

# Genotypic complexity of Fisher's geometric model

---

Sungmin Hwang<sup>1</sup>, Su-Chan Park<sup>2</sup>, Joachim Krug<sup>1</sup>

July 7, 2016

<sup>1</sup>Institut für Theoretische Physik, Universität zu Köln

<sup>2</sup>The Catholic University of Korea / KIAS

## **Fitness landscape and Fisher's geometric model**

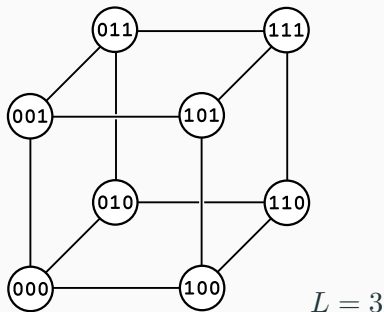
---

# Fitness landscape

- **Fitness**  $f$  is a measure of the **reproductive success** of an organism.
- **Genotype**  $\sigma$  is the genetic representation of an organism which is encoded as a binary sequence of length  $L$  representing the presence (absence) of possible mutations.  
E.g.  $\sigma = (1, 0, 1, \dots, 0)$  and  $\sigma_0 = (0, 0, 0, \dots, 0)$ .  
$$\underbrace{\hspace{10em}}_L$$
- **Fitness landscape** is a mapping from a genotype  $\sigma$  to its fitness, i.e.,  $\tau \rightarrow f(\tau)$ .

## Underlying genotypic structure: Hypercube

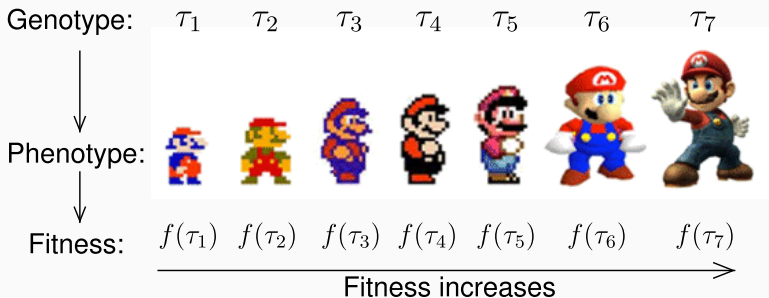
- We call  $\sigma$  and  $\sigma'$  **adjacent** iff  $D(\sigma, \sigma') = 1$ , i.e., if  $\sigma$  can be reached by a single mutation from  $\sigma'$ ,
- The underlying graph structure becomes an  $L$ -dimensional **hypercube** with  $2^L$  nodes.



## Local optima and genotypic complexity

- The genotype  $\sigma$  is a **local optimum**, if  $f(\sigma) > f(\sigma')$  for all **adjacent** genotypes  $\sigma'$ .
- **The total number of local optima**  $\mathcal{N}$  in the landscape is expected to obey  $\mathcal{N} \sim e^{L\Sigma}$ , where  $\Sigma$  is called the genotypic complexity.
- Obviously,  $0 \leq \Sigma \leq \ln 2$ .

# Phenotypic fitness landscape and Fisher's geometric model<sup>1</sup>



<sup>1</sup>R. Fisher (1930).

## Phenotypic fitness landscape and Fisher's geometric model<sup>2</sup>

- A **phenotype** is the set of an organism's **observable** traits.
- e.g.) weight, height, color,  $\dots$ .
- The phenotypes are assumed to be **continuous** variables  $\vec{z}(\sigma) \in \mathbb{R}^n$  where  $n$  is the number of traits.
- $n$  is referred to as **phenotypic complexity** in biology.
- The fitness depends on genotypes only via **phenotypes**, i.e.,  $\sigma \rightarrow \vec{z}(\sigma) \rightarrow f(\vec{z}(\sigma))$ .

---

<sup>2</sup>R. Fisher (1930).

# Phenotypic fitness landscape and Fisher's geometric model<sup>2</sup>

- A **phenotype** is the set of an organism's **observable** traits.
- e.g.) weight, height, color,  $\dots$ .
- The phenotypes are assumed to be **continuous** variables  $\vec{z}(\sigma) \in \mathbb{R}^n$  where  $n$  is the number of traits.
- $n$  is referred to as **phenotypic complexity** in biology.
- The fitness depends on genotypes only via **phenotypes**, i.e.,  $\sigma \rightarrow \vec{z}(\sigma) \rightarrow f(\vec{z}(\sigma))$ .

