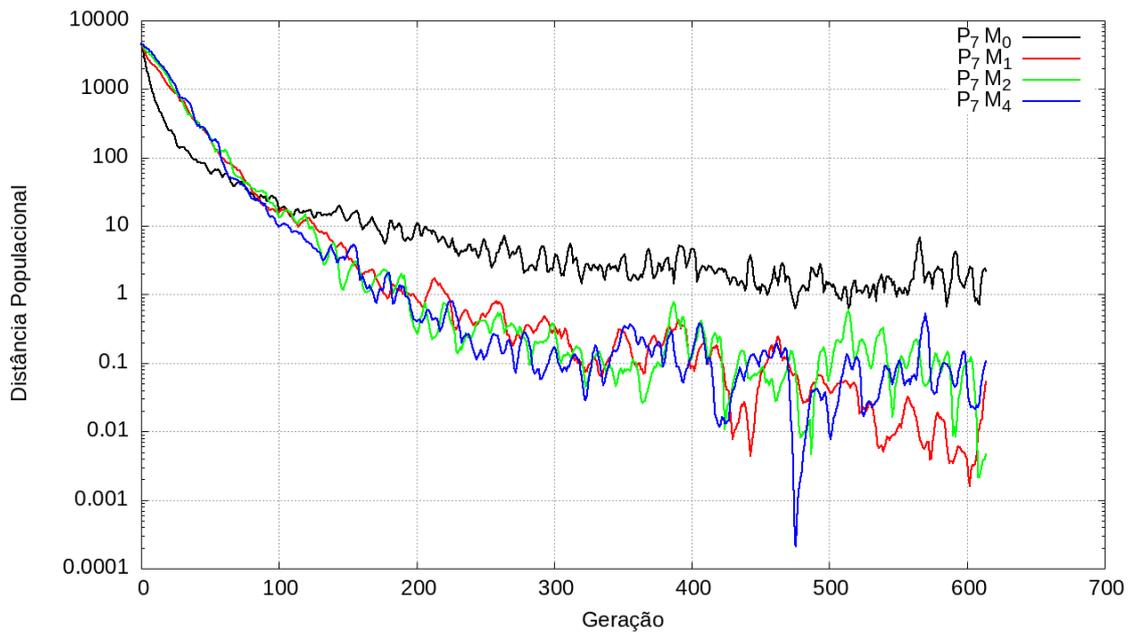


Figura 10. Comparação entre a Distância Populacional de IBEA's com 25 Ilhas para a Função 3

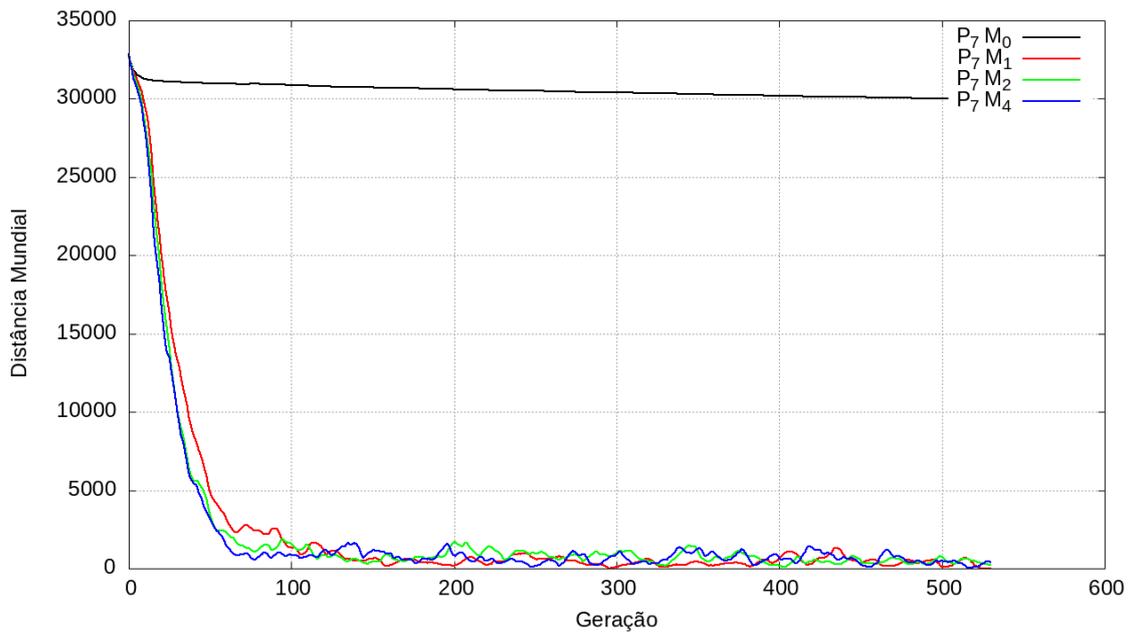


Figura 11. Comparação entre a Distância Mundial de IBEA's com 25 Ilhas para a Função 2

