**Introduction**

Species migration can result in the following: (i) it allows individuals and species to colonise new areas and create new subpopulations. Over time, ecological events cause species to become reproductively isolated. This is known as allopatric speciation and it leads to the splitting of lineages, differentiating both related populations long-term \cite{barber1999patterns,coyne1992genetics}. This differentiation also increases with geographic distance. Restricting gene flow allows both sets of species are able to rapidly evolve in their local optimums /cite{garcia1997genetic}. Overall, species that are reproductively isolated have more pronounced modifications which can be observed phenotypically and genotypically \cite{pongratz2002genetic,sato2006effect}. (ii) Migration can allow isolated species to attempt to colonise each other’s habitats. If species are capable of interbreeding, this introduces new sets of alleles into an environment and interspecies reproduction passes on varying heritable genes which changes the developing genetic makeup of local species. The latter case is gene flow which helps to maintain the genetic diversity in an area but has been shown to homogenize populations over long periods of time, through the recombination of genes \cite{sato2006effect}. Advancements in genotypic techniques now enable us to study the genotypic effects and further our understanding of phenotype-genotype relationship. As organisms evolve, phenotypic evolution is assisted with genotypic evolution. Collective expression of genes through pathways influence the morphologies we observe in species \cite{hinman2009evolution}. Hereditary genome alterations through random changes in molecular mechanisms change varying aspects of the species \cite{chandrasekaran2008origins} These molecular changes induced by mutation and recombination lead to the variation of descending species \cite{chandrasekaran2008origins,ohno1999gene,brown2002genomes}. Over time, evolutionary forces involving genetic drift and selection acts on these polymorphisms and those most fit passes their variant genes and phenotype as a result, to future generations. This is the foundation of Darwin’s theory natural selection.\\

To better understand species evolution, we can see the development of their genome networks. Orr showed that there is variation with respect to genetic differences or gene influence on phenotype. The effects of adaptive and non-adaptive processes vary among species where there is no common set of genes involved, nor is the effects and interactions of the genes similar for species \cite{orr1998population} Although long temporal period has shaped a myriad of genetic function and interactions, what can be considered is the pattern at which these genetic processes develop over time. Genetic network simulations can be used to understand these patterns of evolution and the effect on phenotype-genotype relationships. Long temporal periods allow genetic interactions within a network to robustly develop, canalising the network \cite{orr1998population,lynch2007evolution}. \texit{Lynch et al} highlighted the significance of non-adaptive processes as well in shaping genetic networks. The study showed that networks can still evolve its architecture and become redundant even without the influence of natural selection \cite{lynch2007evolution}. Robustness can evolve from the effects of epistasis, additivity and dominance, all of which are connected \cite{omholt2000gene}.\\

Species evolution is non-linear, descending from a common ancestor, with modification and constant splitting of lineages. This continuous process over long temporal periods results in the accumulation of optimal genetic adaptations that results in a robust network structure that are adaptive and resistant to perturbations \cite{hinman2009evolution}. This can be quantified through fitness or reproductive success. There is a balancing act as selection aids to propagate fitter variants in a population, while mutation and environmental change limits such propagation \cite{burt1995evolution}. When migration is included, a balance between migration and selection will influence gene frequencies of future generations \cite{brown1992evolution}. What this study will investigate is how these forces affect the development of genes and the genetic network. Specifically, with the effects of gene flow on a genetic network that has evolved in isolation. The patterns of change that a genetic network undergoes with these adaptive and non-adaptive evolution processes. Even once a robust structure is reached, how does the structure resist change and maintain its network despite perturbations and evolutionary processes. As species evolve, studies have shown that pathways have a safety margin, that make them resistant to change such as mutations \cite{bourguet1999evolution}. Species best suited to their environment will evolve to their local optima, which we can represent as a quantitative trait value. The further apart these values are, what I label as \textit{environmental distance}, the greater the variance of the two species. The concern is how a genetic network responds when these evolutionary forces come into play and seeing the evolving interactions. How the regulatory interactions are modified during the networks evolution and how these changes affect trait values \cite{hinman2009evolution}. I will consider the effects of varying migration rates, two variants of a genetic network, the environmental distance and patterns of migration.