---

<sup>2</sup>R. Fisher (1930).



# Phenotypic fitness landscape and Fisher's geometric model<sup>2</sup>

- A **phenotype** is the set of an organism's **observable** traits.
- e.g.) weight, height, color,  $\dots$ .
- The phenotypes are assumed to be **continuous** variables  $\vec{z}(\sigma) \in \mathbb{R}^n$  where  $n$  is the number of traits.
- $n$  is referred to as **phenotypic complexity** in biology.
- The fitness depends on genotypes only via **phenotypes**, i.e.,  $\sigma \rightarrow \vec{z}(\sigma) \rightarrow f(\vec{z}(\sigma))$ .

---

<sup>2</sup>R. Fisher (1930).

# Phenotypic fitness landscape and Fisher's geometric model<sup>2</sup>

- A **phenotype** is the set of an organism's **observable** traits.
- e.g.) weight, height, color, . . . .
- The phenotypes are assumed to be **continuous** variables  $\vec{z}(\sigma) \in \mathbb{R}^n$  where  $n$  is the number of traits.
- $n$  is referred to as **phenotypic complexity** in biology.
- The fitness depends on genotypes only via **phenotypes**, i.e.,  $\sigma \rightarrow \vec{z}(\sigma) \rightarrow f(\vec{z}(\sigma))$ .

---

<sup>2</sup>R. Fisher (1930).

## Fisher's geometric model: Additional assumptions<sup>3</sup>

1. There is a **unique** phenotypic optimum at the origin.
2. The fitness is a **decreasing** function of phenotypic distance to the origin.
3. **Universal pleiotropy**: Mutations correspond to **random displacements** ( $\vec{\xi}_i \in \mathbb{R}^n$ ) drawn from (isotropic) Gaussian distribution with unit standard deviation.
4. **Additivity** of mutational displacements:

$$\vec{z}(\sigma) = \vec{Q} + \sum_{i=1}^L \sigma_i \vec{\xi}_i,$$

where  $\vec{Q}$  is the wild-type phenotype .

---

<sup>3</sup>O. Tenaillon, *Annu. Rev. Ecol. Evol. Syst.*(2014).

## Fisher's geometric model: Additional assumptions<sup>3</sup>

1. There is a **unique** phenotypic optimum at the origin.
2. The fitness is a **decreasing** function of phenotypic distance to the origin.
3. **Universal pleiotropy**: **Mutations** correspond to **random displacements** ( $\vec{\xi}_i \in \mathbb{R}^n$ ) drawn from (isotropic) Gaussian distribution with unit standard deviation.
4. **Additivity** of mutational displacements:

$$\vec{z}(\sigma) = \vec{Q} + \sum_{i=1}^L \sigma_i \vec{\xi}_i,$$

where  $\vec{Q}$  is the wild-type phenotype .

---

<sup>3</sup>O. Tenaillon, *Annu. Rev. Ecol. Evol. Syst.*(2014).

## Fisher's geometric model: Additional assumptions<sup>3</sup>

1. There is a **unique** phenotypic optimum at the origin.
2. The fitness is a **decreasing** function of phenotypic distance to the origin.
3. **Universal pleiotropy**: **Mutations** correspond to **random displacements** ( $\vec{\xi}_i \in \mathbb{R}^n$ ) drawn from (isotropic) Gaussian distribution with unit standard deviation.
4. **Additivity** of mutational displacements:

$$\vec{z}(\sigma) = \vec{Q} + \sum_{i=1}^L \sigma_i \vec{\xi}_i,$$

where  $\vec{Q}$  is the wild-type phenotype .

---

<sup>3</sup>O. Tenaillon, *Annu. Rev. Ecol. Evol. Syst.*(2014).

