Quantitative Logics for Directed Evolution of an Asexual Population

Kangbien Park[1]

[1] Department of Physics, Yonsei University, Seoul 03722, Republic of Korea

November 2025

1 Abstract

Directed evolution of asexual populations is expected to offer a wide range of benefits to humanity. Achieving efficient directed evolution (DE) requires a quantitative formulation of the experimental methodologies—or logics—that can address potential challenges throughout the evolutionary process. In this article, I introduce ten logics designed for specific purposes in DE, accompanied by their mathematical formulations. To illustrate their application, I consider a hypothetical scenario of evolving an imaginary asexual population limb into a wing, employing a matrix-based discretization method to represent the trait of interest. The results indicate that applying these logics can accelerate the evolutionary process by roughly tenfold, while achieving an average accuracy of approximately 85% in reaching the desired trait across ten simulation iterations. Based on these findings, I discuss key considerations for implementing quantitative logic–based DE, propose strategies to improve alignment between the final outcome and the target trait, and examine factors for making the discretization method practical. Finally, I explore potential connections between quantitatively guided DE processes and emerging technologies, such as artificial intelligence and quantum computation.

2 Introduction

Directed evolution (DE) of asexual populations is an active area of research and is expected to bring diverse benefits to humanity Yuan et al. [2005], Laurent et al. [2023]. Despite substantial experimental progress in implementing DE through biological and chemical techniques Ren et al. [2025], Lee et al. [2020], quantitative modeling remains a crucial component of the DE process LaCroix et al. [2017], Lalejini et al. [2022], Corrao et al. [2024]. In particular, mathematical formulations enable rigorous analysis and the development of strategies to overcome key challenges associated with evolutionary experiments.

For example, when a population becomes trapped at an undesired local optimum on the fitness landscape, methods are needed to steer the population along pathways that descend the landscape. Rigorously achieving this requires precise mathematical modeling of both the trait and its associated potential. Additionally, if the evolutionary process takes too long to reach the target algebra, strategies for accelerating the trajectory become necessary. To this end, predictive tools that estimate the evolutionary timescale based on the mathematically defined experimental design, and that suggest shortcuts informed by such predictions, are essential.

Finally, mathematical formulations also provide insight into how changes in environmental conditions or differences in evolutionary speed influence the overall success probability of the evolutionary process, as these outcomes are ultimately governed by stochastic evolutionary dynamics.

Fortunately, with the growing body of knowledge in evolutionary dynamics and population genetics, along with the rapid increase in computational power driven by advances in computer science and AI, it has become increasingly promising to apply quantitative methodologies to directed evolution (DE) of asexual populations James et al. [2024], Gallup et al. [2024], Lee and Steel [2022]. Numerical approaches that design and simulate the evolutionary trajectory of an asexual population—and that express the corresponding experimental methodologies, or *logics*, in well-defined mathematical terms—can therefore serve as a foundation for systematically guiding and optimizing the DE process.

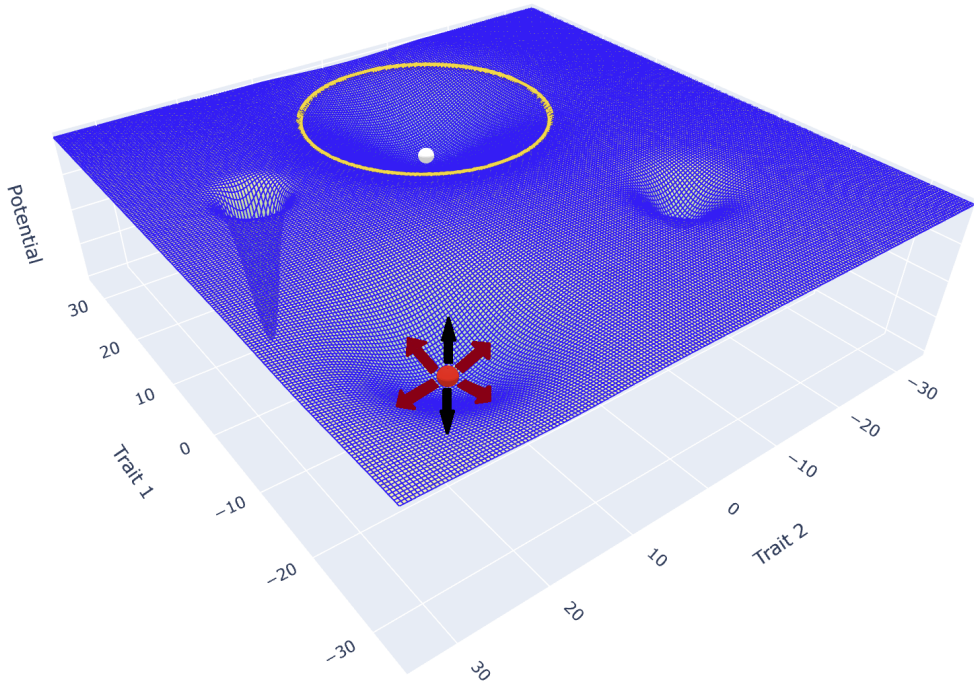


Figure 1: **Illustration of the objective of directed evolution.**

In this article, ten plausible logics for artificially directing evolution are mathematically formulated, and the corresponding reductions in evolutionary timescale are estimated. To accomplish this, an algebraic model based on an $n_1 \times n_2 \times n_3$ matrix is used to represent the trait of interest, capturing the physiological structure of an imaginary asexual population limb in the x-, y-, and z-directions. This representation is constructed using a discretization method that converts a continuous geometric structure into a matrix-based form. The results show that these logics can accelerate the evolutionary process by an order of magnitude. In addition, the potential implementation of these logics in laboratory settings is discussed.

3 Theory and Methodology

For the overall analysis, an algebraic model Park [2025] based on the operator framework is employed Park and Bae [2024]. This model assumes an idealized situation in which mutations of interest occur only within the algebraic set, while the potential is defined independently of the current trait state. Note that I use the concept of the potential $\Phi = F_{max} - F$ where F is the fitness landscape Park [2025]. Under this setting, a two-dimensional vector with continuous components is used to provide an intuitive illustration of why the logics are important and how they can be defined.

Fig 1 illustrates the scheme for directing the evolution of a two-dimensional vector. The red arrows indicate the possible mutation directions within the plane, while the black arrows represent mutations in the z-direction (or any other higher-dimensional direction), which are forbidden under the assumption of a closed algebraic set. In this setting, the goal of DE is to guide the current population state (red ball) toward the target state (white ball), where the radius of each ball denotes the proportion of existing mutations.

$$P_{\text{establishment per generation}} \sim -\Delta\Phi \cdot m \cdot N(t), \tag{1}$$

However, because the potential is convex near the red ball, artificial interventions—or *logics*—are required to push the ball upward along the potential Φ. Otherwise, an establishment event leading to a higher potential level will not occur naturally, since equation (1) implies that $P_{\text{establishment per generation}} < 0$ whenever $\Delta\Phi > 0$. Furthermore, even after the population reaches the flat region, additional guidance is required to reach the

yellow ring. Once the population arrives at the ring, the evolutionary process naturally converges to the target position, as described by equation (2) [Park, 2025], where A_i represents the algebra of interest in the i-th generation, $m_{A_i \leftarrow A_{i-1}}$ is the mutation rate from algebra A_{i-1} to A_i, and $\Delta\Phi_{A_i \leftarrow A_{i-1}}$ is the potential difference between A_{i-1} and A_i.

$$T_{\text{establishment}} = \begin{cases} T_{\text{est}_{A_{i+1}}} = \dfrac{1}{|\Delta\Phi_{A_i \leftarrow A_{i-1}}|} \cdot \ln\left(\dfrac{(\Delta\Phi_{A_i \leftarrow A_{i-1}})^2}{k \cdot m_{A_i \leftarrow A_{i-1}} \cdot |\Delta\Phi_{A_{i+1} \leftarrow A_i}|}\right), \quad i \geq 1, \\ T_{\text{est}_{A_1}} = \dfrac{1}{N \cdot m_{A_1 \leftarrow A_0} \cdot |\Delta\Phi_{A_1 \leftarrow A_0}|}. \end{cases} \tag{2}$$

To achieve this, ten logics are introduced in this article: the logic of switching traits, the logic of towing from linkage, the logic of towing from phenotypic dependence, the logic of towing from pleiotropy, the logic of emergence, the logic of varying mutation rate, the logic of paste, the logic of positive potential, the logic of negative potential, and the logic of explicit modification. While other logics may also be devised to manipulate evolutionary trajectories, this article focuses on these ten.

