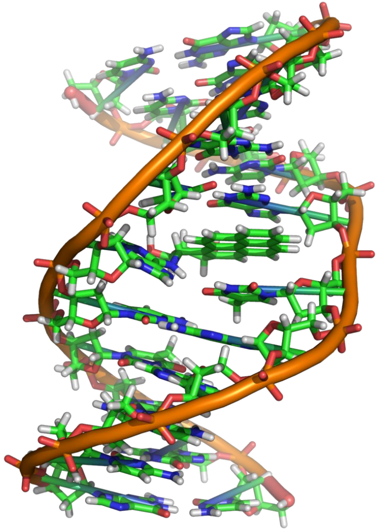
UNDERSTANDING THE DISTRIBUTION OF FITNESS EFFECTS (DFE) ACROSS ENVIRONMENTS

EXPLORING THE STATISTICAL PROPERTIES OF FITNESS LANDSCAPES USING THE FISHER’S GEOMETRIC MODEL

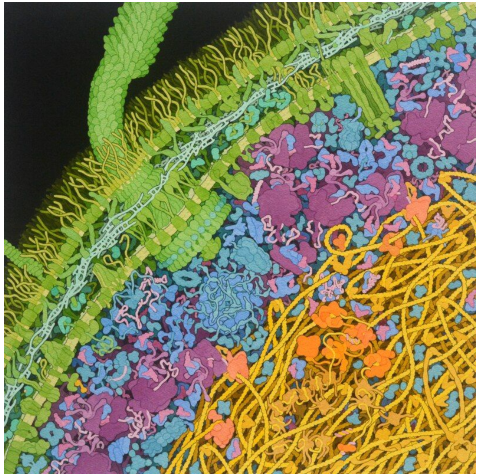
Shahnewaz Ahmed

Fundamental Scales of Living Matter

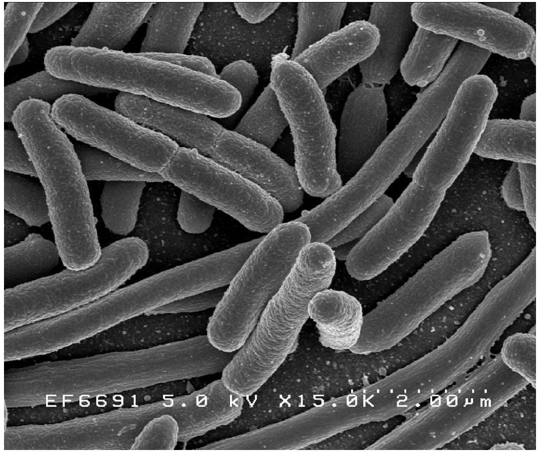
Genotype
(Molecular/Microscopic scale)



Phenotype
(Mesoscopic/Cellular scale)



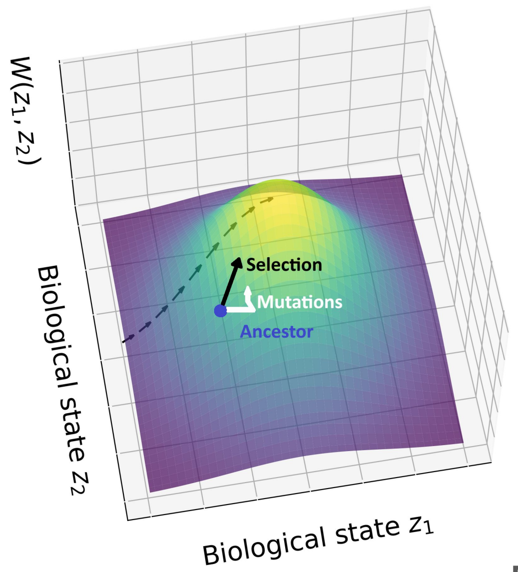
Fitness
(Macroscopic/Population scale)