\\Ecological events can eliminate barriers and allow species to migrate into new environments, introducing new sets of genes in an environment. The presence of variant genes and network structures from gene flow hinders local adaptation and fixation of adaptive genes \cite{burt1995evolution}. Using quantitative trait loci (QTL) we are able to numerically interpret and visualise the patterns of change. Previous research has looked at the effects of gene flow, selection and mutation at generating local adaptation at the phenotypic level, showing how maintenance of alleles and linkage is important in adaptation \cite{yeaman2011genetic}. Even with random perturbations, there are bounds for which selection for canalization can act on, through the aid of genetic modifiers. They also revealed that under migration selection balance, selection for robustness increases with the migration rates \cite{proulx2005opportunity}.

\\This research will be looking at the changes in genetic architecture dynamics and effect of resulting interactions of the varying systems. As a genetic network evolves, there exists a threshold which is actively regulating these homeostatic genes \cite{gjuvsland2007threshold}. As selection for robustness occurs within the local population, it can give insight into the change in architecture and statistically significant interaction \cite{gjuvsland2007statistical}. Using a multi-locus system, I will construct a genetic network and simulate the effects over many generations and see how the output of the network changes, specifically looking at allelic interactions and tracking the fitness over time. Variance in fitness should decrease as a genetic network becomes robust, making it resistant to perturbations. Fitness can be quantified as reproductive success and is represented by passing on quantitative values generated from the alleles. These values are used to derive the trait values of individuals of which phenotypic values are then calculated and used as probabilities for fitness. The expectation is that after migration, a more robust network is formed when compared to before migration. At the start allowing new alleles to enter the population will hinder network development, but other evolutionary forces including selection will counteract the perturbations and result in a robust network more susceptible to perturbations \cite{garcia1997genetic}. Especially when the migrant network is a different structure, gene flow will allow maladaptive alleles to enter and should those be passed on, will impose a fitness cost to individuals \cite{tigano2016genomics}

**Methods**

To understand the effects from the evolutionary forces, I wrote a R script that constructs a genetic network, and a variant form, and simulates its evolutions, allowing migration to occur between two populations. All functions to perform adaptive and non-adaptive processes were written from scratch and implemented in the simulation. The following functions are:

\begin{itemize}

\item Population: initialises the starting populations of specified size where each individual (row) contains 12 allele sites (4 per gene). Since it is a di-allelic model, it is a 2-dimensional array.

\item Fitness: determines the fitness value of each individual based on their trait values and used as a probability for offspring contribution. A heavy tailed Cauchy distribution is used to determine fitness value from trait values. Each individual has a probability of passing on their genotype to the next generation and function randomly samples from the distribution to select parents, representative of genetic drift.

\item Mutation: produces an array same dimensions as the population and random uniformly distribution of values to determine which sites undergo mutation based on inputted mutation rate. Generates a new value using a normal function with current value as the mean and a standard deviation of 0.001.

\item Recombination: randomly chooses which site, and if any consecutive sites downstream, to switch allele values for each individual.

\item Migration: using a uniform distribution, randomly generates values for each individual in the migrant population to determine which individuals will migrate and replace those in the main population. Population is kept constant in both populations, representing a balanced dispersal between immigration and emigration \cite{rice2009evolution,w2004dispersal}.

\end{itemize}

\subsection{The model}

A di-allelic interlocus model from the research of \textit{Omholt et al}. In this case, all the genes are hereditary, representing only the regulatory and coding region which determine protein expression and rate of expression. Studies has shown that mutations along the coding region are known to cause morphological variation within species \cite{stern2009genetic}. Similar to the work of \textit{Omholt et al}, this model structure evolves dominance through epistatic interactions and regulatory effects. Using a system of equilibrium solutions and solved ordinary differential equations (ODE), simulated protein concentrations corresponding to phenotype are measured over time \cite{omholt2000gene}. Here I consider the sites as quantitative factors of protein function, and trait value is determined by protein concentrations. The greater the amount of protein expressed, the larger the trait value. The model consists of three genes, $X\_1$, $X\_2$ and $X\_3$. Let \(j\) represent the genes where $j = {1,2,3}$, each gene $X\_j$ consists of two alleles, $X\_{j1}$ and $X\_{j2}$. This leads to the following formula:

