

# Analysis and Simulation Of A Simple Evolutionary System

Mahendra Duwal Shrestha

# Overview

Background

Research Question 1

Diploid Model

Simulation and Results

Conclusion

Research Question 2

Infinite Population Limits

Simulation and Results

Conclusion

# Overview

Research Question 3

Violation in Mutation

Simulation and Results

Conclusion

Research Question 4

Violation in Crossover

Simulation and Results

Conclusion

# Terms

Crossover : Choose a random point in two parents and exchange subsequences after that point to create two offspring.

Ex: **11001011+11011111** -> {**11001111**, 11011011}

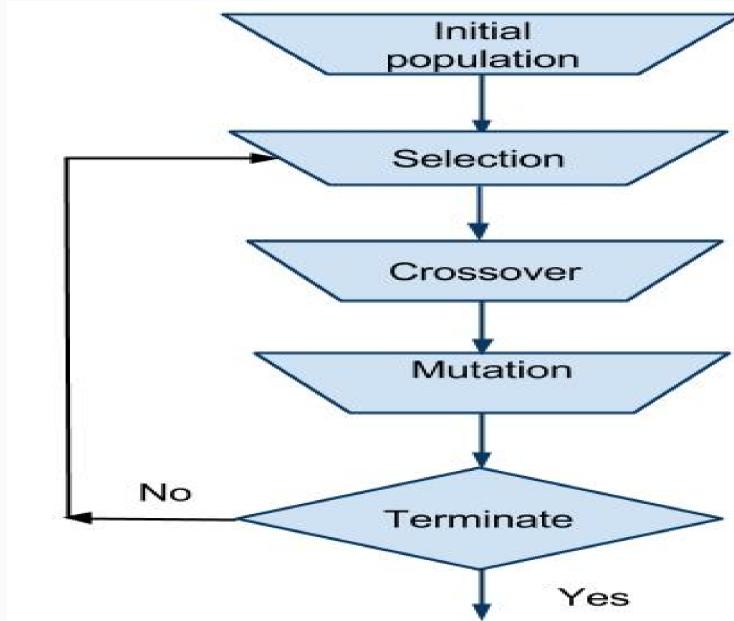
Mutation: Flip bits of an individual with some small probability, the mutation rate.

Ex: 11001001 + 01000000 -> 10001001  
mutation mask

Population vector: Population can be represented by population vector  $\mathbf{p}$  where  $p_j$  is the proportion of string  $j$  in the population.

If population  $P = \{00, 01, 01, 10, 11, 11\}$ , then  $p_3 = 2/6 = 1/3$

# Finite Population GA



Two parents  $u$  and  $v$  selected to produce two offspring  $u'$  and  $v'$  through crossover and mutation

Keep one of  $u'$  and  $v'$  with uniform probability

Replace  $P$  by new generations formed

Process is repeated until system stops to improve or threshold is reached

# Infinite Population Model

Population modeled as population vector  $\mathbf{p}$

$\mathcal{G}$  is the function mapping  $\mathbf{p}$  to the next generation

$\mathcal{G}(\mathbf{p})_j$  = the probability that string  $j$  occurs in the next generation

The evolution of  $\mathbf{p}$  in infinite population model is the sequence

$$\mathbf{p} \rightarrow \mathcal{G}(\mathbf{p}) \rightarrow \mathcal{G}(\mathcal{G}(\mathbf{p})) \rightarrow \dots$$

# History

Several people working in the 1950s and the 1960s – like Box (1957), Friedman(1959), Bledsoe (1961), Bremermann (1962), and Reed, Toombs and Baricelli (1967) in evolution-inspired algorithms

Holland and colleagues formalized and promoted population based algorithms with crossover and mutation

In 1999, Vose developed and formalized genetic algorithm

# Random Heuristic Search

$\tau$  is stochastic transition rule that maps  $\mathbf{p}$  to  $\mathbf{p}'$

Finite population sequence  $\mathbf{p}, \tau(\mathbf{p}), \tau^2(\mathbf{p}), \dots$  forms Markov chain

The next population vector  $\tau(\mathbf{p})$  cannot be predicted by certainty

$\mathcal{G}(\mathbf{p})$  is the expected next generation  $\mathcal{E}(\tau(\mathbf{p}))$

$\mathcal{G}(\mathbf{p})_j =$  the probability that string  $j$  occurs in the next generation

The variance of the next generation (with respect to the expected population) is

$$\mathcal{E}(\|\tau(\mathbf{p}) - \mathcal{G}(\mathbf{p})\|^2) = \frac{1 - \|\mathcal{G}(\mathbf{p})\|^2}{r}$$

# Research Question 1:

## Distance Between Finite Population and Infinite Population

# Distance: Chebyshev's Inequality

Let  $\epsilon = f(r)/\sqrt{r}$ , where  $f(r)$  grows arbitrarily slowly and

$$\lim_{r \rightarrow \infty} f(r) = \infty$$

From Chebyshev's inequality,

$$\lim_{r \rightarrow \infty} P(\|\tau(\mathbf{p}) - \mathcal{G}(\mathbf{p})\| \geq \epsilon) \leq \lim_{r \rightarrow \infty} \frac{1 - \|\mathcal{G}(\mathbf{p})\|^2}{f(r)^2} = 0$$

This suggests distance between  $\tau(\mathbf{p})$  and  $\mathcal{G}(\mathbf{p})$  might decrease as  $1/\sqrt{r}$

# Distance: Population Points

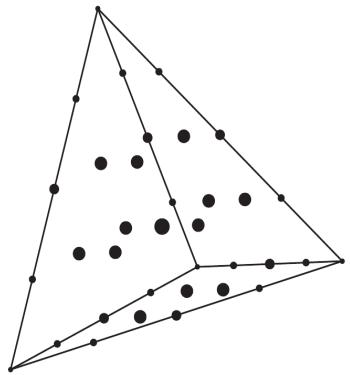


Figure 1.2: Population points

Finite populations are represented by dots

Infinite population can be anywhere in the space

Distance between finite population and infinite population is bounded by  $O(1/\sqrt{r})$

This suggests distance between  $\tau(\mathbf{p})$  and  $\mathcal{G}(\mathbf{p})$  might decrease as  $1/\sqrt{r}$

# Distance: Jensen's Inequality

Let  $\eta$  be the random variable  $||\tau(\mathbf{p}) - \mathcal{G}(\mathbf{p})||$

Then from Jensen's Inequality,

$$\mathcal{E}(||\tau(\mathbf{p}) - \mathcal{G}(\mathbf{p})||) = \mathcal{E}(\eta) \leq \sqrt{\mathcal{E}(\eta^2)} = \frac{\sqrt{1 - \|\mathcal{G}(\mathbf{p})\|^2}}{\sqrt{r}}$$

