

The invasion and dispersal
of *Lavatera arborea*

Tereza Rieglova

Heriot Watt University

Edinburgh

September 2007

Contents

1	Introduction	4
1.1	Motivation	4
1.2	History and facts about <i>Lavatera arborea</i>	4
1.3	Interaction between native Puffin and invasive <i>L. arborea</i>	5
1.4	Project Outline	6
2	Model of the interaction between <i>L. arborea</i>, puffins and rabbits	9
2.1	Model description	9
2.2	Steady states	10
2.3	Linear stability analysis	11
3	Experimental methods to examine the genetic diversity of <i>L. arborea</i>	29
3.1	Background into experimental methods	29
3.1.1	Polymerase Chain Reaction	29
3.1.2	Amplified Fragment Length Polymorphism	30
3.1.3	Gel electrophoresis	31
3.2	Specific techniques applied to <i>L. arborea</i>	32
3.2.1	Sample collection	32
3.2.2	DNA extraction	32
3.2.3	AFLP	32
3.2.4	Gel scoring	34
4	Analysis of genetic data for <i>L. arborea</i>: Techniques and results	35
4.1	ANOVA test on heterozygosity	35
4.1.1	Determining the proportion of Heterozygotes	35
4.1.2	ANOVA and hypothesis testing	40
4.1.3	Results for ANOVA	42
4.1.4	Discussion of ANOVA	42
4.2	AMOVA test of population diversity	43
4.2.1	Calculating distances	43
4.2.2	Results of AMOVA	46
4.2.3	Discussion of AMOVA	47

5 Discussion	48
References	50

Abstract

This study investigates the spread of the invasive plant *Lavatera arborea* (Tree Mallow) on the Islands of the Firth of Forth and its impact on the puffin population. We investigate the relationship between these two species and suggest possible strategies to prevent puffin decline and extinction.

We develop a mathematical model to describe interactions between *L. arborea*, puffins and a herbivore and investigate conditions that lead to the control of *L. arborea* density and therefore the survival of the puffin population.

Experimental genetic tests were also undertaken on samples of *L. arborea* collected from four islands of the Firth of Forth. The resulting genetic data was analysed to understand the amount of genetic variation within and between each island population and to infer possible dispersal patterns between islands.

Acknowledgement

I would like to thank Andy White for his help and encouragement throughout this project.

I am also very grateful to Stephen Cavers from the Centre for Ecology and Hydrology, Edinburgh, for the opportunity to undertake the experimental aspects of this project and instructing me in the laboratory techniques and genetic analysis.

My thanks also to Bob Munro and Sam Davies from the Centre for Ecology and Hydrology, Edinburgh, who helped me in the laboratory experiments.

Finally, I would also like to thank Alex Cook from Heriot Watt University who assisted me in understanding the analysis and Iain Currie from Heriot Watt University for help on the statistical methods.

1 Introduction

1.1 Motivation

An invasive plant has a predisposition to thrive outside its natural range. Invasive plants are a threat to ecosystems with rare native species. They alter habitats and reduce biodiversity and expose native plants and animals to extreme environmental pressure [19]. In this project we will focus on an invasive plant *Lavatera arborea* (Tree Mallow) and its establishment and spread on the Islands of the Firth of Forth.

1.2 History and facts about *Lavatera arborea*

L. arborea is a Mediterranean-Atlantic herb tree. *L. arborea* was originally introduced to coastal locations by humans, from where it has spread and has established habitat in the southwest and west coast of England (see Figure 1). The presence of *L. arborea* has also been reported on Bass Rock, one of the islands of the Firth of Forth, since the 17th century. Recent studies have shown that *L. arborea* is now established on neighbouring islands to Bass Rock. It has reached high abundance on the islands Craighleith and Fidra and smaller populations are found on the islands of Inchcolm and Lamb [21] (see Figure 2).

L. arborea grows on rocky places and does not withstand intensive grazing, therefore is not found in areas with abundant livestock or rabbits. This is highly likely the reason why *L. arborea* does not grow on the Isle of May which is also located in the Firth of Forth as here rabbits are found in a high number [21]. *L. arborea* prefers a nitrogen-rich soil and therefore flourishes among sea-bird colonies which are rich in bird guano. Generally, *L. arborea* can only withstand mild winters. This might be one of the reasons for the recent success of *L. arborea* in the Forth Estuary, as the Edinburgh region has experienced mild winters over the last few successive years [5]. *L. arborea* is a tall plant that grows up to 3m and forms dense formations that cover large areas. The velvety leaves are 5-7 lobed and its flower colour varies from light pink to dark violet (see Figure 3). *L. arborea* lives for two years, in the first year it forms leaves and in the second it flowers and fruits [6].

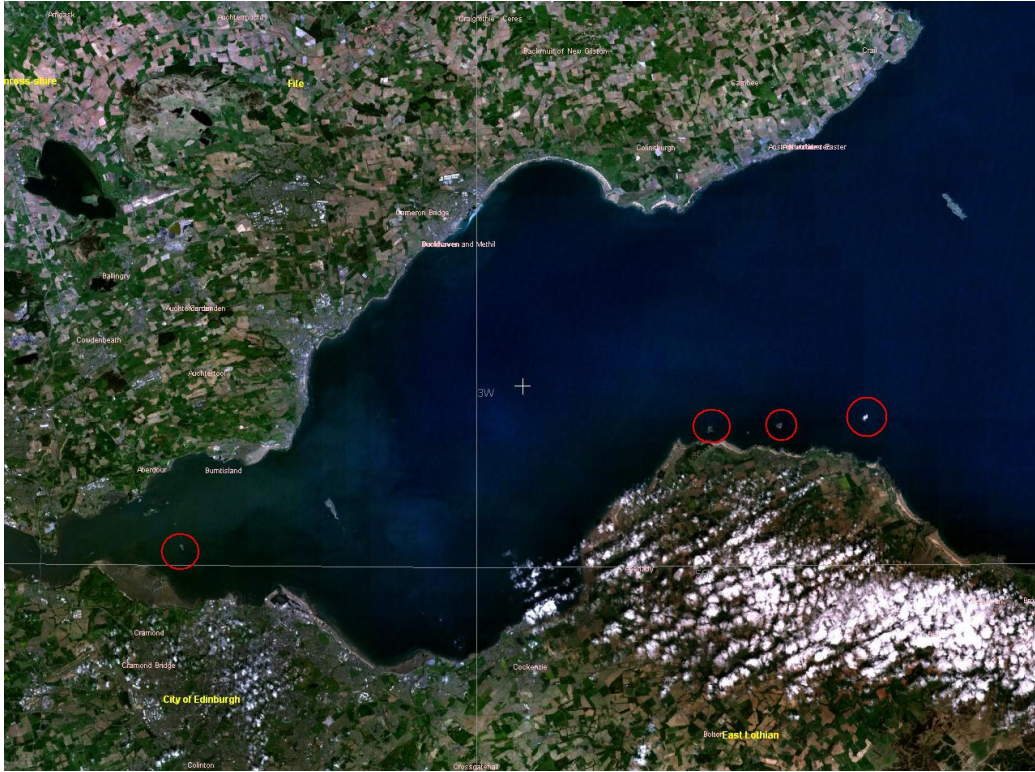


Figure 2: The Islands of the Firth Of Forth. The circled islands are (from the left to the right): Inchcolm, Fidra, Craigeith and Bass Rock. Satellite image, NASA World Wind [15].

