


# A Microfluidic Platform for Directed Evolution of Dynamic Phenotypes

Jessica James

DPhil Thesis

Department of Engineering Science

Supervisor: Harrison Steel

# Acknowledgements

I firstly wish to thank my supervisor Harrison Steel, with whom I feel immensely fortunate to have studied. I have found Harrison's guidance, both on scientific and career matters, to be extremely valuable, and as such he will remain a fundamental role model for me as I move on in my career. Not only do I need to thank Harrison for his guidance, but also for the opportunities he has granted. I believe Harrison's lab is quite unique in the extent to which it enables students to cross interdisciplinary boundaries, and his encouragement to take on entirely unfamiliar challenges has not only allowed me to develop skills I might not otherwise have had the opportunity to, but has also taught me not to be daunted by problems I don't immediately know how to solve.

I next need to thank Idris Kempf, with whom I have worked closely throughout this project. Idris is an additional role model for me for his scientific drive and dedication. We have shared many of the successes and frustrations of this project, including early mornings and late nights in the lab, and he has always been a supportive and encouraging collaborator. On a related note, I am also grateful to the England men's and women's football teams for consistently defeating Switzerland and ensuring that Idris remains indebted to me in pints.

Another key collaborator without whom this DPhil would not be the same is Sebastian Towers. Seb has worked with me on the theory work surrounding the project, which forms the first chapter of this thesis. Seb's computational and mathematical abilities are highly admirable and it has been exciting to build this work alongside him. Another area in which Seb excels is making everyone laugh, so I

wish to thank him also for ensuring the office is always an enjoyable place to be.

I would then like to thank the rest of the lab, for creating an exceptionally fun and exciting work environment, and for forming an invaluable advice network from which I have learned a lot. In addition, thank you to those who support the work in the Engineering Science Department, notably Sara Kingsbury, Alice Robertson, and Lisa Zillig who have enabled our lab to function smoothly, and to Duncan Constable and Peter Walters, who have assisted me in many of my manufacturing needs throughout the project.

I would finally like to thank my family for their continual love and support throughout this work and beyond, as well as my partner Jamie, who has ensured that life outside of work is always a source of happiness and grounding.

# Table of Contents

**Abstract** **1**

**Publication Summary** **2**

**Introduction** **4**

**1 Literature Review** **9**

**2 Theory: Tools for sequence-free optimisation of directed evolution** **26**

2.1 Featured publications . . . . . 26

2.2 Abstract . . . . . 26

2.3 Background . . . . . 27

2.4 DE optimisation tools . . . . . 30

2.4.1 Selection functions . . . . . 30

2.4.2 Population splitting . . . . . 36

2.4.3 Multi-dimensional selection . . . . . 38

2.5 Sequence-free fitness landscape inference with SLIDE . . . . . 41

2.6 A combined pipeline for optimisation of directed evolution . . . . 47

2.7 Methods . . . . . 51

2.7.1 Directed evolution model . . . . . 51

2.7.2 NK landscapes . . . . . 52

2.7.3 Selection functions . . . . . 53

2.7.4 Empirical landscapes . . . . . 54

2.8 Discussion . . . . . 55

2.9 Statement of authorship . . . . . 58

**3 Hardware: Development of an automated platform for light targeting and long-term observation of cells on microfluidic chips 59**

3.1 Featured publications . . . . . 59
3.2 Abstract . . . . . 59
3.3 Background . . . . . 60
3.4 System components . . . . . 63
3.4.1 System overview . . . . . 63
3.4.2 ASI microscope . . . . . 65
3.4.3 Stage . . . . . 66
3.4.4 Incubator . . . . . 68
3.4.5 Optics . . . . . 69
3.4.6 Microfluidics . . . . . 73
3.4.7 System control . . . . . 74
3.5 Establishing chip fabrication capabilities . . . . . 75
3.5.1 Lab setup . . . . . 75
3.5.2 Microfluidic chip fabrication . . . . . 76
3.5.3 Epoxy cast fabrication . . . . . 78
3.6 Long term culture of MG1655 cells in the mother machine . . . . 81
3.6.1 Cell loading . . . . . 81
3.6.2 Long term observation of cells in the mother machine . . . 82
3.6.3 Interacting with cells in the mother machine . . . . . 85
3.7 Discussion . . . . . 87
3.8 Statement of authorship . . . . . 90

**4 Application: Photoselection of cells to enable directed evolution of dynamic phenotypes** **92**

4.1 Featured publications . . . . . 92

4.2 Abstract . . . . . 92

4.3 Background . . . . . 93

4.4 Results . . . . . 97

4.4.1 Characterisation of growth inhibition with light . . . . . 97

4.4.2 Quantification of off-target effects . . . . . 102

4.4.3 Implementation of enrichment by photoselection . . . . . 106

4.4.4 Refinement of photoselection outcomes with growth-based feedback . . . . . 107

4.5 Methods . . . . . 111

4.5.1 Media and Strains . . . . . 111

4.5.2 Microfluidic Experiments . . . . . 113

4.6 Discussion . . . . . 114

4.7 Statement of authorship . . . . . 117

**Discussion** **118**

**Bibliography** **123**

## Abstract

Directed evolution is a strategy for engineering novel biological functions without requiring detailed sequence-to-function knowledge. It consists of iterative rounds of mutagenesis and selection designed to encourage development of a functional goal. Widely used high-throughput selection approaches such as fluorescence-activated cell sorting, magnetic-activated cell sorting, and droplet-based selection are highly effective, but are inherently limited to static or endpoint phenotypes, precluding selection on dynamic cellular behaviours.

The aim of this DPhil was to develop a directed evolution platform that could overcome this limitation and enable directed evolution of dynamic phenotypes. This was achieved using mother machine microfluidic chips, which allow long-term observation of up to $10^6$ individual cells under tightly controlled conditions. Selection in this context was achieved through a custom-built hardware system capable of projecting high-power, spatially targeted UV light to selectively eliminate undesired cells, thereby controlling the development of population properties over many generations. In addition to the above process we call "photoselection", the platform enables interaction with cells via chemical, optogenetic, and more recently magnetic stimuli, providing multiple orthogonal modes of phenotypic perturbation on which we can select.

This thesis first describes a theoretical framework for directed evolution in sorting-based systems with a focus on optimisation without sequencing data. It then presents the design and implementation of the integrated hardware and software platform, and finally

demonstrates its application to single-cell photoselection in the mother machine. Together, this work establishes a new platform for evolving dynamic phenotypes - such as biosensors and complex gene circuits - in a manner that is inaccessible to current state-of-the-art methodologies.

***Keywords—*** Directed evolution - Microfluidics - Mother Machine

# List of Abbreviations

**DE** . . . . . Directed Evolution

**MiPS** . . . . . Microscopic PhotoSelection

**FACS** . . . . . Fluorescence Activated Cell Sorting

**MACS** . . . . . Magnetic Activated Cell Sorting

**DMD** . . . . . Digital Micromirror Device

**SLIDE** . . . . . Sequence-free Landscape Inference for Directed Evolution

**UV** . . . . . Ultraviolet

**LED** . . . . . Light Emitting Diode

**DNA** . . . . . Deoxyribonucleic Acid

**RNA** . . . . . Ribonucleic Acid

**MB** . . . . . Methylene Blue

# Publication Summary

This thesis includes work that has contributed to the following publications:

## Primary research outputs

- **Jessica James***, Sebastian Towers*, Jakob Foerster, Harrison Steel. "Optimisation strategies for directed evolution without sequencing". *PLOS Computational Biology*. Dec 2024. DOI: 10.1371/journal.pcbi.1012695.
- Sebastian Towers*, **Jessica James***, Harrison Steel, Idris Kempf. "Sequence-free landscape inference for directed evolution". *BioRxiv, under review at PLOS Computational Biology*. Oct 2025. DOI: 10.1101/2025.10.08.
- **Jessica James***, Idris Kempf*, Kirill Sechkar, Jingyu Wang, Gabriel Abrahams, Sebastian Towers, Harrison Steel. "Microscopic PhotoSelection (MiPS) of single cells in mother machine microfluidic devices". *BioRxiv*. Dec 2025. DOI: 10.64898/2025.12.03.692065.

*Co-first author

## Review articles

- Idris Kempf *et al.* "High-throughput, quantitative single-cell methods for next-generation diagnostics and therapeutics in bacteria". (under review at Advanced Science).

# Co-authored publications

- Gabriel Abrahams *et al*. "Quantum spin resonance in engineered proteins for multimodal sensing". *Nature*. Jan 2026. DOI:10.1038/s41586-025-09971-3.
- Georgie Hau Sorensen *et al*. "Structure-guided design of protein attachment points for functional augmentation of complex molecular machines". *BioRxiv, under review at Nature Communications*. Jun 2025. DOI: 10.1101/2025.06.22.660945.

# Introduction

Engineered biological systems hold immense potential for application across industries including medicine, manufacturing, and agriculture [1–3]. In recent decades, protein engineering in particular has demonstrated the potential of natural biological elements to be adapted for new functionalities. Once such method for protein engineering is known as "directed evolution" [4]. Directed evolution techniques have delivered products across a range of applications, from cancer and autoimmune disorder drugs [5] to enzymes for converting cooking oil into bio diesel [6].

In a process that mimics nature, directed evolution consists of iteratively introducing random variation by mutagenesis, followed by selection biased toward desirable user-defined traits (Fig. 1). Directed evolution methods are highly varied, and exist for both *in vivo* and *in vitro* settings. Selection of living cells can be achieved broadly via two families of approaches; first, those that couple a trait-of-interest to growth of a host organism [7, 8], or second, those that couple a trait to

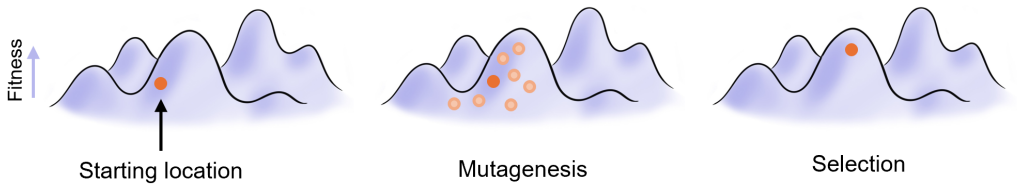



Figure 1: **Mutagenesis and selection.** Iterative rounds of mutagenesis and selection are the driving forces behind directed evolution. This schematic represents directed evolution as navigation over a fitness landscape, where higher positions represent higher fitness. Mutagenesis allows spreading out on that landscape, and selection allows the population to be narrowed back down to the new highest fitness point(s).

a measurable output (e.g. expression of fluorescent proteins). In the latter, cells are screened to identify and isolate those with desirable output expression levels for future mutagenesis and propagation [9]. Each approach has different strengths and weaknesses. Growth-coupled selection utilises (comparatively) straightforward growth-based assays but requires engineering of trait-to-growth coupling, which is both challenging and can lead to "cheating" behaviour [10]. Meanwhile, screening-based methods only require a trait be *measurable* (i.e. they do not require coupling to growth), but in turn necessitate more complex experimental approaches to implement the selection for the measured winners (e.g. Fluorescence-Activated Cell Sorting, FACS [9]).

FACS is a widely used high-throughput method in which individual cells flow past a laser, are measured for fluorescence and scatter properties, and are physically sorted into separate populations based on predefined thresholds. Although FACS has enabled major advances in high-throughput biological research, it remains fundamentally limited to snapshot measurements of each cell. As a result, dynamic responses on fast timescales cannot be profiled. For example, MagLOV proteins exhibit diverse response kinetics at sub-second timescales [11], and these behaviours are inaccessible to conventional FACS measurements.

Snapshot measurements also restrict experiments to a single stimulus per selection round. In biosensor development, however, one may wish to select simultaneously for sensitivity and specificity by testing a sequence of compounds. Consider a panel of compounds A-F: in a conventional FACS workflow, cells are first screened against compound A, and any biosensors showing an undesirable response are discarded. This discarded population may nevertheless contain variants

that would have performed well overall when evaluated across compounds B-F. Because decisions are made from incomplete information, potentially valuable candidates can be lost early in the selection process. These limitations highlight a gap in current high-throughput selection methods that this work seeks to address.

Microfluidics, specifically the "Mother machine" [12], offers a potential solution to these issues. The mother machine design consists of a central channel, lined with one-cell-thick "trenches". Once loaded into these trenches, bacterial cells can grow stably in exponential phase for hundreds of generations [13]. It is therefore possible to observe cells for long time periods and monitor their responses to stimuli in real time, something that is not possible with FACS. While this capability enables high-throughput screening (i.e. measurement and classification) of dynamic phenotypes, a further challenge is to convert such measurements into a true selection platform, in which desired cells are enriched while undesired cells are removed. A key goal of this project was therefore to implement selection in this format, enabling downstream directed evolution of dynamic phenotypes that are invisible to FACS.

The primary challenge of this work, therefore, was to reliably extract cells from the mother machine based on their properties, in an automated fashion. Extraction of cells from a mother machine has been achieved before in a low-throughput, manual format, using optical tweezers [14], and laser ablation [15]. Taking inspiration from this former work, this platform was designed to enable automated, high-throughput selection of cells in such a way that is capable of selecting the top e.g. 5% of a $10^6$ cell population, similar to a FACS regime. It was achieved using a Digital Micromirror Device (DMD), which is an array of controllable mi-

cromirrors that can be used to dictate the pattern of light that shines onto cells (and target individual cells). DMDs have been used in a mother machine context before, to deliver optogenetic signals [16], however in this case a DMD is being used to deliver high-power UV light to the sample to destroy unwanted cells. The entire system is optimised for the purpose with custom hardware and software.

The final goal of this work would be to apply it in a fully-automated, iterative directed evolution platform (Fig. 2). Mutagenesis would first be induced in a population of cells carrying the gene of interest, to produce an array of variants. These variants would then be measured in the mother machine via a fluorescent reporter, selection applied with UV, and the new population of cells proliferated ahead of the next round. This thesis, however, focuses on optimising the selection component, and full closure of the loop shown remains for future work.

This thesis is divided into three main chapters: Theory, Hardware, and Application. The Theory chapter describes work from two papers that was conducted alongside the development of the experimental platform. Because the platform is well suited to continuous directed evolution - that is, evolution without iterative DNA sequencing - this work explored how directed evolution can be optimised in sequence-agnostic contexts. The Hardware chapter describes the process of building the platform, and concludes with several examples of bacterial cells being successfully cultured, analysed, and interacted with in the mother machine. Finally, the Application chapter describes full characterisation of the UV selection process that this platform is optimised to perform.

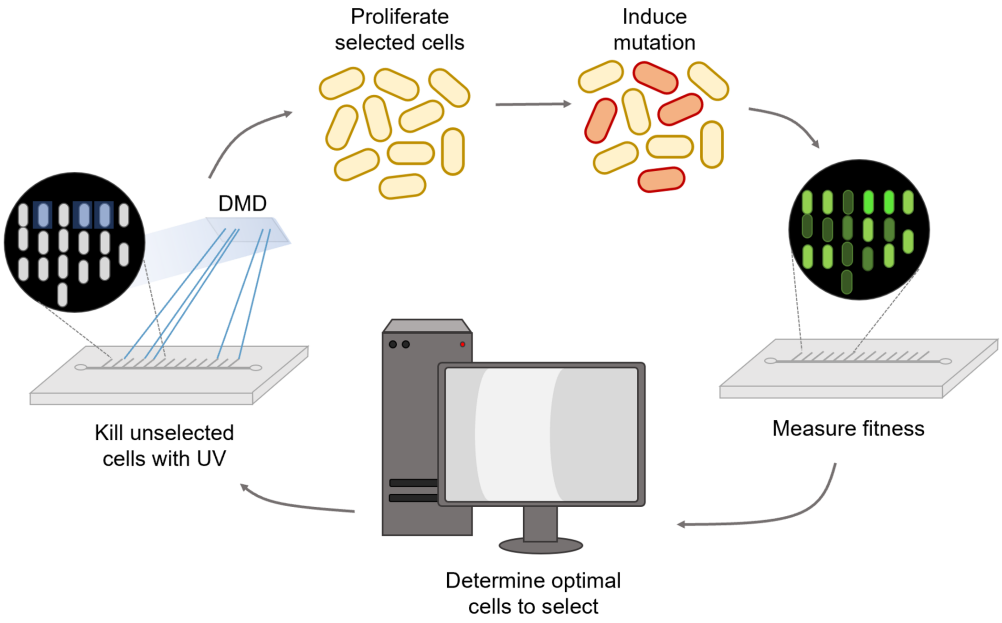



Figure 2: **A future automated platform for directed evolution.** In this platform, cells carrying the gene of interest would have mutagenesis induced to produce an array of variants. Via a fluorescent reporter, the fitness of these cells would be measured in the mother machine microfluidic chip, which leads directly on to automated selection with targeted UV. The surviving population which flows out of the chip (daughter cells of the selected population that continue to proliferate on the chip) would become the next population to be proliferated ahead of the following round.

# 1 | Literature Review

## Protein engineering and directed evolution

All proteins are composed of 20 available building blocks, known as amino acids. Proteins vary enormously in size, and the amino acid repertoire encompasses a wide range of chemical properties (e.g. charge, acidity, and hydrophobicity). Consequently, the possible structures and functions that proteins can adopt are extraordinarily diverse. Given that no man-made technology can match the efficiency and versatility of proteins at such minute scale, harnessing their abilities is of considerable interest to industries including medicine, manufacturing and agriculture.

Advances in AI tools have dramatically accelerated the rate of progress in *de novo* protein design. Within the time frame of this DPhil alone, the field has progressed from breakthroughs in structure prediction with AlphaFold 2 [17] and RoseTTAFold [18], to active protein *design* using powerful generative models such as RFdiffusion [19] and ProteinMPNN [20], which now form core components of protein design pipelines. However, at this time the output of AI pipelines are still considered *predictions* as opposed to guaranteed functional proteins. By contrast, directed evolution is a laboratory-based protein engineering approach in which success is defined by validated functionality. It involves iterative rounds of mutation and selection designed to steer a protein towards a desired functional outcome. Directed evolution can be used in conjunction with AI-driven design pipelines or as a standalone strategy.

Early directed evolution experiments focused on relatively modest modifications of enzymes - a class of proteins that catalyse chemical reactions [21]. These modifications included increasing reaction rates and improving thermostability, properties that are fundamental to the economic viability of large-scale enzymatic processes [22]. Since then, the scope of applications has expanded considerably across manufacturing and therapeutics [5, 6, 23]. The first marketed drug developed using directed evolution was Humira (adalimumab), a treatment for conditions such as rheumatoid arthritis and Crohn's disease. Humira is now prescribed to over two million people per year in the United States alone and remains one of the best-selling drugs of all time [24].

# Methods for directed evolution

## Mutagenesis methods

A directed evolution experiment consists of two key elements: a method of mutagenesis and a method of selection. Mutagenesis strategies increase the rate at which changes (mutations) are introduced into the DNA sequences encoding proteins, thereby accelerating evolutionary processes. The most basic approaches are global *in vivo* methods, such as exposure to UV light or DNA-damaging agents, which induce mutations across the entire genome. If a specific gene is to be targeted, mutagenesis is traditionally performed outside the cell before the modified DNA is reintroduced via transformation. Such *in vitro* techniques - including error-prone PCR, DNA shuffling, and randomised oligonucleotide pools [25] - require greater manual effort and resources. In addition, efficient transformation of

DNA into cells presents a common technical bottleneck, as uptake efficiency is often low. This limitation is particularly problematic in directed evolution experiments, where entire libraries of variant genes are generated and loss of diversity is highly undesirable.

By manipulating cellular machinery to target mutations to specific regions of DNA, scientists have created *in vivo* methods of mutagenesis that offer the specificity of *in vitro* methods while eliminating the need to work with DNA outside of the cell [26]. Such methods include MutaT7 [27], EvolvR [28], BacORep [29] and CTRLE [30]. Given the huge reduction in labour and resources per iteration afforded by these approaches, they allow for many more iterations of mutation-selection to be carried out in the same time frame.

MutaT7 is a targeted *in vivo* mutagenesis system that uses a modified T7 RNA polymerase and T7 promoter. A promoter is a DNA sequence upstream of a gene that regulates its expression by recruiting the cellular machinery responsible for protein production. Transcription, the first stage of protein synthesis, is catalysed by RNA polymerases, which bind to the promoter before processing along the DNA strand. In the MutaT7 system, the target gene is placed under a T7 promoter, which is specifically bound by a T7 RNA polymerase. In MutaT7 this T7 RNA polymerase is tethered to a deaminase enzyme, which introduces mutations to the DNA as the RNA polymerase processes along. The MutaT7 system has been modified in a number of ways since its original implementation. Such modifications include increasing the mutation rate [31], using a dCas9 module to terminate mutation mid-gene [32], using a library of deaminases to widen the mutational range [33], re-engineering it to work in yeast [34] and mammalian

cells [35], or enabling dynamic control of deaminase attachment [36]. The main advantage of MutaT7 is that there is a well-defined mutational window that can encompass entire genes (up to at least 3 kb [37]). The main disadvantage is that it is only capable of transition mutations (A <-> G, C <-> T), and cannot access the entire spectrum of possible mutations.

In the EvolvR system, a DNA-cutting enzyme known as nCas9 makes a nick in the DNA at the site of interest (guided by an RNA strand that binds to the specific DNA region). The site is then repaired by an error-prone DNA polymerase enzyme, which introduces mutations during the process [28]. EvolvR has been shown able to access all possible mutation types, and has been expanded from *E. coli* to yeast [38] and mammalian cell applications [39]. The main disadvantage of EvolvR, however, is that the mutational window is small (<350 bp). It is possible to increase this by using a range of overlapping guide RNAs, but that adds a significant degree of technical complication.

Similarly to EvolvR, the recently-developed CTRLE system also leverages a nCas9-based system to accelerate mutagenesis. In CTRLE, however, the nCas9 nick is repaired by a highly processive error-prone phage polymerase that can induce mutations in much larger stretches than that implemented with EvolvR [30]. As a result, the mutational window achieved is up to 2kb, overcoming the key limitation of EvolvR while maintaining the ability to cover all possible mutation types.

Plasmids are usually rings of DNA that are separate to an organism's main genome. In the BacORep system (which is a variant of the original OrthoRep system), the target gene is placed on a "linear plasmid", which is specifically targeted by an error-prone DNA polymerase for replication. The system is capable of large mu-

| Method | Host compatibility | Mutation rate (substitutions per base) | Mutational window | Mutational spectrum | Defined window | *E. coli* Ref |
|---|---|---|---|---|---|---|
| MutaT7 | *E. coli*, yeast, mammalian cells | $9.4 \times 10^{-5}$ | $<$ 3kb | 4/12 | Yes | [33] |
| EvolvR | *E. coli*, yeast, mammalian cells | $10^{-6}$ to $10^{-5}$ | $<$ 350bp | 12/12 | No | [28] |
| CRTLE | *E. coli, B. subtilis, K. lactis* | $10^{-7}$ | $<$ 2kb | 12/12 | No | [30] |
| BacORep (OrthoRep) | *E. coli, B. subtilis, B. thuringiensis*, yeast | $10^{-8}$ to $10^{-5}$ | $<$ 15kb | 12/12 | No | [40] |
| UV or DNA damaging agent | Non-specific | $10^{-7}$ to $10^{-5}$ | Whole genome | 12/12 | No | [42] |

Table 1.1: Comparison of *E. coli in vivo* mutagenesis methods.

tational windows (up to 15kb) and can access all 12 types of mutations [40]. BacORep is compatible with *E. coli, B. subtilis* and *B. thuringiensis*, and the original OrthoRep system is compatible with yeast [37, 41].

Factors to consider when selecting a mutagenesis method for directed evolution include host compatibility, mutation rate, the mutation window (size and definition) and mutational spectrum. These factors are summarised in Table 1.1 for the *E. coli* variants of each technology.

## Selection methods

The second component of a directed evolution experiment is selection. Early directed evolution studies relied on laboriously screening variants in multi-well plates [4]. Typically, each variant was cultured in a separate well and evaluated using a colorimetric or fluorometric assay. Although laboratory automation has substantially increased throughput [43], plate-based screening remains orders of magnitude less scalable than pooled selection strategies.

Today, the most widely used method for selecting cells from a pooled population is Fluorescence Activated Cell Sorting (FACS) [9]. In FACS, cells are engineered to fluoresce in response to a property of interest. Sorting is performed by creating a stream of charged droplets that each contains approximately one cell. The stream is intercepted by a laser that interrogates whether fluorescence is above a defined threshold, and the charged droplet is diverted into the appropriate container depending on the result. This method is capable of sorting many thousands of cells per second [44]. To contrast this to a screen in 96-well plates, one study calculated that the number of variants that can be screened in a week using FACS would take 80 years (or roughly 20 PhDs) in 96-well plates [45].

Where FACS measures intracellular fluorescent outputs, MACS (Magnetic-Activated Cell Sorting) enables interrogation of *external* markers [46]. In MACS, magnetic beads coated with target-specific binders attach to cells displaying the marker of interest. When passed through a magnetised column, labelled cells are retained while unbound cells are washed away. Related surface-based approaches include display technologies, in which protein variants are presented externally - often on viral or cellular surfaces - allowing direct selection for binding affinity, catalytic activity, or stability [47–50]. Display techniques have proved extremely successful particularly due to their massive throughput (library sizes of $10^8$ to $10^{10}$) [51], and are responsible for the development of the aforementioned Humira [52].

The ability to perform selection on secreted phenotypes has been enabled by droplet-based microfluidics [53]. In droplet-based microfluidics, cells are encapsulated in small droplets, which can act as micro-reactors. Additional compounds can then be introduced to the droplets, for instance to investigate the performance

| Method | Throughput | Optimal for | Ref |
|---|---|---|---|
| 96WP screening | 96 | Small libraries, complex readouts | [43] |
| FACS | $10^9$ | Static intracellular phenotypes | [9] |
| MACS | $10^9$ | External markers in natural cell context | [46] |
| Display techniques | $10^{10}$ | Ultra-high throughput affinity, binding, or stability analysis | [47] |
| Droplet-based microfluidics | $10^6$ | Secreted phenotypes | [53] |
| Growth-coupled | $10^9$ | Phenotypes that can be robustly linked to growth or viability | [54] |
| **MiPS (this thesis)** | $10^6$ | **Dynamic intracellular phenotypes** | [56] |

Table 1.2: Comparison of common selection methods.

of a secreted enzyme [45].

A widely used selection strategy that does not require specialised equipment is growth-coupled directed evolution. In this approach, the desired phenotype is linked to cell survival or proliferation, such that high-performing variants are automatically enriched within the population [54]. Once established, the system is relatively low-maintenance and integrates effectively with *in vivo* mutagenesis platforms for an entirely continuous process [55]. However, implementing growth coupling can present practical challenges, as engineering a robust link between a designed phenotype and cellular fitness is not always straightforward. Furthermore, these systems are vulnerable to the emergence of "cheater" populations - variants that bypass the intended selection mechanism - which can rapidly dominate due to the exponential nature of growth [10].

Similarly to FACS, this technology presented in this thesis (MiPS) performs selection based on intracellular fluorescence readouts. However, using microfluidic technologies, it overcomes the key limitation of FACS which is the static nature of readouts. Therefore it joins this suite of selection methods as a new tool that is optimal for evolution of dynamic, intracellular phenotypes (Table 1.2).

# Microfluidics and mother machines

## Development of microfluidics

Microfluidic chips consist of microscopic channels through which small amounts of liquid can flow (as little as $10^{-18}$ litres) [57]. They were originally developed for analytical applications, offering benefits including low reagent cost, short analysis time and high sensitivity. In addition, fluid behaviour at such small scales is far more predictable due to limited convection, enabling precise control over chemical environments [58].

These advantages were clear, but the field was originally limited by an intensive manufacturing procedure, which involved etching microscopic features onto glass or silicon for each device. In 2000, the process of soft lithography was implemented to overcome this limitation [59]. With soft lithography, instead of etching a new device each time, a single silicon "master" is patterned via photolithography as a positive relief of the final devices, which can be produced repeatedly by casting PDMS over the master. Not only is this process more straightforward, but producing the final devices in PDMS leverages several advantageous properties such as being inexpensive, transparent, non-toxic, high-resolution and suitable for plasma bonding to glass slides (which is required to seal channels).

As the field progressed, more sophisticated features were developed to expand the functionality of microfluidic chips. The integration of on-chip valves [60], mixers [61] and pumps [62] transformed devices from passive flow channels into actively controllable microsystems.

## Microfluidics in live cell culture

Microfluidic chips were soon recognised as an ideal platform for studying biological systems. They enable precise spatial and temporal control over cellular microenvironments at a level of regulation that is difficult to achieve using conventional batch culture methods, and their dimensions can be at the scale of single cells. Such precise environmental control has enabled studies of chemotaxis in cancer-immune cell interactions by engineering controlled gradients [63], studies of stem cell differentiation [64] and organ-on-a-chip systems [65].

One of the earliest microfluidic devices developed for bacterial culture consisted of simple linear channels (or "trenches") into which cells were seeded, allowing their growth to be monitored as they expanded to fill the channel [66]. Since then, alternative structures have been designed to accommodate different experimental aims, including monolayers in which bacteria grow in a two-dimensional space, and mother machines [67]. Mother machines are similar to linear trenches but are closed at one end and connected at the other to a continuously flowing media channel. In this structure, the cell at the closed end (termed the "mother cell") remains effectively static for the duration of the experiment, while newly-produced daughter cells are displaced toward the open end and washed away by media flow [12]. For this reason, mother machines are ideal for studying single-cell behaviour over long time scales [68].

Since their introduction, mother machine platforms have undergone substantial architectural refinement. Device throughput has been dramatically increased, with serpentine channel layouts enabling arrays containing up to $10^6$ individual growth

trenches [13]. Design innovations such as secondary "backport" channels - sub-cellular width channels at the end of each trench, which allow flow of media *through* the trench - have further improved media exchange within each trench while and microfluidic loading and unloading [69, 70]. Adaptations in geometry and fabrication have also broadened the biological scope of mother machine systems, enabling experimentation across a range of microbial [70–78] and mammalian cells [79, 80].

Several orthogonal mechanisms enable interaction with cells in the mother machine. Through microfluidic control, compounds can be delivered with high precision and minimal mechanical disturbance, facilitating dynamic environmental perturbations. In addition, a recent work published by our lab demonstrated the integration of magnetic control of cells within the mother machine framework [11]. Alternatively, spatially resolved optical stimulation can be achieved using Digital Micromirror Devices (DMDs), which consist of arrays of individually addressable micromirrors capable of directing patterned light onto selected cells or trenches. This approach has enabled optogenetic regulation of gene expression with single-cell precision [16, 81]. Together, these modalities illustrate the versatility of the platform for manipulating and studying single-cell behaviour. However, beyond stimulation, the project outlined in this thesis will leverage the DMD to deliver high-power UV light to kill cells and perform selection for directed evolution.

# Theory of directed evolution

## Theoretical fitness landscapes

Similarly to many optimisation problems, the process of evolution can be envisaged as navigation across high-dimensional "fitness landscapes". Fitness landscapes map each genotype combination to a measure of fitness [82]. Advances in high throughput sequencing have made it possible to map examples of small fitness landscapes with real data [83–90], however real-world landscapes are generally too large for exhaustive experimental exploration ($20^n$ combinations for a protein length $n$). As the largest empirical landscapes measured to date are only 4 amino acid positions [88–90], theory of directed evolution is generally studied using theoretical fitness landscape models [91].

Many fitness landscape models have been proposed and are appropriate for different contexts. Fisher's geometric model [92, 93], the holey landscape model [94] and the multilinear model [95] are all designed for entire genomes. These models all take as input some description of location on the landscape (e.g. coordinates corresponding to phenotype values or mutations from a reference point) and output a corresponding fitness value.

A central concept in models of single gene evolution is that of a rugged fitness landscape, in which epistatic interactions between loci generate multiple adaptive peaks rather than a single smooth optimum. This idea can be traced back to Wright's shifting balance theory, which emphasised that evolution proceeds on a complex landscape with multiple adaptive peaks [82]. In this view, popula-

tions may become trapped on local fitness peaks, with genetic drift and population division playing a key role in allowing movement across fitness valleys and enabling exploration of alternative adaptive peaks. The NK model [96, 97] can be seen as a formal realisation of this framework, providing a tunable representation of ruggedness through explicit epistatic interactions that generate multiple local optima and thus a mechanistic setting in which shifting balance dynamics may occur.

$$F(G) = \sum_{i=1}^{N} \tilde{f}_i(G) \tag{1.1}$$

$$\tilde{f}_i(G) = f(G_i, G_{k_{i1}}, G_{k_{i2}} ... G_{k_{iK}}) \tag{1.2}$$

In the NK model, a gene $G$ is represented by an array of $N$ sites (each $0$ or $1$). Every site interacts with $K$ other sites in the gene, influencing the resulting fitness. The number of interactions determined by $K$ is what allows ruggedness to be tuned. Mathematically, the fitness $F$ of a gene is the sum of fitnesses of each locus (1.1). The fitness of each locus is dependent on the state of that locus and the state of the $K$ other loci it interacts with (1.2). When $K = 0$, all sites are independent and the model is additive with a single peak. The other extreme, $K = N - 1$, is maximally rugged (and is akin to what's known as the House of Cards model [98]).

$$F(G) = -cD(G, G^*) + \eta(G) \tag{1.3}$$

An alternative representation of landscape ruggedness is Rough Mount Fuji model

[99]. In the Rough Mount Fuji model, fitness of a gene $G$ is inversely proportional to the distance $D$ from the maximum point $G^*$ of the additive landscape. Ruggedness is introduced by addition of a random variable $\eta(G)$ (1.3). The degree of ruggedness in the Rough Mount Fuji model is controlled by changing the ratio of importance of the distance and random variable parts. When the distance-related part is dominant (i.e. large $c$), the model is closer to being additive with a single peak. When the random variable part is dominant (i.e. small $c$), the resulting landscape is again akin to the House of Cards model [98].

Given the limited bank of empirical fitness landscapes, there are not many instances of comparison of these models to real data. In one case, when compared to an empirical dataset, the Rough Mount Fuji model was found to be a better fit than the NK model [87]. The dataset used, however, did not exhibit epistasis (the property of interactions between sites), and thus it is not necessarily a surprising result. Epistasis is frequently identified as an important factor in evolution [88, 100–102] and thus it is surely important to select a model that can account for this. In a similar analysis evaluating suitability of Fisher's geometric model [103], which also does not account for epistasis, it was found that the model could not capture the important aspects of landscape structure that arise from interactions. These interactions present much more interesting challenges for directed evolution, and therefore the NK model was selected as the most appropriate model for this context.

Since the original implementation of the NK model, two alternative versions have been independently proposed: the NKp [104] and NKq [105], both of which add to the model the concept of selective neutrality. They are inspired by the neutral

theory of evolution that was described by Motoo Kimura in 1983. Neutral theory states that evolution is not primarily a "hill-climbing" process driven by positive or negative selection, but rather the majority of mutations are selectively neutral, and populations "drift" until chancing upon a beneficial mutation. Over 50 years after the introduction of neutral theory, it is still firmly under debate (note the article published to strongly discredit it in 2018 [106], and the extensive rebuttal published swiftly afterwards [107]).

