An investigation into whether or not there exists a stable state at which the selection pressure created by the introduction of cholera into a population containing cystic fibrosis carriers is sufficient to allow the CF gene to persist in the population despite the potential for carriers to produce life-limited, infertile offspring.

Jordan Atkinson

School of Computing
Teesside University
M2016954@live.tees.ac.uk

Saul Johnson

School of Computing
Teesside University
M2082166@live.tees.ac.uk

Keywords: AIA: An investigation into whether or not there exists a stable state at which the selection pressure created by the introduction of cholera into a population containing cystic fibrosis carriers is sufficient to allow the CF gene to persist in the population despite the potential for carriers to produce life-limited, infertile offspring.- COM3036-N-BJ1.

Abstract

This paper will investigate the persistence of the cystic fibrosis (CF) carrier genotype in a population under various levels of selection pressure created by the introduction of vibrio cholerae infection causing secretory disease, which CF is thought to confer a resistance to. Experimental evidence from the source paper (Gabriel, 1994) suggests that heterozygotes for the CF gene (carriers) express 50% of the normal amount of the transmembrane chloride channel protein cystic fibrosis transmembrane conductance regulator (CFTR) in the intestinal epithelium and therefore secrete 50% of the normal amount of fluid and chloride ions in response to cholera toxin (CT). This correlation between the amount of CFTR protein and CT-induced chloride ion and fluid secretion suggests that carriers might possess a partial resistance to cholera. While the data from the paper (Gabriel, 1994) is arguably insufficient to draw accurate conclusions about this, a model will be produced to search for V. cholerae parameter levels (e.g. mortality, transmissibility) that allow carriers to persist in the population indefinitely despite their potential to produce life-limited, infertile offspring. The model produced suggests that this stable state does exist.

Problem Introduction

It is suggested in the source paper “Cystic Fibrosis Heterozygote Resistance to Cholera Toxin in the Cystic Fibrosis Mouse Model” (Gabriel 1994) that CF carriers possess a resistance against the disease-casusing effects of CT. Gabriel ran experiments using mice and observed that there was a 50 percent decrease in fluid secretion in carrier mice when compared to genetically normal mice. He concluded that this data suggested the existence of a genetic resistance to cholera based on the level of fluid accumulation during the 6-hour disease incubation period in carrier mice compared to their genetically normal counterparts. The results strongly supported the figure of a 50% reduction in CT-induced fluid accumulation in carrier mice.

The Model

The paper presented by Gabriel provides a range of experimental data which suggests that there is a strong case for a 50% reduction in CT-induced fluid accumulation in CF heterozygotes. Using this statistic, the experimental model will attempt to determine if there exists a stable state in which cholera will persist in the population indefinitely alongside CF carriers and genetically normal individuals. Other supporting papers are used for statistics which are not included or referenced by Gabriel. NetLogo will be used to model a real world population through the use of autonomous small

scale agents. Agents exist in an agent-domain (AD) and have a single purpose which is to live and reproduce. Each time an agent dies, two agents will be picked at random to reproduce. To "reproduce" in this model, an agent is spawned with a probability of being CF-heterozygous, CF-homozygous or genetically normal based on the genotypes of its parents. The AD contains 3 types of agents, Affected (A) who are CF-homozygous (that is, carrying two faulty CF genes), Carriers (C) which are CF heterozygous and Unaffected (U) who carry no faulty CF gene.