1.4 Project Outline

In the project we wish to investigate the invasive properties of *L. arborea* and to do this we will (i) develop a mathematical model to understand the interaction between *L. arborea*, puffins and a herbivore (rabbits) and (ii) produce a DNA data set of *L. arborea* and statistical analysis of the data to explain genetic differences in *L. arborea* on the Islands of the Firth of Forth.

(i) Mathematical Model: Observations indicate that *L. arborea* is detrimental to the abundance of puffins and that herbivores may reduce the density of *L. arborea*. We will develop and analyse a mathematical model to understand these interactions. This will be composed of a system of three non-linear ordinary differential equations (ODEs). We will undertake a steady state and stability analysis to understand when ecosystems



Figure 3: *Lavatera arborea* in blossom. Picture taken from [17].

composed of different combinations of species will occur. We will verify results using numerical simulations.

Key objective: Understand the mechanisms that tend to *L. arborea* establishment and puffin decline/eradication.

(ii) Genetic diversity of *L. arborea* on the Islands of the Firth of Forth: When a few organisms start a new population and are isolated from the parental population, we often see that the new population is less genetically diverse than the original one: a genetic bottleneck [8, 9]. It is often observed that the frequencies of alleles¹ in a subpopulation differ from

¹An allele is one of the possible variations of the DNA sequence found at a specific locus/region on a DNA strand.



Figure 4: A puffin. Picture taken from [18].

the allele frequencies in the original population. The knowledge about allele frequencies gives us information that can be used to establish the population structure [10]. We will examine the DNA variation between each island's population (Bass Rock, Fidra, Craigleith and Inchcolm) to produce information about the spread of this species.

Key objective: To use results from the analysis of the DNA variation of *L. arborea* on the Islands of the Firth of Forth to determine the dispersal route of this invasive plant.

2 Model of the interaction between *L. arborea*, puffins and rabbits

In this section we introduce a mathematical framework that can be used to model the interaction between *L. arborea*, puffins and rabbits. It is a system of three ordinary differential equations for the density of *L. arborea* (T), puffins (P) and rabbits (R). We assume all parameters are positive. The model is represented as follows:

$$\frac{dP}{dt} = r_P P \left(1 - \frac{P}{k_P} \right) - c_T P T - a_P c_P P R \quad (1)$$

$$\frac{dT}{dt} = r_T T \left(1 - \frac{T}{k_T} \right) + a_T c_T P T - c_R T R \quad (2)$$

$$\frac{dR}{dt} = r_R R \left(1 - \frac{R}{k_R} \right) + a_R c_R T R - c_P P R \quad (3)$$

2.1 Model description

Equation (1) describes the dynamics of puffin population. Puffins have maximum growth rate r_P which reduces as puffin density increases towards the carrying capacity k_P . Puffins are known to be negatively affected by competition from *L. arborea* and we assume this effect is proportional to the density of *L. arborea* and puffins with a competition coefficient c_T . We also assume puffins and rabbits compete for territory with a competition coefficient $a_P c_P$ (where $a_p < 1$ reflects that rabbits are inferior competitors and $a_p > 1$ superior).

Equation (2) describes the dynamics of *L. arborea* population. *L. arborea* has maximum growth rate r_T which reduces as *L. arborea* density increases towards the carrying capacity k_T . *L. arborea* benefits from the presence of puffins, this effect is proportional to the rate of competition exerted on puffins by *L. arborea* with a conversion coefficient a_T . *L. arborea* is consumed by rabbits, this effect is proportional to the density of both species with the coefficient c_R .

Equation (3) represents the dynamics of the rabbit population. Rabbits have maximum growth rate r_R which reduces as rabbit density increases towards the carrying capacity k_R . Rabbits benefit from the consumption of

L. arborea, which is converted into births of rabbits with a conversion coefficient a_R . We also assume rabbits and puffins compete for territory with a competition coefficient c_P .

2.2 Steady states

The following steady states may occur depending on parameter values. Apart from the zero steady state, there is a set of non-zero steady states when one population exists in isolation:

$$[P, T, R] = [k_P, 0, 0]; \quad \text{Puffins alone at their carrying capacity.} \quad (4)$$

$$[P, T, R] = [0, k_T, 0]; \quad \text{\textit{L. arborea} alone at its carrying capacity.} \quad (5)$$

$$[P, T, R] = [0, 0, k_R]; \quad \text{Rabbits alone at their carrying capacity.} \quad (6)$$

The second group of steady states represents the coexistence of two populations in the absence of the third.

Coexistence of puffins and *L. arborea*:

$$[P, T, R] = \left[\frac{k_P r_T (r_P - c_T k_T)}{r_P r_T + a_T c_T^2 k_T k_P}, \quad \frac{r_P k_T (r_T + a_T c_T k_P)}{r_P r_T + a_T c_T^2 k_T k_P}, \quad 0 \right]. \quad (7)$$

This equilibrium is only biologically realistic if puffins and *L. arborea* have positive densities. This requires $r_P > c_T k_T$, i.e. that the growth rate of puffins exceeds the costs due to competition from *L. arborea*.

Coexistence of *L. arborea* and rabbits:

$$[P, T, R] = \left[0, \quad \frac{k_T r_R (r_T - c_R k_R)}{r_T r_R + a_R k_R c_R^2 k_T}, \quad \frac{r_T k_R (r_R + a_R c_R k_T)}{r_T r_R + a_R k_R c_R^2 k_T} \right]. \quad (8)$$

L. arborea and rabbit coexistence equilibrium is biologically realistic when $r_T > c_R k_R$, i.e. the growth rate of *L. arborea* must exceed the cost due to consumption from rabbits.

