An engineering theory of evolution

Simeon D. Castle[1], Claire S. Grierson[1,2] and Thomas E. Gorochowski[1,2,*]

[1] School of Biological Sciences, University of Bristol, Tyndall Avenue, Bristol, BS8 1TQ, UK

[2] BrisSynBio, University of Bristol, Tyndall Avenue, Bristol, BS8 1TQ, UK

* Correspondence should be addressed to T.E.G. (thomas.gorochowski@bristol.ac.uk)

Keywords: evolution; evotype; engineering theory; genotype-phenotype maps; directed evolution; selection; synthetic biology; bioengineering; evolvability; self-adaptive systems

Abstract

Biological technologies are fundamentally unlike any other because biology evolves. Bioengineering therefore requires novel design methodologies with evolution at their core. Knowledge about evolution is currently applied to the design of biosystems *ad hoc.* Unless we have a unified engineering theory of evolution, we will neither be able to meet evolution's potential as a design tool, nor understand or limit its unintended consequences on our designs. Our concept of the evotype offers a conceptual framework for engineering the evolutionary potential of biosystems. We show how a biosystem's evolutionary properties might be rationally designed by engineering aspects of genetic variation, designed function, and natural selection. This idea could apply to all biosystems – from individual proteins to communities of whole-cells or even entire ecosystems – whether the goal is to direct evolution in the design process, or to limit its impacts during application. These principles could even be used beyond the realm of bioengineering to design entirely synthetic evolving auto-adaptive technologies.

Introduction

The past few decades have seen a revolution in our ability to engineer biology and create novel living systems [1]. Yet, several hurdles still hinder our ability to harness biology's full potential [2]. These predominantly stem from the fact that engineering the stuff of life is not the same as engineering its properties, because life evolves. Evolution makes engineering living systems a radically different challenge to engineering other mediums. To be effective, we cannot just apply traditional engineering design principles to biology and deal with evolution as a secondary thought. If nothing in biology makes sense except in the light of evolution [3], then evolution must be a central part of an engineering theory of biology.

Evolution poses both a challenge and an opportunity when designing biosystems. On one hand it is a detrimental force that can unpick the meticulous plans of an engineer through mutation [4]. Designed biosystems cannot escape evolution when used and loss of function is a particular concern for engineers, especially as there are often selection pressures for it [5,6]. Thus, it is essential that we are able to build evolutionarily robust biosystems that can continue to operate under unavoidable evolutionary forces.

On the other hand, evolution is an extremely effective problem solver, and engineers have exploited this fact for decades [7–10]. For example, directed evolution can be used to optimise or even generate completely novel traits in single proteins [11] or entire cells [12]. However, these methods rely on the ability of evolution to find solutions in a reasonable length of time. For most systems, the search space is so vast that the starting point in this process must have the potential to generate useful phenotypes relatively quickly. Evolution may even be employed as a feature of the system during operation. For example, being used to develop adaptive systems that evolve in response to environmental cues, or even to create evolvable genetic circuits that can be designed with specific classes of phenotype that are reached as necessary through evolutionary change. To create such systems, it is critical that the biological design is evolvable, having the potential to generate desired phenotypes from a single starting point.

Even more critical, is our moral obligation to a deeper understanding of how synthetic biosystems will continue to evolve if deployed into our bodies or the wider environment [13]. The field has rightly made efforts to develop tools to reduce and mitigate evolution [14], with fail-safes such as kill switches [15] or metabolic dependencies [16]. However, without a good theoretical understanding of how synthetic biosystems might continue to evolve once deployed, we risk these technologies developing unexpected faults with dire, but avoidable, consequences.

Central to many of these issues is the view in traditional engineering disciplines that the engineered artefact is a final destination in the design process. This view breaks for biology. Instead, we believe that a new perspective is needed for a truly effective engineering

of biology; one that sees a designed biosystem as a starting point in a lineage of possibilities. Whilst much of evolutionary biology has concerned itself looking backwards at an organism's history [17], bioengineers must consider the future, and specifically how a biosystem will continue to evolve when used [18]. Here, we describe a framework to enable this transition offering a way to specify, test and conceive the properties of biosystems in terms of their evolutionary potential, and not just their phenotype (**Figure 1**). This provides a means to reimagine biological engineering in a way that works hand-in-hand with the ability for all life to evolve.