This also suggests distance might decrease as  $1/\sqrt{r}$

# Diploid Population Model

Diploid genome  $\alpha = \langle \alpha_0, \alpha_1 \rangle$

Haploid genome length /

$q^n$  → population at generation  $n$        $q_\alpha^n$  → prevalence of diploid  $\alpha$

$t_\alpha(g)$  → probability that gamete  $g$  is produced from parent  $\alpha$

$$q_\gamma^{n+1} = \sum_{\alpha} q_\alpha^n t_\alpha(\gamma_0) \sum_{\beta} q_\beta^n t_\beta(\gamma_1)$$

# Diploid Model Reduction to Haploid Model

$$q_{\langle \gamma_0, \gamma_1 \rangle}^{n+1} = p_{\gamma_0}^{n+1} p_{\gamma_1}^{n+1}$$

Where for any haploid  $\gamma_0$ ,

$$p_{\gamma_0}^{n+1} = \sum_{\alpha} q_{\alpha}^n t_{\alpha}(\gamma_0)$$

Evolution equation in terms of haploid distributions  $p$ ,

$$p_{\gamma_0}^{n+1} = \sum_{\alpha_0, \alpha_1} p_{\alpha_0}^n p_{\alpha_1}^n t_{\langle \alpha_0, \alpha_1 \rangle}(\gamma_0)$$

The weighted count of haploid  $g$  in generation  $n$  is,

$$\begin{aligned} & \sum_{\alpha_0, \alpha_1} q_{\langle \alpha_0, \alpha_1 \rangle}^n ([g = \alpha_0] + [g = \alpha_1]) \\ &= \sum_{\alpha_0, \alpha_1} p_{\alpha_0}^n p_{\alpha_1}^n [g = \alpha_0] + \sum_{\alpha_0, \alpha_1} p_{\alpha_0}^n p_{\alpha_1}^n [g = \alpha_1] \\ &= 2p_g^n \end{aligned}$$

Matrix form:  $p'_g = p^T M_g p$  where  $(M_g)_{u,v} = t_{\langle u,v \rangle}(g)$

# Specialization to Vose's Haploid Model

Mutation:  $\mu_i = (\boldsymbol{\mu})^{\mathbf{1}^T i} (1 - \boldsymbol{\mu})^{\ell - \mathbf{1}^T i}$

Crossover:  $\chi_i = \begin{cases} \chi^{c_i} & \text{if } i > 0 \\ 1 - \chi + \chi^{c_0} & \text{if } i = 0 \end{cases}$

For uniform crossover,  
 $c_i = 2^{-\ell}$

$$t_{\langle u, v \rangle}(g) = \sum_{i \in \mathcal{R}} \sum_{j \in \mathcal{R}} \sum_{k \in \mathcal{R}} \mu_i \mu_j \frac{\chi_k + \chi_{\bar{k}}}{2} [k(u+i) + \bar{k}(v+j) = g] \quad (\text{By Vose \& Wright})$$

$\mathcal{R}$  is set of length  $\ell$  binary strings

# Walsh Basis

$$W_{n,t} = N^{-1/2} (-1)^{n^T t}$$

$$\widehat{A} = WAW$$

$$\widehat{w} = Ww$$

Mixing matrix in Walsh basis

$$\widehat{M}_{u,v} = 2^{\ell-1} [uv = \mathbf{0}] \widehat{\boldsymbol{\mu}}_u \widehat{\boldsymbol{\mu}}_v \sum_{k \in \overline{u+v}\mathcal{R}} \boldsymbol{\chi}_{k+u} + \boldsymbol{\chi}_{k+v}$$

Evolution eqn in Walsh basis

$$\widehat{p}'_g = 2^{\ell/2} \sum_{i \in g\mathcal{R}} \widehat{p}_i \widehat{p}_{i+g} \widehat{M}_{i,i+g} \quad \text{where } g\mathcal{R} = \{gi \mid i \in \mathcal{R}\}$$

# Distance Computation

Naive computation

$$\|\mathbf{f} - \mathbf{q}\| = \sum_{\alpha}^{\frac{1}{2}} (\mathbf{f}_{\alpha} - \mathbf{q}_{\alpha})^2 \longrightarrow 2^{\ell} \cdot 2^{\ell} \text{ terms}$$

Our implementation

$$S_f = \{\alpha \mid f_{\alpha} > 0\}$$

$$\|\mathbf{f} - \mathbf{q}\|^2 = \sum_{\alpha \notin S_f} (\mathbf{f}_{\alpha} - \mathbf{q}_{\alpha})^2 + \sum_{\alpha \in S_f} (\mathbf{f}_{\alpha} - \mathbf{q}_{\alpha})^2$$

Using  $\mathbf{q}_{\alpha} = p_{\alpha_0} p_{\alpha_1}$  and simplifying,

$$\begin{aligned} \|\mathbf{f} - \mathbf{q}\|^2 &= \sum_g^2 (\mathbf{p}_g)^2 + \sum_{\alpha \in S_f} (\mathbf{f}_{\alpha} - \mathbf{q}_{\alpha})^2 - \sum_{\alpha \in S_f} (\mathbf{q}_{\alpha})^2 \\ &= \sum_g^2 (\mathbf{p}_g)^2 + \sum_{\alpha \in S_f} \mathbf{f}_{\alpha}(\mathbf{f}_{\alpha} - 2\mathbf{q}_{\alpha}) \longrightarrow 2^{\ell} + |S_f| \text{ terms} \end{aligned}$$



## Computational advantages of Reduction to Haploid Model and Walsh Transform

Specialization simplifies computation, which otherwise for diploid case would have been impractical

Only one mixing matrix as opposed to  $2^l$  is needed to compute next generation

For  $l = 14$ ,  $2^{14}$  mixing matrices with each having  $2^{14} \times 2^{14}$  entries would require 32 terabytes of memory otherwise; whereas one mixing matrix of 2 gigabytes is required

