A Momentary Glance at the 'Solitary Wave of Asexual Evolution'

David Kessler
Bar-Ilan University

Herbert Levine
UC San Diego

E. Cohen
Bar-Ilan

W. Peng
D. Ridgway
L. Tsimring
UC San Diego

Motivation: Experiments on Evolution of Viral Populations

Novella, et al., PNAS **92, 5841 (1995)**

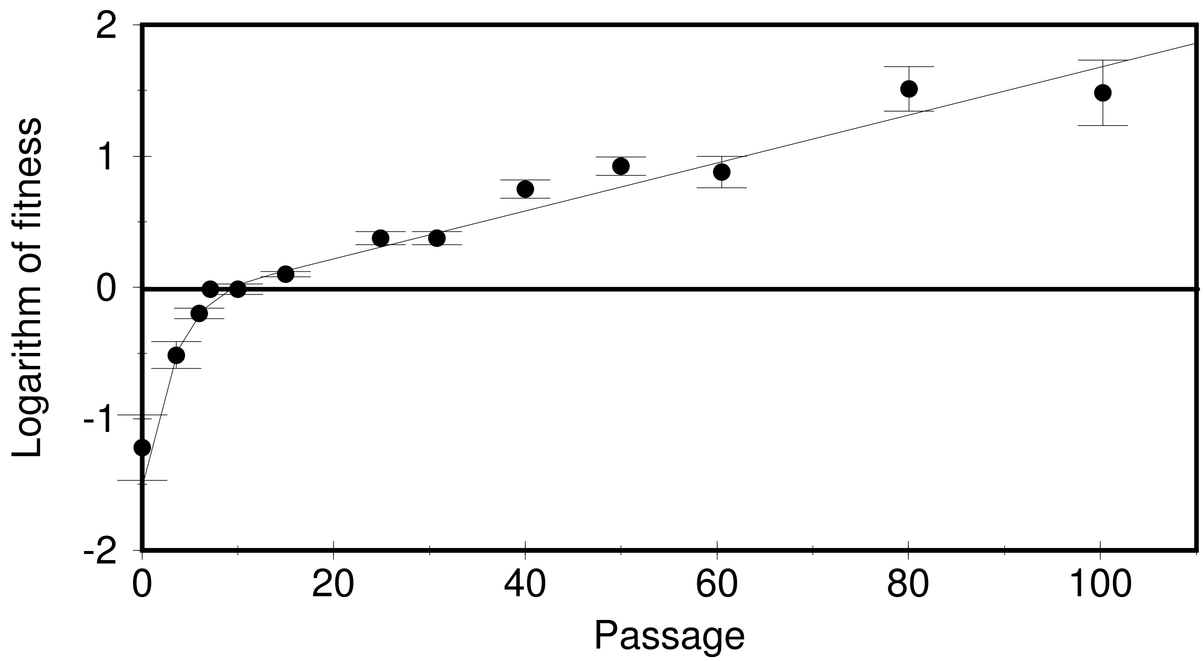
Basic Idea:

- Subject Vesicular Stomach Virus to Novel Environment
- Monitor Changes in Time of Population Characteristics
- Monitored Parameter: ''Instantaneous'' Growth Rate

Details:

- Daily Passages:
  - ▷ *Inject Small Number of Virons, ($N = 10^5$) into Cell Culture*
  - ▷ *Let Grow for 24 Hours, $N = 10^{10}$*
  - ▷ *Select a Small Sample of $N = 10^5$ Virons*
  - ▷ *Repeat*
- Monitor Growth Rate:
  - ▷ *Every Few Days, Take a Small Sample of Virons*
  - ▷ *Measure Exponential Growth Rate, in Competition with Reference Virus*
  - ▷ *Exponential Growth Rate = ''Log Fitness''*

The Results:



Log Fitness 0 $\equiv$ Wild Type

Significant Change in Fitness

Longer Experiments (> 100 Days) Show Saturation of Fitness

Similar Effects Seen in *E. coli*, HIV Infections

How can we understand this?

Basic Model

Genome: L binary genes, 0=Bad, 1=Good

Birth: Individual Reproduces with Rate x = # of 1's in Genome

Mutation: Genes Flip at Rate μ_0 / gene / birth

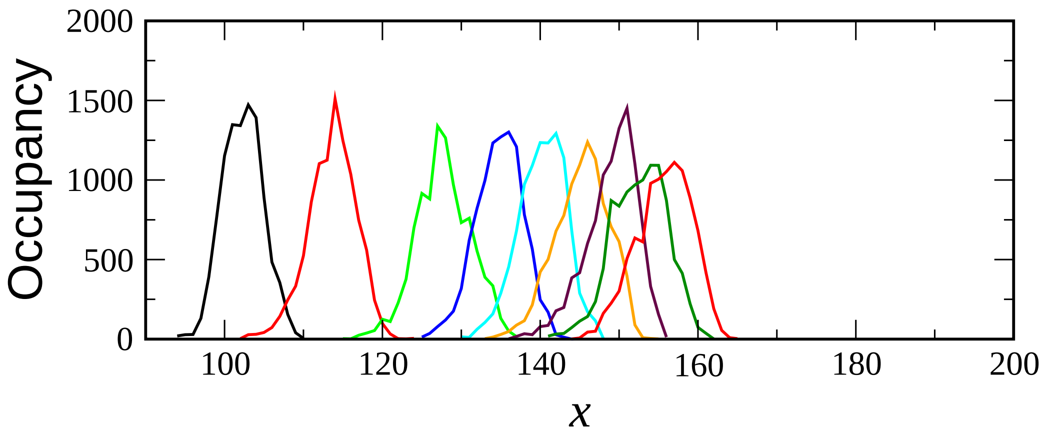
- $\mu \equiv \mu_0 L$, Overall (Genomic) Mutation Rate