Coexistence of puffins and rabbits:

$$[P, T, R] = \left[\frac{r_R k_P (r_P - a_P c_P k_R)}{r_R r_P - a_P k_P c_P^2 k_R}, \quad 0, \quad \frac{k_R r_P (r_R - c_P k_P)}{r_R r_P - a_P k_P c_P^2 k_R} \right]. \quad (9)$$

Rabbits and puffins coexistence equilibrium is biologically realistic in two cases. Let **the first case** be $r_P > a_P c_P k_R$ and $r_R > c_P k_P$ (in which case the

denominator of both densities is automatically positive). This can be interpreted as: puffin growth rate exceeds costs due to competition from rabbits and rabbit growth rate exceeds costs due to competition from puffins. In **the second case** $r_P < a_P c_P k_R$ and $r_R < c_P k_P$ (in which case the denominator of both densities is automatically negative). These conditions suggest coexistence of rabbits and puffins in highly competitive environment.

The final steady state $[P_s, T_s, R_s]$ is the coexistence of all three species. We will discuss the feasibility of this steady state later. The algebraic expression for the densities at this steady state can be determined but is too cumbersome to detail.

2.3 Linear stability analysis

Firstly, we will look at the Jacobian matrix J of the system, which is

$$J = \begin{pmatrix} D_{11} & -c_T P & -a_P c_P P \\ a_T c_T T & D_{22} & -c_R T \\ -c_P R & a_R c_R R & D_{33} \end{pmatrix}$$

where

$$\begin{aligned} D_{11} &= r_P \left(1 - \frac{P}{k_P}\right) - \frac{r_P P}{k_P} - c_T T - a_P c_P R, \\ D_{22} &= r_T \left(1 - \frac{T}{k_T}\right) - \frac{r_T T}{k_T} + a_T c_T P - c_R R, \\ D_{33} &= r_R \left(1 - \frac{R}{k_R}\right) - \frac{r_R R}{k_R} + a_R c_R T - c_P P. \end{aligned}$$

We substitute our steady states into J and find the characteristic polynomials $\det(J - \lambda I) = 0$ in order to calculate the eigenvalues λ of the system (I is the identity matrix). For a steady state to be stable all eigenvalues must have negative real part $Re(\lambda) < 0$ [14].

In some situations determining the eigenvalues can be difficult and in these situations we use the *Routh-Hurwitz (R-H) conditions* [14]. The *R-H conditions* are necessary and sufficient for the requirement $Re(\lambda) < 0$ to hold. For the characteristic polynomial (in our case of order 3)

$$P(\lambda) = \lambda^3 + a_1 \lambda^2 + a_2 \lambda + a_3 = 0$$

the *R-H conditions* are

$$a_1 > 0, \quad a_3 > 0, \quad \det \begin{pmatrix} a_1 & a_3 \\ 1 & a_2 \end{pmatrix} > 0.$$

Let us examine the stability of our steady states one by one:

The trivial steady state $[0, 0, 0]$ has the stability matrix

$$J([0, 0, 0]) = \begin{pmatrix} r_P & 0 & 0 \\ 0 & r_T & 0 \\ 0 & 0 & r_R \end{pmatrix}.$$

The eigenvalues are r_P, r_T, r_R all of which are positive. This equilibrium is always unstable.

Stability of single species steady states

The steady state $[k_P, 0, 0]$ has the stability matrix

$$J([k_P, 0, 0]) = \begin{pmatrix} -r_P & -c_T k_P & -a_P c_P k_P \\ 0 & r_T + a_T c_T k_P & 0 \\ 0 & 0 & r_R - c_P k_P \end{pmatrix}.$$

The eigenvalues can be read from the leading diagonal as $-r_P, r_T + a_T c_T k_P$ and $r_R - c_P k_P$. Because we get at least one positive eigenvalue, the equilibrium is always an unstable steady state.

The steady state $[0, k_T, 0]$ has the stability matrix

$$J([0, k_T, 0]) = \begin{pmatrix} r_P - k_T c_T & 0 & 0 \\ a_T c_T k_T & -r_T & -c_R k_T \\ 0 & 0 & r_R + a_R c_R k_T \end{pmatrix}.$$

The eigenvalues are $r_P - k_T c_T, -r_T$ and $r_R + a_R c_R k_T$, this steady state is always unstable.

The stability matrix of the steady state $[0, 0, k_R]$ is

$$J([0, 0, k_R]) = \begin{pmatrix} r_P - a_P c_P k_R & 0 & 0 \\ 0 & r_T - k_R c_R & 0 \\ -c_P k_R & a_R c_R k_R & -r_R \end{pmatrix}.$$

The eigenvalues are $r_P - a_{PCP}k_R$, $r_T - k_Rc_R$ and $-r_R$. Therefore the steady state $[0, 0, k_R]$ is stable if and only if $r_P < a_{PCP}k_R$ and $r_T < c_Rk_R$, otherwise it is unstable.

The result for the stability analysis of steady states in which one species is present are summarized in Table 1.

Coexistence of puffins (P) and $L. arborea$ (T)

For $[P^*, T^*, 0]$ we get the stability matrix J equal to

$$\begin{pmatrix} r_P(1 - \frac{P^*}{k_P}) - \frac{r_PP^*}{k_P} - c_T T^* & -c_T P^* & -a_{PCP}P^* \\ a_T c_T T^* & r_T(1 - \frac{T^*}{k_T}) - \frac{r_T T^*}{k_T} + a_T c_T P^* & -c_R T^* \\ 0 & 0 & d_{33} \end{pmatrix},$$

where

$$d_{33} = r_R + a_{RCR}T^* - c_P P^*. \quad (10)$$

Here the stability can be decomposed into whether $d_{33} < 0$ and whether 2×2 submodel - for puffins and $L. arborea$ - has negative eigenvalues.

$$\frac{dP}{dt} = r_P P \left(1 - \frac{P}{k_P}\right) - c_T P T \quad (11)$$

$$\frac{dT}{dt} = r_T T \left(1 - \frac{T}{K_T}\right) + a_T c_T P T \quad (12)$$

Here we use the *R-H conditions* for the two species model. We obtain a characteristic polynomial $P(\lambda) = \lambda^2 + a_1\lambda + a_2$, with stability condition $a_1, a_2 > 0$. Here $a_1 = -Trace(J([P^*, T^*]))$ and $a_2 = Determinant(J([P^*, T^*]))$.

$$Det = \frac{r_P r_T (r_P - c_T k_T)(r_T + a_T c_T k_P)}{r_P r_T + a_T c_T^2 k_T k_P}$$

which is positive if and only if $r_P > c_T k_T$ (note this is the condition for this steady state to have positive density). We denote $\Theta = r_P - c_T k_T$ and assume that $\Theta > 0$. Then

$$Trace = -\frac{r_P r_T (\Theta + a_T c_T k_P + r_T)}{r_P r_T + a_T c_T^2 k_T k_P}$$

which is always < 0 .