6.4. Avaliação entre uma população panmítica e um IBEA com elevadas taxas de migração

Nesse experimento utilizamos para comparativo IBEA's com $n \in \{2, 5, 10, 25, 50\}$ ilhas e $p \in \{40\}$ indivíduos por ilha com uma migração de $m \in \{1, 2, 4, 6, 10, 14, 18, 20\}$ indivíduos por geração, além de uma população panmítica correspondente para cada configuração do IBEA, possuindo essa população

panmítica um total de $n \times p$ indivíduos. A Distância Populacional calculada para a população panmítica foi comparada com a Distância Mundial do IBEA, visto que essa medida no cenário PBEA representa a distância entre todos os indivíduos da população. Assim, o objetivo desse experimento consiste em verificar se um aumento no número de migrantes por geração induz uma população panmítica. Com base nos dados obtidos concluímos que uma alta taxa de migração induz uma população panmítica. Para os gráficos considere M_i como sendo o número de migrantes por geração.

Ao compararmos os dados da Distância Populacional da população panmítica com o cálculo da Distância Mundial do IBEA, foi possível observar que, conforme o número de migrantes por geração aumenta, o comportamento da curva do IBEA se torna semelhante ao comportamento da curva da população panmítica, como pode ser observado nas Figuras 12, 13 e 14.

Isso ocorre devido ao alto fluxo gênico criado pelas elevadas taxas de migração, que acabam por espalhar os genes entre as ilhas de forma semelhante a que ocorre em uma população panmítica, anulando assim a barreira existente entre as ilhas e levando o IBEA a evoluir de forma equivalente a uma população panmítica.

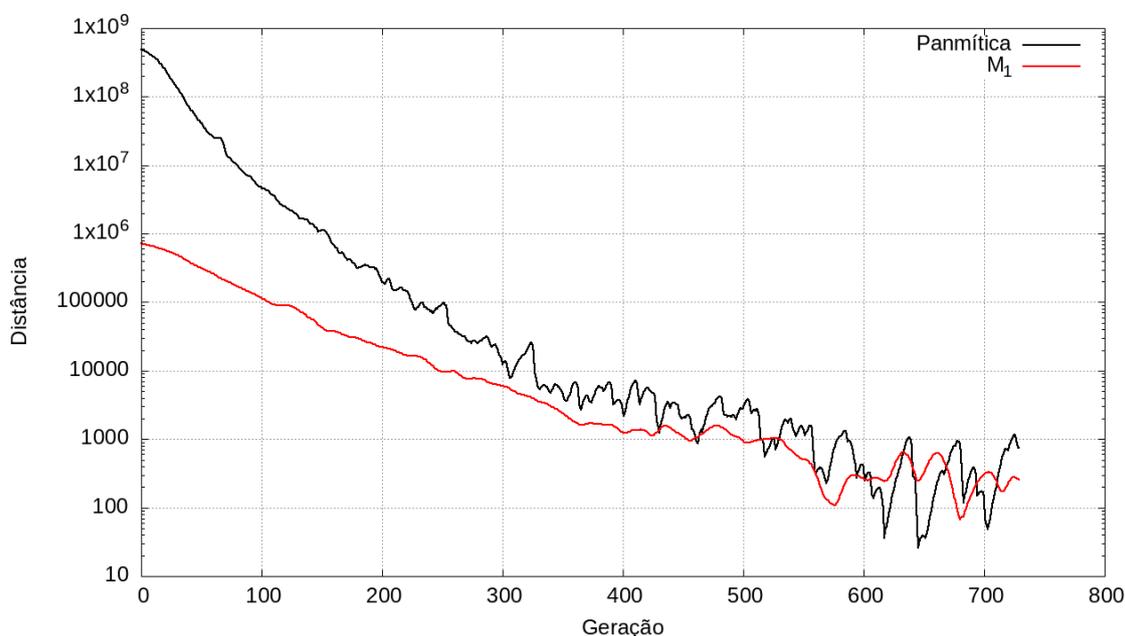


Figura 12. Comparação entre a Distância Mundial de um IBEA com 50 ilhas migrando 1 indivíduo por geração e a Distância Populacional de um PBEA para a função 3