Death: Fixed Number of Individuals = N

- Kill One Person (at Random) For Every Birth

Simulation:

L=200, N=10^4, μ=0.1



'The Solitary Wave of Asexual Evolution'

J. Wakeley

Basic Mechanisms:

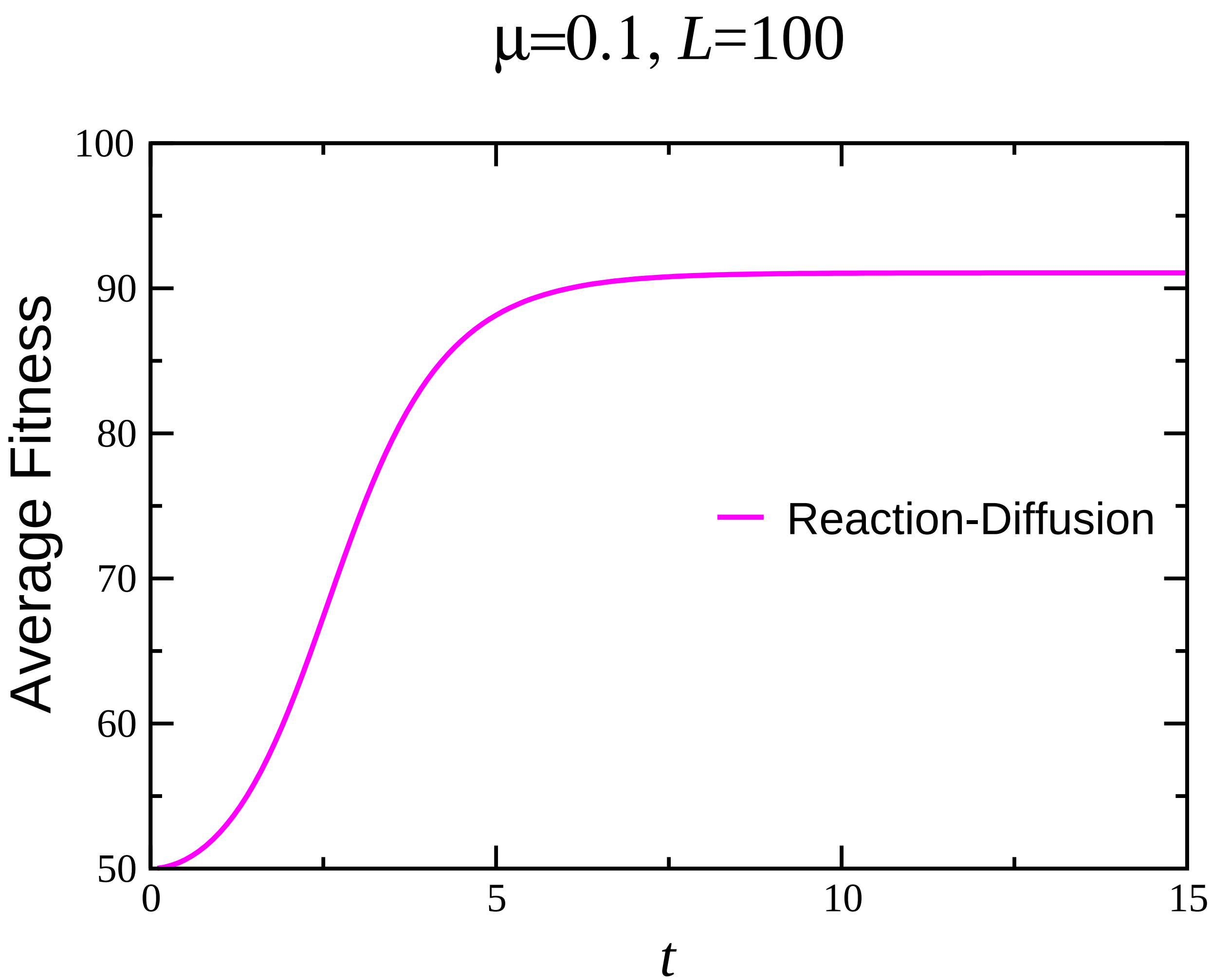
- Selection: Higher Fitness Individuals Grow Faster at Expense of Less Fit
  - ▷ *Drives Waveform to Delta Function at Rightmost Edge*
- Diffusion: Widens the Waveform
  - ▷ *Together with Selection, Drives Population to Higher Fitness*
- Bias: Entropy Favors 50/50 State: $x = L/2$
  - ▷ *Leads to Equilibrium*

Can we solve for this traveling wave?