# Distance: Convergence

$$\begin{aligned} X &= 0.1 \\ \mu &= 0.001 \end{aligned}$$

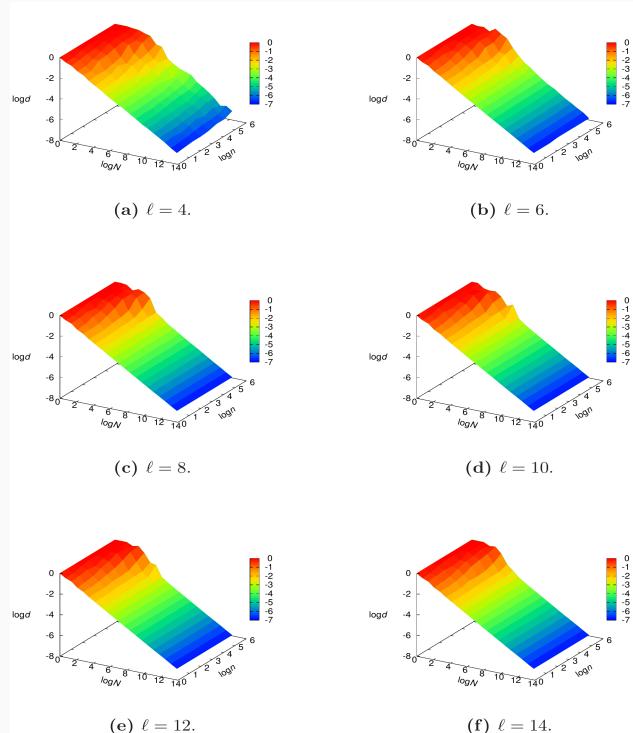
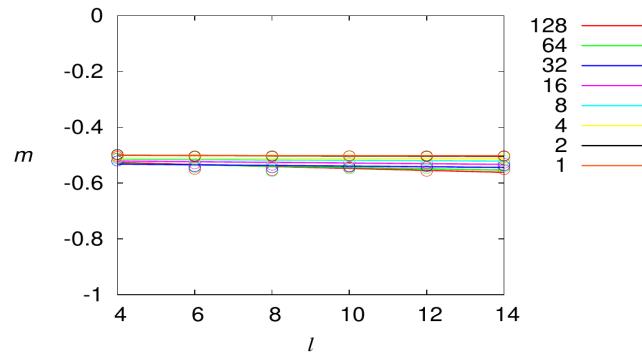


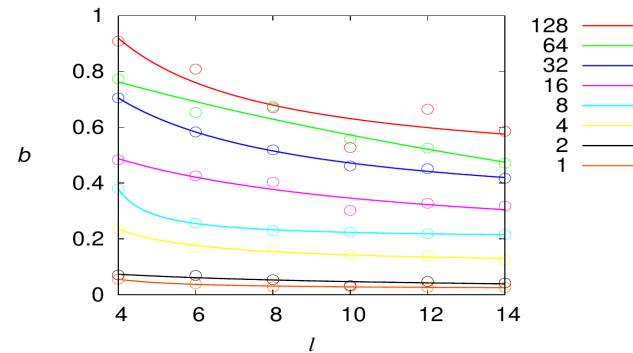
Figure: Convergence of finite population behavior

## Distance: Regression

$$\log d = m \log N + b$$



(a) Slope  $m$ , genome length  $\ell$ .



(b) Intercept  $b$ , genome length  $\ell$ .

Figure: Regression parameters for generation  $n$  in  $\{1, 2, 4, 8, 16, 32, 64, 128\}$

$$d \approx N^m e^b$$

From figure (a) above,  $m \approx -0.5$

$$d \approx k / \sqrt{N}$$

## Distance: Conclusion

Vose's infinite population model can be extended to diploid case

Distance between finite diploid population and infinite diploid population decrease like  $1/\sqrt{N}$

# Research Question 2:

Oscillation in Finite Population Evolution

## Limits

If  $\mathcal{G}$  is continuously differentiable,  
the sequence  $\mathbf{p}, \mathcal{G}(\mathbf{p}), \mathcal{G}^2(\mathbf{p}), \dots$  converges to a fixed point (also called limit)

$$\mathcal{G}(\omega) = \lim_{n \rightarrow \infty} \mathcal{G}^n(\mathbf{p}) = \omega$$

But under some circumstances, the sequence converges to a periodic orbit that oscillate between two fixed points;  $\mathbf{p}^*$  and  $\mathbf{q}^*$

## Necessary and Sufficient Conditions to Converge to a Periodic Orbit

For any  $g \in \mathcal{R}$ ,  $g \neq 0$

$$-1 = \sum_j (-1)^{g^T j} \boldsymbol{\mu}_j$$

$$1 = \sum_{k \in \bar{g}\mathcal{R}} \chi_{k+g} + \chi_k$$

# Does Finite Population Evolution Also Oscillate?

If finite population follows infinite population behavior,  
does finite population also exhibit oscillation?