The NKp model implements neutral evolution by setting a proportion of genotype combinations to a fitness value of $0$. The NKq model implements neutral evolution by sampling fitness values from a discrete distribution, such that many genotypes will possess the same fitness value. In directed evolution, selectively neutral mutations are found to be common and capable of opening up adaptive paths (for instance by increasing protein stability and tolerance of future mutations) [108]. Neutral mutations are also more common in directed evolution than in nature, given that the laboratory-defined definition of fitness is narrower. The NKp and NKq models are therefore likely to be a sensible representation of laboratory-based evolution.

Another important consideration in modelling directed evolution is the dynamic nature of protein fitness landscapes. Fluctuations in the cellular and external environment, as well as inconsistencies in our ability to measure fitness, will alter the momentary shape of a protein fitness landscape. An interesting study released recently also modelled the effects of protein mistranslation on fitness landscapes [109]. This is not an rare occurrence, as 10% of the proteins translated in *E. coli* are estimated to contain an incorrect amino acid. The dynamic nature of protein

fitness landscapes that arises from mistranslation (and likely from other sources) is an effective flattening of the fitness landscape that increases the prominence of genetic drift.

## Computational strategies in directed evolution

Standard practise in directed evolution is simply to take forward and mutate the top fraction of variants with each iteration [110]. This greedy approach is prone to getting trapped in local optima, particularly in rugged landscapes [111]. With the aid of computational methods, it is possible to navigate protein fitness landscapes in a more active way.

In 1992, John Holland introduced the concept of genetic algorithms [112]. Genetic algorithms take inspiration from the process of evolution to solve optimisation problems. Simply put, genetic algorithms take a population of variants, select the top performers, mate them by combining their "genes" and repeat iteratively until they have sufficiently improved. This process is more commonly used to perform optimisation on problems outside of biology (although application of genetic algorithms to directed evolution was explored computationally in [113] and the closest real-life example would be the DNA shuffling method of mutagenesis [114]). The concept of a genetic algorithm did, however, inspire people to think about better ways to navigate protein fitness landscapes.

A key development in directed evolution has been the incorporation of statistical models to guide sequence design. Early approaches such as ProSAR moved beyond simple selection of high-performing variants by analysing correlations between residue identity and measured fitness across a library [115]. This shift

marked the beginning of data-driven protein engineering, in which mutational effects are inferred from sequence-function relationships rather than explored purely through iterative screening. Recent advances in machine learning have further expanded these capabilities, enabling increasingly accurate prediction of mutational fitness effects across large sequence spaces [116–119]. More recently, statistical frameworks have been developed that incorporate evolutionary trajectories, using time-resolved data from directed evolution experiments to infer properties of the underlying fitness landscape and epistatic interactions [120]. Together, these approaches reflect a broader transition from purely empirical optimisation toward model-informed exploration of protein fitness landscapes.

Directed evolution can be guided using data generated during experiments, but evolutionary information is also embedded in naturally occurring protein sequences. Comparative analysis across homologous proteins enables inference of amino acid co-variation, reflecting structural or functional interactions, and evolutionary conservation, indicating functional importance [121, 122]. Such sequence-based models have been used to approximate features of underlying protein fitness landscapes [123], offering a framework for informing and tuning directed evolution strategies.

Together, these advances illustrate a growing capacity to model and navigate protein fitness landscapes in directed evolution. However, many computationally guided strategies depend on extensive sequencing across discrete generations, limiting their compatibility with continuous directed evolution systems. The particular continuous directed evolution system developed in this thesis implements selection in the mother machine, which unlocks the ability to select single cells

based on dynamic phenotypes. In order to optimally implement that platform, therefore, expanding the small body of theory exploring optimisation of directed evolution without sequencing is required.

# 2 | Theory: Tools for sequence-free optimisation of directed evolution

## 2.1 Featured publications

This chapter combines work from two co-first author publications:

- Optimisation strategies for directed evolution without sequencing [124]
- Sequence-free landscape inference for directed evolution [125]

## 2.2 Abstract

Directed evolution can enable engineering of biological systems with minimal knowledge of their underlying sequence-to-function relationships. A typical directed evolution process consists of iterative rounds of mutagenesis and selection that are designed to steer changes in a biological system (e.g. a protein) towards some functional goal. Much work has been done, particularly leveraging advancements in machine learning, to optimise the process of directed evolution. Many of these methods, however, require DNA sequencing and synthesis, making them resource-intensive and incompatible with continuous directed evolution pipelines. Given that the platform under development in this thesis lends itself well to continuous directed evolution, there was a strong motivation to fill the gap in optimisation of directed evolution *without* sequencing. Therefore, this chapter describes three tools for optimisation of directed evolution that do not require sequencing,

and could be employed in the novel platform: Selection functions, population splitting and multi-dimensional selection. To optimally employ these tools, however, one then requires information about the ruggedness of the underlying fitness landscape, which is generally unavailable at the outset of an experiment, and cannot currently be computed using analytical methods. Therefore this chapter secondly describes a method for estimation of landscape ruggedness from a mutating population using only population-level phenotypic data and an estimation of mutation rate. Together, the tools proposed, informed by the landscape inference method, form a framework for selection of high-performing parameters for directed evolution control. In the third and final part of the chapter, the pipeline is used to demonstrate consistent improvement over the standard approach using theoretical $NK$ landscapes and four empirical protein fitness landscapes *in silico*.

## 2.3 Background

As a biological entity undergoes directed evolution, the process can be imagined as navigation across high-dimensional "fitness landscapes" [82]. Fitness landscapes map each genetic sequence to a measure of fitness (with "fitness" being performance for a desired function), and the goal of directed evolution is to find the highest peaks on that landscape. Fitness landscapes are known to exhibit variable degrees of ruggedness, which can create local optima that constrain paths of evolution [88, 126, 127]. Standard practice in directed evolution is to take forward and mutate only the top fraction of variants during each iteration [29, 110, 128–130]. This "greedy" approach is prone to getting trapped in local optima, particularly on rugged landscapes [111]. With the aid of computational methods,

however, it is possible to navigate protein fitness landscapes in a more active way.

One of the earliest examples of this approach is ProSAR, which uses a statistical algorithm to identify residues correlated with high fitness and designs each new generation of variants by combining those predicted to contribute most strongly to fitness [115]. Methods that predict the fitness effects of mutations in this way are now able to accommodate machine learning [116–119] and Bayesian optimisation approaches [131, 132]. Such methods have been built upon by not only utilising the fitnesses of the sequences in isolation, but also time-series mutation data acquired *during* a directed evolution experiment [120]. The vast majority of these methods, however, are based upon the requirement for sequencing information from each generation. This means they are somewhat resource- and labour-intensive, and are not suited to maximising the benefits from continuous directed evolution (where a population of bacteria evolves continuously as opposed to iterative production of variants *in vitro*).

In continuous directed evolution where sequencing information is not employed, the field has established the default approach of selecting only the top variants with each generation [29, 110, 128–130]. This approach may restrict population diversity, leading to a higher propensity to get trapped in local optima. Examples of previous work to address this include alternating between "on" and "off" states of selection [133], as well as modifying selection stringency [134] to increase population diversity. In Section 2.4 the challenge is approached from several new angles. First, probability of selection is applied as a parameterised function of fitness that can be used to tune the balance exploration and exploitation on a fitness landscape. Second, the benefits that can be gained by splitting a population into

sub-populations and allowing their trajectories to diverge is explored. Finally, optimisation of multiple properties simultaneously is explored in theory, something which is not possible with FACS-based selection, but is possible using the platform described in this thesis.

In order to optimally implement tools such as selection functions and population splitting optimally, one requires some information about the underlying ruggedness of the fitness landscape (smooth landscapes benefit from greedy approaches, whereas rugged landscapes are better navigated with exploratory approaches). Various metrics have been proposed to define and quantify the ruggedness of fitness landscapes [135–139]. Although these metrics have been shown to correlate across several empirical landscapes [139], most of them require full genotypic information to compute a ruggedness measure, which is highly resource-intensive and defeats the purpose of directed evolution if the full landscape is already known.

Section 2.5 describes development of the novel method of Sequence-Free Landscape Inference for Directed Evolution (SLIDE) to efficiently estimate the ruggedness of a landscape using phenotypic information and an estimation of mutation rate. Using SLIDE, it is possible to show that the underlying local ruggedness of a fitness landscape can be estimated from the average fitness decay of a population that is randomly mutated from a starting point. On smooth landscapes, the average fitness decays more slowly than on rugged landscapes. This phenomenon can be attributed to fitness correlations between variants, as random mutations are less likely to result in large fitness differences on smooth landscapes. Most importantly, the fitness decay curve can be acquired without large-scale sequencing

data; instead it uses population-level fitness readouts and a mutation rate estimation.

Therefore this chapter concludes by demonstrating a full pipeline for optimisation of directed evolution without sequencing, which consists of initial landscape ruggedness inference with SLIDE, followed by subsequent selection of optimal selection function and population splitting parameters. The method is evaluated *in silico* on theoretical landscapes obtained from the $NK$ model [97] and on four empirical landscapes [84, 88–90]. It is possible to show that the method consistently finds the optimal parameters on $NK$ landscapes, and outperforms or matches the baseline strategy on all empirical landscapes.

## 2.4 DE optimisation tools

### 2.4.1 Selection functions

In order to test selection strategies, a computational model was implemented to mimic the process of directed evolution. Genes in the model are represented by one-dimensional arrays, which iterate through rounds of mutation and selection. In the selection process, the fitness of each gene is calculated using an empirical landscape [84, 88–90] or an NK model [96, 97]. Empirical landscapes are combinatorially complete fitness measurements for all variants of a protein (or protein region). Due to the experimental difficulty of measuring whole empirical landscapes, examples greater than three amino acids are scarce. NK models are computationally generated fitness landscapes that can be used as a tuneable substitute for empirical landscapes. In both NK and empirical landscape implementations,

the gene sequence is taken as input and the corresponding fitness value is given as output.

The first tool for optimisation of DE introduced in this chapter is selection functions. Selection functions are introduced as a means to tune the balance of exploration and exploitation on a fitness landscape. The selection functions proposed here are defined by two parameters: "fitness threshold" and "base chance" (Fig 2.1A). The fitness threshold is the fitness percentile above which variants have a $100\%$ chance of selection, otherwise the chance of selection is equal to the base chance. In this work, selection functions are normalised to select a constant fraction of variants. Fixing the selected proportion of cells reduces the parameter space of selection functions to one dimension, as every base chance has only one fitness threshold value corresponding to a fixed proportion of the population being selected. In a continuous directed evolution experiment, this also ensures that proliferation time remains approximately constant between generations (hence, performance metrics can be considered as improvement in a trait *per unit time*). It was hypothesised that the base chance parameter will improve directed evolution by allowing a population to escape local optima on the fitness landscape. By selecting some cells unconditionally, they are allowed to accumulate more mutations, potentially allowing them reach higher performing variants via deleterious phenotypes.

As mentioned previously, most experiments performed in this section (unless stated otherwise) are using the NK model in the place of empirical fitness landscapes. The NK model describes a class of fitness landscape with tuneable ruggedness [96, 97]. N describes the number of variable sites, and K can be thought of

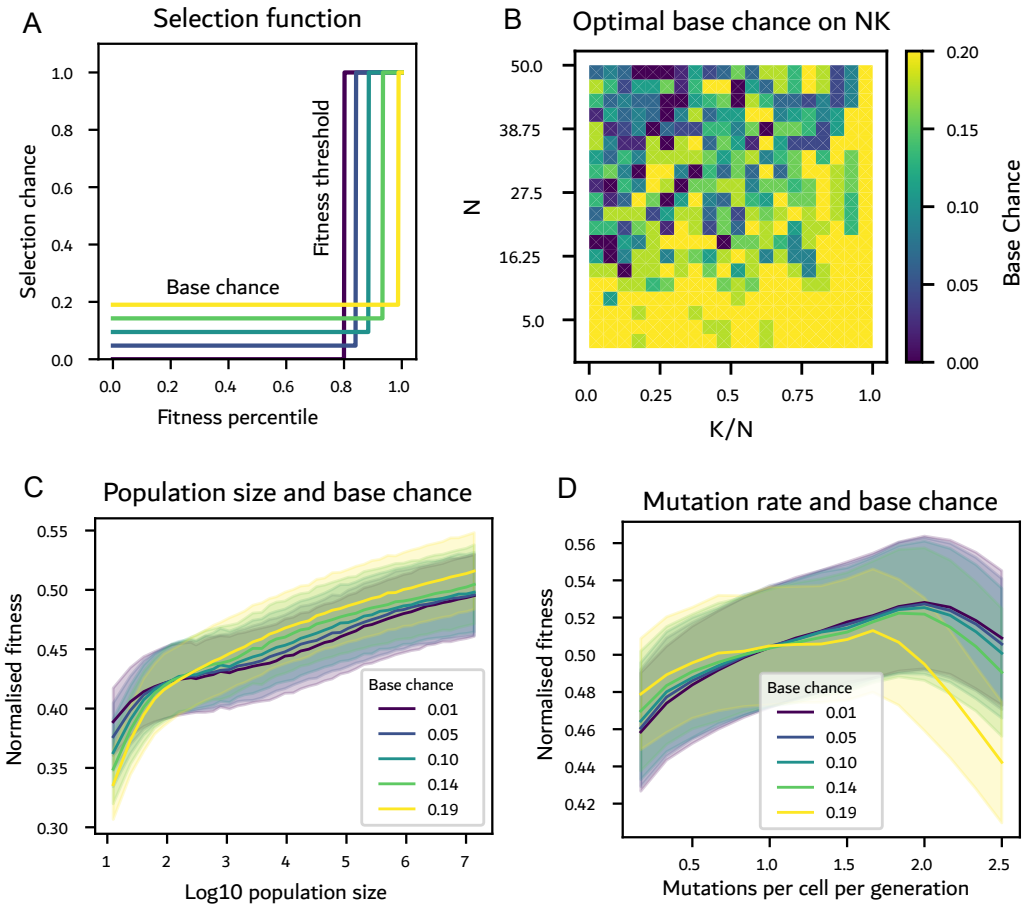



Figure 2.1: **Investigating the performance of selection functions in directed evolution A:** Selection functions define probability of selection as a function of fitness. The selection function used here is determined by two parameters: fitness threshold and base chance. Selection functions were normalised to select 20% of the population. **B:** Optimal base chance values on varying NK landscapes. Dependence of trajectory end-point fitness on **C:** population size, and **D:** mutation rate. "Normalised fitness" is maximum fitness across the population at the final time point (averaged over 100 runs) as a fraction of the global maximum on the landscape. Shaded areas represent standard deviation over repeat runs. Experiments ran for 300 generations. $0.01 \leq$ base chance $\leq 0.19$, N = 25, K = 5, mutations per cell = 0.1, population size = 1000.

as a metric of ruggedness ranging from 0 to $N - 1$, with high K meaning high ruggedness (i.e. more local optima). $K/N$ is used to compare ruggedness across landscapes of different $N$. Fig 2.1B shows the optimal base chance over varying NK landscapes, as estimated via simulation. In particular, Fig 2.1B shows the optimal base chance increasing with respect to $K/N$ (ruggedness), and decreasing with respect to $N$ (dimensionality). Given that base chance is hypothesised to help escape local optima, one explanation for this result is that more local optima are found in rugged landscapes, and they are more difficult to escape in low-dimensional landscapes, which offer fewer paths between any two points. For smooth, high-dimensional landscapes the opposite is true, therefore the function that most favours exploitation over exploration (i.e. base chance = 0) is found to be optimal.

In order to determine a landscape on which to perform the rest of the evaluations, a simple comparison of 0 base chance and 0.1 base chance was implemented on a range of landscapes (Fig. 2.2). By plotting the proportion of "wins" by each strategy when compared directly, it was possible to identify the landscape N = 25, K = 5 as a sensible choice for further exploration, given neither strategy dominated under our "standard" mutation rate and population size conditions. N = 50, K = 10 was also a balanced choice, but was eliminated on the basis that computational memory requirements scale quadratically with respect to N.

Next, the interaction between base chance and population size was explored. Fig 2.1C shows that base chance can improve performance across a range of population sizes, particularly in large populations. Given that large populations can explore more of the landscape, they are likely to encounter more local optima.

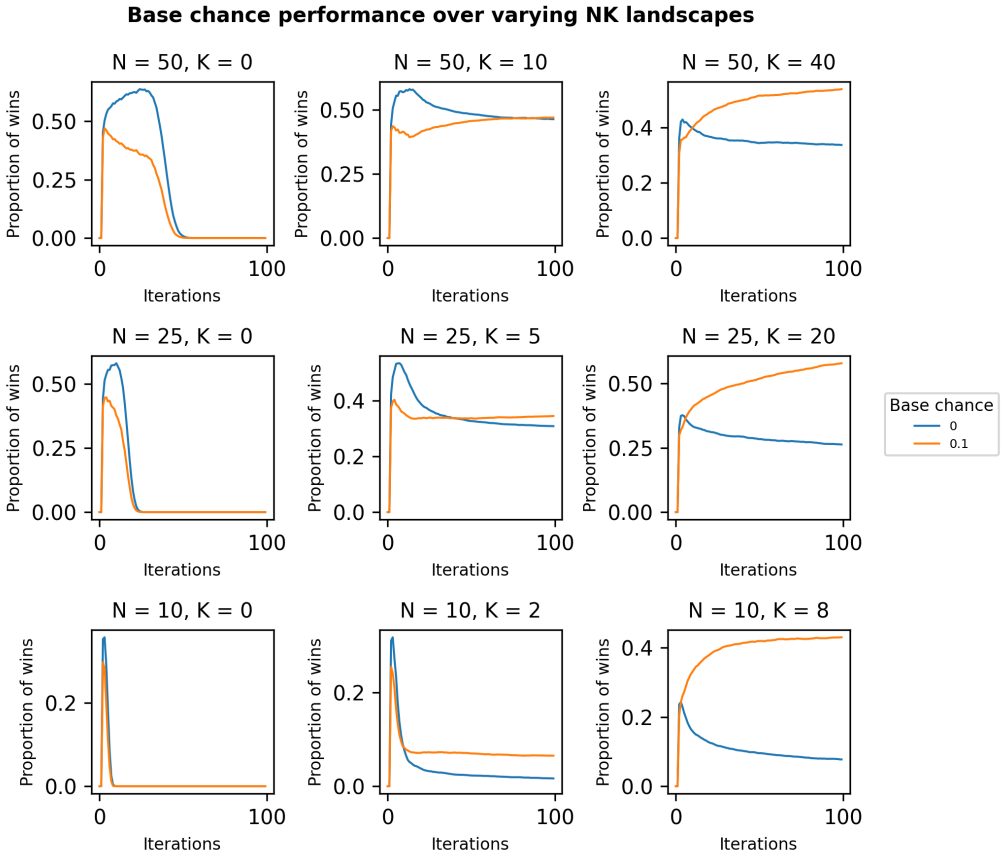



Figure 2.2: **Base chance performance over varying NK landscapes**. Competition plots showing base chance performance over varying NK landscapes. Data shows the proportion of runs that are won by either a 0 base chance, or 0.1 base chance strategy over different landscapes. Where the outcome is tied, the value is 0. Mutations per cell = 0.1, population size = 1000.

The ability for a population to escape a local optimum is dependent on its ability to reach a fitter state via at least one deleterious mutation. Evolution via deleterious mutations is shown to occur more readily in large populations [140, 141] and this dynamic may be promoted by base chance, hence the improvement in performance. In small populations, the cost of including detrimental variants is greater relative to the potential gain, therefore base chance is less beneficial.

The performance of the selection functions with varying mutation rates was also investigated (Fig 2.1D). When mutation rate is low, higher base chance values perform best and vice versa for high mutation rate. One explanation for this is in the balance of exploration and exploitation. Both base chance and mutation rate aid in escaping local optima by increasing the likelihood of a cell undergoing multiple mutations. Although this benefits exploration, *further* increases in mutation rate can come at the cost of not effectively exploiting a position on the landscape. For this reason, base chance performance drops off more quickly in a high mutation rate regime. Given that most directed evolution experiments operate in a low mutation rate regime to avoid detrimental side effects [21], base chance could act as a useful tool for promoting landscape exploration. The benefits of such an approach are that implementing a base chance has no direct impact on top-performing variants, whereas increasing mutation rate impacts all variants.

The impact of measurement noise on the performance of a high base chance strategy was also explored (Fig. 2.3). It was found that as noise increases, the good performance of a higher base chance strategy is relatively dampened. This likely occurs as, similarly to the compounding effects of base chance and mutation rate, noise can have similar effects as base chance in increasing the likelihood of selec-

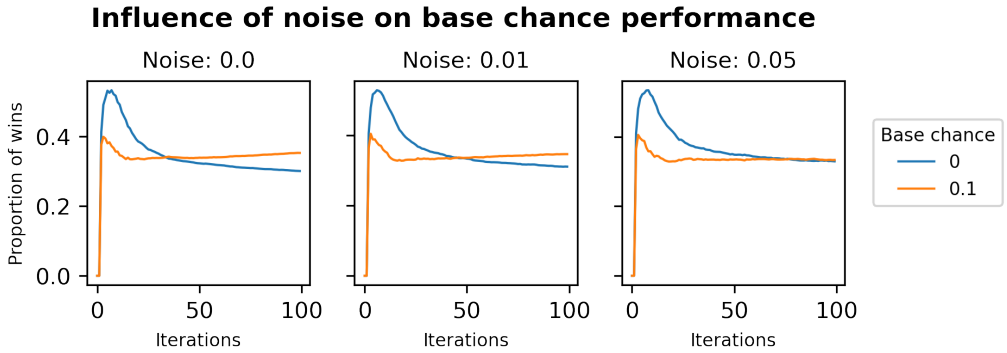



Figure 2.3: **Noise effects**. Competition plots showing the influence of measurement noise on base chance performance over 100 iterations. To introduce noise, values sampled from $\mathcal{N}(0, noise^2)$ are added to fitness values each generation. N = 25, K = 5, mutations per cell = 0.1, population size = 1000.

tion of lower-fitness variants.

### 2.4.2 Population splitting

This work has thus far assumed a single population undergoing directed evolution. However, in practice, one could run multiple, smaller, copies of the same directed evolution experiment by subdividing a population, and take the best outcome across all of them as the final result. Here, this method is referred to as population splitting. An example of such a situation is displayed in Fig 2.4A, where a population of size 500 is split into five equal sub-populations of size 100, leading to fitness improvements. This may be because in a single, mixed population, mutations will eventually drift to fixation or extinction, therefore the population as a whole remains largely on the same trajectory. If one splits the population, sub-populations are able to drift on separate trajectories without cross-competition, effectively mimicking the process of speciation and increasing landscape exploration [142, 143].

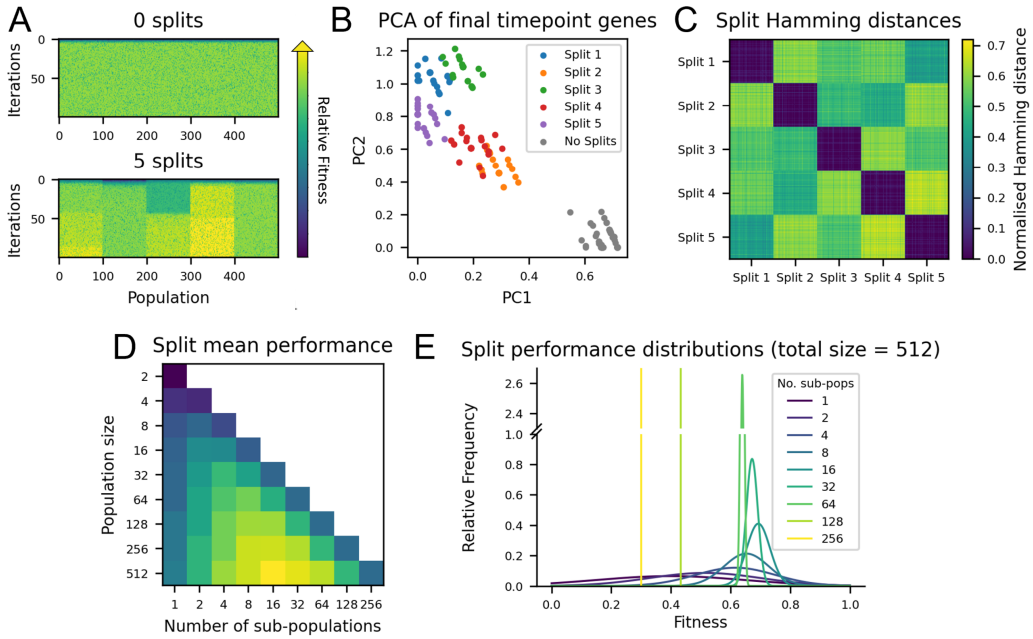



Figure 2.4: **Investigating the effects of population splitting in directed evolution A:** Example of a directed evolution run split into five sub-populations vs a single large population. Colourbar indicates fitness value. **B:** Principal components analysis of the final time point sequences of a split vs non-split population. **C:** Hamming distances between the final timepoint sequences of the split population. **D:** Mean performance of directed evolution with varying total population and sub-population size. **E:** Mean and standard deviation of performance with varying number of sub-populations (fit to a normal distribution). Experiments ran for 100 generations. Mutations per cell = 0.1, base chance = 0, N = 25, K = 5.

Principal components analysis (PCA) was used to verify that sub-populations diverge on separate trajectories. The final timepoint gene sequences from the simulation in Fig 2.4A were collected. PCA was performed on the combined dataset to reduce the 25-dimensional genetic space to just 2 dimensions for visualisation (Fig 2.4B). The result shows that each sub-population forms a cluster, and the overall variation of the split population is significantly more than the non-split population. This is further verified in Fig 2.4C, which displays that the average normalised Hamming distance between sub-populations (0.44) was far greater than that within sub-populations (0.011, similar to the average Hamming distance of 0.009 measured within the large single population).

Population splitting can clearly confer a benefit to performance, however there is a trade-off between splitting the population to maximise exploration, and keeping sub-populations large enough to effectively search around their local position on the landscape. Fig 2.4D summarises the performance of population splitting for different total population sizes, demonstrating that if the (total) population is too small, splitting is instead detrimental to performance. Fig 2.4E shows the distribution of performance outcomes, over 1000 runs, for splitting a large population into increasingly smaller populations. This demonstrates an additional advantage of population splitting, which is that the variance of the final outcome decreases as the number of sub-populations increases.

### 2.4.3 Multi-dimensional selection

The platform described in this thesis enables long-term observation of single cells prior to selection. Not only does this increase the reliability of readings and allow

selection based on complex time-dependent traits, it would also allow for multiple properties (or responses to stimuli) to be measured in a single round of selection. Such multi-dimensional selection is highly applicable in the directed evolution of biosensors, in which one seeks to optimise both specificity and sensitivity [110].

The current standard approach to multi-dimensional selection (e.g. in FACS) is to perform sequential rounds of selection, one for each property [110, 144]. The sub-optimality of this approach is previously highlighted in [145]. This introduces a systematic error with respect to the selection objective, as shown in Fig 2.5A. The blue dashed line divides the true best cells from the population (as could be achieved by single-round selection), whereas the red dashed lines represent the cut-offs of a double-round selection setup. Cells that are poor in one property but excel in another are not selected by double-round selection. As a result, the overall performance of single-round selection is higher (Fig 2.5B).

Not only does this approach improve upon FACS in the simplest case, it also offers the additional ability to *tune* the prioritisation of different properties. When selecting on a single property, the fitness value ($F$) used to determine selection is simply the value of that property. When selecting on multiple properties, however, the overall fitness value used to determine selection is some combination. Given that most directed evolution experiments have a limited amount of time and/or physical resources, it is crucial to consider how much one prioritises each property in this combination. This prioritisation can be implemented by applying a weight ($w_i$) to the value of each property. So, $F = w_1 f_1 + w_2 f_2 + w_3 f_3$, where $f_i$ is the value of property $i$. In the simplest case, all weightings are allowed to be equal (Fig 2.5C). By changing the weightings of the properties one observes

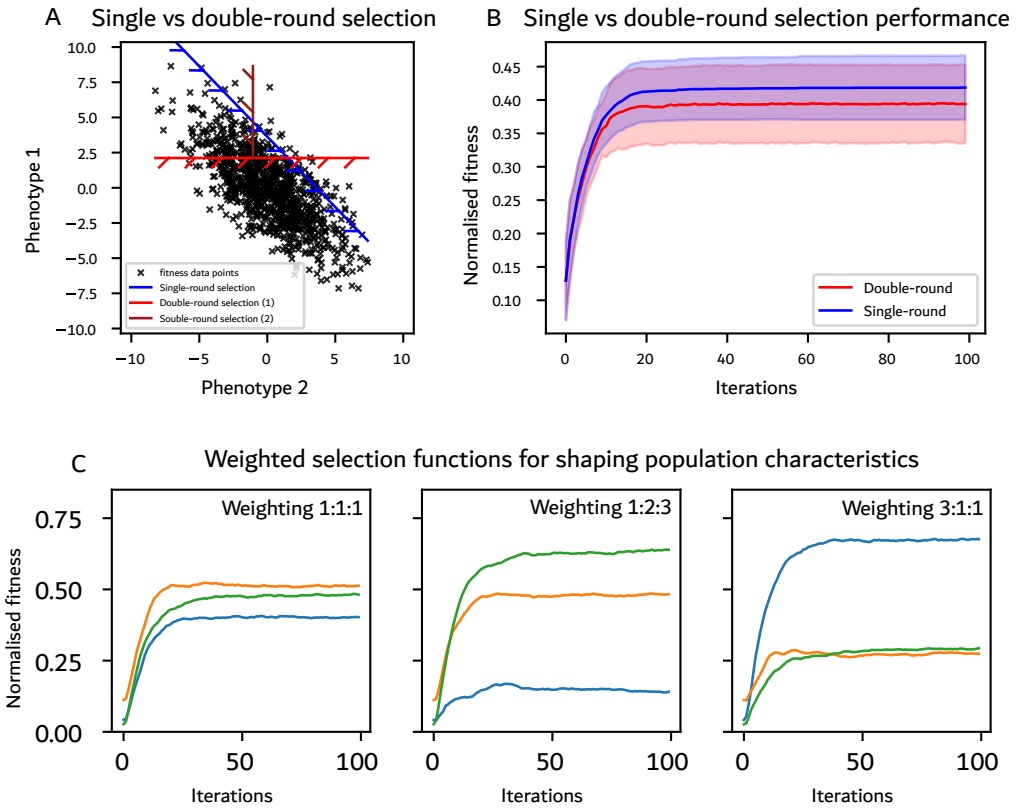



Figure 2.5: **Optimisation of multi-dimensional selection** A: Demonstration of selection patterns with double-round (i.e. flow cytometry selection) and single-round selection. Randomly-generated fitness points normally-distributed around the mean 0 of phenotypes 1 and 2. B: Directed evolution performance of double-round vs single-round selection. Shaded region is standard deviation over repeated runs. C: Performance of weighted selection functions to perform directed evolution over three properties, with weightings of 1:1:1, 1:2:3 and 3:1:1. Experiments ran for 100 generations. N = 25, K = 5, mutations per cell = 0.1, population size = 1000, fitness threshold top 5%, as per [110].

proportional gains in the fitness of each property.

## 2.5 Sequence-free fitness landscape inference with SLIDE

As shown in Fig. 2.1B, the optimal strategy to employ is influenced by the ruggedness of the underlying landscape, therefore in order to implement the above strategies optimally, one must find some way to estimate landscape ruggedness.

Ruggedness of a fitness landscape can be defined in terms of various mathematically distinct metrics (see [135–139]), which aim to capture how fitness changes when the landscape is traversed. These metrics have limited applicability to experimental processes due to requiring complete mapping of the landscape as a whole, which is extremely resource-intensive and undermines the benefits of DE (if one already knows the entire landscape). There is strong motivation, therefore, to develop a method of ruggedness estimation that does not require extensive sequencing, and could add value to DE pipelines.

The process described in this Chapter is SLIDE (Sequence-free Landscape Inference for Directed Evolution). The parameter measured by SLIDE is $\rho \in [0, 1]$, where $\rho = 0$ would be a linear landscape, and $\rho = 1$ would be a maximally-rugged landscape. The ruggedness parameter $\rho$ can be understood through Fourier analysis. One can imagine breaking a fitness landscape down into sub-landscapes of different "frequencies". For a large $\rho$ representing a rugged landscape, higher-frequency terms dominate the landscape, and for a small $\rho$ representing a smooth landscape, lower-frequency terms dominate. $\rho$ itself can be considered a weighted

average of the Fourier terms that comprise this spectrum.

Experimentally, however, the $\rho$ value can conveniently be measured by observing the rate of fitness decay of a population of cells mutating away from a higher-than-average fitness starting point. The full derivation of this relationship was developed by my collaborator Sebastian Towers, and can be seen in the full publication [125]. The equation for measuring $\rho$ from the decay rate is:

$$F_\mu \approx Ce^{-\mu\rho} + c, \tag{2.1}$$

where $F_\mu$ is the population's average fitness after $\mu$ mutations, $c$ is the average landscape fitness, $C$ is a constant that depends on the initial population. For rugged landscapes, a high $\rho$ will be measured, and for smooth landscapes a lower $\rho$ will be measured. Intuitively, this is due to the fact that smooth landscapes should have high fitness variants clustered together, leading to high fitness correlations rugged landscapes should exhibit large fitness differences between adjacent nodes. On a rugged landscape with less fitness correlation between neighbours, average fitness of this population will more quickly converge to the overall fitness landscape mean, whereas on a smooth landscape the rate of convergence to the overall fitness landscape mean will be much slower.