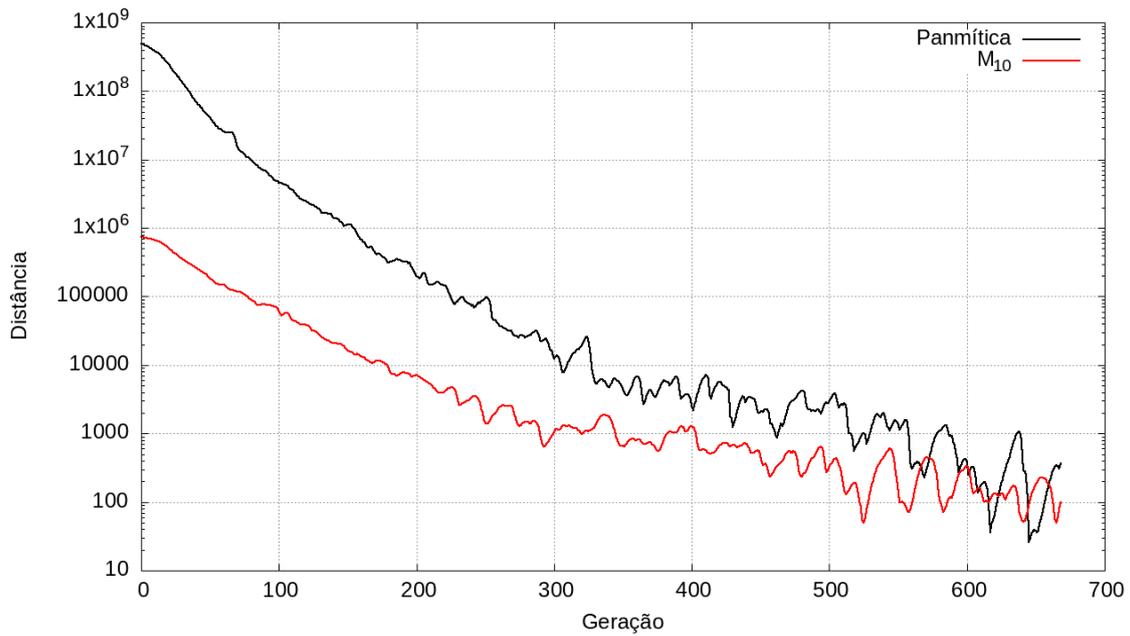


Figura 13. Comparação entre a Distância Mundial de um IBEA com 50 ilhas migrando 10 indivíduos por geração e a Distância Populacional de um PBEA para a função 3

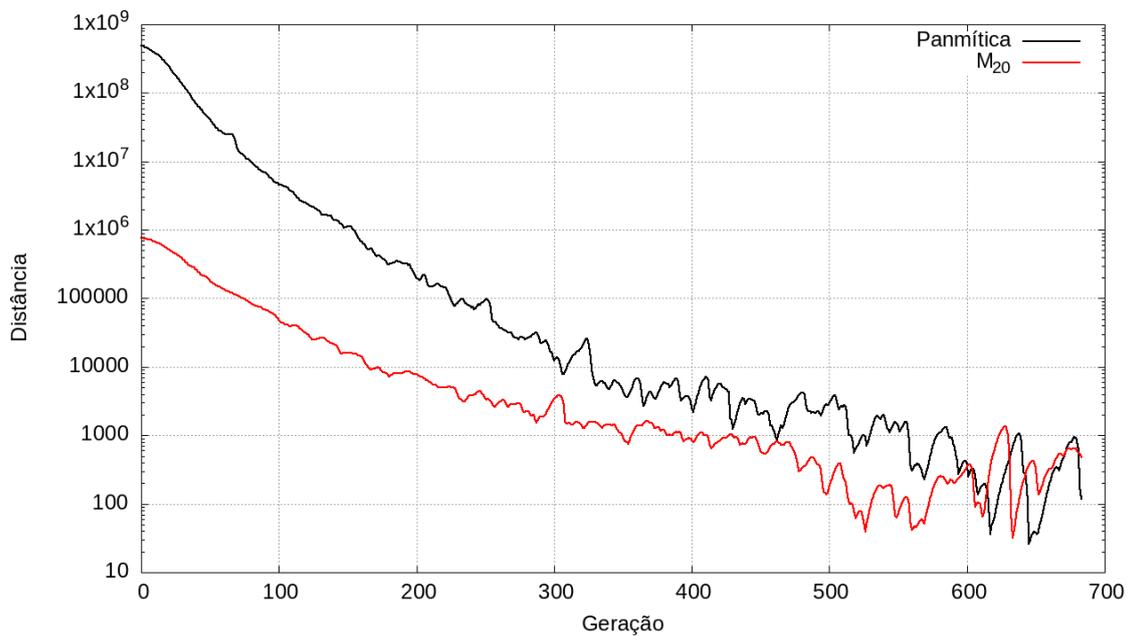


Figura 14. Comparação entre a Distância Mundial de um IBEA com 50 ilhas migrando 20 indivíduos por geração e a Distância Populacional de um PBEA para a função 3