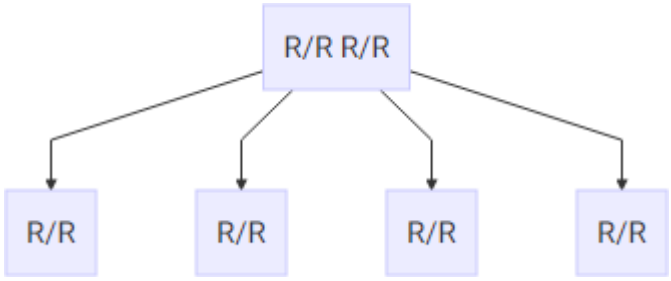
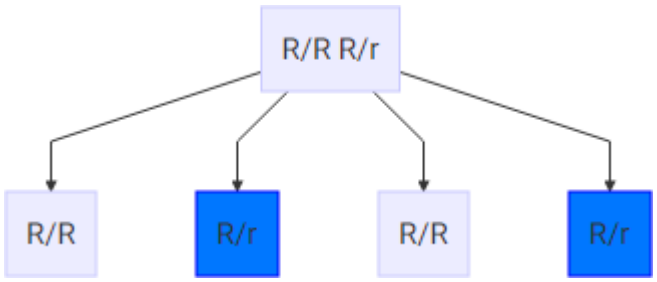
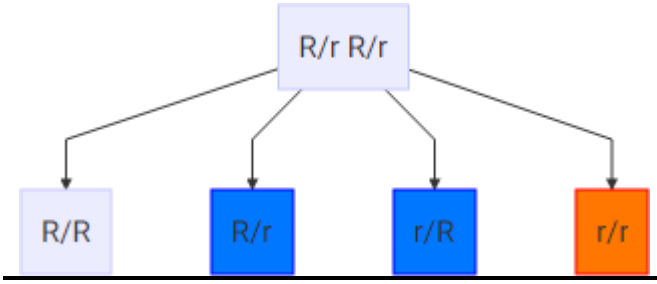
Diagram	*Description*
	If two unaffected individuals reproduce, the offspring have a 0% chance of being affected by or carrying the CF gene. There are no disease genes present.
	If an unaffected individual and a carrier reproduce, the offspring each have a 50% chance of being born as carriers themselves. They have no chance of being born as CF sufferers because one healthy gene is always inherited from the unaffected parent.
	If two carriers reproduce, the offspring each have a 50% chance of being born carriers, a 25% chance of being born unaffected and a 25% chance of being born CF sufferers.

Fig. 1 – Inheritance pattern for the CF gene, with the healthy gene represented as *R* and disease gene represented as *r*. To be affected by CF, one copy of the defective gene must be inherited from each parent.

Vibrio cholerae (VC) will then be introduced into the population with varying parameter levels (e.g. mortality, transmissibility) in order to observe the effect on the population.

If an agent is infected with VC, it will be infected for a period of time during which it cannot become reinfected. At the end of its infection period, it will either become immune or die. The proportions of CF genotypes in the starting population can be varied prior to running the model and the infection can be seeded into a chosen number of agents at once.

An infection occurs when an infected agent is on the same patch as an uninfected agent who is not immune. The uninfected agent will then have a chance of becoming infected, with the probability of transmission being configurable.

Assumptions Made

In order to create this model, some assumptions were made with regard to VC infection as well as the reproduction of agents. This was to ensure that the model was both concise and useful in exploring the hypothesis. Briefly, the following assumptions were made:

Assumption 1 - Affected agents cannot reproduce

This is due to the fact that the large majority of males who are affected with cystic fibrosis are infertile (≈3% fertility) (Hubert, 2006) and therefore would have negligible impact on the model. Affected agents are of limited relevance with respect to the hypothesis which focuses exclusively on CF carriers and not those who are affected by the disease. Notable, however, is the fact that affected people breeding together could potentially become relevant should the number of CF sufferers in the population climb high enough.

Assumption 2 - Infection Source

The VC pathogen is transmitted through poorly sanitised water/food sources (specifically, via the faecal-oral route) and not directly person-to-person. However, adding this into the model introduces unnecessary complexity. While the transmissibility of VC is of relevance, the transmission route is not; the source of the disease does not have any effect with respect to the hypothesis. At the most basic level, the cholera mentioned throughout this project could reasonably be substituted for any transmissible secretory disease.

Assumption 3 – Genetic basis of CF inheritance not fully modelled

The likelihood of an agent being born CF-affected, CF-carrying or unaffected can be accurately modelled as a set of probabilities depending on parental genotype. More advanced genetic models were not necessary.

Assumption 4 – One NetLogo tick represents one ‘day’ in the life of an agent

While largely irrelevant to the research question, one tick in NetLogo is assumed to be representative of one day of life for each agent in the AD.

