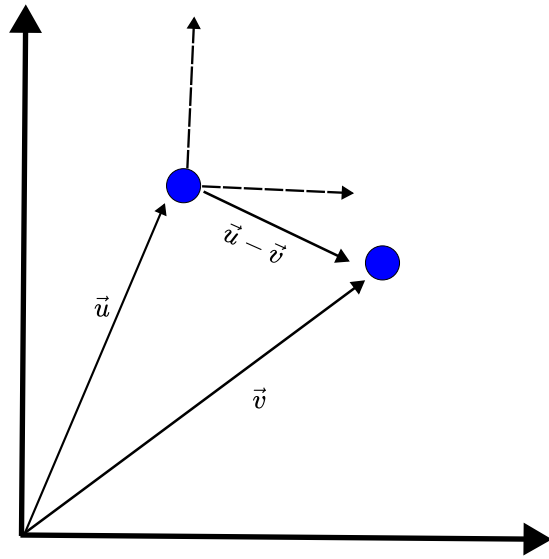


Diffusion in Genotype Space

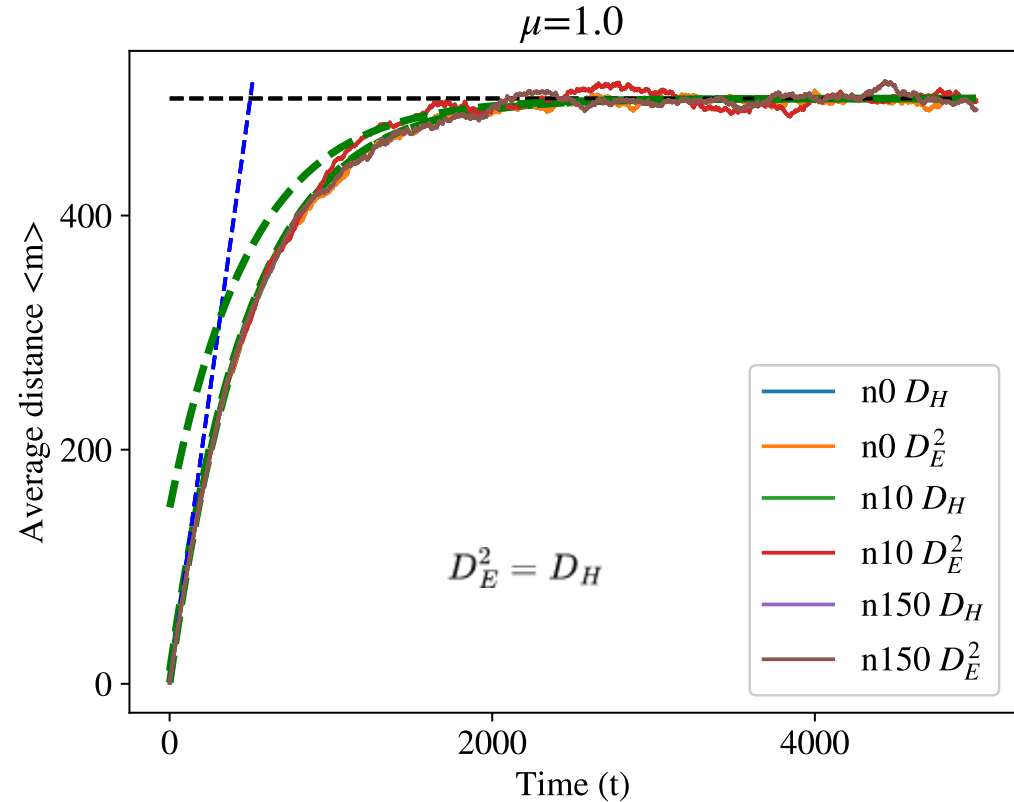
2

Diffusion in Genotype Space



$$D_H = \sum_{i=1}^L |u_i - v_i|$$

$$D_E = \|\vec{u} - \vec{v}\| = \sqrt{\sum_{i=1}^L (u_i - v_i)^2}$$



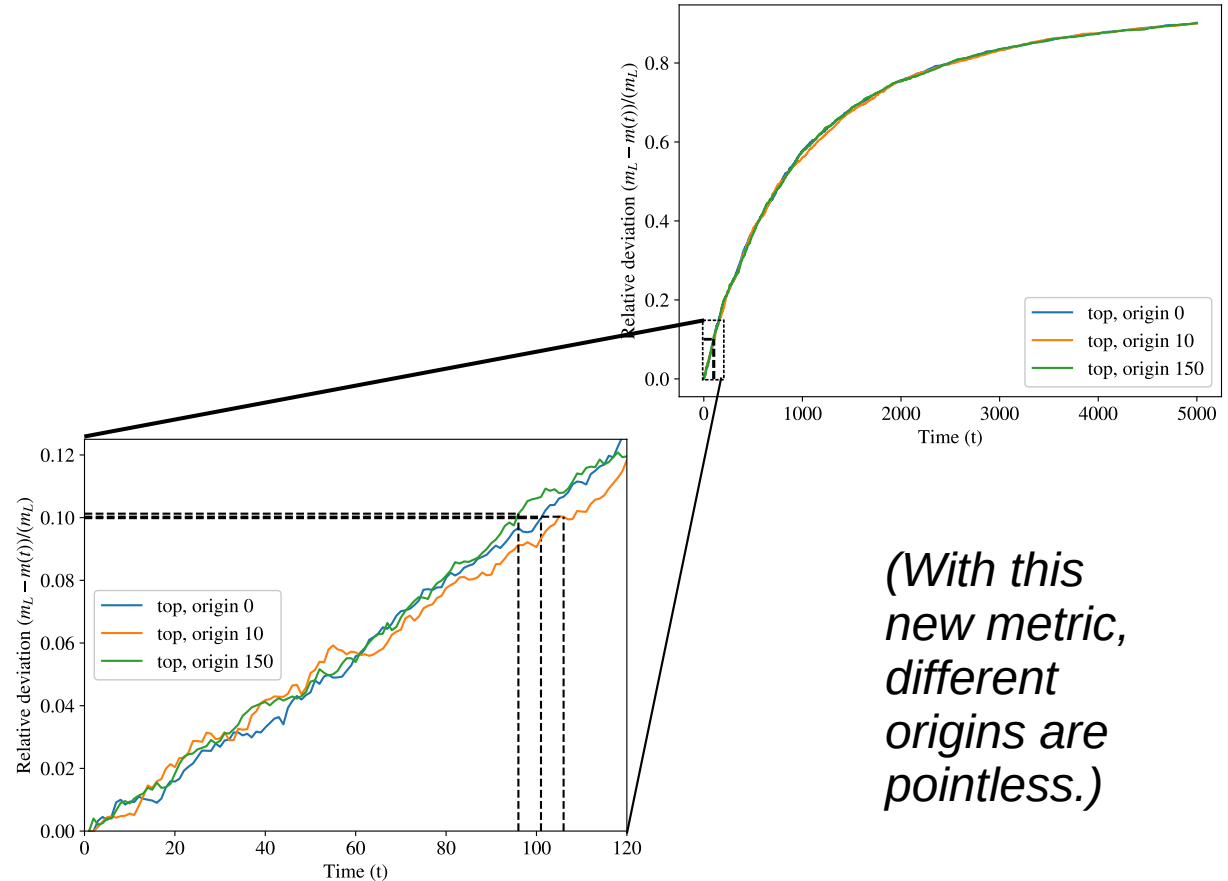
Using this metric only shifts the origin

Diffusion in Genotype Space