The design type and the evotype

To more fully harness the capabilities of biology, it is crucial to have a way of thinking about the evolutionary properties of engineered biosystems. We need to design not only the immediate functionalities of the system (i.e. its phenotypic traits), but also its potential for evolutionary change. Though these are properties of populations yet to exist, they can still be predicted for an individual biosystem. We consider the *design type* as the system that has been engineered consisting of a single genotype. We introduce the concept of the *evotype* to capture the evolutionary properties of a system. The evotype is the set of *evolutionary dispositions* of the design type, analogous to genotype and phenotype being sets of genes and traits, respectively. Unlike a trait, a disposition is not a directly observable property, rather it is a potential property of the system. For example, a protein may have the disposition of instability where its phenotype may change dramatically when mutated. Designing the dispositions of the evotype is a challenge fundamental to engineering biology.

For all but the very simplest biosystems, it is impractical to enumerate every potential evolutionary disposition, just as it is impossible to consider every trait of the phenotype. Instead, an appropriate sample of the evotype must be used for the purpose at hand, just as samples of traits are used when describing the phenotype. How we take this sample, and thus the scope of the evotype covered, should be determined by knowledge of the design type, its intended function, and the context in which it will be used. This could include the size of population, environment, and number of generations over which the system must operate reliably.

The key properties of the evotype can be understood by describing a landscape surrounding the design type in sequence space that captures the interwoven roles of genetic variation and natural selection, as well as the mapping between sequences (i.e. genotypes) and their associated function. The bioengineer's goal is to sculpt this landscape to their specification and ensure the landscape of the evotype has a structure in line with their requirements.

Variation

Not all point mutations are equally likely with transversions and transitions differing in their likelihoods [19], and methylation [20], genomic context [21], and species [22] all influencing local and global mutation rates. Furthermore, *algorithmic mutations* [23] may occur. These are mutations that result in changes of several nucleotides in one mutation event (thus an algorithm can describe the change) and can be thought of as shortcuts through sequence space (**Figure 1B, left**). The likelihood of an algorithmic mutation may be much greater than the sum likelihood of the equivalent sequence of individual point mutations. For example, the chance of an insertion of the two-base motif 'AC' into a tandem repeat region due to slipped strand mispairing may be more likely than two insertion events of 'A' and 'C' occurring independently [24]. Recombination [25] and mobile genetic elements [26] are other examples of biological processes capable of producing algorithmic mutations.

Sequence space is therefore not explored in a uniformly random way. Instead, the path evolution can take is determined by the *variation operator set*, which defines all the different point and algorithmic mutations that can occur in the system. Each variation operator in this set has an associated probability distribution that represents the likelihood of arriving at a given sequence from another (i.e. by this operator acting on the design type). The distributions of the variation operator set can be combined to produce the *variation probability distribution*, which describes the chance of arriving at a given sequence from the design type due to all the processes of genetic variation present in the system (**Figure 1B, right**). As a design type evolves, the variation probability distribution can be recalculated for each lineage in a population to understand further dispositions available to a system.

The variation operator set depends on the specifics of the biosystem being engineered, and the set to be used in practice is dependent on available knowledge of the system. For example, the variation operator set for a plasmid with many repeated parts may be said to include transition mutations, transversion mutations, and homologous recombination. A sample population can be generated by applying the operator set to the design type. This population, with the design type at its centre, may be named a *quasispecies*; as is used for the related concept in viral evolution [27].

The variation probability distribution can be considered in all stages of the engineering process. Global and local mutation rates could be specified in a design and standardised mutation rates could be listed in part datasheets and it is likely that improvements to the prediction of mutation probabilities will be made with the increasing availability of sequence data and improvements in computational methods. Design rules for influencing the local genetic variability are already known (e.g. avoiding the use of repeated parts to reduce homologous recombination and avoiding repeat sequences to avoid sequence mutation) [28], and global mutation rates can also be rationally engineered and manipulated [12,29]. Design

types could even be constructed with specific sets of variation operators. For example, making a system capable of recombination of certain genetic parts but not others. Highly specific recombination systems can also be directly built into the genome of a cell, such as the SCRaMbLE system found in the synthetic yeast Sc2.0 [30]. In the future it may be possible to design and build systems capable of carefully specified algorithmic mutations by controlling combinations of biochemical mechanisms such as recombination, CRISPR mediated modifications, and methylation to name but a few. Finally, advances and increased accessibility and throughput of quantitative sequencing [31] will enable better characterisation of a system's genetic variation in detail.

Function

Each genotype maps to a phenotype. However, these are not necessarily distributed evenly throughout sequence space, and physical and biological constraints mean that not all conceivable phenotypes may be possible. This means the topology of the genotype-phenotype map has an underlying structure, which influences evolvability [32]. When talking about phenotypes, we are really talking about traits or subsets of traits. As engineers, the trait we are interested in is the design type's *function* – the behaviour or properties specified by the designer. Thus, we are interested in the genotype-function map of the region surrounding our design type. We may group functions into distinct classes, or work with a continuous function space (**Figure 1C**). This could be a literal mathematical function like a logic function, a physical characteristic like colour or size, or a combination of several properties.