## Genotypic complexity of FGM for fixed $n$

---

# Main question of the talk

- We calculate the mean number of **local optima**  $\mathcal{N}$  out of  $2^L$  possible genotypes as a measure of **ruggedness**.
- **Mapping to a spin model**

$$\begin{aligned}\mathcal{H} = -f(\sigma) &\propto |\vec{z}(\sigma)|^2 = \left( \vec{Q} + \sum_{i=1}^L \sigma_i \vec{\xi}_i \right)^2 \\ &= |\vec{Q}|^2 - \sum_{i=1}^L H_i \sigma_i - \sum_{i=1}^L \sum_{j=1}^L J_{ij} \sigma_i \sigma_j\end{aligned}$$

where  $J_{ij} = -\vec{\xi}_i \cdot \vec{\xi}_j$  and  $H_i = -2\vec{Q} \cdot \vec{\xi}_i$

- c.f.)

$$\mathcal{H}^{\text{Hopfield}} = - \sum_{i,j} \vec{\xi}_i \cdot \vec{\xi}_j \tau_i \tau_j$$

# Main question of the talk

- We calculate the mean number of **local optima**  $\mathcal{N}$  out of  $2^L$  possible genotypes as a measure of **ruggedness**.
- **Mapping to a spin model**

$$\begin{aligned}\mathcal{H} = -f(\sigma) &\propto |\vec{z}(\sigma)|^2 = \left( \vec{Q} + \sum_{i=1}^L \sigma_i \vec{\xi}_i \right)^2 \\ &= |\vec{Q}|^2 - \sum_{i=1}^L H_i \sigma_i - \sum_{i=1}^L \sum_{j=1}^L J_{ij} \sigma_i \sigma_j\end{aligned}$$

where  $J_{ij} = -\vec{\xi}_i \cdot \vec{\xi}_j$  and  $H_i = -2\vec{Q} \cdot \vec{\xi}_i$

- c.f.)

$$\mathcal{H}^{\text{Hopfield}} = - \sum_{i,j} \vec{\xi}_i \cdot \vec{\xi}_j \tau_i \tau_j$$



## Mean Number of optima of FGM

- Additivity of random variables:

$$\langle \mathcal{N} \rangle = \sum_{s=0}^L \binom{L}{s} R_s(L),$$

where  $R_s(L)$  is the prob. that a genotype carrying  $s$  mutations is a local optimum.

## Mean Number of optima of FGM

$$R_s(L) = \int_{\mathbb{R}^n} d\vec{z} \delta \left( \vec{z} - \vec{Q} - \sum_{i=1}^s \vec{\xi}_i \right) \times \\ \left[ \prod_{i=1}^s \int_{\mathcal{D}(\vec{z})} d\vec{\xi}_i p(\vec{\xi}_i) \right] \left[ \prod_{i=s+1}^L \int_{\mathcal{D}(-\vec{z})} d\vec{\xi}_i p(\vec{\xi}_i) \right],$$

where  $\mathcal{D}(\vec{x}) \equiv \{\vec{y} \in \mathbb{R}^n \mid |\vec{y} - \vec{x}| > |\vec{x}|\}$ .

- $\delta \left( \vec{z} - \vec{Q} - \sum_{i=1}^s \vec{\xi}_i \right)$  constrains  $\vec{z}$  to be the position of the phenotype of  $\sigma$ .
- For  $\sigma_i = 1$ , the optimum condition requires  $|\vec{z}| < |\vec{z} - \vec{\xi}_i|$ .
- For  $\sigma_i = 0$ , it requires  $|\vec{z}| < |\vec{z} + \vec{\xi}_i|$ .
- If  $z = 0$ , they are automatically satisfied.

## Mean Number of optima of FGM

$$R_s(L) = \int_{\mathbb{R}^n} d\vec{z} \delta \left( \vec{z} - \vec{Q} - \sum_{i=1}^s \vec{\xi}_i \right) \times \\ \left[ \prod_{i=1}^s \int_{\mathcal{D}(\vec{z})} d\vec{\xi}_i p(\vec{\xi}_i) \right] \left[ \prod_{i=s+1}^L \int_{\mathcal{D}(-\vec{z})} d\vec{\xi}_i p(\vec{\xi}_i) \right],$$

where  $\mathcal{D}(\vec{x}) \equiv \{\vec{y} \in \mathbb{R}^n \mid |\vec{y} - \vec{x}| > |\vec{x}|\}$ .

- $\delta \left( \vec{z} - \vec{Q} - \sum_{i=1}^s \vec{\xi}_i \right)$  constrains  $\vec{z}$  to be the position of the phenotype of  $\sigma$ .
- For  $\sigma_i = 1$ , the optimum condition requires  $|\vec{z}| < |\vec{z} - \vec{\xi}_i|$ .
- For  $\sigma_i = 0$ , it requires  $|\vec{z}| < |\vec{z} + \vec{\xi}_i|$ .
- If  $z = 0$ , they are automatically satisfied.

