

LI Detector: a framework for sensitive colony-based screens regardless of the distribution of fitness effects

CEBOM
PITTSBURGH

LI Detector: a framework for sensitive colony-based screens regardless of the distribution of fitness effects

Saurin Bipin Parikh, Nelson Castilho Coelho, and Anne-Ruxandra Carvunis
Department of Computational & Systems Biology, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA
Center for Evolutionary Biology & Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA



Introduction

Colony-based high-throughput screens

CEBOM

LI Detector

Methods

The LI Detector framework for spatial bias correction

The experimental pipeline (Fig. 2) follows a four-step pipeline protocol that covers four purposes:

1. It isolates colony size differences that arise during the plating process, and
2. adds a reference colony grid [10] on every plate.

Figure 2. Schematic of colony-based high-throughput screen.

The problems of spatial bias

The environment can easily be contaminated consistently within a single plate, let alone across multiple plates, leading to systematic colony growth differences that are similar to infected colonies (Fig. 3).

Figure 3. Experimental pipeline of the LI Detector framework.

Results

Empirical strategy for LI Detector performance evaluation

We compare the performance of LI Detector with one of the most accurate and robust tools available for assessing spatial bias, ICIM [2].

We utilized the validation experiments in constructed datasets where the underlying CEBOM sets, known hot colony sites, were randomly affected by spatial bias, noise, and other technical artifacts of CEBOM (Fig. 4a,b).

1. To evaluate specificity, we constructed a “condition negative” dataset consisting of colony size measurements of our CEBOM diversity colonies at eleven time points (Fig. 4a).
2. To evaluate sensitivity, we constructed a “condition positive” dataset consisting of two sets of “virtual plates”:
 - The first set combined colony size estimates of the mock references and random bias in different locations so that the resulting virtual plates had isolated colony size distributions (Fig. 4b).
 - The second set combined the reference distribution from a single time point with random colony sizes from randomly chosen time points (Fig. 4c).

It is important to note that colony sizes of virtual plates retain the spatial biases in colony sizes because they simulate the original plate layout.

Figure 4. Experimental pipeline of the LI Detector framework.

Discussion

Salient features of the LI Detector

1. LI Detector is a CEBOM framework (Fig. 2, Fig. 3) that generates highly sensitive and sensitive fitness estimations without being dependent on a priori assumptions of the CEBOM (Fig. 3, Fig. 4a, Fig. 4b).
2. LI Detector is specifically designed to observe small deviations and beneficial changes in fitness. By observing 20% of the plate in reference unbiased fashion, LI Detector can detect fitness effects an increase 5% with 100% sensitivity (Fig. 4a, Fig. 4b).
3. LI Detector’s superior performance over existing methods (Fig. 4a, Fig. 4b) comes at the cost of having to integrate a reference colony grid, and therefore uses a higher number of plates to measure the same number of mutant colonies. However, the parallel nature of the LI Detector pipeline allows for a large number of replicates per strain to be handled according to the user’s requirement (Fig. 4c).
4. Lastly, by not relying on a priori assumptions of fitness distributions, which is ICIM [2] (Fig. 4a), LI Detector provides a flexible approach that can be applied to CEBOM independent of their scale.

Possible applications of the LI Detector

1. LI Detector can be a valuable method for improving the current gene-gene, gene-environment, and protein-protein interaction networks for colony-based fitness estimation for its ability to provide reliable and consistent fitness measurements.

LIVE SESSION CHAT CLOSED AUTHOR INFORMATION ABSTRACT REFERENCES CONTACT AUTHOR RATE ★ ★ ★ ★ ★ GET POSTER

Saurin Bipin Parikh, Nelson Castilho Coelho, and Anne-Ruxandra Carvunis

Department of Computational & Systems Biology, University of Pittsburgh School of Medicine,
Pittsburgh, PA, USA

Center for Evolutionary Biology & Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA,
USA



PRESENTED AT:

EMBL Conference

Molecular Mechanisms in Evolution and Ecology

EMBL



INTRODUCTION

Colony-based high-throughput screens

Colony-based high-throughput screens (CBHTS) have become increasingly popular due to the availability of model organism mutant collections and robotic equipment that help facilitate quicker and larger screens. Such screens have been used to explore gene-gene[1,2], gene-environment[3] and protein-protein[4] interactions, and very recently evolutionary biology[5].

A typical experiment generates multiple agar plates which are used to record growth using imagery that is then analyzed to result in a quantitative output of colony size as a proxy for fitness (**Fig. 1**).