Any system has a degree of *utility* – the extent to which the system fulfils a specified/desired function. The sole goal of a traditional engineering design process is to maximise the utility of the design type. However, the topology of the function landscape surrounding the design type, and thus of the utility landscape, is also important. It may be highly variable with the function rapidly changing across sequence space, or it may have large neutral regions, where function changes little or remains constant. It may be rugged, where the function changes chaotically and unpredictably, or it may be smooth, where gradients can be predicted across larger distances in sequence space. These properties are a key design consideration: what is the functional range to be covered by the design type's evotype? Should it be large (e.g. for directed evolution), or limited (e.g. for robust devices)? Most likely the function landscape should be smooth and predictable, but how is this best achieved? Which regions of function space must be avoided, and which can be tolerated?

Designs may have identical functionality but occupy regions of function space with very different topological properties. If a system is designed without considering its surrounding function landscape, a design with an undesirable evotype is a likely outcome. For example, a rugged chaotic landscape where function is unstable, or a landscape where the desired

functions are inaccessible to directed evolution. We might engineer as much redundancy into our system as possible, so that mutations are less likely to result in dramatic changes of function (i.e. a flat or gently sloping landscape). Alternatively, we may want to maximise variability, however, this may be constrained to favour certain classes of function. For example, it may be necessary to reduce irrelevant or harmful functions as much as possible (e.g. in a diagnostic application regions of function space that causes false negatives must be avoided, whilst false positives can be tolerated). Engineering this landscape is not trivial: it requires significant knowledge of the physical, biochemical, and organisational properties of the biosystem. However, awareness of this challenge will enable engineers to think beyond the utility of the individual and consider the performance of future lineages as a whole.

Selection

Selection is the force that gives the otherwise random (but constrained) processes of genetic variation a 'direction', driving a population up the slopes of the adaptive landscape [33]. Uniquely, an engineered biosystem is a result of two forms of selection: natural selection and the design process. Natural selection acts on reproductive fitness of the biosystem, and the design process can be thought of as a sophisticated form of artificial selection acting on its utility. Understanding the interplay between these two processes is critical for good evotype design, since there is often a tension between utility and fitness (**Figure 1D**). If the two are uncorrelated, then natural selection will always strive to undo the work of the engineer. However, if fitness and utility are highly correlated, then natural selection will also drive up utility. It should be noted that natural selection here is meant as the process that acts on the reproductive ability of the biosystem. Neither the environment nor biosystem need to be natural (e.g. the organisms could be engineered to make use of non-canonical amino acids and grown within a bioreactor). The critical distinction is that natural selection acts on survival of the biosystem without the input of the engineer.

The aim of a bioengineer is to maximise *fitneity* – defined as a function that combines utility and fitness (**Figure 1D**). Biosystems with either high utility but low fitness or high fitness but low utility will both have low fitneity. To maximise fitneity both fitness and utility must be considered. How these two are weighted, and the exact function that makes up fitneity is dependent on the use-case and the properties of the system to be engineered. In some cases, utility may be prioritised, in others, the system may only require some minimal level of functionality, as long as it can survive under high natural selection pressures.

Careful design could ease the tension between utility and fitness by reducing the selection pressures of a design, for example by reducing metabolic load. The interplay between utility and natural selection can also be controlled, for instance by correlating the two in a directed evolution experiment. Evotype engineering then, is to sculpt the *fitneity landscape*

of the design type. The design type itself should have maximal fitneity, but a robust design must also be in a region of sequence space with high fitneity. Similarly, an evolvable design ideally has a fitneity landscape with smooth slopes and single peak that can be climbed by natural selection.

Conclusion

Like the genotype and phenotype, the evotype is a further way to think about the properties of engineered biosystems and how they relate to each other (**Table 1**). It is a framework for thinking about an important but often overlooked property: the role the biosystem itself plays in its future evolution. As engineered biosystems are the result of both human thought and natural adaptation, a holistic consideration of both the roles of design and evolution is necessary. The evotype helps us do this by explicitly considering the effects that variation, function, and selection will have on a design (**Figure 1**).