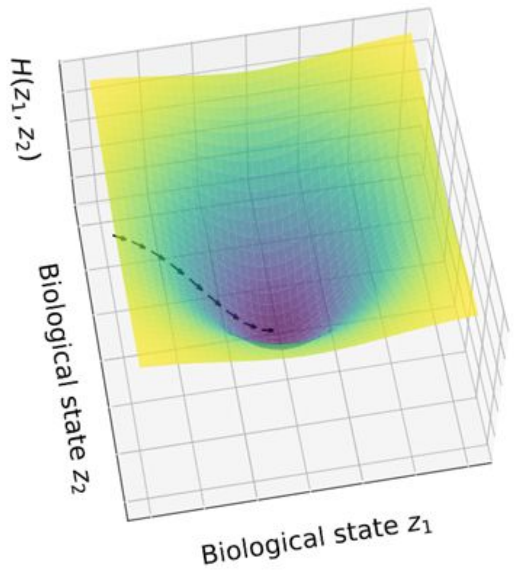
Source: Wikipedia

Evolutionary Analog of Statistical Mechanics

Fitness function



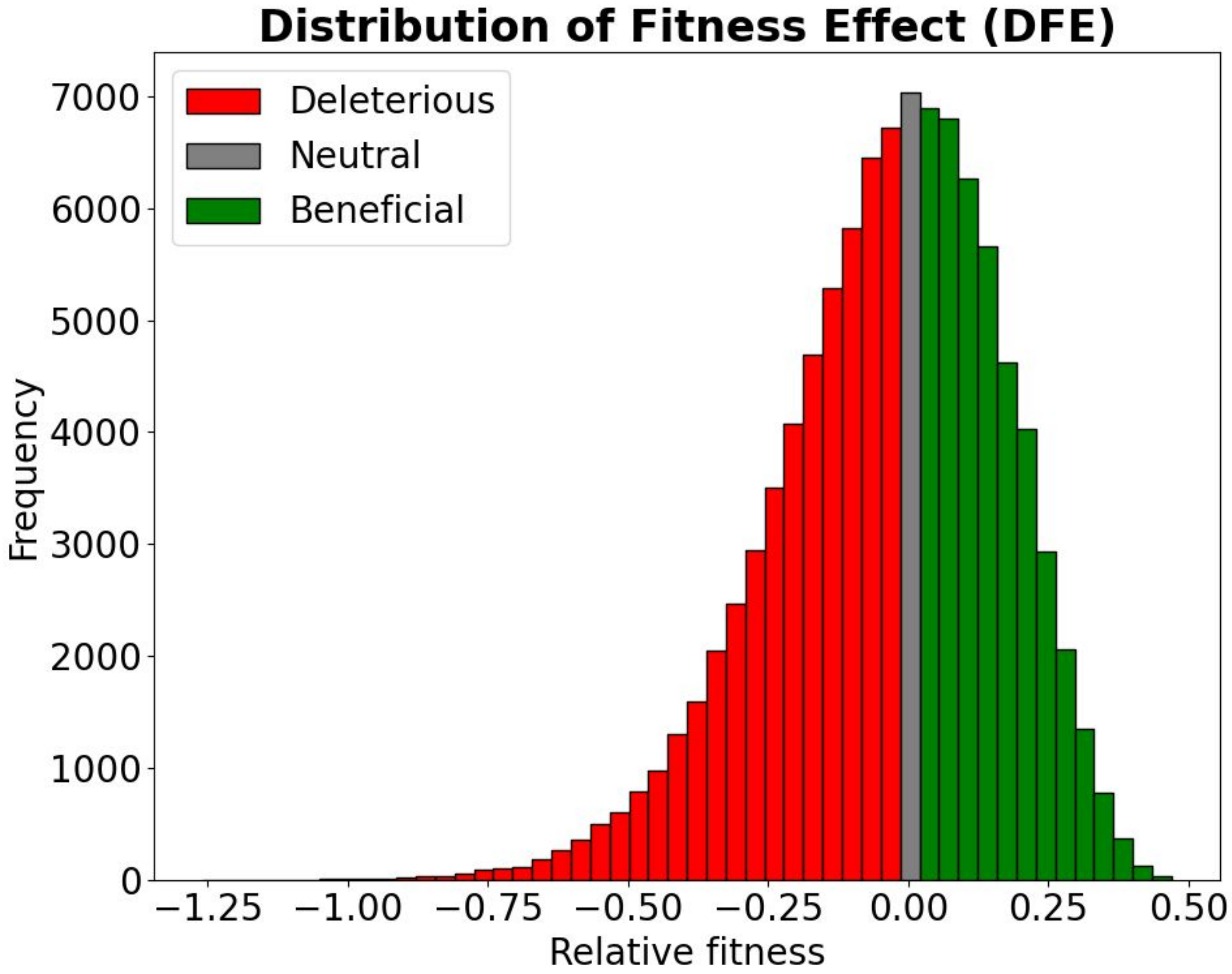
Potential function



Fitness ⟺ Energy

Mutations ⟺ Entropic force

Finite size fluctuation ⟺ Thermal fluctuation



Distribution of Fitness Effect (DFE)
Deleterious
Neutral
Beneficial
Frequency
0
1000
2000
3000
4000
5000
6000
7000
−1.25
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
Relative fitness

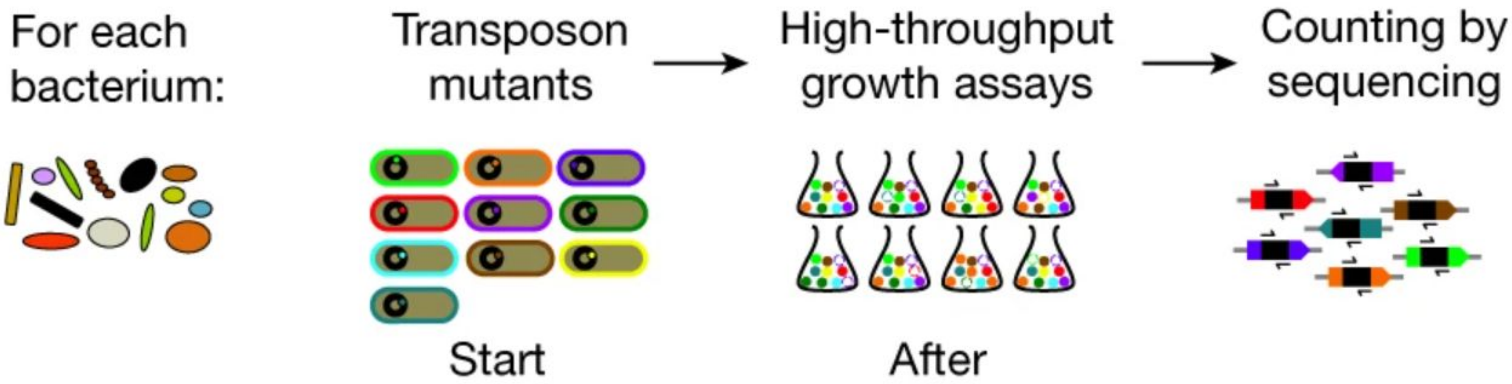
Importance of DFE

- Understanding what kind of effect mutation have on species especially on humans
- Central to questions in evolutionary biology
- Quantitative genetic variation

Key Questions

* How does the distribution of fitness effects (DFE) change across different environments?
* Are the mean and variance of the DFE more sensitive to environmental changes than the overall shape?
* What is the correlation of fitness effects for mutations across different environments?