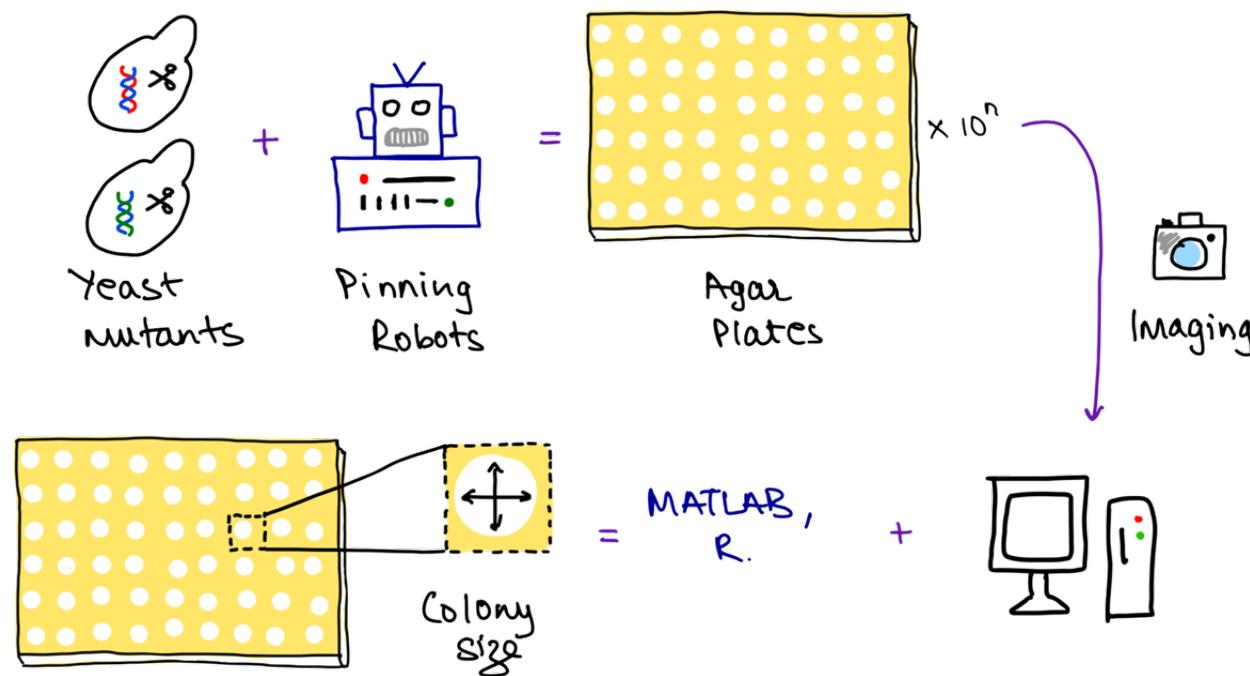


Figure 1. Schematic of a colony-based high-throughput screen.

The problem of spatial bias

The environment can rarely be maintained constant within a single plate, let alone across multiple plates, leading to systematic colony growth differences that arise due to technical reasons (**Fig. 2**).

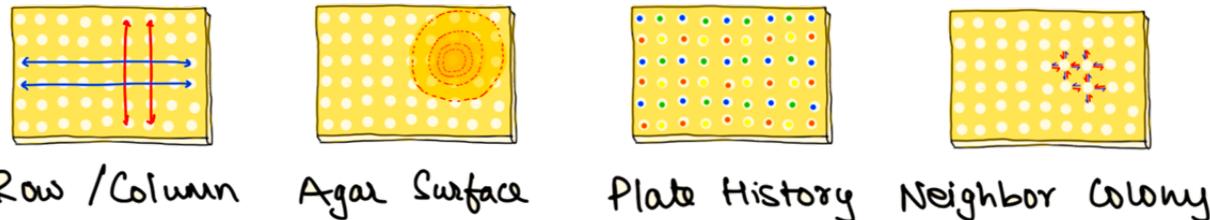


Figure 2. Various spatial biases that affect colony growth.

Spatial bias correction

Existing methods[6,7,8,9] that correct for spatial bias rely on the following assumptions (**Fig. 3**) about the distribution of fitness effects (DFE):

1. Genetic manipulations rarely cause significant fitness deviation from wildtype[7], and
2. The colony sizes in an experiment are normally distributed.

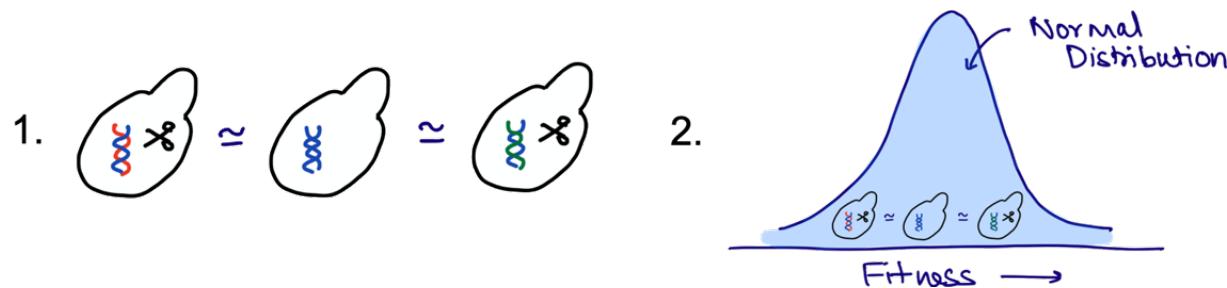


Figure 3. Two assumptions of existing methods for spatial bias correction.

Although reasonable in certain applications these assumptions are restrictive and limit the scope of the scientific examination, especially when it comes to detecting small increases in fitness.

With these limitations in mind we developed the linear interpolation-based detector (LI Detector or LID) framework. LI Detector consists of complementary experimental and analytical CBHTS pipelines that are specifically designed to control spatial bias and sensitively detect small significant changes in fitness, without making a priori assumptions about the underlying DFE of tested strains.

METHODS

The LI Detector framework for spatial bias correction

The experimental pipeline (**Fig. 4**) follows a pin-copy-upscale protocol that serves two purposes:

1. It reduces colony size differences that arise during the pinning process, and
2. adds a reference colony grid[10] on every plate.