Compromises Made

To minimise the complexity of the model while retaining its usefulness in exploring the hypothesis, there were some compromises made while designing the model:

Compromise 1 – Each agent can be represented as a small population or colony rather than a discrete individual

It is understood that VC thrives well within dense populations (e.g. family units, tribes) due to close contact between individuals and shared sanitation resources (Penrose, 2010). One agent might therefore be better thought of as a small, dense population of individuals becoming infected as one unit.

Compromise 2 – Population cap of 350 agents is used

A population threshold of 350 is enforced to ensure that there are always 350 agents within the AD. If this is not enforced, the infection may kill every agent before meaningful results can be gathered, or the population may grow out of control due to reproduction and negatively impact the performance of the model. Enforcing a constant population allows the factor of the effect of cholera on the prevalence of the CF gene in the population to be effectively isolated.

Compromise 3 – Disregard of maternal/paternal differences in inheritance

It was understood that when a carrier breeds there is a chance of the following gene pairs:

1. Unaffected/Unaffected (P_U 100%, P_C 0%, P_A 0%)
2. Unaffected/Carrier (P_U 50%, P_C 50%, P_A 0%)
3. Carrier/Unaffected (P_U 50%, P_C 50%, P_A 0%)

4. Carrier/Carrier (P_U 25%, P_C 50%, P_A 25%)

Where P_X is the probability that an agent of genotype X is produced as offspring.

In the model, pairs 2 and 3 are treated identically without regard to any differences between maternal and paternal inheritance of the CF gene. As there is no need to take into account which parent the gene came from, this has been extracted out.

The Results

Testing the Model

From fig. 2 it is apparent that without the selection pressure exerted by VC infection, the less genetically successful C agents seeded into the model initially gradually fall to zero in number. This is because C agents, while possessing a 50% chance to pass the CF gene on to their offspring, also have a chance to produce infertile A offspring when reproducing with each other. This is useful test of the model in which being a carrier for the CF gene demonstrably leads to a less successful line of offspring which ultimately results in complete gene extinction.

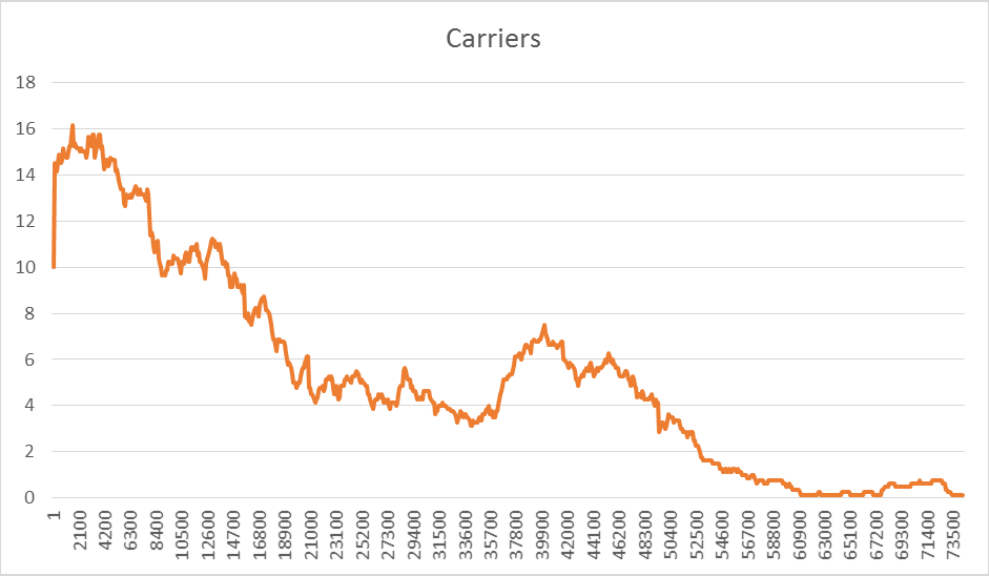


Fig. 2 – A graph showing the decrease (and eventual extinction) of C agents within the population with no VC infection. Only carriers are shown here as they are of primary relevance to the hypothesis (average of 8 model runs).