In the absence of rabbits $[P^*, T^*]$ is stable provided it has a positive density. When rabbits are included, $[P^*, T^*, 0]$ is stable if it is positive ($r_P > c_T k_T$) and provided $d_{33} < 0$ which prevents rabbits from invading. We can investigate this condition further by substituting P^* and T^* into d_{33} . Then $d_{33} =$

$$\frac{r_T r_P (r_R - c_P k_P) + r_R a_T c_T^2 k_P k_T + a_R c_R r_P k_T T T + a_R c_R r_P k_P k_T a_T c_T + c_P k_P T T c_T k_T}{r_P r_T + a_T c_T^2 k_T k_P},$$

and d_{33} is definitely positive when $r_R > c_P k_P$. When $[P^*, T^*, 0]$ is positive it is unstable if $r_R > c_P k_P$ or $d_{33} > 0$ which represents the fact that rabbits can invade and increase in numbers.

In the absence of rabbits, if we start with a pure puffin population and introduce a small number of *L. arborea* then we leave the unstable steady state $[k_P, 0, 0]$ and the puffins density will start decreasing as *L. arborea* increases. Whether the coexistence steady state is stable or puffins die out depends on parameter values: if $r_P > c_T k_T$ then the long-term coexistence is possible, if $r_P < c_T k_T$ puffins will die out. Both cases are demonstrated by numerically solving the ODEs (1)-(3) in Figure 5.

Coexistence of *L. arborea* (T) and rabbits (R)

For $[0, T^*, R^*]$ we get the stability matrix J equal to

$$\begin{pmatrix} d_{11} & 0 & 0 \\ a_T c_T T^* & r_T \left(1 - \frac{T^*}{k_T}\right) - \frac{r_T T^*}{k_T} - c_R R^* & -c_R T^* \\ -c_P R^* & a_R c_R R^* & r_R \left(1 - \frac{R^*}{k_R}\right) - \frac{r_R R^*}{k_R} + a_R c_R T^* \end{pmatrix},$$

where

$$d_{11} = r_P - c_T T^* - a_P c_P R^*. \quad (13)$$

Here the stability can be decomposed into whether $d_{11} < 0$ and whether 2×2 submodel - for *L. arborea* and rabbits - has negative eigenvalues.

$$\frac{dT}{dt} = r_T T \left(1 - \frac{T}{K_T}\right) - c_r T R \quad (14)$$

$$\frac{dR}{dt} = r_R R \left(1 - \frac{R}{K_R}\right) + a_R c_R T R \quad (15)$$

Using the *R-H conditions* for the two species model we obtain $a_1 = -\text{Trace}(J([T^*, R^*]))$ and $a_2 = \text{Determinant}(J([T^*, R^*]))$.

$$\text{Det} = \frac{r_T r_R (r_T - c_R k_R) (r_R + a_R c_R k_T)}{r_T r_R + a_R c_R^2 k_R k_T}$$

which is positive if and only if $r_T > c_R k_R$, (this is a breach of one of the stability conditions for the steady state $[0, 0, k_R]$). We denote $\Theta = r_T - c_R k_R$ and assume that $\Theta > 0$. Then

$$\text{Trace} = -\frac{r_T r_R (\Theta + a_R c_R k_T + r_R)}{r_T r_R + a_R c_R^2 k_R k_T}$$

which is always < 0 .

In the absence of puffins $[T^*, R^*]$ is stable provided it has a positive density. When puffins are included $[0, T^*, R^*]$ is stable if it is positive ($r_T > c_R k_R$) and provided $d_{11} < 0$ which prevents puffins from invading. We can investigate this condition by substituting T^* and R^* into d_{11} .

$d_{11} =$

$$\frac{a_R k_T (r_P c_R k_R - a_P c_P k_R r_T) + r_R r_T (r_P - a_P c_P k_R) + c_T k_T r_R (c_R k_R - r_T)}{r_T r_R + a_R k_R c_R^2 k_T}$$

While the positive density condition holds ($r_T > c_R k_R$), d_{11} is negative as long as $r_P < a_P c_P k_P$ or $d_{11} < 0$. This represents the fact that puffins cannot invade and increase in numbers (therefore allowing $[0, T^*, R^*]$ to be a stable steady state).

In the absence of puffins, if we start with a pure *L. arborea* population and introduce a small number of rabbits, we leave the steady state $[0, k_T, 0]$ and the rabbit density increases and *L. arborea* decreases. Will we end up at some coexistence level or will we end up with no *L. arborea* left? It depends on parameter values: If $r_T > c_R k_R$ the coexistence is possible, if $r_T < c_R k_R$ *L. arborea* will disappear. Both cases are demonstrated in Figure 6.

Coexistence of puffins (P) and rabbits (R)

For $[P^*, 0, R^*]$ we get the stability matrix J equal to

$$\begin{pmatrix} r_P \left(1 - \frac{P^*}{k_P}\right) - \frac{r_P P^*}{k_P} - a_P c_P R^* & -c_T P^* & -a_P c_P P^* \\ 0 & d_{22} & 0 \\ -c_P R^* & a_R c_R R^* & r_R \left(1 - \frac{R^*}{k_R}\right) - \frac{r_R R^*}{k_R} - c_P P^* \end{pmatrix},$$

where

$$d_{22} = r_T - c_R R^* + a_T c_T P^*. \quad (16)$$

Here the stability can be decomposed into whether $d_{22} < 0$ and whether 2×2 submodel - for puffins and rabbits - has negative eigenvalues.

$$\frac{dP}{dt} = r_P P \left(1 - \frac{P}{k_P}\right) - a_P c_P P R \quad (17)$$

$$\frac{dR}{dt} = r_R R \left(1 - \frac{R}{k_R}\right) - c_P P R \quad (18)$$

Using the *R-H conditions* for the two species model we obtain $a_1 = -\text{Trace}(J([P^*, R^*]))$ and $a_2 = \text{Det}(J([P^*, R^*]))$.

$$\text{Det} = \frac{r_P r_R (r_P - a_P c_P k_R)(r_R - c_P k_P)}{r_P r_R - a_P c_P^2 k_P k_R}.$$