## Mean Number of optima of FGM

$$R_s(L) = \int_{\mathbb{R}^n} d\vec{z} \delta \left( \vec{z} - \vec{Q} - \sum_{i=1}^s \vec{\xi}_i \right) \times \\ \left[ \prod_{i=1}^s \int_{\mathcal{D}(\vec{z})} d\vec{\xi}_i p(\vec{\xi}_i) \right] \left[ \prod_{i=s+1}^L \int_{\mathcal{D}(-\vec{z})} d\vec{\xi}_i p(\vec{\xi}_i) \right],$$

where  $\mathcal{D}(\vec{x}) \equiv \{\vec{y} \in \mathbb{R}^n \mid |\vec{y} - \vec{x}| > |\vec{x}|\}$ .

- $\delta \left( \vec{z} - \vec{Q} - \sum_{i=1}^s \vec{\xi}_i \right)$  constrains  $\vec{z}$  to be the position of the phenotype of  $\sigma$ .
- For  $\sigma_i = 1$ , the optimum condition requires  $|\vec{z}| < |\vec{z} - \vec{\xi}_i|$ .
- For  $\sigma_i = 0$ , it requires  $|\vec{z}| < |\vec{z} + \vec{\xi}_i|$ .
- If  $z = 0$ , they are automatically satisfied.

## Mean Number of optima of FGM

$$R_s(L) = \int_{\mathbb{R}^n} d\vec{z} \delta \left( \vec{z} - \vec{Q} - \sum_{i=1}^s \vec{\xi}_i \right) \times \\ \left[ \prod_{i=1}^s \int_{\mathcal{D}(\vec{z})} d\vec{\xi}_i p(\vec{\xi}_i) \right] \left[ \prod_{i=s+1}^L \int_{\mathcal{D}(-\vec{z})} d\vec{\xi}_i p(\vec{\xi}_i) \right],$$

where  $\mathcal{D}(\vec{x}) \equiv \{\vec{y} \in \mathbb{R}^n \mid |\vec{y} - \vec{x}| > |\vec{x}|\}$ .

- $\delta \left( \vec{z} - \vec{Q} - \sum_{i=1}^s \vec{\xi}_i \right)$  constrains  $\vec{z}$  to be the position of the phenotype of  $\sigma$ .
- For  $\sigma_i = 1$ , the optimum condition requires  $|\vec{z}| < |\vec{z} - \vec{\xi}_i|$ .
- For  $\sigma_i = 0$ , it requires  $|\vec{z}| < |\vec{z} + \vec{\xi}_i|$ .
- If  $z = 0$ , they are automatically satisfied.

## Mean Number of optima of FGM

$$R_s(L) = \int_{\mathbb{R}^n} d\vec{z} \delta \left( \vec{z} - \vec{Q} - \sum_{i=1}^s \vec{\xi}_i \right) \times \\ \left[ \prod_{i=1}^s \int_{\mathcal{D}(\vec{z})} d\vec{\xi}_i p(\vec{\xi}_i) \right] \left[ \prod_{i=s+1}^L \int_{\mathcal{D}(-\vec{z})} d\vec{\xi}_i p(\vec{\xi}_i) \right],$$

where  $\mathcal{D}(\vec{x}) \equiv \{\vec{y} \in \mathbb{R}^n \mid |\vec{y} - \vec{x}| > |\vec{x}|\}$ .

- $\delta \left( \vec{z} - \vec{Q} - \sum_{i=1}^s \vec{\xi}_i \right)$  constrains  $\vec{z}$  to be the position of the phenotype of  $\sigma$ .
- For  $\sigma_i = 1$ , the optimum condition requires  $|\vec{z}| < |\vec{z} - \vec{\xi}_i|$ .
- For  $\sigma_i = 0$ , it requires  $|\vec{z}| < |\vec{z} + \vec{\xi}_i|$ .
- If  $z = 0$ , they are automatically satisfied.