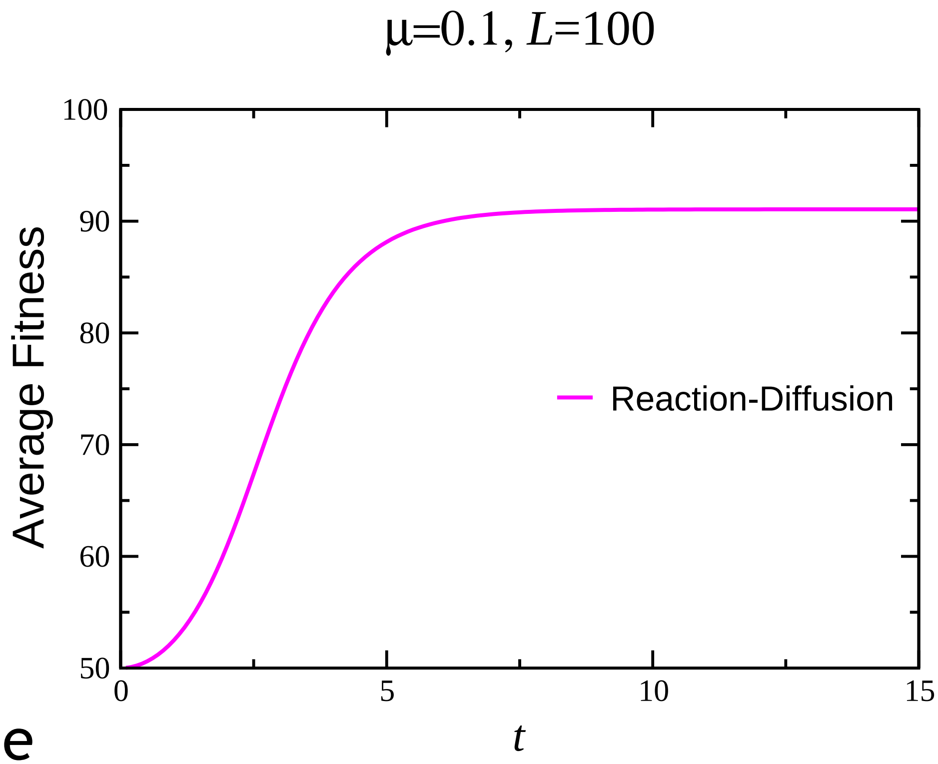
Reaction-Diffusion Approach

Eigen, Shuster

$$\dot{P}(x,t) = (x-\overline{x})P + \mu(xP)'' + \mu\left(1 - 2\frac{x}{L}\right)(xP)'$$