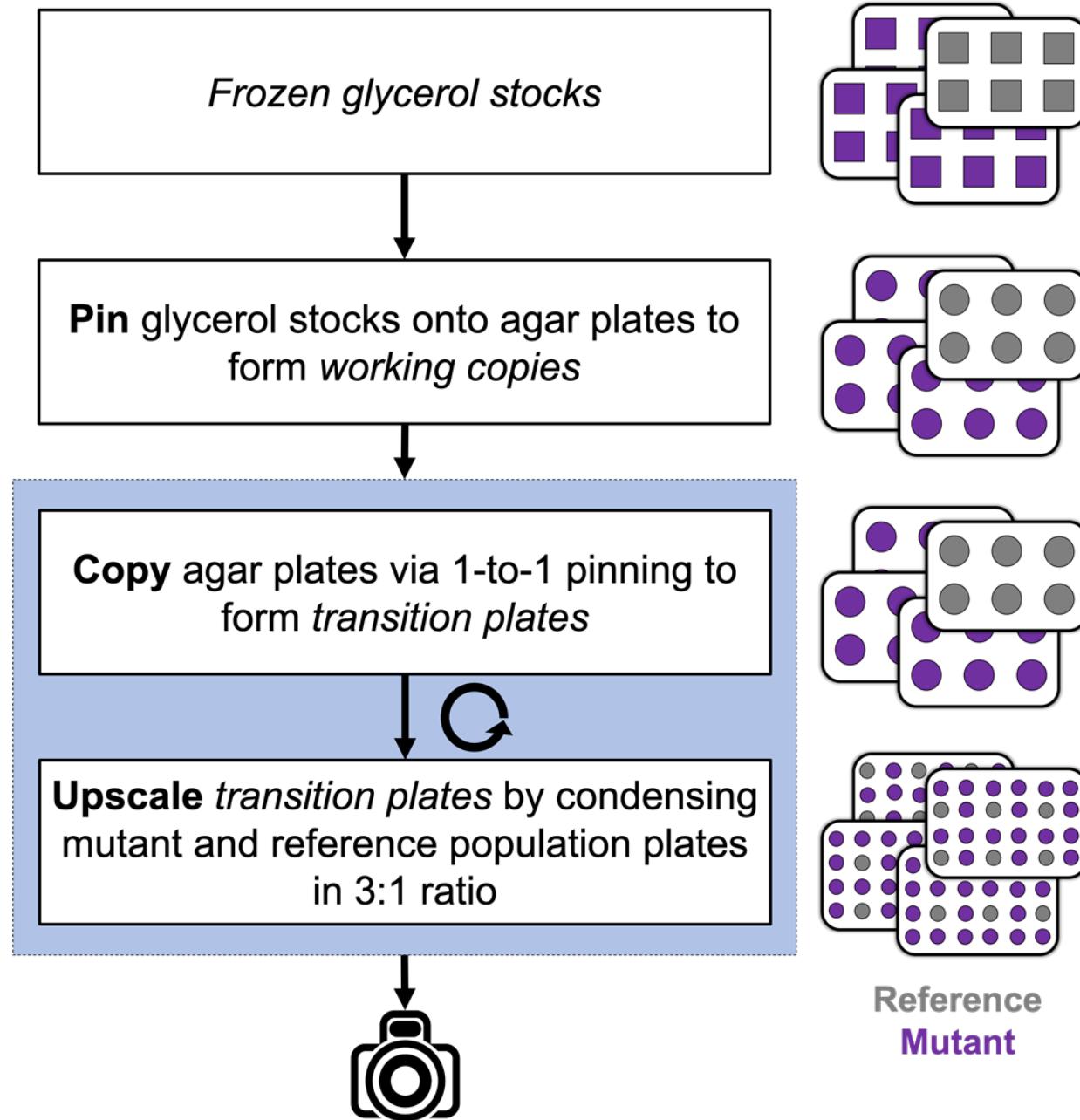


Figure 4. Experimental pipeline of the LI Detector framework.

The reference colony grid is utilized by the analytical pipeline to correct spatial bias and infer the relative fitness of mutant strains.

The analytical pipeline (Fig. 5) consists of five main steps:

1. Local artifact correction,
2. Source normalization,
3. Reference strain based background colony size estimation using a 2-dimensional linear interpolant,
4. Spatial bias correction by dividing the local artifact corrected colony sizes with the estimated background colony sizes and providing a measure of relative fitness, and
5. The relative fitness distribution of the reference strain is used as a null distribution to assign empirical p-values to mutant strains.