In the first case, see discussion of (8), we have $r_P > a_P c_P k_R$ (this is a breach of one of the stability conditions for the steady state $[0, 0, k_R]$), and $r_R > c_P k_P$. We denote $\Theta_1 = r_P - a_P c_P k_R$ and $\Theta_2 = r_R - c_P k_P$. Assume $\Theta_1, \Theta_2 > 0$, then

$$\text{Trace} = -\frac{r_P r_R (\Theta_1 + \Theta_2)}{r_P r_R - a_P c_P^2 k_P k_R} \quad \text{which is always } < 0.$$

Therefore $[P^*, 0, R^*]$ is a stable steady state if $r_P > a_P c_P k_R, r_R > c_P k_P$ and $d_{22} < 0$ which prevents *L. arborea* from invading. **In the second case, see discussion of (8)**, we have $r_P < a_P c_P k_R$ and $r_R < c_P k_P$. Here *Determinant* < 0 , even though *Trace* remains negative. Therefore $[P^*, 0, R^*]$ with $r_P < a_P c_P k_R, r_R < c_P k_P$ is not a stable steady state.

In the absence of *L. arborea*, $[P^*, R^*]$ is stable provided it has a positive density. When *L. arborea* is included, $[P^*, 0, R^*]$ is stable if it is positive and $d_{33} < 0$ which prevents *L. arborea* from invading. We can substitute P^* and R^* into d_{22} .

$d_{22} =$

$$\frac{r_P r (r_T - c_R k_R) + c_P k_P (c_R k_R r_P - a_P c_P k_R r_T) + a_T c_T r_R k_P (r_P - a_P c_P k_R)}{r_P r_P - a_P c_P^2 k_P k_R}$$

The condition $d_{22} < 0$ prevents *L. arborea* from invading the puffin, rabbit coexistence. In the absence of *L. arborea*, what happens when we introduce

few rabbits to a pure puffin population? If $r_R < c_P k_P$ then in the long-term nothing will change. Otherwise, if $r_R > c_P k_P$, then puffin density will decline with increasing rabbit numbers. Will these dynamics stop at some coexistence level or will puffins die out? This depends on the inequality $r_P > a_P c_P k_R$. If it holds, then coexistence of both species will occur, but if $r_P < a_P c_P k_R$ puffins will die out (Figure 7).

Table 1. The summary of the one-species and two species steady states:

	Positive density condition:	Stability condition:
Puffins	always	never stable
<i>L. arborea</i>	always	never stable
Rabbits	always	$r_P < a_P c_P k_R$ and $r_T < c_R k_R$
Puffins and <i>L. arborea</i>	$r_P > c_T k_T$	$r_P > c_T k_T$ and $r_R + a_R c_R T^* - c_P P^* < 0$
<i>L. arborea</i> and rabbits	$r_T > c_R k_R$	$r_T > c_R k_R$ and $r_P - c_T T^* - a_P c_P R^* < 0$
Puffins and rabbits	$r_P > a_P c_P k_R$ and $r_R > c_P k_P$ $r_P < a_P c_P k_R$ and $r_R < c_P k_P$	$r_P > a_P c_P k_R$ and $r_R > c_P k_P$ and $r_T + a_T c_T P^* - c_R R^* < 0$ always unstable

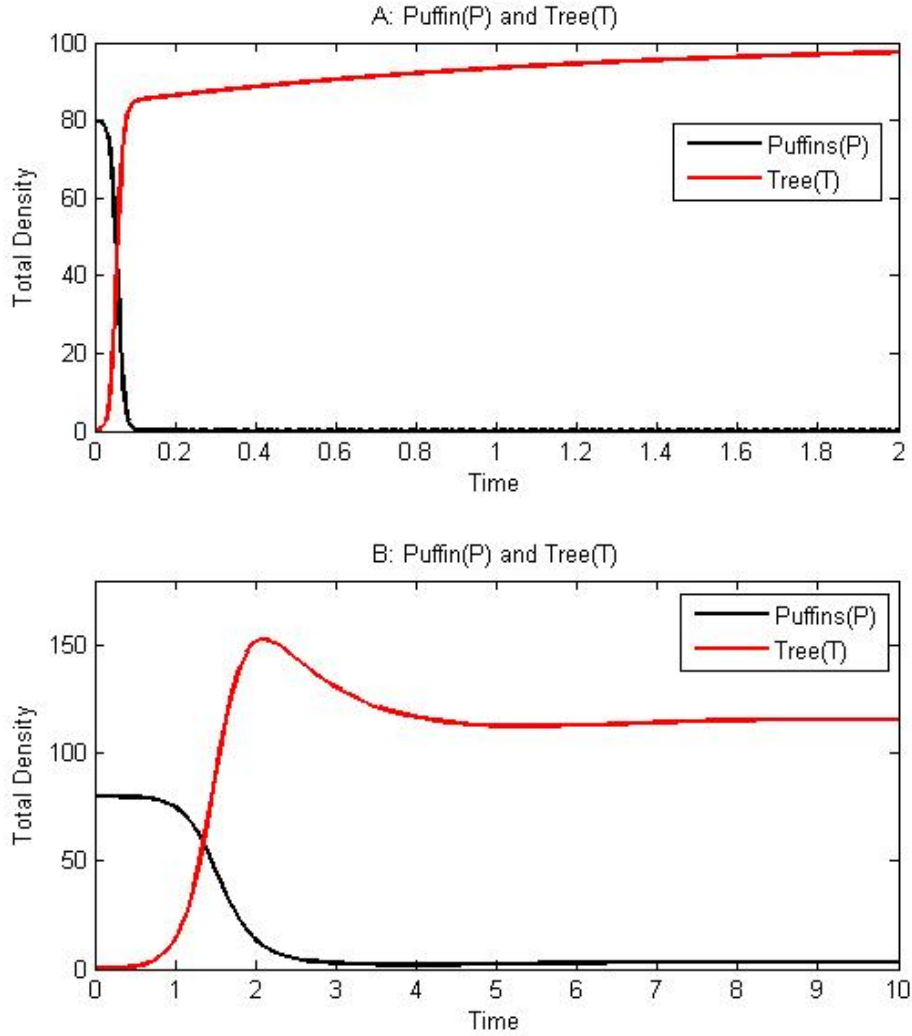


Figure 5: A: Dynamics of puffins and *L. arborea*: $r_P < c_T k_T$. Parameter values: $r_P = 6$, $k_P = 80$, $c_T = 1.5$, $r_T = 1$, $k_T = 100$, $a_T = 1$. In the absence of rabbits, if we introduce *L. arborea* to a pure puffin population, puffins will die out. B: Dynamics of puffins and *L. arborea*: $r_P > c_T k_T$. Parameter values are the same as in A except c_T : $c_T = 0.05$. In the absence of rabbits, if we introduce *L. arborea* to a pure puffin population, a coexistence steady state will evolve.