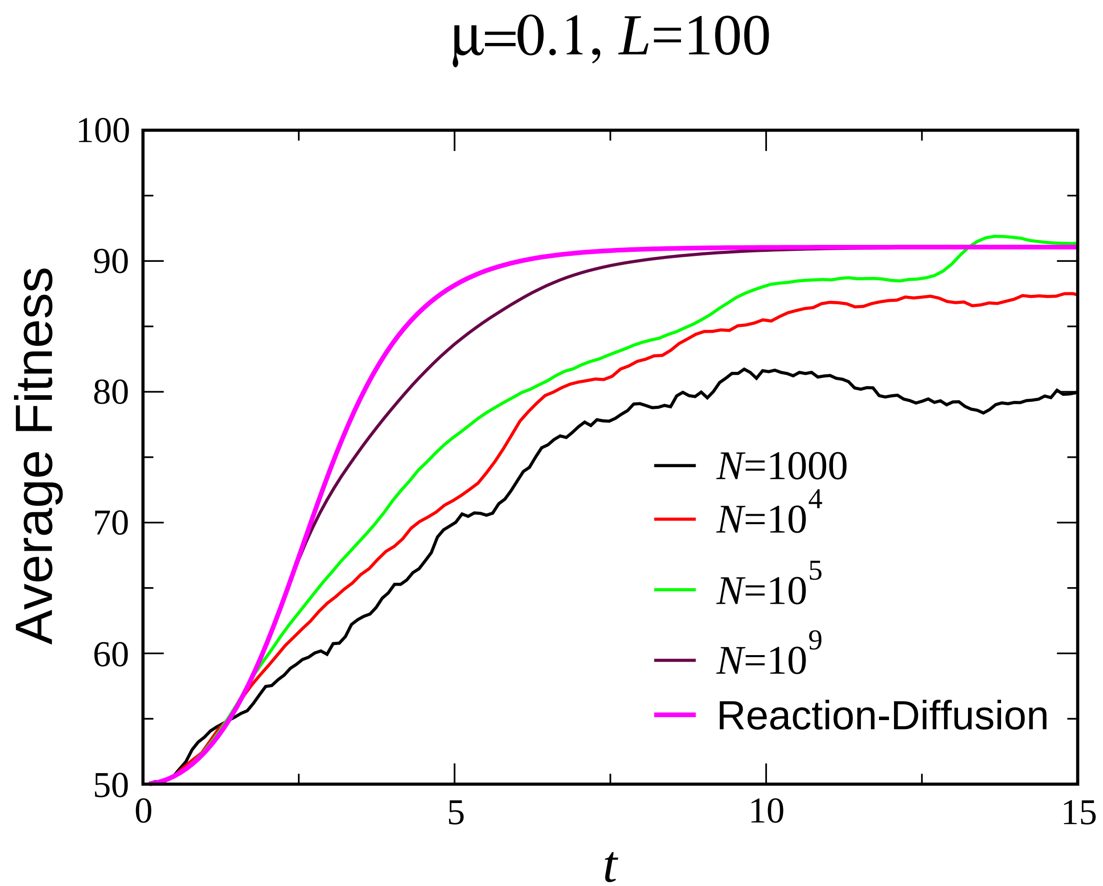
Reaction-Diffusion Approach, continued



- Exactly Solvable
- Velocity increases exponentially for short times
- Velocity peaks at $O(L)$ in times of order $\ln L$
- Velocity falls exponentially to 0, also in times of order $\ln L$

But does it agree with simulation?

Reaction-Diffusion Approach, continued



- Agrees with Simulation, BUT only for $N \sim O(e^L)$, HUGE!!

For Reasonable N, Need Different Approach

The Single-Locus Model, *aka* The $\mu \ll 1$ Limit

Kimura, 1968; Kessler, Levine, Ridgway, Tsimring, 1997

If $\mu \ll 1$, almost all individuals have same fitness

- Selection collapses distribution much faster than mutation can widen it

Assume we start from a state with $N-1$ individuals of fitness x, and one of fitness y.

- i.e., a mutation event just occured

Then, what is the probability that type y will become fixed?

- Ignore additional mutations -- very unlikely
  - ▷ *Two absorbing states: 1) All x ; 2) All y*

Answer: Prob.= $\frac{\left(\frac{y}{x}\right)^{N-1}\left(\frac{y}{x}-1\right)}{\left(\frac{y}{x}\right)^{N}-1}$

- Limits:
  - ▷ $|N(y-x)/x| \ll 1 \quad \Rightarrow \quad \frac{1}{N} + \frac{y-x}{2x}$
  - ▷ $N(y-x)/x \gg 1 \gg (y-x)/x > 0 \quad \Rightarrow \quad \frac{y-x}{x}$
  - ▷ $(y-x)/x \gg 1 \quad \Rightarrow \quad 1 - \frac{x}{y}$
  - ▷ $N(y-x)/x \ll -1 \ll (y-x)/x < 0 \quad \Rightarrow \quad \frac{x-y}{x} e^{-N(x-y)/x}$, *Exponentially small*

The Single-Locus Model (continued)

Implications for velocity:

- $|y-x|=1, x \gg 1 \Rightarrow$ Prob. = $\frac{\left(1\pm\frac{1}{x}\right)^{N-1}\left(\pm\frac{1}{x}\right)}{\left(1\pm\frac{1}{x}\right)^{N}-1} \approx \frac{1}{x(1-e^{N/x})}$, Depends of ratio N/x

- For $N/x \ll 1$
  - ▷ *Rate of up mutations:* $\mu(1-x/L)Nx(\frac{1}{N}+\frac{1}{2x})$
  - ▷ *Rate of down mutations:* $\mu(x/L)Nx(\frac{1}{N}-\frac{1}{2x})$
  - ▷ *Overall average drift speed:* $\mu N/2+\mu x(1-2x/L)$*, increases linearly with* N
  - ▷ *Diffusion constant for center of mass:* μx*, independent of* N*, much larger than drift speed*

- For $N/x \gg 1$
  - ▷ *Rate of up mutations:* $\mu(1-x/L)Nx\left(\frac{1}{x}\right)$
  - ▷ *Rate of down mutations: 0*
  - ▷ *Overall average speed:* $\mu N(1-x/L)$*, increases linearly with* N
  - ▷ *Diffusion constant for center of mass:* μN*, same as drift speed*

The Single-Locus Model (continued)

Variance of distribution = $\mu N/2$, widens with N

- As N increases, single-locus assumption breaks down
- Not enough time for selection to collapse distribution completely before next mutation

For large N, need different approach

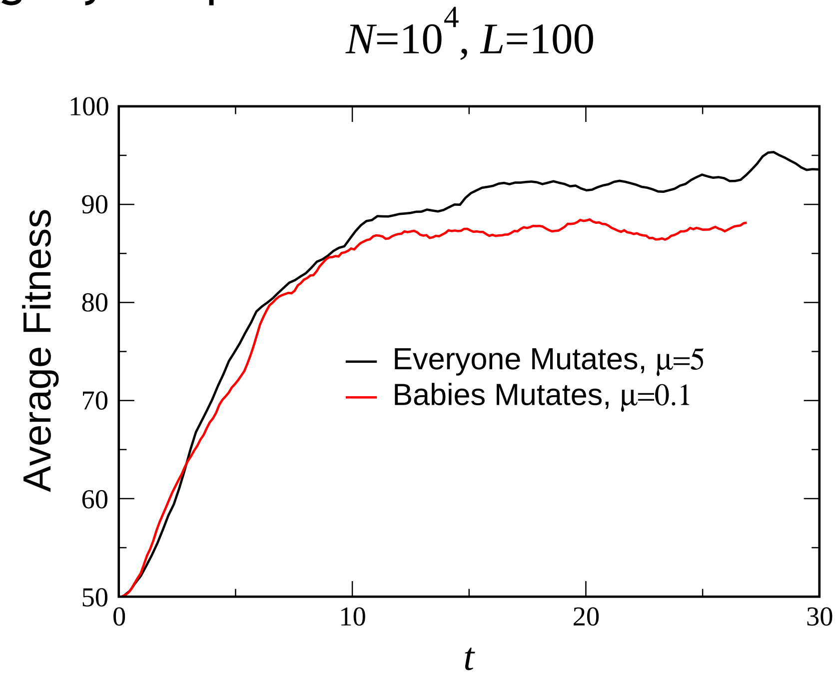
But first...