\begin{equation\*}

y\_j = X\_{j1} + X\_{j2} \label{eq:Protein Expression} \tag{1}

\end{equation\*}

\begin{wrapfigure}{l} {0.7\textwidth}

\begin{center}

\includegraphics[scale=0.35]{../Results/Model\_diagram.jpg}

\end{center}

\caption{Diagram showing the genetic model and two variants used to represent the migrant population. (a) Interlocus model of the population in focus. Lines labelled with mathematical symbols showing the interactions between genes. Gene $X\_1$ interacts with both gene $X\_2$ and $X\_3$, positively regulating both of them. To limit site values below infinity, gene $X\_2$ is reponsible for negatively autoregulating $X\_1$. There is an output for each gene where $j = {1,2}$ and $y\_j = x\_{j1} + x\_{j2}$. Gene $X\_3$ contains the trait values for each individual, which is the output. Circle represents phenotype which is determined from trait values using a Cauchy Distribution. (b) and (c) represent models for the migrant population. (b) is the same pathway and regulation as (a) however (c) is switched where gene $X\_1$ negatively autoregulates $X\_2$, and gene $X\_2$ postively regulates $X\_1$ and $X\_3$. Again, output of gene $X\_3$ are the trait values used to derive fitness.}

\label{fig:Starting parameters}

\end{wrapfigure}

Where $y\_j$ is the total protein concentration at each gene. There are four sites which represent the different factors affecting protein production. These are \textit{\textalpha}, \textit{\textgamma}, \textit{\texttheta} and $P$. \textit{\textalpha} is the protein production rate while \textit{\textgamma} is the degradation rate \cite{omholt2000gene}. For both sets of populations, a single gene, $X\_3$ determines the trait value for individuals and quantifiably differentiates the populations in terms of morphology \cite{orr2001genetics}. For the population in focus, gene $X\_1$ positively regulates gene $X\_2$ and gene $X\_3$, and gene $X\_1$ is negatively regulated by gene $X\_2$. This is to regulate trait value and prevent the value from exceeding to infinity. As gene $X\_2$ increases in expression, it decreases $X\_1$ expression, negatively autoregulating the system and limiting its value. Let $j = {1,2,3}$ and $i = {1,2}$, from the separate researches of \textit{Omholt et al}, and \textit{Gjuvsland et al}, $R\_{j}$ is a regulatory Hill Function representing a Michaelis-Menten mechanism, where $S(y\_j,\theta,P) = \frac{y\_j^P}{y\_j^P+\theta^P}$. The Hill Function explains the relationship between regulator and producer, where \texttheta is the amount of regulator needed for 50\% production rate and P affects the steepness of the curve \cite{gjuvsland2007statistical,omholt2000gene}. Should the network be negatively regulated, it leads to the following equation:

\begin{equation\*}

R\_{j}(y) = 1 - S(y, \theta\_j , P\_j), j = {1, 2} \label{eq:Negative autoregulation function} \tag{2}

\end{equation\*}

And if positively regulated:

\begin{equation\*}

R\_{j}(y) = S(y, \theta\_j, P\_j), j = {1, 2} \label{eq:Positive autoregulation function} \tag{3}

\end{equation\*}

Again, letting $j = {1,2,3}$, as gene $X\_1$ positively autoregulates gene $X\_2$ and gene $X\_3$, and gene $X\_2$ negatively autoregulates gene $X\_1$, this results in the following equations:

\begin{equation\*}

R\_{1j}(y\_2) = 1 – S(y\_2, \theta\_{2j}, P\_{2j}) \label{eq:X1 negative autoregulation function} \tag{4.1},

\end{equation\*}

\begin{equation\*}

R\_{2j}(y\_1) = 1 – S(y\_1, \theta\_{1j}, P\_{1j}) \label{eq:X2 positive autoregulation function} \tag{4.2},

\end{equation\*}

\begin{equation\*}

R\_{2j}(y\_1) = 1 – S(y\_1, \theta\_{3j}, P\_{3j}) \label{eq:X3 positive autoregulation function} \tag{4.3}

\end{equation\*}

\textit{\textmu} is the ratio of \textalpha and \textgamma per locus. Using the equilibrium solutions, total protein concentration is calculated by the following equations:

\begin{equation\*}

y\_1 = \mu\_{11}(1 – S(y\_2, \theta\_{21}, P\_{21})) + \mu\_{12}(1 – S(y\_{2}, \theta\_{22}, P\_{22})) \label{eq:y1 function} \tag{5.1}