The brief spike in carrier numbers around step 39900 is of particular note. The chance increase in C agents in the population leads to a greater probability of interbreeding between them and production of infertile A agents, leading to an overall decline in CF gene prevalence.

Testing the Hypothesis

In order to test for the existence of a stable state at which VC infection creates a selection pressure in favour of the CF gene that is in balance with the reduced genetic success of C agents, VC had to be introduced into the model environment. The survival rates were controlled using sliders in the model interface:

- Percentage of carriers that survive – This slider controls the percentage chance of survival of all C agents after becoming infected with VC.
- Percentage of unaffected that survive – This slider controls the percentage chance of survival for all U Agents after becoming infected with VC.
- A Agents were assumed to be completely immune to the disease-causing effects of VC.

The percentage of carriers that survive will be kept at exactly double the percentage of unaffected that survive, which is consistent with the experimental findings in the source paper (Gabriel 1994).

Gabriel identifies this 2:1 survival ratio of CF carriers to genetically normal individuals through experiments that suggest that the 50% reduction in fluid secretion found in CF carriers equates with roughly double the survival rate. Despite this, it is unclear how linear the relationship between fluid secretion and survival rate would be *in vivo*. External factors such as age, general health, living conditions and other illnesses have not been considered which would affect amount of fluid loss individuals would be able to physiologically tolerate and still survive.

In order to investigate how a range of agent parameters affect the stability of the CF gene in the population, the following agent and environment attributes are configurable via the model user interface:

- Number of infections to seed – This value represents the number of random agents which will be infected with VC when the infection is seeded. This is useful in identifying any tipping point with regard to the minimum number of initial infections required for VC to become established amongst the population.
- Duration of illness – This allows the duration of VC infection to be adjusted.
- Immune days – This allows for simulation of acquired immunity in agents that survive VC infection for a number of days (ticks).

Initially the model was run in its base condition of 10 C agents and 240 U agents. This reflects the prevalence of CF gene carriers *in vivo* within the Caucasian population (Cystic Fibrosis UK, 2013) and the infectiousness rate was set at 5%. Mortality was set at 80% for unaffected agents and 40% for carriers. The initial number of infected individuals was seeded at 20.

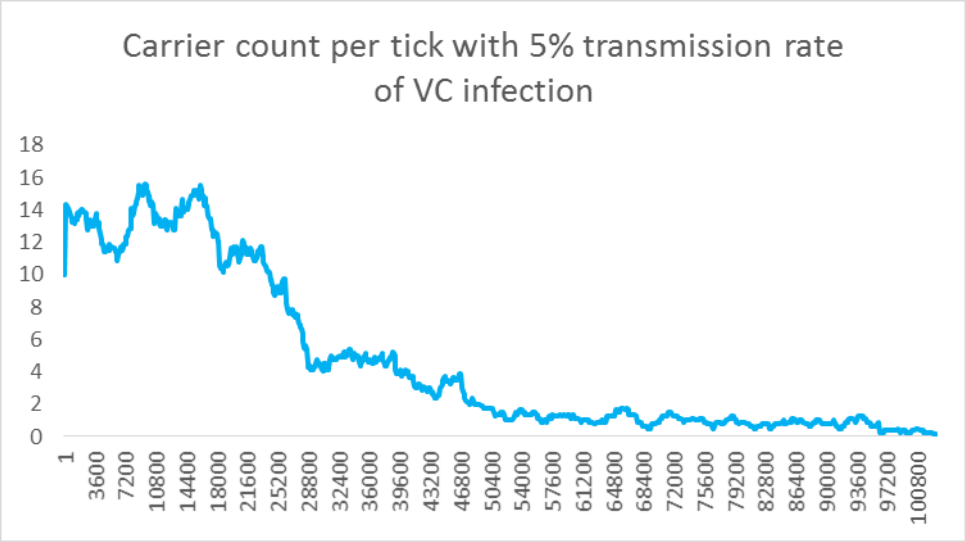


Fig. 3 – A 5% transmission rate was not sufficient to allow the infection to persist in the population, which led to extinction of the VC infection and subsequent extinction of the CF gene.

The survival rate was increased in 5% increments to see if a stable state could be maintained and it was found that at 40% VC transmission rate, the infectious disease and the CF gene appeared to stabilise and remain in the population indefinitely.