A Technical Interlude

In the Service of Truth in Advertising

We go over from Babies Mutate Model to Everyone Mutates Model

- Now μ is a *rate* of mutation, not a *probability* of mutation
- Basically, $\mu \rightarrow \mu/\overline{x}$
- No Major Difference in Physics
- Makes Things Slightly Simpler



Cutoff Reaction-Diffusion Equation

Tsimring, Levine, Kessler, 1997; Rouzine, Wakeley, Coffin, 2003

Problem is Mistreatment of Leading Edge

- Leading Edge Has Fastest Growth Rate
- Velocity Controlled by Statistical Fluctuations at Edge
- The Exponentially Small Number of Particles in the Leading Edge in the Reaction-Diffusion Equation Have a Huge Effect
- Reaction-Diffusion Equation Only Valid if $x = L$ has Finite Occupation
  - ▷ *Explains* $N \sim e^L$ *Requirement*
- Similar Physics Occurs in Diffusion-Limited Aggregation

Cutoff Reaction-Diffusion Equation (continued)

The Answer: Cut Off the Growth in the Leading Edge

- No Growth if Average Occupancy < 1 Individual

Various Cutoff Schemes Possible

- Replace $(x - \overline{x})P(x)$ with $\theta(P(x) - 1)(x - \overline{x})P(x)$
  - ▷ $v \sim \ln^{1/3} N$, *not consistent with simulation*
- $P_{k*+1} = 0$ where $k*$ is last site such that $P_{k*} > \frac{1}{\mu}$.
  - ▷ $v \sim \ln N$, *fair agreement with simulation*
- Replace $(x - \overline{x})P(x)$ with $x\theta(P(x) - \epsilon)P(x) - \overline{x}P(x)$
  - ▷ $v \sim \ln N$, *qualitatively consistent with simulation, quantitative agreement if* ϵ *is roughly* $1/4$
- Countless other possibilities

Problem: No way to justify cutoff *a priori*

Would like a more rigorous approach

The Moment Hierarchy

Consider Average Fitness, $E_1 \equiv \frac{1}{N}\sum_{i=1}^{N} x_i$

Eqn. of Motion for $\langle E_1 \rangle$: ($\langle O \rangle \equiv$ Ensemble Average of O)

$$\langle \dot{E}_1 \rangle = \mu \left(1 - 2\frac{\langle E_1 \rangle}{L} \right) + \langle C_2 \rangle$$

- C_2 is the variance of the fitness: $C_2 \equiv \frac{1}{N}\sum x_i^2 - E_1^2 \equiv E_2 - E_1^2$

To proceed, we need Eqn. of Motion for $\langle C_2 \rangle$:

$$\langle \dot{C}_2 \rangle = \mu \left(1 - 4\frac{\langle C_2 \rangle}{L} \right) + \langle C_3 \rangle - \frac{\langle C_3 + 2E_1 C_2 \rangle}{N}$$

- C_3 is the skewness of x: $C_3 \equiv E_3 - 3E_1E_2 + 2E_1^3$

N is large, so maybe we can drop the $\frac{1}{N}$ term?

The Moment Hierarchy (continued)

$$\langle \dot{E}_1 \rangle = \mu \left(1 - 2\frac{\langle E_1 \rangle}{L}\right) + \langle C_2 \rangle$$

$$\langle \dot{C}_2 \rangle = \mu \left(1 - 4\frac{\langle C_2 \rangle}{L}\right) + \langle C_3 \rangle - \frac{\langle C_3 + 2E_1C_2 \rangle}{N}$$

N is large, so maybe we can drop the $\frac{1}{N}$ term?

- NO! If we do, we recover (after considering all $\dot{C}_k$), EXACTLY the Reaction-Diffusion Equation!
  - ▷ *C_3 is driven by C_4, C_4 by C_5, etc.*
  - ▷ *Produces exponential growth of all moments*
  - ▷ *Continues until $1/L$ terms set it*
  - ▷ *All C's grow to size $O(L)$, and then decay*
- $\frac{1}{N}$ term is a *singular* perturbation
  - ▷ *Responsible for cutting off the growth of the C's*

Consider Next Moment Equation:

$$\langle \dot{C}_3 \rangle = \mu \left(1 - \frac{\langle 2E_1 + 6C_3 \rangle}{L}\right) + \langle C_4 \rangle - \frac{\langle 3C_4 + 6E_1C_3 + 6E_2^2 \rangle}{N}$$