\end{equation\*}

\begin{equation\*}

y\_2 = \mu\_{21}(S(y\_{1}, \theta\_{11}, P\_{11})) + \mu\_{22}(S(y\_{1}, \theta\_{12}, P\_{12})) \label{eq:y2 function} \tag{5.2}

\end{equation\*}

\begin{equation\*}

y\_3 = \mu\_{31}(S(y\_{1}, \theta\_{31}, P\_{31})) + \mu\_{32}(S(y\_{1}, \theta\_{32}, P\_{32})) \label{eq:y3 function} \tag{5.3}

\end{equation\*}

\subsection{Migrant network}

For the first motif, the genetic network will be the same as the main population, just evolving to a different local optimum value. For the second motif however, the difference is that gene $X\_1$ negatively regulates gene $X\_2$, while gene $X\_3$ and gene $X\_1$ are positively regulated by gene $X\_2$. The formulas used to derive $y\_1$, $y\_2$ and $y\_3$ values for the migrant population are as follows:

\begin{equation\*}

y\_1 = \mu\_{11}(S(y\_2, \theta\_{21}, P\_{21})) + \mu\_{12}(S(y\_2, \theta\_{22}, P\_{22})) \label{eq:y1 function} \tag{6.1}

\end{equation\*}

\begin{equation\*}

y\_2 = \mu\_{21}(1 – S(y\_1, \theta\_{11}, P\_{11})) + \mu\_{22}(1 – S(y\_1, \theta\_{12}, P\_{22})) \label{eq:y2 function} \tag{6.2}

\end{equation\*}

\begin{equation\*}

y\_3 = \mu\_{31}(S(y\_2, \theta\_{31}, P\_{31})) + \mu\_{32}(S(y\_2, \theta\_{32}, P\_{32})) \label{eq:y3 function} \tag{6.3}

\end{equation\*}

This is to represent the concept of speciation but can still integrate in the other population and interbreed.

\subsection{The simulation}

A total of 44 permutations based on conditions in \textit{Appendix I} of environmental distance, genetic network structure, migration rates and migration patterns were simulated for 1,200 generations each run. For the effect of genetic drift and to account for the large deviations of values, a Cauchy distribution is used to generate fitness probabilities per generation. Since the Cauchy distribution is characterized for its heavy tails, it is able to account for values that greatly deviate from desired trait value. The values entered in the Cauchy distribution are the desired trait values. It is important to note that environment is kept constant both spatially and temporally. Both populations were kept constant at 500 individuals. The main population evolved to a trait value of 50 with a standard deviation 8, while the migrant population alternated between 65 and 80 with standard deviation 10. The large standard deviations characterise the varying forms of morphology that can be noticed in species. The trait values represent the environments of both populations and the local optimums they evolve to. The probabilities extracted from the Cauchy distribution are for parental contribution to offspring for the next generation. As the network evolves to a stable state, the feedback loops should maintain the homogeneity of the system.

\\For the simulation we assume that both populations have the same size and stay constant, with migrants replacing individuals. There is no spatial structure and all individuals have an equal chance of being replaced. Both populations undergo divergent selection, stabilising in their own environments to different specified trait values, thus differentiating the populations over time \cite{sato2006effect}. Alleles for each individual can either be homogenous, using a uniform distribution to determine starting value between 0.1 and 0.3 for both populations, or heterogenous, using a uniform to randomly generate the starting allele values, again between 0.1 and 0.3. To get a value from equation 2, there must be a value for $y$, so uniformly distributed starting values were generated for the three genes. If the population is homogenous, each individual in the population started with the same value at each gene, otherwise have differing values if heterogenous.

Recombination is equal chance at any locus and interchanges the alleles and everything downstream. Mutation can occur at each locus by randomly deviating from the current value. The probability is the same constant for both populations where each locus has an equal chance of mutating. Mutation probability is kept constant at 0.0011 per site. A cis-mutation in the second gene will affect the corresponding value in the first gene as in equation 3, gene $y\_2$ negatively autoregulates gene $y\_1$, while a trans-mutation in gene $y\_1$ will affect the expression in gene $y\_3$. Since all genes in the model represent regulatory and coding regions, mutations in any site can be considered to affect phenotype, for its pleiotropic effects \cite{rice2019evolution; landry2007genetic}.\\