## Small $q$ phase

- **Naturally**, we can guess that  $R_s$  is dominated around  $z = 0$ .  
Around this point, we found

$$R_s(L) \approx \frac{s^{-n/2} \exp[-Q^2/(2s)]}{s \exp[-Q^2/(2s^2)] + L - s}.$$

- Interpretation: Let  $P_r$  be prob. that  $|\vec{z}| < r$  with  $r \ll 1$ .

$$P_r = \frac{V_n r^n}{(2\pi)^{n/2}} s^{-n/2} \exp[-Q^2/(2s)]$$

- Typical size of an optimal phenotype:

$$\frac{V_n r^n}{(2\pi)^{n/2}} \simeq (s \exp[-Q^2/(2s^2)] + L - s)^{-1}$$

## Small $q$ phase

- Naturally, we can guess that  $R_s$  is dominated around  $z = 0$ . Around this point, we found

$$R_s(L) \approx \frac{s^{-n/2} \exp[-Q^2/(2s)]}{s \exp[-Q^2/(2s^2)] + L - s}.$$

- Interpretation: Let  $P_r$  be prob. that  $|\vec{z}| < r$  with  $r \ll 1$ .

$$P_r = \frac{V_n r^n}{(2\pi)^{n/2}} s^{-n/2} \exp[-Q^2/(2s)]$$

- Typical size of an optimal phenotype:

$$\frac{V_n r^n}{(2\pi)^{n/2}} \simeq (s \exp[-Q^2/(2s^2)] + L - s)^{-1}$$



## Small $q$ phase

Setting  $q = Q/L$  and  $\rho = s/L$ , we get

$$\begin{aligned}\mathcal{N} &= \sum_{s=0}^L \binom{L}{s} R_s(L) \\ &\approx \int_0^1 d\rho \frac{1}{L^{n/2} \rho^{n/2}} \frac{e^{L\Sigma(\rho)}}{\sqrt{2\pi L\rho(1-\rho)}} \frac{1}{1-\rho + \rho e^{-\frac{q^2}{2\rho^2}}},\end{aligned}$$

where  $\Sigma(\rho) \equiv -\rho \ln \rho - (1-\rho) \ln(1-\rho) - q^2/2\rho$ .

$$\mathcal{N} \approx \frac{1}{L^{1+n/2}} \sqrt{\frac{1}{1+(1-\rho^*)(q/\rho^*)^2}} \frac{(\rho^*)^{-n/2} e^{L\Sigma(\rho^*)}}{1-\rho^* + \rho^* e^{-\frac{q^2}{2(\rho^*)^2}}}.$$

$\mathcal{N}$  decreases with  $n$ .

## Small $q$ phase

Setting  $q = Q/L$  and  $\rho = s/L$ , we get

$$\begin{aligned}\mathcal{N} &= \sum_{s=0}^L \binom{L}{s} R_s(L) \\ &\approx \int_0^1 d\rho \frac{1}{L^{n/2} \rho^{n/2}} \frac{e^{L\Sigma(\rho)}}{\sqrt{2\pi L\rho(1-\rho)}} \frac{1}{1-\rho + \rho e^{-\frac{q^2}{2\rho^2}}},\end{aligned}$$

where  $\Sigma(\rho) \equiv -\rho \ln \rho - (1-\rho) \ln(1-\rho) - q^2/2\rho$ .

$$\mathcal{N} \approx \frac{1}{L^{1+n/2}} \sqrt{\frac{1}{1+(1-\rho^*)(q/\rho^*)^2}} \frac{(\rho^*)^{-n/2} e^{L\Sigma(\rho^*)}}{1-\rho^* + \rho^* e^{-\frac{q^2}{2(\rho^*)^2}}}.$$

$\mathcal{N}$  decreases with  $n$ .

## Small $q$ phase

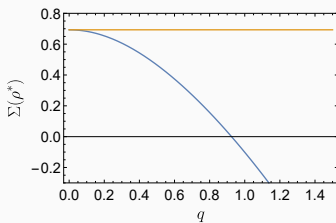
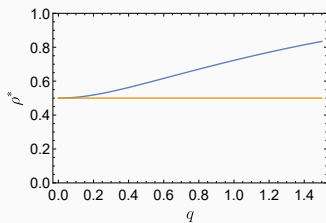
Setting  $q = Q/L$  and  $\rho = s/L$ , we get

$$\begin{aligned}\mathcal{N} &= \sum_{s=0}^L \binom{L}{s} R_s(L) \\ &\approx \int_0^1 d\rho \frac{1}{L^{n/2} \rho^{n/2}} \frac{e^{L\Sigma(\rho)}}{\sqrt{2\pi L\rho(1-\rho)}} \frac{1}{1-\rho + \rho e^{-\frac{q^2}{2\rho^2}}},\end{aligned}$$