On $NK$ landscapes, the epistatic interaction parameter $K$ and $\rho$ can be shown to be related by $\rho \approx (K+1)/N$ [125]. This makes the NK model a highly useful context in which to test SLIDE, as there is a clearly-defined ground truth for evaluating the output. The trend from equation (2.1) is illustrated in Fig 2.6A, which shows the mean population fitness decay from mutating popula-

tions on two $NK$ landscapes of differing ruggedness, and the approximation from equation (2.1). The figure shows that on landscapes with greater ruggedness ($(K+1)/N = 0.75$), the rate of fitness decay is higher than that of the smooth

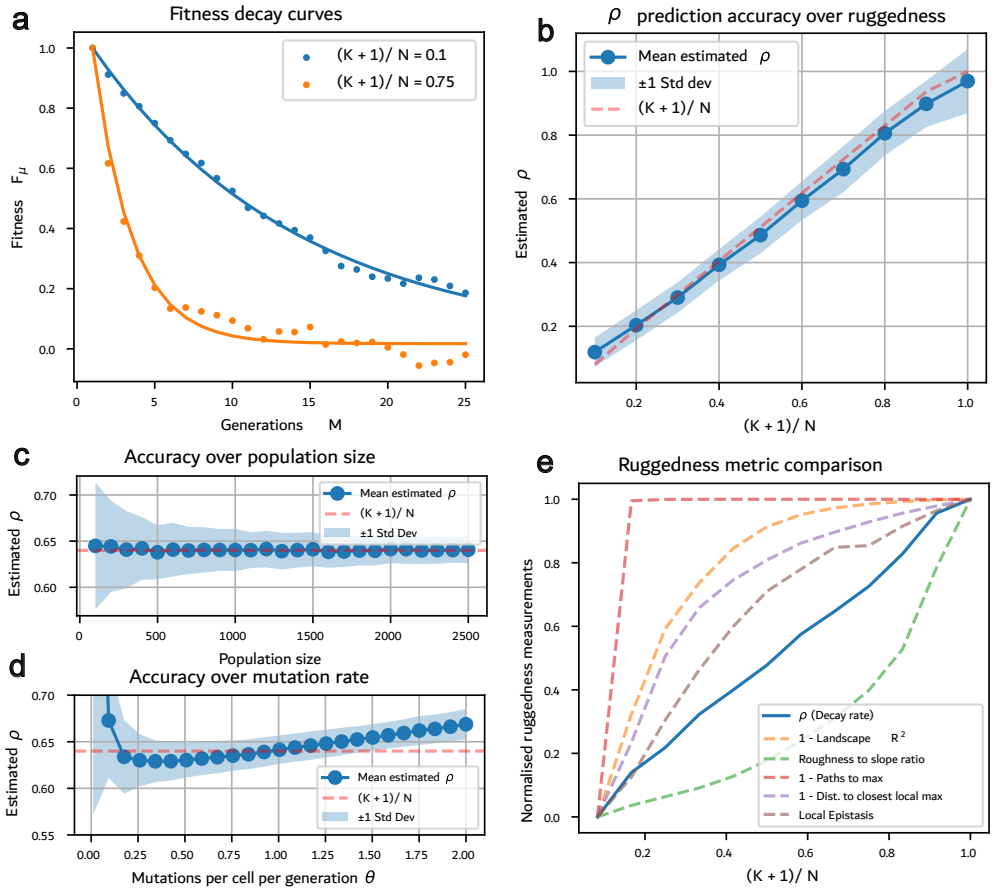



Figure 2.6: **Fitness decay rate of a randomly mutating population, as a metric for landscape ruggedness**. **A:** Simulations of two populations on two $NK$ landscapes ($(K+1)/N$ = 0.1 and 0.75, respectively, with N = 20) mutating away from a starting point (population size = 300, mutation rate $\theta$ = 0.5 mutations per generation). Exponential fits to equation 1 are indicated by solid lines. **B:** Estimated $\rho$ values across landscapes of increasing ruggedness using 25 generations and a mutation rate of 0.5 per generation. The figure shows mean and standard deviation for 300 estimates. **C–D:** Accuracy of $\rho$ over increasing population size (C) and mutation rate (D) for 500 estimates. **E:** Comparison of ruggedness metrics on 600 $NK$ landscapes of increasing ruggedness ($(K+1)/N$). All metrics are normalised between 0 (least rugged) and 1 (most rugged).

landscape ($(K+1)/N = 0.1$). This is further validated in Fig 2.6B, which shows that the estimated $\rho$ corrlates strongly with $(K+1)/N \in \{0.1, 0.2, \ldots, 1\}$.

Due to the noise associated with observing a randomly mutating population, the reliability of $\rho$ estimates also depends on the parameters of the observed population, including the population size and the mutation rate. Fig 2.6C shows $\rho$ estimates over increasing population sizes. For each population size, $\rho$ is estimated 500 times (Fig 2.6). For small population sizes, the ruggedness estimate exhibits higher variability due to noise in the mean population fitness. For population sizes exceeding 2,000, the standard deviation is less than $5\,\%$ of the "true" decay rate. The effect of mutation rate is analysed in Fig 2.6D, which shows mean and standard deviation of $\rho$ for an increasing mutation rate $\theta$. For small $\theta$, only a small fraction of the fitness decay curve in Fig 2.6A is captured, leading to a large variability of the estimate from equation (2.1). For very large $\theta$, the mean population fitness decays more rapidly to the landscape average, again leading to a large variability of the estimate from equation (2.1).

To further validate equation (2.1) and its characteristic decay rate $\rho$, Fig 2.6E provides a comparison between $\rho$ and other ruggedness measures across $NK$ landscapes. The other ruggedness measures are the landscape $R^2$ measure [139], the roughness to slope ratio [135], number of paths to the global maximum from the opposite position on the landscape [138], the distance to the nearest local maximum [137], and the local epistasis metric [139]. Figure 2.6E shows that although all metrics increase with $(K+1)/N$, they express either concave (roughness to slope ratio), linear (decay rate), or convex curves (all other metrics). Because all metrics increase with $K$, this allows one to compare the ordering (w.r.t. rugged-

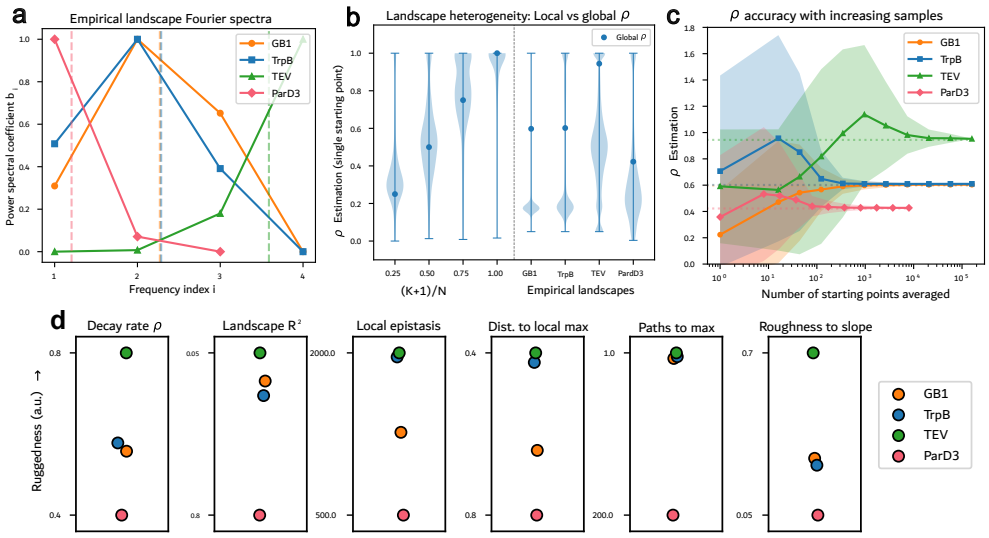



Figure 2.7: **Ruggedness estimation on empirical landscapes. A:** Fourier spectra of each complete landscape. Centre of mass of the spectrum ($\rho$) indicated by dashed line. **B:** Violin plots showing distribution of $\rho$ estimates with the fitness decay computed from individual starting points. The blue dot is the global $\rho$ value. **C:** Estimations of $\rho$ with increasing numbers of starting points. Starting points were sampled uniformly across the space, with mutation rate 0.1 per generation, population size = 2500. **D:** Comparison of ruggedness metrics applied to the empirical landscapes.

ness) of landscapes across metrics, but the differences in curve shapes prohibits comparison of distances (i.e. quantitative ruggedness differences) between landscapes across metrics. However, one can mathematically prove the decay rate $\rho$ directly correlates with $(K+1)/N$ [125], and the same relationship holds regardless of $N$. Overall this demonstrates a second benefit of the methodology - that it provides an interpretable measure of ruggedness that has immediate meaning in the widely-studied $NK$ landscapes.

The performance of SLIDE on real-world empirical landscapes was assessed by testing it on four experimentally-measured landscapes: GB1 (4 sites) [88], ParD3 (3 sites) [84], TrpB (4 sites) [89] and TEV (4 sites) [90]. First, a "ground truth"

$\rho$ value was determined by performing a Fourier transform of the landscape and taking a weighted average of the components (as indicated by the dashed line), Fig 2.7A. The SLIDE process was then performed from all possible starting points on the landscape and used to estimate $\rho$, with the distribution of these outcomes being shown in Fig 2.7B. The ground truth $\rho$ value from Fourier is displayed as a blue dot on the violin plot. What can be seen is that on NK landscapes, the distribution of SLIDE estimates from each starting point (which we can consider local $\rho$ estimates) is firmly centred over the ground truth "global" $\rho$ value. This indicates that NK landscapes are highly homogeneous, and the SLIDE $\rho$ estimate from any one location is representative of the landscape in general. On empirical landscapes, however, local SLIDE $\rho$ estimates are highly inconsistent with the global value, indicating significant heterogeneity. In order to overcome this heterogeneity, it was necessary to estimate $\rho$ from multiple starting points as opposed to just one. This can be achieved by sampling, squaring, and averaging fitness decays from multiple starting points [125]. In order to explore this the global $\rho$ value was computed from an increasing number of starting points in Fig 2.7C. While GB1, TrpB and ParD3 rapidly converge upwards of 100 starting points, TEV appears highly heterogeneous and requires vastly more starting points before converging to a reliable estimate.

Fig 2.7D compares the different ruggedness measures across the empirical landscapes, using 10,000 starting points for $\rho$. Given the variation in scaling and linearity between metrics, comparisons should be focused on the ordering of landscapes within each metric, as opposed to comparisons of raw values between metrics. Local epistasis, distance to local maximum, and paths to global maximum

agree with $\rho$ in terms of their ordering of all the empirical landscapes tested. Landscape $R^2$ and roughness to slope ratio agree with the extremities (TEV as the most rugged, ParD3 as the least), however place GB1 and TrpB in a different order. Similarly, our metric $\rho$ groups together TrpB and GB1 landscapes as having a similar, intermediate ruggedness, whereas other methods separate these landscapes, or locate them very close to the TEV landscape. Given that other ruggedness metrics generally agree with the ordering of $\rho$ on empirical landscapes provides increased confidence in the metric, which has the major advantage over all others of being experimentally measurable without extensive sequencing (mutation rate estimation only).

## 2.6 A combined pipeline for optimisation of directed evolution

The final part of this chapter demonstrates how the output of the landscape inference method SLIDE can be used to optimally select parameters from Section 2.4 for improved DE outcomes. The pipeline described begins with an initial ruggedness inference phase, followed by selection of DE parameters from a look-up table, and finally implementation of the selected parameters (Fig. 2.8).

In order to explore the relationship between decay rate and optimal strategy and produce a look-up table, a large parameter sweep was conducted over 100 different $NK$ landscapes at fixed mutation rate and population size. For each landscape, 14,700 DE simulations were conducted over 49 strategy combinations. Fig 2.9A (dotted line) displays the optimal base chance and splitting values over landscapes

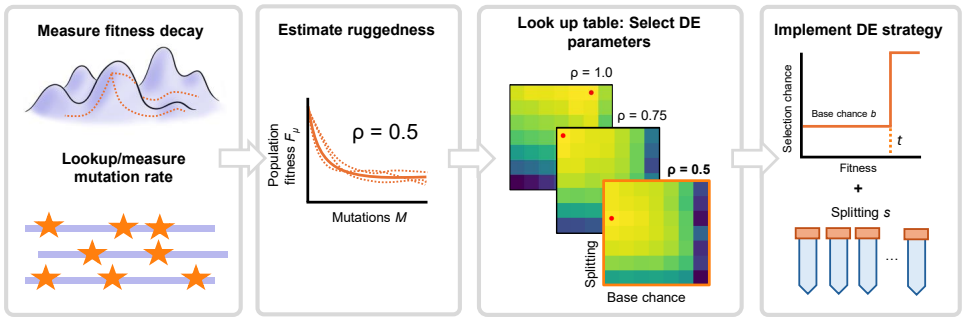



Figure 2.8: **Schematic outline of a future DE pipeline.** The initial phase of SLIDE involves random mutation away from a starting point, and measurement of the resulting fitness decay curve, which can be used to estimate the decay rate $\rho$ as a measure of ruggedness. A look-up table matches $\rho$ values of NK landscapes to optimal DE strategy parameters. This look-up table can be used to select appropriate parameters for the measured $\rho$. Parameters explored in this paper include base chance (applying a constant chance of selection that is irrespective of fitness) and population splitting.

of increasing epistatic parameter $K$. As $\rho \approx (K+1)/N$, this data can be used as a SLIDE look-up table to select an appropriate combination of DE parameters from $\rho$. Note that the shape of this look-up table is influenced by the experimental parameters such as mutation rate and generations of directed evolution (Fig 2.10), and thus in order to optimally implement SLIDE one should generate this look-up table using the known parameters of their own experiment and the publicly-available SLIDE codebase [125].

Two examples of DE curves on $NK$ landscapes are displayed in Fig 2.9B–C. Fig 2.9B displays a smooth landscape ($\rho \approx 0.02$), where the strategy selected by SLIDE is highly similar to the baseline strategy (which is already near the optimum). Fig 2.9C, however, displays a more rugged landscape ($\rho \approx 0.55$), on which the baseline strategy is highly sub-optimal and gets trapped at a local optimum within 10 generations. SLIDE, on the other hand, selects the optimal

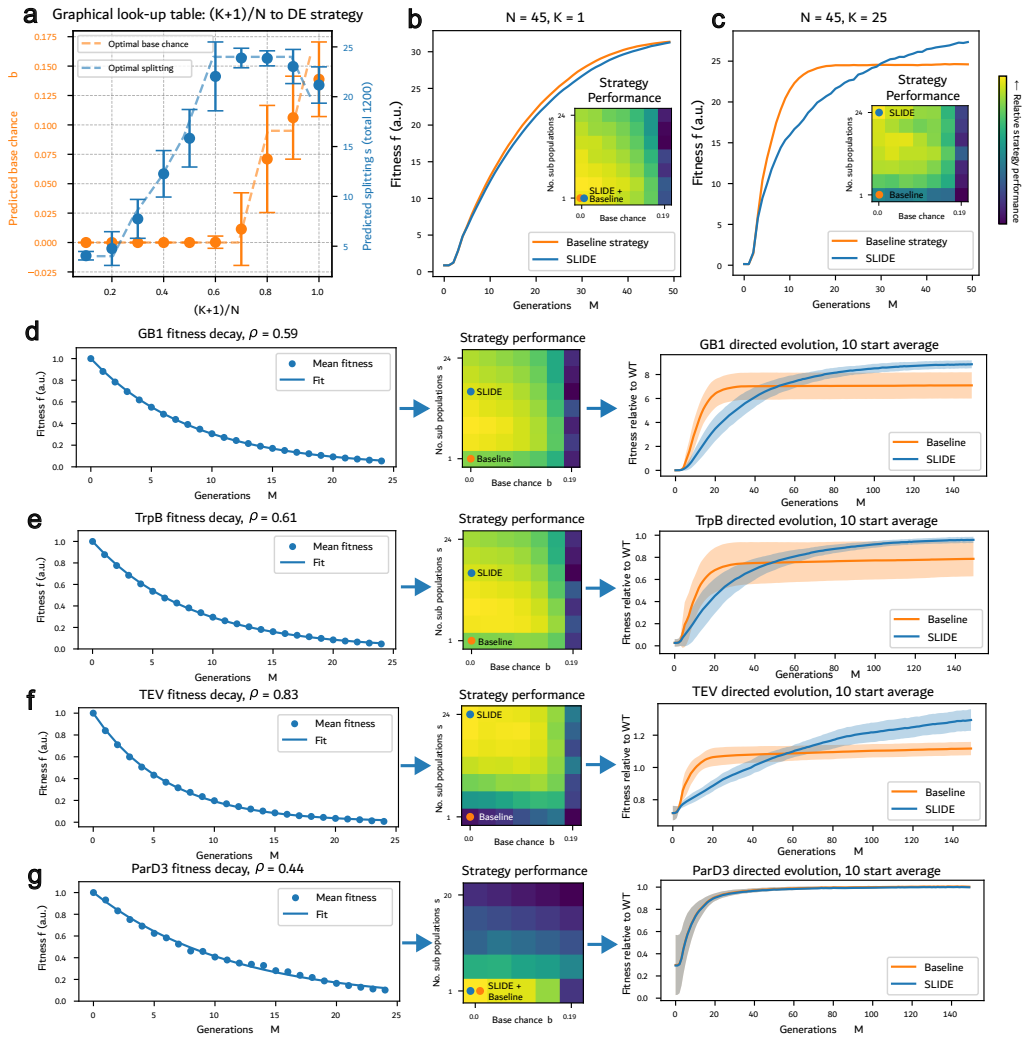



Figure 2.9: **DE using SLIDE on $NK$ and empirical landscapes. A:** Representation of the SLIDE look up table for $NK$. Generated from 100 $NK$ landscapes, 14,700 simulations over 49 strategy combinations. Base chance options = $\{0.00, 0.03, 0.06, 0.10, 0.13, 0.16, 0.19\}$, splitting options = $\{24, 20, 16, 12, 8, 4, 1\}$. Error bars = std over 250 simulations. **B-C:** SLIDE on an example smooth and rugged $NK$ landscape, respectively. Population size = 1200, mutations = $0.1/N$ per generation. SLIDE on **D:** GB1, **E:** TrpB, and **F:** TEV. Decay curves from 10,000 starts, mutations = 0.025 per generation, population size = 2500. SLIDE strategy look-up table generated for $A = 20, N = 4$. DE simulations conducted from 10 random starts, population size = 1200, mutations = 0.01 per generation. **G:** ParD3. Parameters as above, except base chance options = $\{0.00, 0.05, 0.10, 0.14, 0.19\}$, splitting options = $\{20, 15, 10, 5, 1\}$. SLIDE strategy look-up table generated for $A = 20, N = 3$. DE population size = 60.

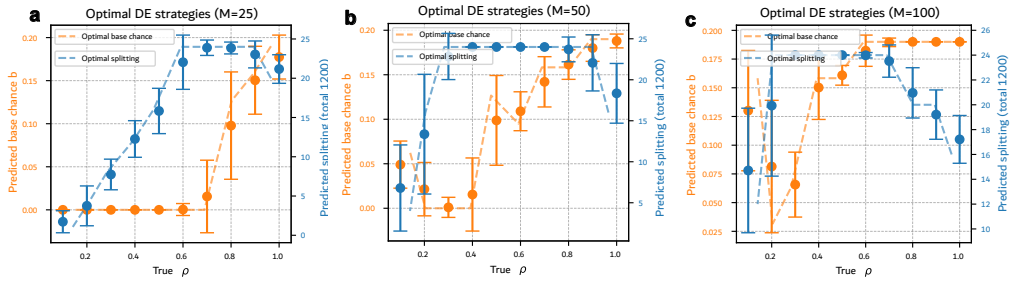



Figure 2.10: **The influence of DE generations on the SLIDE look-up table.** Displaying the optimal base chance and splitting parameters with respect to $\rho$ for **A:** 25, **B:** 50 and **C:** 100 generations (M) of DE.

strategy within the strategy space and achieves $> 15\,\%$ higher fitness than the baseline strategy after 50 generations.

A similar approach was then applied to the four combinatorially-complete empirical landscapes GB1, TrpB, TEV and ParD3 (Fig 2.9D-G). The strategy look up, as displayed in Fig 2.9A, was re-generated using $NK$ landscapes of the same dimensions as the landscape: $A = 20$, $N = 4$ (or $N = 3$ for ParD3). As well as this, it was necessary to increase the number of starting points sampled to generate an accurate estimation (as described in the previous section). SLIDE out-performed the baseline strategy on all examples, apart from ParD3, for which the baseline strategy was optimal (and therefore SLIDE selected the same parameters). On GB1, TrpB and TEV, it achieved relative fitness gains of 24.9% ($p < 0.001$), 22.6% ($p < 0.01$), and 44.2% ($p < 0.001$) respectively over the baseline approach, which commonly gets trapped at local maxima.

## 2.7 Methods

### 2.7.1 Directed evolution model

Selection algorithms were tested *in silico* using simulations of directed evolution, which can be applied to either synthetic or empirical landscapes. Each genetic variant is modelled within a population using a one-dimensional array of $N$ integer values, which is initialised at a random starting location, creating a population of size $P$. To calculate the corresponding fitness value of a gene, this array is either used as input to the NK model, or used as coordinates to look up fitness in an empirical dataset. The simulation algorithm proceeds with three cyclic steps; selection, proliferation, and mutation.

Individual cell fitness values are input to a selection function, which translates relative fitness into probability of selection. Then, each cell is selected (or not) based only on its respective probability of selection. Cells that are selected are proliferated to bring the population back up to its original size. To perform proliferation, a *new* population of size $P$ is created by randomly sampling (with replacement) from the previously selected cells. This introduces a degree of stochasticity mirroring experimental error and biological variation, as described in other models of evolutionary processes [146].

Once a population of size $P$ has been selected and proliferated, mutations are introduced by making random changes to genes in the population. For every cell's genetic code (an array), each residue (i.e. nucleotide or amino acid) has a random chance, $p_I$, of changing into another random *different* value (i.e. each other

possibility is equally likely). Given the gene length $N$, the expected number of changes across the entire gene (the "mutations per cell") is given by $\mu = Np_I$. This quantity is useful to work with, as a fixed $\mu$ will give comparable results as $N$ varies.

The final result is a new population of size $P$, and the process of selection, proliferation, and mutation may be repeated again. In this work, it is assumed that this runs for a fixed number of iterations, and the final result (by which methods are compared) is the maximum fitness across the final population. "Normalised fitness" divides this measure by the global maximum of the landscape.

### 2.7.2 NK landscapes

The NK model is a widely used approach for generating synthetic fitness landscapes with tuneable ruggedness [96, 97]. In the NK model, a gene is represented by an array of $N$ sites, each of which has $A$ possible values. In Section 2.4, $A = 2$ such that each entry has two possibilities (i.e. binary 1 or 0). In Sections 2.5 and 2.6, $A$ and $N$ are defined to match the shape of any empirical landscape they are being compared to (e.g. In comparison to GB1, $A = 20$ and $N = 4$). Every entry in the gene corresponds to a "locus", which interacts with $K$ other loci in the gene. The fitness contribution of each locus is dependent on the state of that locus and the state of the $K$ other loci it interacts with . The fitness $F$ of a gene $G$ is the sum of the fitnesses of each locus. When $K = 0$, all loci are independent and the model is linear (and hence has a single peak). The other extreme, $K = N - 1$, in which every locus interacts with every other locus, is maximally unstructured; the fitness landscape consisting of only random noise.

The NK model defines a parameterised *distribution* over functions $F : A^N \to R$. To define how to generate samples from this distribution, first let $L : N \times K \to N$ be a "locus function", where $L(a, i)$ is the $i$th site that locus $a$ interacts with. We require each locus to interact with precisely $K$ other, uniformly random, sites, independent of all other loci. We then also have $N \cdot A^{K+1}$ independent, identically distributed standard normal ($\mathcal{N}(0, 1)$) random variables, which we denote as $X^a_{i_0,i_1,i_2...,i_K}$, where $a \in \{1...N\}$ and $i_{0,1,...,K} \in \{1...A\}$. These represent the possible fitness contributions of each loci, with $a$ being the index of the loci in question, and $i_{0,1,...,K}$ being the values of that loci and the $K$ other loci it interacts with, plus itself ($i_0$). Then, $F$ is given as the following, where $g \in A^N$, and $g[i]$ denotes the $i$th site in $g$:

$$F(g) = \sum_{a=1}^{N} X^a_{g[a],g[L(a,1)],g[L(a,2)],...,g[L(a,K)]} \tag{2.2}$$

Computationally, the algorithm used explicitly generates and stores $L$; in particular as a $N \times N$ binary matrix. However, it does not store the $N \cdot A^{K+1}$ random variables explicitly as doing so would require large amounts of memory. Instead, every time the value of $X^a_{i_0,i_1,i_2...,i_K}$ is required, $a, i_0, i_1, ..., i_K$ are used as inputs to a pre-defined deterministic pseudorandom generation algorithm.

### 2.7.3 Selection functions

In this work, the concept of a "selection function" was introduced. This function takes a cell's fitness percentile, defined as the percentage of other cells in the population it is fitter than, and outputs its chance of being selected to be in the next

population. The selection function used in this work is defined by two parameters: a "fitness threshold" and a "base chance"(Fig 2.1A). Above the "fitness threshold" the selection function outputs 1, otherwise it outputs the "base chance" (Equation 2.3). In this work, selection functions are normalised to select a constant fraction of cells (20% unless stated otherwise). By fixing the selected proportion (given by the integral of the selection function), our parameter search space is reduced to one dimension, as every base chance has only one possible corresponding fitness threshold value (Equation 2.4). In addition, it ensures that in a real continuous directed evolution experiment, the time required for proliferating cells between iterations remains effectively constant such that a "fair" comparison is made between strategies in terms of performance per-unit-time.

$$\textit{selection chance} = \begin{cases} \textit{base chance} & \text{if } \textit{fitness percentile} < \textit{fitness threshold} \\ 1 & \text{if } \textit{fitness percentile} \geq \textit{fitness threshold} \end{cases} \tag{2.3}$$

$$\textit{fitness threshold} = \frac{1 - \textit{selected fraction}}{1 - \textit{base chance}} \tag{2.4}$$

### 2.7.4 Empirical landscapes

Empirical fitness landscapes are real data sets produced by measuring the fitness of all possible sequential variants of a protein (or region of a protein). The empirical landscapes used in this work are GB1 [88], TrpB [127], TEV [90] and ParD3 [84]. In this work, the landscapes (each three or four amino acid positions) are stored as

three- or four-dimensional arrays, where each dimension corresponds to a variable residue. By assigning every amino acid a number from 1 to 20, each sequence of amino acids can therefore be mapped to a set of coordinates that points to the corresponding fitness value in the array.

## 2.8 Discussion

As well as providing the theoretical support for future applications of the hardware in this thesis, this chapter demonstrates that even in the absence of sequencing information, there are approaches that can be used to improve directed evolution outcomes. The complete pipeline presented begins with sequence-free landscape inference using SLIDE, followed by informed implementation of tools such as selection functions and population splitting to maximise directed evolution outcomes. The pipeline was validated on both synthetic $NK$ landscapes and empirical datasets, showing that it can result in up to a $44.2\,\%$ improvement in directed evolution performance compared to the most common baseline directed evolution strategy.

There are several ways in which the methods outlined in this paper could be implemented experimentally. For SLIDE, one example is random *in vitro* mutagenesis, which allows targeted exploration of specific regions of the genotype space. Mutant libraries generated in this way can be expressed in cells, with fitness linked to a measurable phenotype like fluorescence. By iteratively measuring average fluorescence across generations and estimating mutation accumulation (or by simultaneously producing samples with increasing levels of mutation), it becomes possible to construct fitness decay curves and apply SLIDE in a laboratory set-

ting. To enable a sufficient number of starting points for a reliable estimate, such an assay could be conducted across several 96-well plates (or more, if combined with automated pipetting robots), or reduced to a pooled format using mother machine microfluidics [12, 13]. Mutation rate estimation might already be known for well-established workflows, or if not can be estimated by sequencing a sample of variants that is orders of magnitude smaller than the sequencing requirements of a combinatorially-complete empirical landscape.

With respect to the tools for directed evolution itself, "base chance" parameters could be implemented by pipetting the appropriate portion of pre-selection cells into a post-selection population. A more sophisticated approach could incorporate the selection function into the Fluoresence Activated Cell Sorting (FACS) methodologies, or of course into the microfluidic selection platform MiPS described in this thesis (Chapter 4) [56]. It should also be noted that a certain degree of base chance is likely naturally incorporated into many directed evolution experiments as a result of selection inaccuracy. Population splitting is the most intuitive to implement, and in many cases is already performed, in the form of biological replicates. Effective population splitting could also be implemented using segregated channels on microfluidic chips. Multi-dimensional selection in the form described here can only be implemented with sorting technology that allows for multiple properties to measured prior to selection (i.e. the microfluidic platform described in this thesis).

One limitation of this work is the reliance on NK models. NK landscapes were used for testing because empirical landscapes, such as that of GB1 and TrpB, are scarce. NK models do not perfectly represent the statistical properties of nat-

ural fitness landscapes (as was made very clear with respect to heterogeneity, Fig. 2.7B), therefore in order to improve the reliability of simulations such as these, future work could tailor the NK model to the specific application of protein fitness landscapes. For instance, by allowing K to be drawn from a distribution for each amino acid position, as opposed to a constant value, or by integrating the information offered by PAM substitution matrices into fitness estimations.

Further improvements to the landscape inference component could see integration of pre-existing sequence data (when available). By utilising multiple sequence alignment (MSA) information, it is possible to identify interacting residues from patterns of co-evolution (the frequency of interacting residues being an indicator of ruggedness). This approach is employed by EVmutation [121] and DeepSequence [122] to predict the effects of novel mutations. It is also employed in SLIP [123], which uses MSA information to generate synthetic landscapes. These methods are not currently accurate enough in isolation to predict sequence-to-function fitness peaks reliably, but if used in combination with SLIDE (i.e. to strengthen ruggedness-informed selections of base chance and splitting, as well as to provide promising starting points), they could produce a superior approach overall.

In summary, the pipeline presented here offers a novel method for estimating the ruggedness of protein fitness landscapes using average phenotypic data and known mutation rate, and using the resulting prediction to inform experimental strategies in directed evolution. This body of theory work provides essential support for the hardware platform described in subsequent chapters of this thesis, which is an optimal platform for continuous DE.

## 2.9 Statement of authorship

Much of the work in this chapter was a joint collaboration with Sebastian Towers. In the first paper represented (*Optimisation strategies for directed evolution without sequencing*), I was responsible for the initial conceptualisation of the work, as well as the initial implementation of the pipeline in Python. Sebastian Towers offered extensive expertise to re-work the pipeline for implementation on the GPU via JAX, which allowed for significant increases in computational efficiency. Figs 2.1, 2.2 and 2.3 were a collaborative effort, during the production of which a lot of fundamental aspects of the codebase were refined. The work represented in Fig 2.4 was primarily my own, and that represented in Fig 2.5 primarily that of Sebastian. This paper was supervised by Harrison Steel, who provided valuable guidance throughout, as well as Jakob Foerster, who offered additional support.

With respect to the second paper represented here (*Sequence-free landscape inference for directed evolution*), the initial conceptualisation was led by Idris Kempf and Sebastian Towers, who together laid the mathematical ground work for this project (much of which has been excluded from this thesis). I joined the project at the implementation stage and conducted a large proportion of the computational work. Specifically, Fig 2.6 was a collaborative effort between Sebastian and I, during the production of which we again refined many aspects of the computational pipeline. Fig 2.7 was again a collaborative effort. The work represented in Fig 2.9 was conducted by myself, and Fig 2.10 was primarily developed by Idris Kempf using my codebase from Fig 2.9. Idris Kempf was the primary supervisor on this work, with additional guidance provided by Harrison Steel.

# 3 | Hardware: Development of an automated platform for light targeting and long-term observation of cells on microfluidic chips

## 3.1 Featured publications

This chapter primarily consists of work from this co-first author paper:

- Microscopic PhotoSelection (MiPS) of single cells in mother machine microfluidic devices [56]

## 3.2 Abstract

The vast majority of microbial experiments in biological sciences are conducted on a population level, for instance using flasks or well-plates. The advent of flow cytometry unlocked the ability to take high-throughput phenotypic measurements on a single-cell basis, however limited to single time point measurements of each cell. The emergence of microfluidics has since allowed continuous culturing of cells in spatially controlled formats, such that single cells can be observed and tracked for long time periods. One specific design, known as the "mother machine", has proved transformative for single-cell phenotypic research. This design consists of a central channel lined with one-cell-thick trenches in which bacteria can grow stably in exponential phase. This chapter describes the development of a custom platform that is specialised for mother machine-based experiments,

including an integrated Digital Micromirror Device (DMD) that allows patterned projection of selected LED wavelengths down to sub-cellular resolution. Such a platform allows for real-time observation of single cells in response to various chemical, optical or magnetic stimuli, on time scales from milliseconds up to several hundred generations.

## 3.3 Background

In biological systems, populations are often highly heterogeneous due to stochastic gene expression, cell cycle differences, or micro-environmental variations [147]. For this reason, single-cell measurements offer large advantages over population-level measurements, for which such details are averaged out. For short-term measurements, single cells can be observed on microscope slides, however they will soon exhibit stress-related responses due to nutrient limitation. This time period can be extended with the use of agar pads, which provide a nutrient source for cells, however this context still produces significant spatial variability in growth conditions (particularly after a few generations when cells begin competing for resources), and can prove challenging for lineage tracking [148]. Microfluidics, on the other hand, offers highly controlled environmental conditions, and the ability to spatially control growth channels for optimal lineage tracking [67]. Continuous supply of nutrients in microfluidic chips also extends the time period of observation to several hundred generations [13].

In general, microfluidic chips are comprised of a PDMS layer into which custom growth chambers and channels are cast, which is then bonded to a glass coverslip, sealing the channels and allowing for observation of their contents. The earliest

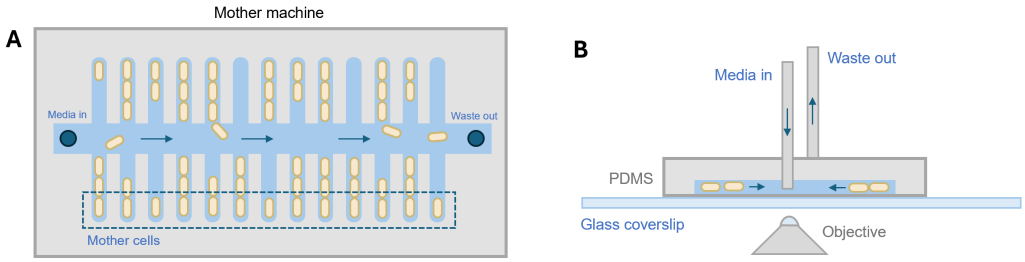



Figure 3.1: **Schematic representation of a mother machine microfluidic chip. A:** Top down view of mother machine, showing the central channel through which media flows, lined with trenches on either side within which bacteria grow. The cells at the end of each trench are known as the "mother cells". **B:** Side view of the mother machine, showing the fusion of PDMS to coverslip, and the entry points of needles to deliver media.

example of observing bacterial growth in PDMS channels was to investigate bacterial persistence [66]. In this design, bacteria were grown in long, linear channels that were closed at either end, and fed media through a membrane. The spatial limitation of that design (which would only allow bacterial growth until the channels were filled), was overcome by the introduction of the "mother machine" [12] (Fig. 3.1 A-B). In the mother machine, the "trenches" containing cells are open at one end, which connects to a central media channel that both allows continuous supply of nutrients, and clears bacterial daughter cells away once they have been pushed out of the trench.

Mother machine designs are continually evolving to become ever more sophisticated. The throughput of designs has been increased to up to $10^6$ trenches by replacing the linear channel with a serpentine design [13]. The addition of backports to individual trenches enables improved flow of nutrients to cells, but also unlocks the ability to load and unload cells from trenches using the microfluidics (as opposed to centrifugation) [69, 70].

Modifications in chip design have also expanded the breadth of cell types and applications that can be studied in this format. Originally developed for *E. coli* [71], it is now possible to culture Gram-positive bacteria such as *B. subtilis*, pathogens such as *V. cholerae*, and eukaryotes such as yeast and mammalian cells in mother machine style chips. Applications span studies of cell size control [72], bacterial aging [73], stress responses [74–76], sporulation [72, 77, 78], and virulence regulation [70]. In mammalian cells, mother machines have been used to investigate chemotherapy-induced cell-cycle dynamics [79] and population heterogeneity in cancer [80]. For medical applications, bacterial mother-machine studies are particularly relevant to diagnostics, drug-response profiling, and the development of controllable therapeutic strains [149].