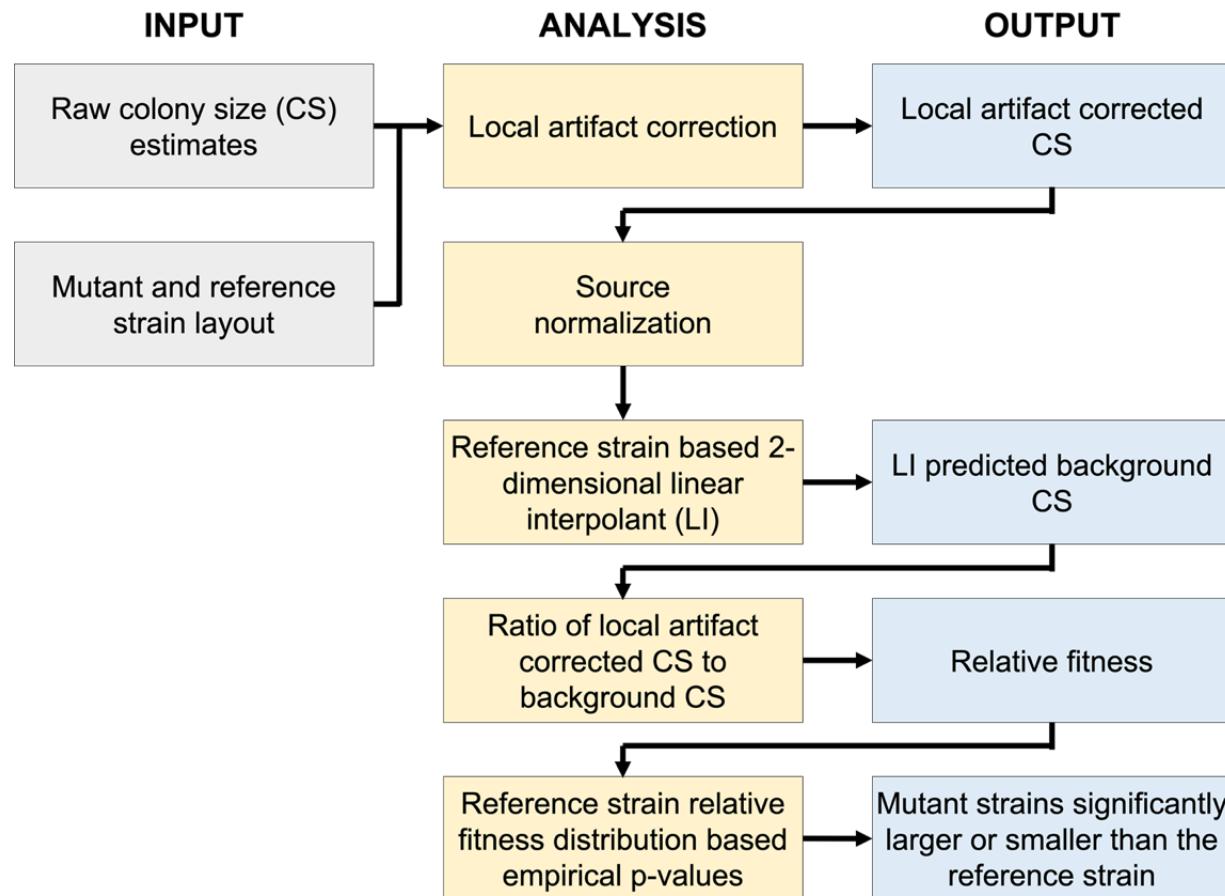


Figure 5. Analytical pipeline of the LI Detector framework.

Validation experiment

We applied the pin-copy-upscale experimental pipeline of our framework (**Fig. 4**) starting with four 384-well glycerol stock plates, each containing replicate frozen cultures of the same *Saccharomyces cerevisiae* strain (FY4).

This procedure generated agar plates (**Fig. 6a**) containing 16 replicate colonies for each individual culture in the starting glycerol stock plates. The sizes of these colonies were measured at eleven time points while they grew to saturation. **The colonies originating from one of the glycerol stock plates were treated as reference and the rest were treated as mutants (Fig. 6b).**

We then assembled these measurements into a series of datasets that we used to evaluate the specificity and sensitivity of the LI Detector.

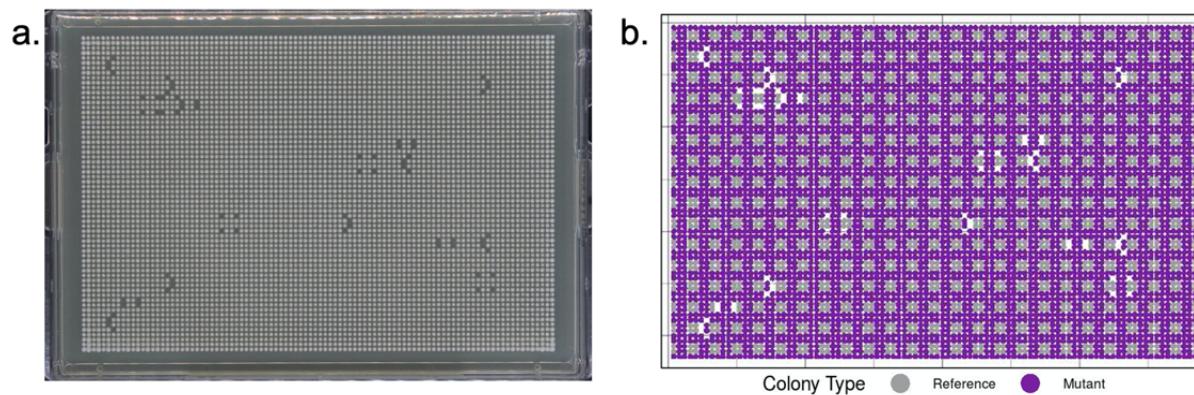


Figure 6. a. Picture of an agar plate with 6144 colony density, and b. its colony layout.

RESULTS

Empirical strategy for LI Detector performance evaluation

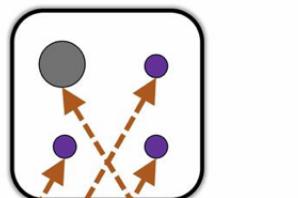
We compare the performance of LI Detector with one of the most versatile and robust tools available for correcting spatial bias, MCAT[6].

We utilized the validation experiment to construct datasets where the underlying DFE was known but colony sizes were realistically affected by spatial biases and other technical artifacts of CBHTS (**Fig. 7a,b,c**).