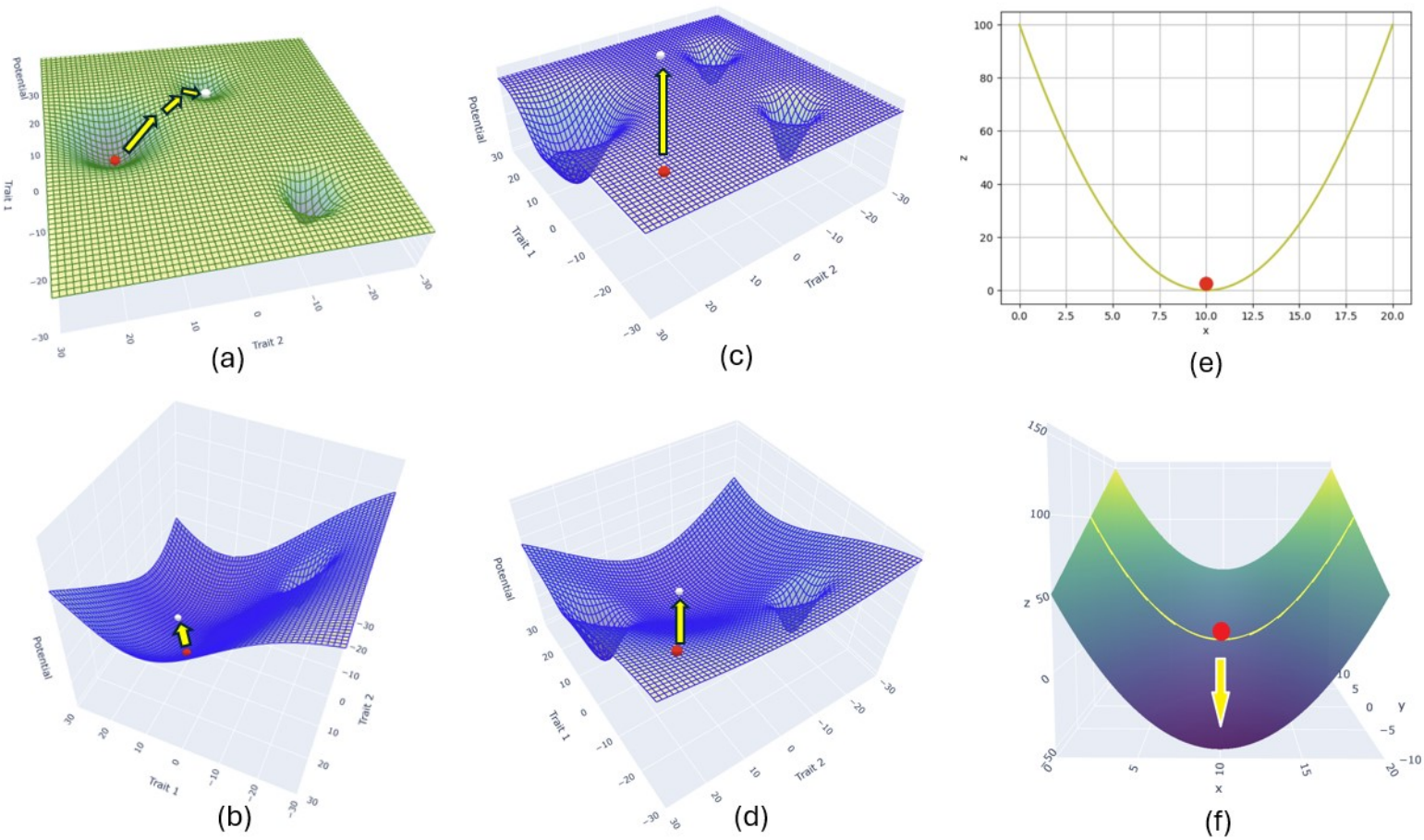


Figure 2: **Pictorial representation for the logics with 2D vector case.** (a) and (b) represent the logic of switching traits as well as the three variants of the towing logic. Also, (c) and (d) illustrate the logic of positive potential, and panels (e) and (f) demonstrate the logic of paste.

Starting with the logic of switching traits, represented by Fig 2 (a) and (b), consider a particular trait represented by a two-dimensional vector that can adopt two different forms depending on the environmental pH. When the pH is neutral, the potential for the vector corresponds to form (a); when the pH is acidic, the potential shifts to form (b) because the trait adopts a different configuration. In this scenario, even if

the state becomes trapped in an undesired local minimum under neutral pH, it can still evolve toward the target position when the environment turns acidic. The estimated time for the algebra to reach the objective position can be calculated using relation (2), with the potential differences derived from form (b).

Second, the logic of positive potential, represented from Fig 2 (c) and (d) concerns altering the significance, or weight, of a particular environmental factor for survival, which induces an overall change in the potential. This change can be represented by form (c) as the initial potential and form (d) as the new potential after applying the logic of positive potential. The time required for the evolutionary reaction can then be estimated using relation (2), with the potential in this case taken from form (d).

Third, the logic of paste is essentially about expanding the explorable range of the algebra. For an intuitive example, consider an asexual population with only 100 base pairs of DNA, initially positioned at a minimum potential. By adding 100 more base pairs—analogous to "pasting" in sculpture—the population acquires additional genetic material, enabling further evolutionary exploration. In algebraic terms, the logic of paste typically involves increasing the dimension of the algebra. For instance, the initial algebra represented by a one-dimensional variable in Fig 2 (e) can be expanded into a two-dimensional space, as shown in Fig 2 (f). While no further optimization was possible in the original one-dimensional algebra, the logic of paste allows optimization in the higher-dimensional representation. Note that the logic of paste preserves all original algebra elements within the expanded set.

Fourth, the logic of towing from linkage relates to the evolution of linked genes, where multiple genes are physically located on the same plasmid or the main chromosome of the microorganism. For illustration of the logic, consider two genes encoding two 2-dimensional vectors that are linked within a plasmid. Even if the first vector should move downward along its potential, it may be constrained if the selective advantage of the second vector is much larger, preventing the first vector from moving as expected. In this scenario, even if the first vector becomes trapped in a local minimum, as illustrated in Fig 2 (a), mutations corresponding to a higher potential can still establish due to the dominant evolutionary influence of the linked second vector. It is important to note that this approach can also stabilize the position of the first vector, effectively *desensitizing* it to potential differences because of the dominant dynamics of the second vector.

Fifth, the logic of towing from phenotypic dependence concerns two interdependent algebras, which do not necessarily share genes. In this scenario, the evolution of the first algebra can alter the potential of the second algebra. While generation-independent potentials are generally preferable for ease of calculation, this logic requires accounting for the generation-dependent nature of the potential. To simplify the analysis, one could experimentally restrain the evolution of the second algebra until the first algebra completes its influence on the second algebra's potential, and then allow evolution to proceed. Under this approach, the total generation for the evolutionary reaction would include the additional duration required for the potential change.

Sixth, the logic of towing from pleiotropy exploits the fact that a single gene can influence multiple algebras. This logic is structurally similar to the logic of linked genes, where the dominant algebra determines the evolutionary trajectory of a less strongly selected algebra. However, unlike the linked case, the effect in pleiotropy is more direct: the weaker-selected algebra must move whenever the dominant algebra moves. The evolutionary time in this scenario can be estimated by considering the dominant algebra's evolution via relation (2) and its effect on the DNA sequence, which subsequently drives the evolution of the target algebra. For precise modeling, the relationship between the DNA sequence and the algebraic representation of the corresponding phenotypes must be clearly defined.

Note that the three variants of the logic of towing are illustrated across Fig 2 (a) and (b): Fig 2 (a) depicts the primary algebra of interest, while (b) shows the secondary algebra that facilitates towing.

Seventh, the logic of negative potential is inspired by the behavior of antibiotic-resistant bacteria. When a mutation allows survival in a harsh environment, it is almost certainly fixed, regardless of whether it was initially established. This harsh environment is represented as a negative potential, illustrated in (h) and (i) of Fig 2. In (h), the radius of the ball represents the proportion of the mutation, showing that only a few vectors occupy the position (-16, 16). After applying the negative potential, which permits survival only at (-16, 16), the original population around (-20, 20), shown by the dotted yellow sphere, is driven to (-16, 16), illustrated by the solid red sphere. The benefits of the negative potential logic are twofold. First, it accelerates evolution by bypassing the time required for beneficial mutations to be established and fixed. Second, in directed evolution, key intermediate states are often needed but may be difficult to establish. In such cases, negative potential can eliminate competing variants, leaving only the algebra of interest and thus

