**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. This differentiation also increases with geographic distance. Without gene flow, 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, parapatric or peripatric speciation may occur, depending on distance. This introduces new sets of alleles into an environment and as species interact and pass on its heritable genes, it changes the developing genetic makeup of local species. The latter case includes 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 certain genes through pathways assist in the morphology and behaviour we observe in species. 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 drift and selection acts on these polymorphisms and those most fit passes their variant genome, 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 focus on the development of their genome network. 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 generalizations cannot be made 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}. Lynch highlighted the significance of non-adaptive processes as well in shaping genetic networks. His 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 with modification and constant splitting from lineages of a common ancestor. This continuous process over long temporal periods results in the accumulation of optimal genetic adaptations that results in a robust network structure. 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}. What this study focuses on 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 value. The further apart these values are, what I label as environmental distance, the greater the variance of the two species. The concern is on how a network responds when these evolutionary forces come into play and seeing the evolving genetic interactions. Investigating the effects of changing migration rates, two variant genetic networks, environmental distance and patterns of migration.

Ecological events 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 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}. It was shown that with random perturbations and aid of genetic modifiers, there are bounds for which selection for canalization can act on, leading to evolution of robustness. They also showed that under migration selection balance, selection for robustness increases with the migration rates \cite{proulx2005opportunity}. Not just migration rate but pattern must be taken into account as the more migrants enter, the ability to pass on more foreign alleles affects network evolution \cite {w2004dispersal}. This research will be looking at the changes in genetic architecture dynamics and the 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 result in a less robust network and 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} However, over time, the network should adapt and become resistant to such perturbations.

**Methods**

To understand the effects from the evolutionary forces, I wrote a R script that constructs genetic networks and simulates their 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:

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

**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 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\*}

Which is the total protein concentration at each gene. There are four sites which represent the different factors affecting protein production. These are \alpha, \gamma\, \theta and $P$, where \alpha is the protein production rate while \gamma is the degradation rate \cite{omholt2000gene}.

For the 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 $S\_{y\_{j},\theta,P\_{j}} = \frac{y\_{j}^P\_{j}}{y\_{j}^P\_{j} + \theta^P\_{j}}$. The Hill Function explains the relationship between regulator and producer, where [theta] 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}

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

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

\end{equation\*}

\mu is the ratio of \alpha and \gamma 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}

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}

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\*}

**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}

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}

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.

**The simulation**

A total of 44 different permutations 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. The main population was kept at a constant trait value 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. In addition to this, these trait values also 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 stabilising selection towards different specified trait values, thus homogenizing the population over time \cite{sato2006effect}. These trait values represent the different environments. 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 value 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 $y\_3$. 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 ensures varying combination of parents are produced.

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 after the 80th generation. For the random pattern, a uniform distribution would randomly choose a migration rate between 1% and 5% should migration occur.

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. With migration, I wanted to see its effect on a robust network and how long it takes for the network to recover after migration. As the migration rate and patterns were set, can track allele values in-between periods of migration and analysing recovery. By the 700th generation, the network should be more robust that it recovers faster compared to when migration starts in the 80th generation. 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. Comparing them, variance after migration should be less as the network has had more time for evolutionary processes to act on and evolve.

**Results**

As trait values are distributed along a Cauchy distribution, the values must not be too far apart that probability is 0, meaning that migrant genes are not passed on to future generations, and see how the network develops with invading alleles. Therefore, the environmental distance has overlap as seen from the distributions in figure 1. Therefore foreign alleles has a significant chance of passing on their values to the next generation.

-highest possible fitness value was around 0.8

-network evolved slowly especially if homogenous and starting at a low value

-this is because the highest probability (median) from the Cauchy distribution is not that high

-migration would aid to make the network more fit

-even with migration, the network managed to stay around peak value of 0.8

-heterogenous populations would evolve more rapidly and reach the peak higher

**Discussion**

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}.

Predictions

-no ecological barriers restricting migration

-with no gene flow, population evolves rapidly to their environment (Garcia,tigano)- balance of selection and drift

-recombination and epistasis maintain adaptiveness even with gene flow (tigano)

-selection favours allele combinations that are more tightly linked as recombination increases with distance (tigano)

-heterogeneous system evolves more rapidly than homogenous

Shortcomings

**Conclusion**

**Looking forward**

In just a few centuries, societal development has caused many anthropogenic changes, many of which will have harmful long-term effects to many species. This coupled with geological changes, will lead to future shifts in the distribution of land as well as alteration of habitat states. One effect of this change can be migration. Whether it is forced migration or new route opportunities, the distribution and movement of species will also change in generations to come. These migrations can cause the introduction of novel genes into local habitats, implicating future evolutionary effects in these local habitats. Evolution of both the local species and migratory species. 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.