As mentioned before fitness is determined by phenotypic value as the offspring contribution per generation. Each individual per generation has no limit as to how many times they can be a parent, however the standard deviation of 8 and 10 in the Cauchy distribution attempts to produce varying combination of parents. Migration rates varied between 1\%, 3\% and 5\%. As migrant individuals enter the population, they randomly replace individuals in the population. With constant population size, this represents immigration and emigration. Furthermore, low migration rates were used to prevent migration population from completely replacing the original population and allowing the network to be able to adapt to the new values. Both populations have a burn-in period of 80 generations to evolve in their own environments before migration can happen. Also, migration can only occur until the 700th generation. The remaining 500 generations were to assess how the network responds to the migration. Pattern of migration was also considered, varying between each generation, every 10 generations, every 5 generations and random (between 1\% and 5\% each occurrence) after the 80th generation.

\subsection{Analysis}

Analysis was done on the recorded fitness, trait values and population arrays. At the end of each simulation, fitness is normalised by dividing fitness probabilities with the medians of Cauchy distributions, with 1.0 being the highest possible fitness value. Firstly, control conditions of no migration were simulated to see how rapidly isolated networks evolve. Without migration, I expect rapid evolution of allele values \cite{garcia1997genetic}. However, which of the two starting genetic network makeup: homogenous or heterogenous evolve faster to their optimums. To analyse robustness, a comparison was made by calculating a Robustness ratio. The fittest individuals before and 10 generations after migration were recorded and replicated such that there were 4 separate mutations per site. Trait values were again calculated and inputted into the Cauchy distribution to determine fitness values. The robustness ratio is then the variance in fitness after migration divided by variance in fitness before migration. A robust network means that the ratio should be less than 1. Ratios were then log transformed as to linearise and make it less skewed. In addition to this, the length of time foreign alleles persists in the environment as how migration patterns affect this \cite{w2004dispersal}. Finally comparing robustness and variance in before and after migration and getting the ratio of the variance. The expectation is that the ratio should be a value less than 1 as the network has had more time for evolutionary processes to act on and evolve. Modelling using linear regression was used to see the trends for changing migration rates, and box plots to compare the migration patterns.

**Results**

\subsection{Cauchy Distribution}

\begin{wrapfigure}{r} {0.7\textwidth}

\begin{center}

\includegraphics[scale=0.2]{../Results/Cauchy\_Distribution.pdf}

\end{center}

\caption{Plot of the Cauchy Distribution used to derive fitness probabilities (for reproduction) and fitness values (normalising) of individuals. Distribution is also representative of the environmental distances respective populations evolved to. The main population (black) evolved to a trait value of 50, where the peak trait value has a reproductive probability of 3.98\%. Migrant populations either evolved to a trait value of 65 (blue) or 80 (red).}

\label{fig:Cauchy Distribution}

\end{wrapfigure}

For the distribution of trait values and fitness, a Cauchy distribution was selected. The distribution is similar to a normal distribution as seen in \textit{figure 2}. The peak represents the median, in this case the maximum probability a trait has of passing on to the next generation. The distribution shows how the trait values are distributed where even small deviations away from the median can have reduce the probability signficantly. These desired trait values represent the environment and the distance between them is the environmental distance. The values must not be too far apart that the probability is 0, meaning that migrant genes are passed on to future generations and we see how the network develops with invading alleles. Since the Cauchy distribution is long tailed, this allowed for varying alleles to potentially persist in the population, as seen from the probability values along the y-axis.

\subsection{No Migration}

\begin{figure}[h]

\centering

\includegraphics[width=0.9\textwidth]{../Results/early\_no\_migration.jpg}

\caption{graphs showing the early fitness evolution without migration for two different starting genetic makeups. (a) shows the evolution of a network that is homogenous network, where all starting allele values are the same per site per individual, randomly generated using a uniform distribution between 0.10 and 0.30. (b) shows how a heterogenous network evolves where starting values (between 0.10 and 0.30) for each site are different. Heterogenous populations had a mean fitness of 0.165, standard deviation of 0.035 while the homogenous populations had an average fitness value of 0.955 with a standard deviation of 0.324. Variation helps the heterogenous population evolve while homogenous rely on beneficial mutation and recombination’s.}