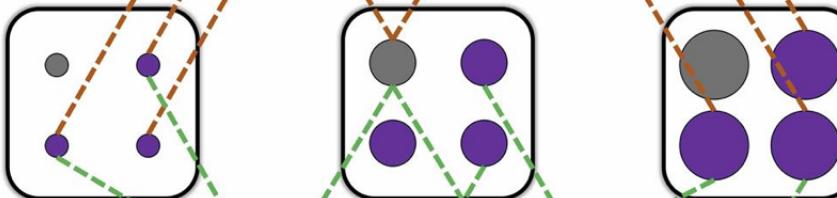
1. To estimate specificity, we assembled a "condition negative" dataset consisting of colony size measurements of our 6144-density plates at eleven time points (**Fig. 7b**).
2. To estimate sensitivity, we constructed a "condition positive" dataset consisting of two sets of "virtual plates":
 - o The first set combined colony size estimates of the mock references and mutants from two different time points, so that the resulting virtual plates had bimodal colony size distributions (**Fig. 7a**).
 - o The second set combined the reference distribution from a single time point with mutant colony sizes from randomly chosen time points (**Fig. 7c**).

It is important to note that both sets of virtual plates retain the spatial biases in colony sizes because they maintain the original plate layout.

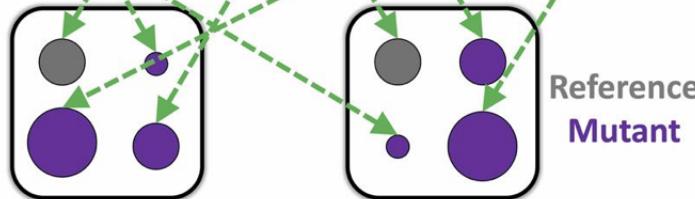
a. Virtual Plates with bimodal colony size distribution ($t_R > t_M$ or $t_R < t_M$) – 440 plates



b. Condition Negative Dataset ($t_R = t_M$) – 44 plates



c. Virtual Plates with random colony size distribution ($t_M \geq t_R > t_M$) – 44 plates



t_R = Reference colony-size time, t_M = Mutant colony-size time

Figure 7. Condition positive and negative datasets for empirical evaluation of the LI Detector.

The mutant strains were classified into beneficial, deleterious or neutral phenotypes depending on whether their relative fitness was identified as significantly higher, significantly lower or unchanged as compared to the reference distribution.

The LI Detector has high specificity

Specificity was calculated as the proportion of mutant strains that were correctly classified as neutral in the condition negative dataset.

LI Detector's specificity was more than 98% for a p-value cut off of 0.05 and remained above 95% when that cut-off was increased to a p-value of 0.1 (Fig. 8).

For comparison, MCAT[6] showed a maximum specificity of 94.5% for a p-value cut off of 0.05 using the same dataset (Fig. 8).

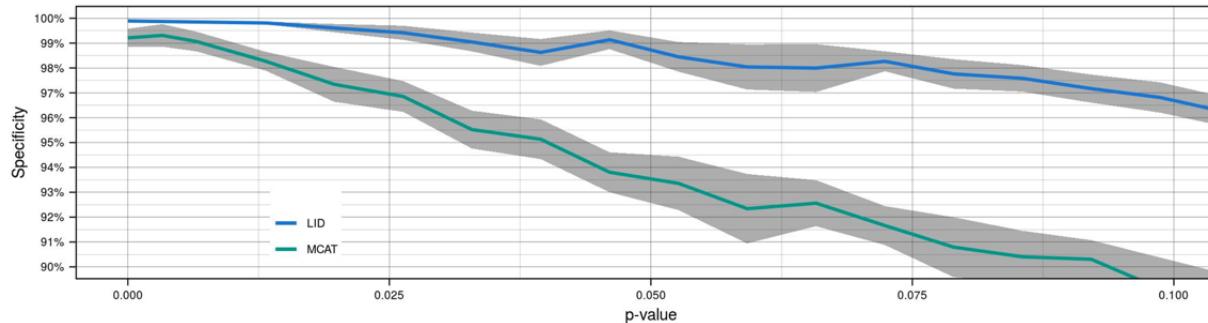


Figure 8. The LI Detector has high specificity.

The LI Detector identifies small fitness effects with high sensitivity

Sensitivity was estimated as the proportion of mock mutant strains correctly classified as either beneficial or deleterious at a false positive rate of 5% using our condition positive dataset with bimodal fitness distribution. The fitness effect were measured as a difference in the mean colony-sizes of the two distributions as a percentage of the reference distribution mean colony-size. This allowed us to evaluate sensitivity for a broad range of fitness effects.

LI Detector's sensitivity was higher than 95% for beneficial and deleterious fitness effects of 5%, and reached 100% for fitness effects of about 7% (Fig. 9a).

We also performed the same analysis using MCAT[6]. MCAT[6] was 80% sensitive in detecting 5% fitness decreases, and 40% sensitive when it came to 5% fitness increases (Fig. 9b).