Simulations run for both haploid and diploid population simulation

Genome lengths  $l \in \{8, 10, 12, 14\}$  were used

Random initial population

Finite population size 4096, 40960, 81920 considered

Distance between fixed points and population used to visualize oscillation

# Results: Oscillation



Figure: Oscillation for  $l = 10$

# Results: Oscillation

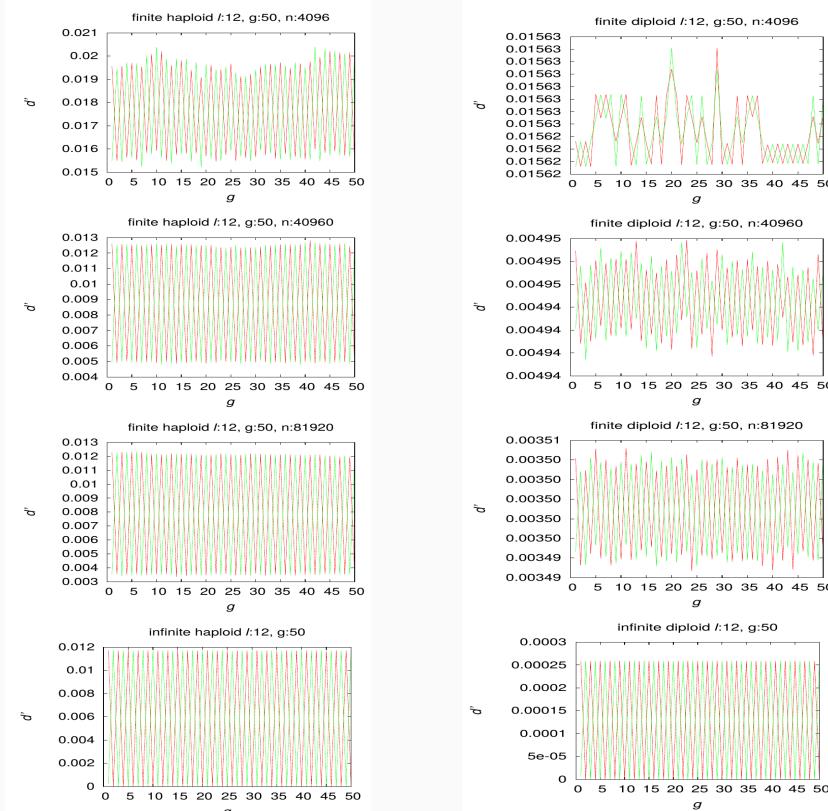


Figure: Oscillation for  $l = 12$

## Results: Oscillation

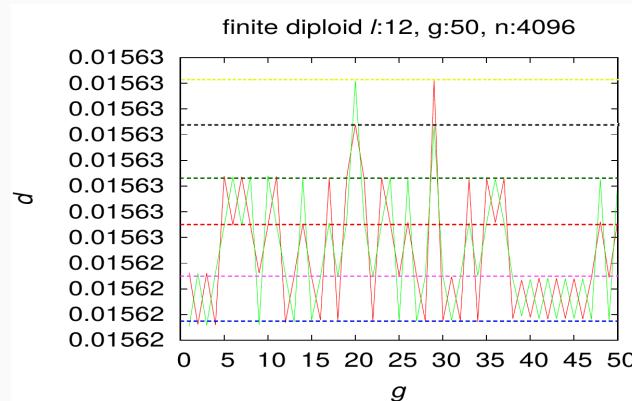


Figure: Jumps in oscillation

# Results: Oscillation Amplitude

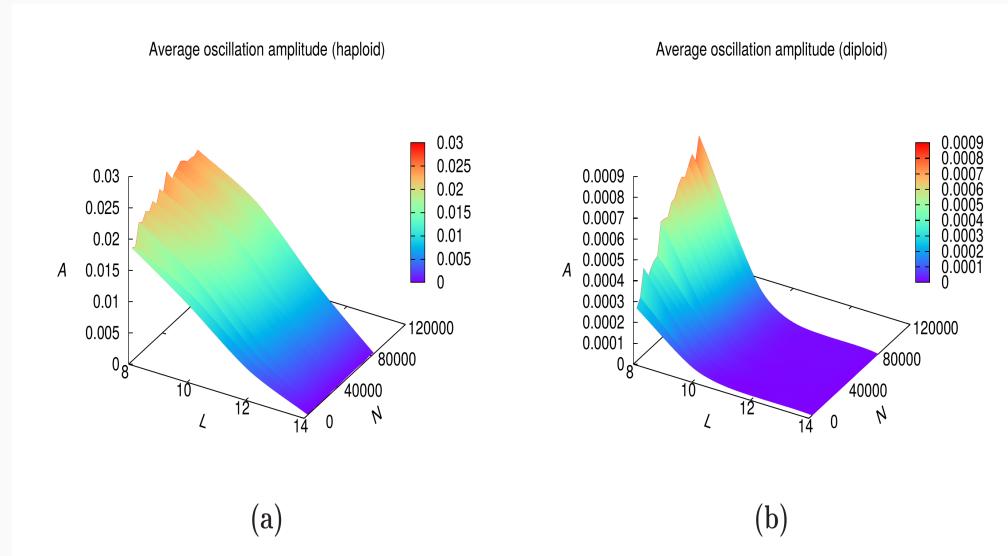


Figure: Average oscillation amplitude

Oscillation amplitude increases with increase in population size

Amplitude of oscillation decreases with increase in  $L$

## Oscillation: Conclusion

Finite population evolution exhibit approximate oscillation

Larger population exhibit better oscillation

# Research Question 3:

Oscillation Under Violation in Mutation

## Robustness of Finite Population

A Markov chain is said to be irreducible if it is possible to get to any state from any state

A Markov chain is aperiodic if it can return to state  $i$  at irregular times

Markov chain is regular if it is both irreducible and aperiodic

Steady state distribution exists if Markov chain is regular

No periodic orbit exists for infinite population

Can finite population exhibit approximate oscillation?

## Violation in Mutation

Violation  $\epsilon$  is introduced in  $\mu$

$$\mu_0 = \epsilon$$

$$\mu_i := (1 - \epsilon)\mu_i$$

This modification makes the Markov chain regular

## Simulation

$\epsilon = \{0.01, 0.1, 0.5\}$

$l = \{8, 10, 12, 14\}$

$N = \{4096, 40960, 81920\}$

Distances of population to limits without violation  $p^*$  **and**  $q^*$  are plotted