There are multiple ways by which one can interact with cells in the mother machine. The studies above primarily utilise chemical stimuli - simply adding chemical inducers into the media that flows into the mother machine and observing single-cell responses in real time. This can be done in a highly controlled fashion, and as media is continually being replenished, has the other significant advantage of being able to swap chemical inducers rapidly, with minimal undesired disturbance to the system. One limitation of chemical stimuli is that dosage can not be independently determined on a cell-by-cell level. This can be achieved, however, using projected light via a DMD. A DMD is an array of many thousands of small mirrors that can each be switched between on and off positions to determine the pattern of light that shines onto the sample. By integrating a DMD into the path of an LED to the microfluidic chip, one can dictate the pattern of light that shines onto cells, and therefore target individual cells with different dosages. This is

shown in [16], in which the authors implement closed-loop control expression of an optogenetic gene circuit by delivering targeted signals to independent cells in response to their current expression levels. Recently, a third mode of interaction with cells has been investigated in the mother machine context, which is via a magnetic field [11]. In this study, the platform developed in this thesis was used to demonstrate it possible to modulate the brightness of a fluorescent protein by exposure to a magnetic field.

This chapter will describe the development of a platform that is fully optimised for DMD-integrated mother machine experiments. The system allows for continuous observation of thousands of single-cells in the mother machine, and interaction via chemical, optical and magnetic signals. This chapter outlines both the custom hardware and software that was developed, alongside the extensive protocol development that was carried out to establish a functioning system. The final product is a system applicable to tasks such as large-scale single-cell strain profiling, directed evolution via photoselection (see Chapter 4), or data collection for machine learning, and the applicability can be expanded by novel microfluidic chip designs and microscopy tools.

## 3.4 System components

### 3.4.1 System overview

A schematic of the key components of the system is shown in Fig. 3.2. The purpose of the system is to maintain and observe cells in the mother machine for long time periods, with the ability to interact with them primarily via chemical

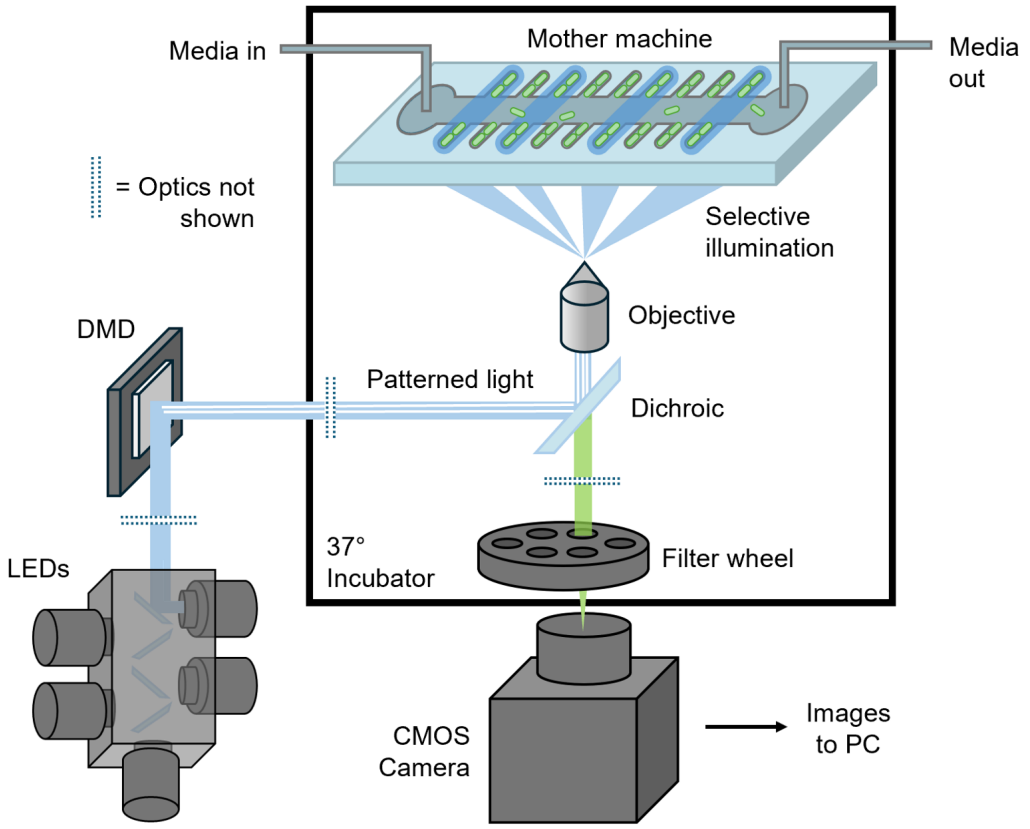



Figure 3.2: **Hardware Overview**. Schematic view of the system, which aims to deliver targeted patterned light to cells on the mother machine. Light from a selection of LEDs travels via a DMD, producing patterned light. Patterned light is reflected onto the sample via a dichroic, through which emission wavelength light from the sample returns and is transmitted via a filter wheel to a high speed camera. Excluded optics are displayed in Fig. 3.6.

and individual-cell targeted optical signals. Images of the process of assembly are shown in Fig. 3.3A and B. An external view of the completed system is displayed in Fig. 3.3C. A high-performance computer runs the custom control software, which is displayed on a monitor beside the system. The software enables manual or pre-programmed control of the LEDs, DMD, camera, microscope stage, and autofocus (not shown). An independent temperature control unit maintains the internal temperature of the custom-built incubator. An internal view of the incubator is displayed in Fig. 3.3D. The chip is mounted via a custom-built stage insert that allows for rotation and clamping. An overhead light is positioned for bright field imaging. Additional remote monitoring (i.e. for microfluidic leaks) can be achieved via an internally mounted camera. The microfluidics is controlled by an additional custom software, which runs independently on a Windows machine (not shown).

### 3.4.2 ASI microscope

The custom built microscopy platform is based on the Rapid Automated Modular Microscope System (RAMM) from Applied Scientific Instrumentation (ASI). The configuration includes a Nikon PLAN APO $\lambda$D 40x/0.95 N.A. objective (Nikon: MRD70470), a Teledyne Kinetix sCMOS Camera, a five-band fluorescence filter set (405/445/514/561/640$\mathrm{nm}$ Quinta Band Set, Chroma part no. 89904), and Nikon Ti-2 200mm tube lens (36mm clear aperture), resulting in a pixel size of $0.1625\,\mu\mathrm{m} \times 0.1625\,\mu\mathrm{m}$ and a field of view (for the camera's 3200 $\times$ 3200 sensor of $6.5\,\mu\mathrm{m}$ pixels) of 0.52 mm $\times$ 0.52 mm. The stage, filter wheel, and hardware autofocus system (ASI CRISP, augmented with periodic software/camera-based

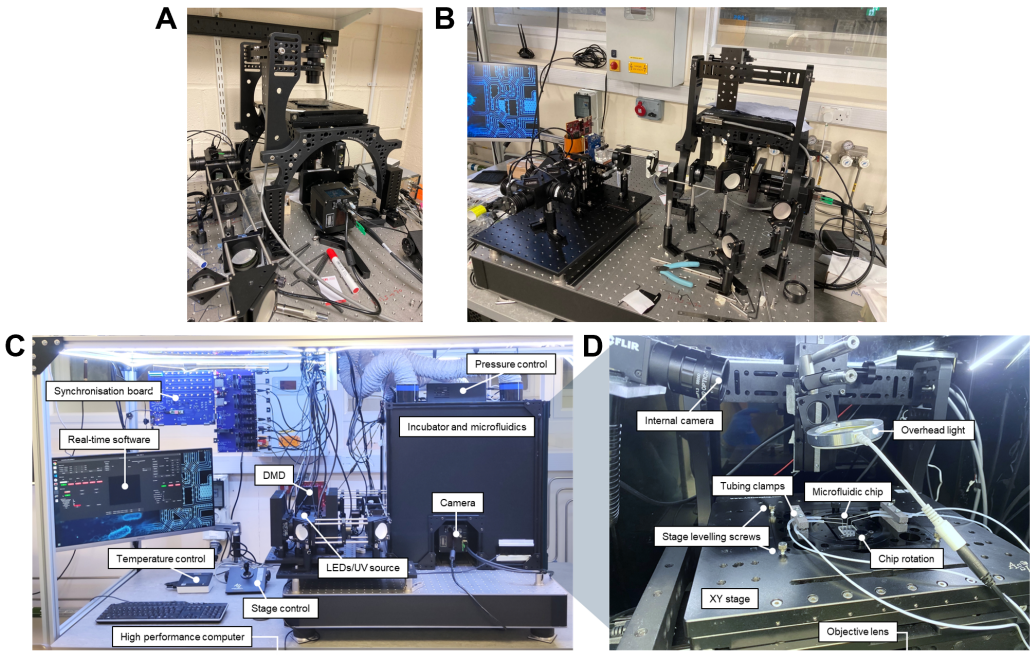



Figure 3.3: **Assembly of the microscope**. **A:** Initial assembly of the RAMM system (13th April 2023). **B:** Optics alignment and LED integration (23rd August 2023). **C:** External view of the system, showing the control software on the left, which interacts with the LEDs, DMD, microscope stage and camera via a custom built syncronisation board. The mother machine is housed within a custom blacked out incubator. **D:** Internal view of the system, showing the mother machine microfluidic chip mounted on the custom stage insert, with microfluidic tubing connected. An overhead light allows for bright field imaging, and an internal camera allows for remote monitoring of leaks or similar events.

autofocus checks) are driven by an ASI TG-1000 Controller.

### 3.4.3 Stage

In order to optimise the hardware for mother machine experiments, a custom stage insert was designed to be compatible with the RAMM system (Fig. 3.4A). This stage insert allows for rotation of the microfluidic chip (to account for the common occurrence of mother machines not being fused perfectly straight during bonding).

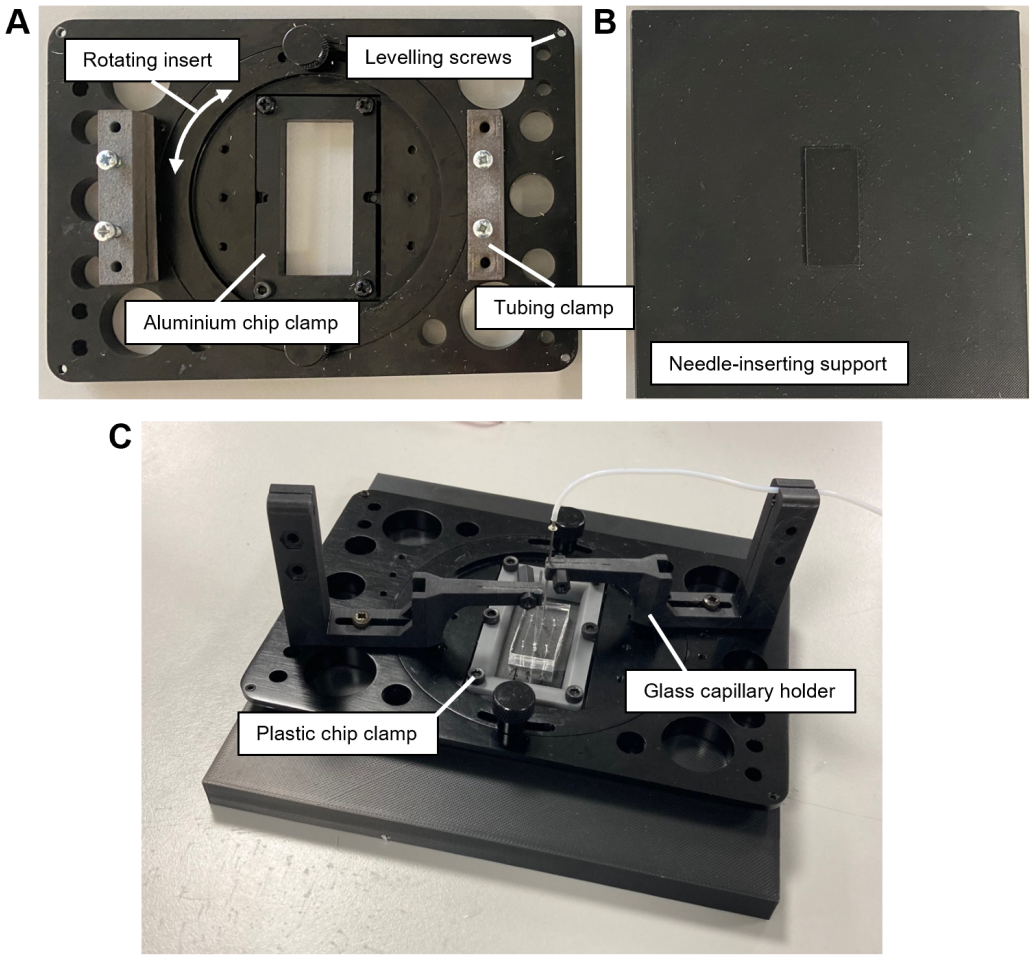



Figure 3.4: **Custom stage insert and support**. **A:** Stage insert allowing for chip rotation, and clamping of the chip itself as well as microfluidic tubing. **B:** Needle-inserting support is positioned underneath the stage insert during insertion of needles. This presses against the back of the chip surface to prevent breakage. **C:** Stage inert and support together, alongside 3D printed holders for glass capillary tubes, which enable projection of UV into the inlet and outlet holes.

It also has a compatible clamp which holds the chip down and reduces drift. In addition, it has several threaded holes on each side to attach holding clamps for different kinds of microfluidic needles, as well as a threaded hole on each corner for levelling screws. The stage insert was designed in Solidworks and manufactured in three parts via CNC in black anodised aluminium. It can be seen in context in Fig. 3.3D.

When the stage insert is placed on a bench there is an approximate 1 mm gap between the mother machine coverslip and the bench (as a small amount of metal is required to hold the chip underneath when mounted). For this reason it was necessary to manufacture a support that could be used to insert microfluidic needles without cracking the microfluidic chip. This 3D printed support is shown in Fig. 3.4B, and simply has a raised section that presses up against the back of the microfluidic chip while needles are being inserted.

Both components can be seen together in Fig. 3.4C. In addition, alternative tubing clamps were manufactured via Solidworks and 3D printing which support the use of glass capillary tubes for microfluidics as opposed to metal needles. The purpose of using glass capillary tubes is to enable projection of UV into the inlet and outlet holes, to try and eliminate contamination (used primarily in Chapter 4).

### 3.4.4 Incubator

A custom incubator was built to the correct dimensions of the RAMM microscope, while excluding the camera to reduce risk of overheating (Fig. 3.5). The frame was assembled from scratch using 25 mm black anodised aluminium railing, cut to size and tapped with M6 holes to allow joining via construction cubes (Thorlabs:

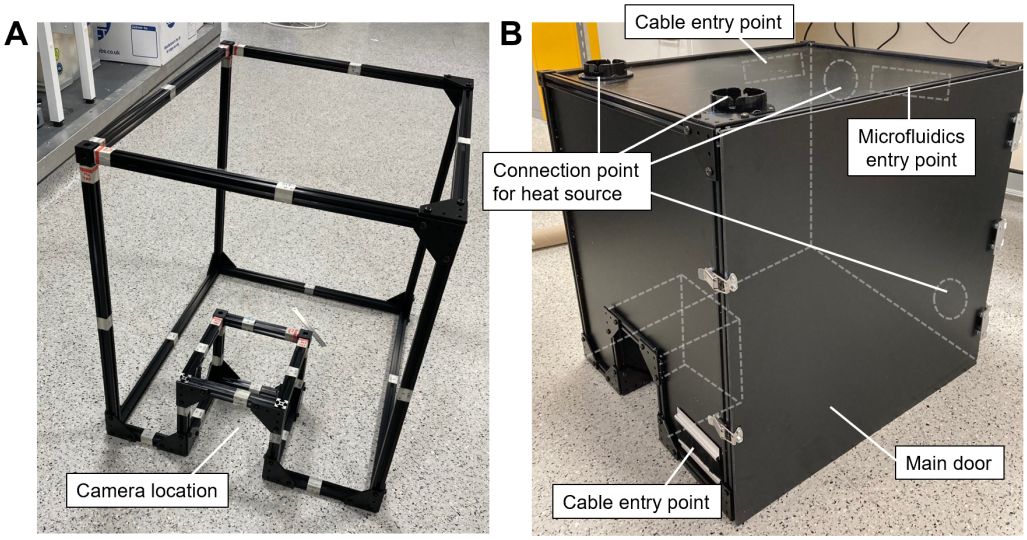



Figure 3.5: **Custom Incubator A:** View of incubator prior to panel insertion, showing the inset region where the camera is positioned to minimise risk of overheating. **B:** View of incubator with panels, showing the attachment of the main door, as well as entry points for cables and microfluidics, and connection points for the heat source.

RM1S) and assorted gusset plates. To ensure light was excluded, large 3 mm sheets of acrylic were cut to the correct dimensions and covered with black film coating. Cut outs were made for the heat source connection points, as well as for microfluidics and cable entry points. To minimise heat loss from these entry points, brush draught excluders were cut to size and attached to each hole via magnetised strips for ease of attachment and removal. Foam strips were also used to minimise heat loss around all edges of the main door. Finally, the system was connected to an Okolab temperature control unit.

### 3.4.5 Optics

A schematic representation of the optics can be seen in Fig. 3.6. The light source includes a custom-built synchronisation electronics board (not shown, based on

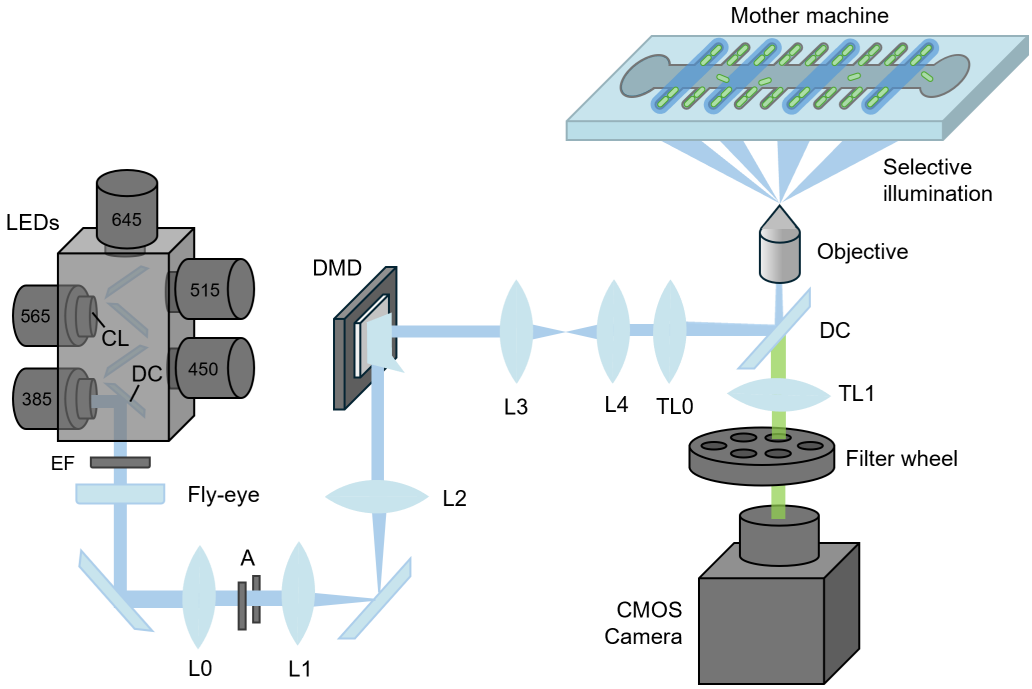



Figure 3.6: **Schematic representation of the optics setup.** Light travels from the LEDs via condenser lenses (CL) that collimate the light and compatible dichroics (DC) which direct light outwards. An excitation filter (EF) ensures the correct wavelengths for the system are passing through, and a fly-eye lens maximises the uniformity of the light. A series of lenses (L) and tube lenses (TL) direct light in the correct proportions onto and away from the DMD, which creates the pattern. Another dichroic directs excitation light onto the sample via the objective, and emission light passes back through via another tube lens and a filter wheel to the camera sensor.

the Teensy 4.1 ARM Cortex-M7 microcontroller) and a custom-built LED driver system which can provide up to $70\,\mathrm{W}$ of power to each of up to eight LEDs channels with hardware-based feedback control of LED current or optical output power. The system was configured with five LEDs with wavelength centres at 385/450/515/565/645 $\mathrm{nm}$ (Luminus Devices: CBM-50X-UV-Y31-FA380-22, SBR-70-B-R75-KG301, PT-39-G-L51, CBT-90-CG-L11-G100, and PT-39-DR-L51-BD100 respectively). LEDs are first collimated using aspheric condenser lenses (CL) (Thorlabs: ACL25416U-A), and then combined with dichroics (DC) with cut-off wavelengths at $600\,\mathrm{nm}$, $525\,\mathrm{nm}$, $470\,\mathrm{nm}$, $425\,\mathrm{nm}$ respectively. The combined beam passes through the excitation filter (EF) of the five-band filter set (Chroma, described above) before reaching a custom made two-sided microlens array / fly-eye to create a uniform rectangular illumination profile [150–152] of approximately the same aspect ratio as the DMD.

The fly-eye is placed at one focal length (FL) from lens L0 ($180\,\mathrm{mm}$ FL Thorlabs AC508-180-A) with a broadband dielectric mirror between (Thorlabs BB2-E02). An adjustable, 3D-printed rectangular aperture (A) is placed at the focal point between L0 and L1 ($200\,\mathrm{mm}$ FL Thorlabs ACT508-200-A), allowing the beam shape to be cut to ensure the DMD is not over-filled. This is followed by another lens (L2, $100\,\mathrm{mm}$ FL Thorlabs AC508-100-A), with the relative FLs of L1/L2 chosen to create a 4-f relay to de-magnify the fly-eye lens image to fill the DMD. Light arrives onto the DMD through a TIR Prism [153] (adapted from a large industrial data projector) allowing a setup similar to [154]. Use of a prism here ensures light hits the DMD mirrors in the "ON" position the correct angle, and light reflected from mirrors in the "OFF" position does not propagate any further

through the system. With L3 (75 mm FL Thorlabs: AC508-075-A), L4 (Thorlabs: AC508-150-A) is chosen to create another 4-f relay system such that, when combined with TL0 (160 mm FL Edmunds Optics # 49-382), the DMD's image is magnified to slightly over-fill the microscope field-of-view visible to the camera's 20.8 mm square sensor.

| Position | Filter | Wavelength (nm) | Optimal for |
|---|---|---|---|
| 0 | ZET405-445-514-561-640m | All non-excitation | Non-specific |
| 1 | ET480/30m | 465–495 | CFP |
| 2 | ET540/24m | 527–551 | YFP, GFP* |
| 3 | ET609/34m | 592–624 | mCherry, RFP |
| 4 | Empty | NA | NA |
| 5 | Beam block | NA | NA |

Table 3.1: **Filter wheel configuration**. Details of the filters present in each position of the filter wheel, alongside information on optimality for various fluorescent proteins. *Note that the system is sub-optimal for GFP (as shown in Fig. 3.7), however good quality measurements can still be obtained with medium to high expression.

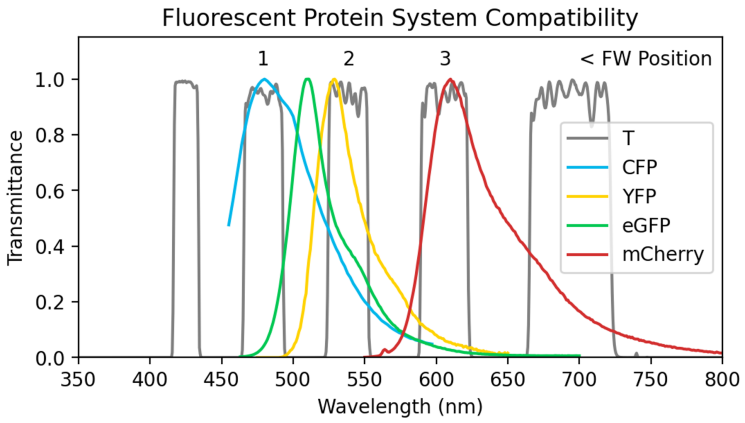



Figure 3.7: **Fluorescent protein and filter compatibility**. Emission wavelengths that are transmitted by the system dichroic (T) are shown in grey. The emission spectra of CFP, YFP, eGFP and mCherry are overlaid, showing the proportions of the emission that can be captured by the system, as well as the optimal band-specific filters (FW position) for isolating individual FP signals.

Prior to reaching the camera, however, light emitted from the sample travels back through the objective, through the dichroic and through the filter wheel. Several filters are present in the filter wheel, which are detailed in Table 3.1. Position 0 contains a filter for all emission wavelengths coming through the dichroic. This is shown in Fig. 3.7 in grey. The remaining filters are specific to individual bands, so the filters in FW position 1, 2 and 3 allow light from the 2nd, 3rd and 4th bands respectively to pass through. These band-specific filters can be used to increase selectivity for specific fluorescent protein signals. For instance, CFP is best isolated with the filter in position 1. It is also possible to see from Fig. 3.7 that the system is not optimised for GFP, however due to the high intensity of the protein it is still possible to achieve adequate signal to noise ratio with medium to high GFP expression levels.

### 3.4.6 Microfluidics

The microfluidics is controlled by an OB-1 MK3+ Flow Controller (Elve Flow). The OB-1 delivers pressurised air into a 50 ml reservoir containing growth media, which is mounted outside of the incubator at room temperature to reduce the impact of any contaminations. When pressurised air from the OB-1 enters the reservoir, media is pushed out into the system, which consists of an MFS Microfluidic Thermal Flow Sensor (Elve Flow) which provides flow measurements for feedback control, followed by a 5 cm 100 $\mu$m inner diameter resistance element (Darwin Microfluidics: LVF-KTU-05) for more robust flow control (Fig. 3.8). After the resistance element, a 0.2 $\mu$m filter is attached to prevent contamination reaching the chip or output. From the filter, the tubing connects directly into the chip via

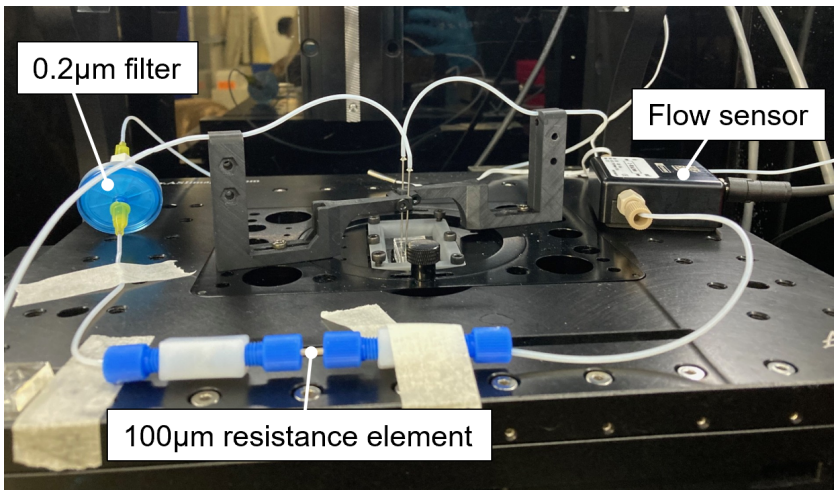



Figure 3.8: **Internal microfluidics setup**. Media from the reservoir first passes through a flow sensor, which enables feedback for flow control. A 100$\mu$m resistance element increases the system resistance for greater control. A 0.2$\mu$m filter reduces risk of blocking the chip.

glass capillary tubes or metal syringe needles, depending on the application. The system can also be integrated with a Microfluidic Distribution valve (Elve Flow) for automated switching between media sources. Custom software was developed by Kirill Sechkar to control this system from a Windows PC, and overcome many of the issues with the supplier's original software.

### 3.4.7 System control

Control of the overall system (camera, stage, LEDs, DMD, autofocus) was primarily developed by Idris Kempf. The main process is implemented in Python on a PC equipped with an NVIDIA RTX 3090 Ti GPU (24 GB VRAM) and 256 GB RAM. It interfaces a modified high-throughput version of the DeLTA pipeline for cell segmentation [155, 156], and automates and coordinates the microscopy hardware and software components from Fig. 3.2. The camera is controlled using

pycro-manager [157]. The stage, filter wheel, and the autofocus system are controlled using a Python serial interface with the ASI TG-1000 Controller (building on the open-source repository [158]). A similar interface is used to connect to the synchronisation board (itself running in C++), which controls the LEDs and connects hardware triggers. The DMD is operated using the Video Pattern Mode of the DLP LightCrafter software [159] and connected to the control PC via 4K HDMI. On the control PC, this HDMI connection appears as a second monitor onto which the patterns are displayed using a lightweight display window, implemented in C using Simple Direct Media Layer (SDL3) and interfaced via a local websocket with the main Python control software. All elements together can be controlled via a custom GUI. Users can either control the microscope manually from the GUI, or upload pre-defined scripts with strategy information (e.g. imaging every 3 minutes at 565 nm).

## 3.5 Establishing chip fabrication capabilities

### 3.5.1 Lab setup

To enable this project, the capability for chip fabrication in our lab was established. This required both acquiring and assembling the necessary equipment, as well as developing protocols for chip fabrication. The final lab set up for chip fabrication is displayed in Fig. 3.9A. The Henniker HPT-100 is the plasma oven used for bonding PDMS to glass coverslips. The plasma oven is connected to the vacuum pump (Edwards: nXDS) via a three-way ball valve (Pfeiffer Vacuum: 120VKD025-T.A), in a custom setup that enables flow to be redirected to

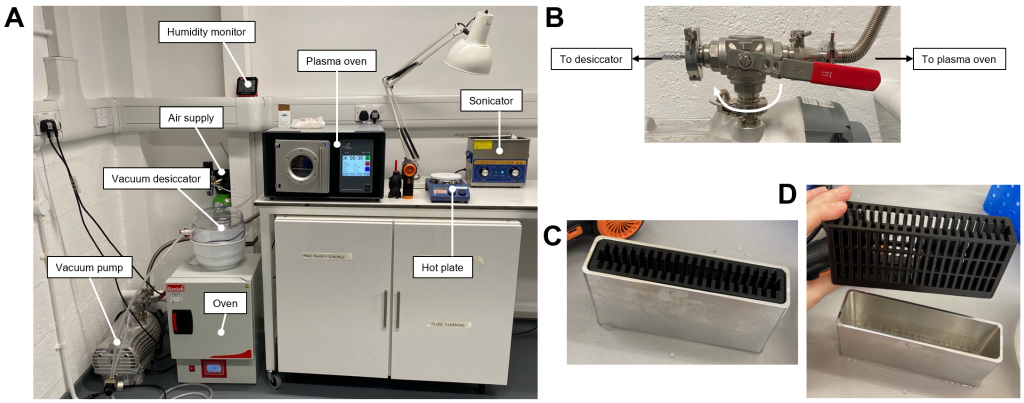



Figure 3.9: **Microfluidics lab. A:** Overview of the lab setup, indicating the core components. **B:** A three-way valve setup on the vacuum pump was designed to increase versatility of the system, with ease of switching between the plasma oven and vacuum desiccator. **C-D** A custom-designed microscope slide cleaner enables simultaneous cleaning of 20 slides. An inner casing 3D printed in nylon holds the slides, and is resistant to IPA and acetone. An outer casing produced from aluminium via CNC is used to contain the solvents.

a desiccator (Sigma-Aldrich: Z119016) when required. In addition, a hot plate is positioned near the plasma oven for post-bonding heat treatment, as well as an oven. A sonicator is used for cleaning coverslips prior to plasma treatment.

### 3.5.2 Microfluidic chip fabrication

The first step in microfluidic chip fabrication is to produce a PDMS cast of the device. For this, a silicon wafer with the mother machine design etched on by lithography was used. The wafer used in this project was a gift from Stephen Uphoff, with the basic design resembling that in Fig. 3.1A.

In order to make the PDMS cast, a temporary container is formed tightly around the silicon wafer using foil, with walls approximately 1 cm high. The PDMS is then made by combining elastomer and curing agent from the SYLGARD 184

Silicone Elastomer Kit (Dow: 4019862) in a 10:1 ratio, and mixing thoroughly (until opaque with bubbles) with an electric whisk. The mixture is then poured over the silicon wafer and degassed in a vacuum desiccator until the bubbles are removed. The PDMS is then cured at 65 °C for 2.5 hours. After removing from the oven, the foil can be removed and PDMS peeled off the wafer. Each mother machine is cut from the PDMS with a razor blade, and inlet/outlet holes carefully punched with a 0.75mm hole punch (Darwin Microfluidics: PT-T983-07-25). The final PDMS mother machines are prepared for bonding by removing dust with with scotch tape (3M: 7100235316), washing twice in IPA and drying with air.

High-precision coverslips (Thorlabs: CG15KH1) were prepared for bonding by sonicating in acetone for $20\,\mathrm{min}$, washing twice with deionised water, then sonicating in isopopropanol for $20\,\mathrm{min}$. A custom slide holder was designed to enable cleaning of 20 coverslips in parallel. The inner component (Fig. 3.9C, D) was designed in Solidworks and 3D-printed in nylon (for resistance to isopropanol and acetone). Due to the permeability of nylon, an aluminium outer component was manufactured to retain liquid. The design is such that the inner component can be removed from the outer casing after sonication, with gaps allowing airflow for efficient drying.

To bond mother machines to coverslips, both are treated in a Henniker HPT-100 plasma oven at 50% power, 0.3 mbar for $30\,\mathrm{s}$. Bonding is performed by transferring each coverslip treated face up to a hot plate at $80\,^{\circ}\mathrm{C}$, then placing the mother machine face down on top and applying even pressure. Bonded chips are then baked at 95 °C for one hour.

When bonding microfluidic chips, it is crucial to maintain cleanliness of the plasma

oven. An indication of poor function is if an increasingly low SCCM setting is required to achieve the desired 0.3 mbar pressure, followed by bonding failure. It was found during this project that wiping the chamber of the plasma oven with 70% IPA cleanroom wipes after bonding helps maintain good function of the machine. In addition, regular deep cleans should be carried out, which consist of scrubbing the inside chamber with soapy water, wiping out with IPA and then running the machine at 100% power for 30 mins.

Several other methods were tested for chip bonding. It was found that chips *can* be bonded with no coverslip cleaning whatsoever, which can save time but risks the appearance of unwanted dust inside the channels. Mixed results were obtained by using scotch tape only for cleaning - This can be effective for cleaning PDMS, but was found to leave behind residue when cleaning the glass coverslip (as well as adding risk of breakage when removing tape). For the protocol described above, sonication times can be reduced to no less than 10 minutes to ensure adequate removal of solvents (residue appears as grey streaks on drying).

### 3.5.3 Epoxy cast fabrication

The silicon wafers used in chip fabrication are expensive and easily broken. In order to replace the silicon wafer in the chip fabrication protocol, one can take an epoxy cast of the PDMS mother machine [160]. This can be a useful process both to reduce the risk of damage to the silicon wafer from repeat use, but also to produce "custom" casts with more duplicate designs than the original wafer, increasing manufacturing capacity.