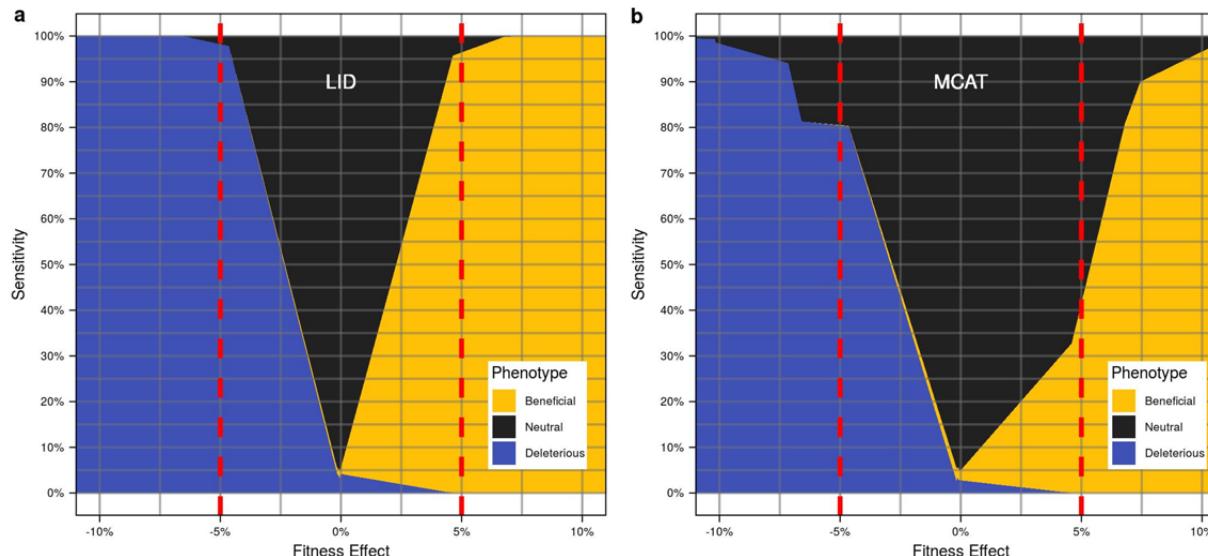


Figure 9. The LI Detector identifies small fitness effects with high sensitivity.

These findings show that, not only is the LI Detector highly sensitive in observing small fitness effects, but it is equally sensitive towards both increases and decreases in fitness.

The LI Detector maintains high sensitivity when the DFE is random

We designed the LI Detector to be highly sensitive regardless of the underlying DFE. To evaluate the LI Detector's performance when the underlying DFE is random, we used our virtual plates with random colony size distribution (**Fig. 7c**).

These virtual plates combined contained 41.60% beneficial, and 50.26% deleterious mutants with sixteen replicate colonies each (**Fig. 10a**).

We found that the LI Detector was 98.93% sensitive, successfully identifying 98.65% beneficial and 99.20% deleterious mutants (**Fig. 10b**).

In comparison, MCAT[6] was 83.08% sensitive and successful in identifying 82.76% beneficial and 83.40% deleterious (**Fig. 10c**).

It was also observed that the LI Detector's neutral calls were mostly limited to fitness effects of 5% or smaller, whereas MCAT[6] neutral calls covered a wider range of fitness effects (data not shown).

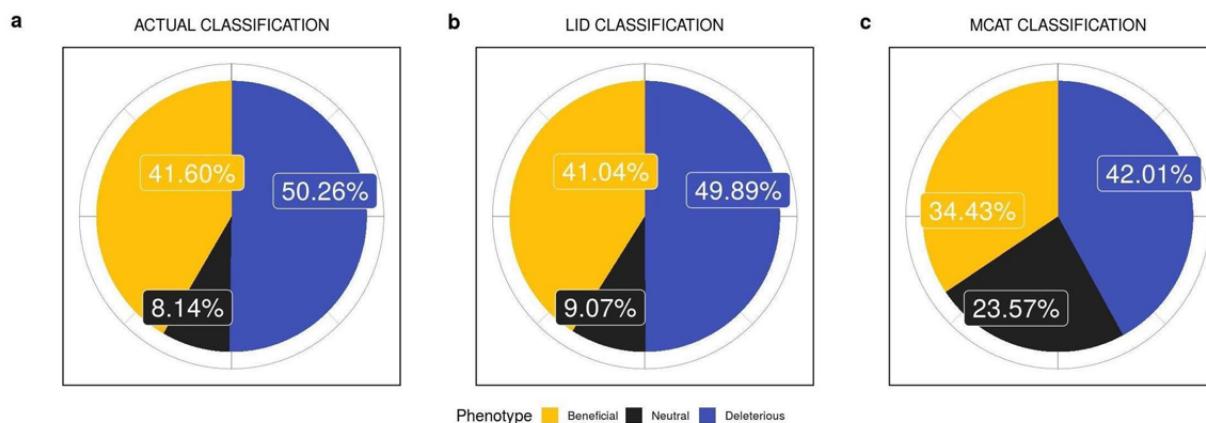


Figure 10. The LI Detector maintains high sensitivity even when underlying DFE is random.

That MCAT[6] was significantly less accurate than the LI Detector in this scenario was not surprising, since a random underlying distribution of fitness effects violates the assumptions of MCAT[6] and other existing methods.

The LI Detector's sensitivity increases with increasing the number of references and replicates

We analyzed how the number of references per plate and the number of replicates per strain affected the sensitivity of the LI Detector. To do this, we computationally masked portions of the reference-colony grid and replicates, and then reanalyzed the condition positive dataset.

We observed that the LI Detector's sensitivity in detecting 5% fitness effects increased with an increasing proportion of reference colonies per plate and the number of replicates per strain in both sets of virtual plates (**Fig. 11**). Unsurprisingly, sensitivity was higher for detecting a fitness effect of 7% (data not shown).

Increasing number of replicates was most powerful when there are more references on the plate. The sensitivity was, in general, highest in the virtual plates with bimodal distribution (**Fig. 11**).