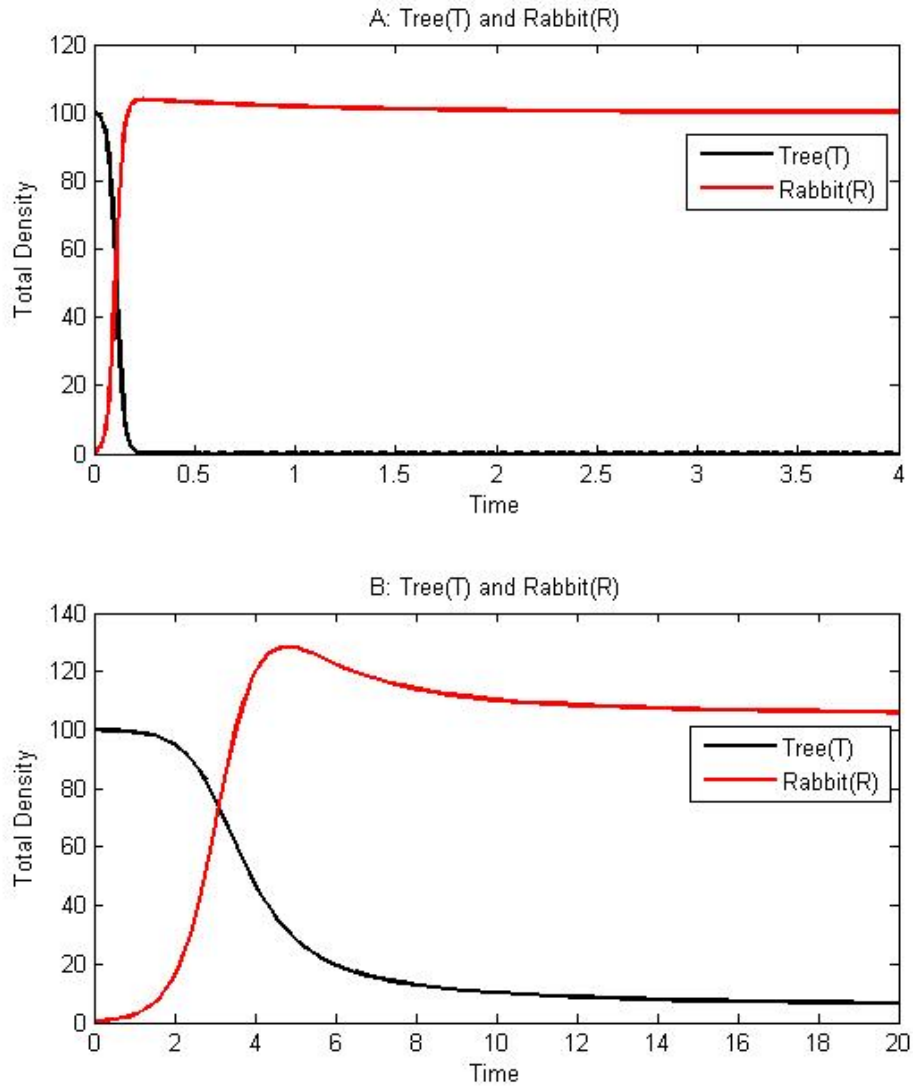


Figure 6: A: Dynamics of *L. arborea* and rabbits: $r_T < c_R k_R$. Parameter values: $r_T = r_R = 1$, $k_T = k_R = 100$, $c_R = 0.5$, $a_R = 1$. In the absence of puffins, if we introduce rabbits to a pure *L. arborea* population, *L. arborea* will disappear. B: Dynamics of *L. arborea* and rabbits: $r_T > c_R k_R$. Parameter values are the same as in A except c_R : $c_R = 0.009$. In the absence of puffins, if we introduce rabbits to a pure *L. arborea* population, a coexistence steady state will evolve.