\label{fig:No Migration}

\end{figure}

The prediction was that a population reproductively isolating would evolve rapidly \cite{garcia1997genetic}. In addition to this, a population that begun as heterogenous would evolve faster as there was greater variation in the allele values, allowing for better combinations to be selected for. The results show that when there is no migration happening, the genetic network does evolve rapidly to the desired trait values. Especially for a starting population of heterozygous individuals, in a very short time period, it reaches an average fitness of 0.955. This can be seen from \textit{figure 3} to the right. Normalising the fitness, it shows how rapidly the trait values rises to 50, plateauing around 1 within 100 generations. Although there are cases of the fitness deviating, overall the rise to the average fitness is steep.\\

\textit{Figure 3} also plots the average fitness values of homogenous and heterogenous population, they both show rapid evolution and have very similar patterns of sharp gradients in the early generations. For the heterozygous population, starting gradient is very sharp and population does not vary until it begins to plateau. This likely due to mutations and recombination. However, unlike a population of heterozygous, the homozygous population has a lot more noise, even at the start. There were instances where the population did not even increase in fitness and stayed near 0. There are a few instances where fitness sharply rose to around 1 and stayed there. Furthermore, this rapid evolution is seen in \textit{figure 6} which is a boxplot showing the distribution of the number of generations needed to reach the average fitness for with and without migration.

\subsection{Migration}

\begin{figure}[h]

\centering

\includegraphics[width=0.9\textwidth]{../Results/migration.jpg}

\caption{Plots showing the effect of migration on fitness over time. Migration occurs between generations 80 and 700, where the network gets 500 generations afterwards to try and recover. The graphs show how migration caused fluctuations in fitness values (a) shows the effect of migrant network with different regulations from the main population, evolving to a trait value of 65 (Environmental Distance of 15). It is a variation where gene $y\_1$ negatively regulates gene $y\_1$, and gene $y\_2$ positively regulates the other genes, including gene $y\_3$ which is the trait values. Mean fitness was 0.917, with a standard deviation of 0.069. (b) is a migrant network same as in graph (a) where the regulations are different, instead evolving to 80 (Environmental Distance of 30). Mean fitness value was 0.070, standard deviation of 0.070. (c) shows the effect of a migrant population but with the same regulations as in the main population. Therefore, gene $y\_2$ is negatively regulated by gene $y\_1$, and gene $y\_1$ positively regulates the other genes. Similar to graph (a) it is evolving to a trait value of 65. Mean fitness of 0.917 and standard deviation of 0.071. (d) same regulation as the migrant network in graph (c) however evolves to a trait value of 80. Mean fitness of 0.916 and standard deviation of 0.071.}

\label{fig:With Migration}3.

\end{figure}

With migration, the genetic network evolved the same as without migration. It took an average of 57.4 generations (standard deviation of 15.55) to reach an average fitness value of 0.918 (standard deviation of 0.05). This is to be expected however as the genetic network did not encounter gene flow until the 80th generation. The mean speeds were the same but differed was a slightly in terms of variance. This is shown in the boxplot in \textit{figure 6} to the right. The expectation was that migration would hinder the structures ability to evolve and it would take more generations to reach the average fitness. The effect of migration was seen in the variance of fitness from the plots in \textit{figure 5}. When there was migration, the fitnesses was fluctuating and going below 0.50. This was to be expected as maladaptive foreign alleles cause the trait value and fitness to deviate from the local optimum. The variation between the different genetic structures and environmental distances was very low as conditions generated similar fitnesses and variance during periods of migration. In the periods of migration (80th generation – 700th generation), the average fitness 0.916 and a standard deviation 0.071.

\begin{wrapfigure}{r} {0.7\textwidth}

\begin{center}

\includegraphics[scale=0.4]{../Results/boxplot\_migration.jpg}

\end{center}

\caption{Boxplot showing the number of generations taken to reach the average fitness of both conditions, with and without migration. When there is no migration, it takes an average of 57.4 generations (standard deviation of 15.58) to reach an average fitness value of 0.919 (standard deviation of 0.05). When migration is present, it takes an average of 57.4 generations (standard deviation of 15.55) to reach an average fitness value of 0.918 (standard deviation of 0.05).}

\label{fig:Speed to average fit}

\end{wrapfigure}

Although there were no significant effects of environmental distance, when viewing \textit{figure 5}, alleles did persist longer. When the migrant network is the same as the main population, the alleles persist longer compared to the migrant network having a different regulation structure. Specifically, when the migrant population is evolving to a trait value of 15 (environmental distance of 65), it took an average of 3.34 generations (standard deviation 6.58) for the fitness value to equal or rise above the average fitness. This compares to an environmental distance of 30 (evolving to trait value of 80) with a 1.38 generation recovery average, fluctuating between 1.01 generations. For the different genetic structure migrant population, the average recovery time was 0.95 and 1.24 for environmental distance of 15 and 30 respectively, with a standard deviation of 0.67 and 1.15 generations.