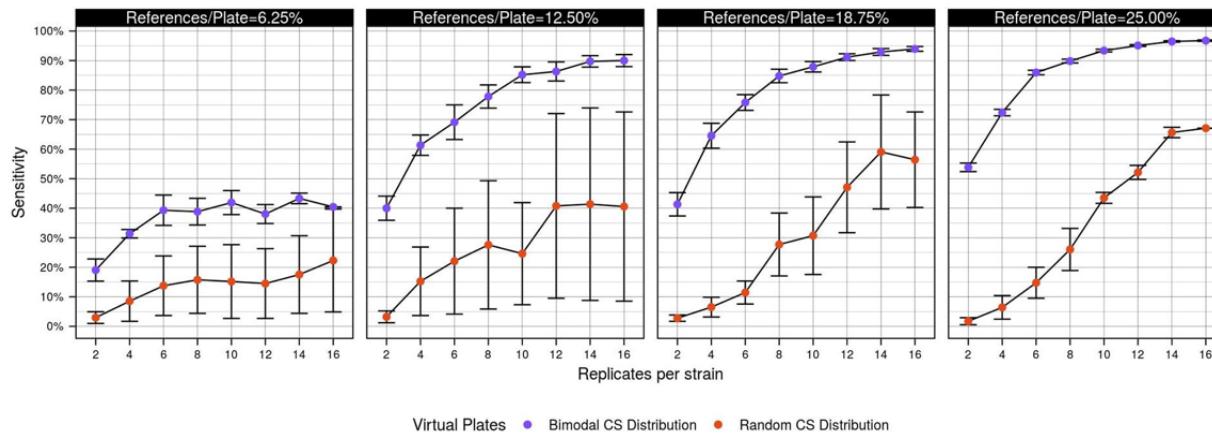


Figure 11. Sensitivity is directly related to the number of references and replicates.

It must be noted that the cost of reducing the number of references is lower for detecting more substantial fitness effects.

DISCUSSION

Salient features of the LI Detector

1. LI Detector is a CBHTS framework (**Fig. 4**, **Fig. 5**) that generates highly specific and sensitive fitness estimations without being dependent on *a priori* assumptions of the DFE (**Fig. 8**, **Fig. 9a**, **Fig. 10b**).
2. LI Detector is specifically designed to observe small deleterious and beneficial changes in fitness. By dedicating 25% of the plate to reference colonies and having 16 replicates per strain, we show that LI Detector can observe fitness effects as low as 5% with 95% sensitivity (**Fig. 9a**, **Fig. 11**).
3. LI Detector's superior performance over existing methods (**Fig. 9a**, **Fig. 10b**) comes at the cost of having to integrate a reference colony grid, and therefore use a higher number of plates to screen the same number of mutant colonies. However, the proportion of references per plate and the number of replicates per strain can be tunable according to the user's requirement (**Fig. 11**).
4. Lastly, by not relying on *a priori* assumptions of distribution of fitness effects (DFE) (**Fig. 3**), LI Detector provides a flexible approach that can be applied to CBHTS independent of their scale.

Possible applications of the LI Detector

1. LI Detector can be a valuable method for improving the current gene-gene, gene-environment and, protein-protein interaction networks for colony-forming-microorganisms due to its ability to provide reliable and well-resolved fitness measurements.
2. LI Detector's ability to detect small increases in fitness in particular makes it a favorable method to examine gain-of-function mutations, questions of evolutionary biology, and pharmacological screens of adaption and resistance[3,5,11-16].
3. LI Detector offers a tunable system where users may choose the number of references and replicates adequate for their purposes depending on the fitness effects they expect to observe and the sensitivity they aim to achieve.
4. Freedom of scale increases the applicability of CBHTS to situations where a large number of mutants or genome-wide assays aren't required. For example, LI Detector can be used as efficiently for a highly biased screen of non-synonymous mutations in a single gene to identify important residues[17-22], as for a genome-wide synthetic genetic array used to infer the genetic interactions[2,23-26].

Summary

The LI Detector framework experimentally introduces a reference population grid on plates whose colony size estimates are used to correct for spatial bias independently of the underlying DFE.

It offers a tunable system that has the potential to expand the utility of CBHTS by making them:

- Independent of scale,
- Sensitive towards small fitness effects, and
- Equally sensitive in detecting increases and decreases in fitness.

These features make LI Detector an effective method to detect small effects, including adaptive ones, that can help reveal interesting evolutionary phenomenon.

Although developed and validated using yeast, it can be applied to any colony-forming-microorganisms, including clinically relevant isolates, as long as they can be grown in the laboratory.

LI Detector preprint is available on bioRxiv. (<https://doi.org/10.1101/2020.06.27.175216>)

AUTHOR INFORMATION

Saurin Bipin Parikh, MBBS, MS

PhD Candidate, Integrative Systems Biology Program (<https://www.isb.pitt.edu>)

Department of Computational and Systems Biology (<https://www.csb.pitt.edu>)

University of Pittsburgh School of Medicine (<https://www.medschool.pitt.edu>), Pittsburgh, PA, USA

email - sbp29@pitt.edu

twitter - [@sauriiin](https://twitter.com/sauriiin) (<https://twitter.com/sauriiin>)

github - [sauriiin](https://github.com/sauriiin) (<https://github.com/sauriiin>)

Nelson Castilho Coelho, MS

Senior Research Specialist

Department of Computational and Systems Biology (<https://www.csb.pitt.edu>)

University of Pittsburgh School of Medicine (<https://www.medschool.pitt.edu>), Pittsburgh, PA, USA

email - nelsoncastilho@gmail.com

twitter - [@NelsioC](https://twitter.com/NelsioC) (<https://twitter.com/NelsioC>)

Anne-Ruxandra Carvunis, PhD

Assistant Professor

Department of Computational and Systems Biology (<https://www.csb.pitt.edu>)

University of Pittsburgh School of Medicine (<https://www.medschool.pitt.edu>), Pittsburgh, PA, USA

email - carvunis@gmail.com

twitter - [@carvunis](https://twitter.com/carvunis) (<https://twitter.com/carvunis>)

website - www.carvunis.com (<https://www.csb.pitt.edu/faculty/carvunis/>)