Prior to starting the protocol, the Epotek-301 (Kremer Pigmente: 97241) Part A

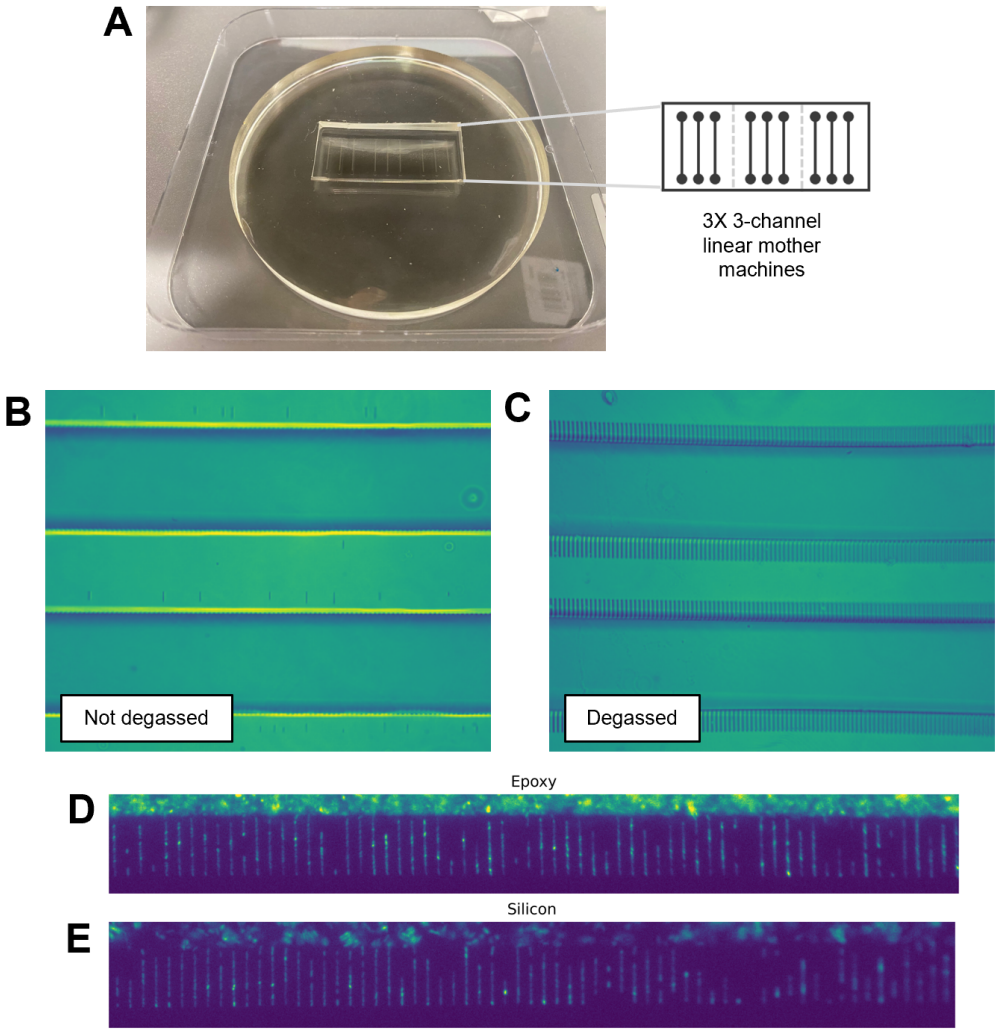



Figure 3.10: **Epoxy cast of mother machines.** **A:** Image of an epoxy cast featuring three 3-channel linear mother machines. **B:** Brightfield image of a mother machine produced from an epoxy cast without degassing the epoxy, and **C:** with degassing the epoxy in a centrifuge. Comparison of loading in **D:** a chip made with the epoxy cast, and **E:** a chip made with the original silicon wafer.

was warmed up to 65 °C for 3 hours. Then 60g heated Part A was combined with 15g Part B and mixed thoroughly. The mixture was separated into two 50 ml falcons and spun in a centrifuge at 2000 RPM for 3 mins to remove bubbles. This step is essential to enable proper formation of trenches as can be seen in 3.10B and C. Degassing can also be achieved in a vacuum, but if the vacuum is too strong it will cause the epoxy to boil, which should be avoided. To prepare the PDMS for casting, each mother machine to be replicated was placed (pattern side up) in a 4-inch diameter silicon baking mould. The mould containing the PDMS was then placed in a vacuum desiccator and degassed repeatedly for at least 10 minutes, to remove any air between the silicon and the PDMS. The epoxy mixture was then poured gently over the top, and it was left to harden for 24 hours at room temperature.

Prior to first use, the epoxy should be silanized. This makes the surface hydrophobic, which protects it and prevents PDMS from sticking during casting. This entire process should be conducted inside a fume hood due to the reagents being hazardous and volatile. Firstly, a vacuum desiccator should be placed inside the fume hood (and connected to a vacuum if possible, or sealed shut if not). The epoxy cast to be silanized should be placed inside the desiccator, and next to it, a small sheet of aluminium onto which 2-3 drops of trichloro(1H,1H,2H,2H-perfluorooctyl)silane (Sigma Aldrich: 448931) should be pipetted. If a vacuum is connected, it can be *briefly* switched on, avoiding complete aspiration of the trichloro(1H,1H,2H,2H-perfluorooctyl)silane, and then left for 1 hour. If there is no vacuum, it can simply be left in the desiccator overnight. The next day, the desiccator should be opened, and left open for several hours to overnight in the

fume hood to ensure all vapours are gone. The epoxy cast should now be silanized and ready for casting.

A prototype epoxy cast is displayed in Fig. 3.10A. A brightfield image of a mother machine produced from the cast without degassing can be seen in Fig. 3.10B, showing improper trench formation. A similar brightfield image that was produced using a cast with degassing is shown in Fig. 3.10C, where all trenches are intact. No significant difference in loading efficiency was observed between a chip manufactured with the epoxy cast (Fig. 3.10D) and with the silicon wafer (Fig. 3.10E).

## 3.6 Long term culture of MG1655 cells in the mother machine

### 3.6.1 Cell loading

Loading of cells into the trenches of the original mother machine design is achieved via centrifugation. Two different holders were designed and 3D printed to enable loading at different angles (Fig. 3.11A, B). The holders were designed to be compatible with a 96-well plate centrifuge rotor (Fig. 3.11C), such that when they spin, the chip is oriented at a specific angle (60° or 10°), which encourages cells to enter the trenches (Fig. 3.11D). The rotor at 10° (spinning at 3000 RPM for 3 minutes) produced the better results of the two with exponential phase cells (Fig. 3.11E, F). It was found that loading could be made significantly easier by loading instead with stationary phase cells (Fig. 3.11G), due to their smaller size.

Loading must then be followed by 3-4 hours growth in the mother machine prior to starting any experiment, to allow cells to re-enter exponential phase.

Common failures with loading were found to be related to properties of the strain, for instance RecA knockout strains (e.g. DH5$\alpha$, DH10B) are prone to filamentation due to their lack of DNA repair, and hence sensitivity to stress. This causes challenges with loading because highly filamented cells are too large to enter the trenches. In addition, high-motility strains present loading challenges (or have high escape rates post-loading). Novel chip designs with backports in each trench allow flow of media through trenches, and can help overcome issues with high motility cells. Alternatively, one can consider genomic knockouts of flagella genes (FliC, FliA) to reduce motility.

### 3.6.2 Long term observation of cells in the mother machine

Once loaded into the mother machine, the chip is connected to the microfluidics and positioned on the microscope stage in the incubator where cells can grow stably at 37°. Fig 3.12 displays a simple experiment, imaging a GFP-expressing population of 3000 MG1655 *E. coli* cells every 3 minutes for 7.5 hours.

Fig. 3.12A shows an image section from the mother machine, and Fig 3.12B shows how the trenches in the image can be identified using the DeLTA-based analysis pipeline. Within each trench, individual cells can also be segmented, as shown in Fig. 3.12C. From this point, it is possible to take measurements of cell length (Fig. 3.12D), which can be translated into growth rate measurements by counting events where the length sharply decreases (as indicated by grey dashed lines in the plot). It is also possible to take intensity measurements from each cell

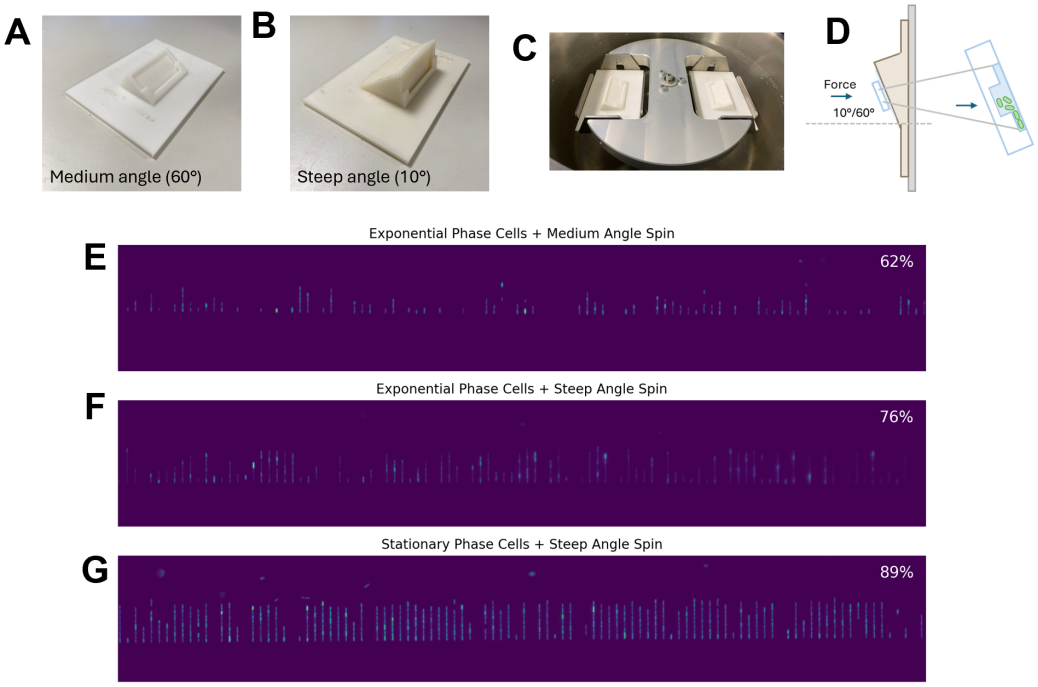



Figure 3.11: **Loading of MG1655 cells into the mother machine by centrifugation.** 3D-printed centrifuge holders with angles of **A:** 60° or **B:** 10°. **C:** Centrifuge holders in the 96-well plate centrifuge holder. **D:** Loading schematic, showing that the angle of centrifugation encourages cells to enter the trenches. Images of loading efficiency with exponential phase cells spun at an **E:** medium (60°) and **F:** steep angle (10°). **G:** Stationary phase cells spun at a steep angle (10°).

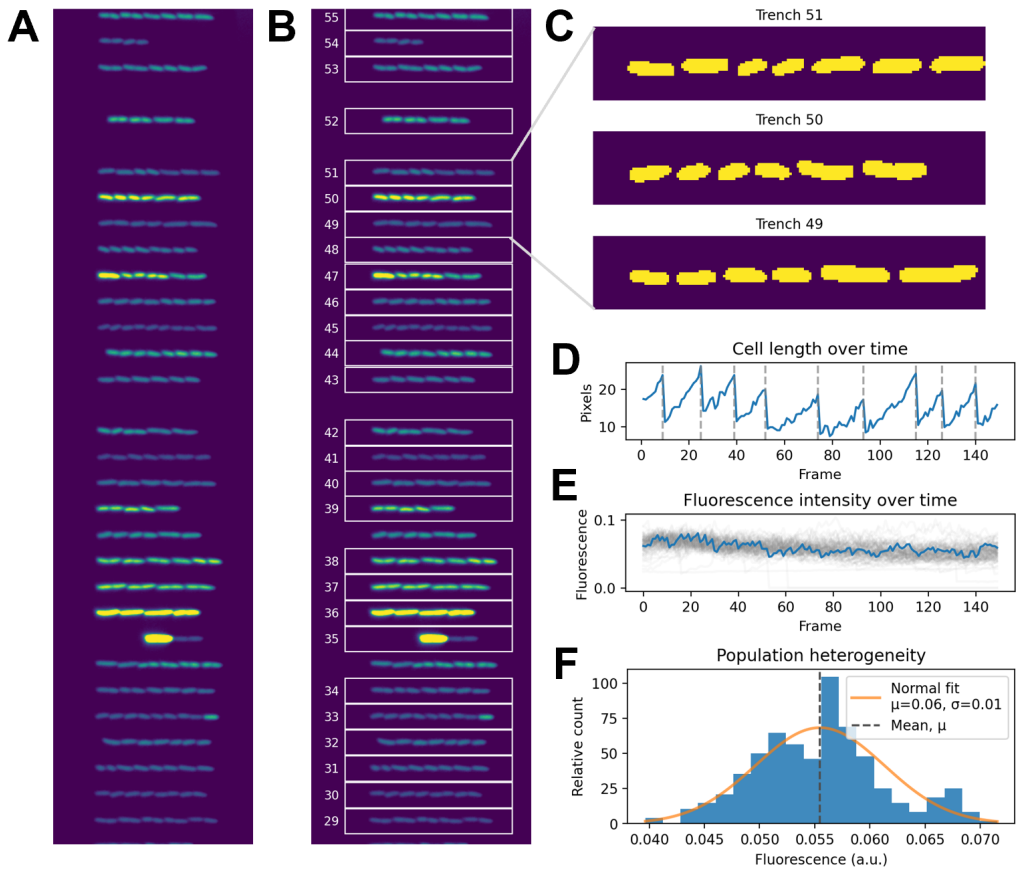



Figure 3.12: **Long term observation of single cells in the mother machine**. **A:** Image of GFP-expressing MG1655 cells. **B:** Automated identification of filled trenches using DeLTA. **C:** Automated segmentation of cells in individual trenches using DeLTA. **D:** Length measurements over time, which can be used to identify division events (dashed grey lines). **E:** Fluorescence intensity over time. Grey lines are individual cell fluorescence values, and blue line is one example cell. **F:** Population heterogeneity according to the distribution of fluorescence values across all 3000 trenches.

(Fig. 3.12E).

One of the key advantages of mother machine experiments, which is visually evident in Fig. 3.12A, is the ability to capture population heterogeneity. Where measurements in a pooled format offer a single average measurement, now we have access to the 3000 individual measurements that comprise that average, and can gain a greater understanding of the underlying composition of the population (Fig. 3.12F).

### 3.6.3 Interacting with cells in the mother machine

This platform was designed, however, not only to observe cells but also to interact with them. The most straightforward mechanism of doing so is by chemical induction, which is shown in Fig. 3.13. In this experiment, cells with IPTG-inducible GFP expression were loaded into the mother machine, cultured for a short time period without IPTG, and then switched to IPTG media to observe induction in real time. Media can be switched using an automated valve system, meaning there is little to no physical disturbance to the cells during the process. Individual frames of a single trench taken at 3 minute intervals are displayed in Fig. 3.13A, within which one can clearly see expression being induced. Using the DeLTA-based segmentation pipeline, it is possible to extract fluorescence readings, which clearly show the induction curve, either from the population as a whole (Fig. 3.13B), or on a cell-by-cell basis (Fig. 3.13C).

This system enables real-time observation of cell responses to chemical stimuli, with minimal undesired disturbance to the system. Here, images are taken every 3 minutes, but the system is capable of taking measurements on time frames

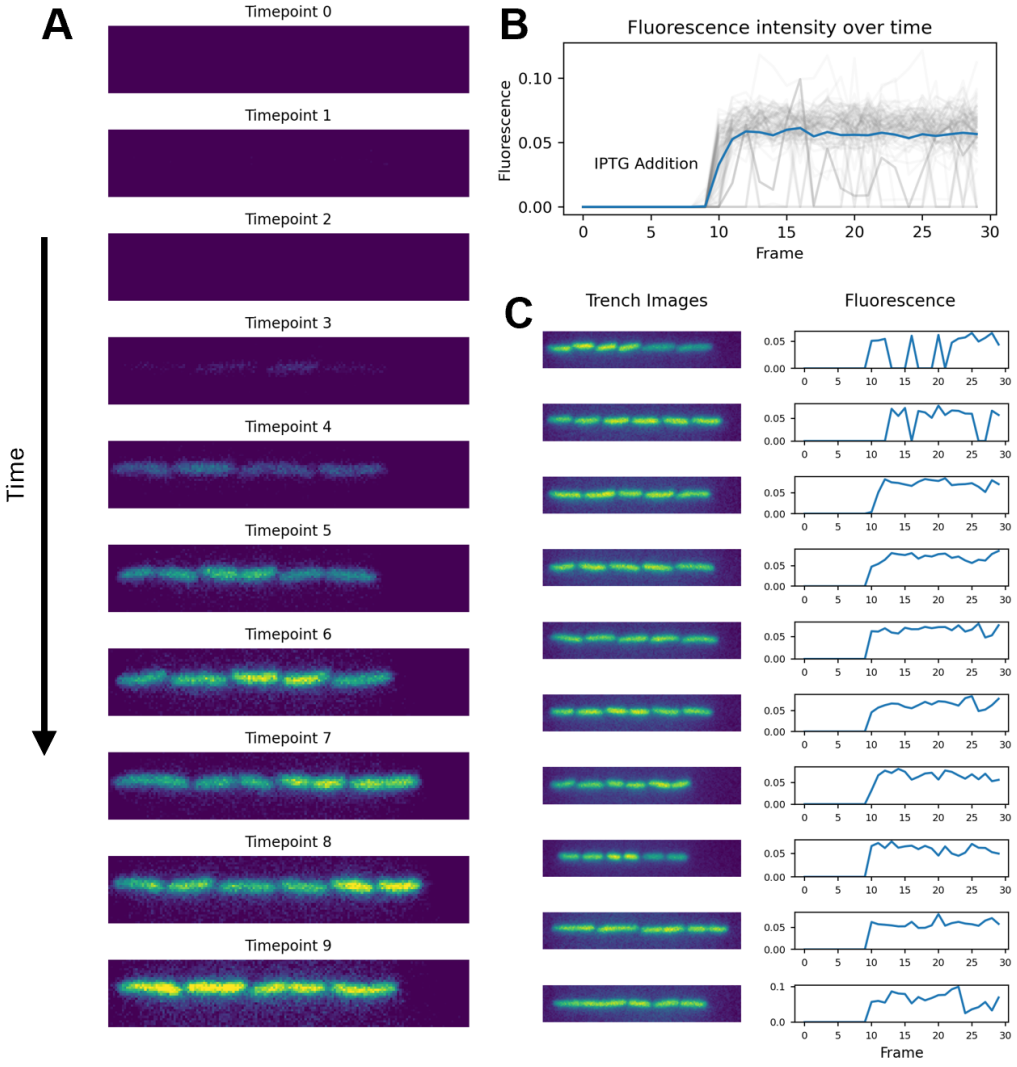



Figure 3.13: **Single-cell observation of IPTG induction in the mother machine**. **A:** Time frames of a single trench over the course of induction. **B:** Fluorescence intensity over time, with individual trench information shown in grey, and whole population average shown in blue. **C:** A random selection of individual trench IPTG response data.

anywhere from milliseconds to hundreds of cell generations.

This section demonstrates interaction with cells via chemical stimuli, however the system is also optimised to stimulate cells with targeted beams of light via the DMD. This capability is demonstrated in detail in the following chapter in the case of delivering UV, however it could also be used to deliver optogenetic signals to control gene expression, as shown in [16]. A third mode of interaction with cells is via a magnetic field. It has recently been shown, using the system presented in this chapter, that it is possible to modulate the fluorescence of a protein using a magnetic field [11]. This presents an additional mode of interacting with cells that is mother machine compatible, opening the door to even further possibilities for dynamic single-cell based studies.

## 3.7 Discussion

This chapter describes the development of a fully integrated experimental platform optimised for long-term, high-throughput single-cell studies in mother machine microfluidic devices, with the ability to interact with cells via chemical, optical or magnetic signals. Many aspects of both the hardware and software have been custom designed to maximise the versatility, throughput, and performance of the platform. As a result, the system is broadly applicable across diverse single-cell studies and has already underpinned two conceptually distinct publications [11, 56], demonstrating its wide applicability and robustness.

The throughput of the system shown in this chapter is approximately $6 \times 10^3$ trenches using the linear mother machine design from Stephen Uphoff's lab. This

could, however, be significantly increased up to $10^6$ by implementing more condensed serpentine designs [13], which would not require any additional modifications to the system (beyond the wafer used to cast the chip). The imaging demonstrated here also occurs every 3 minutes, however due to the integration of a high-speed camera that could be feasibly be decreased to sub-second intervals for a single FOV, as shown in [11].

There are several other ways in which the system capabilities could be meaningfully expanded using only modifications to the microfluidics. Firstly, with the addition of backports to the back of trenches, loading can be conducted without the need for centrifugation. By reversing the flow through backports, cells can additionally be unloaded, enabling a new iterative class of experiments (e.g. directed evolution). One further advantage of backports is improvement of the flow of nutrients to cells, as flow occurs not just past but *through* each trench.

One significant challenge with microfluidic experiments is blockages, primarily at the inlet and outlet of the chip. In our system, this has been partially rectified by utilising the high-power UV LED to kill cells in the area of risk. Alternatively, double inlet chips have been designed to specifically reduce cell accumulation at the inlet [161]. However, more sophisticated microfluidic designs exist that enable closing off of the inlet to the main channel and subsequent washing with bleach, which destroys even cells that would be inaccessible to our UV-targeting method [14]. With respect to further increasing the capabilities of this system with microfluidics alone, different designs can also enable culturing of different species, including other types of bacterial cells as well as eukaryotic cells [70, 72–80, 149].

This system is capable of collecting vast amounts of highly valuable phenotypic data, however it is currently incapable of relating that information back to the corresponding genotype. Overcoming this limitation in the mother machine context enables an unparalleled screening capability, which is to observe dynamic behaviours of genes on a single-cell level and relate those behaviours directly back to sequence. Such a setup has already been achieved by other groups using fluorescence *in situ* hybridisation (FISH) [13, 162]. In this approach, cells are chemically fixed at the end of an experiment and permeabilised to allow fluorescently labelled oligonucleotide probes to diffuse into the cytoplasm. These probes hybridise specifically to complementary nucleic acid sequences, enabling direct identification of genes or transcripts present within each individual cell. By imaging the resulting fluorescence patterns and spatially registering them with previously recorded time-lapse data, phenotypic trajectories can be retrospectively linked to genotype at single-cell resolution.

The mother machine in combination with FISH represents a powerhouse for generation of datasets for machine learning, as one can collect phenotypic behavioural data alongside genomic data for up to a million cells at a time. In the context of directed evolution, which is the ultimate focus of this thesis, these also represent powerful datasets for mapping fitness landscapes, and training models for ML-based variant design [116, 118, 119]. As cells in FISH are chemically fixed, however, they are no longer viable for downstream growth, therefore FISH is not compatible with a *continuous* directed evolution strategy (which offer several advantages including lower labour and resource requirements). The final chapter of this thesis, therefore, will present a method of selecting live cells from the

mother machine with a view to implementing a future continuous directed evolution pipeline.

To conclude this chapter, the machine that was developed during this PhD enables automated, high-throughput analysis of single cell behaviour. It has broad applications across general screening and directed evolution, and will hopefully represent a valuable technology for machine learning dataset generation.

## 3.8 Statement of authorship

The overall design of this platform was conceptualised by Harrison Steel prior to the start of my DPhil, who was also responsible for the optics design, syncronisation board and RAMM microscope design, as well as assisting in the assembly process. Jingyu Wang additionally assisted me with microscope assembly in the early assembly stages. Gabriel Abrahams and Sebastian Towers assisted in the optics alignment. I was responsible for the design and fabrication of other custom components including the incubator, stage insert, chip centrifuge holders, needle insertion support and other minor components. Idris Kempf was responsible for the main body of software supporting the system (specifically the GUI, underlying control mechanisms and DeLTA-integrated analysis pipeline), with myself writing strategy code and producing custom data analysis pipelines for specific experiments. I was responsible for setting up microfluidic capabilities in our lab, including acquiring equipment and developing protocols for chip fabrication, epoxy cast fabrication, cell loading, and general experimental procedure. Kirill Sechkar and I worked together to establish the initial microfluidic control set up, and Kirill was responsible for the custom microfluidic control software that is now in place.

Other members of the lab have since made valuable contributions to chip fabrication and experimental protocols as our collective experience has grown, including Idris Kempf, Agrima Deedwania, Vicente Trelles Fernandez and Gabriel Abrahams. Meanwhile, I have spent significant time training many people in the use of this platform which has gone on to benefit their own projects. The initial process of chip fabrication development was additionally aided by members of Stephen Uphoff lab, most notably Chloe Cassaro, Valentine Lagage and Fiona Sargison, who shared their expertise and materials. Experiments shown in this chapter were executed collaboratively by myself alongside Idris Kempf, and I am responsible for the subsequent analysis.

# 4 | Application: Photoselection of cells to enable directed evolution of dynamic phenotypes

## 4.1 Featured publications

This chapter primarily consists of work from this co-first author paper:

- Microscopic PhotoSelection (MiPS) of single cells in mother machine microfluidic devices [56]

## 4.2 Abstract

Techniques for selecting or sorting single cells within large populations of genetic variants are central to synthetic biology and biotechnology. Widely-used methods such as Fluorescence Activated Cell Sorting (FACS) enable rapid processing of large libraries, but are restricted to low-dimensional measurements taken at a single time point. As a result, sorting based on dynamic or multi-trait phenotypes - such as transient properties and properties that occur on fast timescales or in response to dynamic actuating signals - remains fundamentally challenging. This chapter introduces Microscopic PhotoSelection (MiPS), which employs the platform described in Chapter 3 for single-cell selection based on multiple dynamic criteria, directly on microfluidic mother machine devices. The system enables fully automated selection using high-intensity UV light or alternative wave-

lengths with the addition of photosensitisers. By targeting many cells in parallel, MiPS overcomes throughput limitations of existing microfluidic-based selection technologies such as optical tweezers or droplet-based methods and provides a direct approach to select cells based on time-resolved, multi-trait phenotypes. This chapter demonstrates the ability to perform *in vivo* selection and outlines how iterative, feedback-based selection strategies can refine selective enrichment across multiple rounds. Taken together, this work establishes a high-throughput selection framework integrated into microfluidic devices, enabling applications in directed evolution - the focus of this thesis - as well as biosensor optimisation, circuit engineering, and diagnostics, where selection based on dynamic, multi-trait phenotypes is essential.

## 4.3 Background

High-throughput analysis or enrichment of libraries, in the form of "screening", underpins contemporary biological engineering. Screening capabilities have been dramatically enhanced by single-cell selection technologies, enabling a shift from traditional arrayed screening - where individual variants are isolated in multiwell plates - to pooled screening approaches, in which many millions of variants can be analysed from a single tube. This shift has transformed multiple areas of biology [163]. For example, the vast increases in throughput have been applied for selection-based directed evolution [9] (contrasting growth-coupled directed evolution [164]), which relies on iterative rounds of selection and mutagenesis to alter the properties of a protein.

One of the most widely used high-throughput selection methods is Fluorescence-

Activated Cell Sorting (FACS). In FACS, single cells pass at high speed - up to 100,000 per second [165] - through a laser that excites fluorescent molecules and measures a per-cell intensity. Cells are subsequently sorted into separate collections depending on their fluorescence level. An alternative method, Magnetic-Activated Cell Sorting (MACS), allows interrogation of *external* markers [46]. In MACS, magnetic beads coated with a target-specific binder (e.g. antibody, lectin) are mixed with the sample, where they bind only to cells expressing the target on their surface. Selection is subsequently performed by passing the sample through a magnetised column. Both FACS and MACS are capable of screening individual cells at high throughput, however they are restricted to static phenotypes (as they only measure each cell once), and selecting based on multiple and/or dynamic traits remains a major challenge [166, 167]. Where FACS is often applied for intracellular phenotypes, and MACS for surface-based phenotypes, *secreted* phenotypes can be selected upon using droplet-based microfluidics [53]. In droplet-based microfluidics, cells are encapsulated in small droplets that act as microreactors. These droplets can also be combined with various substrates, making this system a suitable tool for the directed evolution of secreted enzymes [45]. Examples exist of droplet based-microfluidics being used to measure dynamic responses by immobilising drops [168], however, at significantly lower throughput. Long-term observation of cells in droplets presents a challenge due to the finite amount of contained nutrients, as well the complexity of tracking of single cells in three dimensions.

Given the rich phenotypic data that can be collected from mother machine studies (as shown in Chapter 3), physically selecting variants is of significant interest

for unlocking downstream applications. It has been shown possible to link phenotypic to genotypic data in the mother machine using dCas9 libraries [162] as well fluorescence *in situ* hybridisation (FISH) [169]. These methods are effective for screening, however they require genetic modifications, extensive sequencing, and in the case of FISH, the cells themselves are destroyed in the process. One approach to extracting cells without the need for genetic modification or destruction is with the use of optical tweezers [14]. This method is able to select a small number of cells from the mother machine, however the process requires a long time period per cell and has only been achieved at low-throughput (on the order of tens of cells), making it incompatible with high-throughput sorting. It was later shown possible to use strategies such as laser-ablation and optogenetically-induced antibiotic resistance to select cells, paving the way for integration with future high-throughput pipelines [15].

This thesis introduces Microscopic PhotoSelection (MiPS) as a method for fully automated, high-throughput selection of cells from mother machine devices using targeted UV light, as well as using targeted red light in combination with the photosensitiser methylene blue. These methods both utilise the system Digital Micromirror Device (DMD) to project targeted, patterned light onto cells at $< 0.5\,\mu\mathrm{m}$ resolution, resulting in selected cell death or growth inhibition. Connected to high-powered LEDs, it is shown possible to use the DMD to perform selection on populations in mother machines with $> 10^4$ trenches, with the potential to expand throughput to the limits of microfluidic device sizes (up to $10^6$ [13]). When compared to FACS - the method currently used for similar challenges - this method unlocks the ability to perform selection based on the rich, multi-factorial

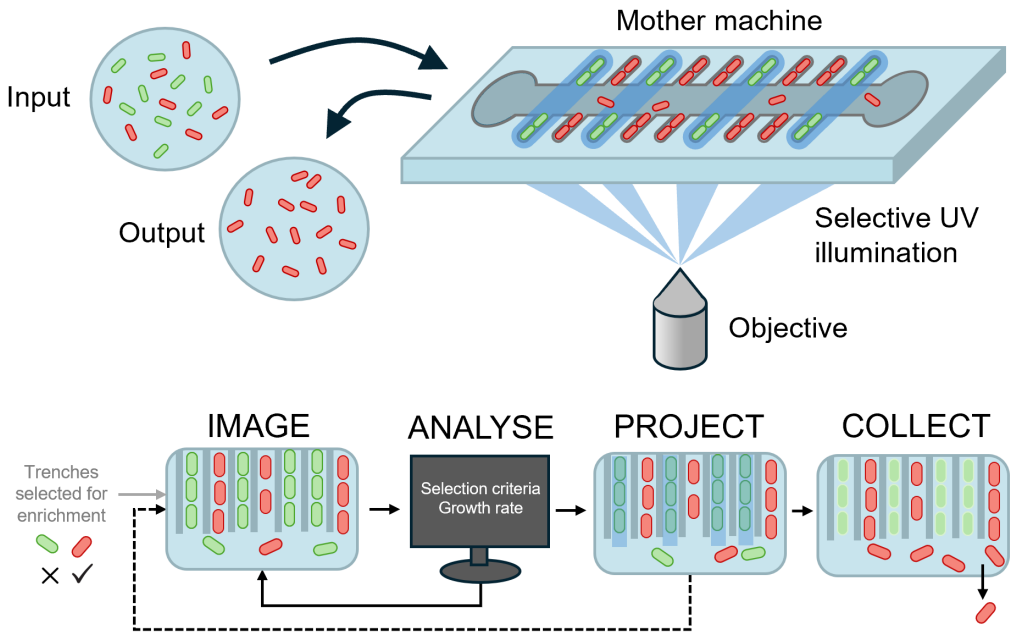



Figure 4.1: **Schematic of the MiPS pipeline.** Trenches are imaged and analysed to assess selection criteria, as well as growth rate. After a sufficient data has been collected, trenches to be destroyed are targeted with high-power UV. The untargeted cells continue to proliferate and can be collected from the output of the microfluidics.

data available from mother machine experiments, including dynamic responses [170], tracking of sub-cellular localisation [171], or responses to whole arrays of chemical, magnetic and/or optogenetic stimuli.

## 4.4 Results

### 4.4.1 Characterisation of growth inhibition with light

To optimise cell selection protocols, patterned light dosing was applied to trenches at different intensities subsequent reductions in cell viability were measured. In all experiments cell trenches in the mother machine were located and segmented using a modified version of the DeLTA pipeline [155] (Fig. 4.2A). The coordinates for each identified trench were used to generate a DMD pattern that projects light onto targeted trenches. Figure 4.2B shows a scenario in which UV was projected onto alternating trenches (8.0 $\mathrm{Wcm}^{-2}$ intensity and $300\,\mathrm{s}$ duration), subsequently reducing the division rate of targeted cells to on average $\sim 10\,\%$ that of untargeted cells.