We can now design and build genotypes with great precision, but we must account for the inevitable processes of genetic variation that will follow. The statistical structure of variation is unique to the biosystem and something we have control over. Yet understanding the details of genetic variation is insufficient if we don't also understand how this will manifest in changes of the designed function of the biosystem. Even a system with low mutation rates can be evolutionarily unstable if function changes wildly with small sequence alterations. Directed evolution will not be successful, despite the mutation strategy, if desired functions are simply not accessible from the starting point. If the biosystem's utility (i.e. its success as a design) and its fitness (i.e. its success as a biological replicator) are at odds, well designed dispositions for variation or function might not save the design from the pressure of natural selection. This must also be understood as conflict between utility and fitness landscapes across sequence space surrounding the original design type. It is clear then that all three of the aspects of the evotype – variation, function, and selection – must be considered together, and all offer significant scope for engineering. For instance, imagine a large genetic circuit that places an unavoidably high metabolic burden on the host cell. If it is crucial that the function of the circuit is maintained over long periods of time, then redundancy could be used to accommodate unavoidable mutations. However, if the dent to reproductive fitness is severe, this may still not be enough. Therefore, combining redundancy in the design with a hyper-stable host cell (e.g. one where all mobile genetic elements have been deleted and efficient DNA repair mechanisms are present [34]) might be the only way to achieve the desired goal for the system.

Designing biosystems with evolution in mind is a step towards a more complete engineering theory of biology. However, to be practical, supporting tools must exist that can provide key information regarding the genetic variation, genotype-function mapping and selective pressures within a biosystem. Advances in sequencing offer a means to

quantitatively measure millions of genotypes in parallel [35] and when combined with techniques such as fluorescence activated cell sorting (FACS) make it possible to infer simplified genotype-function maps [36]. Even so, the vastness of evotype landscapes and the need for functions calculated from many outputs of a system mean that new methods with greater throughputs are also necessary, especially those able to measure many characteristics of each cell simultaneously (e.g. via automated high-content microscopy [37] or high-throughput Raman spectroscopy [38]). Parallel to these experimental methods, a promising direction to bypass the need to directly measure these properties are the development of sufficiently comprehensive computational models (e.g. encompassing whole cells [39]) to allow for a mechanistic understanding of the biases in processes related to variation and reproductive rate. In these cases, and if sufficiently accurate, the evotype could be predicted and used within computer aided-design workflows [40] to reduce the need to physically implement every possible design.

In addition to characterising evotypes, tools for bioengineers to directly sculpt their landscapes must also be available. (**Figure 2**). Here, we have touched upon the numerous ways that biomolecular components have been repurposed and genome engineering performed to alter the types of possible variation and selective pressures that are present. However, the spectacular diversity of molecular machines dedicated to manipulating genetic information found in the natural world lends support to the idea that we will likely require a large library to allow for precise modifications we might like to shape evotype landscapes any way that we want.

An engineering theory of evolution is both a new way of looking at evolutionary theory, and a new way of thinking about what it is that engineers do, and what the design process is. The concept of the evotype, with some modifications, may also find use in evolutionary science, where it offers a framework for considering the mechanistic constraints of evolution and a way of talking about the evolutionary characteristics of organisms. It may also be applied beyond biological engineering fields to create new auto-adaptive technologies. Here the framework could be applied to ask how we design technologies to evolve, and not just how to engineer systems that already do.

Acknowledgements

S.D.C. was supported by the EPSRC/BBSRC Centre for Doctoral Training in Synthetic Biology grant EP/L016494/1. T.E.G. was supported by a Royal Society University Research Fellowship grant UF160357. T.E.G. and C.S.G. were supported by BrisSynBio, a BBSRC/EPSRC Synthetic Biology Research Centre grant BB/L01386X/1. This study did not involve any underlying data.

Author Contributions

Whilst T.E.G. and C.S.G. framed questions and provided direction, the core concepts and new terminology here are mostly the work of S.D.C., who also wrote the first draft. S.D.C. and T.E.G. developed the figures with input from C.S.G. T.E.G. and C.S.G edited the manuscript and supervised the work.

Declaration of Interest

The authors declare no competing financial interests.

References

1. Andrianantoandro, E. *et al.* Synthetic biology: New engineering rules for an emerging discipline. *Molecular Systems Biology* **2**, 2006.0028 (2006).

2. Grozinger, L. *et al.* Pathways to cellular supremacy in biocomputing. *Nature Communications* 1–11 (2019) doi:10.1038/s41467-019-13232-z.

3. Dobzhansky, T. Nothing in Biology Makes Sense except in the Light of Evolution. *The American Biology Teacher* **35**, 125–129 (1973).

4. Renda, B. A., Hammerling, M. J. & Barrick, J. E. Engineering reduced evolutionary potential for synthetic biology. *Mol. BioSyst.* **10**, 1668–1678 (2014).

5. Ellis, T. Predicting how evolution will beat us. *Microbial Biotechnology* **12**, 41–43 (2019).

6. Fernandez-Rodriguez, J., Yang, L., Gorochowski, T. E., Gordon, D. B. & Voigt, C. A. Memory and Combinatorial Logic Based on DNA Inversions: Dynamics and Evolutionary Stability. *ACS Synthetic Biology* **4**, 1361–1372 (2015).