Experiment

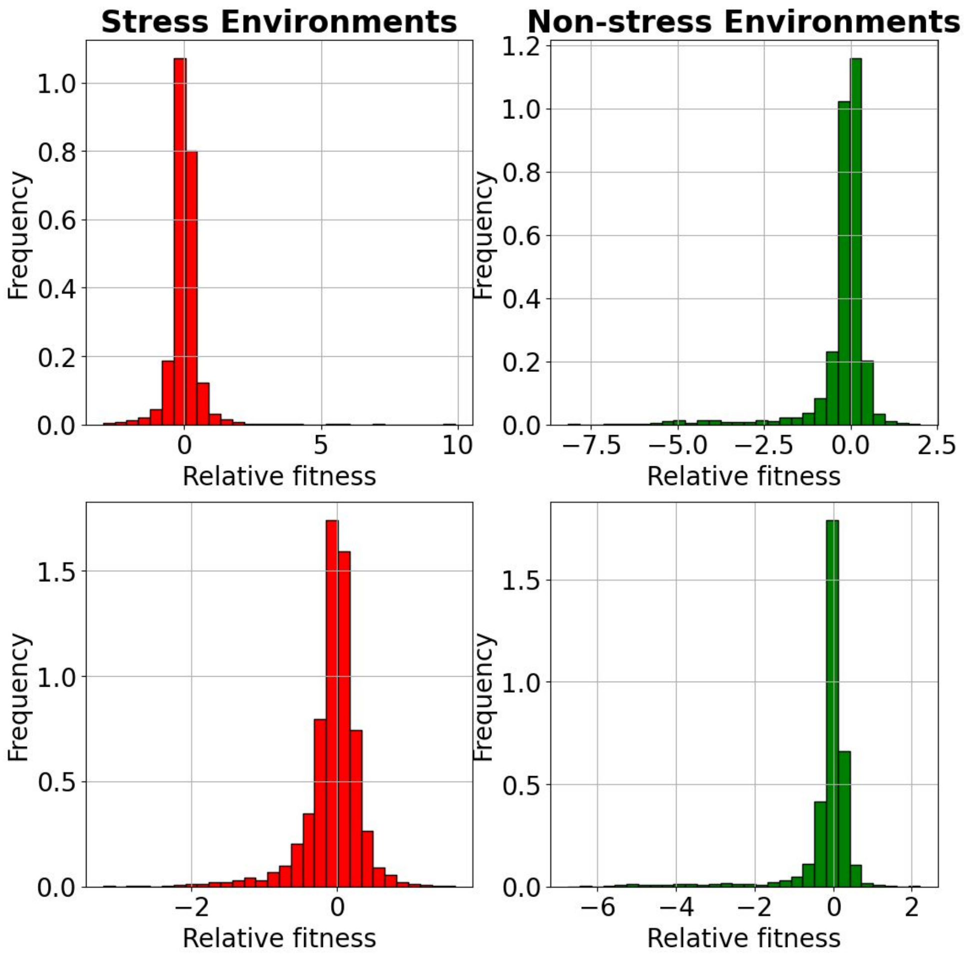


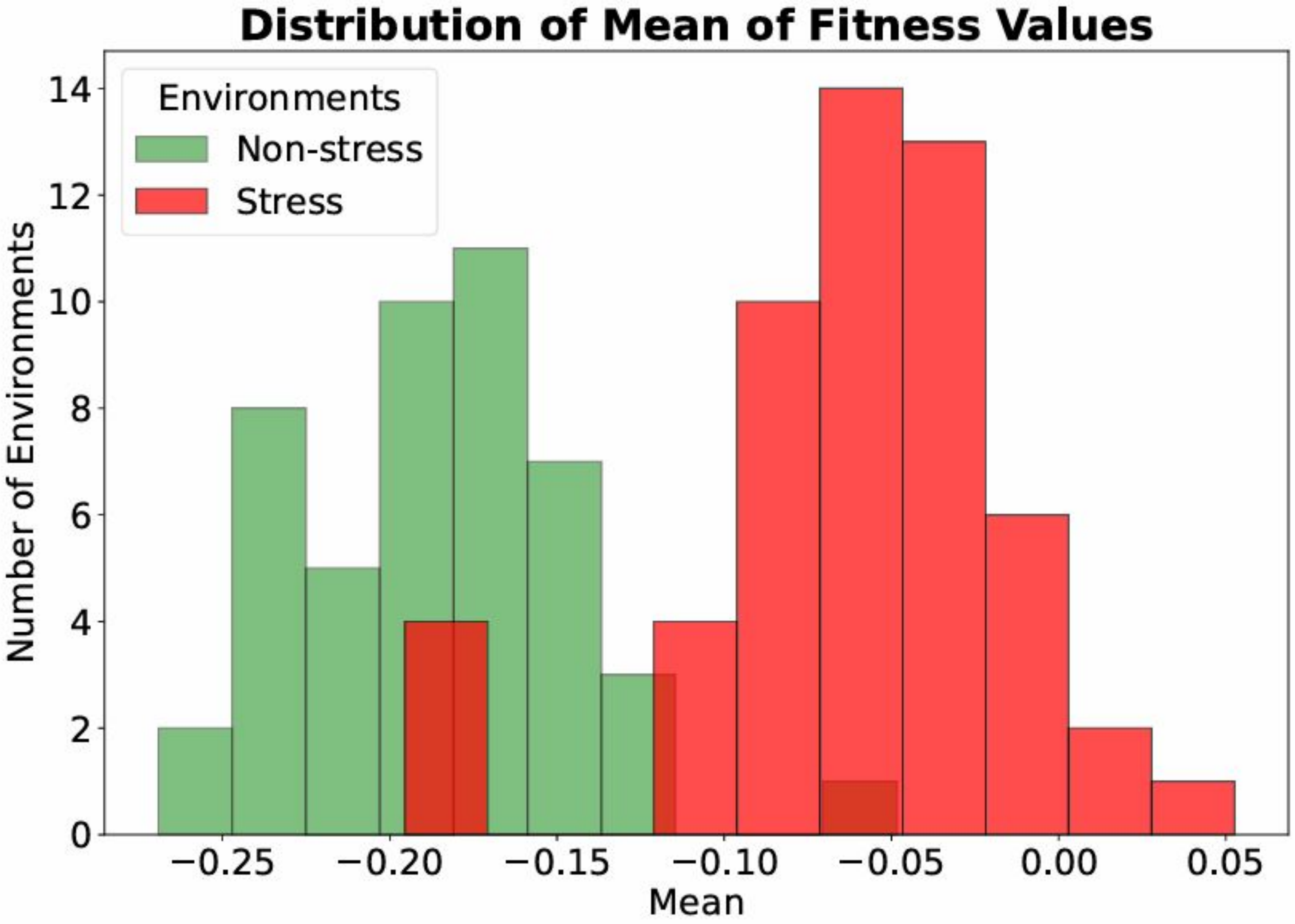
$$\text{Mutant fitness} = \log_2\left(\frac{\text{abundance after}}{\text{abundance at the start}}\right)$$

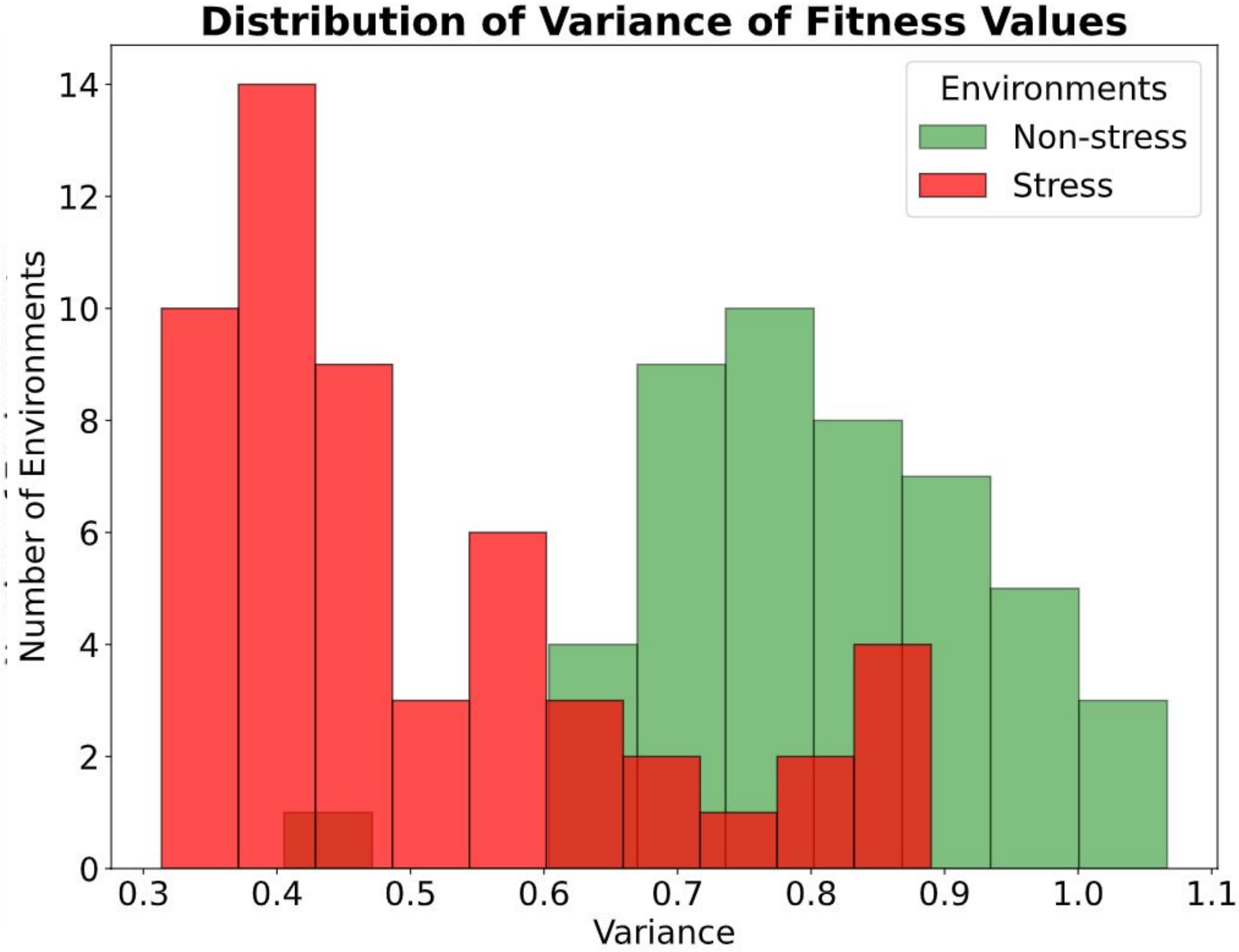
Gene Name	D-Glucose (C)	Spectinomycin 0.0125 mg/ml
talB	-0.001	-2.712
yaaJ	-0.127	0.275
polB	-0.055	-0.328