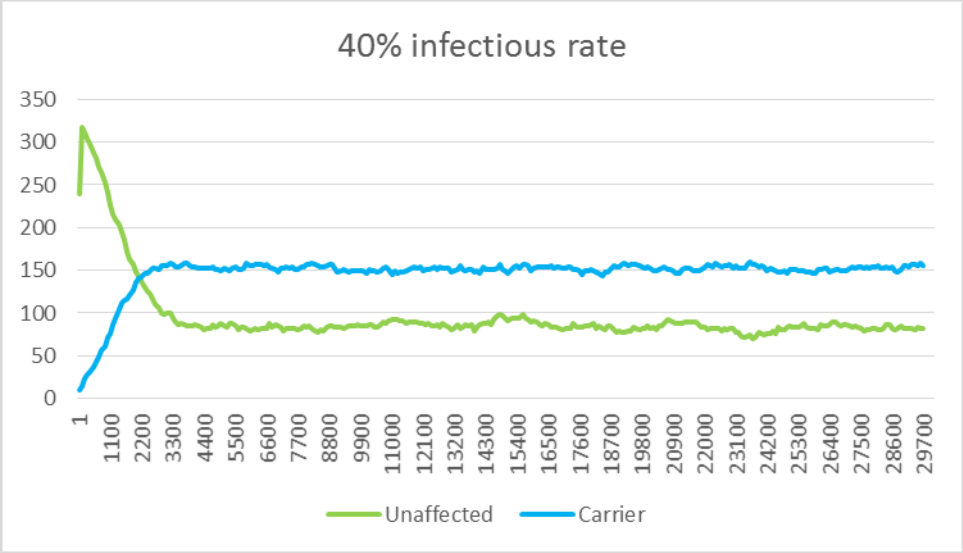


Fig. 4 – At the 40% transmissibility rate, C agents became the dominant genotype despite being less genetically successful, before the proportions of C agents and U agents levelled out.

Testing for hidden behaviours and Further Investigation

Seeding more than 20 agents when infection is introduced

The first further investigation examined the impact of introducing VC infection to more than just 20 agents initially. It was discovered that there was no long-term impact on the model, with the only change being the number of agents which die within the model immediately following the introduction of infection. This is expected due to the fact that more agents were seeded with the infection and therefore mortality will be more pronounced before the population recovers due to reproduction.

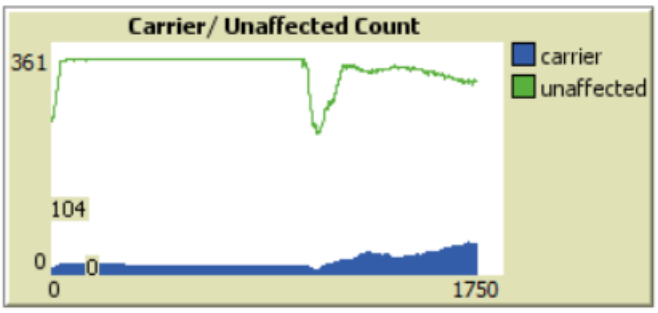


Fig. 5 – The impact on the population when 20 agents are initially infected with VC.

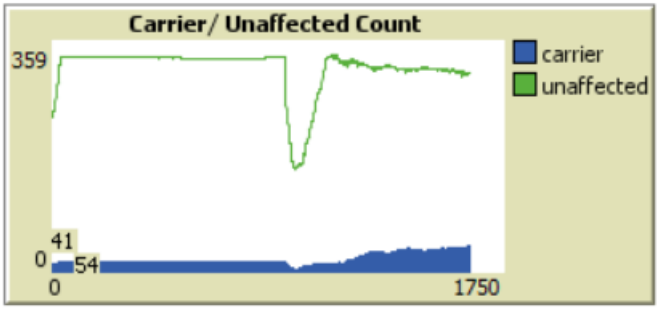


Fig. 6 – As for fig. 5, but with 100 agents initially infected.

Reducing the number of days in which an agent is immune

A further experiment was conducted to explore the effect of length of acquired VC immunity on time taken to reach a stable state of CF gene prevalence within the population.