6.5. Comparação entre a divergência genética de ilhas isoladas e uma população panmítica

Para a realização desse experimento e avaliação da divergência genética nas populações foram utilizados IBEA's com $n \in \{2, 5, 10, 25, 50\}$ ilhas, $p \in \{40\}$ in-

divíduos para cada ilha, onde as ilhas foram mantidas isoladas durante todo o processo evolutivo, além de uma população panmítica correspondente para cada configuração do IBEA, possuindo essa população panmítica um total de $n \times p$ indivíduos. Observando os resultados obtidos para esse experimento foi constatado que várias ilhas isoladas possuem uma divergência genética maior que uma população panmítica. Para os gráficos considere M_i como sendo o número de migrantes por geração.

Ao analisarmos a Distância Mundial do IBEA com a Distância Populacional da população panmítica (Figura 15), podemos observar que ilhas isoladas possuem uma aproximação genética extremamente lenta com o passar das gerações quando comparadas a uma população panmítica, que tende a convergir muito mais rapidamente. Isso se deve ao fato das ilhas isoladas tenderem a convergir cada uma para um ótimo local diferente. Sendo assim, após um determinado número de gerações cada ilha acaba sofrendo com o efeito da depressão endogâmica localmente, o que as leva a estagnação. Portanto, a divergência genética entre elas se mantém próxima ao longo das gerações, e tende a diminuir devido ao operador de mutação que atribui uma variação genética nas populações de tempos em tempos.

Por outro lado, os indivíduos em uma população panmítica estão próximos sem nenhum tipo de isolamento, levando a uma troca genética entre todos os indivíduos da população, fazendo com que eles convirjam juntos a um mesmo ótimo local o que os torna geneticamente próximos.

Assim, com o passar das gerações a população panmítica acaba sofrendo com a depressão endogâmica, enquanto que no IBEA as ilhas isoladas acabam sofrendo com a depressão endogâmica localmente e permanecendo geneticamente distâtes umas das outras, o que leva a uma maior divergência genética entre as ilhas isoladas quando comparado a uma população panmítica.

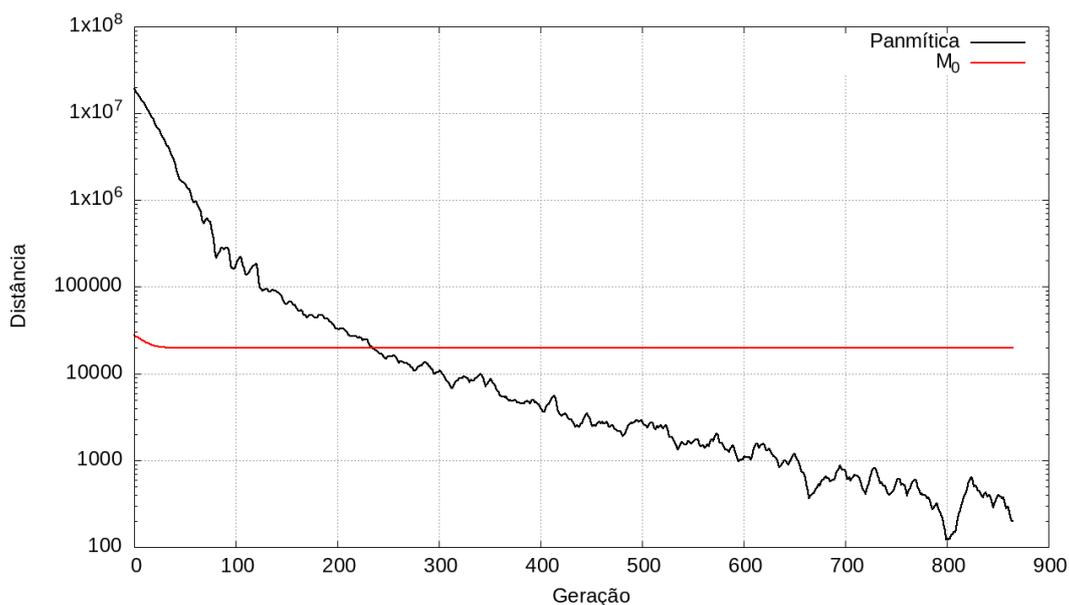


Figura 15. Comparação entre a Distância Mundial de um IBEA com 10 ilhas e a Distância Populacional de um PBEA para a função 14

6.6. Resumo dos experimentos

Na tabela abaixo temos um resumo para cada um dos experimentos realizados.