ABSTRACT

Natural selection cannot act and adaptation cannot occur without changes in fitness. Colony-based high-throughput screens use microbial growth on agar plates as a proxy for fitness, enabling the parallel analysis of thousands of individual strains. However, fitness estimation is complicated by spatial factors affecting colony growth, including uneven nutrient distribution, agar surface irregularities, and batch effects. To correct for these spatial biases, several analytical methods have been developed which rely on the following assumptions: i) that fitness effects are normally distributed, and ii) that most genetic perturbations lead to minor changes in fitness. These assumptions, although reasonable for many genome-wide screens, are not always warranted and can reduce the ability to detect small fitness effects. Beneficial fitness effects, in particular, are notoriously difficult to detect using these assumptions. Here, we developed the linear interpolation-based detector (LI Detector) framework to enable sensitive colony-based screening without making prior assumptions about the underlying distribution of fitness effects. The LI Detector uses a grid of reference colonies on each agar plate to predict and correct spatial biases. It assigns a relative fitness to all colonies as a measure of their size compared to the neighboring reference colonies. We show that the LI Detector can identify both increases and decreases in fitness effects as low as 5% with more than 95% sensitivity and specificity. These findings make LI Detector an effective method to detect small effects, including adaptive ones, that can help reveal interesting evolutionary phenomenon.

REFERENCES

1. Roguev A, et al. Conservation and rewiring of functional modules revealed by an epistasis map in fission yeast. *Science*. 2008;322(5900):405-10.
2. Costanzo M, et al. The genetic landscape of a cell. *Science*. 2010;327(5964):425-31.
3. Parsons AB, et al. Integration of chemical-genetic and genetic interaction data links bioactive compounds to cellular target pathways. *Nat Biotechnol*. 2004;22(1):62-9.
4. Yang F, et al. Development and application of a recombination-based library versus library high- throughput yeast two-hybrid (RLL-Y2H) screening system. *Nucleic Acids Res*. 2018;46(3):e17
5. Vakirlis N, et al. De novo emergence of adaptive membrane proteins from thymine-rich genomic sequences. *Nat Comm*. 2019. (in press)
6. Bean GJ, et al. Development of ultra-high-density screening tools for microbial "omics". *PLoS One*. 2014;9(1):e85177.
7. Wagih O, et al. SGAtools: one-stop analysis and visualization of array-based genetic interaction screens. *Nucleic Acids Res*. 2013;41(Web Server issue):W591-6.
8. Young BP, et al. Balony: a software package for analysis of data generated by synthetic genetic array experiments. *BMC Bioinformatics*. 2013;14:354.
9. Dittmar JC, et al. ScreenMill: a freely available software suite for growth measurement, analysis and visualization of high-throughput screen data. *BMC Bioinformatics*. 2010;11:353.
10. Zackrisson, M., et al., Scan-o-matic: High-Resolution Microbial Phenomics at a Massive Scale. *G3 (Bethesda)*, 2016. 6(9): p. 3003-14.
11. Piotrowski, J.S., et al., Functional annotation of chemical libraries across diverse biological processes. *Nat Chem Biol*, 2017. 13(9): p. 982-993.
12. Hoepfner, D., et al., High-resolution chemical dissection of a model eukaryote reveals targets, pathways and gene functions. *Microbiol Res*, 2014. 169(2-3): p. 107-20.
13. Lee, A.Y., et al., Mapping the cellular response to small molecules using chemogenomic fitness signatures. *Science*, 2014. 344(6180): p. 208-11.
14. Nichols, R.J., et al., Phenotypic landscape of a bacterial cell. *Cell*, 2011. 144(1): p. 143-56.
15. Tong, A.H., et al., Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science*, 2001. 294(5550): p. 2364-8.
16. Helsen, J., et al., Gene loss predictably drives evolutionary adaptation. *Mol Biol Evol*, 2020.
17. Sharon, E., et al., Functional Genetic Variants Revealed by Massively Parallel Precise Genome Editing. *Cell*, 2018. 175(2): p. 544-557 e16.
18. Bao, Z., et al., Genome-scale engineering of *Saccharomyces cerevisiae* with single-nucleotide precision. *Nat Biotechnol*, 2018. 36(6): p. 505-508.

19. Garst, A.D., et al., Genome-wide mapping of mutations at single-nucleotide resolution for protein, metabolic and genome engineering. *Nat Biotechnol*, 2017. 35(1): p. 48-55.
20. Guo, X., et al., High-throughput creation and functional profiling of DNA sequence variant libraries using CRISPR-Cas9 in yeast. *Nat Biotechnol*, 2018. 36(6): p. 540-546.
21. Sadhu, M.J., et al., Highly parallel genome variant engineering with CRISPR-Cas9. *Nat Genet*, 2018. 50(4): p. 510-514.
22. Roy, K.R., et al., Multiplexed precision genome editing with trackable genomic barcodes in yeast. *Nat Biotechnol*, 2018. 36(6): p. 512-520.
23. Costanzo, M., et al., A global genetic interaction network maps a wiring diagram of cellular function. *Science*, 2016. 353(6306).
24. Zimmermann, C., et al., Mapping the Synthetic Dosage Lethality Network of CDK1/CDC28. *G3 (Bethesda)*, 2017. 7(6): p. 1753-1766.
25. Klobucar, K. and E.D. Brown, Use of genetic and chemical synthetic lethality as probes of complexity in bacterial cell systems. *FEMS Microbiol Rev*, 2018. 42(1).
26. Kuzmin, E., et al., Exploring whole-genome duplicate gene retention with complex genetic interaction analysis. *Science*, 2020. 368(6498)