# Results:

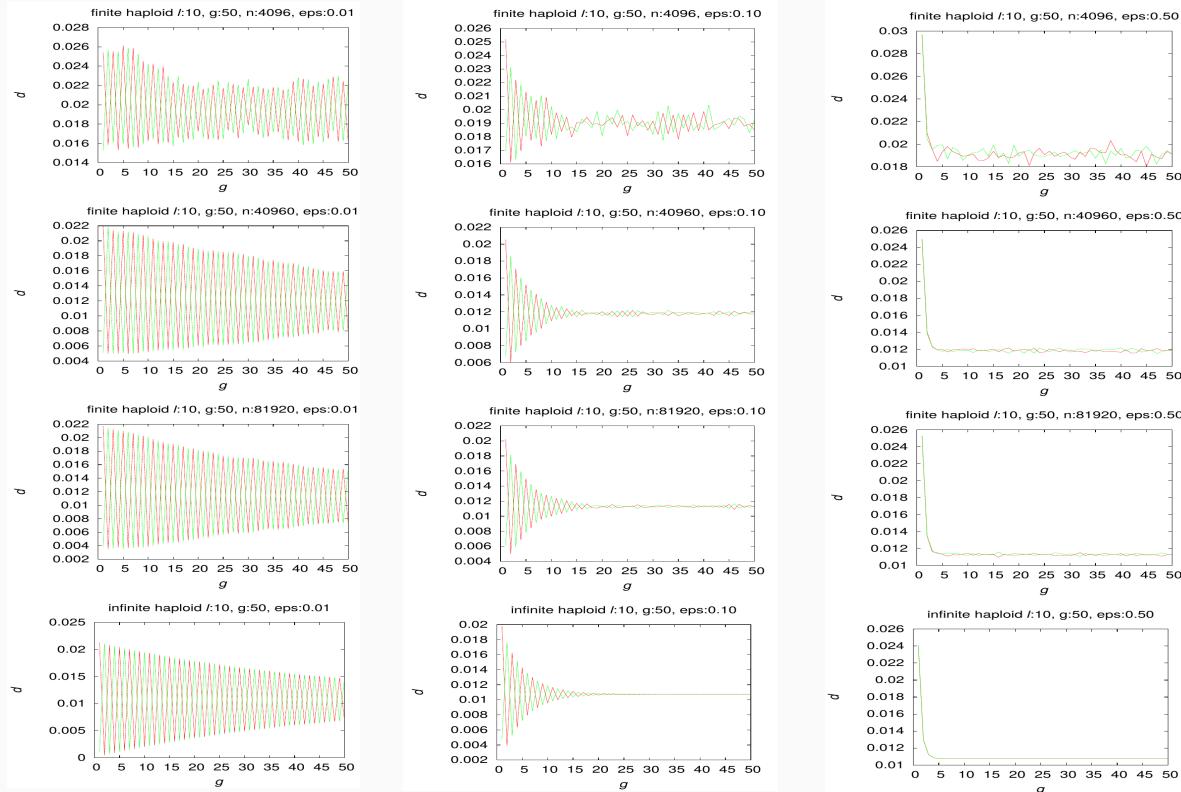


Figure: Oscillation under violation in mutation for haploids,  $l = 10$

# Results:

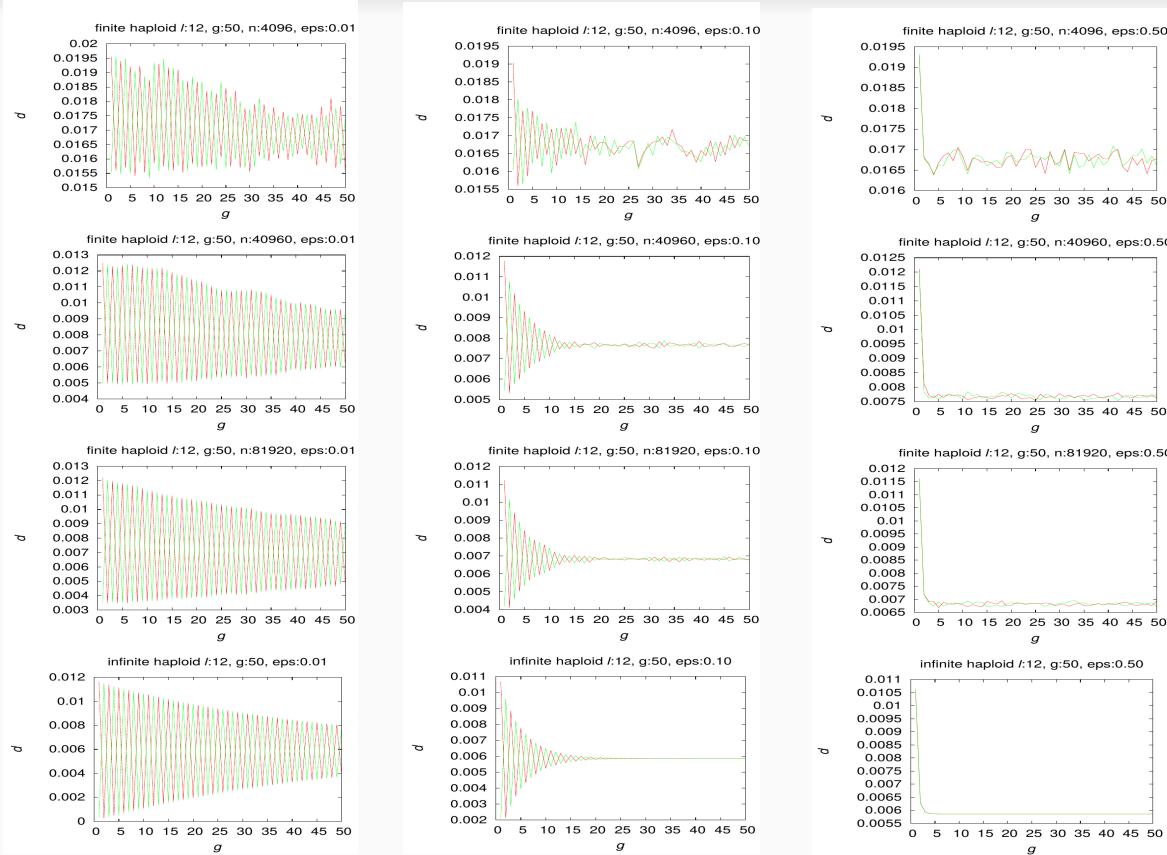


Figure: Oscillation under violation in mutation for haploids,  $l = 12$

# Results

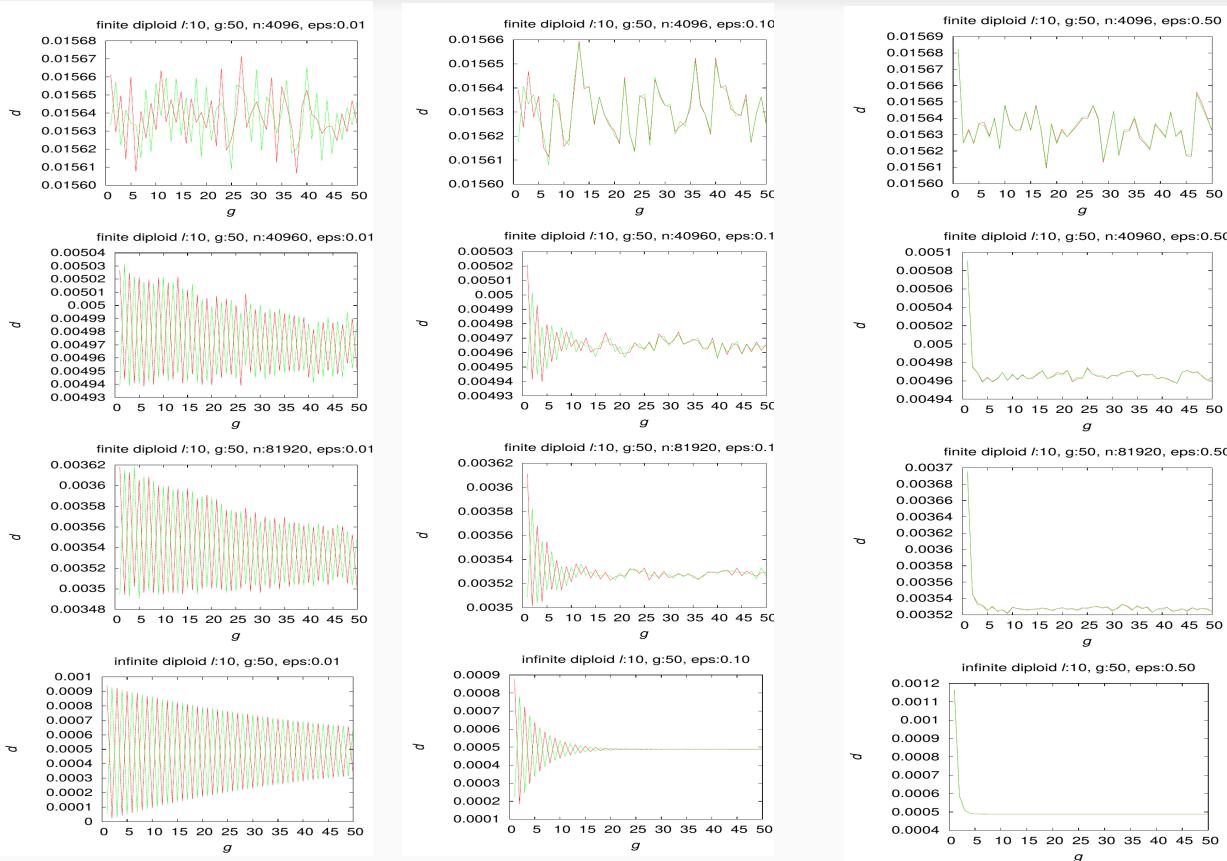


Figure: Oscillation under violation in mutation for diploids,  $l = 10$

# Results

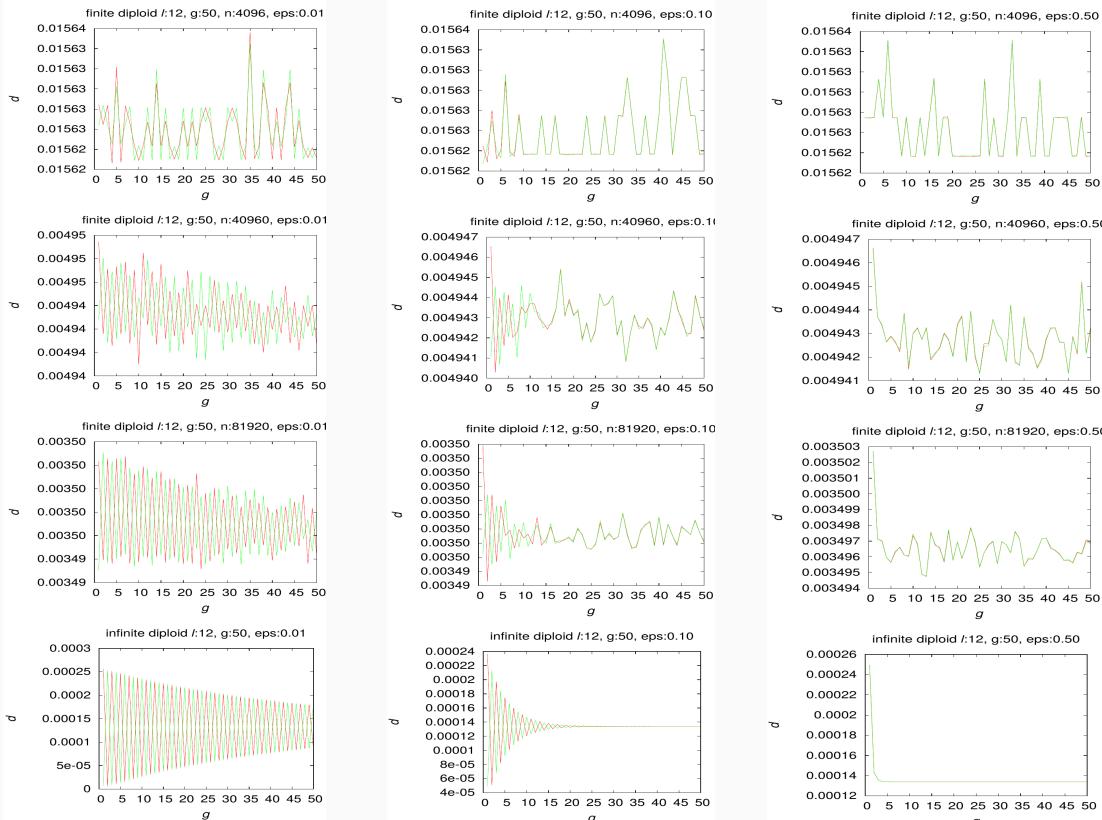


Figure: Oscillation under violation in mutation for diploids,  $l = 12$

## Violation in Mutation: Conclusion

Finite population evolution exhibits approximate oscillation if the violation level is small

If violation is large, then finite population ceases to oscillate

As string length increases, oscillation degrades

As larger population size increases, better oscillations observed

But also increase in population size shows convergence to infinite population behavior

# Research Question 4:

Oscillation under Violation in Crossover

## Robustness of Finite Population

Violation in crossover condition means no periodic orbit exists for infinite population

Can finite population exhibit approximate oscillation?

## Violation in Crossover

Violation  $\epsilon$  is introduced in  $\chi$

$$\chi_i := (1 - \epsilon)\chi_i$$

$$\chi_j = \epsilon \quad \text{where } \chi_j = 0$$

## Simulation

$\epsilon = \{0.01, 0.1, 0.5\}$

$l = \{8, 10, 12, 14\}$

$N = \{4096, 40960, 81920\}$

Distances of population to limits without violation  $p^*$  **and**  $q^*$  are plotted