Experimento	Resultados	Conclusões
Otimização de hiperparâmetros	Hiperparâmetros otimizados: Tamanho da população: 3229 Taxa de mutação: 13% Taxa de cruzamento: 8%	Os parâmetros otimizados foram usados nos demais experimentos.
Regra OMPG em ilhas pequenas	Depressão endogâmica mitigada com a aplicação da regra OMPG. Cenário OMPG: aproximação genética rápida entre as ilhas. Cenário de isolamento: As ilhas permanecem geneticamente distantes.	A regra OMPG mitiga a depressão endogâmica, mas não é capaz de evitá-la devido à baixa variabilidade genética presente nas ilhas pequenas.
Ilhas pequenas com altas taxas de migração	Um aumento no número de migrantes resulta em uma maior diversidade genética inicialmente. Uma alta taxa de migração acelera a depressão endogâmica após a aproximação genética entre as ilhas.	As altas taxas de migração têm um grande impacto em ilhas pequenas, acelerando o processo de depressão endogâmica.
População panmítica vs IBEA com altas taxas de migração	À medida que o número de migrantes aumenta, o comportamento de um IBEA torna-se similar ao de uma população panmítica.	Um alto fluxo gênico entre as ilhas acaba tornando a evolução da população do IBEA equivalente a de uma população panmítica, anulando a barreira existente entre as ilhas.
Divergência genética de ilhas isoladas vs população panmítica	Ilhas isoladas: convergem para ótimos locais diferentes, permanecendo geneticamente distantes. População panmítica: todos os indivíduos convergem para um mesmo ótimo local, tornando os indivíduos geneticamente próximos	Várias ilhas isoladas possuem uma divergência genética maior que uma população panmítica. Isto ocorre porque as ilhas isoladas tendem a convergir para ótimos locais diferentes, mantendo-se geneticamente distantes durante o processo evolutivo.

Tabela 3. Resumo dos experimentos, resultados e conclusões

7. Considerações finais

No campo da pesquisa científica, a simulação da divergência genética de populações usando algoritmos evolutivos tem se mostrado uma abordagem promissora e de grande relevância. Essa técnica visa analisar os aspectos da evolução genética por meio da implementação de algoritmos, contribuindo significativamente para o avanço do conhecimento nessa área.

A simulação computacional desempenha um papel fundamental no estudo da evolução biológica, fornecendo uma ferramenta poderosa para analisar cenários diversos e compreender os processos de divergência genética. Através dessa abordagem, espera-se obter resultados que contribuam para uma melhor compreensão da dinâmica evolutiva e permitam avanços significativos no campo da genética.

Os resultados esperados dessa simulação podem revelar informações valiosas sobre os mecanismos subjacentes à divergência genética. Ao analisar a interação entre as populações e aplicar os princípios dos algoritmos evolutivos, é possível investigar a maneira que fatores como mutação, seleção, recombinação e migração influenciam a evolução genética. Essas descobertas podem ter implicações tanto no contexto teórico, ajudando a aprimorar modelos e teorias existentes, quanto no âmbito prático, fornecendo *insights* para áreas como genética populacional e conservação de espécies.

Em conclusão, a simulação da divergência genética de populações por meio de algoritmos evolutivos representa uma abordagem promissora e relevante para o estudo da evolução genética. Diante dessas perspectivas, questiona-se se a pesquisa nesse campo pode ser estendida para explorar outros aspectos da evolução, como a interação entre os genes e o ambiente, ou se há possibilidade de integrar abordagens genômicas e computacionais para uma compreensão mais abrangente dos processos evolutivos.

Trabalhos futuros neste tema devem explorar a simulação computacional para responder as outras perguntas levantadas na Seção 5.1. Além disso, outras questões podem ser levantadas com o contínuo estudo de genética de populações, estreitando ainda mais os laços entre estas áreas.