With the goal of finding an optimal trade off between UV dosage and timing, an array of UV intensities and durations was tested (Fig. 4.3). The highest performing selection of conditions were found to be of $300\,\mathrm{s}$ duration (right hand column). Cells with a division rate of less than 0.4 per hour were deemed to be killed. After targeting with UV, populations generally exhibited bimodal distributions with respect to division rate (one non-growing population and one with similar-to-initial growth population, and few intermediate rates of growth). This is unsurprising in the light of prior investigations *E. coli*'s response to oxidative

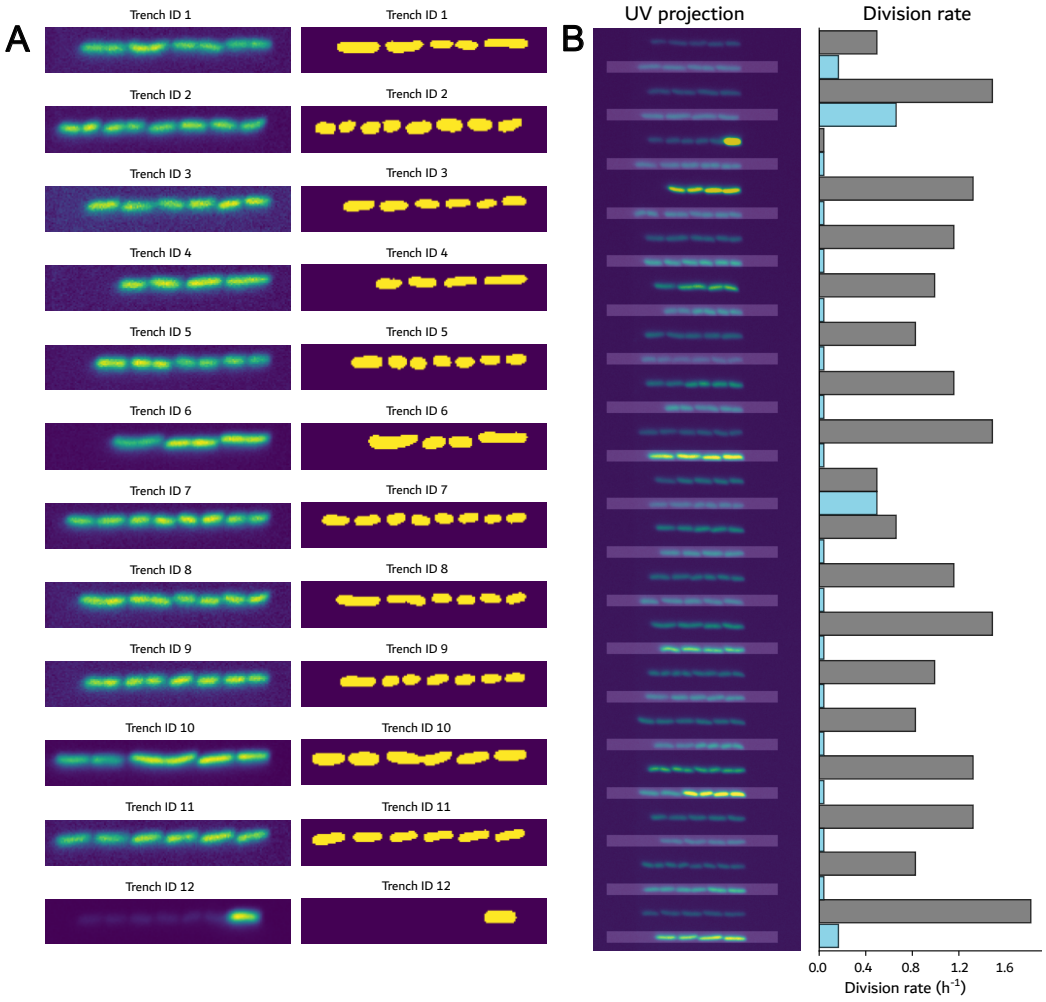



Figure 4.2: **Photoselection with UV A:** Segmentation of cells in the mother machine by DeLTA, allowing for identification and localisation of cells for UV projection. **B:** Targeting of alternating trenches with UV light projection $\approx 8\,\mathrm{W\,cm^{-2}}$ ($90\,\%$ maximum deliverable power) for $300\,\mathrm{s}$ (left), and subsequent division rates for each cell (right). Targeted cells division rates highlighted in blue, untargeted in grey.

stress [172, 173], which is a key mechanism of cell killing by UV light [174]. Expression of stress response genes in bacterial populations is highly heterogeneous [172], while growth rates stay consistent for a wide range of antioxidant defence enzyme activities before sharply falling for very low levels [173].

As the proportion of targeted cells increases, off-target effects also lead to an increase in the proportion of untargeted cells that are killed; this ratio of on:off target killing is displayed in Fig. 4.5A. The optimal dose in the tested conditions was 8.0 $\mathrm{Wcm}^{-2}$ UV $300\,\mathrm{s}$, which resulted in a 2.78$\times$ increase in on:off target ratio with respect to cells killed.

It was then explored whether the effectiveness of cell killing could be further improved with the use of a photosensitiser chemical. Methylene blue (MB) has previously been shown to cause production of singlet oxygen species in the presence of red light, which can kill *E. coli* in sufficiently high concentrations [175]. In order to test the efficiency of methylene blue, cells in the mother machine were treated with M9 supplemented with $2\,\mu\mathrm{M}$ MB for $1\,\mathrm{h}$, followed by projection with red light ($645\,\mathrm{nm}$) following the same protocol as used in the UV dosage experiments (Fig. 4.4). It was found that cells were generally far more sensitive to this protocol than UV.

The resultant differences in on:off target killing are displayed in Fig. 4.5B. By utilising the photosensitiser MB, it was possible to achieve up to 5.13$\times$ on:off target killing, a 1.8$\times$ improvement over UV, although the optimum for UV appearing on the boundary of achievable conditions suggests higher UV dosages may lead to improved performance. This difference is likely due to the fact that MB has a steeper dose response curve (Fig. 4.6), and thus greater differences between on

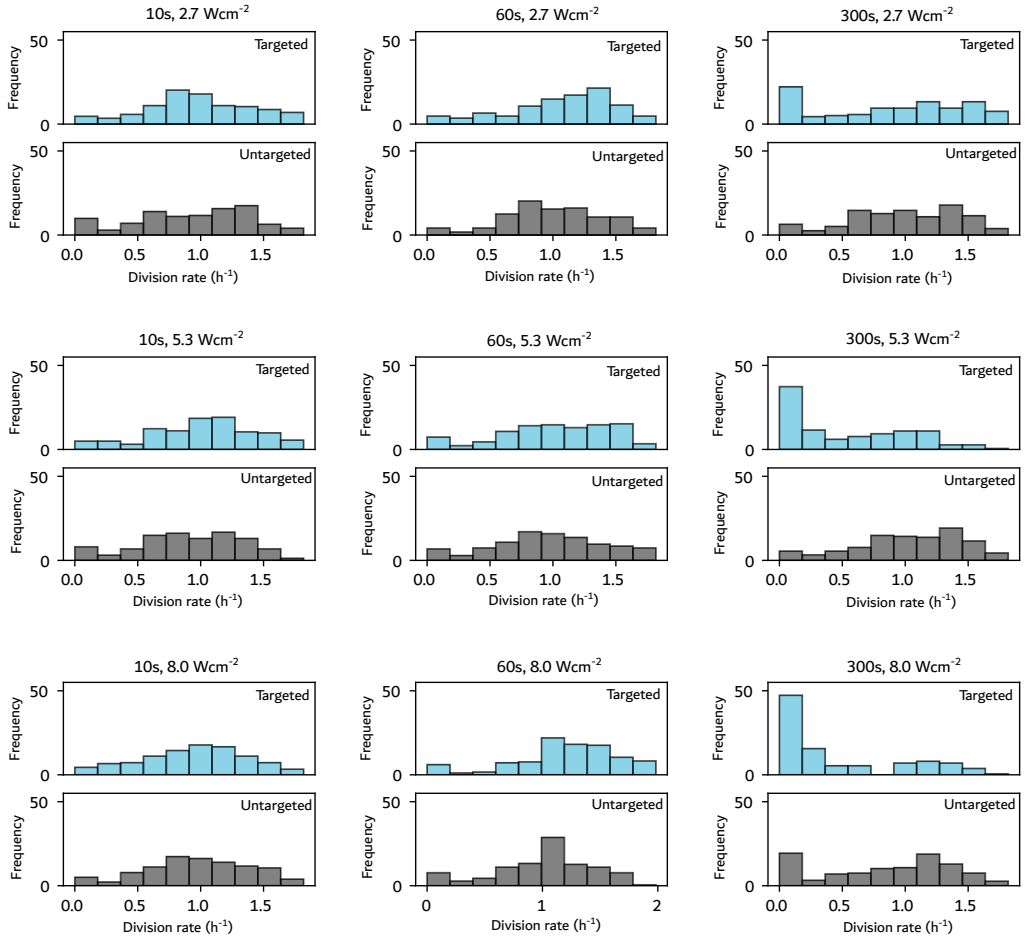



Figure 4.3: **UV dosage testing**. Alternating trenches in the mother machine were targeted with UV light (385 nm) at the intensities and durations above. The subsequent division rates of targeted (blue) and untargeted (grey) cells are compared.

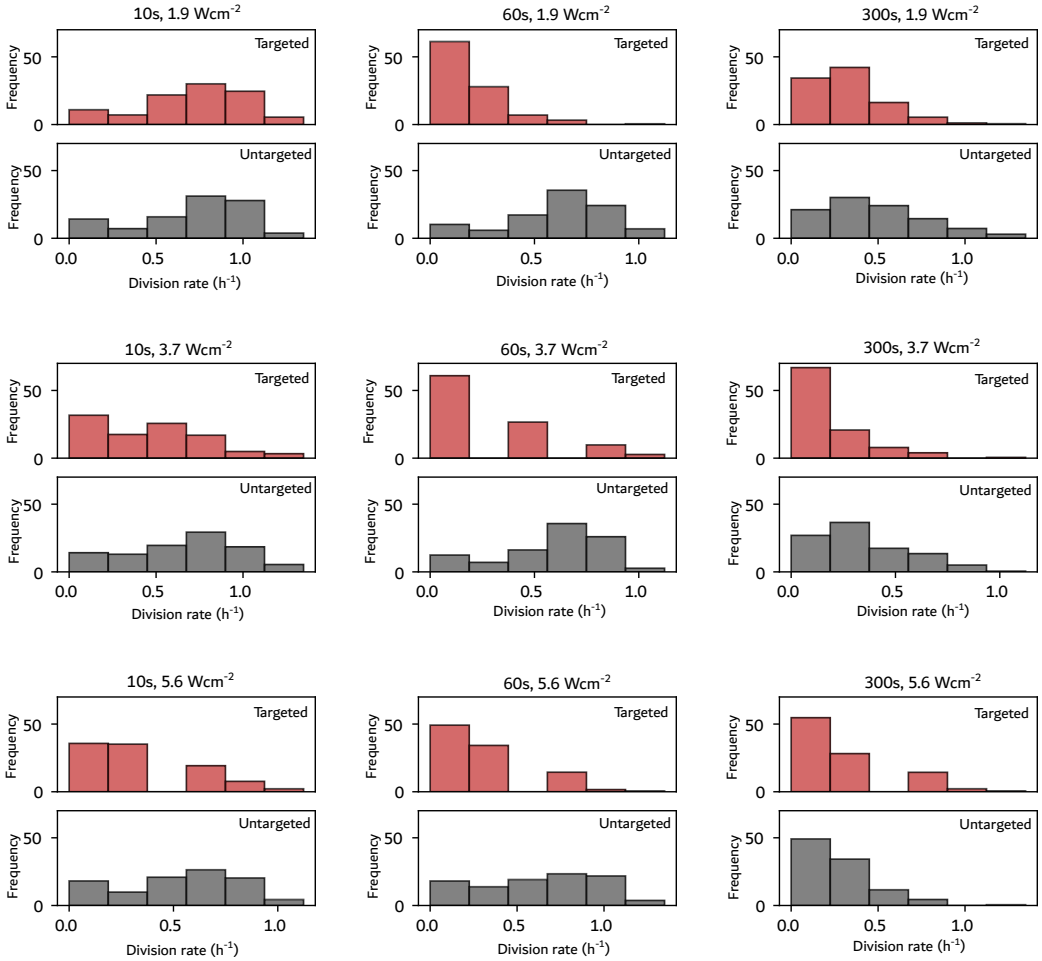



Figure 4.4: **Red light and methylene blue dosage testing**. Cells were first treated with 2$\mu$M methylene blue for 1 hour, then alternating trenches in the mother machine were targeted with red light (645 nm) at the intensities and durations above. The subsequent division rates of targeted (blue) and untargeted (grey) cells are compared. 90% intensity $\approx$ 5.6 W/cm$^2$

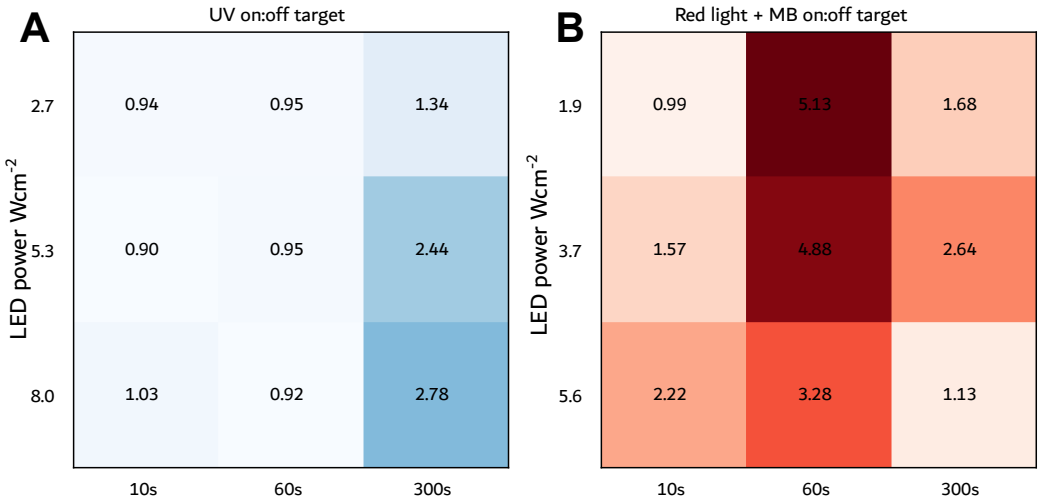



Figure 4.5: **Relative on:off target killing** using **A:** 385 nm UV, and **B:** MB ($2\,\mu$M) with 645 nm red light. Higher values indicate a better on:off target ratio.

and off-target effects can be achieved at a given pair of on and off target dosages. In addition, the optimal settings at which this is achieved were far lower, requiring only $60\,\mathrm{s}$ and 1.9 $\mathrm{Wcm}^{-2}$ LED intensity, a $5\times$ reduction in duration relative to the UV optimum. However, it was also observed that addition of MB to the cells causes a growth rate reduction of $\approx 40\,\%$ even in cells untreated with red light, and therefore this apparent sensitivity could lead to further issues if subsequent rounds of growth and selection are intended.

### 4.4.2 Quantification of off-target effects

From these experiments it is clear that off-target effects cause unwanted death of untargeted cells. Therefore, several experiments were conducted to gain a greater understanding of these effects. First, to understand whether these effects are primarily local (i.e., limited by contrast and accuracy of the projected DMD image) or global (i.e., device-wide scattering or delivery of off-target light), an exper-

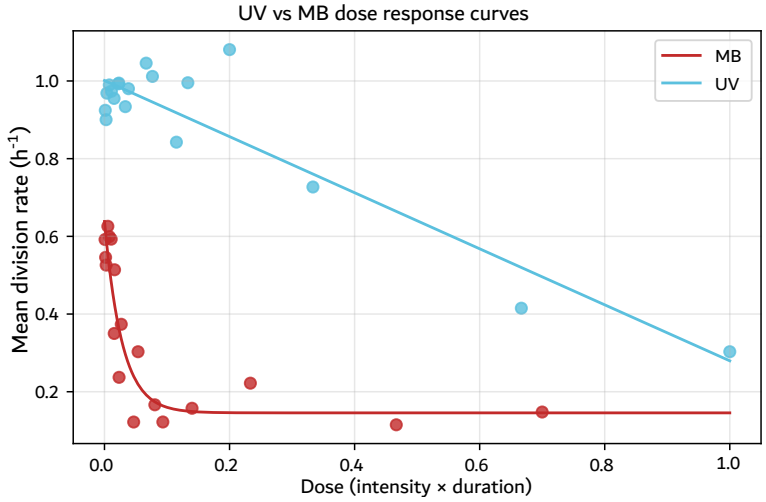



Figure 4.6: **A comparison of dose response curves between UV and red light with MB**. For MB, cells were first treated with 2$\mu$M methylene blue for 1 hour, then alternating trenches in the mother machine were targeted with red light (645 nm). No prior treatment was required for the UV case, in which alternating trenches were targeted with UV light (385 nm). Dosages have been adjusted for the difference in LED power (90% 385 nm $\approx$ 8 W/cm$^2$, whereas 90% 645 nm $\approx$ 5.6 W/cm$^2$).

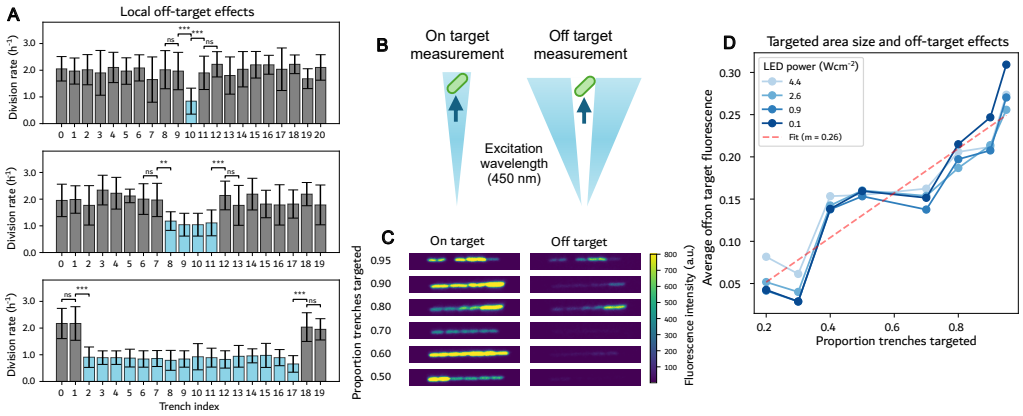



Figure 4.7: **Characterisation of off-target effects. A:** Experiment to explore local off-target effects, in which an increasingly large strip of UV light was projected down the centre of a sample, and the division rates of the immediate neighbours measured. Blue = targeted. **B:** Schematic of an experiment to explore global off-target effects using the excitation wavelength of GFP. An on-target measurement is from illumination of that cell only. An off-target measurement is illumination of an area away from that cell, but subsequent measurement of any excitation that is a result of off-target light. **C:** Demonstration that as the proportion of trenches illuminated increases, the off-target illumination also increases. **D:** Quantification of the relationship between targeted area size and off-target light. Gradient (m) w.r.t proportion cell trenches targeted = 0.26.

iment was conducted in which a strip of UV was projected on one or multiple neighbouring trenches, and the division rate of the immediate neighbours assessed (Fig. 4.7A). It was found that there was no significant difference in division rate between a cell one position away from the targeted trenches versus two positions away (or more), suggesting that off-target effects are *not* primarily caused by local effects arising from blurring of DMD light delivery, and instead occur at a larger or global scale. This provides validation for the design of the DMD and optical system, namely that sufficient local (i.e. pixel to pixel) contrast is achieved in image projection to make this factor insignificant compared to other sources of stray light (e.g. reflection and scattering).

A second experiment to quantify the amount of stray light was conducted using selective targeting of fluorescence excitation light ($450\,\mathrm{nm}$) to GFP expressing cells, allowing a direct measurement (through fluorescence) of light delivery to each trench (i.e., rather than one mediated by cellular sensitivity to UV measured through growth rate). The assay consists of pairing on-target and off-target measurements of the same trench; in this case an on-target measurement comes from illumination of the target trench (and nothing else), while an off-target measurement comes from illumination of a known area elsewhere in the FOV, and then measuring fluorescence of the target trench arising from excitation by this off-target light (Fig. 4.7B). The off:on target ratio is then computed by comparing the fluorescence of each trench by off-target light to its on-target fluorescence measurements. A sample of such comparisons is displayed in Fig. 4.7C. This procedure was conducted across 9 different FOVs, in which an increasing proportion of trenches was illuminated in the off-target tests. 20 individual trenches

were measured in each of the FOVs, and the results are shown in Fig. 4.7D. The ratio of off:on target effects increases approximately linearly with respect to the area of light delivered, and are constant with respect to LED intensity. The exact off:on target ratio expected can be calculated using the proportion of cell trenches targeted, using scale factor 0.26 (the gradient of the linear fit in Fig. 4.7D). For instance, if $99\,\%$ trenches are targeted, then the untargeted $1\,\%$ will receive the 99 $\times$ 0.26 $\approx 25.7\,\%$ of the on-target dose.

### 4.4.3 Implementation of enrichment by photoselection

In order to test whether the process of selective UV illumination can effectively lead to downstream enrichment of a desired phenotype, an experiment was set up to test whether mCherry cells could be selectively enriched from a mixed mCherry/GFP population. mCherry cells were chosen to to be the selected population on account of being the slower growing of the two (GFP cells were measured to grow 11% faster and thus would become naturally enriched by default). Immediately before loading into the mother machine, cells were combined in an approximate 1:1 ratio. After loading, cells were allowed to grow for several hours in order to achieve homogeneous trenches (i.e. each trench either contains GFP- *or* mCherry-expressing cells). The mother machine device was then imaged both at $450\,\mathrm{nm}$ (for GFP) and $565\,\mathrm{nm}$ (for mCherry) excitation, with emission filters complementary to the respective fluorescent proteins. A composite image was used to perform segmentation with DeLTA, and the resulting masks for each trench were compared in both imaging wavelengths to classify trenches as GFP or mCherry (Fig. 4.8A). The resulting classification was used to generate a projected image

specifically targeting GFP (or empty and mixed) trenches (Fig. 4.8B) for killing. UV was projected in this image at 5.3 $\mathrm{Wcm}^{-2}$ intensity for $5\,\mathrm{min}$. The resulting mother cell length traces in Fig. 4.8C show targeted growth inhibition of GFP cells, indicated by a flat line after the projection at time point 40.

Following validation that growth could be selectively inhibited on the mother machine device, the population flowing out of the device was analysed. In order to minimise contamination from unselected cells, all microfluidic tubing (post-filter) was replaced with fresh, sterile tubing. In addition, inlet and outlet holes of the chip were projected with maximum intensity UV to destroy any cells accumulating there. The population flowing out of the device was collected for a period of $30\,\mathrm{min}$ before being centrifuged to concentrate cells and plated on agar. Figure 4.8D displays the resulting colony counts from $48\,\mathrm{h}$ post-experiment, and shows a shift in population composition from 1:1 mCherry:GFP, to 3.8:1 after a single round of targeted selection. In an ideal system *no* targeted cells would survive selection (in which case this ratio would be much larger). However, as illustrated by our characterisation experiments above this outcome is limited by the off-target delivery – if the UV dose is increased further, it will increasingly kill the desired (non targeted) cells and hence not necessarily improve the enrichment ratio.

### 4.4.4 Refinement of photoselection outcomes with growth-based feedback

The previous sections illustrate the limitations imposed by off-target effects; below demonstrates using simulation how these *could* be overcome by implementing

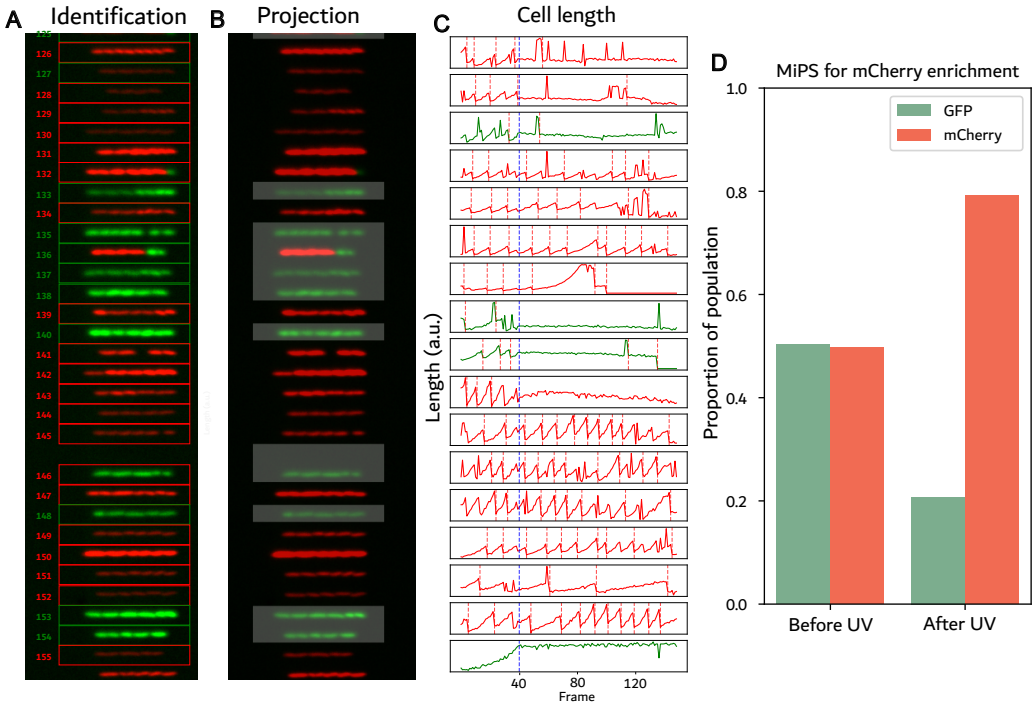



Figure 4.8: **Enrichment of mCherry cells from a mixed mCherry/GFP population by photoselection**. **A:** Identification of GFP and mCherry cells was carried out by imaging with a 450 nm LED (GFP) and 565 nm LED (mCherry), segmenting the images with DeLTA and comparing the brightness in each channel. **B:** The resulting classifications were converted into a projected image. Unidentified or mixed trenches were projected on by default. **C:** Length of the mother cell in each trench can be used to measure divisions (at which point the cell length sharply decreases, as indicated by dashed red line). When UV projection occurs at time point 40 (dashed blue line), GFP cell length becomes static, indicating no more division, while mCherry cells continue to divide. **D:** Comparison of the population composition within the mother machine device (pre-enrichment), to the population composition flowing out of the device after UV (post-enrichment).

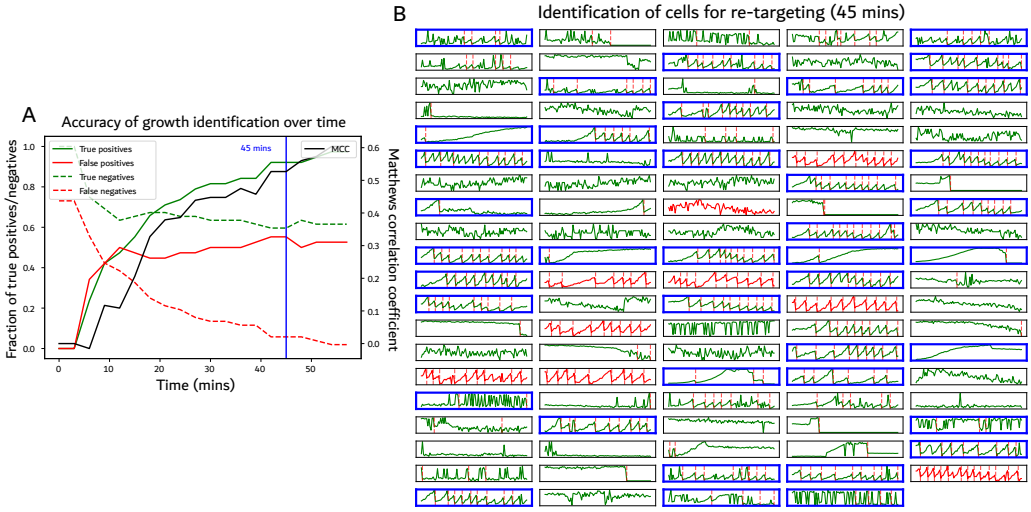



Figure 4.9: **Identification of cells for re-targeting**. Cells are identified as growing if the cell length $L$ is on average increasing: $L_{t+1} > L_t$ in over 50% time points. **A:** Accuracy of identification over time. MCC = Matthews Correlation Coefficient. **B:** Cells identified for re-targeting after $45\,\mathrm{min}$ are shown with a blue border. Green = GFP (targeted population), red = mCherry (untargeted population).

multiple rounds of selection to refine the output, i.e. the final ratio of targeted to untargeted cells. Between each round, the division rate of remaining cells can be assessed. Within $45\,\mathrm{min}$, $97\,\%$ of surviving targeted cells can be correctly identified (Fig. 4.9). Fig. 4.9A displays the true positive/negative rate of growth identification over time, and Fig. 4.9B shows the individual traces from which this identification is performed in which targeted (GFP) cells identified as survivors are highlighted with a blue outline. Classification into growing or not is done by computing whether the length of the cell is on average increasing (i.e. $L_{t+1} > L_t$ in over 50% time points). For subsequent rounds, therefore, only those targeted cells that were not killed need be re-targeted, hence reducing off-target impacts on untargeted cells.

Fig. 4.10 demonstrates how this information can be used to refine the final output. Using growth information, a feedback-based approach succeeds due to fewer cells being targeted in each successive round (e.g. only the survivors), meaning *total* UV dose is reduced each round, preventing off-target effects from killing the desired population (Fig. 4.10A).

To enable simulations of an iterative selection process, the UV killing process was modelled using a stochastic framework. The probability of a cell surviving a UV dose is a function of the dose received (a constant reflecting on target and off target contributions), and the inherent tolerance of a cell to UV (assumed to be normally distributed). To combine these variables, we introduce a metric termed "survival potential", which combines dose and tolerance, and represents the population in such a way that the proportion of the population that is below 0 in survival potential is dead following treatment, and that which is above 0 is alive.

Known survival probabilities from the dosage experiments in Fig. 4.3 were used to fit a sigmoid UV dose-response curve (Fig. 4.10B), which can be translated into our survival potential representation for any given dose. Representations of survival potential for three successive selection rounds are shown in Fig. 4.10C, visualising the output of the simulation for a per-round UV dose of $3\,\mathrm{min}$ at 5.3 $\mathrm{Wcm}^{-2}$ UV and an initial population ratio of 10:1 targeted (green) to untargeted (red) cells. The ratio of alive to dead cells is visualised using the normally-distributed survival potential, where the area over the positive domain corresponds to the fraction of alive cells. Given the high proportion of targeted cells, the ratio of off-target to on-target light at the beginning is high, at >$10\,\%$ (calculated from Fig. 4.7D). With each round, as the proportion of targeted cells remaining

decreases, the amount of off-target light received by untargeted cells becomes smaller, but the targeted population continues to receive *approximately* the same, high dose of UV. This is why the populations appear to move apart with subsequent rounds. After three selection rounds, the fraction of targeted cells remaining in the population is $5.8\,\%$ (a total enrichment of $\sim 170$-fold compared to the initial 10:1 ratio), with $0.5\,\%$ and $81.9\,\%$ of targeted and untargeted cells still being alive. The fraction of remaining untargeted cells could be increased by decreasing the per-round UV dose, at the expense of running additional selection rounds to meet the same output refinement.

This trade-off is further explored in Figs. 4.10D–E for a starting population of 10:1. With 8.8 $\mathrm{Wcm}^{-2}$ per-round UV intensity (the maximum deliverable by our system), one can achieve almost $100\,\%$ untargeted cells after two rounds (Fig. 4.10D), however over $20\,\%$ untargeted cells are lost in the process (Fig. 4.10E). In contrast, with 2.2 $\mathrm{Wcm}^{-2}$ per-round UV intensity, it takes 8 rounds to reach the same population composition, but results in only $14\,\%$ loss of untargeted cells. Therefore, one should consider whether their priority is minimising loss of cells desired for selection, or time ($45\,\mathrm{min}$ between rounds for reliably evaluating growth, up to $2\,\mathrm{h}$ per round for UV projection).

## 4.5 Methods

### 4.5.1 Media and Strains

All strains used in this work were MG1655, transformed with pVS-01-01-mCherry (a constitutive mCherry expressing plasmid) or pVS-01-03-eGFP (a constitutive

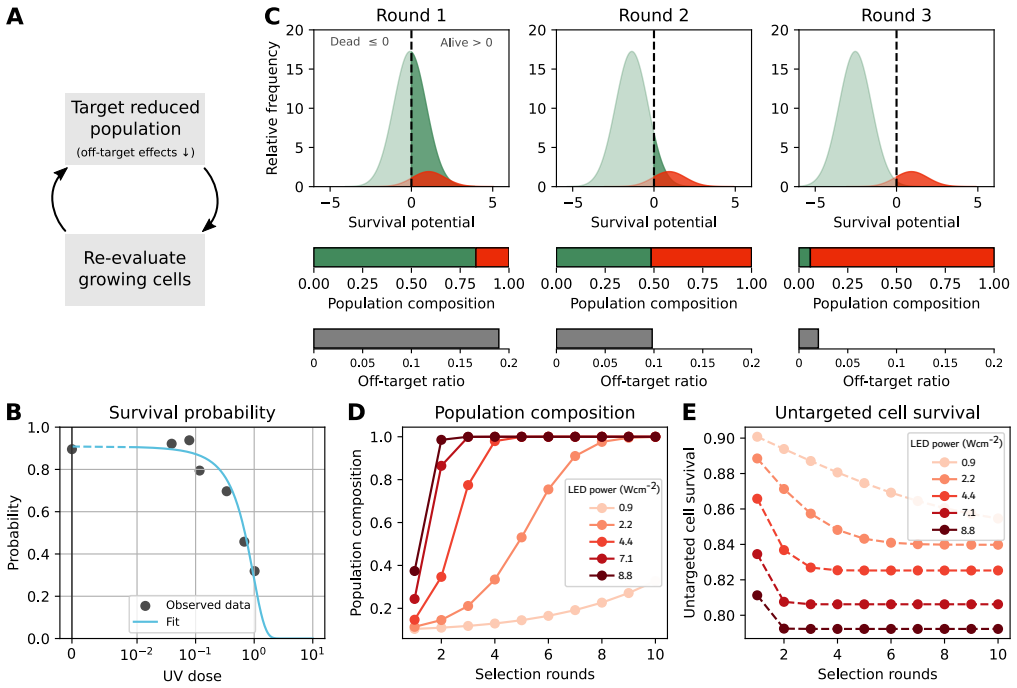



Figure 4.10: **Simulating multiple rounds of targeted selection**. **A:** Illustrating the two steps of each selection cycle, noting that off-target effects decrease with each round. **B:** Survival probability fit from UV dosage experiments. One unit of UV dose corresponds to $\approx 8\,\mathrm{W\,cm^{-2}}$ for $300\,\mathrm{s}$. **C:** Visualising three rounds of simulated selection, including population composition and off-target ratio results from each round. **D:** Population composition and **E:** untargeted cell survival over 10 rounds with different UV intensities.

eGFP expressing plasmid). All M9 media (Formedium: MMS0102) used in this work was supplemented with 50 $\mu$g/ml Carbenicillin (Formedium: CAR0005), 0.4% glucose (Formedium: GLU03), 2 mM $MgSO_4$ (Sigma Aldrich: M7506-500G), 0.1 mM $CaCl_2$ (Sigma Aldrich: C5670-500G), 0.2% casamino acids (Formedium: CAS03), 0.2g/l Pluronic F-127 (Sigma Aldrich: P2443-250G), 50 mg/l EDTA $Na_2.2H_2O$ (Sigma Aldrich: E4884-100G), 30.8 $\mu$M $FeCl_3$ (Sigma Aldrich: 157740-100G), 6.16 $\mu$M $ZnCl_2$ (Sigma Aldrich: 208086-100G), 0.76 $\mu$M $CuCl_2.2H_2O$ (Sigma Aldrich: 307483-500G), 0.42 $\mu$M $CoCl_2$ (Sigma Aldrich: 232696-100G), 1.62 $\mu$M $H_3BO_3$ (Sigma Aldrich: B0394-500G) and 0.081 $\mu$M $MnCl_2.4H_2O$ (Sigma Aldrich: 221279-500G).

### 4.5.2 Microfluidic Experiments

Overnight cell cultures were prepared in 5 ml M9 media. The following day, mother machines were injected with 15$\mu$l of 8.5 mg/ml Pluronic F-127 (Sigma Aldrich: P2443-250G), using a gel loading tip (Fisher: 11927734). After approximately $10\,\mathrm{min}$, the overnight cell cultures were injected into the chip in the same manner. In order to load cells into trenches, the chip was spun in a centrifuge at 3000 RPM for $3\,\mathrm{min}$ on each side, at an angle of $10^{\circ}$. This was achieved using a 3D-printed chip holder designed to be compatible with a 96-well plate centrifuge rotor.