Source: Price et. al. 2018

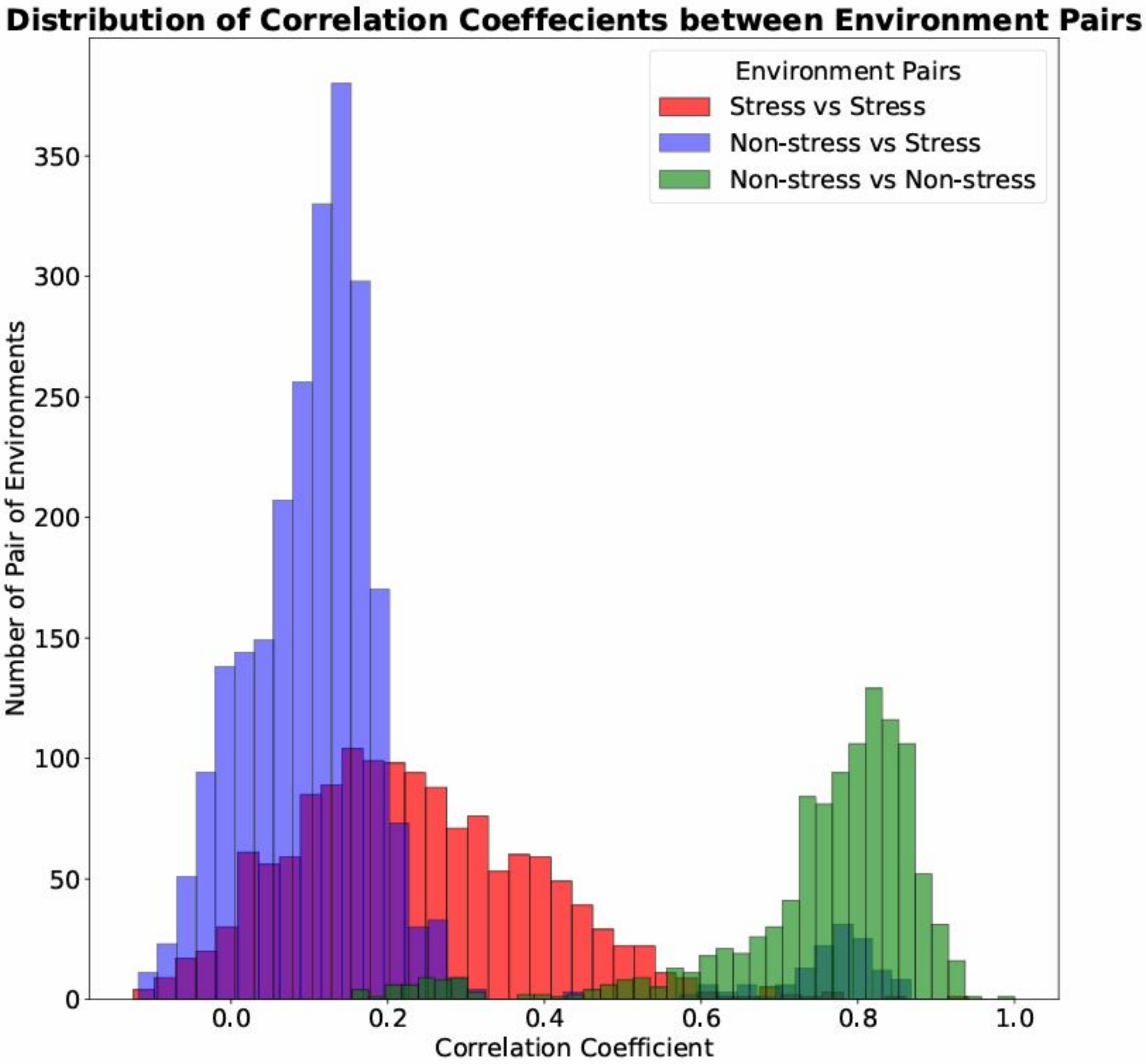
Distribution of Fitness Effect from Experimental Data