The Moment Hierarchy (continued)

$$\langle \dot{E}_1 \rangle = \mu \left(1 - 2\frac{\langle E_1 \rangle}{L} \right) + \langle C_2 \rangle$$

$$\langle \dot{C}_2 \rangle = \mu \left(1 - 4\frac{\langle C_2 \rangle}{L} \right) + \langle C_3 \rangle - \frac{\langle C_3 + 2E_1C_2 \rangle}{N}$$

Consider Next Moment Equation:

$$\langle \dot{C}_3 \rangle = \mu \left(1 - \frac{\langle 2E_1 + 6C_3 \rangle}{L} \right) + \langle C_4 \rangle - \frac{\langle 3C_4 + 6E_1C_3 + 6E_2^2 \rangle}{N}$$

- $\frac{1}{N}$ suppression term is stronger

Ansatz: Depending on N, all C_k are suppressed beyond some $k^*(N)$

- $k^*(N) \sim \ln N$

$\Rightarrow$ Can truncate hierarchy beyond $k^*(N)$!

- Physically, Dynamics can not be sensitive to very high moments of initial condition
- $\frac{1}{N}$ plays the role of surface tension in Saffman-Taylor

The Moment Hierarchy (continued)

$$\langle \dot{E}_1 \rangle = \mu \left(1 - 2\frac{\langle E_1 \rangle}{L} \right) + \langle C_2 \rangle$$

$$\langle \dot{C}_2 \rangle = \mu \left(1 - 4\frac{\langle C_2 \rangle}{L} \right) + \langle C_3 \rangle - \frac{\langle C_3 + 2E_1 C_2 \rangle}{N}$$

$$\langle \dot{C}_3 \rangle = \mu \left(1 - \frac{\langle 2E_1 + 6C_3 \rangle}{L} \right) + \langle C_4 \rangle - \frac{\langle 3C_4 + 6E_1 C_3 + 6E_2^2 \rangle}{N}$$

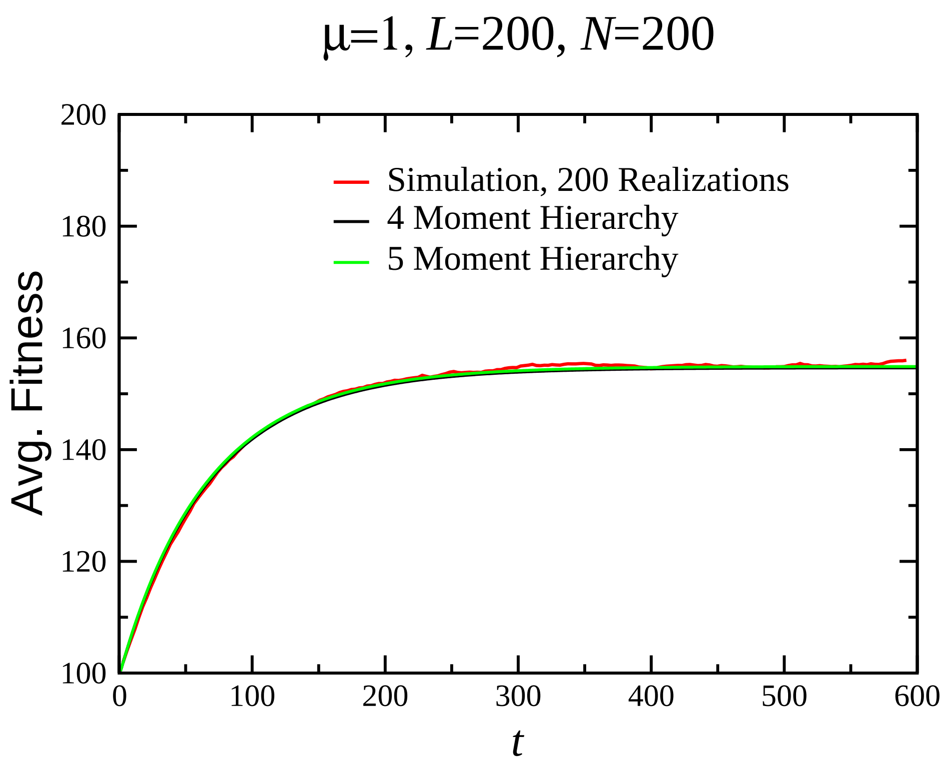
1 Remaining Problem: What to do with Products $\langle E_1 C_2 \rangle$, $\langle E_1 C_3 \rangle$, *etc.*?

Ansatz #2: Factorize: $\langle E_1 C_2 \rangle \to \langle E_1 \rangle \langle C_2 \rangle$, *etc.*