7. Yokobayashi, Y., Weiss, R. & Arnold, F. H. Directed evolution of a genetic circuit. *Proceedings of the National Academy of Sciences of the United States of America* **99**, (2002).

8. Giver, L., Gershenson, A., Freskgard, P.-O. & Arnold, F. H. Directed evolution of a thermostable esterase. *PNAS* **95**, 12809–12813 (1998).

9. Boder, E. T., Midelfort, K. S. & Wittrup, K. D. Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity. *Proceedings of the National Academy of Sciences* **97**, 10701–10705 (2000).

10. Esvelt, K. M., Carlson, J. C. & Liu, D. R. A system for the continuous directed evolution of biomolecules. *Nature* **472**, 499–503 (2011).

11. Guntas, G., Mansell, T. J., Kim, J. R. & Ostermeier, M. Directed evolution of protein switches and their application to the creation of ligand-binding proteins. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 11224–11229 (2005).

12. Wang, H. H. *et al.* Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894–898 (2009).

13. Anderson, J. *et al.* Engineering and ethical perspectives in synthetic biology. Rigorous, robust and predictable designs, public engagement and a modern ethical framework are vital to the continued success of synthetic biology. *EMBO Rep* **13**, 584–590 (2012).

14. Wright, O., Stan, G.-B. & Ellis, T. Building-in biosafety for synthetic biology. *Microbiology (Reading, Engl.)* **159**, 1221–1235 (2013).

15. Chan, C., Lee, J., Cameron, E., Bashor, C. & Collins, J. 'Deadman' and 'Passcode' microbial kill switches for bacterial containment. *Nat Chem Biol* **12**, 82–86 (2015).

16. Mandell, D. J. *et al.* Biocontainment of genetically modified organisms by synthetic protein design. *Nature* **518**, 55–60 (2015).

17. Pigliucci, M. Are ecology and evolutionary biology "soft" sciences? *Ann. Zool. Fennici* **39**, 87–98 (2002).

18. Bartley, B. A., Kim, K., Medley, J. K. & Sauro, H. M. Synthetic Biology: Engineering Living Systems from Biophysical Principles. *Biophys J* **112**, 1050–1058 (2017).

19. Stoltzfus, A. & Norris, R. W. On the Causes of Evolutionary Transition:Transversion Bias. *Mol Biol Evol* **33**, 595–602 (2016).

20. Jones, P. A., Rideout, W. M., Shen, J. C., Spruck, C. H. & Tsai, Y. C. Methylation, mutation and cancer. *Bioessays* **14**, 33–36 (1992).

21. Zhu, Y. O., Siegal, M. L., Hall, D. W. & Petrov, D. A. Precise estimates of mutation rate and spectrum in yeast. *Proc. Natl. Acad. Sci. U.S.A.* **111**, E2310-2318 (2014).

22. Drake, J., Charlseworth, B., Charlseworth, D. & Crow, J. Rates of Spontaneous Mutation. *Genetics* **148**, 1667–1686 (1998).

23. Chaitin, G. *Proving Darwin: Making Biology Mathematical.* (Vintage Books, 2013).

24. Levinson, G. & Gutman, G. A. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* **4**, 203–221 (1987).

25. Vos, M. Vos M.. Why do bacteria engage in homologous recombination? Trends Microbiol 17: 226-232. *Trends in microbiology* **17**, 226–32 (2009).

26. Frost, L. S., Leplae, R., Summers, A. O. & Toussaint, A. Mobile genetic elements: the agents of open source evolution. *Nat. Rev. Microbiol.* **3**, 722–732 (2005).

27. Eigen, M. On the nature of virus quasispecies. *Trends Microbiol.* **4**, 216–218 (1996).

28. Jack, B. R. *et al.* Predicting the Genetic Stability of Engineered DNA Sequences with the EFM Calculator. *ACS Synth Biol* **4**, 939–943 (2015).

29. Geng, P., Leonard, S. P., Mishler, D. M. & Barrick, J. E. Synthetic Genome Defenses against Selfish DNA Elements Stabilize Engineered Bacteria against Evolutionary Failure. *ACS Synth. Biol.* **8**, 521–531 (2019).

30. Dymond, J. & Boeke, J. The Saccharomyces cerevisiae SCRaMbLE system and genome minimization. *Bioeng Bugs* **3**, 168–171 (2012).

31. Gorochowski, T. E. *et al.* Absolute quantification of translational regulation and burden using combined sequencing approaches. *Mol. Syst. Biol.* **15**, e8719 (2019).

32. Ahnert, S. E. Structural properties of genotype– phenotype maps. *Journal of The Royal Society Interface* **14**, 20170275 (2017).