When can we say that
linear approximation is
valid?

$$\varepsilon := \frac{m_L - m(t)}{m_L} = 0.1$$

$$m_L = \mu t$$



*(With this
new metric,
different
origins are
pointless.)*

Diffusion in Genotype Space

Developing $m(t)$ to second order, one can obtain that

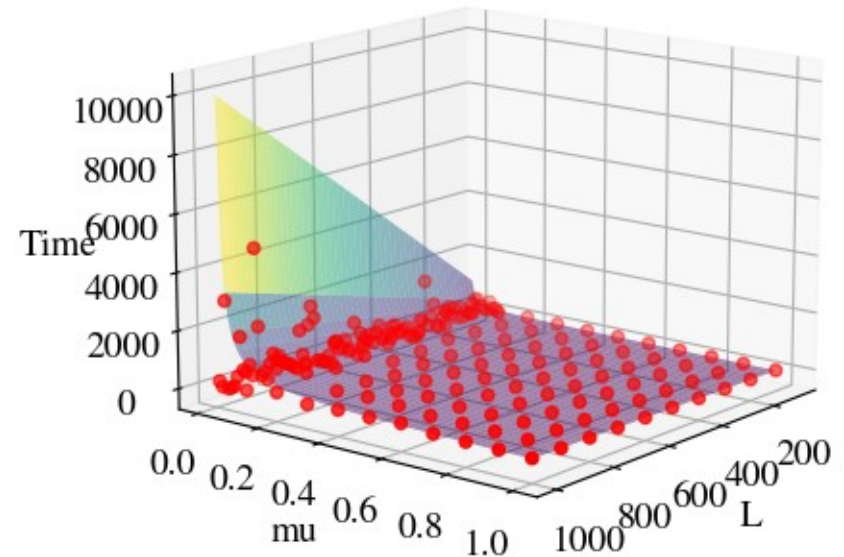
$$t_{\varepsilon} = \frac{2\lambda\varepsilon}{\mu}$$