- Formally, Connected Correlators should be $\sim \frac{1}{N}$

Algorithm: Truncate Moment Hierarchy at same point, Factorize, Solve

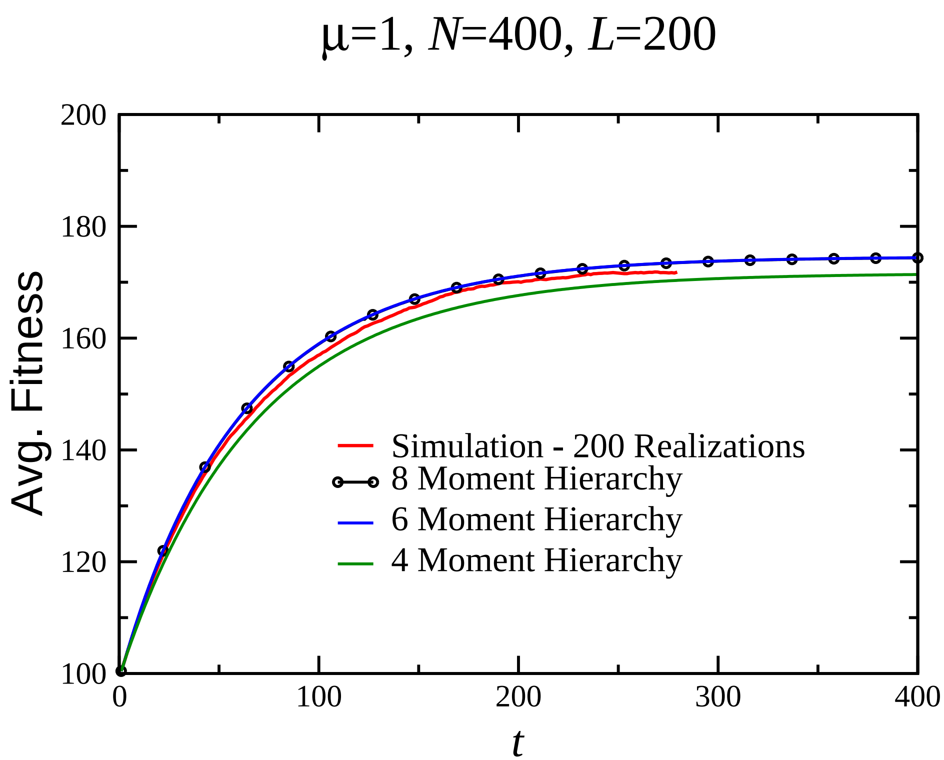
And the Results are . . .



Adding 5'th Moment to Hierarchy Has Essentially No Effect!

Agreement with Simulation: Perfect!

And the Results are . . .



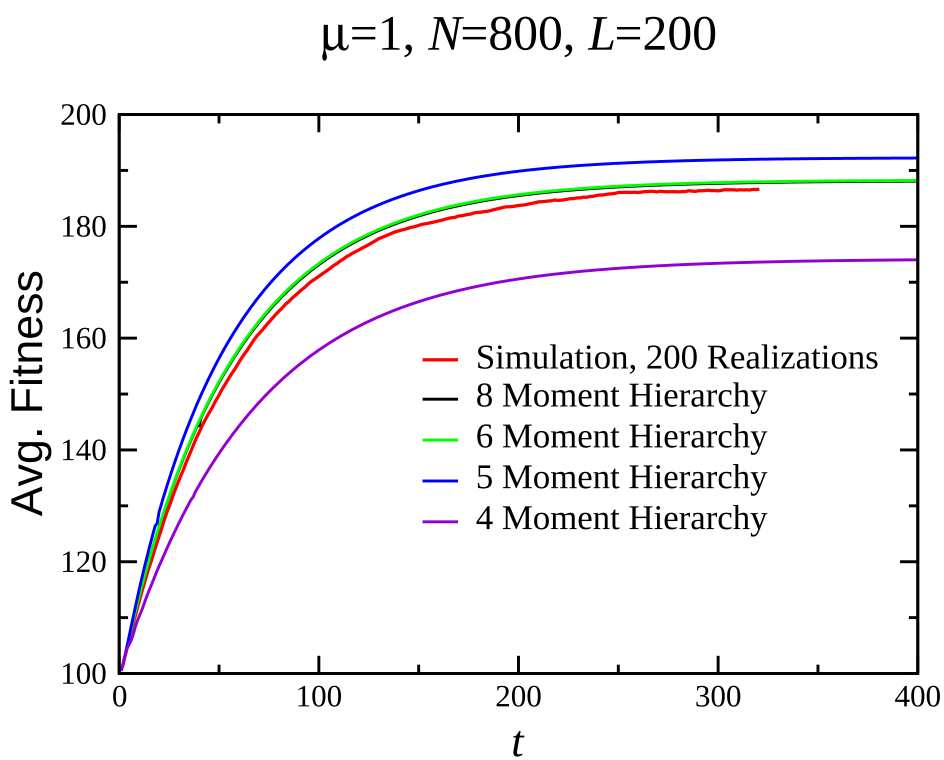
With Larger N, Have to go to 6'th Moment

Going to 8'th Moment Has Essentially No Effect!

Agreement with Simulation:

Perfect!