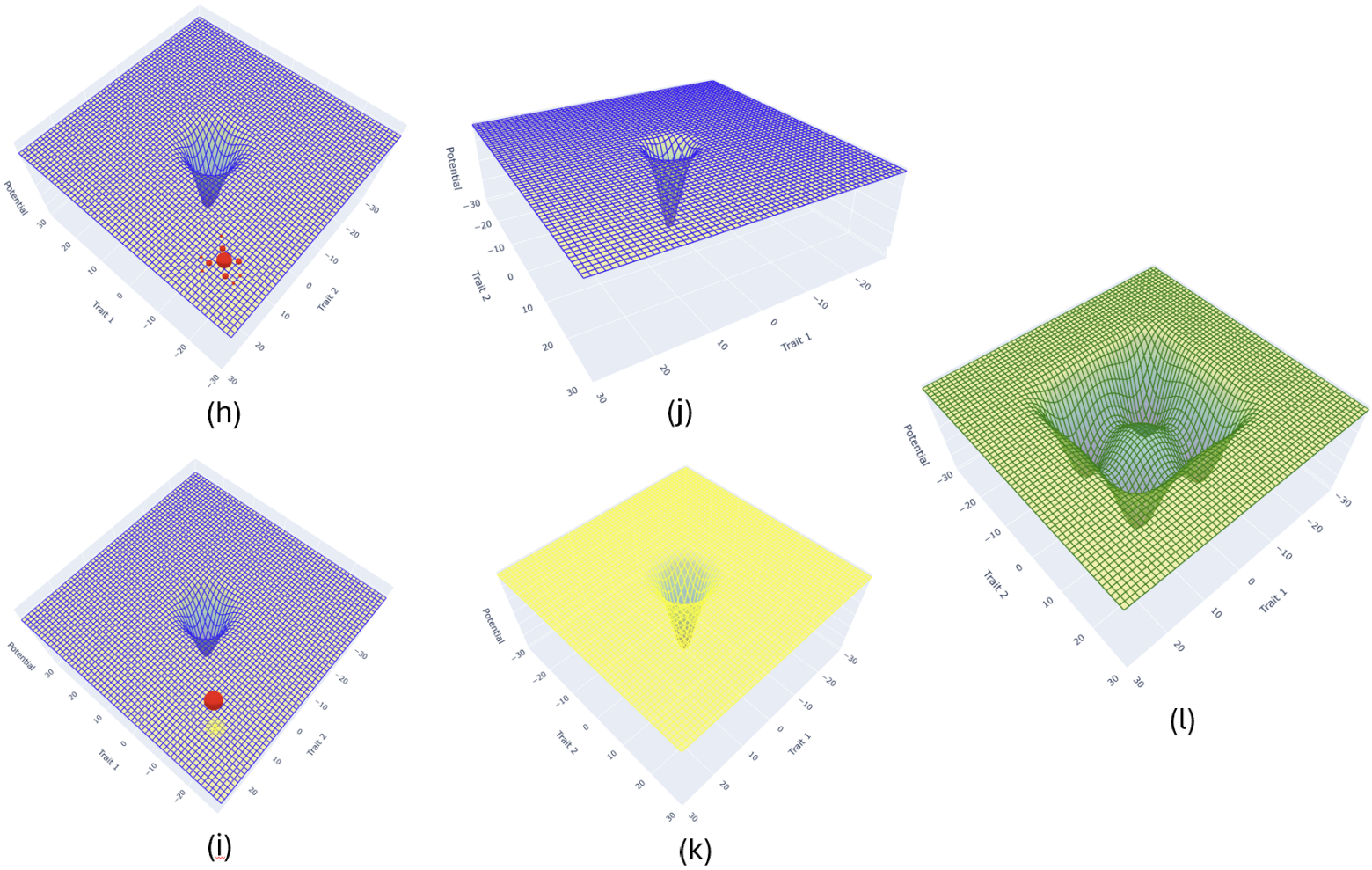


Figure 2: **Continuous pictorial representation for the logics with 2D vector case.**(h) and (i) represent the logic of negative potential. Also, (j), (k), and (l) illustrate the logic of emergence.

contributing to the overall success of the evolutionary reaction. The total evolutionary time can then be estimated as $t = \log_2(N/N_{\text{initial}})$, representing the time required for the surviving mutation population to double from N_{initial}.

Eighth, the logic of emergence concerns the combination of potentials that produces emergent effects. It is often assumed that when two potentials, Φ_1 and Φ_2, have the forms shown in Fig 2 (j) and (k), their sum, Φ_{sum}, can be represented simply as a linear addition, $\Phi_{\text{sum}} = \Phi_1 + \Phi_2$. However, some potentials may exhibit emergent properties, such that $\Phi_{\text{sum}} = \Phi_1 + \Phi_2 + f(\Phi_1, \Phi_2)$, where $f(\Phi_1, \Phi_2)$ is an emergent function, resulting in the form illustrated in (l). To apply the logic of emergence in directed evolution, extensive experimental data are required to accurately combine the potentials and design the emergent potential. The expected evolutionary time can then be estimated using relation (2), with the emergent potential taken into account.

Ninth, the logic of varying the mutation rate concerns adjusting the mutation rate of the population. Since the mutation rate strongly influences the evolutionary trajectory, strategically increasing or decreasing it at specific stages can help guide the population along the desired path more efficiently.

Tenth, the logic of explicit modification involves directly moving the red ball to or near the objective position. In practice, this can be realized in the laboratory through genetic engineering methods, effectively placing the population closer to the desired trait.

Using the introduced logics, I evolve the traits of interest represented by an $n_1 \times n_2 \times n_3$ matrix M,

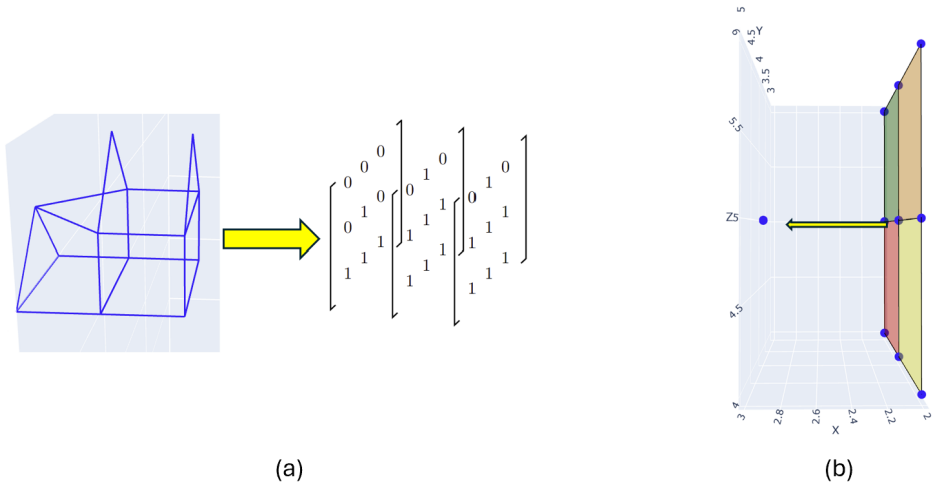


Figure 3: (a) illustrates the discretization example, while (b) represents the square-based mutation rule.

demonstrating an application of algebraic modeling to directed evolution (DE) [Park, 2025]. This matrix contains only 0s and 1s, where 1s correspond to points defining the physical shape in three-dimensional space. Rather than using the DNA sequence as the basis for mutation, mutations are applied directly to the algebra M representing the phenotypic trait, such that an element of M changes from 0 to 1 after mutation. Mathematically, the algebra is defined as

$$A = \Big(\big\{M_{i,j,k} \in \{0,1\} : i \in \{0,\ldots,n_1\}, j \in \{0,\ldots,n_2\}, k \in \{0,\ldots,n_3\}\big\}, +\Big).$$

Although other algebraic representations could be used to model the traits of interest, this $n_1 \times n_2 \times n_3$ matrix was chosen to intuitively capture the continuous physiological geometry of the traits in three-dimensional space via discretization, as illustrated in Fig 3.

With the defined algebra, consider a scenario in which the limb of an imaginary asexual population—sharing evolutionary characteristics similar to bacteria—gradually evolves into a wing. For this process, research on the evolutionary development of wings in birds, insects, and bats was used to define the appropriate potentials. Specifically, the development of the patagium enabling thermoregulation [Kingsolver and Koehl, 1985, Shockney and Rummel, 2025], the semi-wing structure allowing proto-birds to generate thrust for climbing trees [Bundle and Dial, 2003], and the wing structure enabling lift for flight [Tobalske, 2007] were all considered.

For the matrix, the mutation law is defined as follows: a mutation occurs when an original element 0 turns into 1. This mutation occurs only if there is an existing background structure in the direction of the mutation. To illustrate with an example, when the matrix evolves in the x direction at position $M_{3,4,5}$, one of the following sets of elements must already be 1:

$$M_{2,4,5} = M_{2,3,5} = M_{2,3,4} = M_{2,4,4}, \quad M_{2,4,5} = M_{2,3,5} = M_{2,3,6} = M_{2,4,6},$$
$$M_{2,4,5} = M_{2,4,6} = M_{2,5,6} = M_{2,5,5}, \quad M_{2,4,5} = M_{2,5,5} = M_{2,5,4} = M_{2,4,4}$$

as illustrated in Fig 3 (b). While other mutation rules could be applied, this choice intuitively represents a scenario in which a new mutation occurs only when there is sufficient prior "paste" in the matrix.