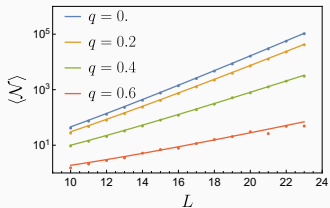
where  $\Sigma(\rho) \equiv -\rho \ln \rho - (1-\rho) \ln(1-\rho) - q^2/2\rho$ .

$$\mathcal{N} \approx \frac{1}{L^{1+n/2}} \sqrt{\frac{1}{1+(1-\rho^*)(q/\rho^*)^2}} \frac{(\rho^*)^{-n/2} e^{L\Sigma(\rho^*)}}{1-\rho^* + \rho^* e^{-\frac{q^2}{2(\rho^*)^2}}}.$$

$\mathcal{N}$  decreases with  $n$ .



$\Sigma$  becomes negative when  $q > 0.925$ .



## Large $q$ phase

- For  $z \gg 1$ , it turns out that the largest contribution comes from  $z^* = q - q_0$  with  $q_0 = 1/\sqrt{2\pi}$  and  $\rho^* = 1/2$ .

For  $q > q_0$ ,

$$\mathcal{N} \approx \left( \frac{q - q_0}{q} \exp \left[ \frac{q_0}{q - q_0} \right] \right)^{n-1}.$$

- $\mathcal{N}$  does **not** depend on  $L$ , implying  $\Sigma = 0$ .
- If  $n = 1$ ,  $\mathcal{N} = 1$ , implying there is **only one** global optimum.
- $\mathcal{N}$  increases with  $n$ .

## Large $q$ phase

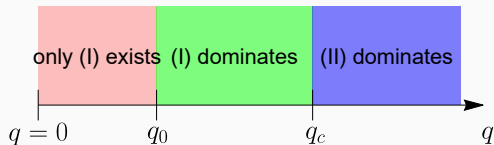
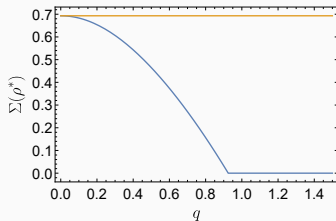
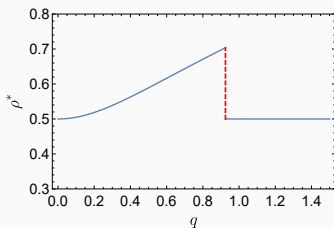
- For  $z \gg 1$ , it turns out that the largest contribution comes from  $z^* = q - q_0$  with  $q_0 = 1/\sqrt{2\pi}$  and  $\rho^* = 1/2$ .  
For  $q > q_0$ ,

$$\mathcal{N} \approx \left( \frac{q - q_0}{q} \exp \left[ \frac{q_0}{q - q_0} \right] \right)^{n-1}.$$

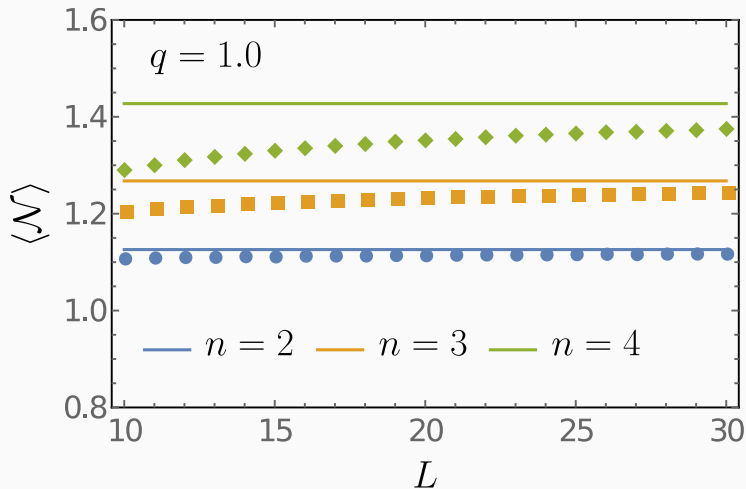
- $\mathcal{N}$  does **not** depend on  $L$ , implying  $\Sigma = 0$ .
- If  $n = 1$ ,  $\mathcal{N} = 1$ , implying there is **only one** global optimum.
- $\mathcal{N}$  increases with  $n$ .