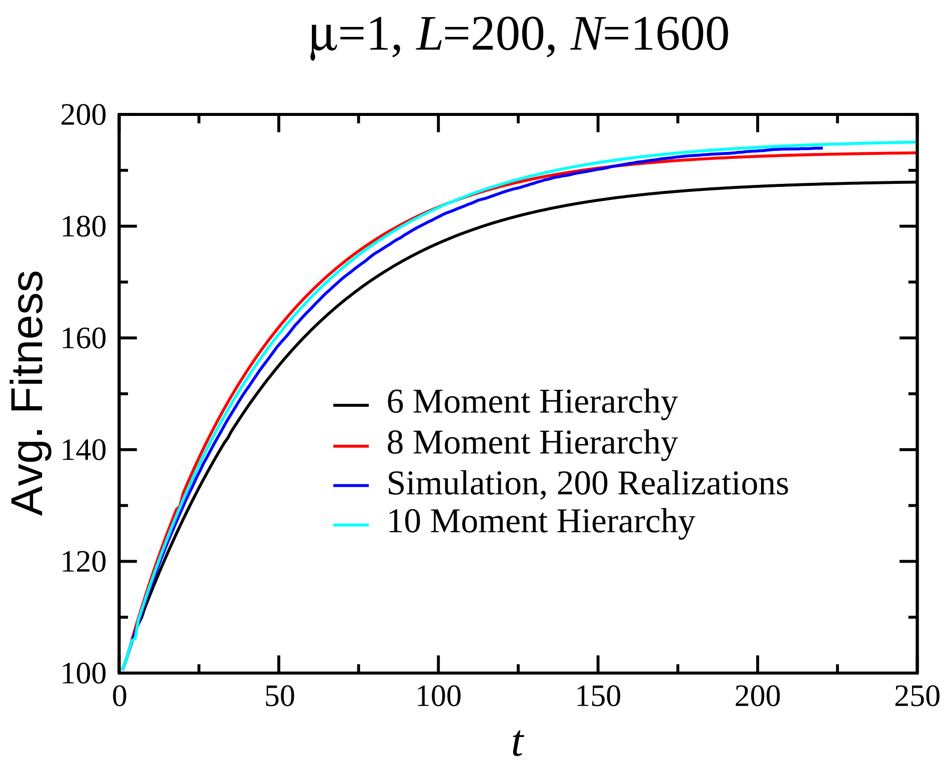
And the Results are . . .



6'th and 8'th Moments Agree

Agreement with Simulation:
Not Quite Perfect!!??!!

And the Results are ...



Now have to go to 10th Moment

Agreement with Simulation:

Not Quite Perfect!!??!!

What is Going On?

Appears to Be Due to Failure of Factorization

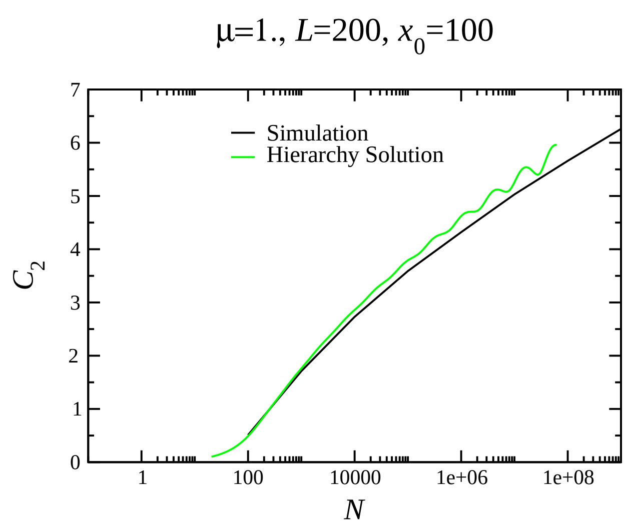
- Surprisingly, Problem Gets Worse with Increasing N

Is There a Way to Doctor This Up?

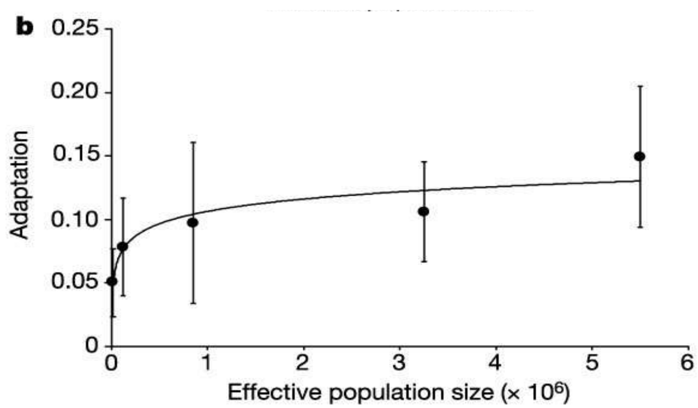
Wait and See!

Nevertheless, Basic Trends OK

Velocity $\sim \ln N$



Agrees with Experiment?



Chlamydomonas
From Colegrave
Nature (2002)

Fit is to $N^{.16}$
Clearly Log is OK

References

Evolution:

- Phys. Rev. Lett. **76**, 4440 (1996).
- J. Stat. Phys. **87**, 519, (1997).
- J. Stat. Phys. **90**, 191 (1998).
- Phys. Rev. Lett. **80**, 2012 (1998).

Other Examples of Flucutation-Dominated Systems

- Philos. Mag. **B77**, 1313 (1998).
- Phys. Rev. **E58**, 107, (1998).
- Nature **394**, 556 (1998).

Take Home Messages

- Multi-Locus Evolution Model Rich and Interesting Problem
- Finite Population Acts as a Singular Cutoff
- Truncated, Factorized Moment Expansion Captures Essential Physics
- Full Solution Still Awaits Us