Now, with a well-defined matrix and its mutation law, the evolutionary path depends on the geometry of the potentials. To list the potentials that are used, first the thermal regulation potential is approximately modeled as

$$\Phi_{thermal} - \Phi_{\text{thermal max}}$$

$$\approx \begin{cases} -\sum_{i=0}^{N_i}\sum_{k=0}^{N_k}\sum_{j=0}^{N_j} M_{i\,j\,k}\cdot(y_{\text{criteria, thermal}}-j)^{\frac{1}{4}}\cdot w_{consecutive\,y,\,thermal}\cdot s\cdot w_{\text{thermal, thermal}} & if M_{i,\,j+1,\,k}=1 \text{ or } M_{i,\,j-1,\,k}=1, \\ -\sum_{i=0}^{N_i}\sum_{k=0}^{N_k}\sum_{j=0}^{N_j} M_{i\,j\,k}\cdot(y_{\text{criteria, thermal}}-j)^{\frac{1}{4}}\cdot w_{else,\,thermal}\cdot s\cdot w_{\text{thermal, thermal}} & \text{otherwise.} \end{cases}$$

The second potential, the thrust potential, is defined as

$$\Phi_{\text{thrust}}(M_{ijk})-\Phi_{\text{thrust max}}\approx\left(-\sum_{j=0}^{N_j}(\text{Weighted Area}_x)_j\cdot w_{\text{thrust, thrust}}+\frac{m}{2}\cdot g\right)s,$$

where the weighted area is the x-plane projected area computed according to the following rule:

$$\text{Weighted Area}_x=\begin{cases} 0, & (z_{\max}-z_{\min})_j<5, \\ N(x\text{–plane projected points})+\sum_k\sum_j w_{\text{consecutive y, thrust}}\cdot M_{i,j,k}\cdot M_{i,j+1,k} & \\ \quad+\sum_k\sum_j w_{\text{wide z, thrust}}\cdot M_{i,j,k}\cdot|k-z_{\text{middle}}|, & (z_{\max}-z_{\min})_j\geq 5. \end{cases}$$

In detail, the weighted area is computed as follows. First, the total number of x-plane projected points is counted. Then, for each pair of consecutive y-values at the same z-value, a weight w_y is added, and for each integer z-value, a weight $w_z\cdot(z-z_{\text{middle}})$ is also summed, where z_{middle} denotes the midpoint index of the matrix along the z-axis. These considerations are introduced to approximately capture the action–reaction principle for real gases: when the wing is insufficiently broad, the gas tends to slip sideways rather than exerting thrust.

Moreover, the mass m is defined to be proportional to the total number of matrix elements, given by

$$m=N(\text{total elements})\cdot w_{\text{mass factor}}.$$

Next, the lift potential is defined as

$$\Phi_{\text{lift}}(M_{ijk})-\Phi_{\text{lift max}}\approx\left(-\sum_{j=0}^{N_j}\left(\Delta P\cdot\text{Area}_x\right)_j+m\cdot g\right)\cdot s,$$

where the pressure difference for each j is given by

$$(\Delta P)_j=\left[\frac{\rho_{air}}{2}\left(v_{\text{fast}}^2-v_{\text{slow}}^2\right)\right]_j\approx\left[\frac{\rho_{air}}{2}\left(\frac{l_{\text{long}}^2}{l_{\text{short}}^2}-1\right)v^2\right]_j,$$

following Bernoulli's principle, based on the wing's y-axis cross-sectional structure shown in Fig 5 (h). Here, l_{long} and l_{short} denote the longer and shorter perimeters of the wing, respectively, and v is the velocity of the moving creature, approximated as $v\approx w_{\text{velocity, lift}}$. Moreover, $\rho_{air}=1.225$ (omitting an explicit physical unit kg/m^3) is set to represent the realistic value of the air density.

The quantity Area_x is the x-plane projected area for each j, obtained by projecting all matrix elements with $y=j$ onto the x-plane and counting the number of projected points. The following constraints are applied. First, if the wing is too thick at the edge-specifically, if $\Delta x\geq 3$ for a fixed y and for either $z_{\min}$ or $z_{\max}$-then that y-plane component does not contribute to the lift force, based on the assumption that excessively thick edges prevent the airflow from splitting cleanly enough to generate Bernoulli lift. Second, if the angle between adjacent points within a fixed y-plane exceeds $45°$, the corresponding component is also excluded from contributing to lift, under the intuitive assumption that airflow may not smoothly follow such abrupt geometric changes.

Lastly, the towing potential is modeled in a straightforward manner, taken to be proportional to the total number of matrix elements that exceed a specified threshold in the x, y, or z direction:

$$\Phi_{\text{towing}} = N(k \geq k_{\text{criteria}}) \cdot s \cdot w_{\text{towing}}.$$

The emergent potential is similarly modeled in a simplified but nonlinear form:

$$\Delta\Phi_{\text{emergence}} = w_{\text{thermal, emergence}} \cdot \left(\Delta\Phi_{\text{thermal}}\right)^2 + w_{\text{mobility, emergence}} \cdot \left(\Delta\Phi_{\text{mobility}}\right)^2 + w_{\text{both, emergence}} \cdot \left(\Delta\Phi_{\text{thermal}} \cdot \Delta\Phi_{\text{mobility}}\right),$$

where the weight vector is defined as

$$\vec{w}_{\text{emergence}} = \left[w_{\text{thermal, emergence}},\ w_{\text{mobility, emergence}},\ w_{\text{both, emergence}}\right].$$

Within this potential framework, several observations can be made. When only the thermal potential Φ_{thermal} is active, the optimal matrix configuration is one in which all elements with indices satisfying $j \leq y_{\text{criteria, thermal}}$ are equal to 1. When only the thrust potential Φ_{thrust} is applied, multiple optimal configurations exist, each maximizing the x-plane projected area. Finally, when the lift potential Φ_{lift} is the only active potential, the matrix evolves toward a configuration that maximizes the ratio between l_{long} and l_{short}.

Importantly, these optimal matrices differ substantially from one another, underscoring the necessity of adjusting the activated potentials at appropriate moments to guide the system toward the target matrix shown in Fig 3 (b).

The total potential combining these effects is defined as

$$\Phi_{\text{total}} = w_{\text{mobility, total}} \cdot \Phi_{\text{mobility}} + w_{\text{thermal, total}} \cdot \Phi_{\text{thermal}} + \Phi_{\text{towing}} + \Phi_{\text{emergence}},$$

where

$$\Phi_{\text{mobility}} = w_{\text{thrust, mobility}} \cdot \Phi_{\text{thrust}} + w_{\text{lift, mobility}} \cdot \Phi_{\text{lift}}.$$

Weight	**Value**
$w_{mass\,factor}$	1/50
g	10
$w_{thrust,\,mobility}$	1/5
$w_{lift,\,mobility}$	1/40
$w_{thermal,\,total}$	1/3
$w_{mobility,\,total}$	2/3
$y_{criteria,\,thermal}$	10
$w_{thermal,\,thermal}$	1/2
$w_{consecutive\,y,\,thermal}$	10
$w_{else,\,thermal}$	5
$w_{thrust,\,thrust}$	12.5
$w_{consecutive\,y,\,thrust}$	1/5
$w_{wide\,z,\,thrust}$	1/10
$w_{velocity,\,lift}$	1
$w_{towing,\,x}$	0
$w_{towing,\,y}$	0
$w_{towing,\,z}$	0
$\vec{w}_{emergence}$	$\vec{0}$

Table 1: Weights are set such that mutations can decrease the potential on the order of $\Delta\Phi \sim -0.02$.

The table 1 shows the initial weight values that determine the total potential. These weights are experimentally adjustable parameters, tuned according to the quantitative logics introduced above. They are set so that each mutation approximately lowers the potential by $\Delta\Phi \approx -0.01 \sim -0.1$, with an average

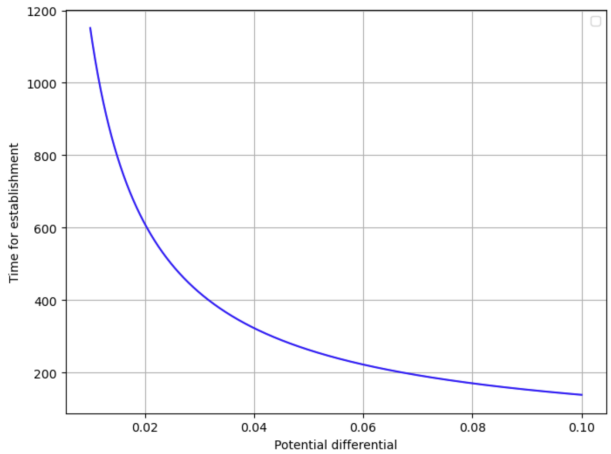


Figure 4: **Graph representing the relation between $\Delta\Phi$ and time taken for establishment.**