Referências

- [Al-Gazali et al. 2006] Al-Gazali, L., Hamamy, H., and Al-Arrayad, S. (2006). Genetic disorders in the arab world. *British Medical Journal*, 333(7573):831–834.
- [Alba 2005] Alba, E. (2005). *Parallel metaheuristics: a new class of algorithms*, volume 47. John Wiley & Sons.
- [Alba and Luque 2005] Alba, E. and Luque, G. (2005). Theoretical models of selection pressure for deas: topology influence. In *2005 IEEE Congress on Evolutionary Computation*, volume 1, pages 214–221. IEEE.
- [Alba and Troya 1999] Alba, E. and Troya, J. M. (1999). A survey of parallel distributed genetic algorithms. *Complexity*, 4(4):31–52.
- [Alba and Troya 2000] Alba, E. and Troya, J. M. (2000). Influence of the migration policy in parallel distributed gas with structured and panmictic populations. *Applied Intelligence*, 12(3):163–181.

- [Araujo and Merelo 2011] Araujo, L. and Merelo, J. J. (2011). Diversity through multiculturalism: Assessing migrant choice policies in an island model. *IEEE Transactions on Evolutionary Computation*, 15(4):456–469.
- [Balloux and Goudet 2002] Balloux, F. and Goudet, J. (2002). Statistical properties of population differentiation estimators under stepwise mutation in a finite island model. *Molecular Ecology*, 11(4):771–783.
- [Cantú-Paz 1998] Cantú-Paz, E. (1998). A survey of parallel genetic algorithms. *Calculateurs paralleles, reseaux et systems repartis*, 10(2):141–171.
- [Cantú-Paz 2001] Cantú-Paz, E. (2001). Migration policies, selection pressure, and parallel evolutionary algorithms. *Journal of heuristics*, 7(4):311–334.
- [Cotterman 1940] Cotterman, C. W. (1940). *A calculus for statistico-genetics*. PhD thesis, Ohio State University.
- [Da Silveira et al. 2021] Da Silveira, L. A., Soncco-Alvarez, J. L., De Lima, T. A., and Ayala-Rincon, M. (2021). Heterogeneous parallel island models. In *2021 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 1–8. IEEE.
- [Devi et al. 1987] Devi, A., Rao, N. A., and Bittles, A. (1987). Inbreeding and the incidence of childhood genetic disorders in karnataka, south india. *Journal of medical genetics*, 24(6):362–365.
- [Girardin and Débarre 2021] Girardin, L. and Débarre, F. (2021). Demographic feedbacks can hamper the spatial spread of a gene drive. *Journal of Mathematical Biology*, 83(6-7):67.
- [Gordon and Whitley 1993] Gordon, V. S. and Whitley, D. (1993). Serial and parallel genetic algorithms as function optimizers. In Forrest, S., editor, *Proceedings of the 5th International Conference on Genetic Algorithms*, pages 177–183. Morgan-Kaufmann.
- [Gustafson and Burke 2006] Gustafson, S. and Burke, E. K. (2006). The speciating island model: An alternative parallel evolutionary algorithm. *Journal of Parallel and Distributed Computing*, 66(8):1025–1036.
- [Hartl and Clark 2007] Hartl, D. and Clark, A. (2007). *Principles of Population Genetics*. Sinauer Associates.
- [Holland et al. 1992] Holland, J. H. et al. (1992). *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. MIT press.
- [Jost 2008] Jost, L. (2008). G_{ST} and its relatives do not measure differentiation. *Molecular Ecology*, 17(18):4015–4026.
- [Kardos et al. 2021] Kardos, M., Armstrong, E. E., Fitzpatrick, S. W., Hauser, S., Hedrick, P. W., Miller, J. M., Tallmon, D. A., and Funk, W. C. (2021). The crucial role of genome-wide genetic variation in conservation. *Proceedings of the National Academy of Sciences*, 118(48):e2104642118.
- [Keller and Waller 2002] Keller, L. F. and Waller, D. M. (2002). Inbreeding effects in wild populations. *Trends in Ecology & Evolution*, 17(5):230–241.
- [Lässig and Sudholt 2013] Lässig, J. and Sudholt, D. (2013). Design and analysis of migration in parallel evolutionary algorithms. *Soft Computing*, 17(7):1121–1144.
- [Lässig and Sudholt 2014] Lässig, J. and Sudholt, D. (2014). General upper bounds on the runtime of parallel evolutionary algorithms. *Evolutionary computation*, 22(3):405–437.