Simulations were carried and stopped when the mean number of mutations of the population was

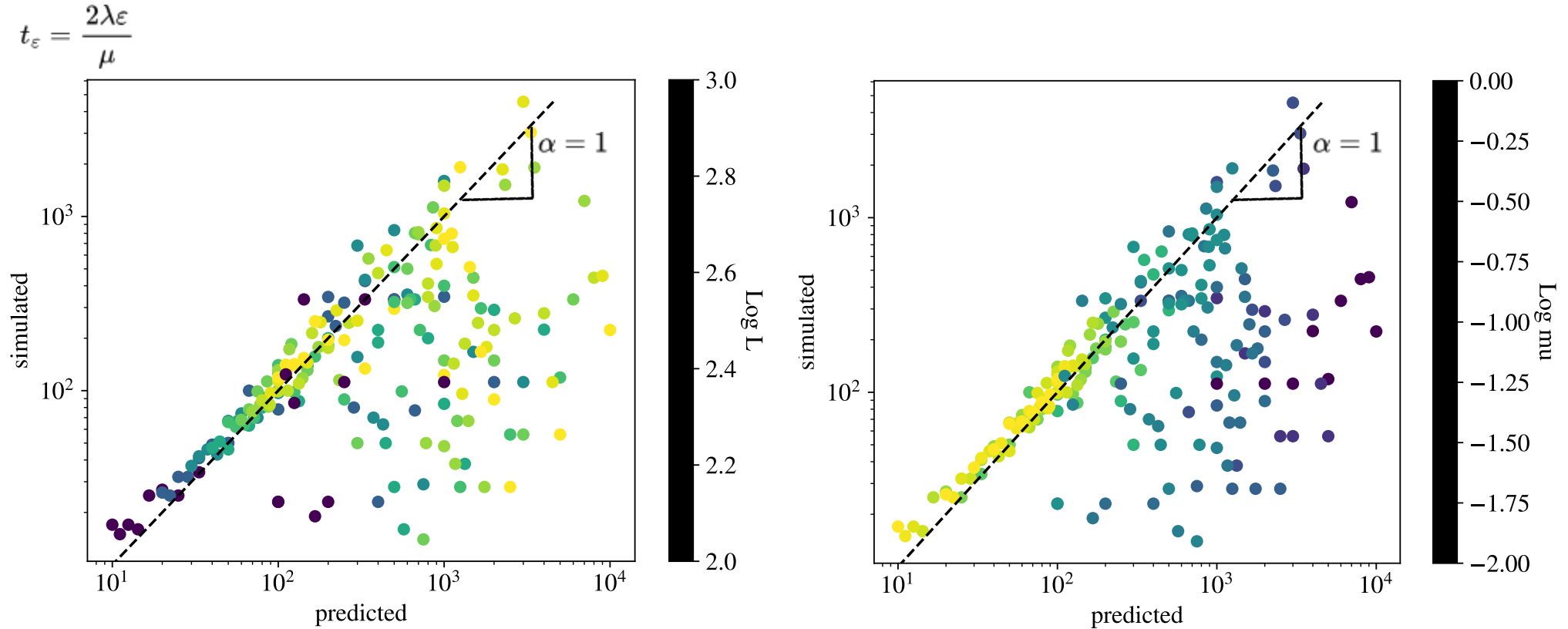
$$m(t) \leq (1 - \varepsilon)\mu t$$

then saving the value of t .

Surface → Theoretical result
Points → Simulated



Diffusion in Genotype Space



Although we have a good agreement, prediction fails for small values of μ

Open questions

The upcoming phase involves the duplication of sequences with a specified probability, denoted as "p." In instances where the probability is $1-p$, sequences do not duplicate but may undergo mutations.

I have a struggle deciding which implementation we want for the process.

1-) Does each sequence undergo a predetermined duplication? (my most voted)

This scenario suggests that certain sequences possess advantages for replication based on their genotype.

2-) Is the duplication of each sequence a random event (with probability p)?

In this case, the replication process appears to be a stochastic one within the population, regardless of the specific genotypes present.