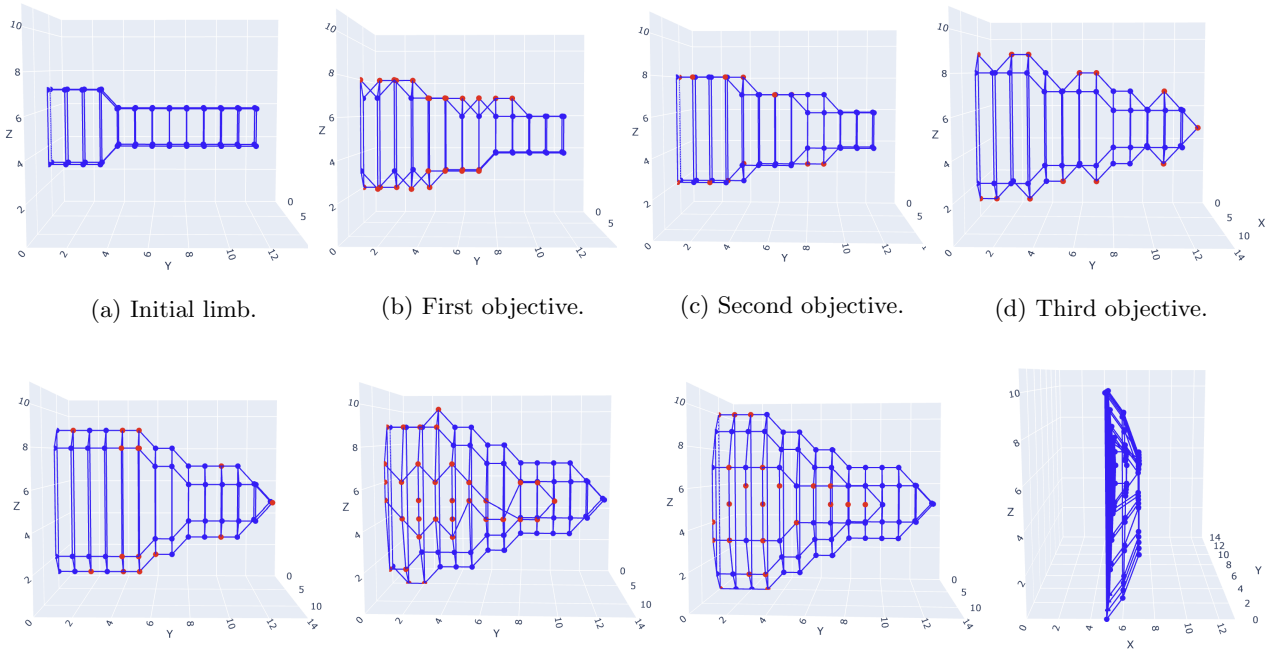
value of $\langle\Delta\Phi\rangle \approx -0.02$, consistent with empirical observations of natural evolution [Imhof and Schlötterer, 2001]. Under these conditions, each newly established mutation requires approximately $T \approx 841$ generations according to equation (2), assuming a mutation rate of $m \sim 10^{-9}$ [Imhof and Schlötterer, 2001] and $k \approx 1$ [Park, 2025]. Now, the initial limb shown in Fig 5 (a) contains 94 elements equal to 1, whereas the final wing in Fig 5 (b) contains 214 such elements. Thus, a rough estimate indicates that the transition requires approximately $214 - 94 = 120$ mutational establishments, corresponding to about $120 \times 841 \sim 10^5$ generations.

Even assuming a generation time of roughly 30 minutes for an asexual population, the entire process would then require about 70 months, making it economically impractical. Furthermore, the evolutionary trajectory is not guaranteed to reach the target matrix, since many alternative mutational paths exist and evolutionary outcomes are irreversible.

To overcome such inefficiency in the number of required generations, the logics introduced above can be employed to design the evolutionary pathway. In doing so, it is important to recognize several fundamental principles. First, the matrix is more likely to evolve in the direction that maximally decreases the potential, i.e., along $-\nabla\Phi$, as implied by equation (1). Second, to minimize the reaction time, the evolutionary path should ideally follow a trajectory of roughly uniform potential decrease [Park, 2025], unless the logics provide novel shortcuts. Based on these principles, the evolutionary pathway guided by the logics is designed to achieve potential decreases of $\Delta\Phi \approx -0.1 \sim -0.4$. Such a setting could be experimentally feasible given the relatively strong selection that bacteria can experience [Sidorova et al., 2020]. Moreover, a beneficial mutation rate of $m = 10^{-7}$ is chosen, as it falls between previously reported empirical values—ranging from $m \sim 10^{-9}$ [Imhof and Schlötterer, 2001] to $m \sim 10^{-5}$ [Perfeito et al., 2007]—and thus represents a realistic and experimentally attainable value. With these parameters, the estimated evolutionary time decreases to approximately 3,000–16,500 generations, corresponding to roughly 3–11.5 months, which is about one-tenth or less of the time required under natural evolutionary conditions, while the accuracy of reaching the objective is also expected to improve.

Note that the population evolution in this setting appears to deviate from the power-law behavior of selection coefficient accumulation, even though the environment is time-independently fixed [Wiser et al., 2013], and instead exhibits an approximately linear accumulation of selection coefficients. Despite this seemingly apparent discrepancy, the evolution still fundamentally follows the power law; the observed linear behavior corresponds only to the initial range of the power-law curve. Over a much larger number of generations, the matrix evolution will eventually converge to the expected power-law trend.

The evolutionary path is illustrated in Fig 5. It begins by allowing thrust force to be the dominant environmental pressure ($(a) \to (b)$, 3,000 generations), followed by thermal regulation becoming the primary selective factor ($(b) \to (c)$, 2,000 generations). These pressures are then repeated for one more cycle ($(c) \to (d)$, 1,500 generations; $(d) \to (e)$, 1,500 generations), before the environment increasingly requires lift force ($(e) \to (f) \to (g)$, 5,000 generations). To realize such qualitative design, the logic of towing and logic



(a) Initial limb. (b) First objective. (c) Second objective. (d) Third objective.

(e) Fourth objective. (f) Fifth objective. (g) Final wing. (h) Y-axis projected wing.

Figure 5: The graphical representation of milestones in the directed evolution of the discretized matrix. The red dots indicate elements that evolved from the preceding objective.

of positive potential were actively used while the logic of negative potential was often used. To achieve this, quantitative parameters were adjusted to approximately maximize the potential drop along the desired mutation direction, thereby increasing the probability of establishment in that direction. The specific weights used are provided in the Appendix.

For the simulation, all logics related to towing were incorporated into the logic of towing defined above, while the logic of switching traits was included in the logic of positive potential. The logics of explicit modification, varying mutation rate, paste, and emergence were not applied in this instance, as they were unnecessary for evolving the wing. However, these options remain available and can be simulated using the posted code. Note that the simulation was developed based on the operator model [Park and Bae, 2024], and ChatGPT was utilized to assist in developing sections of the plotting code and to support the design and debugging of specific parts of the computational elements.

4 Results and Discussion

The simulation results for a 13,000-generation evolution followed the predesigned trajectory, with the target wing structure shown in Fig 5 (g) (the resulting wing morphologies are provided in the appendix). The matrix corresponding to this figure was selected as the reference final matrix among several plausible divergent structures because it preserved the key morphological features intended to define a wing. Moreover, the morphology in (g) aligns with a rough quantitative expectation: if the weights are set such that commonly observed new mutations produce a potential change of approximately $\Delta\Phi \approx -0.15$, then relation (2) predicts that the final structure would incorporate about 120 new points over roughly 13,000 generations. Therefore, the matrix represented by Fig 5 (g) serves as an informed reference that both reflects the desired wing shape and aligns with predictions from a rough calculation. Moreover, to refine this reference, preliminary iterations were carried out to predict the final wing shape under weight sets that typically produce mutations with $\Delta\Phi \sim -0.1$ to -0.4 and steer the evolution toward a morphology resembling Fig 5 (g). The outcomes of these preliminary simulations were then used to estimate the expected final wing structure, ultimately

supporting the choice of Fig 5 (g) as the representative form. Although this process of performing preliminary iterations and anticipating the final form may seem somewhat ad hoc, it illustrates a realistic approach that computational DE can follow—namely, an expectation-maximization (EM) algorithm [Dempster et al., 1977]. Specifically, if the experimenter identifies a desired trait with certain degrees of freedom, adjusting the weights can help constrain that freedom and identify the optimal solutions among the diverse candidates sharing the trait. This is particularly important because, in realistic scenarios, objective traits often allow a wide range of functional or morphological variations that cannot be fully specified by the experimenter's criteria.

Comparing Quantity	**Min (%)**	**Max (%)**	**Average (%)**
Included Points	80	89	85
Total Number of 1s	-6	9	2
Lift Force	11	52	29
Thrust Force	-2	20	9
Thermal Regulation	-24	8	-5

Table 2: Quantities were measured from 10 independently gevolved wings, each obtained through a separate iteration of the procedure. The term "included points" refers to the percentage of points associated with the objective trait that appear in the resulting matrix, whereas the "total number of 1s" denotes the percentage difference between the total count of 1s in the resulting matrix and that in the objective matrix. Percentage differences for the potential values of lift force, thermal regulation, and thrust force were calculated using the weights listed in Table A3, which correspond to the environmental conditions at the final stage of the evolutionary process. Here, the percentage difference was computed using the formula $100 \times \frac{(Resulting\ matrix\ quantity - Objective\ matrix\ quantity)}{Objective\ matrix\ quantity}\%$.

The results in Table 2 demonstrate that an informed approximation of the objective was indeed feasible, and that the matrix can evolve into traits resembling the target characteristics. In particular, the resulting wings recovered, on average, up to 85% of the objective points, and the total number of 1s differed only slightly from the expected values. Table 2 also suggests that, for the system to approach the intended morphology, the evolutionary pressures associated with lift and thrust should gradually decrease, whereas the pressure associated with thermal regulation should increase.

Despite this, the outcome does not perfectly match the desired objective for two principal reasons: the limited computational capability to precisely track the evolutionary trajectory, and the restricted number of variables available for optimization. If the simulation were designed to evaluate potential changes for all plausible mutations at each generation of the directed evolution process—while simultaneously adjusting the associated potentials using optimized weights—the accuracy with which the target trait is achieved would improve substantially. Such an enhanced system, capable of handling large-scale calculations, would enable the design of far more sophisticated evolutionary pathways. Nonetheless, the primary goal of this work was to demonstrate that directed evolution guided by quantitative principles is feasible. Accordingly, the implemented code was developed only to the extent needed for this proof of concept.