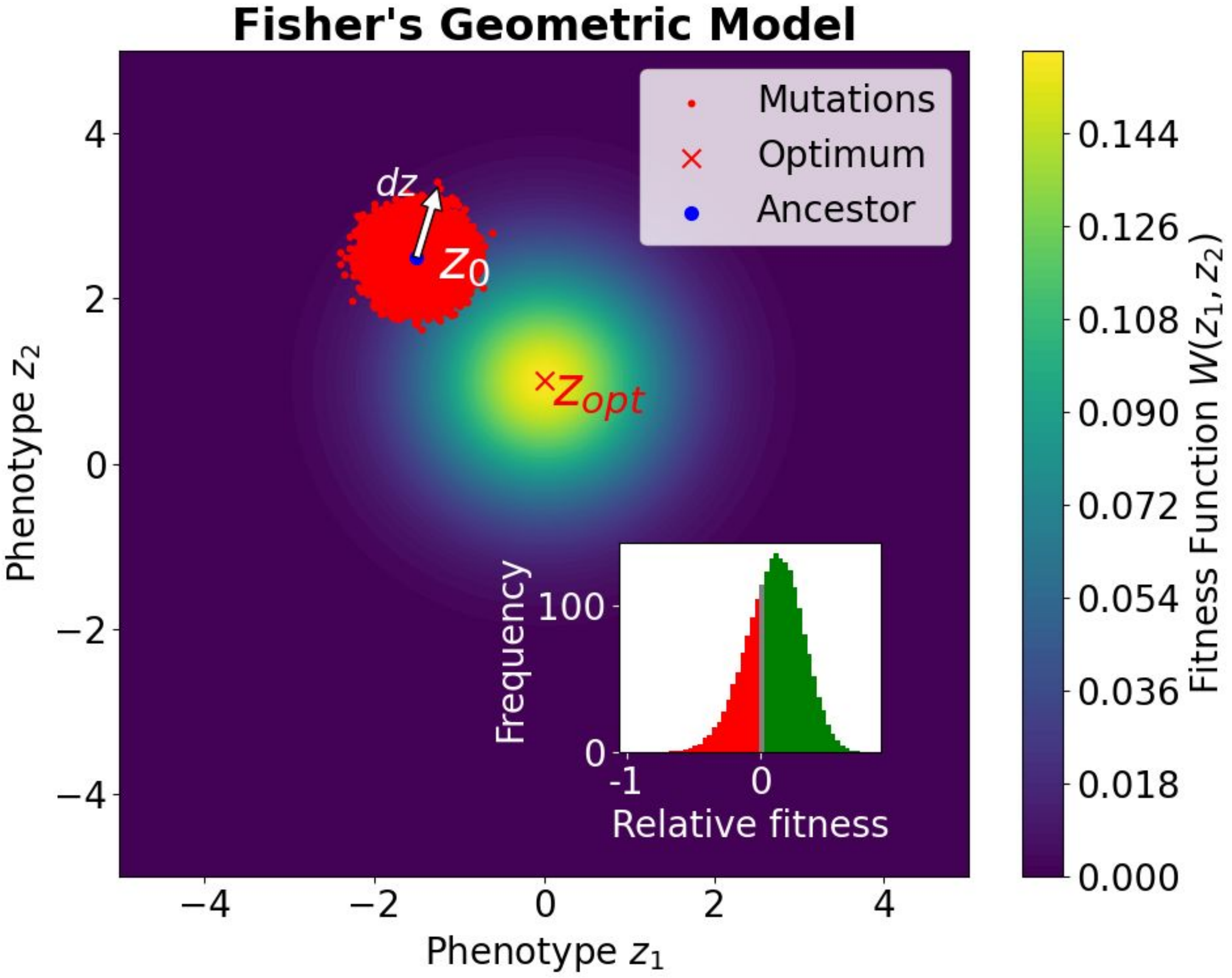

Distribution of Mean of Fitness Values
Environments
Non-stress
Stress
Number of Environments
Mean
0
2
4
6
8
10
12
14
−0.25
−0.20
−0.15
−0.10
−0.05
0.00
0.05



Distribution of Variance of Fitness Values
Environments
Non-stress
Stress
Number of Environments
Variance
0
2
4
6
8
10
12
14
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
1.1



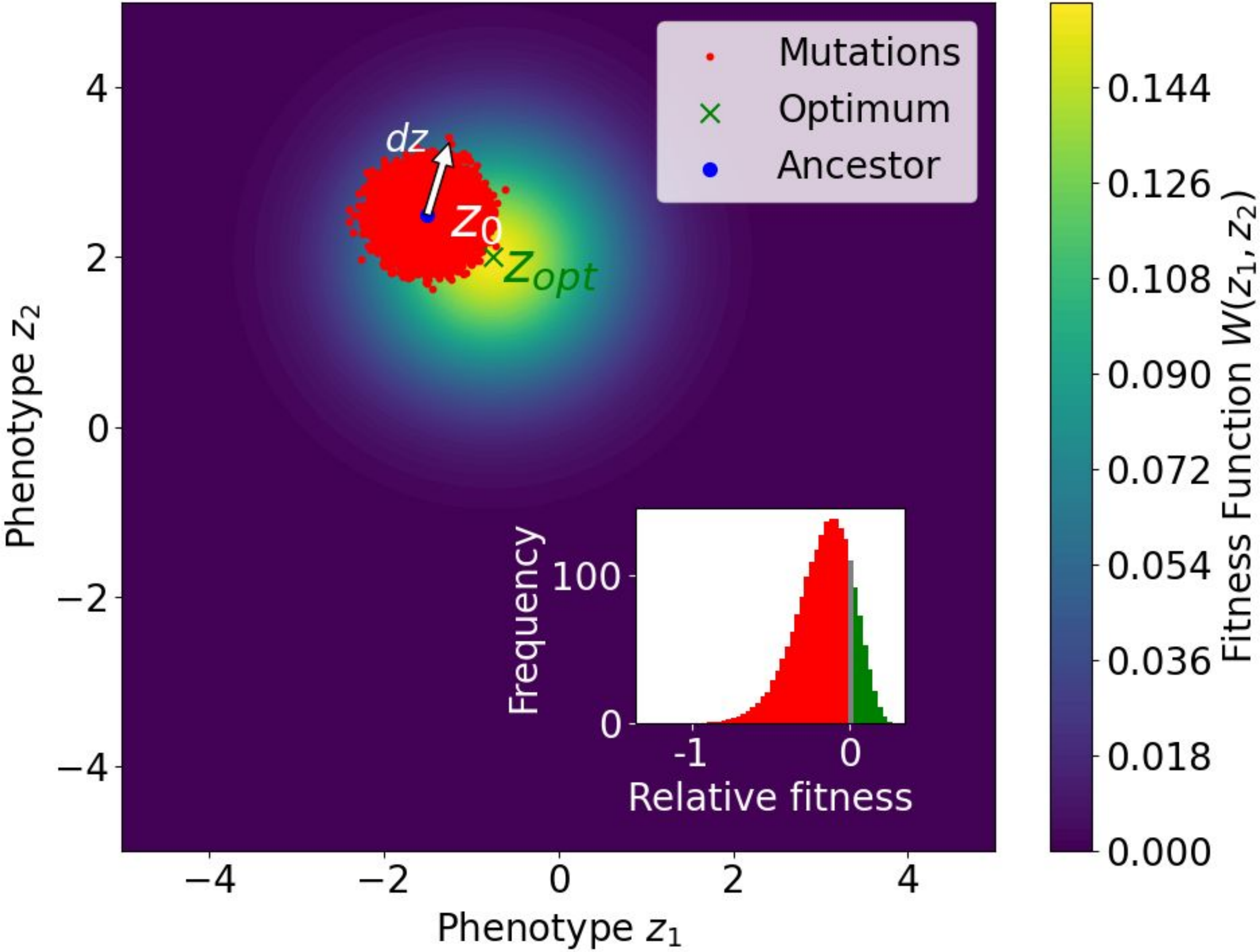
Distribution of Correlation Coeffecients between Environment Pairs
Environment Pairs
Stress vs Stress
Non-stress vs Stress
Non-stress vs Non-stress
Number of Pair of Environments
Correlation Coefficient
0
50
100
150
200
250
300
350
0.0
0.2
0.4
0.6
0.8
1.0