33. Kauffman, S. A. *The Origins of Order: Self-Organization and Selection in Evolution.* (Oxford University Press, USA, 1993).

34. Umenhoffer, K. *et al.* Genome-Wide Abolishment of Mobile Genetic Elements Using Genome Shuffling and CRISPR/Cas-Assisted MAGE Allows the Efficient Stabilization of a Bacterial Chassis. *ACS Synth. Biol.* **6**, 1471–1483 (2017).

35. Reuter, J. A., Spacek, D. V. & Snyder, M. P. High-throughput sequencing technologies. *Mol. Cell* **58**, 586–597 (2015).

36. Sharon, E. *et al.* Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. *Nature Biotechnology* **30**, 521–530 (2012).

37. Pepperkok, R. & Ellenberg, J. High-throughput fluorescence microscopy for systems biology. *Nat. Rev. Mol. Cell Biol.* **7**, 690–696 (2006).

38. Goda, K. *et al.* High-throughput single-microparticle imaging flow analyzer. *PNAS* **109**, 11630–11635 (2012).

39. Marucci, L. *et al.* Computer-Aided Whole-Cell Design: Taking a Holistic Approach by Integrating Synthetic With Systems Biology. *Front. Bioeng. Biotechnol.* **8**, 942 (2020).

40. Nielsen, A. A. K. *et al.* Genetic circuit design automation. *Science* **352**, aac7341– aac7341 (2016).

41. Mozhaev, V. V. & Martinek, K. Structure-stability relationships in proteins: new approaches to stabilizing enzymes. *Enzyme and Microbial Technology* **6**, 50–59 (1984).

42. Archetti, M. Genetic robustness and selection at the protein level for synonymous codons. *Journal of Evolutionary Biology* **19**, 353–365 (2006).

Tables

Table 1: Role of genotype, phenotype and evotype when describing biological systems

	Genotype	Phenotype	Evotype
Describes	Informational/ Hereditary properties	Physical/Environmental properties	Evolutionary properties
Is a set of...	Genes/Sequences	Traits	Dispositions
Pig example	Hox gene	Four legs	Chance of evolving wings
Protein example	Codon	Structural stability under temperature [a]	Structural stability under mutation [b]

a.   For example, strengthened hydrophobic interactions in the interiors of thermostable proteins [41].

b.   For example, the robustness of the genetic code to amino acid mutations due to synonymous codons [42].