# Results

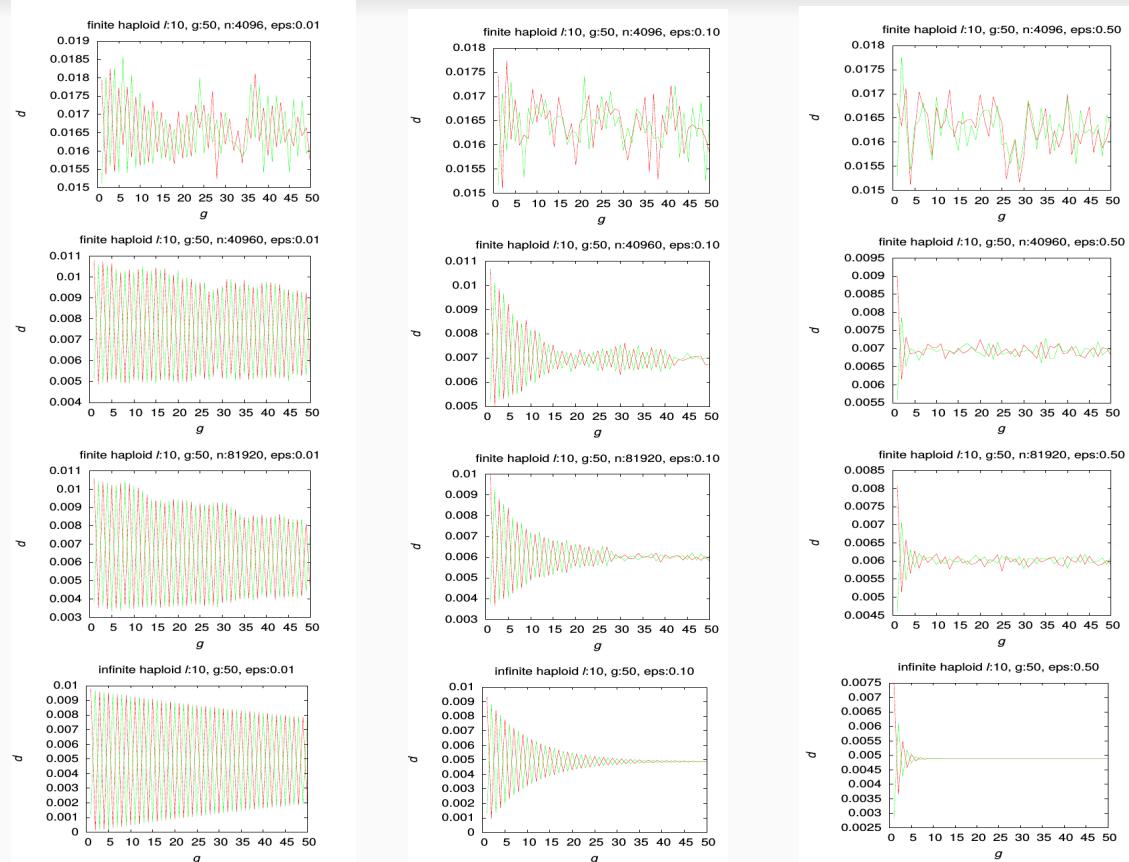


Figure: Oscillation under violation in crossover for haploids,  $l = 10$

# Results

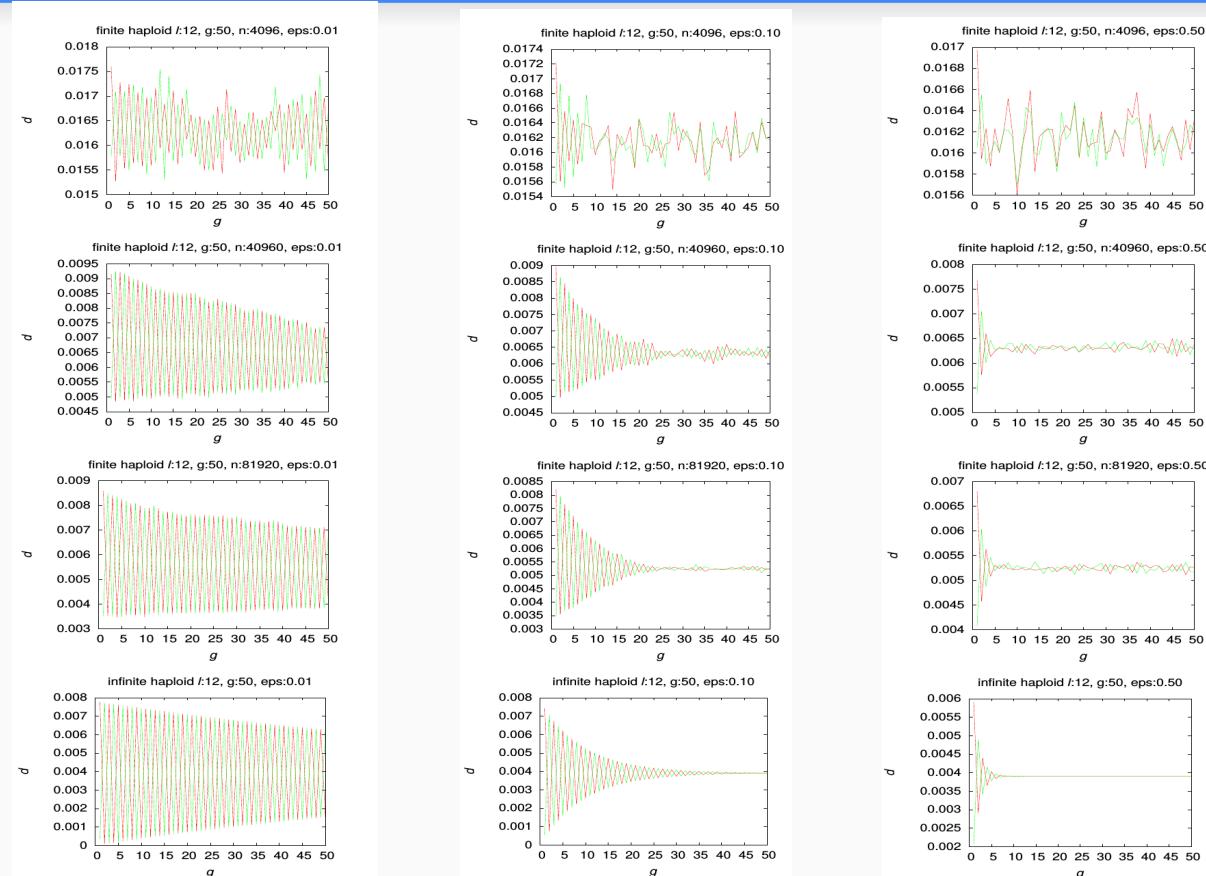


Figure: Oscillation under violation in crossover for haploids,  $l = 12$

# Results

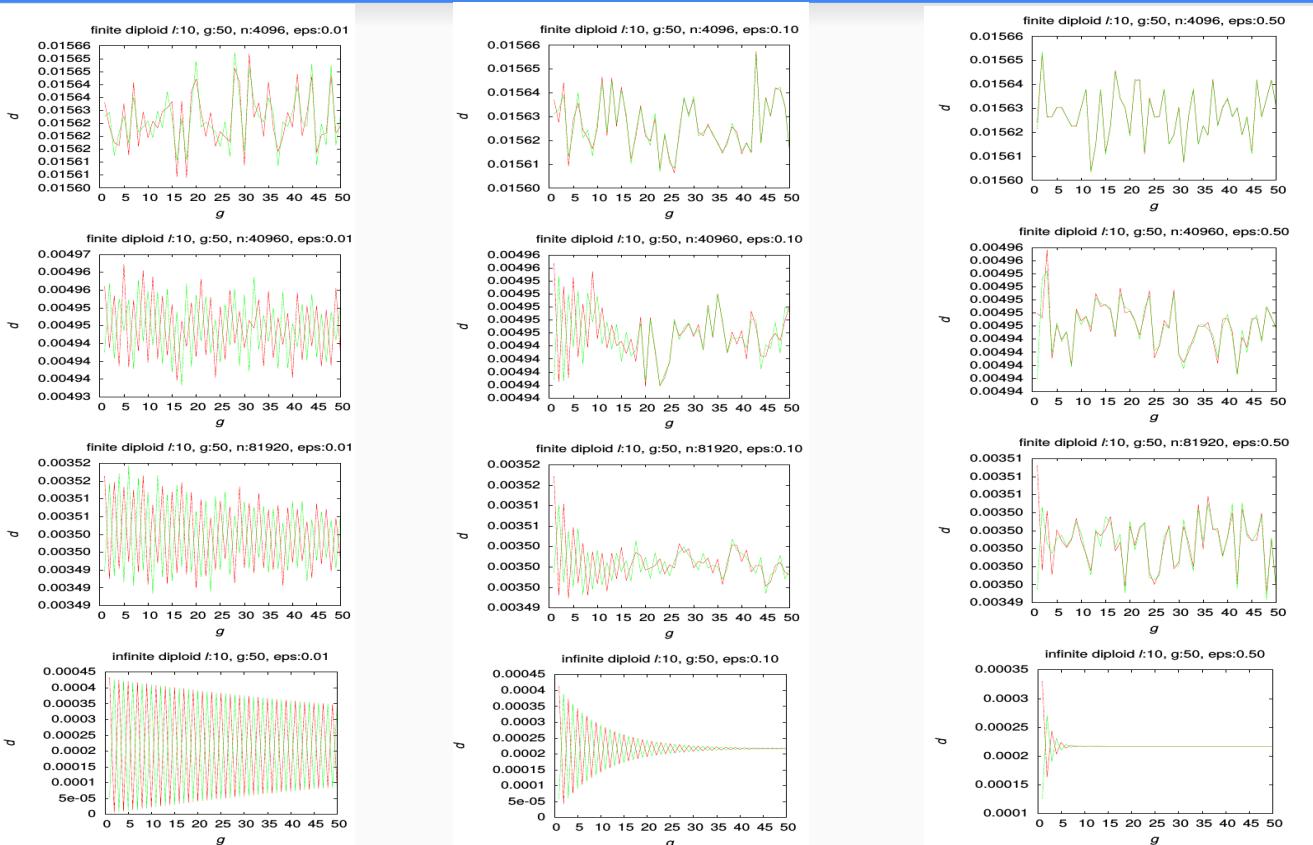


Figure: Oscillation under violation in crossover for diploids,  $l = 10$