Moreover, controlling only the three potentials listed above imposes limitations on achieving the exact target matrix. This is because multiple divergent evolutionary pathways exist, and in many cases, undesired traits can yield a similar or even greater potential decrease than the objective trait under the preset weights. For instance, the symmetry evident in Fig 5 (g) cannot be easily ensured, as the listed potentials do not regulate the additional degrees of freedom required to produce a symmetric wing geometry. Similarly, evolution along the x- and z-axes was difficult to control with the given parameters. This is because preparing combinations of parameters that maintain $\Delta\Phi$ between -0.1 and -0.4 in the desired direction required selecting settings that were among the more persuasive options available (i.e., the weights listed in the appendix), yet these settings still allowed deviations along the x- and z-directions. In summary, sophisticated computational tracking and additional potentials that provide more degrees of freedom are essential for higher-precision DE, and would further enhance the accuracy of the resulting outcomes.

Also, even when the $|\Delta\Phi|$ is fairly large (e.g., around $0.3 \sim 0.4$) for mutations, the evolutionary reaction still requires a minimum amount of time for any mutation to become established. Consequently, if the objective traits must go through many such establishment steps from the original traits, a significant minimum

time is inevitably required. To reduce this time, one approach is to begin with traits already available in nature that are closest to the desired objective traits, or to utilize the logic of explicit modification. In such a genetic engineering approach, providing only a rough framework for the desired trait is sufficient, as it still allows numerous possibilities to evolve into the fully developed desired trait. On the other hand, when the experimenter cannot fully predict how the engineered genes will behave due to numerous emergent effects within the biological system [Zhang et al., 2016, Şimşek et al., 2023], DE methodologies can be employed to refine and stabilize the inserted genes so that they function as intended. Thus, genetic engineering and directed evolution can complement each other, working synergistically to achieve the desired traits.

As demonstrated, evolutionary modeling based on trait-geometry discretization can provide significant benefits for DE calculations. To apply a discretized trait matrix in practical DE experiments, it is essential to establish a well-defined mutation law that governs how the matrix evolves from DNA mutations. Ideally, this could be implemented within a continuous *in vivo* DNA mutation system [Tong et al., 2025, DeBenedictis et al., 2022, Wong et al., 2018, Steel et al., 2020], in which the evolution of matrix elements follows predictable rules.

One possible strategy is to introduce a *parallelly evolvable genetic system*—a combination of DNA or RNA sequences with regulatory pathways—that satisfies three major characteristics: orthogonality and modularity [Wang et al., 2011, Szenk et al., 2020] of gene elements, inducibility [Tominaga et al., 2024, Cui et al., 2016, Wiechert et al., 2020] of the control system, and malleability (or evolvability) [Pines et al., 2017, Wagner, 2023] of genes.

Specifically, orthogonality ensures that multiple, mutually independent DNA sequences collectively contribute to the formation of the discretized matrix. A key advantage of this design is that, by avoiding long serial DNA constructs, mutations are less dependent on preceding sequences, allowing greater flexibility without compromising the integrity of already-evolved segments. This modularity also helps prevent unwanted epistasis among elements. Regarding inducibility, if a DNA fragment in the parallel system remains dormant until specific combinations of other elements are active, this behavior may serve as a robust foundation for implementing a matrix-based discretized evolution framework. Finally, gene malleability ensures that DNA or RNA sequences can be readily modified under evolutionary pressure. In sum, using a parallelly evolvable genetic system is analogous to introducing independent layers of malleable clay: initially shapeless, they are progressively molded as evolution proceeds, while the discretized matrix framework guides the system's evolutionary dynamics. Furthermore, insights from genotype–phenotype mapping [Otwinowski and Nemenman, 2013, Ascensao and Desai, 2025] are expected to be valuable for designing and optimizing such an approach.

Moreover, the example above treated the traits defining the algebra as the spatial physiological structure along the x, y, and z directions, which allowed efficient calculation of physical potentials since both classical mechanics and the trait space could be naturally formulated in three-dimensional physical space. However, when considering more general traits in arbitrary mathematical spaces, it becomes necessary to develop both theoretical and empirical methodologies to define a well-posed map that links matrix elements to the corresponding potential. In summary, theoretical research aimed at devising abstract mathematical geometries that efficiently model traits of interest—while ensuring that mutations of the geometry remain predictable based on preset DNA mutation technologies—would be of critical importance.

Finally, note that since the design of the evolutionary path is highly sensitive to even small changes in the potentials, a rigorous formulation that accurately reflects realistic environmental effects is imperative. In practice, because the potentials continuously change as the trait evolves, they will likely need to be remeasured at regular intervals during the evolution. Additionally, to achieve the required level of precision in potential measurement, massive datasets from both theoretical calculations and experiments are necessary, and the mechanisms by which environmental factors influence the potential must be well understood.

With such data in hand, the remaining task becomes the tuning of weights to design an evolutionary path that maximizes the success rate while minimizing the time required to reach the objective algebra using the logics described above. For the purpose, artificial intelligence (AI) methodologies, such as gradient descent [Cauchy et al., 1847] and backpropagation [Rumelhart et al., 1986], can be effectively employed to adjust parameters and minimize deviations from the intended path. Similarly, if sufficient data exists on how environmental factor determines the weights, the DE process could potentially be automated via AI-driven computation. Here, since the Expectation–Maximization algorithm has a close conceptual connection with self-supervised learning [Zhao et al., 2018, Si et al., 2010], situations involving a large number of possible evolutionary paths toward an initially vague final trait can benefit from self-supervised approaches, which help

converge on the desired trait while simultaneously adjusting the weights. Furthermore, for optimizations involving very high-dimensional potentials, techniques such as quantum annealing [Finnila et al., 1994, Kadowaki and Nishimori, 1998] or QAOA [Peruzzo et al., 2014] from future fault-tolerant quantum computing frameworks [Aharonov and Ben-Or, 1997] could be useful. Here, the mathematical structure of population states distributed across diverse potential positions may allow leveraging the superposition property of qubits to facilitate quantum optimization.

5 Conclusion

From the quantitative modeling of the logics corresponding to diverse experimental methodologies for addressing issues in directed evolution (DE), it is possible to calculate how the evolutionary path can be more effectively designed. Using a discretized matrix representing limb geometry, along with well-defined mutation rules and physical potentials, it was observed that these logics indeed allow the DE process to proceed more efficiently. Specifically, the speed of evolution increased by nearly tenfold, while the accuracy of the final trait relative to the objective reached approximately 85% on average. Encouraged by these results, further advancement in the formulation of quantitative logics for DE, or a more rigorous conceptualization of trait discretization, may provide deeper insights into how DE should be conducted. Moreover, such a mathematical representation of DE is expected to benefit from emerging technologies, including artificial intelligence and quantum computing, in various applications.

6 Data availability

The code for the simulation could be found from `https://github.com/TheDEnotes/Quantitative-Logics-for-DE`.

7 Acknowledgments

This work was largely conducted at the Department of Physics, Yonsei University, and at DAMTP, University of Cambridge. I am grateful for the calm and encouraging environments offered by both institutions. Furthermore, ChatGPT was utilized to help improve the clarity and readability of the English text.

8 Funding

This is a self-funded research.

9 Conflicts of interest

There is no conflict of interest to declare.

References

D. Aharonov and M. Ben-Or. Fault-tolerant quantum computation with constant error. In *Proceedings of the twenty-ninth annual ACM symposium on Theory of computing*, pages 176–188, 1997.

J. A. Ascensao and M. M. Desai. Experimental evolution in an era of molecular manipulation. *Nature Reviews Genetics*, pages 1–15, 2025.

M. W. Bundle and K. P. Dial. Mechanics of wing-assisted incline running (wair). *Journal of Experimental Biology*, 206(24):4553–4564, 2003.

A. Cauchy et al. Méthode générale pour la résolution des systemes d'équations simultanées. *Comp. Rend. Sci. Paris*, 25(1847):536–538, 1847.

M. Corrao, G. Abrahams, and H. Steel. Optimised control of adaptive evolution with competing selective pressures. *bioRxiv*, pages 2024–04, 2024.

W. Cui, L. Han, J. Cheng, Z. Liu, L. Zhou, J. Guo, and Z. Zhou. Engineering an inducible gene expression system for bacillus subtilis from a strong constitutive promoter and a theophylline-activated synthetic riboswitch. *Microbial Cell Factories*, 15(1):199, 2016.

E. A. DeBenedictis, E. J. Chory, D. W. Gretton, B. Wang, S. Golas, and K. M. Esvelt. Systematic molecular evolution enables robust biomolecule discovery. *Nature Methods*, 19(1):55–64, 2022.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society: series B (methodological)*, 39(1):1–22, 1977.

A. B. Finnila, M. A. Gomez, C. Sebenik, C. Stenson, and J. D. Doll. Quantum annealing: A new method for minimizing multidimensional functions. *Chemical physics letters*, 219(5-6):343–348, 1994.

O. Gallup, K. Sechkar, S. Towers, and H. Steel. Computational synthetic biology enabled through jax: A showcase. *ACS Synthetic Biology*, 13(9):3046–3050, 2024.

M. Imhof and C. Schlötterer. Fitness effects of advantageous mutations in evolving escherichia coli populations. *Proceedings of the National Academy of Sciences*, 98(3):1113–1117, 2001.