# Phase transition

First order transition occurs at  $q = q_c \simeq 0.925$ .



## Numerical check





## Summary and Discussion

- Despite the simplicity of the model, it presents rich nontrivial features of various complexity measures.
- As reported in different contexts, complexity is generated in the FGM by two different mechanisms.
- Unlike the **common wisdom** of evolutionary biologists, the role of  $n$  is highly nontrivial.
- When  $n = O(L)$ , new method is necessary
  - The first correction of the first phase is order of  $L^{-1/n}$
  - $\mathcal{N}$  in the second phase increases exponentially with  $n$ .

# **Genotypic complexity of FGM for joint limit**

---

$n, L \rightarrow \infty$  **with**  $\alpha = n/L$  **fixed**

It turns out that solving  $\langle \mathcal{N} \rangle$  can be reduced to a three-dimensional variational problem, i.e.,

$n, L \rightarrow \infty$  **with**  $\alpha = n/L$  **fixed**

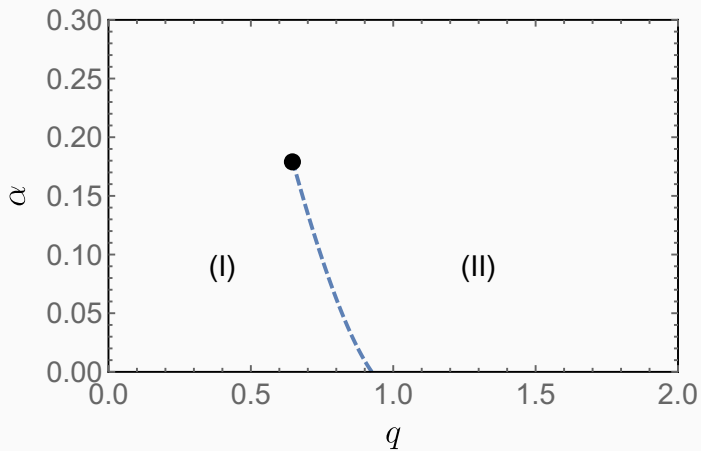
It turns out that solving  $\langle \mathcal{N} \rangle$  can be reduced to a three-dimensional variational problem, i.e.,

$$\langle \mathcal{N} \rangle = \mathcal{C}(a^*, b^*, c^*) e^{L \Sigma^{\text{red}}(a^*, b^*, c^*)} \left( 1 + O\left(\frac{1}{L}\right) \right), \quad (1)$$

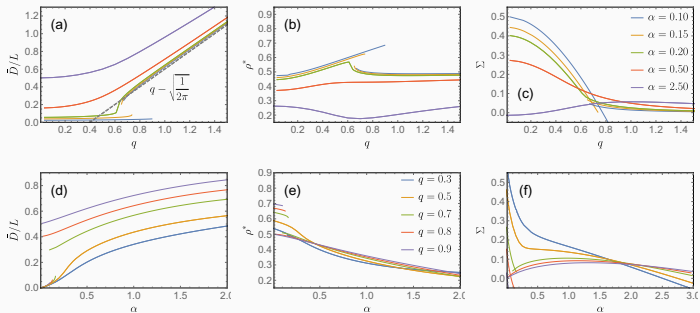
where the exponent  $\Sigma^{\text{red}}(a, b, c)$  is given by

$$\begin{aligned} \Sigma^{\text{red}}(a, b, c) = & \\ & - \frac{\alpha}{2} \log \left( \frac{\alpha \left( \alpha + \sqrt{\alpha^2 - 16cq^2} \right)}{2(ac + b^2)} \right) + \frac{\alpha + 2b + \sqrt{\alpha^2 - 16cq^2}}{2} \\ & + \log \left( \frac{1}{2} \left( e^{-2c} \left( \operatorname{erf} \left( \frac{\alpha + 2b}{\sqrt{2}\sqrt{a}} \right) + 1 \right) + \operatorname{erf} \left( \frac{\alpha}{\sqrt{2}\sqrt{a}} \right) + 1 \right) \right) \end{aligned} \quad (2)$$

# Phase diagram



# Phase diagram



## Large $q$ phase