- [Lewontin et al. 1974] Lewontin, R. C. et al. (1974). *The genetic basis of evolutionary change*, volume 560. Columbia University Press New York.
- [Li and Horvitz 1953] Li, C. C. and Horvitz, D. G. (1953). Some methods of estimating the inbreeding coefficient. *American journal of human genetics*, 5(2):107.
- [Liang et al. 2014] Liang, J., Qu, B., Suganthan, P., and Chen, Q. (2014). Problem definitions and evaluation criteria for the cec 2015 competition on learning-based real-parameter single objective optimization. Technical Report Technical Report201411A, Zhengzhou University and Nayang Technological University.
- [Lima-Costa et al. 2015] Lima-Costa, M. F., Rodrigues, L. C., Barreto, M. L., Gouveia, M., Horta, B. L., Mambrini, J., Kehdy, F. S., Pereira, A., Rodrigues-Soares, F., Victora, C. G., et al. (2015). Genomic ancestry and ethnoracial self-classification based on 5,871 community-dwelling brazilians (the epigen initiative). *Scientific reports*, 5.
- [Lin et al. 1994] Lin, S.-C., Punch, W. F., and Goodman, E. D. (1994). Coarse-grain parallel genetic algorithms: Categorization and new approach. In *Parallel and Distributed Processing, 1994. Proceedings. Sixth IEEE Symposium on*, pages 28–37. IEEE.
- [Lindenmayer and Lacy 1995] Lindenmayer, D. B. and Lacy, R. C. (1995). Metapopulation viability of leadbeater’s possum, *gymnobelideus leadbeateri*, in fragmented old-growth forests. *Ecological Applications*, pages 164–182.
- [López-Ibáñez et al. 2016] López-Ibáñez, M., Dubois-Lacoste, J., Cáceres, L. P., Berrattari, M., and Stützle, T. (2016). The irace package: Iterated racing for automatic algorithm configuration. *Operations Research Perspectives*, 3:43–58.
- [Magalhães et al. 2015] Magalhães, T. T., Krempser, E., and Barbosa, H. J. C. (2015). Migration policies to improve exploration in parallel islands model for optimization via metaheuristics. In *Proceedings of the XXXVI Ibero-Latin American Congress on Computational Methods in Engineering*, page 20p. Brazilian Association of Computational Methods in Engineering.
- [Malécot 1948] Malécot, G. (1948). *Les mathématiques de l’heredite*. Masson et Cie, Paris.
- [Mambrini and Sudholt 2015] Mambrini, A. and Sudholt, D. (2015). Design and analysis of schemes for adapting migration intervals in parallel evolutionary algorithms. *Evolutionary computation*, 23(4):559–582.
- [Martin et al. 1997] Martin, W. N., Lienig, J., and Cohoon, J. P. (1997). Island (migration) models: evolutionary algorithms based on punctuated equilibria. In Baeck, T., Fogel, D., and Michalewicz, Z., editors, *Handbook of Evolutionary Computation*, pages 387–402. Taylor & Francis.
- [Masel 2011] Masel, J. (2011). Genetic drift. *Current Biology*, 21(20):R837–R838.
- [Matsumoto and Nishimura 1998] Matsumoto, M. and Nishimura, T. (1998). Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 8(1):3–30.
- [Migicovsky et al. 2016] Migicovsky, Z., Sawler, J., Money, D., Eibach, R., Miller, A. J., Luby, J. J., Jamieson, A. R., Velasco, D., von Kintzel, S., Warner, J., et al. (2016). Genomic ancestry estimation quantifies use of wild species in grape breeding. *BMC genomics*, 17(1):478.