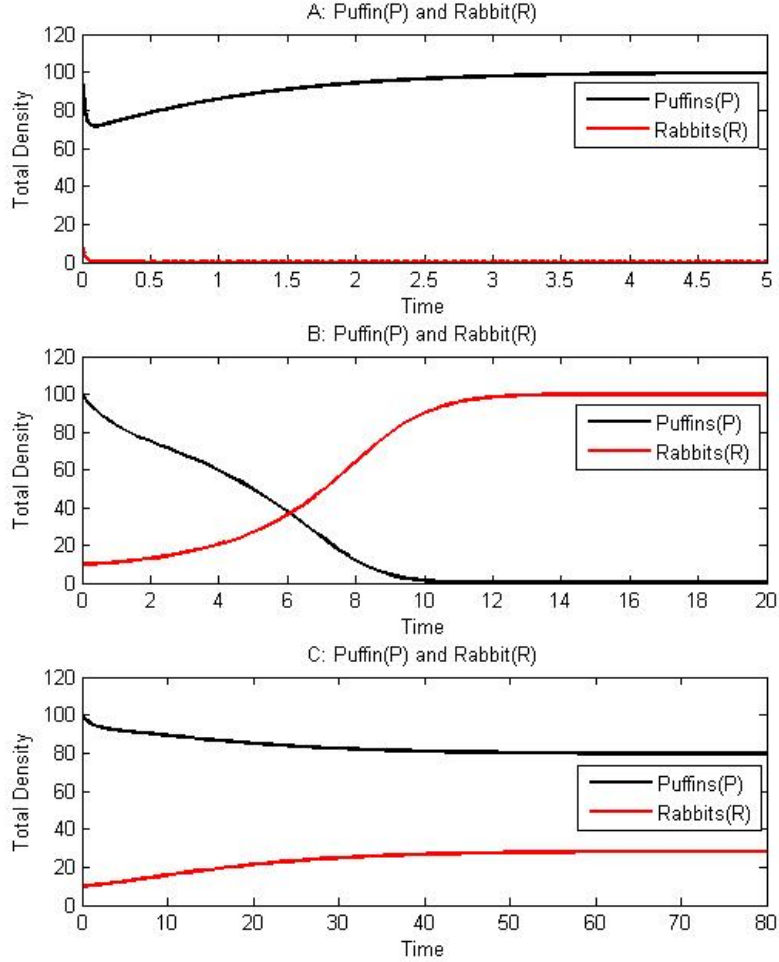


Figure 7: A: Dynamics of puffins and rabbits: $r_R < c_P k_P$. The parameters $r_P = r_R = 1$ and $k_P = k_R = 100$ are common parameter values to A, B and C. Specific parameter values: $a_P = 3$, $c_P = 0.6$. In the absence of *L. arborea*, if we introduce rabbits to a pure puffin population, puffins will reach their carrying capacity - in the long term. B: Dynamics of puffins and rabbits: $r_R > c_P k_P$ and $r_P < a_P c_P k_R$. Specific parameter values: $a_P = 3$, $c_P = 0.009$. In the absence of *L. arborea*, if we introduce rabbits to a pure puffin population, puffins will die out. C: Dynamics of puffins and rabbits: $r_R > c_P k_P$ and $r_P > a_P c_P k_P$. Specific parameter values: $a_P = 0.8$, $c_P = 0.009$. In the absence of *L. arborea*, if we introduce rabbits to a pure puffin population, a coexistence steady state will evolve.

Coexistence of puffins (P), $L. arborea$ (T) and rabbits (R)

From the steady state conditions we know that at the three species equilibrium the following holds:

$$\begin{aligned} r_P(1 - \frac{P}{k_P}) - c_T T - a_P c_P R &= 0 \quad \text{which gives} \quad \frac{r_P P}{k_P} = \underbrace{r_P - c_T T - a_P c_P R}_{d_{11}}, \\ r_T(1 - \frac{T}{k_T}) + a_T c_T P - c_R R &= 0 \quad \text{which gives} \quad \frac{r_T T}{k_T} = \underbrace{r_T + a_T c_T P - c_R R}_{d_{22}}, \\ r_R(1 - \frac{R}{k_R}) + a_R c_R T - c_P P &= 0 \quad \text{which gives} \quad \frac{r_R R}{k_R} = \underbrace{r_R + a_R c_R T - c_P P}_{d_{33}}. \end{aligned}$$

For $[P_s, T_s, R_s]$ to be positive requires that $d_{11} > 0$, $d_{22} > 0$ and $d_{33} > 0$ (referring to (10), (13), (16)). Therefore the two species coexistence equilibria become unstable whenever the three species coexistence has a positive density.

For $[P_s, T_s, R_s]$ we get the stability matrix

$$J = \begin{pmatrix} -\frac{r_P P_s}{k_P} & -c_T P_s & -a_P c_P P_s \\ a_T c_T T_s & -\frac{r_T T_s}{k_T} & -c_R T_s \\ -c_P R_s & a_R c_R R_s & -\frac{r_R R_s}{k_R} \end{pmatrix},$$

We will use the R - H conditions to derive stability conditions. The characteristic polynomial is $P(\lambda) = \lambda^3 + a_1 \lambda^2 + a_2 \lambda + a_3$ with

$$a_1 = -\text{Trace}(J) = \frac{r_P P_s}{k_P} + \frac{r_T T_s}{k_T} + \frac{r_R R_s}{k_R}$$

which is positive whenever this steady state has a positive density,

$$\begin{aligned} a_3 &= \frac{P_s T_s R_s}{k_P k_T k_R} (r_P r_T r_R - a_P c_P^2 k_P k_R r_T + r_P c_R^2 k_T a_R k_R + a_P c_P k_P a_T c_T k_T a_R c_R k_R + \\ &\quad + a_T c_T^2 k_T k_P r_R + c_T k_T c_P k_P c_R k_R) \\ &= \frac{P_s T_s R_s}{k_P k_T k_R} (r_T (r_P r_R - a_P c_P^2 k_P k_R) + r_P c_R^2 k_T a_R k_R + a_P c_P k_P a_T c_T k_T a_R c_R k_R + \\ &\quad + a_T c_T^2 k_T k_P r_R + c_T k_T c_P k_P c_R k_R), \end{aligned}$$

which can attain positive or negative values. In the case when $r_P r_R > a_P c_P^2 k_P k_R$ it is definitely a positive number (and this is true whenever the

two species puffins, rabbits equilibrium has a positive density).

The last condition for the stability is $a_1 a_2 - a_3 > 0$. This condition is difficult to analyse. When this condition holds the three species equilibrium is stable. If it fails but $a_1 > 0, a_3 > 0$, we expect the population to exhibit limit cycles in all three species.

Figure 8 demonstrates what happens when the third species is introduced at low density to a two species coexistence steady state. (All parameter values were chosen so that they allow the three species equilibrium $[P_s, T_s, R_s]$ to have positive densities.)

Conservation of puffins

A key motivation for analysing a theoretical system for puffin, *L. arborea*, rabbit interactions is to understand the conditions that prevent puffins from becoming extinct. Since evidence [21] shows that in the absence of rabbits, puffin densities are reduced by *L. arborea*, we will assume the extreme case where *L. arborea* will outcompete and replace puffins ($r_P < c_T k_T$). Evidence [21] also suggests that rabbits will totally exclude *L. arborea* in the absence of puffins, so we assume $r_T < c_R k_R$.

Clearly with this parameter set-up in the absence of rabbits, *L. arborea* will exclude puffins and in the absence of puffins rabbits will exclude *L. arborea*. It is of interest to investigate whether when all three species are introduced this will allow coexistence and therefore prevent extinction of puffins, while controlling *L. arborea* density. To examine this we conduct numerical simulations. In all cases we use initial conditions $P = \frac{1}{2} k_P, T = \frac{1}{2} k_T$. We then introduce rabbits at a low density and examine the temporal dynamics of the system for different combinations of the interaction between rabbits and puffins (i.e. for all the combinations of when inequalities $r_P > a_P c_P k_R$ and $r_R > c_P k_P$ hold and do not hold).

In Figure 9 we assume that $r_P > a_P c_P k_R$ and $r_R < c_P k_P$. This means that puffins would outcompete rabbits in the absence of *L. arborea*. For these criteria the three species coexistence steady state is the only equilibrium that can potentially be stable (Table 1). Figure 9 indicates that the three species coexistence equilibrium is approached by a process of damped oscillation. If the interaction parameters are altered, in particular a_P , then the three species equilibrium becomes unstable and the population exhibits periodic oscillations.