The loaded chip was then mounted on a custom-designed microscope stage insert that supports the cover slide around all edges. To insert needles for microfluidics, a custom support was designed that provides a backing for the chip to prevent cracking. Once connected to the microfluidic pressure flow control system, M9 media

with 25% of the standard antibiotic concentration (12.5 $\mu$g/ml Carbenicillin) was flowed continuously at a rate of $20\,\mu\mathrm{L\,min^{-1}}$. The pressure is controlled by an OB-1 MK3+ Flow Controller (Elve Flow). The OB-1 delivers pressurised air into a 50 ml reservoir containing M9 media. As a result, media is pushed out into the system, which consists of an MFS Microfluidic Thermal Flow Sensor (Elve Flow) which provides flow measurements for feedback control, followed by a 5 cm 100 $\mu$m inner diameter resistance element (Darwin Microfluidics: LVF-KTU-05) for more robust flow control. After the resistance element, a 0.2 $\mu$m filter is attached to prevent contamination reaching the chip or output. From the filter, the tubing connects directly into the chip via glass capillary tubes (DWK Life Sciences: 764500-0002). Cells were maintained at $37\,^{\circ}\mathrm{C}$ by a custom, blacked-out microscope incubator, and allowed to grow for several hours before initialising the experiment.

## 4.6 Discussion

This chapter demonstrates Microfluidic PhotoSelection (MiPS) - a method for applying selection to microfluidic chips that leverages real-time data processing and actuation afforded by high capability robotics and real-time image processing algorithms. It was shown possible with a DMD to produce sufficient resolution in a projected image to selectively kill (or inhibit growth of) single cells with light. This capability unlocks selection based on dynamic phenotypes in a mother machine context - features fundamentally inaccessible to state-of-the-art selection methods such as FACS or MACS.

The current implementation of this platform is applied to chips with approxi-

mately $8 \times 10^3$ trenches across the whole mother machine device ($\sim 400$ trenches per FOV). This is a limitation of microfluidic device capacity, which can feasibly be increased to the order of $10^6$ by using multiple devices in parallel or more densely-packed serpentine designs as opposed to the standard linear channel [13]. Serpentine chips have been designed to contain twice as many cells per FOV than the linear designs, and with a lower-magnification objective one could fit up to 1500 cells in a single FOV. While the proof-of-concept selection experiment demonstrated here used $300\,\mathrm{s}$ of UV light for selection, it was shown possible to decrease this time to $60\,\mathrm{s}$ using the photosensitiser methylene blue (MB) alongside targeted red light.

The emergence of two surviving and dying *E. coli* subpopulations in our photoselection step is mirrored by prior studies of oxidative stress [172], caused both by UV and red light in presence of photosensitisers [174, 175]. This is explained by populational heterogeneity immediately *prior* to the onset of adverse conditions, with a fraction of cells in favourable environments expressing more stress-response protein at the cost of slower growth [172]. In the future, our platform's ability to observe cells over time and measure their growth rates could help identify targeted cells which are stress-resistant. This may allow to overcome the photoselection trade offs, for example having trench-specific non-uniform light intensity delivered by the DMD in proportion to their (estimated) stress-resistance, which would minimise total light delivered (and hence off-target effects) while optimising targeted killing.

While this proof-of-concept work focused on selection of cells based on a categorical variable (mCherry or GFP expression), future development would apply the

technique for selection of cells based on a continuous variable (e.g. fluorescence of a variant library). In this, we believe a further advantage over FACS will be unlocked, which is the time over which cells can be observed. Gene expression fluctuates naturally over time (e.g. as a function of cell cycle phase, intracellular resource fluctuations, stochastic noise in gene expression), and thus single time-point measurements will be inherently noisy. By observing cells over longer durations and averaging the measurements, resulting in a readout closer to the "true" selected phenotype.

Beyond simple gene expression measurements, this system also enables exploration of dynamic cellular responses from real-time perturbations. Particularly useful in the context of directed evolution of biosensors is also the ability to challenge cells with an array of stimuli before performing selection. These stimuli could consist of chemical stimuli flowed in via microfluidics, as well as optogenetic stimuli delivered via the DMD. When applying MiPS to such applications, particularly when the fitness landscapes on which selection is being performed are highly rugged, running the selection process in a regime where some de-selected cells survive each selection round may even *improve* the overall performance of the directed evolution process, acting as an effective "base chance" as described in Chapter 2 of this thesis.

Overall these results demonstrate that high-throughput selection from mother machine microfluidic chips can be achieved by integrating a DMD with automated data processing and control algorithms. While mother machines have proved immensely effective tools for understanding dynamic and long-term cell behaviours at the single-cell level, this approach can equivalently be applied to cells on a

standard microscope slide when desired phenotypes are compatible with such a configuration. Taken together, these advantages highlight MiPS as an enabler of high-throughput data collection and multi-parametric directed evolution of diverse properties for engineered biotechnologies.

## 4.7 Statement of authorship

Authorship with regards to hardware and software supporting this body of work is described in the Statement of Authorship of the previous chapter. Most of the experiments (conceptualisation and execution) conducted in this chapter were a collaborative effort between Idris Kempf and myself. The analysis shown was conducted by myself, using the DeLTA-integrated analysis pipeline developed by Idris.

# Discussion

This thesis describes the development of an experimental platform optimised for DE of dynamic phenotypes. The system enables long-term, high-throughput observation of up to $10^6$ individual cells cultured under highly controlled microfluidic conditions, with imaging frequencies ranging from milliseconds to several hundred generations [13]. Precise environmental control allows rapid introduction of chemical stimuli with minimal unwanted perturbation to cellular growth or imaging. In addition, an integrated DMD provides spatially targeted light delivery at sub-cellular resolution ($< 0.5, \mu$m), enabling precise interaction with individual cells (Chapter 2). This capability can be used both to deliver optogenetic signals and to perform fully automated, on-chip selection via high-power UV illumination (Chapter 3).

Together, these features unlock the ability to perform high-throughput DE on a class of phenotypic information that is uniquely accessible in the mother machine: dynamic responses over time, multi-dimensional phenotypes arising from sequential or combined stimuli, and single-cell responses to chemical, optical, and magnetic perturbations. Importantly, the experimental capabilities are supported by a parallel thread of theoretical work, which explores how DE outcomes can be maximised in sequence-free (i.e. continuous) selection contexts, as well as in scenarios where multiple phenotypes are to be optimised simultaneously - conditions that are naturally enabled by this platform (Chapter 2).

DE consists of two main elements: selection and mutagenesis. This platform is primarily a method of selection, therefore in order to finalise the DE pipeline as a

whole it must be integrated with a mutagenesis method that can also be performed iteratively. Mutagenesis can be performed *in vitro*, by synthesising gene libraries and transforming them into cells for evaluation. Such methods have the advantage of offering control over library design, but are resource- and labour-intensive, requiring manual intervention with each round that limits the number of rounds that can be performed. This platform lends itself well to *continuous* DE (as will be described later), in which the same population of cells is mutated *in vivo* after selection and returned to the mother machine.

Methods for *in vivo* mutagenesis can be global, such as UV or chemical mutagens - neither of which would require any additional modifications to the platform. While highly convenient and suitable for evolution of a strain as a whole, these methods are less suitable for evolution of a specific gene or protein, as off-target mutations may influence fitness measurements. An emerging family of more sophisticated targeted *in vivo* mutagenesis methods allows this limitation to be overcome, by targeting mutagenesis primarily within that region. Such methods, including EvolvR [176], MutaT7 [27], BacORep [29] and CTRLE [30], have already begun to be adopted by the wider community, and are continually undergoing further refinement for increased performance [26, 36].

This platform is also entirely compatible with non-continuous directed evolution (i.e. by selecting from a library of variants, sequencing the output, and synthesising the next library of variants). One could either sequence the outcome of MiPS-based selection (Chapter 4) that flows out of the chip, or in a more sophisticated fashion perform on-chip sequencing with FISH, which was described in detail in the Discussion of Chapter 3. FISH involves fixation and permeabili-

sation of cells, followed by sequential washing of fluorescent probes which can reveal the locations of known barcodes, and can be later linked to the genotypes of the cells containing each barcode. The advantages of these methods is that one has complete insight into the genetic composition of their population at any time, and can utilise ML DE algorithms to accelerate their process [116, 118, 119], as well as having greater information to gain mechanistic understanding.

Alternatively, the process can be accelerated in continuous DE pipelines, by eliminating the need for manual intervention and therefore enabling many more DE cycles of selection and mutagenesis. In this platform, the process of selection is already automated with MiPS, in which one can select a defined proportion of the population with targeted UV light (or other light-based selection mechanisms). In the format shown in Chapter 4, the method for collecting selected cells was to collect the progeny from the output of the chip, and unselected cells remain in the trenches. Although this was shown to work for one round, it requires that the chip be replaced with each iteration and therefore does not yet *fully* lend itself to automation. With the introduction of trench backports, which can be used to load and unload cells from the trenches [69, 70], one could completely automate this process. For instance, after selection, the cells should be expelled from the chip into a small reservoir where survivors can expand, and then recirculated back into the same chip to initialise the next round. *In vivo* mutagenesis could be implemented at any point in this process.

Given the unique ability of this system to evolve dynamic phenotypes, initial targets for DE could focus on biosensors. Not only can the system measure single-cell biosensor responses in a way not possible with FACS, it can also be used to

investigate specificity by sequentially flowing in alternative compounds and observing the responses in a single round of selection. Promising initial targets could include allosteric transcription factors (aTFs) [110, 177, 178], which are single-gene transcription factors that can activate gene expression (via their unique promoter) in response to small molecules. aTFs are simple to convert into fluorescent reporters (i.e. by placing a fluorescent protein downstream of the aTF promoter), and are ubiquitous in nature, offering many starting points for DE.

Finally, the platform can be used to experimentally explore the theory hypotheses outlined in Chapter 2. With a fully-automated DE platform, it should be much simpler to run several replicates of a directed evolution experiment, to compare the average outcomes with and without base chance implementation, for instance. Freedom to alter chip design also enables a modularity to the platform that allows it easily to be tailored to population splitting (e.g. by physically segregating several mother machine channels). In addition, much of the DE theory work presented could be enhanced by growing the bank of available fitness landscapes, which this platform could be applied to. The largest empirical landscapes existing at present are 160,000 variants, and given that chip size can be expanded to up to 1,000,000, characterisation of similar fitness landscapes could be achieved using FISH, without the need for excessive manual labour.

Despite these capabilities, several practical limitations of the platform should be acknowledged. First, potential accumulation of mutations in selected cells from repeated exposure to off-target UV/methylene blue needs investigating. If prevalent, this could prove limiting with respect to the amount of rounds of directed evolution that can take place, and/or produce misleading artefacts in sequencing.

To reduce these effects, it would be possible to further target light (i.e. by using segmentation masks of cells in a trench, as opposed to targeting the entire trench). Second, long-term operation of the microfluidic system can be affected by practical issues such as channel blockages due to accumulation of cells and debris. Cell blockages within the chip itself can be identified in the imaging and subsequently targeted with UV, although currently this is a manual process that could be vastly improved by automation within the software itself. Thirdly, although this system can image up to $10^6$ cells (limited by microfluidic chip size), it should be noted that with increases in fields-of-view to image come inherent reductions in the frequency with which cells can be imaged, due to the time taken to move between fields-of-view. Finally and most broadly, this system allows us to look more closely at real time behaviour of single cells on fast time scales, but we still only have several readouts - in the form of emission wavelengths - with which to interpret that information. With further advances in the suite of fluorescence microscopy tools, perhaps it will become possible to gain even greater information about the inner workings of the cells we observe with this system.

In summary, this thesis presents the development of a fully integrated experimental platform that enables directed evolution on dynamic, single-cell phenotypes within the mother machine microfluidic framework. By combining long-term, high-throughput observation with automated selection and multiple modes of cellular perturbation, this work extends the scope of directed evolution beyond static or endpoint measurements into a regime where time-resolved, multidimensional behaviours can be evolved directly. In parallel, the accompanying theoretical framework clarifies how selection can be optimised in continuous

and multi-phenotype contexts, providing guidance for the practical deployment of such systems. Together, these contributions offer solutions to previous challenges of biosensor engineering with applicability to a host of other topics such as strain characterisation and fitness landscape mapping.

# Bibliography


[1] Xu Yan et al. “Applications of synthetic biology in medical and pharmaceutical fields”. In: *Signal Transduction and Targeted Therapy* 8 (May 2023), p. 199. ISSN: 2095-9907. DOI: 10.1038/s41392-023-01440-5. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10173249/ (visited on 03/05/2024).

[2] Corinne D. Scown and Jay D. Keasling. “Sustainable manufacturing with synthetic biology”. en. In: *Nature Biotechnology* 40.3 (Mar. 2022), pp. 304–307. ISSN: 1546-1696. DOI: 10.1038/s41587-022-01248-8. URL: https://www.nature.com/articles/s41587-022-01248-8 (visited on 03/05/2024).

[3] Demi Sargent et al. “Synthetic biology and opportunities within agricultural crops”. en. In: *Journal of Sustainable Agriculture and Environment* 1.2 (2022), pp. 89–107. ISSN: 2767-035X. DOI: 10.1002/sae2.12014. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/sae2.12014 (visited on 03/05/2024).

[4] Frances H. Arnold. “Design by Directed Evolution”. In: *Accounts of Chemical Research* 31.3 (Mar. 1998), pp. 125–131. ISSN: 0001-4842. DOI: 10.1021/ar960017f. URL: https://doi.org/10.1021/ar960017f (visited on 02/28/2023).

[5] Andrew E Nixon, Daniel J Sexton, and Robert C Ladner. “Drugs derived from phage display”. In: *mAbs* 6.1 (Jan. 2014), pp. 73–85. ISSN: 1942-0862. DOI: 10.4161/mabs.27240. URL: https://www

.ncbi.nlm.nih.gov/pmc/articles/PMC3929457/ (visited on 07/14/2023).

[6] Bradley S. Heater et al. "Directed evolution of a genetically encoded immobilized lipase for the efficient production of biodiesel from waste cooking oil". In: *Biotechnology for Biofuels* 12.1 (June 2019), p. 165. ISSN: 1754-6834. DOI: 10.1186/s13068-019-1509-5. URL: https://doi.org/10.1186/s13068-019-1509-5 (visited on 07/14/2023).

[7] Martin Neuenschwander et al. "A simple selection strategy for evolving highly efficient enzymes". en. In: *Nature Biotechnology* 25.10 (Oct. 2007), pp. 1145–1147. ISSN: 1546-1696. DOI: 10.1038/nbt1341. URL: https://www.nature.com/articles/nbt1341 (visited on 02/19/2024).

[8] Yinan Wu et al. "Advanced strategies and tools to facilitate and streamline microbial adaptive laboratory evolution". en. In: *Trends in Biotechnology* 40.1 (Jan. 2022), pp. 38–59. ISSN: 0167-7799. DOI: 10.1016/j.tibtech.2021.04.002. URL: https://www.sciencedirect.com/science/article/pii/S0167779921000834 (visited on 01/03/2023).

[9] Guangyu Yang and Stephen G. Withers. "Ultrahigh-throughput FACS-based screening for directed enzyme evolution". eng. In: *Chembiochem* 10.17 (Nov. 2009), pp. 2704–2715. ISSN: 1439-7633. DOI: 10.1002/cbic.200900384.

[10] Vikas D. Trivedi et al. "Cheating the cheater: Suppressing false positive enrichment during biosensor-guided biocatalyst engineering". In: *ACS synthetic biology* 11.1 (Jan. 2022), pp. 420–429. ISSN: 2161-5063. DOI: 10

.1021/acssynbio.1c00506. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9375551/ (visited on 07/26/2023).

[11] Gabriel Abrahams et al. "Quantum spin resonance in engineered proteins for multimodal sensing". en. In: *Nature* 649.8099 (Jan. 2026), pp. 1172–1179. ISSN: 1476-4687. DOI: 10.1038/s41586-025-09971-3. URL: https://www.nature.com/articles/s41586-025-09971-3 (visited on 02/18/2026).

[12] Ping Wang et al. "Robust growth of Escherichia coli". In: *Current Biology* 20.12 (June 2010), pp. 1099–1103. ISSN: 0960-9822. DOI: 10.1016/j.cub.2010.04.045. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2902570/ (visited on 09/05/2025).

[13] Daniel S. Eaton et al. "Essentialome-Wide Multigenerational Imaging Reveals Mechanistic Origins of Cell Growth Laws". In: *bioRxiv* (June 2025), p. 2025.06.10.658525. DOI: 10.1101/2025.06.10.658525. (Visited on 09/09/2025).

[14] Scott Luro et al. "Isolating live cells after high-throughput, long-term, time-lapse microscopy". en. In: *Nature Methods* 17.1 (Jan. 2020), pp. 93–100. ISSN: 1548-7105. DOI: 10.1038/s41592-019-0620-7. URL: https://www.nature.com/articles/s41592-019-0620-7 (visited on 09/05/2025).

[15] Scott Luro. "Quantifying and Screening Dynamic Phenotypes in Bacteria". https://dash.harvard.edu/server/api/core/bitstreams/b228b84f-ef09-4d34-92e9-bf5cc5015d6b/content. PhD thesis. Cambridge, Massachusetts, USA: Harvard University, 2021. URL: https://dash.harvard.edu/server/api/

core/bitstreams/b228b84f-ef09-4d34-92e9-bf5cc5015d6b/content.

[16] Jean-Baptiste Lugagne, Caroline M. Blassick, and Mary J. Dunlop. "Deep model predictive control of gene expression in thousands of single cells". In: *Nature Communications* 15.1 (2024), p. 2148. DOI: 10.1038/s41467-024-46361-1.

[17] John Jumper et al. "Highly accurate protein structure prediction with AlphaFold". en. In: *Nature* 596.7873 (Aug. 2021), pp. 583–589. ISSN: 1476-4687. DOI: 10.1038/s41586-021-03819-2. URL: https://www.nature.com/articles/s41586-021-03819-2 (visited on 07/31/2023).

[18] Minkyung Baek et al. "Accurate prediction of protein structures and interactions using a three-track neural network". In: *Science* 373.6557 (Aug. 2021), pp. 871–876. DOI: 10.1126/science.abj8754. URL: https://www.science.org/doi/10.1126/science.abj8754 (visited on 10/31/2023).

[19] Joseph L. Watson et al. "De novo design of protein structure and function with RFdiffusion". In: *Nature* 620 (2023), pp. 153–160. DOI: 10.1038/s41586-023-06415-8.

[20] J. Dauparas et al. "Robust deep learningâbased protein sequence design using ProteinMPNN". In: *Science* 378.6615 (Oct. 2022), pp. 49–56. DOI: 10.1126/science.add2187. URL: https://www.science.org/doi/10.1126/science.add2187 (visited on 10/31/2023).

[21] Lara SellÃ©s Vidal et al. "A primer to directed evolution: current methodologies and future directions". en. In: *RSC Chemical Biology* 4.4 (2023),

pp. 271–291. DOI: 10.1039/D2CB00231K. URL: https://pubs.rsc.org/en/content/articlelanding/2023/cb/d2cb00231k (visited on 07/14/2023).

[22] Edward G. Hibbert and Paul A. Dalby. "Directed evolution strategies for improved enzymatic performance". In: *Microbial Cell Factories* 4.1 (Oct. 2005), p. 29. ISSN: 1475-2859. DOI: 10.1186/1475-2859-4-29. URL: https://doi.org/10.1186/1475-2859-4-29 (visited on 08/10/2023).

[23] Asako Sawano and Atsushi Miyawaki. "Directed evolution of green fluorescent protein by a new versatile PCR strategy for site-directed and semi-random mutagenesis". In: *Nucleic Acids Research* 28.16 (Aug. 2000), e78. ISSN: 0305-1048. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC108465/ (visited on 07/14/2023).

[24] Jason Gibbons, Micaela Laber, and Charles Bennett. "Humira: The First $20 Billion Drug". en. In: February 2023 29 (Feb. 2023). URL: https://www.ajmc.com/view/humira-the-first-20-billion-drug (visited on 07/24/2023).

[25] Michael S. Packer and David R. Liu. "Methods for the directed evolution of proteins". en. In: *Nature Reviews Genetics* 16.7 (July 2015), pp. 379–394. ISSN: 1471-0064. DOI: 10.1038/nrg3927. URL: https://www.nature.com/articles/nrg3927 (visited on 03/07/2023).

[26] Rosana S. Molina et al. "In vivo hypermutation and continuous evolution". en. In: *Nature Reviews Methods Primers* 2.1 (May 2022), pp. 1–22. ISSN: 2662-8449. DOI: 10.1038/s43586-022-00119-5. URL: https:

//www.nature.com/articles/s43586-022-00119-5 (visited on 10/21/2022).

[27] Christopher L. Moore, Louis J. Papa, and Matthew D. Shoulders. "A Processive Protein Chimera Introduces Mutations Across Defined DNA Regions In Vivo". In: *Journal of the American Chemical Society* 140.37 (Sept. 2018), pp. 11560–11564. ISSN: 0002-7863. DOI: 10.1021/jacs.8b04001. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6166643/ (visited on 07/26/2023).

[28] Shakked O. Halperin et al. "CRISPR-guided DNA polymerases enable diversification of all nucleotides in a tunable window". en. In: *Nature* 560.7717 (Aug. 2018), pp. 248–252. ISSN: 1476-4687. DOI: 10.1038/s41586-018-0384-8. URL: https://www.nature.com/articles/s41586-018-0384-8 (visited on 03/07/2023).

[29] D. Tian et al. "Cell Sorting-Directed Selection of Bacterial Cells in Bigger Sizes Analyzed by Imaging Flow Cytometry during Experimental Evolution". In: *International Journal of Molecular Sciences* 24.4 (2023), p. 3243. DOI: 10.3390/ijms24043243.

[30] Shuaili Chen et al. "CRISPRâDNA Polymerase Assisted Targeted Mutagenesis for Regulable Laboratory Evolution". In: *Adv Sci (Weinh)* 12.45 (Sept. 2025), e11448. ISSN: 2198-3844. DOI: 10.1002/advs.202511448. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC12677585/ (visited on 02/11/2026).

[31] Hyojin Park and Seokhee Kim. "Gene-specific mutagenesis enables rapid continuous evolution of enzymes in vivo". In: *Nucleic Acids Research* 49.6 (Apr. 2021), e32. ISSN: 0305-1048. DOI: 10.1093/nar/gkaa1231.

URL: https://doi.org/10.1093/nar/gkaa1231 (visited on 01/03/2023).

[32] Beatriz Ãlvarez et al. “In vivo diversification of target genomic sites using processive base deaminase fusions blocked by dCas9”. en. In: *Nature Communications* 11.1 (Dec. 2020), p. 6436. ISSN: 2041-1723. DOI: 10.1038/s41467-020-20230-z. URL: https://www.nature.com/articles/s41467-020-20230-z (visited on 10/11/2022).

[33] Amanuella A. Mengiste et al. “Expanded MutaT7 toolkit efficiently and simultaneously accesses all possible transition mutations in bacteria”. eng. In: *Nucleic Acids Research* (Jan. 2023), gkad003. ISSN: 1362-4962. DOI: 10.1093/nar/gkad003.

[34] Aaron Cravens et al. “Polymerase-guided base editing enables in vivo mutagenesis and rapid protein engineering”. en. In: *Nature Communications* 12.1 (Mar. 2021), p. 1579. ISSN: 2041-1723. DOI: 10.1038/s41467-021-21876-z. URL: https://www.nature.com/articles/s41467-021-21876-z (visited on 07/26/2023).

[35] Haiqi Chen et al. “Efficient, continuous mutagenesis in human cells using a pseudo-random DNA editor”. eng. In: *Nature Biotechnology* 38.2 (Feb. 2020), pp. 165–168. ISSN: 1546-1696. DOI: 10.1038/s41587-019-0331-8.

[36] Georgie Hau SÃ¸rensen et al. *Structure-guided design of protein attachment points for functional augmentation of complex molecular machines*. en. Pages: 2025.06.22.660945 Section: New Results. June 2025. DOI: 10.1101/2025.06.22.660945. URL: https://www.biorxiv

.org/content/10.1101/2025.06.22.660945v1 (visited on 01/16/2026).

[37] Jorge D GarcÃa-GarcÃa et al. “Using continuous directed evolution to improve enzymes for plant applications”. In: *Plant Physiology* 188.2 (Feb. 2022), pp. 971–983. ISSN: 0032-0889. DOI: 10.1093/plphys/kiab500. URL: https://doi.org/10.1093/plphys/kiab500 (visited on 07/26/2023).

[38] Connor J. Tou, David V. Schaffer, and John E. Dueber. “Targeted Diversification in the *S.* cerevisiae Genome with CRISPR-Guided DNA Polymerase I”. en. In: *ACS Synthetic Biology* 9.7 (July 2020), pp. 1911–1916. ISSN: 2161-5063, 2161-5063. DOI: 10.1021/acssynbio.0c00149. URL: https://pubs.acs.org/doi/10.1021/acssynbio.0c00149 (visited on 07/26/2023).

[39] Juan E. Hurtado et al. “Nickase fidelity drives EvolvR-mediated diversification in mammalian cells”. en. In: *Nat Commun* 16.1 (Apr. 2025), p. 3723. ISSN: 2041-1723. DOI: 10.1038/s41467-025-58414-0. URL: https://www.nature.com/articles/s41467-025-58414-0 (visited on 02/11/2026).

[40] Rongzhen Tian et al. “Engineered bacterial orthogonal DNA replication system for continuous evolution”. en. In: *Nature Chemical Biology* (July 2023), pp. 1–9. ISSN: 1552-4469. DOI: 10.1038/s41589-023-01387-2. URL: https://www.nature.com/articles/s41589-023-01387-2 (visited on 07/26/2023).

[41] Gordon Rix et al. “Scalable continuous evolution for the generation of diverse enzyme variants encompassing promiscuous activities”. en. In: *Na-*

*ture Communications* 11.1 (Nov. 2020), p. 5644. ISSN: 2041-1723. DOI: 10.1038/s41467-020-19539-6. URL: https://www.nature.com/articles/s41467-020-19539-6 (visited on 03/07/2023).

[42] Atsushi Shibai et al. "Mutation accumulation under UV radiation in Escherichia coli". en. In: *Scientific Reports* 7.1 (Nov. 2017), p. 14531. ISSN: 2045-2322. DOI: 10.1038/s41598-017-15008-1. URL: https://www.nature.com/articles/s41598-017-15008-1 (visited on 08/04/2023).

[43] Alan P. Watt et al. "Higher Throughput Bioanalysis by Automation of a Protein Precipitation Assay Using a 96-Well Format with Detection by LCâMS/MS". In: *Analytical Chemistry* 72.5 (Mar. 2000), pp. 979–984. ISSN: 0003-2700. DOI: 10.1021/ac9906633. URL: https://doi.org/10.1021/ac9906633 (visited on 07/26/2023).

[44] Stefan Becker et al. "Ultra-high-throughput screening based on cell-surface display and fluorescence-activated cell sorting for the identification of novel biocatalysts". eng. In: *Current Opinion in Biotechnology* 15.4 (Aug. 2004), pp. 323–329. ISSN: 0958-1669. DOI: 10.1016/j.copbio.2004.06.001.

[45] Ariane Stucki et al. "Droplet Microfluidics and Directed Evolution of Enzymes: An Intertwined Journey". en. In: *Angewandte Chemie International Edition* 60.46 (2021), pp. 24368–24387. ISSN: 1521-3773. DOI: 10.1002/anie.202016154. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.202016154 (visited on 03/15/2023).

[46] S. Miltenyi et al. “High gradient magnetic cell separation with MACS”. eng. In: *Cytometry* 11.2 (1990), pp. 231–238. ISSN: 0196-4763. DOI: 10.1002/cyto.990110203.

[47] I. Saggio and R. Laufer. “Biotin binders selected from a random peptide library expressed on phage”. eng. In: *The Biochemical Journal* 293 ( Pt 3).Pt 3 (Aug. 1993), pp. 613–616. ISSN: 0264-6021. DOI: 10.1042/bj2930613.

[48] Hayato Yanagida, Tomoaki Matsuura, and Tetsuya Yomo. “Ribosome display for rapid protein evolution by consecutive rounds of mutation and selection”. eng. In: *Methods in Molecular Biology (Clifton, N.J.)* 634 (2010), pp. 257–267. ISSN: 1940-6029. DOI: 10.1007/978-1-60761-652-8_19.

[49] Robert E Speight et al. “A new plasmid display technology for the in vitro selection of functional phenotypeâgenotype linked proteins”. en. In: *Chemistry & Biology* 8.10 (Oct. 2001), pp. 951–965. ISSN: 1074-5521. DOI: 10.1016/S1074-5521(01)00066-7. URL: https://www.sciencedirect.com/science/article/pii/S1074552101000667 (visited on 03/15/2023).

[50] A. D. Keefe. “Protein selection using mRNA display”. eng. In: *Current Protocols in Molecular Biology* Chapter 24 (May 2001), Unit 24.5. ISSN: 1934-3647. DOI: 10.1002/0471142727.mb2405s53.

[51] Chang Ren et al. “Selection and screening strategies in directed evolution to improve protein stability”. In: *Bioresources and Bioprocessing* 6.1 (Dec. 2019), p. 53. ISSN: 2197-4365. DOI: 10.1186/s40643-019

-0288-y. URL: https://doi.org/10.1186/s40643-019-0288-y (visited on 08/10/2023).

[52] Sally Turner. *Humira: the highs and lows of the worldâs most successful drug*. en-US. Publication Title: Pharmaceutical Technology. Sept. 2018. URL: https://www.pharmaceutical-technology.com/features/humira-abbvie-drug/ (visited on 07/24/2023).

[53] Linas Mazutis et al. "Single-cell analysis and sorting using droplet-based microfluidics". en. In: *Nature Protocols* 8.5 (May 2013). Publisher: Nature Publishing Group, pp. 870–891. ISSN: 1750-2799. DOI: 10.1038/nprot.2013.046. URL: https://www.nature.com/articles/nprot.2013.046 (visited on 07/25/2025).

[54] Shuke Wu et al. "A growth selection system for the directed evolution of amine-forming or converting enzymes". en. In: *Nature Communications* 13.1 (Dec. 2022), p. 7458. ISSN: 2041-1723. DOI: 10.1038/s41467-022-35228-y. URL: https://www.nature.com/articles/s41467-022-35228-y (visited on 03/16/2023).

[55] Bryan J. Leong and Andrew D. Hanson. "Continuous Directed Evolution of a Feedback-Resistant Arabidopsis Arogenate Dehydratase in Plantized Escherichia coli". In: *ACS Synthetic Biology* 12.1 (Jan. 2023), pp. 43–50. DOI: 10.1021/acssynbio.2c00511. URL: https://doi.org/10.1021/acssynbio.2c00511 (visited on 07/26/2023).

[56] Jessica James et al. *Microscopic PhotoSelection (MiPS) of single cells in mother machine microfluidic devices*. en. ISSN: 2692-8205 Pages: 2025.12.03.692065 Section: New Results. Dec. 2025. DOI: 10.64898/2025.12.03

.692065. URL: https://www.biorxiv.org/content/10.64898/2025.12.03.692065v1 (visited on 01/28/2026).

[57] George M. Whitesides. "The origins and the future of microfluidics". en. In: *Nature* 442.7101 (July 2006), pp. 368–373. ISSN: 1476-4687. DOI: 10.1038/nature05058. URL: https://www.nature.com/articles/nature05058 (visited on 02/12/2026).

[58] Dirk Janasek, Joachim Franzke, and Andreas Manz. "Scaling and the design of miniaturized chemical-analysis systems". en. In: *Nature* 442.7101 (July 2006), pp. 374–380. ISSN: 1476-4687. DOI: 10.1038/nature05059. URL: https://www.nature.com/articles/nature05059 (visited on 02/12/2026).

[59] J. Cooper McDonald et al. "Fabrication of microfluidic systems in poly(dimethylsiloxane)". en. In: *ELECTROPHORESIS* 21.1 (2000). _eprint: https://analyticalsciencejournals.onlinelibrary 2683%2820000101%2921%3A1%3C27%3A%3AAID-ELPS27%3E3.0.CO%3B2-C, pp. 27–40. ISSN: 1522-2683. DOI: 10.1002/(SICI)1522-2683(20000101)21:1<27::AID-ELPS27>3.0.CO;2-C. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291522-2683%2820000101%2921%3A1%3C27%3A%3AAID-ELPS27%3E3.0.CO%3B2-C (visited on 02/12/2026).

[60] Jong Wook Hong and Stephen R. Quake. "Integrated nanoliter systems". en. In: *Nat Biotechnol* 21.10 (Oct. 2003), pp. 1179–1183. ISSN: 1546-1696. DOI: 10.1038/nbt871. URL: https://www.nature.com/articles/nbt871 (visited on 02/12/2026).

[61] Nam-Trung Nguyen and Zhigang Wu. "Micromixersâa review". en. In: *J. Micromech. Microeng.* 15.2 (Dec. 2004), R1. ISSN: 0960-1317. DOI:

10.1088/0960-1317/15/2/R01. URL: https://doi.org/10.1088/0960-1317/15/2/R01 (visited on 02/12/2026).

[62] D. J. Laser and J. G. Santiago. "A review of micropumps". en. In: *J. Micromech. Microeng.* 14.6 (Apr. 2004), R35. ISSN: 0960-1317. DOI: 10.1088/0960-1317/14/6/R01. URL: https://doi.org/10.1088/0960-1317/14/6/R01 (visited on 02/12/2026).

[63] Jiqing Sai et al. "Study of Chemotaxis and Cell-Cell Interactions in Cancer with Microfluidic Devices". eng. In: *Methods Enzymol* 570 (2016), pp. 19–45. ISSN: 1557-7988. DOI: 10.1016/bs.mie.2015.09.023.

[64] Tongcheng Qian, Eric V. Shusta, and Sean P. Palecek. "Advances in microfluidic platforms for analyzing and regulating human pluripotent stem cells". eng. In: *Curr Opin Genet Dev* 34 (Oct. 2015), pp. 54–60. ISSN: 1879-0380. DOI: 10.1016/j.gde.2015.07.007.

[65] Dongeun Huh et al. "Microfabrication of human organs-on-chips". eng. In: *Nat Protoc* 8.11 (Nov. 2013), pp. 2135–2157. ISSN: 1750-2799. DOI: 10.1038/nprot.2013.137.

[66] Nathalie Q. Balaban et al. "Bacterial Persistence as a Phenotypic Switch". en. In: *Science* 305.5690 (Sept. 2004), pp. 1622–1625. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1099390. URL: https://www.science.org/doi/10.1126/science.1099390 (visited on 07/27/2023).

[67] Laurent Potvin-Trottier, Scott Luro, and Johan Paulsson. "Microfluidics and single-cell microscopy to study stochastic processes in bacteria". In: *Current Opinion in Microbiology* 43 (June 2018), pp. 186–192. ISSN:

1369-5274. DOI: 10.1016/j.mib.2017.12.004. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6044433/ (visited on 09/05/2025).