- [Mills and Allendorf 1996] Mills, L. S. and Allendorf, F. W. (1996). The one-migrant-per-generation rule in conservation and management. *Conservation Biology*, pages 1509–1518.
- [Mitchell 2014] Mitchell, M. (2014). *An Introduction to Genetic Algorithms*. Complex Adaptive Systems. MIT Press.
- [Osorio et al. 2013] Osorio, K., Alba, E., and Luque, G. (2013). Using theory to self-tune migration periods in distributed genetic algorithms. In *2013 IEEE Congress on Evolutionary Computation*, pages 2595–2601. IEEE.
- [Osorio et al. 2011] Osorio, K., Luque, G., and Alba, E. (2011). Distributed evolutionary algorithms with adaptive migration period. In *Intelligent Systems Design and Applications (ISDA), 2011 11th International Conference on*, pages 259–264. IEEE.
- [Paixão et al. 2015] Paixão, T., Badkobeh, G., Barton, N., Çörüş, D., Dang, D.-C., Friedrich, T., Lehre, P. K., Sudholt, D., Sutton, A. M., and Trubenová, B. (2015). Toward a unifying framework for evolutionary processes. *Journal of Theoretical Biology*, 383:28–43.
- [Petit et al. 1993] Petit, R., Kremer, A., and Wagner, D. (1993). Finite island model for organelle and nuclear genes in plants. *Heredity*, 71(6):630–641.
- [Price et al. 2006] Price, K., Storn, R. M., and Lampinen, J. A. (2006). *Differential evolution: a practical approach to global optimization*. Springer Science & Business Media.
- [Prout 1981] Prout, T. (1981). A note on the island model with sex dependent migration. *Theoretical and Applied Genetics*, 59(5):327–332.
- [Provine 1989] Provine, W. B. (1989). *Sewall Wright and evolutionary biology*. University of Chicago Press.
- [Reeves 2010] Reeves, C. R. (2010). Genetic algorithms. In *Handbook of metaheuristics*, pages 109–139. Springer.
- [Reid et al. 2021] Reid, J. M., Arcese, P., Nietlisbach, P., Wolak, M. E., Muff, S., Dickel, L., and Keller, L. F. (2021). Immigration counter-acts local microevolution of a major fitness component: Migration-selection balance in free-living song sparrows. *Evolution letters*, 5(1):48–60.
- [Roux 1974] Roux, C. (1974). Hardy-weinberg equilibria in random mating populations. *Theoretical population biology*, 5(3):393–416.
- [Ruciński et al. 2010] Ruciński, M., Izzo, D., and Biscani, F. (2010). On the impact of the migration topology on the island model. *Parallel Computing*, 36(10):555–571.
- [Schaap 1980] Schaap, T. (1980). The applicability of the hardy-weinberg principle in the study of populations. *Annals of human genetics*, 44(2):211–215.
- [Schwehm 1996] Schwehm, M. (1996). Parallel population models for genetic algorithms. *Universität Erlangen-Nürnberg*, pages 2–8.
- [Scliar et al. 2014] Scliar, M. O., Gouveia, M. H., Benazzo, A., Ghirotto, S., Fagundes, N. J., Leal, T. P., Magalhães, W. C., Pereira, L., Rodrigues, M. R., Soares-Souza, G. B., et al. (2014). Bayesian inferences suggest that amazon yunga natives diverged from andeans less than 5000 ybp: implications for south american prehistory. *BMC evolutionary biology*, 14(1):1.
- [Sekaj 2004] Sekaj, I. (2004). Robust parallel genetic algorithms with re-initialisation. In *Parallel Problem Solving from Nature-PPSN VIII*, pages 411–419. Springer.

- [Servedio and Kirkpatrick 1997] Servedio, M. R. and Kirkpatrick, M. (1997). The effects of gene flow on reinforcement. *Evolution*, pages 1764–1772.
- [Shi et al. 2001] Shi, Y. et al. (2001). Particle swarm optimization: developments, applications and resources. In *evolutionary computation, 2001. Proceedings of the 2001 Congress on*, volume 1, pages 81–86. IEEE.
- [Skakovski and Jędrzejowicz 2019] Skakovski, A. and Jędrzejowicz, P. (2019). An island-based differential evolution algorithm with the multi-size populations. *Expert Systems with Applications*, 126:308–320.
- [Skolicki and De Jong 2005] Skolicki, Z. and De Jong, K. (2005). The influence of migration sizes and intervals on island models. In *Proceedings of the 7th annual conference on Genetic and evolutionary computation*, pages 1295–1302. ACM.
- [Spieth 1974] Spieth, P. (1974). Gene flow and genetic differentiation. *Genetics*, 78(3):961–965.
- [Stern 1943] Stern, C. (1943). The hardy-weinberg law. *Science*, 97(2510):137–138.
- [Talbi 2009] Talbi, E.-G. (2009). *Metaheuristics: from design to implementation*, volume 74. John Wiley & Sons.
- [Tomassini 2005] Tomassini, M. (2005). *Spatially Structured Evolutionary Algorithms: Artificial Evolution in Space and Time*. Springer.
- [Vargas et al. 2015] Vargas, D. V., Murata, J., Takano, H., and Delbem, A. C. B. (2015). General subpopulation framework and taming the conflict inside populations. *Evolutionary computation*, 23(1):1–36.
- [Wakeley 1999] Wakeley, J. (1999). Nonequilibrium migration in human history. *Genetics*, 153(4):1863–1871.
- [Wakeley 2000] Wakeley, J. (2000). The effects of subdivision on the genetic divergence of populations and species. *Evolution*, 54(4):1092–1101.
- [Whitley et al. 1998] Whitley, D., Rana, S., and Heckendorn, R. B. (1998). The island model genetic algorithm: On separability, population size and convergence. *Journal of Computing and Information Technology*, 7:33–48.
- [Wright 1922] Wright, S. (1922). Coefficients of inbreeding and relationship. *American Naturalist*, pages 330–338.
- [Wright 1932] Wright, S. (1932). The roles of mutation, inbreeding, crossbreeding, and selection in evolution. In *Proceedings of the Sixth International Congress of Genetics*, pages 356–366.