\subsection{Robustness}

\begin{table}[h!]

\centering

\includegraphics[scale=0.40]{../Results/robustness\_anova.jpg} \caption{Table showing the analysis of variance results of the regression of the log of the variance ratio and migration rate, migration pattern and genetic makeup of both populations} \label{tab: Regression}

\end{table}

A total of 4 factors that could affect this ratio was measured. These are: main population genetic makeup, migrant population genetic makeup, migration rate and migration pattern which were all considered. Ratio of robustness results show that migration rates and migration patterns have a statistically significant effect. Linear regression and analysis of variance of the result was used to see which of the factors was statistically significant. The data was log transformed to make it less skewed and linearise it. Results show that there were no statistical significance however migration rate showed a slight effect on the robustness ratio as seen in \textit{table 1}. Linear regression computed an F-statistic of 2.27 with a corresponding p-values of 0.08, \textalpha level of 0.05. Early results suggest that migration rate could have a potential significance on robustness, but further runs must be done.

**Discussion**

\subsection{No Migration}

Without gene flow, populations did evolve rapidly to their environment due to a balance between selection and drift \cite{garcia1997genetic,tigano2016genomics,barber1999patterns}. The prediction was that a heterogenous genetic network would evolve a lot quicker compared to the homogenous structure due to variation of allele values. This was the case as variation of allele values at the start allowed for better combinations to propagate quickly in the population. Rapid rise can be first attributed to the fitness distribution and reproductive success of the system. The most fit individuals per generation were likely to have been selected for reproduction many times allowing their combination of alleles to be passed on at a high frequency. Deviation around a high average fitness of 0.955 likely due to mutation and recombination.\\

In the homogenous case, the instances where fitness rose close to 1 is likely due to beneficial mutations and recombination to create better combination of allele values, and epistasis which all maintain adaptiveness, which has even been shown to happen with gene flow \cite{tigano2016genomics}. When a mutation or recombination occurs that brings a fitness within the peak of the Cauchy distribution, it can propagate rapidly in a population. Additionally, \textit{figure 2} shows the magnitude in difference of deviating away from the peak of the distribution. For the desired trait value of 50 with a standard deviation of 8, the maximum probability an individual has of being chosen is 3.98\%. Although this is a small value, an individual that is 10 trait values away will have a probability of 1.55\%, a three-fold difference between them. With the assumption that there is no limit to parental contribution, this allows high number of combinations of fitter individuals to be parents and pass on their values; hence also the low standard deviation for the heterogenous situation. Therefore, the variation in the network allows it to rapidly evolve as those that were close to the peak likely passed on their values to many individuals in the next generation, regardless of genetic makeup. As a result, this decreased the variation in alleles over a short temporal period.

\subsection{Migration and Robustness}

Although there were no statistically significant results shown in recovery times or the log ratio of robustness, a lot can be learned in the pattern seen. More specifically how the genetic network was able to overcome disruptions caused by gene flow and prevent homogenizing the populations. While gene flow caused variation in the system (as seen in fitness), the other evolutionary effects especially selection counteracts gene flow and maintain the adaptiveness of the network \cite{feder2012genomics,burt1995evolution}. When migrant alleles enter the population, recombination should allow them to persist \cite{feder2012genomics}. Despite the presence of migrant alleles, divergent selection heavily favours local alleles in particular acted to remove them \cite{tigano2016genomics}. The fitness distribution in \textit{figure 2} shows how there is a balance between gene flow which would hinder adaptation of the network, and selection which would allow fitter values to propagate \cite{brown1992evolution}. Not just migrant allele values, but also complementary combination of migrant alleles that could allow it to persist \cite{feder2012genomics }. Compared to a normal distribution, a Cauchy distribution is much narrower, thus migrants had very low chance of persisting. Selection favouring fit individuals with better combination of alleles. \textit{Feder et al} study showed how there exists a friction between divergent selection, and gene flow and recombination. Results here show how when selection effects are greater than gene flow and recombination, it acts to swiftly remove migrant alleles in the population.\\