Figures and captions
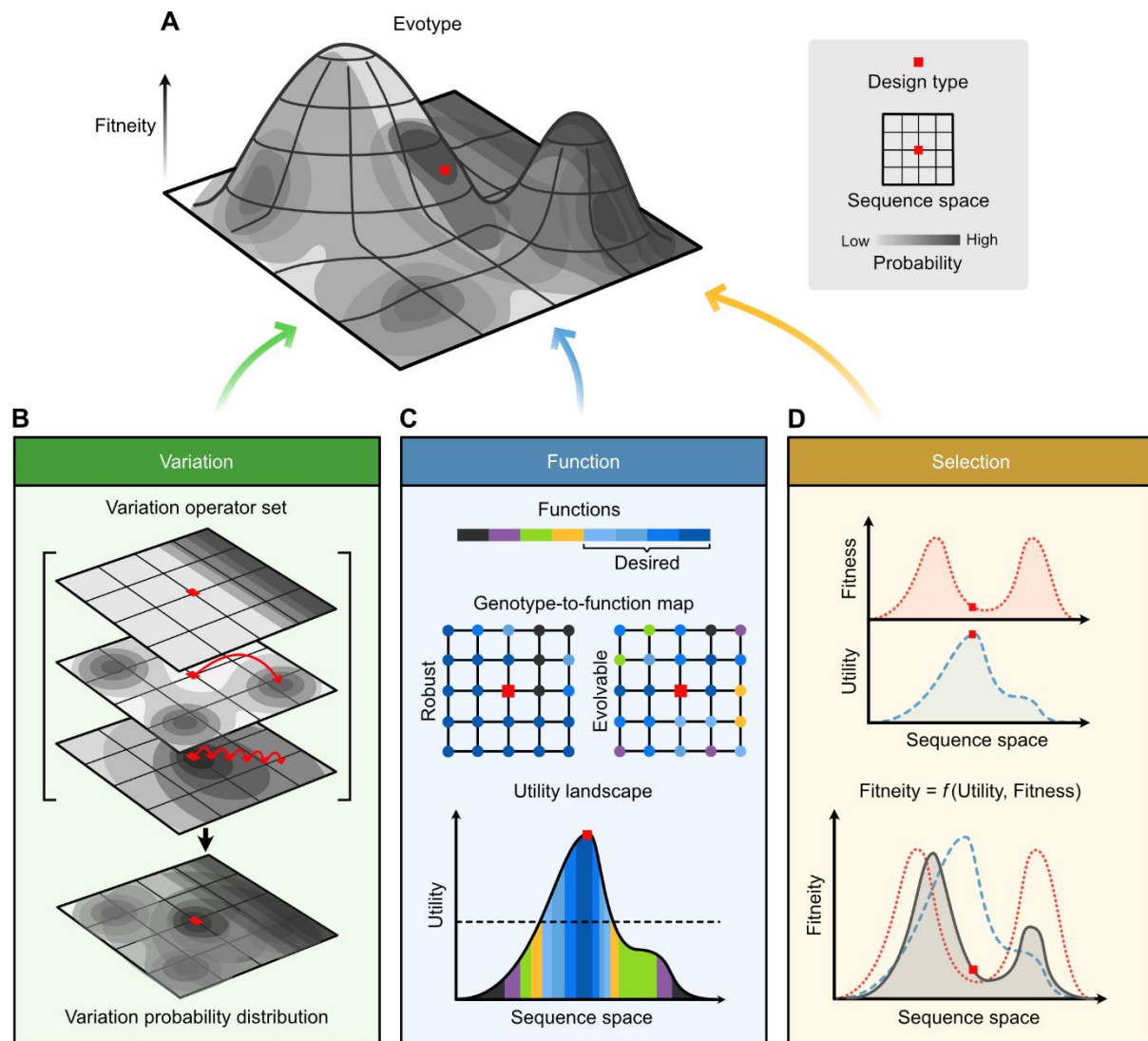
Figure 1: The evotype and its key properties. (**A**) The evotype visualised as a landscape surrounding the design type (red square), where fitneity (the combined function of fitness and utility) is plotted as a vertical axis against a 2D plane of sequence space with the probability of evolution exploring regions of sequence space overlaid in grey. The properties of this landscape are determined by the interaction of three components: variation, function, and selection. (**B**) A variation probability distribution can be projected onto sequence space, which represents the likelihood of exploring a given sequence through genetic variation. Darker regions represent regions of higher probability. This is the sum of the distributions of the individual variation operators present in the system (variation operator set). For example, point mutation (bottom layer in set), recombination of homologous regions (middle layer in set), and slip-strand mutation (top layer in set). Red arrows in middle and bottom layers represent algorithmic and point mutations respectively. (**C**) How phenotypic functions are distributed in sequence space surrounding the design type is critical. Function space may be considered as