[68] Matthias Kaiser et al. "Monitoring single-cell gene regulation under dynamically controllable conditions with integrated microfluidics and software". en. In: *Nature Communications* 9.1 (Jan. 2018), p. 212. ISSN: 2041-1723. DOI: 10.1038/s41467-017-02505-0. URL: https://www.nature.com/articles/s41467-017-02505-0 (visited on 07/28/2023).

[69] Özden Baltekin et al. "Antibiotic susceptibility testing in less than 30 min using direct single-cell imaging". In: *Proceedings of the National Academy of Sciences* 114.34 (2017), pp. 9170–9175. DOI: 10.1073/pnas.1708558114.

[70] Anna L. Calkins et al. "Achieving Single-Molecule Tracking of Subcellular Regulation in Bacteria During Real-Time Environmental Perturbations". In: *Analytical Chemistry* 95.2 (Jan. 2023), pp. 774–783. ISSN: 0003-2700. DOI: 10.1021/acs.analchem.2c02899. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10292772/ (visited on 09/08/2025).

[71] Ping Wang et al. "Robust Growth of *Escherichia coli*". In: *Current Biology* 20.12 (2010), pp. 1099–1103. ISSN: 0960-9822. DOI: 10.1016/j.cub.2010.04.045.

[72] Sattar Taheri-Araghi et al. "Cell-Size Control and Homeostasis in Bacteria". English. In: *Current Biology* 25.3 (Feb. 2015), pp. 385–391. ISSN: 0960-9822. DOI: 10.1016/j.cub.2014.12.009. URL: https:

//www.cell.com/current-biology/abstract/S0960-9822(14)01573-5 (visited on 09/08/2025).

[73] Ulrich Karl Steiner. "Senescence in Bacteria and Its Underlying Mechanisms". In: *Frontiers in Cell and Developmental Biology* 9 (June 2021), p. 668915. ISSN: 2296-634X. DOI: 10.3389/fcell.2021.668915. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8249858/ (visited on 09/08/2025).

[74] Maxence S Vincent and Stephan Uphoff. "Cellular heterogeneity in DNA alkylation repair increases population genetic plasticity". In: *Nucleic Acids Research* 49.21 (Dec. 2021), pp. 12320–12331. ISSN: 0305-1048. DOI: 10.1093/nar/gkab1143. (Visited on 09/08/2025).

[75] Valentine Lagage, Victor Chen, and Stephan Uphoff. "Adaptation delay causes a burst of mutations in bacteria responding to oxidative stress". In: *EMBO Reports* 24.1 (Nov. 2022), e55640. ISSN: 1469-221X. DOI: 10.15252/embr.202255640. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9827559/ (visited on 09/08/2025).

[76] Divya Choudhary et al. "Phenotypic heterogeneity in the bacterial oxidative stress response is driven by cell-cell interactions". In: *Cell Reports* 42.3 (Mar. 2023), p. 112168. ISSN: 2211-1247. DOI: 10.1016/j.celrep.2023.112168. URL: https://www.sciencedirect.com/science/article/pii/S2211124723001791 (visited on 09/08/2025).

[77] Matthew T. Cabeen and Richard Losick. "Single-cell Microfluidic Analysis of Bacillus subtilis". In: *Journal of Visualized Experiments : JoVE* 131 (Jan. 2018), p. 56901. ISSN: 1940-087X. DOI: 10.3791/56901.

URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5908703/ (visited on 09/08/2025).

[78] John T. Sauls et al. "Control of Bacillus subtilis replication initiation during physiological transitions and perturbations". In: *mBio* 10.6 (Dec. 2019), 10.1128/mbio.02205–19. DOI: 10.1128/mbio.02205-19. (Visited on 09/08/2025).

[79] Sivan Pearl Mizrahi et al. "Persistence to anti-cancer treatments in the stationary to proliferating transition". In: *Cell Cycle* 15.24 (Nov. 2016), pp. 3442–3453. ISSN: 1538-4101. DOI: 10.1080/15384101.2016.1248006. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5224467/ (visited on 09/08/2025).

[80] Akihisa Seita et al. "Intrinsic growth heterogeneity of mouse leukemia cells underlies differential susceptibility to a growth-inhibiting anticancer drug". eng. In: *PloS One* 16.2 (2021), e0236534. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0236534.

[81] Remy Chait et al. "Shaping bacterial population behavior through computer-interfaced control of individual cells". en. In: *Nature Communications* 8.1 (Nov. 2017), p. 1535. ISSN: 2041-1723. DOI: 10.1038/s41467-017-01683-1. URL: https://www.nature.com/articles/s41467-017-01683-1 (visited on 10/08/2022).

[82] Sewall Wright. "The Roles of Mutation, Inbreeding, crossbreeding and Selection in Evolution." In: *Proceedings of the XI International Congress of Genetics* 8 (1932), pp. 209–222.

[83] Daniel M. Weinreich et al. "Darwinian Evolution Can Follow Only Very Few Mutational Paths to Fitter Proteins". In: *Science* 312.5770 (Apr. 2006),

pp. 111–114. DOI: 10.1126/science.1123539. URL: https://www.science.org/doi/10.1126/science.1123539 (visited on 03/06/2023).

[84] Thuy-Lan V Lite et al. "Uncovering the basis of protein-protein interaction specificity with a combinatorially complete library". In: *eLife* 9 (Oct. 2020). Ed. by Christian R Landry and Olga Boudker, e60924. ISSN: 2050-084X. DOI: 10.7554/eLife.60924. URL: https://doi.org/10.7554/eLife.60924 (visited on 06/19/2023).

[85] Aisha I. Khan et al. "Negative Epistasis Between Beneficial Mutations in an Evolving Bacterial Population". In: *Science* 332.6034 (June 2011), pp. 1193–1196. DOI: 10.1126/science.1203801. URL: https://www.science.org/doi/10.1126/science.1203801 (visited on 03/06/2023).

[86] Hsin-Hung Chou et al. "Mapping the Fitness Landscape of Gene Expression Uncovers the Cause of Antagonism and Sign Epistasis between Adaptive Mutations". en. In: *PLOS Genetics* 10.2 (Feb. 2014), e1004149. ISSN: 1553-7404. DOI: 10.1371/journal.pgen.1004149. URL: https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1004149 (visited on 03/06/2023).

[87] Claudia Bank et al. "On the (un)predictability of a large intragenic fitness landscape". In: *Proceedings of the National Academy of Sciences* 113.49 (Dec. 2016), pp. 14085–14090. DOI: 10.1073/pnas.1612676113. URL: https://www.pnas.org/doi/full/10.1073/pnas.1612676113 (visited on 03/06/2023).

[88] Nicholas C. Wu et al. “Adaptation in protein fitness landscapes is facilitated by indirect paths”. eng. In: *eLife* 5 (July 2016), e16965. ISSN: 2050-084X. DOI: 10.7554/eLife.16965.

[89] Kadina E. Johnston et al. “A combinatorially complete epistatic fitness landscape in an enzyme active site”. In: *Proceedings of the National Academy of Sciences* 121.32 (2024), e2400439121. DOI: 10.1073/pnas.2400439121. eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.2400439121. URL: https://www.pnas.org/doi/abs/10.1073/pnas.2400439121.

[90] Boqiang Tu, Vikram Sundar, and Kevin M. Esvelt. “An ultra-high-throughput method for measuring biomolecular activities”. In: *bioRxiv* (2024). DOI: 10.1101/2022.03.09.483646. eprint: https://www.biorxiv.org/content/early/2024/03/25/2022.03.09.483646.full.pdf. URL: https://www.biorxiv.org/content/early/2024/03/25/2022.03.09.483646.

[91] InÃªs Fragata et al. “Evolution in the light of fitness landscape theory”. en. In: *Trends in Ecology & Evolution* 34.1 (Jan. 2019), pp. 69–82. ISSN: 0169-5347. DOI: 10.1016/j.tree.2018.10.009. URL: https://www.sciencedirect.com/science/article/pii/S0169534718302660 (visited on 01/04/2023).

[92] O. Tenaillon. “The Utility of Fisherâs Geometric Model in Evolutionary Genetics”. In: *Annual review of ecology, evolution, and systematics* 45 (Nov. 2014), pp. 179–201. ISSN: 1543-592X. DOI: 10.1146/annurev-ecolsys-120213-091846. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4699269/ (visited on 07/25/2023).

[93] R. A. Fisher. *The Genetical Theory of Natural Selection: A Complete Variorum Edition*. en. OUP Oxford, Oct. 1999. ISBN: 978-0-19-850440-5.

[94] S. Gavrilets. "Evolution and speciation on holey adaptive landscapes". eng. In: *Trends in Ecology & Evolution* 12.8 (Aug. 1997), pp. 307–312. ISSN: 0169-5347. DOI: 10.1016/S0169-5347(97)01098-7.

[95] Thomas F. Hansen and GÃ¼nter P. Wagner. "Modeling Genetic Architecture: A Multilinear Theory of Gene Interaction". en. In: *Theoretical Population Biology* 59.1 (Feb. 2001), pp. 61–86. ISSN: 0040-5809. DOI: 10.1006/tpbi.2000.1508. URL: https://www.sciencedirect.com/science/article/pii/S0040580900915089 (visited on 07/25/2023).

[96] Stuart Kauffman and Simon Levin. "Towards a general theory of adaptive walks on rugged landscapes". en. In: *Journal of Theoretical Biology* 128.1 (Sept. 1987), pp. 11–45. ISSN: 0022-5193. DOI: 10.1016/S0022-5193(87)80029-2. URL: https://www.sciencedirect.com/science/article/pii/S0022519387800292 (visited on 10/27/2022).

[97] S. A. Kauffman and E. D. Weinberger. "The NK model of rugged fitness landscapes and its application to maturation of the immune response". eng. In: *Journal of Theoretical Biology* 141.2 (Nov. 1989), pp. 211–245. ISSN: 0022-5193. DOI: 10.1016/s0022-5193(89)80019-0.

[98] J. F. C. Kingman. "A Simple Model for the Balance between Selection and Mutation". In: *Journal of Applied Probability* 15.1 (1978), pp. 1–12. ISSN: 0021-9002. DOI: 10.2307/3213231. URL: https://www.jstor.org/stable/3213231 (visited on 07/25/2023).

[99] Johannes Neidhart, Ivan G Szendro, and Joachim Krug. “Adaptation in Tunably Rugged Fitness Landscapes: The Rough Mount Fuji Model”. In: *Genetics* 198.2 (Oct. 2014), pp. 699–721. ISSN: 1943-2631. DOI: 10.1534/genetics.114.167668. URL: https://doi.org/10.1534/genetics.114.167668 (visited on 07/25/2023).

[100] Michael S. Breen et al. “Epistasis as the primary factor in molecular evolution”. en. In: *Nature* 490.7421 (Oct. 2012), pp. 535–538. ISSN: 1476-4687. DOI: 10.1038/nature11510. URL: https://www.nature.com/articles/nature11510 (visited on 03/06/2023).

[101] Tyler N. Starr and Joseph W. Thornton. “Epistasis in protein evolution”. In: *Protein Science : A Publication of the Protein Society* 25.7 (July 2016), pp. 1204–1218. ISSN: 0961-8368. DOI: 10.1002/pro.2897. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4918427/ (visited on 11/30/2022).

[102] Charlotte M. Miton, Karol Buda, and Nobuhiko Tokuriki. “Epistasis and intramolecular networks in protein evolution”. en. In: *Current Opinion in Structural Biology*. Engineering and Design â Membranes 69 (Aug. 2021), pp. 160–168. ISSN: 0959-440X. DOI: 10.1016/j.sbi.2021.04.007. URL: https://www.sciencedirect.com/science/article/pii/S0959440X21000610 (visited on 01/04/2023).

[103] FranÃ§ois Blanquart and Thomas Bataillon. “Epistasis and the Structure of Fitness Landscapes: Are Experimental Fitness Landscapes Compatible with Fisherâs Geometric Model?” In: *Genetics* 203.2 (June 2016), pp. 847–862. ISSN: 1943-2631. DOI: 10.1534/genetics.115.182691.

URL: https://doi.org/10.1534/genetics.115.182691 (visited on 01/23/2023).

[104] L. Barnett. "Ruggedness and neutralityâthe NKp family of fitness landscapes". In: Aug. 1998. URL: https://www.semanticscholar.org/paper/Ruggedness-and-neutrality%E2%80%94the-NKp-family-of-fitness-Barnett/f3cf719e35b8ea7a0ced458591651f3c (visited on 07/25/2023).

[105] M. E. J. Newman and Robin Engelhardt. *Effects of neutral selection on the evolution of molecular species*. Dec. 1997. DOI: 10.48550/arXiv.adap-org/9712005. URL: http://arxiv.org/abs/adap-org/9712005 (visited on 07/25/2023).

[106] Andrew D Kern and Matthew W Hahn. "The Neutral Theory in Light of Natural Selection". In: *Molecular Biology and Evolution* 35.6 (June 2018), pp. 1366–1371. ISSN: 0737-4038. DOI: 10.1093/molbev/msy092. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5967545/ (visited on 07/25/2023).

[107] Jeffrey D. Jensen et al. "The importance of the Neutral Theory in 1968 and 50 years on: A response to Kern and Hahn 2018". In: *Evolution; international journal of organic evolution* 73.1 (Jan. 2019), pp. 111–114. ISSN: 0014-3820. DOI: 10.1111/evo.13650. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6496948/ (visited on 07/25/2023).

[108] Jesse D. Bloom and Frances H. Arnold. "In the light of directed evolution: Pathways of adaptive protein evolution". In: *Proceedings of the National Academy of Sciences* 106.supplement_1 (June 2009), pp. 9995–10000.

DOI: 10.1073/pnas.0901522106. URL: https://www.pnas.org/doi/10.1073/pnas.0901522106 (visited on 07/25/2023).

[109] Michael Schmutzer. "Not Quite Lost in Translation: Mistranslation Alters Adaptive Landscape Topography and the Dynamics of Evolution". en. In: ().

[110] Leopoldo F. M. Machado, Andrew Currin, and Neil Dixon. "Directed evolution of the PcaV allosteric transcription factor to generate a biosensor for aromatic aldehydes". In: *Journal of Biological Engineering* 13.1 (Nov. 2019), p. 91. ISSN: 1754-1611. DOI: 10.1186/s13036-019-0214-z. URL: https://doi.org/10.1186/s13036-019-0214-z (visited on 11/11/2022).

[111] Philip A. Romero and Frances H. Arnold. "Exploring protein fitness landscapes by directed evolution". en. In: *Nature Reviews Molecular Cell Biology* 10.12 (Dec. 2009), pp. 866–876. ISSN: 1471-0080. DOI: 10.1038/nrm2805. URL: https://www.nature.com/articles/nrm2805 (visited on 10/24/2022).

[112] John H. Holland. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. en. The MIT Press, Apr. 1992. ISBN: 978-0-262-27555-2. DOI: 10.7551/mitpress/1090.001.0001. URL: https://direct.mit.edu/books/book/2574/Adaptation-in-Natural-and-Artificial-SystemsAn (visited on 07/31/2023).

[113] Ekaterina Myasnikova and Alexander Spirov. "Transfer of Genetic Algorithms to Directed Evolution of Macromolecules: Tests in Silico". In: *2020*

*Cognitive Sciences, Genomics and Bioinformatics (CSGB)*. July 2020, pp. 297–300. DOI: 10.1109/CSGB51356.2020.9214734.

[114] W. P. Stemmer. "DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution". eng. In: *Proceedings of the National Academy of Sciences of the United States of America* 91.22 (Oct. 1994), pp. 10747–10751. ISSN: 0027-8424. DOI: 10.1073/pnas.91.22.10747.

[115] Richard Fox et al. "Optimizing the search algorithm for protein engineering by directed evolution". In: *Protein Engineering, Design and Selection* 16.8 (Aug. 2003), pp. 589–597. ISSN: 1741-0126. DOI: 10.1093/protein/gzg077. URL: https://doi.org/10.1093/protein/gzg077 (visited on 07/31/2023).

[116] Jason Yang et al. "Active learning-assisted directed evolution". en. In: *Nature Communications* 16.1 (Jan. 2025). Publisher: Nature Publishing Group, p. 714. ISSN: 2041-1723. DOI: 10.1038/s41467-025-55987-8. URL: https://www.nature.com/articles/s41467-025-55987-8 (visited on 03/11/2025).

[117] Zachary Wu et al. "Machine learning-assisted directed protein evolution with combinatorial libraries". EN. In: *Proceedings of the National Academy of Sciences* 116.18 (Apr. 2019), pp. 8852–8858. DOI: 10.1073/pnas.1901979116. URL: https://www.pnas.org/doi/abs/10.1073/pnas.1901979116 (visited on 10/21/2022).

[118] Bruce J. Wittmann, Yisong Yue, and Frances H. Arnold. "Informed training set design enables efficient machine learning-assisted directed protein evolution". English. In: *Cell Systems* 12.11 (Nov. 2021), 1026–1045.e7.

ISSN: 2405-4712. DOI: 10.1016/j.cels.2021.07.008. URL: https://www.cell.com/cell-systems/abstract/S2405-4712(21)00286-6 (visited on 03/07/2023).

[119] Kevin K. Yang, Zachary Wu, and Frances H. Arnold. "Machine-learning-guided directed evolution for protein engineering". en. In: *Nature Methods* 16.8 (Aug. 2019), pp. 687–694. ISSN: 1548-7105. DOI: 10.1038/s41592-019-0496-6. URL: https://www.nature.com/articles/s41592-019-0496-6 (visited on 07/31/2023).

[120] Sameer DâCosta et al. "Inferring protein fitness landscapes from laboratory evolution experiments". en. In: *PLOS Computational Biology* 19.3 (Mar. 2023), e1010956. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1010956. URL: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1010956 (visited on 07/31/2023).

[121] Thomas A. Hopf et al. "Mutation effects predicted from sequence co-variation". en. In: *Nature Biotechnology* 35.2 (Feb. 2017), pp. 128–135. ISSN: 1546-1696. DOI: 10.1038/nbt.3769. URL: https://www.nature.com/articles/nbt.3769 (visited on 03/07/2023).

[122] Adam J. Riesselman, John B. Ingraham, and Debora S. Marks. "Deep generative models of genetic variation capture the effects of mutations". en. In: *Nature Methods* 15.10 (Oct. 2018), pp. 816–822. ISSN: 1548-7105. DOI: 10.1038/s41592-018-0138-4. URL: https://www.nature.com/articles/s41592-018-0138-4 (visited on 03/07/2023).

[123] Neil Thomas et al. *Tuned Fitness Landscapes for Benchmarking Model-Guided Protein Design*. en. Oct. 2022. DOI: 10.1101/2022.10.28.514293. URL: https://www.biorxiv.org/content/10.1101/2022.10.28.514293v1 (visited on 07/25/2023).

[124] Jessica James et al. "Optimisation strategies for directed evolution without sequencing". en. In: *PLOS Computational Biology* 20.12 (Dec. 2024), e1012695. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1012695. URL: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1012695 (visited on 03/11/2025).

[125] Sebastian Towers et al. *Sequence-free landscape inference for directed evolution*. en. ISSN: 2692-8205 Pages: 2025.10.08.681123 Section: New Results. Oct. 2025. DOI: 10.1101/2025.10.08.681123. URL: https://www.biorxiv.org/content/10.1101/2025.10.08.681123v1 (visited on 01/15/2026).

[126] Andrei Papkou et al. "A rugged yet easily navigable fitness landscape". In: *Science* 382.6673 (Nov. 2023), eadh3860. DOI: 10.1126/science.adh3860. URL: https://www.science.org/doi/10.1126/science.adh3860 (visited on 12/06/2023).

[127] Kadina E. Johnston et al. *A combinatorially complete epistatic fitness landscape in an enzyme active site*. en. June 2024. DOI: 10.1101/2024.06.23.600144. URL: https://www.biorxiv.org/content/10.1101/2024.06.23.600144v1 (visited on 07/16/2024).

[128] Ryan A. LaCroix, Bernhard O. Palsson, and Adam M. Feist. "A Model for Designing Adaptive Laboratory Evolution Experiments". In: *Applied and Environmental Microbiology* 83.8 (Mar. 2017), e03115–16. DOI: 10

.1128/AEM.03115-16. URL: https://journals.asm.org/doi/10.1128/AEM.03115-16 (visited on 08/14/2023).

[129] Yumeng Tan et al. "Directed evolution of an Î±1,3−*fucosyltransferaseusingasingle*−*cellultrahigh*−*throughputscreeningmethod*". eng. In: *Science Advances* 5.10 (Oct. 2019), eaaw8451. ISSN: 2375-2548. DOI: 10.1126/sciadv.aaw8451.

[130] Ran Tu et al. "A Flow CytometryâBased Screening System for Directed Evolution of Proteases". EN. In: *Journal of Biomolecular Screening* 16.3 (Mar. 2011). Publisher: SAGE Publications Inc STM, pp. 285–294. ISSN: 1087-0571. DOI: 10.1177/1087057110396361. URL: https://doi.org/10.1177/1087057110396361 (visited on 03/11/2025).

[131] Trevor S. Frisby and Christopher James Langmead. "Bayesian optimization with evolutionary and structure-based regularization for directed protein evolution". In: *Algorithms for Molecular Biology* 16.1 (July 2021), p. 13. ISSN: 1748-7188. DOI: 10.1186/s13015-021-00195-4. URL: https://doi.org/10.1186/s13015-021-00195-4 (visited on 07/31/2023).

[132] Ruyun Hu et al. *Protein engineering via Bayesian optimization-guided evolutionary algorithm and robotic experiments*. en. Aug. 2022. DOI: 10.1101/2022.08.11.503535. URL: https://www.biorxiv.org/content/10.1101/2022.08.11.503535v1 (visited on 01/03/2023).

[133] Alexander C. Carpenter et al. "Have you tried turning it off and on again? Oscillating selection to enhance fitness-landscape traversal in adaptive laboratory evolution experiments". en. In: *Metabolic Engineering Com-*

*munications* 17 (Dec. 2023), e00227. ISSN: 22140301. DOI: 10.1016/j.mec.2023.e00227. URL: https://linkinghub.elsevier.com/retrieve/pii/S221403012300010X (visited on 08/14/2023).

[134] B. A. Alpay and M. M. Desai. "Effects of selection stringency on the outcomes of directed evolution". In: *PLoS ONE* (2024). DOI: 10.1371/journal.pone.0311438.

[135] Takuyo Aita, Masahiro Iwakura, and Yuzuru Husimi. "A cross-section of the fitness landscape of dihydrofolate reductase". In: *Protein Engineering* 14.9 (Sept. 2001), pp. 633–638. ISSN: 0269-2139. DOI: 10.1093/protein/14.9.633. eprint: https://academic.oup.com/peds/article-pdf/14/9/633/18546067/140633.pdf. URL: https://doi.org/10.1093/protein/14.9.633.

[136] Peter F. Stadler. "Landscapes and their correlation functions". en. In: *Journal of Mathematical Chemistry* 20.1 (Mar. 1996), pp. 1–45. ISSN: 1572-8897. DOI: 10.1007/BF01165154. URL: https://doi.org/10.1007/BF01165154 (visited on 11/03/2022).

[137] Frank J. Poelwijk et al. "Reciprocal sign epistasis is a necessary condition for multi-peaked fitness landscapes". eng. In: *Journal of Theoretical Biology* 272.1 (Mar. 2011), pp. 141–144. ISSN: 1095-8541. DOI: 10.1016/j.jtbi.2010.12.015.

[138] Daniel M. Weinreich, Richard A. Watson, and Lin Chao. "Perspective: Sign epistasis and genetic constraint on evolutionary trajectories". eng. In: *Evolution; International Journal of Organic Evolution* 59.6 (June 2005), pp. 1165–1174. ISSN: 0014-3820.

[139] Ivan G Szendro et al. “Quantitative analyses of empirical fitness landscapes”. en. In: *Journal of Statistical Mechanics: Theory and Experiment* 2013.01 (Jan. 2013), P01005. ISSN: 1742-5468. DOI: 10.1088/1742-5468/2013/01/P01005. URL: https://iopscience.iop.org/article/10.1088/1742-5468/2013/01/P01005 (visited on 07/25/2023).

[140] Yoh Iwasa, Franziska Michor, and Martin A Nowak. “Stochastic tunnels in evolutionary dynamics.” In: *Genetics* 166.3 (Mar. 2004), pp. 1571–1579. ISSN: 0016-6731. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1470783/ (visited on 08/16/2023).

[141] Ian E. Ochs and Michael M. Desai. “The competition between simple and complex evolutionary trajectories in asexual populations”. In: *BMC Evolutionary Biology* 15.1 (Mar. 2015), p. 55. ISSN: 1471-2148. DOI: 10.1186/s12862-015-0334-0. URL: https://doi.org/10.1186/s12862-015-0334-0 (visited on 08/14/2023).

[142] Sergey Gavrilets. “Perspective: Models of Speciation: What Have We Learned in 40 Years?” In: *Evolution* 57.10 (2003), pp. 2197–2215. ISSN: 0014-3820. URL: https://www.jstor.org/stable/3448772 (visited on 02/14/2024).

[143] Z. Chen and L. Kang. “Multi-Population Evolutionary Algorithm for Solving Constrained Optimization Problems”. In: *Artificial Intelligence Applications and Innovations (IFIP Conference)*. 2005.

[144] S. Chen et al. “Ultrahigh-throughput screening-assisted in vivo directed evolution for enzyme engineering”. In: *Biotechnology for Biofuels and Bioproducts* 19.9 (2024). DOI: 10.1186/s13068-024-02457-w.

[145] S. J. Rahi and V. Gligorovski. “Directed evolution of dynamic, multi-state, and computational proteins”. In: *Cell Press* 123.3 (2024). DOI: 10.1016/j.bpj.2023.11.335.

[146] Michael M. Desai and Daniel S. Fisher. “Beneficial Mutation-Selection Balance and the Effect of Linkage on Positive Selection”. In: *Genetics* 176.1 (2007), pp. 1759–1798. DOI: 10.1534/genetics.106.067678.

[147] Nadia M. V. Sampaio et al. “Dynamic gene expression and growth underlie cell-to-cell heterogeneity in Escherichia coli stress response”. In: *Proceedings of the National Academy of Sciences* 119.14 (Apr. 2022), e2115032119. DOI: 10.1073/pnas.2115032119. URL: https://www.pnas.org/doi/10.1073/pnas.2115032119 (visited on 10/08/2022).

[148] Paige Allard, Fotini Papazotos, and Laurent Potvin-Trottier. “Microfluidics for Long-Term Single-Cell Time-Lapse Microscopy: Advances and Applications”. In: *Frontiers in Bioengineering and Biotechnology* 10 (Oct. 2022), p. 968342. ISSN: 2296-4185. DOI: 10.3389/fbioe.2022.968342. (Visited on 09/11/2025).

[149] Thomas Lehnert and Martin A. M. Gijs. “Microfluidic systems for infectious disease diagnostics”. In: *Lab on a Chip* 24 (2024), pp. 1441–1493. DOI: 10.1039/D4LC00117F.

[150] Kyle M. Douglass et al. “Super-resolution Imaging of Multiple Cells by Optimized Flat-field Epi-illumination”. In: *Nature Photonics* 10.11 (2016), pp. 705–708. DOI: 10.1038/nphoton.2016.200.

[151] Dora Mahecic et al. “Homogeneous multifocal excitation for high-throughput super-resolution imaging”. In: *Nature methods* 17.7 (2020), pp. 726–733.

[152] Xueqiong Bai, Xiaoli Jing, and Ningfang Liao. “Design method for the high optical efficiency and uniformity illumination system of the projector”. In: *Optics Express* 29.8 (2021), pp. 12502–12515.

[153] Jui-Wen Pan and Hsiang-Hua Wang. “High contrast ratio prism design in a mini projector”. In: *Applied optics* 52.34 (2013), pp. 8347–8354.

[154] Dan Dan et al. “DMD-based LED-illumination Super-resolution and Optical Sectioning Microscopy”. In: *Scientific Reports* 3.1 (2013), p. 1116. DOI: 10.1038/srep01116.

[155] Owen M. O’Connor et al. “DeLTA 2.0: A deep learning pipeline for quantifying single-cell spatial and temporal dynamics”. In: *PLOS Computational Biology* 18.1 (Jan. 2022). ISBN: 1111111111, e1009797. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1009797.

[156] I. Kempf et al. *Fluorescence Images and Masks for Segmentation using DeLTA*. [Data set]. Zenodo, 2025. DOI: 10.5281/zenodo.17569655. URL: https://doi.org/10.5281/zenodo.17569655.

[157] Henry Pinkard et al. “Pycro-Manager: open-source software for customized and reproducible microscope control”. In: *Nature Methods* 18.3 (2021), pp. 226–228. DOI: 10.1038/s41592-021-01087-6. URL: https://doi.org/10.1038/s41592-021-01087-6.

[158] Herophilus. *ASITiger: A Python Interface for ASI Tiger Controllers*. https://github.com/herophilus/asitiger. Accessed: 2025-01-24. 2021.

[159] Texas Instruments Incorporated. *DLP LightCrafter Software and Firmware Bundle (DLPC900REF-SW)*. https://www.ti.com/tool/DLPC900REF

-SW [accessed: 2025-01-01]. Texas Instruments Incorporated. 2024. URL: https://www.ti.com/tool/DLPC900REF-SW.

[160] J. W. Kamande, Y. Wang, and A. M. Taylor. "Cloning SU8 silicon masters using epoxy resins to increase feature replicability and production for cell culture devices". In: *Biomicrofluidics* 9.3 (June 2015), p. 036502. ISSN: 1932-1058. DOI: 10.1063/1.4922962. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC4482805/ (visited on 12/17/2025).

[161] Simone Scalise et al. *A Mother-Machine Microfluidic Device for Non-Adherent Mammalian Cells Reveals the Population Growth Strategies*. Sept. 2025. DOI: 10.48550/arXiv.2509.06113. arXiv: 2509.06113 [q-bio]. (Visited on 09/09/2025).

[162] Michael J. Lawson et al. "In situ genotyping of a pooled strain library after characterizing complex phenotypes". eng. In: *Molecular Systems Biology* 13.10 (Oct. 2017), p. 947. ISSN: 1744-4292. DOI: 10.15252/msb.20177951.

[163] William G. Telford. "Flow cytometry and cell sorting". In: *Frontiers in Medicine* Volume 10 - 2023 (2023). DOI: 10.3389/fmed.2023.1287884. URL: https://www.frontiersin.org/journals/medicine/articles/10.3389/fmed.2023.1287884.

[164] Zhengqun Li, Yuting Deng, and Guang-Yu Yang. "Growth-coupled high throughput selection for directed enzyme evolution". eng. In: *Biotechnology Advances* 68 (Nov. 2023), p. 108238. ISSN: 1873-1899. DOI: 10.1016/j.biotechadv.2023.108238.

[165] Yuanyuan Han et al. “Review: imaging technologies for flow cytometry”. en. In: *Lab on a Chip* 16.24 (Nov. 2016), pp. 4639–4647. ISSN: 1473-0189. DOI: 10.1039/C6LC01063F. URL: https://pubs.rsc.org/en/content/articlelanding/2016/lc/c6lc01063f (visited on 07/25/2025).

[166] Nima Aghaeepour et al. “Critical assessment of automated flow cytometry data analysis techniques”. en. In: *Nature Methods* 10.3 (Mar. 2013), pp. 228–238. ISSN: 1548-7105. DOI: 10.1038/nmeth.2365. URL: https://www.nature.com/articles/nmeth.2365 (visited on 07/25/2025).

[167] Enrico Lugli, Mario Roederer, and Andrea Cossarizza. “Data analysis in flow cytometry: The future just started”. en. In: *Cytometry Part A* 77A.7 (2010), pp. 705–713. ISSN: 1552-4930. DOI: 10.1002/cyto.a.20901. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/cyto.a.20901 (visited on 07/25/2025).

[168] Etienne Fradet et al. “Measuring Fast and Slow Enzyme Kinetics in Stationary Droplets”. In: *Analytical Chemistry* 87.23 (Dec. 2015), pp. 11915–11922. ISSN: 0003-2700. DOI: 10.1021/acs.analchem.5b03567. URL: https://doi.org/10.1021/acs.analchem.5b03567 (visited on 07/25/2025).

[169] George Emanuel, Jeffrey R. Moffitt, and Xiaowei Zhuang. “High-throughput, image-based screening of pooled genetic-variant libraries”. en. In: *Nature Methods* 14.12 (Dec. 2017), pp. 1159–1162. ISSN: 1548-7105. DOI: 10.1038/nmeth.4495. URL: https://www.nature.com/articles/nmeth.4495 (visited on 07/25/2025).

[170] Maria Cristina Cannarsa et al. “Light-driven synchronization of optogenetic clocks”. en. In: *eLife* 13 (Aug. 2024). DOI: 10.7554/eLife.97754.2. URL: https://elifesciences.org/reviewed-preprints/97754 (visited on 07/25/2025).

[171] Stephan Uphoff. “Real-time dynamics of mutagenesis reveal the chronology of DNA repair and damage tolerance responses in single cells”. In: *Proceedings of the National Academy of Sciences* 115.28 (July 2018), E6516–E6525. DOI: 10.1073/pnas.1801101115. (Visited on 09/05/2025).

[172] Om Patange et al. “Escherichia coli can survive stress by noisy growth modulation”. In: *Nature Communications* 9.1 (Dec. 2018), p. 5333. ISSN: 2041-1723. DOI: 10.1038/s41467-018-07702-z. URL: http://www.nature.com/articles/s41467-018-07702-z.

[173] Amy Strohmeier Gort and James A. Imlay. “Balance between Endogenous Superoxide Stress and Antioxidant Defenses”. In: *Journal of Bacteriology* (1998). DOI: 10.1128/jb.180.6.1402-1410.1998.

[174] Reto Brem, Melisa Guven, and Peter Karran. “Oxidatively-generated damage to DNA and proteins mediated by photosensitized UVA”. In: *Free Radical Biology and Medicine* 107 (2017), pp. 101–109. DOI: 10.1016/j.freeradbiomed.2016.10.488.

[175] LÃlian S. Peloi et al. “Photodynamic effect of light-emitting diode light on cell growth inhibition induced by methylene blue”. en. In: *Journal of Bioscience* 33.2 (June 2008), pp. 231–237. ISSN: 0250-5991, 0973-7138. DOI: 10.1007/s12038-008-0040-9. URL: http://link.springer.com/10.1007/s12038-008-0040-9 (visited on 11/05/2025).

[176] S. O. Halperin et al. “CRISPR-guided DNA polymerases enable diversification of all nucleotides in a tunable window”. In: *Nature* 560 (2018), pp. 248–252. DOI: 10.1038/s41586-018-0384-8.

[177] YaoYao Liang et al. “Directed evolution of the PobR allosteric transcription factor to generate a biosensor for 4-hydroxymandelic acid”. eng. In: *World Journal of Microbiology & Biotechnology* 38.6 (May 2022), p. 104. ISSN: 1573-0972. DOI: 10.1007/s11274-022-03286-5.

[178] Mohamed A Nasr, Vincent J J Martin, and David H Kwan. “Divergent directed evolution of a TetR-type repressor towards aromatic molecules”. In: *Nucleic Acids Research* 51.14 (Aug. 2023), pp. 7675–7690. ISSN: 0305-1048. DOI: 10.1093/nar/gkad503. URL: https://doi.org/10.1093/nar/gkad503 (visited on 04/23/2025).