J. James, S. Towers, J. Foerster, and H. Steel. Optimisation strategies for directed evolution without sequencing. *PLOS Computational Biology*, 20(12):e1012695, 2024.

T. Kadowaki and H. Nishimori. Quantum annealing in the transverse ising model. *Physical Review E*, 58 (5):5355, 1998.

J. G. Kingsolver and M. Koehl. Aerodynamics, thermoregulation, and the evolution of insect wings: differential scaling and evolutionary change. *Evolution*, 39(3):488–504, 1985.

R. A. LaCroix, B. O. Palsson, and A. M. Feist. A model for designing adaptive laboratory evolution experiments. *Applied and Environmental Microbiology*, 83(8):e03115–16, 2017.

A. Lalejini, E. Dolson, A. E. Vostinar, and L. Zaman. Artificial selection methods from evolutionary computing show promise for directed evolution of microbes. *Elife*, 11:e79665, 2022.

J. M. Laurent, A. Jain, A. Kan, M. Steinacher, N. E. Casimiro, S. Stavrakis, A. J. deMello, and A. R. Studart. Directed evolution of microorganisms for engineered living materials. *arXiv preprint arXiv:2311.13342*, 2023.

M. Lee, H. J. Rozeboom, E. Keuning, P. de Waal, and D. B. Janssen. Structure-based directed evolution improves s. cerevisiae growth on xylose by influencing in vivo enzyme performance. *Biotechnology for Biofuels*, 13(1):5, 2020.

T. A. Lee and H. Steel. Cybergenetic control of microbial community composition. *Frontiers in Bioengineering and Biotechnology*, 10:957140, 2022.

J. Otwinowski and I. Nemenman. Genotype to phenotype mapping and the fitness landscape of the e. coli lac promoter. *PloS one*, 8(5):e61570, 2013.

K. Park. Using general algebra to model the directed evolution of an asexual population. 2025. doi: 10.1101/2025.11.22.689897.

K. Park and Y. Bae. Operator model for evolutionary dynamics. *BioSystems*, 237:105130, 2024.

L. Perfeito, L. Fernandes, C. Mota, and I. Gordo. Adaptive mutations in bacteria: high rate and small effects. *science*, 317(5839):813–815, 2007.

A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'brien. A variational eigenvalue solver on a photonic quantum processor. *Nature communications*, 5(1): 4213, 2014.

G. Pines, J. Winkler, A. Pines, and R. Gill. Refactoring the genetic code for increased evolvability. mbio, 8 (6), e01654–17. *DOI*, 10:01654–17, 2017.

Y. Ren, E. Celinska, P. Cai, and Y. J. Zhou. Combing directed enzyme evolution with metabolic engineering to develop efficient microbial cell factories. *Chem & Bio Engineering*, 2(8):449–459, 2025.

D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning representations by back-propagating errors. *nature*, 323(6088):533–536, 1986.

A. C. Shockney and A. D. Rummel. A hairy situation: The diversity of bat wing pelage and its functional implications. *Annals of the New York Academy of Sciences*, 1551(1):129–139, 2025.

Z. Si, H. Gong, S.-C. Zhu, and Y. N. Wu. Learning active basis models by em-type algorithms. 2010.

A. Sidorova, N. Levashova, A. Garaeva, and V. Tverdislov. A percolation model of natural selection. *Biosystems*, 193:104120, 2020.

E. Şimşek, Y. Yao, D. Lee, and L. You. Toward predictive engineering of gene circuits. *Trends in Biotechnology*, 41(6):760–768, 2023.

H. Steel, R. Habgood, C. L. Kelly, and A. Papachristodoulou. In situ characterisation and manipulation of biological systems with chi. bio. *PLoS biology*, 18(7):e3000794, 2020.

M. Szenk, T. Yim, and G. Balázsi. Multiplexed gene expression tuning with orthogonal synthetic gene circuits. *ACS synthetic biology*, 9(4):930–939, 2020.

B. W. Tobalske. Biomechanics of bird flight. *Journal of Experimental Biology*, 210(18):3135–3146, 2007.

M. Tominaga, Y. Shima, K. Nozaki, Y. Ito, M. Someda, Y. Shoya, N. Hashii, C. Obata, M. Matsumoto-Kitano, K. Suematsu, et al. Designing strong inducible synthetic promoters in yeasts. *Nature Communications*, 15(1):10653, 2024.

K. Tong, S. Datta, V. Cheng, D. J. Haas, S. Gourisetti, H. L. Yopp, T. C. Day, D. T. Lac, A. S. Khalil, P. L. Conlin, et al. Genome duplication in a long-term multicellularity evolution experiment. *Nature*, pages 1–9, 2025.

A. Wagner. Evolvability-enhancing mutations in the fitness landscapes of an rna and a protein. *Nature Communications*, 14(1):3624, 2023.

B. Wang, R. I. Kitney, N. Joly, and M. Buck. Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology. *Nature communications*, 2(1):508, 2011.

J. Wiechert, C. Gatgens, A. Wirtz, and J. Frunzke. Inducible expression systems based on xenogeneic silencing and counter-silencing and design of a metabolic toggle switch. *ACS synthetic biology*, 9(8): 2023–2038, 2020.

M. J. Wiser, N. Ribeck, and R. E. Lenski. Long-term dynamics of adaptation in asexual populations. *Science*, 342(6164):1364–1367, 2013.

B. G. Wong, C. P. Mancuso, S. Kiriakov, C. J. Bashor, and A. S. Khalil. Precise, automated control of conditions for high-throughput growth of yeast and bacteria with evolver. *Nature biotechnology*, 36(7): 614–623, 2018.

L. Yuan, I. Kurek, J. English, and R. Keenan. Laboratory-directed protein evolution. *Microbiology and Molecular Biology Reviews*, 69(3):373–392, 2005.

C. Zhang, R. Tsoi, and L. You. Addressing biological uncertainties in engineering gene circuits. *Integrative biology*, 8(4):456–464, 2016.

W. Zhao, S. Wang, Z. Xie, J. Shi, and C. Xu. Gan-em: Gan based em learning framework. *arXiv preprint arXiv:1812.00335*, 2018.

10 Appendix

Weight	**Value**
$w_{mass\,factor}$	1/50
g	10
$w_{thrust,\,mobility}$	1/3
$w_{lift,\,mobility}$	1/40
$w_{thermal,\,total}$	1/3
$w_{mobility,\,total}$	2/3
$y_{criteria,\,thermal}$	10
$w_{thermal,\,thermal}$	3
$w_{consecutive\,y,\,thermal}$	6
$w_{else,\,thermal}$	3
$w_{thrust,\,thrust}$	3
$w_{consecutive\,y,\,thrust}$	6/10
$w_{wide\,z,\,thrust}$	3/10
$w_{velocity,\,lift}$	5
$w_{towing,\,x}$	0
$w_{towing,\,y}$	8
$w_{criteria,\,towing,\,y}$	4
$w_{towing,\,z}$	0
$\vec{w}_{emergence}$	$\vec{0}$

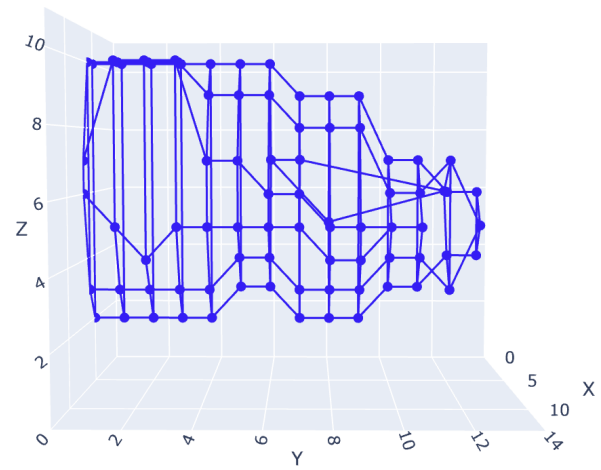
Table A.1: Weights for the thrust force step.

Weight	**Value**
$w_{mass\,factor}$	1/50
g	10
$w_{thrust,\,mobility}$	1/12
$w_{lift,\,mobility}$	1/40
$w_{thermal,\,total}$	1/3
$w_{mobility,\,total}$	2/3
$y_{criteria,\,thermal}$	10
$w_{thermal,\,thermal}$	4
$w_{consecutive\,y,\,thermal}$	10
$w_{else,\,thermal}$	5
$w_{thrust,\,thrust}$	3
$w_{consecutive\,y,\,thrust}$	1/10
$w_{wide\,z,\,thrust}$	3/10
$w_{velocity,\,lift}$	5
$w_{towing,\,x}$	0
$w_{towing,\,y}$	4
$w_{criteria,\,towing,\,y}$	0
$w_{towing,\,z}$	0
$\vec{w}_{emergence}$	$\vec{0}$

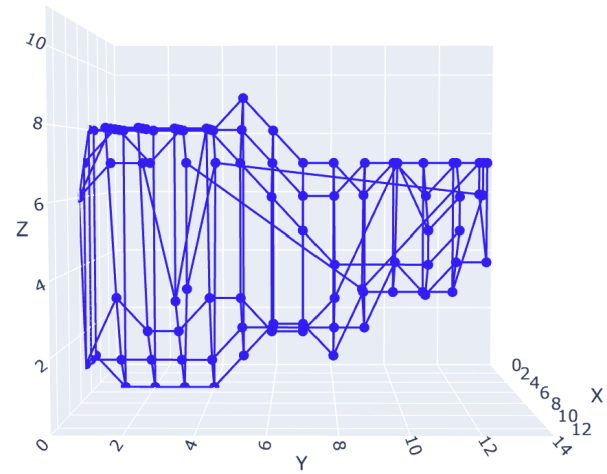
Table A.2: Weights for the thermal regulation step.