As mentioned before, the fitness distribution and assumption of parental contribution and reproductive success prevented foreign alleles from persisting in the network during migration. This can be seen in the recovery times as the network was quick to recover, taking an overall average of 1.73 generations (standard deviation of 3.52 generations). This infers that invading alleles were quickly eliminated in the system. However, during the periods of migration (between 80th and 700th generation), gene flow has an observable effect on fitness, which is consistent with previous studies. Specifically, gene flow leads to large deviations in the fitness. Poorly adaptive traits to the environment would delay the system ability to evolve to local optimum as the values deviate far from it \cite{garcia1997genetic}. This was to be expected as periods where the foreign alleles far from the peak of the distribution. trait value and fitness to deviate from the local optimum. Referring back to \textit{figure 4}, the mean standard deviation is 0.069, which is larger compared to a mean standard deviation of 0.04 when there is no gene flow occurring.

\subsection{Shortcomings and Future Work}

Due to time constraints and difficult circumstances, major assumptions were made in the model and simulation. Few trials of the different conditions and design of the simulation prevented significant results from appearing. Firstly, the model did not take into account spatial or temporal aspects. In reality, species are distributed along a habitat and the environment itself can change due to perturbations i.e. vicariant events \cite{garcia1997genetic}. Any individual in the population will be replaced regardless of fitness. In reality, this is not the case. With more time, I would have designed a more realistic model, taking into account dispersal habits, species distribution, environment perturbations, and migration limits. Species dispersal along with varying environmental conditions would have created a more complex system and delayed the quick homogenization of the network to allow for investigation of gene flow effects on fitness \cite{garcia1997genetic,barber1999patterns,sato2006effect}. This also would better maintain the variation in the population to better analyse the impacts. Since population size was kept constant, migration rates were limited so that the main population would not be completely replaced. To fix this, I would change it so that population size can fluctuate, as migrants would not replace individuals, but also some form of competitive interaction would occur. This would prevent the population being completely replaced in the current model for higher migration rates. Moreover, I could investigate the patterns at which migration, selection and drift affect population growth, analysing effect of dominance on robustness \cite{rice2009evolution, otto1999balanced}. With regards to environmental spatial dynamics, I would have made it so that different areas would have different local optimums. The model assumed there was no ecological barrier restricting migration and individuals were always replaced by migrants. Random environmental perturbations that would alter local optimums in addition to gene flow would further delay network evolution. Thus the effects of selection and recombination would not so quickly recover the system and can better investigate how network responds to disturbances. This would create more varying network structures and allow further investigation into network development with varying motifs.\\

The current simulation assumed that there is no limit to parental contribution. Since there was no limit, the fitter individuals quickly pass on their trait values in the simulation. This rapidly canalised the system and decreased the variation in the environment. To improve this, I would limit parental contribution i.e. individuals reproducing can only be parents to five next generation individuals. This would prevent the system from stabilising so rapidly and allowing for more diverse alleles to continue to propagate in the population. Testing with other genetic structures and motifs could give insight to more complex interactions between evolutionary forces. The changes mentioned would allow for alleles to persist for longer periods of time and could allow for investigation into the lineage aspect of the genetic structure. For example, seeing how long migrant alleles persist in the population and if they help in development of a robust structure. Lastly, it would also be interesting to investigate the opposite case as to how migration patterns are influenced by these evolutionary forces \cite{w2004dispersal}.

**Conclusion**

In just a few centuries, anthropogenic changes by societal development will have ecological consequences. Many of which will likely have long-term effects to many ecological interactions. These ecological changes could then have an impact on the environment and species. There could be vicariant events that isolate species or the opposite where barriers are removed. The presence of these novel suboptimal genes, along with the interaction of species in the environment will lead to evolutionary change. After many generations, this change will be seen in a phenotypic level as well as genotypic. This study was to see the scenario where gene flow allows related species, that could have originally been reproductively isolated, to interbreed, and the impacts on species morphology and development. Although results did not reveal any statistically significant effect, the interactions of environmental forces observed can give further insight into the interplay between micro genetic mechanisms and macro evolutionary processes. Improvement to the current simulation design and repeated trials can contribute to such studies.