Relative Fitness

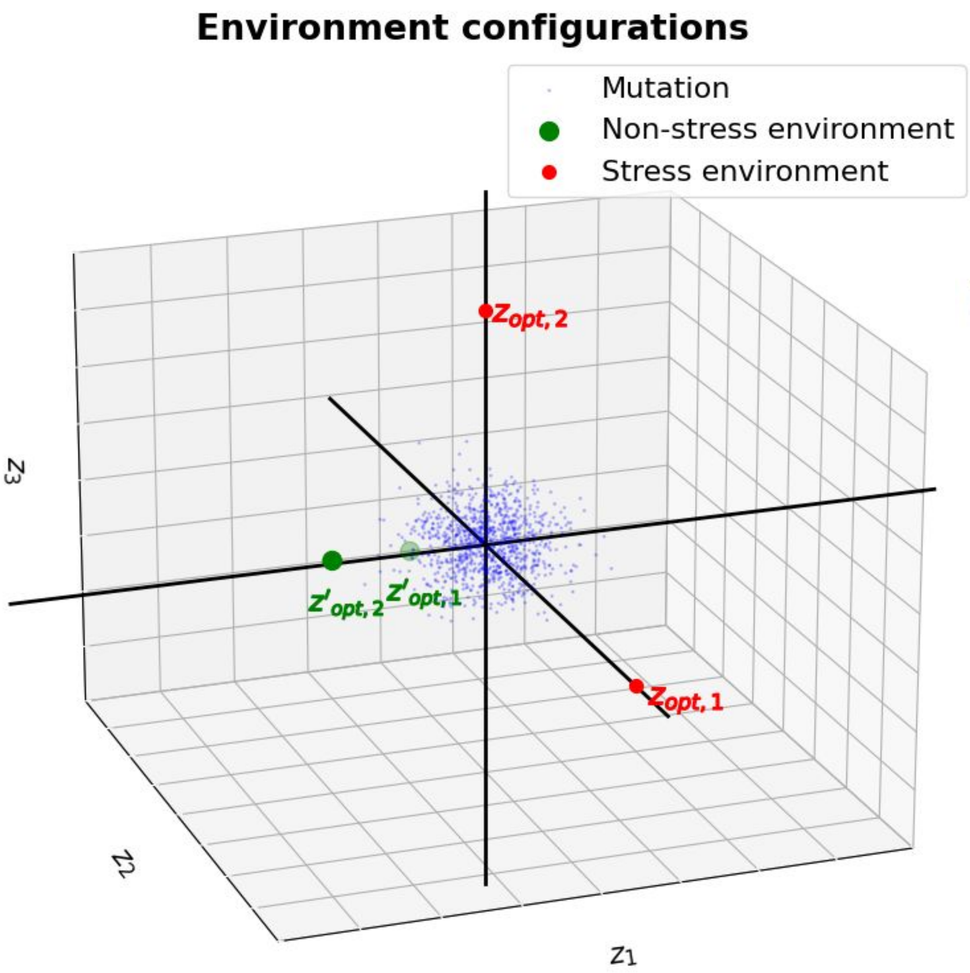
$$H(\mathbf{z}) = \ln\left(\frac{W(\mathbf{z})}{W(\mathbf{z_0})}\right)$$

Fisher's Geometric Model



Relative Fitness

$$H(\mathbf{z}) = \ln\left(\frac{W(\mathbf{z})}{W(\mathbf{z_0})}\right)$$

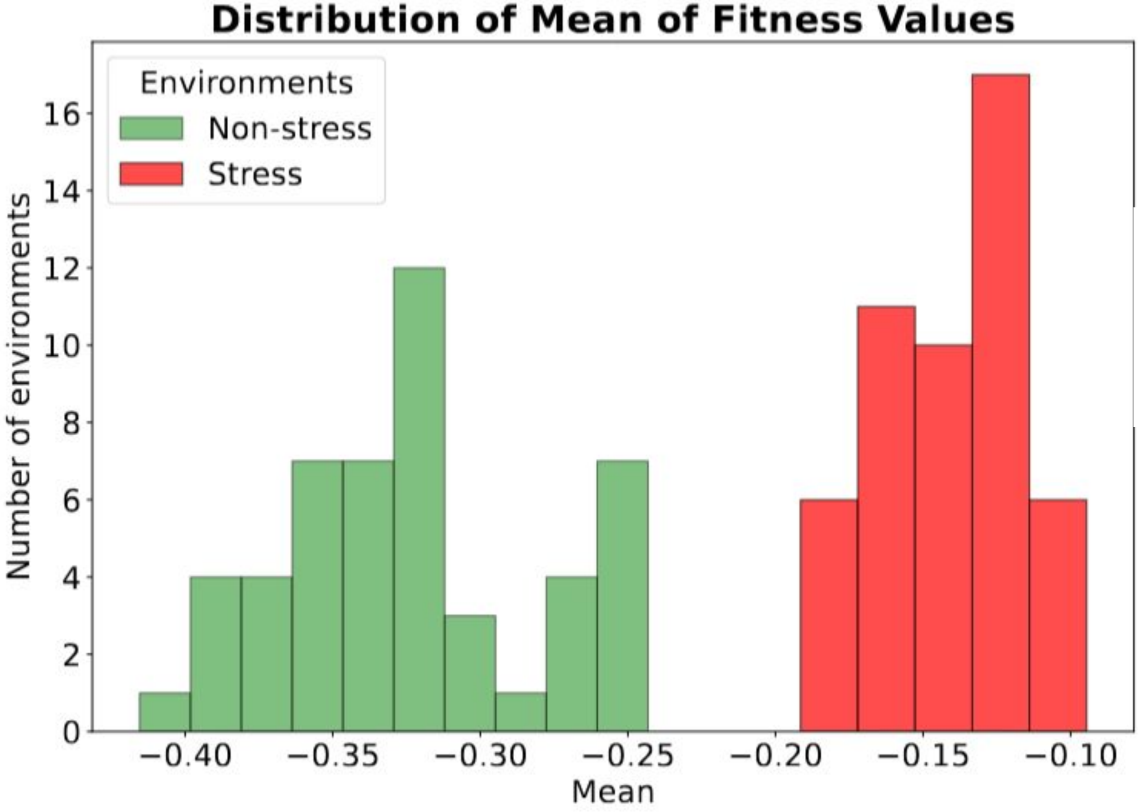


Fitness Function

$$W(||\mathbf{z}-\mathbf{z}_{\text{opt}}||) = e^{\alpha||\mathbf{z}-\mathbf{z}_{\text{opt}}||^Q}$$