# Results

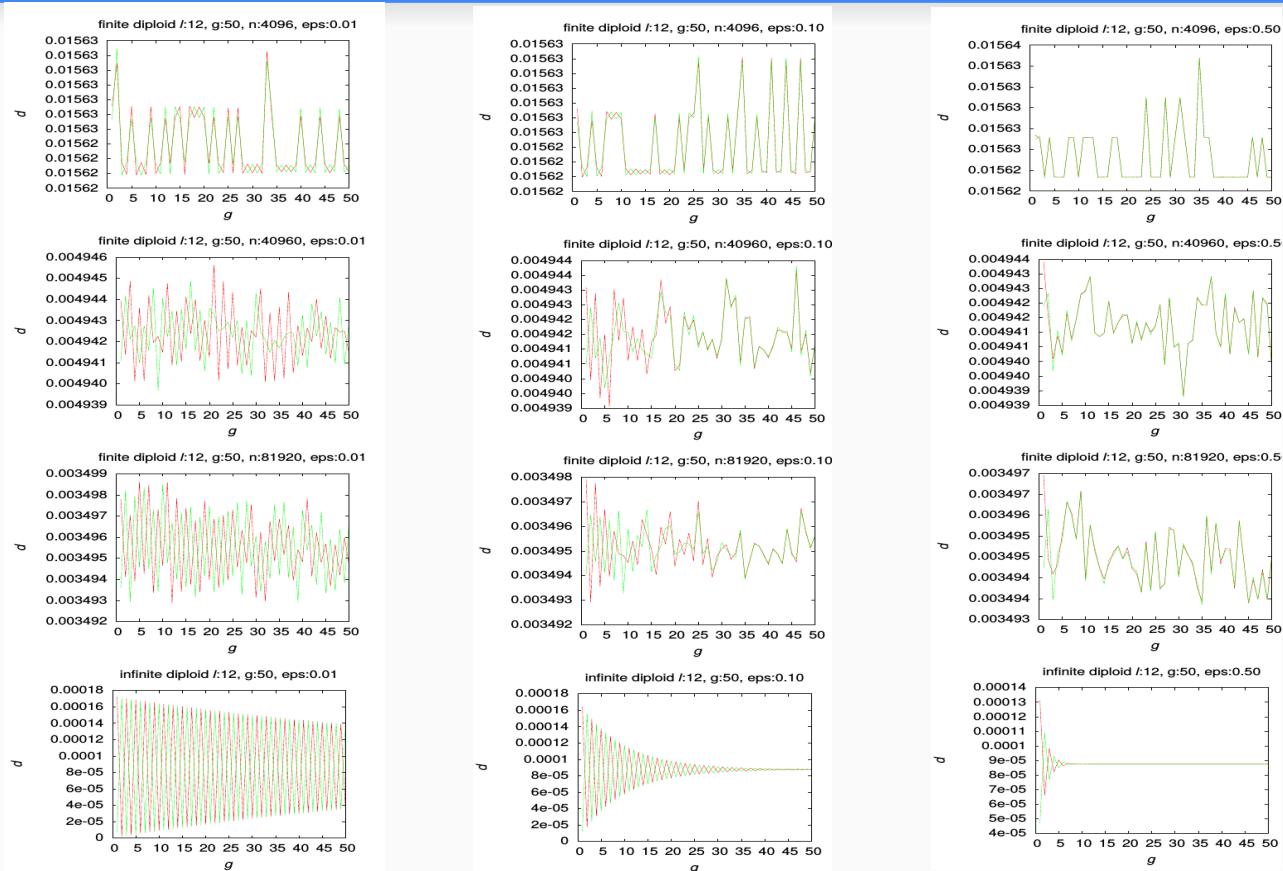


Figure: Oscillation under violation in crossover for diploids,  $l = 12$

## Violation in Crossover: Conclusion

Crossover also affects oscillating behavior of population evolution

Finite population evolution exhibits approximate oscillation if the violation level is small