discrete (top), where the space may have high genotypic robustness (left grid) or high variability (right grid). A continuous utility space (bottom) plotted against a 1D projection of sequence space. The colour under the curve represents the discrete function associated with that region of sequence space and the utility that each has as a continuous value. For example, if the goal is to produce blue-like functions, dark blue may have the highest utility, followed by lighter variants in the spectrum. The bioengineer must define a minimal threshold (dashed line), below which the design is deemed to be a failure (e.g. non-desired function is exhibited). (**D**) Sequences differ in their reproductive fitness. This is the driver of natural selection and can be plotted across sequence space as an adaptive landscape (red dotted line). Utility (blue dashed line) may or may not correlate with reproductive fitness across sequence space. The fitneity (grey solid line) is the combination of the fitness and utility. Bioengineers must optimise fitneity both for the design type and throughout the landscape.
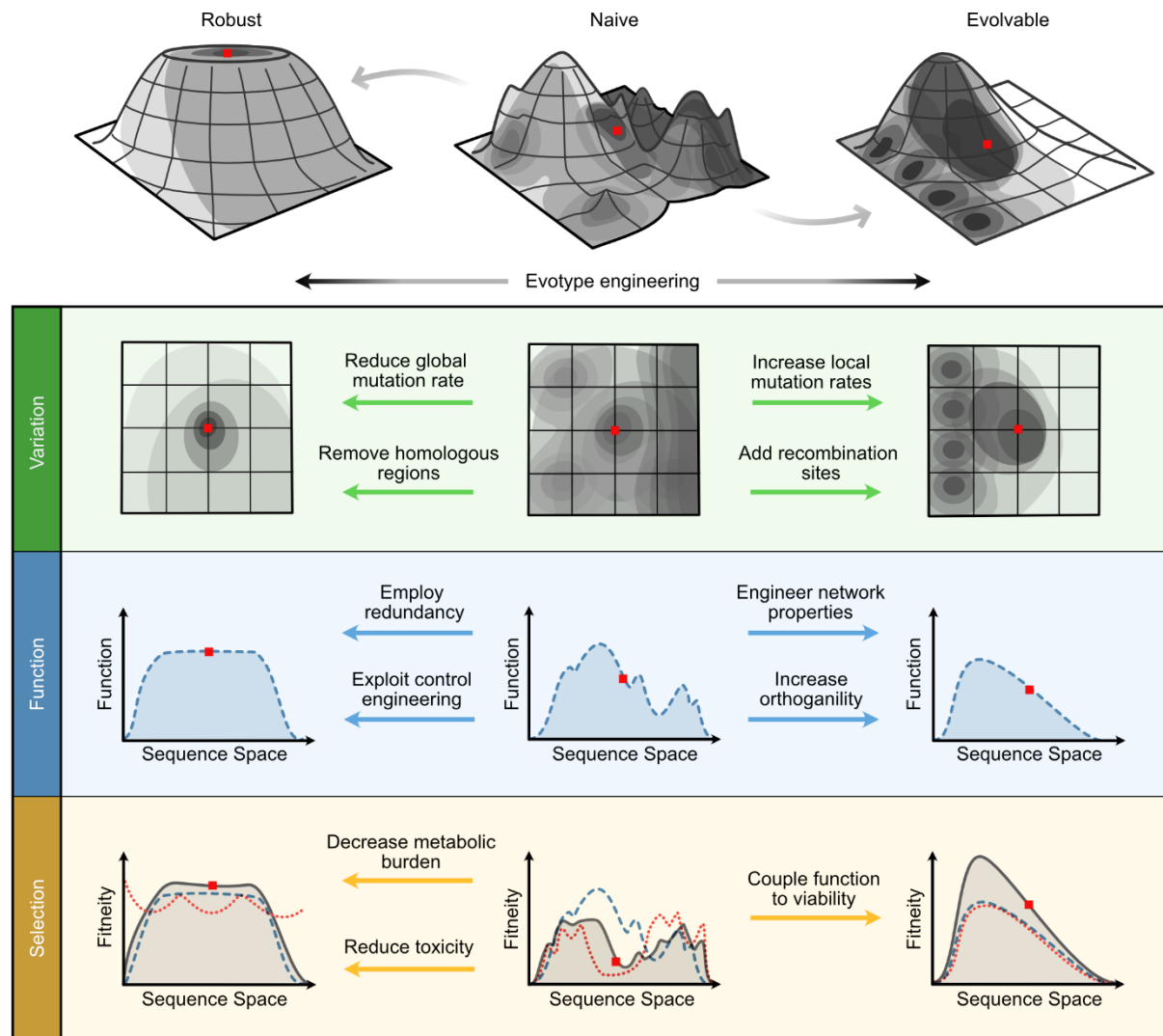
Figure 2: Engineering evotypes by sculpting their landscapes. Different biosystem designs may share the same phenotype but have very different evotypes (top row). Rational engineering approaches could be used to transform a naive design (middle column), where evotype has not been considered, into either robust (left column) or evolvable (right column) evotypes, which are characterised by their fitneity landscapes. Bioengineers can sculpt the evotype by modifying three major factors: variation, function and selection. Variation (green row): In a naive design, a mixture of variation operators may be in play. This might create a system that can reach many different regions of sequence space. It could be made more robust by reducing global mutation rates (e.g. host strain engineering), or by removing homologous regions to reduce the chance of recombination. Conversely, a naive design might be made more evolvable by increasing mutation rates in focused areas of sequence space (e.g. via methylation) and incorporating site-specific recombination or gene-shuffling (e.g. the SCaMbLE system). Function (blue row): A naive design may have high utility, however, if its function changes rapidly and chaotically across sequence space, it may be inherently unstable. A robust evotype has large neutral regions in function space. This may be

engineered by designing in redundancy (e.g. redundant genes), by employing part equivalence (e.g. amino acid mutations with similar properties), or through the application of control engineering principles to accommodate variability. Conversely a design can be made more evolvable if it can access a large range of new phenotypes, of a specific class (e.g. produce a colour), and the landscape may be smoothed (e.g. through removing crosstalk between features) and thus made amenable to evolutionary search. Selection (orange row): If, as in the naive design, reproductive fitness (red dotted line) and utility (blue dashed line) are highly uncorrelated, then the design type may have a strong selection pressure acting against it, and regions where both fitness and utility are maximised may be rare or non-existent, so high fitneity (grey solid line), may not be achievable. In a robust design, one might act to reduce the effects of natural selection through global increases in fitness (e.g. through reducing metabolic load of a genetic circuit), by reducing toxicity of gene products, or by smoothing either fitness or utility landscapes (e.g. through orthogonal parts). A naive design can be made more evolvable by closely correlating fitness and utility (e.g. through coupling function to survival). This means natural selection will act to drive up the utility of the design: the precise goal of a directed evolution experiment.