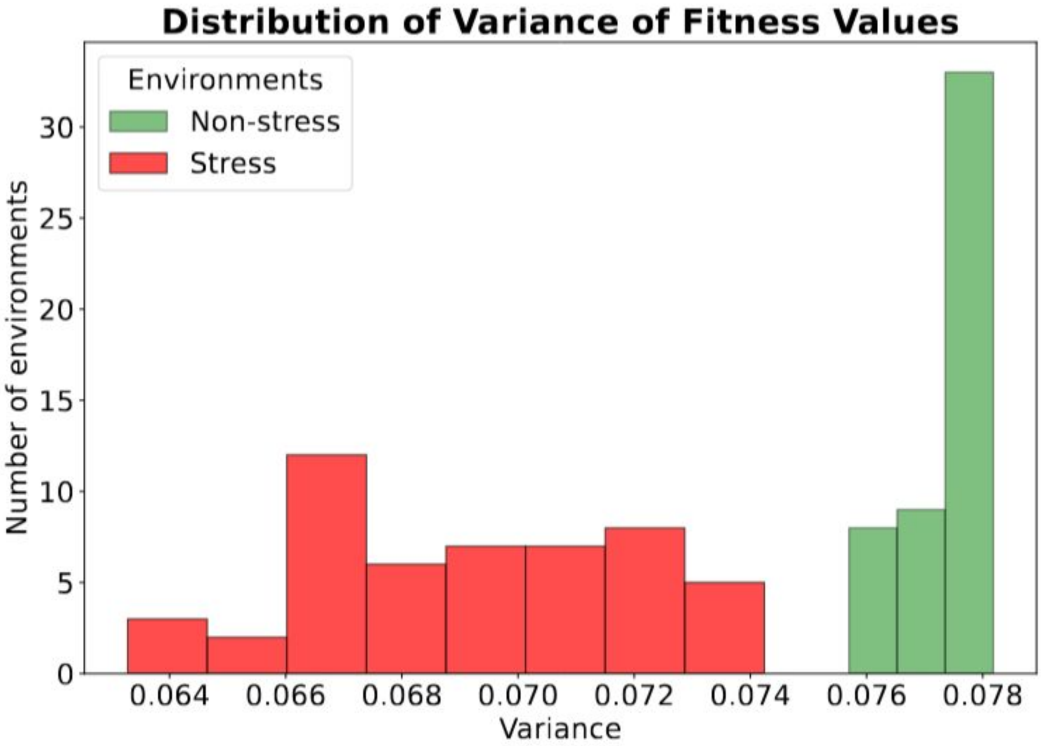
Mutant Distribution

$$d\mathbf{z} = \{dz_i\}_{i\in[1,n]} \quad MVN(0,\sigma_m^2\mathbf{I}_n)$$



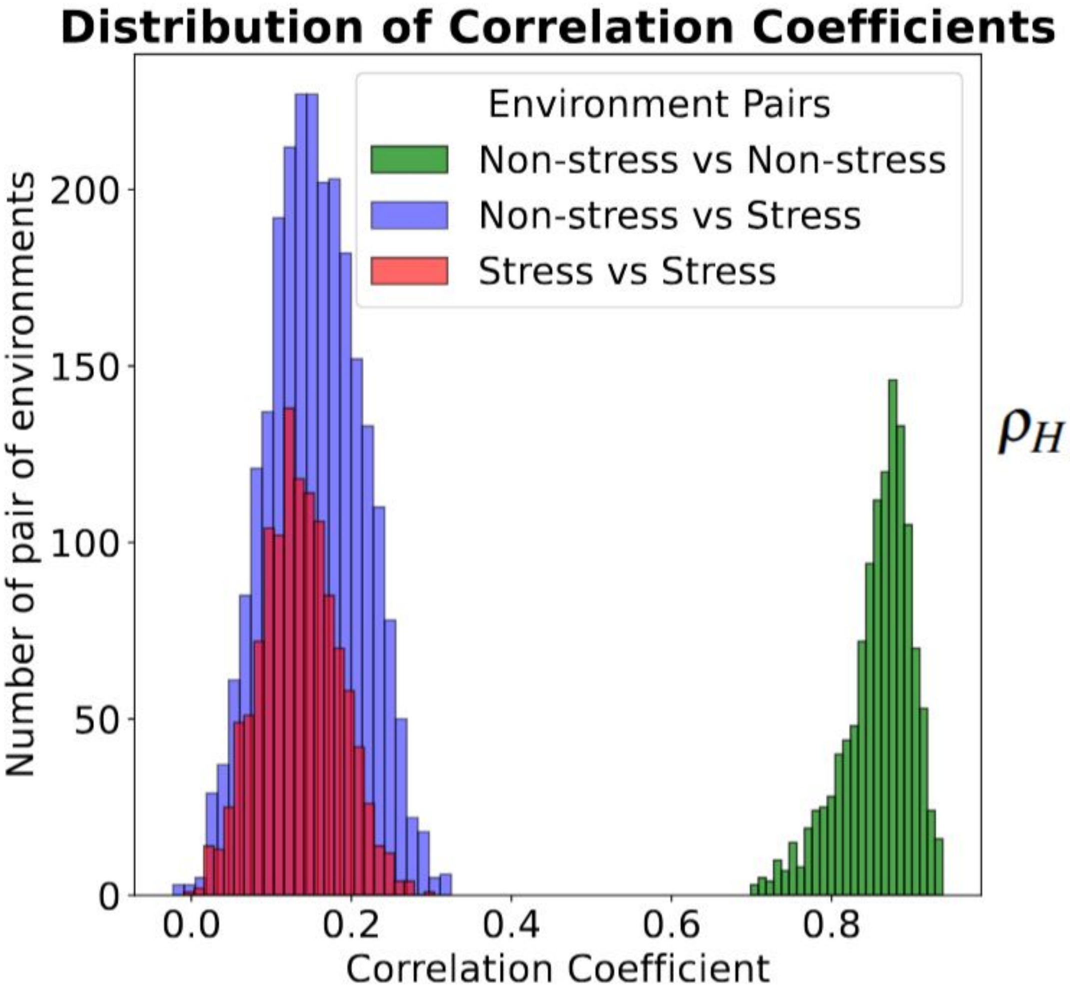
Mean of DFE

$$M \approx \frac{\alpha \sigma_{\mathrm{mut}}^2}{2} Q(Q+n-2) \times ||\mathbf{z}_0 - \mathbf{z}_{\mathrm{opt}}||^{Q-2} + o[E[dz^3]]$$



Variance of DFE

$$V \approx \alpha^2 \sigma_{\mathrm{mut}}^2 Q^2 \times ||\mathbf{z}_0 - \mathbf{z}_{\mathrm{opt}}||^{2Q-2} + o[E[dz^3]]$$



Correlation Coefficient

$$\rho_{H,H'} \approx \cos\left(\theta_{\mathbf{z}_{\text{opt}},\mathbf{z}'_{\text{opt}}}\right) + o[E[dz^3]]$$

Conclusion

1. Mechanistic Models: Link DFEs to molecular mechanisms (e.g., protein stability, metabolic fluxes) to move beyond phenomenological descriptions.