Weight	**Value**
$w_{mass\,factor}$	1/50
g	10
$w_{thrust,\,mobility}$	1/3
$w_{lift,\,mobility}$	1/10
$w_{thermal,\,total}$	1/3
$w_{mobility,\,total}$	2/3
$y_{criteria,\,thermal}$	1
$w_{thermal,\,thermal}$	3
$w_{consecutive\,y,\,thermal}$	5
$w_{else,\,thermal}$	5
$w_{thrust,\,thrust}$	1
$w_{consecutive\,y,\,thrust}$	3/10
$w_{wide\,z,\,thrust}$	6/10
$w_{velocity,\,lift}$	30
$w_{towing,\,x}$	8
$w_{criteria,\,towing,\,x}$	0
$w_{towing,\,y}$	0
$w_{towing,\,z}$	0
$\vec{w}_{emergence}$	$\vec{0}$

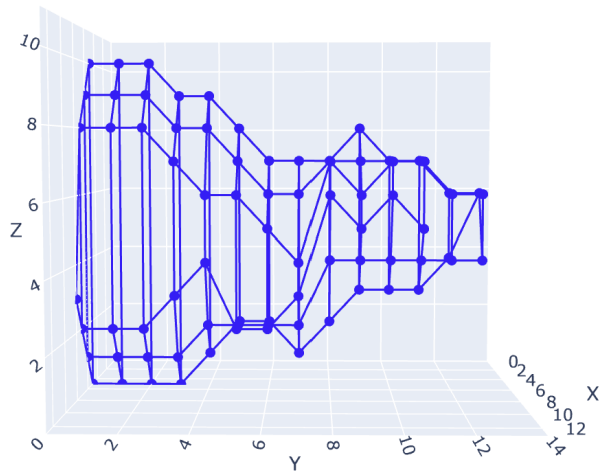
Table A.3: Weights for the lift force step.



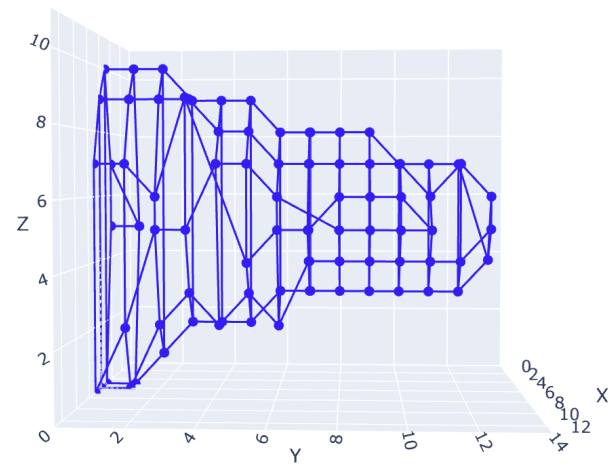
(a) 1st wing.



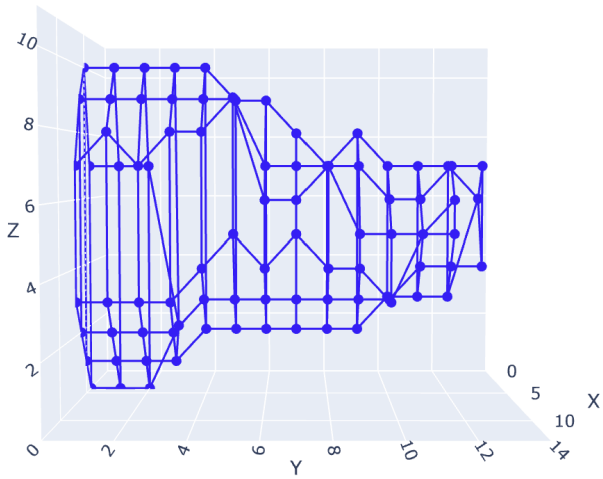
(b) 2nd wing.



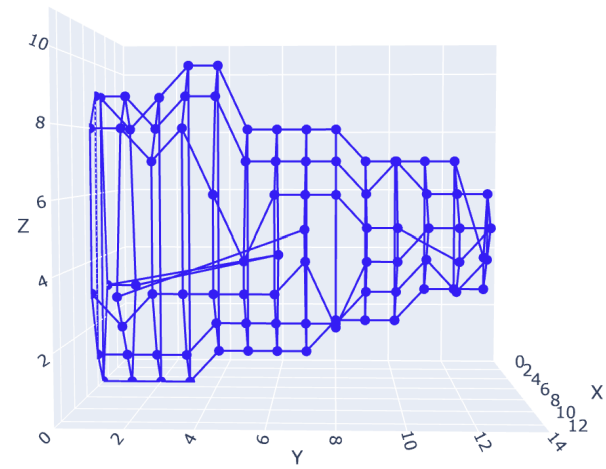
(c) 3rd wing.



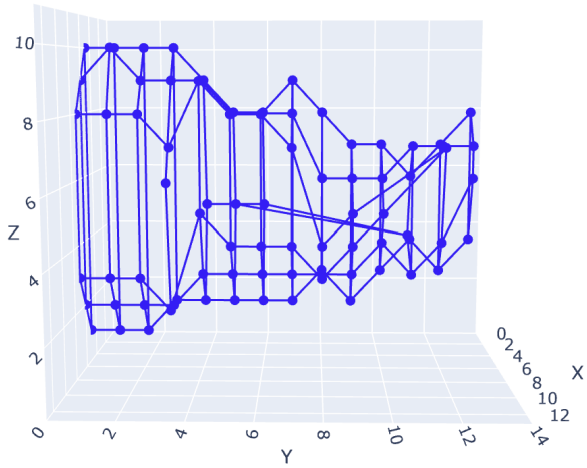
(d) 4th wing.



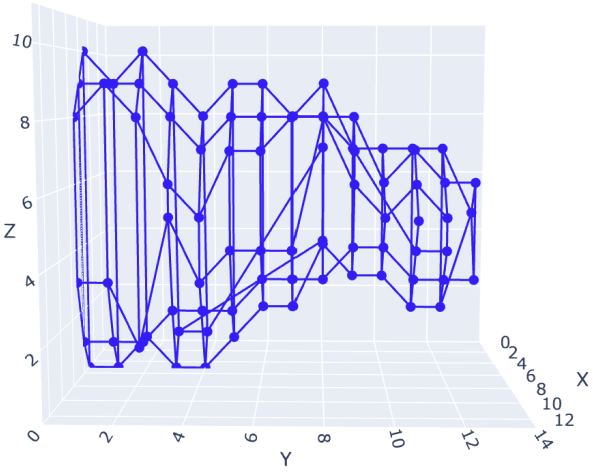
(e) 5th wing



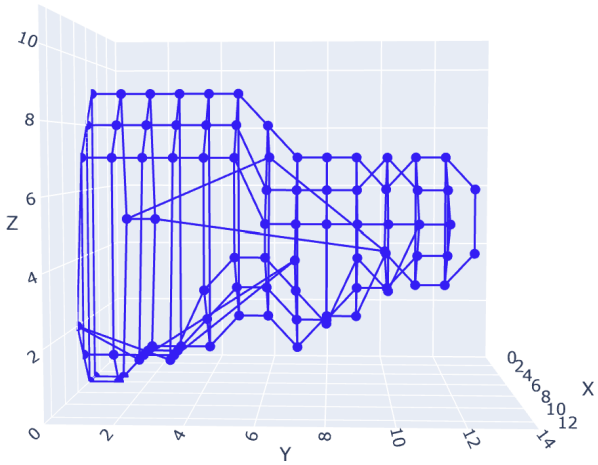
(f) 6th wing.



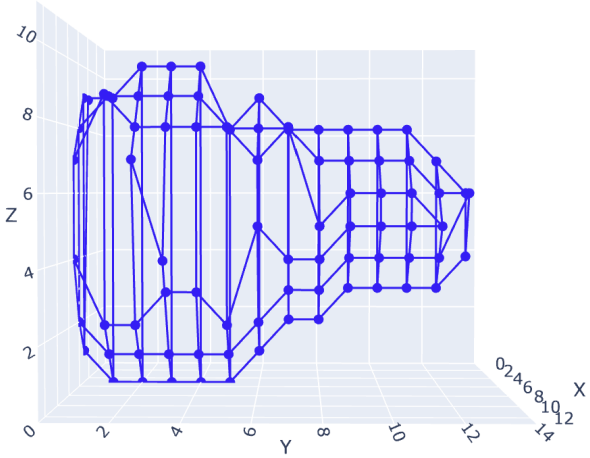
(g) 7th wing.



(h) 8th wing.



(i) 9th wing.



(j) 10th wing.

Figure A1: The resulting wing morphologies from DE.