A key assumption in exploring the hypothesis was that an individual surviving VC infection becomes immune to subsequent infection for 365 ticks (representing one year in time). Investigation revealed that reducing the length of immunity by anywhere between 1% and 50% produced no significant difference in the time taken to reach a stable state.

Reducing the length of immunity by 75% resulted in the stable state being reached 30% faster and again by 40% when the length of immunity was reduced by 80%. This tipping behaviour is theorised to be due to the infection providing a heavier selection pressure with reduced length of immunity.

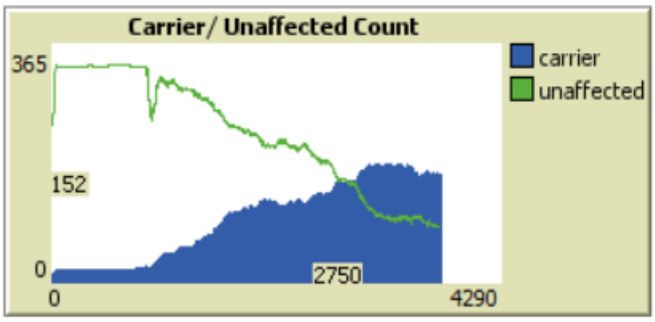
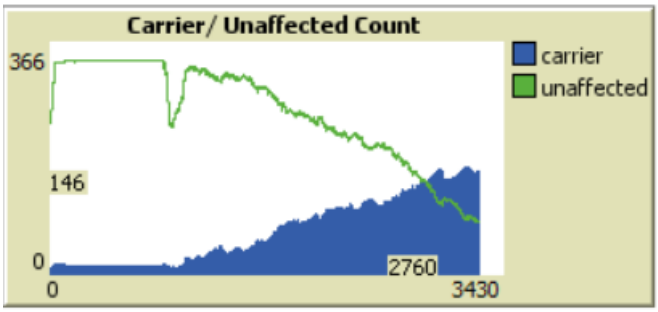
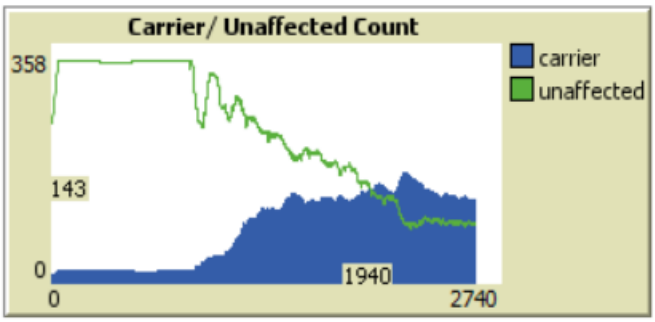
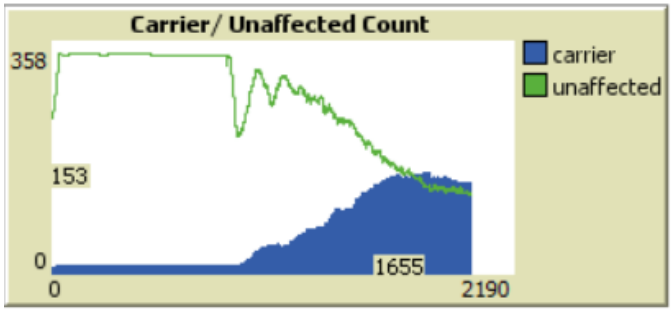
Ticks taken to reach beginning of stable state	*Immune-days number*
	365 (100%)
	183 (50%)
	92 (25%)
	73 (20%)

Fig. 7 – The comparison between the times taken to reach the stable state when the immune-days value is reduced by a certain percentage.

Percentage survival set at 100 percent

A test was performed to investigate the effect on the population when the survival rate for CF Heterozygotes is set at 100% (Unaffected therefore set at 50%). The model results showed an initial spike in the C agents and a slight decrease in U agents however eventually the infection becomes extinct within the population and the CF gene dies out as a result. This is thought to be due to the greater level of acquired immunity within the population inhibiting the spread of VC infection sufficiently to cause it to die out completely.