2. Environment Dimensionality: Develop scalable methods to handle high-dimensional environmental spaces (e.g., multi-nutrient gradients, biotic interactions).

3. Epistasis and Dynamics: Explore how DFEs change as populations evolve (i.e., epistasis) and incorporate temporal environmental fluctuations.

4. Cross-Species Frameworks: Test whether DFE patterns generalize across bacteria, eukaryotes, or viruses, revealing universal evolutionary constraints.

Thank You

Key Assumptions

7. Fitness potential:

$$H(\mathbf{z}) = \log\left(\frac{W(\mathbf{z}_0 + d\mathbf{z})}{W(\mathbf{z}_0)}\right)$$

Relative fitness:

$$dw = \frac{W(\mathbf{z}_0 + d\mathbf{z}) - W(\mathbf{z}_0)}{W(\mathbf{z}_0)}$$

$$dw \approx \ln(1 + dw) = \ln W(\mathbf{z}_0 + d\mathbf{z}) - \ln W(\mathbf{z}_0)$$

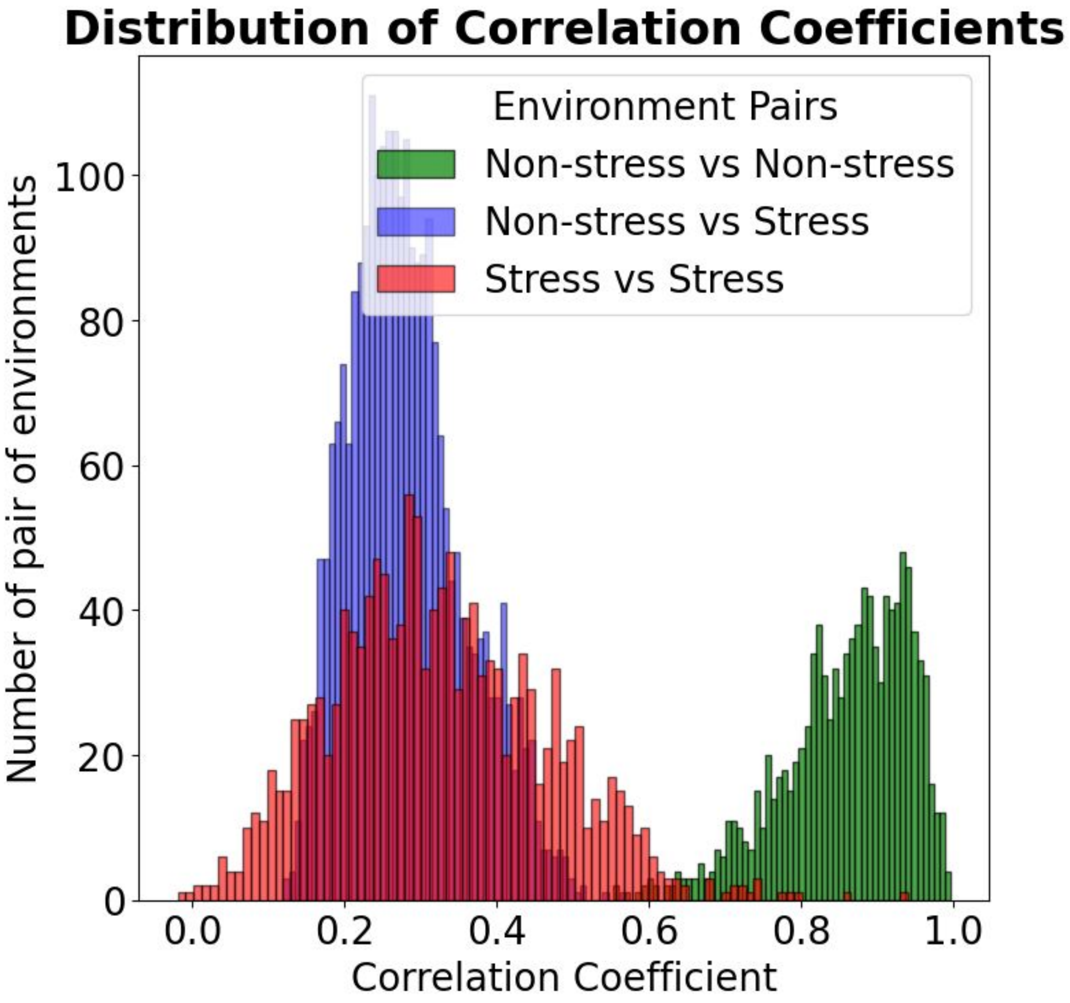
Key Assumptions

1. **Phenotype Space:** $\mathbf{z} = \{z_i\}_{i \in [1,n]}$
2. **Genotype to Phenotype Relationship:** $d\mathbf{z} = \{dz_i\}_{i \in [1,n]} \quad MVN(0, \sigma_m^2 \mathbf{I}_n)$
3. **Phenotypic Fitness Landscape:** $W(\mathbf{z}_0)$
4. **Monotonicity of Fitness Function**
5. **Environmental Change**
6. **Stress and Non-stress Environment**

Dataset

Gene Name	Description	D-Glucose (C)	L-Arginine (N)	Bacitracin 0.5 mg/ml	Spectinomycin 0.0125 mg/ml	Aluminum chloride 5 mM
talB	transaldolase B (NCBI)	-0.001	-0.027	-0.812	-2.712	-0.429
yaaJ	predicted transporter (NCBI)	-0.127	0.123	0.200	0.275	0.008
polB	DNA polymerase II (NCBI)	-0.055	-0.079	-0.169	-0.328	-0.113
ddlB	D-alanylalanine synthetase (NCBI)	-0.509	-0.018	-1.366	-0.670	0.198
yacG	zinc-binding protein (NCBI)	-0.909	-0.525	0.309	0.546	0.750

Source: Price et. al. 2018



Distribution of Correlation Coefficients
Environment Pairs
Non-stress vs Non-stress
Non-stress vs Stress
Stress vs Stress
Number of pair of environments
0
20
40
60
80
100
0.0
0.2
0.4
0.6
0.8
1.0
Correlation Coefficient

Distribution of the number of fitness maxima in Fisher's geometric model

Su-Chan Park[1], **Sungmin Hwang**[2] **and Joachim Krug**[3,4]

[1] Department of Physics, The Catholic University of Korea, Bucheon 14662, Republic of Korea
[2] Capital Fund Management, 23-25 Rue de l'Université, 75007 Paris, France
[3] Institute for Biological Physics, University of Cologne, Zülpicher Strasse 77, 50937 Köln, Germany

E-mail: jkrug@uni-koeln.de

Received 25 February 2020, revised 20 May 2020
Accepted for publication 28 May 2020
Published 21 August 2020

CrossMark