Further Investigation

It is recommended that studies from outside of the Computing field be carried out to investigate how linear the relationship is between survival rates and liquid secretion in VC infection. While Gabriel's findings suggest a 50% reduction in fluid loss equates to double the chance of survival, this linear relationship seems counter-intuitive. If this further investigation was performed, this model could then be reviewed and changes implemented to re-test the hypothesis to reinforce or dispute its validity.

Further investigation could also be done to review other recessive genetic disorders that might confer resistance against infectious disease in order to explore how the genetic disorder might have initially become established within the population.

Critique

While Gabriel's data is suggestive of a 50% reduction in fluid secretion in response to CT in CF heterozygotes creates a twofold increase in survival chance for CF carriers, it is unlikely that the relationship between liquid secretion and survival rate is that linear. This assumption is fundamental to the accuracy of results produced using the model, and warrants further investigation.

The fact that CF-affected agents within the model are incapable of reproduction is not truly reflective of the real world. CF males have around a 3% chance of being fertile and as the proportion of CF sufferers within the population increases, this becomes more and more relevant. A more accurate model would take this into account.

Females affected with CF have a slightly higher rate of fertility, in the region of around 20% of that of genetically normal females (Edenborough, 2001) and a reduction in life expectancy is likely to have a profound effect on the likelihood of either sex to opt in to parenthood. Once again, this is a shortcoming of the model which could be addressed once more data is available.

Aside from this, factors such as mortality rate and initial prevalence of CF heterozygotes within the population conformed as closely as possible to available data on real-world parameters.

Conclusion

Findings from experimentation with the model strongly suggest that a stable state exists at which the selection pressure created by VC infection in favour of the CF gene is sufficient to outweigh the greater reproductive success of genetically normal individuals. Once VC infection was introduced, there was a gradual and consistent increase in CF heterozygote agents within the population until the proportion of U agents to C agents levelled out and became stable.

The hypothesis presented has been proven to a degree which seems satisfactory. The linear relationship between survival rate and liquid secretion levels is still of concern and this is something which we suggest should be investigated further before using the model for serious research purposes.

Glossary of Terms

Acronym	*Term*	*Definition*
U	Unaffected agent	An agent which does not suffer from cystic fibrosis.
C	Carrier agent	An agents who is a carrier of cystic fibrosis – i.e. cystic fibrosis heterozygote.
A	Affected agent	An agent who is affected by cystic fibrosis.
CF	Cystic fibrosis	A disease which affects the quantity of liquid secreted through the urinary tract/secretion tract.
VC	Vibrio cholerae	A bacterium which causes the disease cholera. Cholera is a secretory disease which causes increased fluid loss due to diarrhoea.
FA	Fluid accumulation	The amount of fluid which is accumulated in the body whilst a person is infected with cholera.
AD	Agent domain	The world representation used in the model where all agents reside.
CFTR	Cystic fibrosis transmembreane conductance regulator	Protein in the intestinal epithelium responsible for the controlling liquid secretion and transport throught the body.

References

Gabriel, S.E. (1994). Cystic fibrosis heterozygote resistance to cholera toxin in the cystic fibrosis mouse model. Science. 266(5182), 107-109.

Ryan, E.T. (2000). Cholera Vaccines. Clinical Infectious Diseases, 31(2), 561-565.

Duguépéroux, I. (2006). Paternity in men with cystic fibrosis: a retrospective survey in France. Journal of Cystic Fibrosis, 5(4), 215-221.

Rutter, M. (2008). Child and Adolescent Psychiatry. Blackwell Publishers, 5(1).

Penrose, K. (2010). Informal urban settlements and cholera risk in Dar es Salaam, Tanzania. PLOS, 4(3).

What is Cystic Fibrosis? (2013). Retrieved from Cystic Fibrosis UK: https://www.cysticfibrosis.org.uk/media/132906/CB_Cystic%20fibrosis_what_exactly_Mar_13%20(PDF%20330KB).pdf

Edenborough, F.P. (2001) Women with cystic fibrosis and their potential for reproduction http://thorax.bmj.com/content/56/8/649.full