Oscillations arise as high initial *L. arborea* density benefits rabbits which increase and drive *L. arborea* to a low density. High rabbit density and low *L. arborea* benefit puffins which increases and drive rabbits to a low density. High puffin density then benefits *L. arborea* which increase to high density and the cycle repeats. The periodic oscillations arise since each species can outcompete the other in the absence of the third species (an ecological paper-scissors-stone game).

In Figure 10 we assume that $r_P > a_P c_P k_R$ and $r_R > c_P k_P$. This means that puffins would coexist with rabbits in the absence of *L. arborea*. There are now two possibilities. We can observe a stable three species equilibrium, puffins and rabbits coexist with rabbits controlling *L. arborea* density which prevents the exclusion of puffins (Fig. 10A). Alternatively, rabbits can exclude *L. arborea* and then the rabbits and puffins coexist at a two species steady state (Fig. 10B).

In Figure 11A we assume that $r_P < a_P c_P k_R$ and $r_R < c_P k_P$. These conditions enable the single rabbit species to be a stable steady state, as $r_P < a_P c_P k_R$ and $r_T < c_R k_R$ (Table 1). A coexistence of puffins and rabbits under these conditions is unstable. As *L. arborea* grows puffin density declines and introduction of rabbits to the island will result into *L. arborea* numbers declining and the puffin population will not be able to compete with abundant rabbit population that has benefited from rich *L. arborea* presence.

In Figure 11B we assume that $r_P < a_P c_P k_R$ and $r_R > c_P k_P$. The same dynamics appears. A single rabbit species stable steady state is possible, as $r_P < a_P c_P k_R$ and $r_T < c_R k_R$. Under these conditions none of the two species steady states is possible (we never get two species steady state densities positive at the same time).

Discussion

We constructed a system of 3 ODEs to examine the interaction of puffins, *L. arborea* and a herbivore (which we assume to be rabbits). We undertook a stability analysis to explain when the various single, two and three species steady states were stable and how they related to each other. We then fixed a certain parameter combination based on observations in the field. This assumed that *L. arborea* exclude puffins and rabbits exclude *L. arborea*. We then tested how habitats composed of *L. arborea* and puffins, in which puffins would be driven to extinction would respond to the introduction of the herbivore. If the herbivore was a weaker competitor than puffins then the system

could exhibit a three species coexistence or three species periodic oscillations. If puffins and rabbits could coexist in the absence of *L. arborea* then a three species coexistence or a two species coexistence without *L. arborea* was observed. In these cases puffins could be prevented from extinction. If the herbivore is a superior competitor to puffin then puffins would be excluded by rabbits and rabbits would survive in a single species equilibrium.

Although this theory has been undertaken in a simplified, general theoretical framework, it does highlight that a possible conservation strategy to preserve puffins would be to introduce a suitable herbivore of *L. arborea* onto the Islands of the Firth of Forth where *L. arborea* is detrimentally affecting the puffin density.

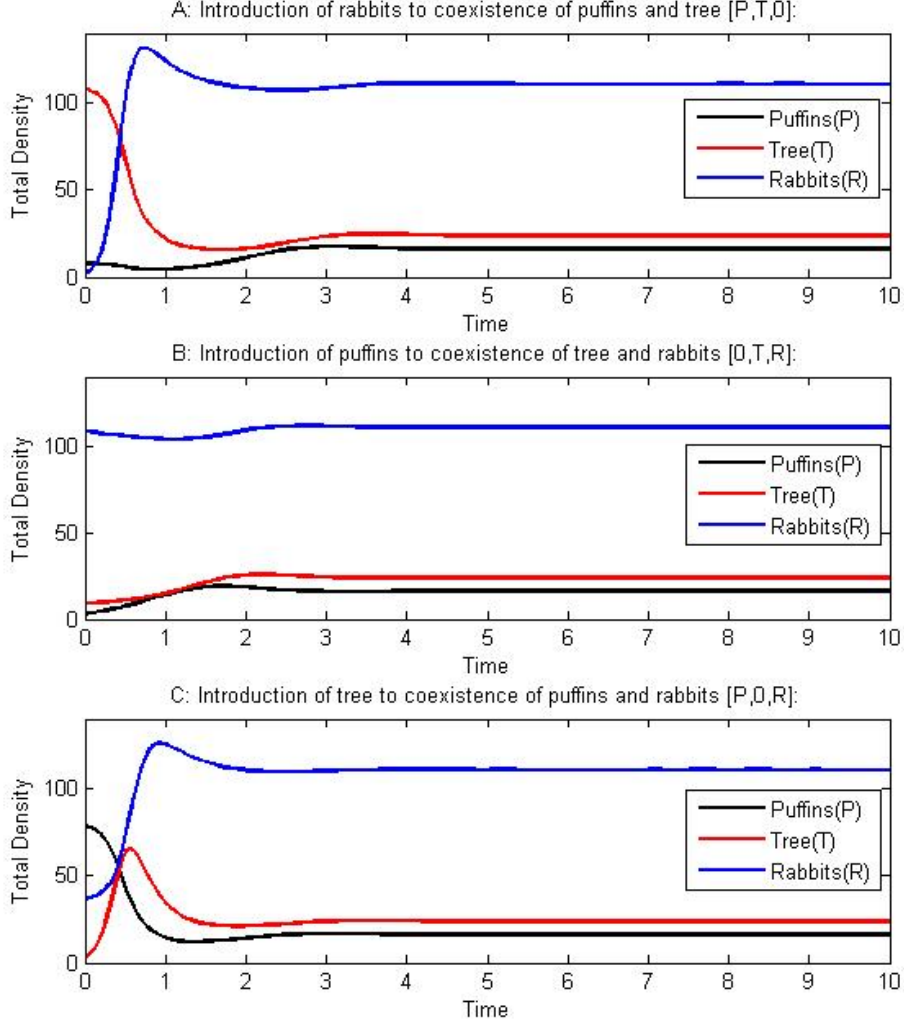


Figure 8: The temporal dynamics of the puffins, *L. arborea*, rabbits system when a third species is introduced to a two species equilibrium. In A a small number of rabbits was introduced to a long-term coexistence levels of puffins and *L. arborea* $[P^*, T^*, 0]$. In B a small number of puffins was introduced to a long-term coexistence levels of *L. arborea* and rabbits $[0, T^*, R^*]$. In C a small number of *L. arborea* was introduced to long-term coexistence levels of puffins and rabbits $[P^*, 0, R^*]$. Parameter values: $r_P = 7$, $r_T = 6$, $r_R = 5$, $c_P = 0.04$, $c_T = 0.06$, $c_R = 0.05$, $k_P = k_T = k_R = 100$, $a_P = a_T = a_R = 1$.

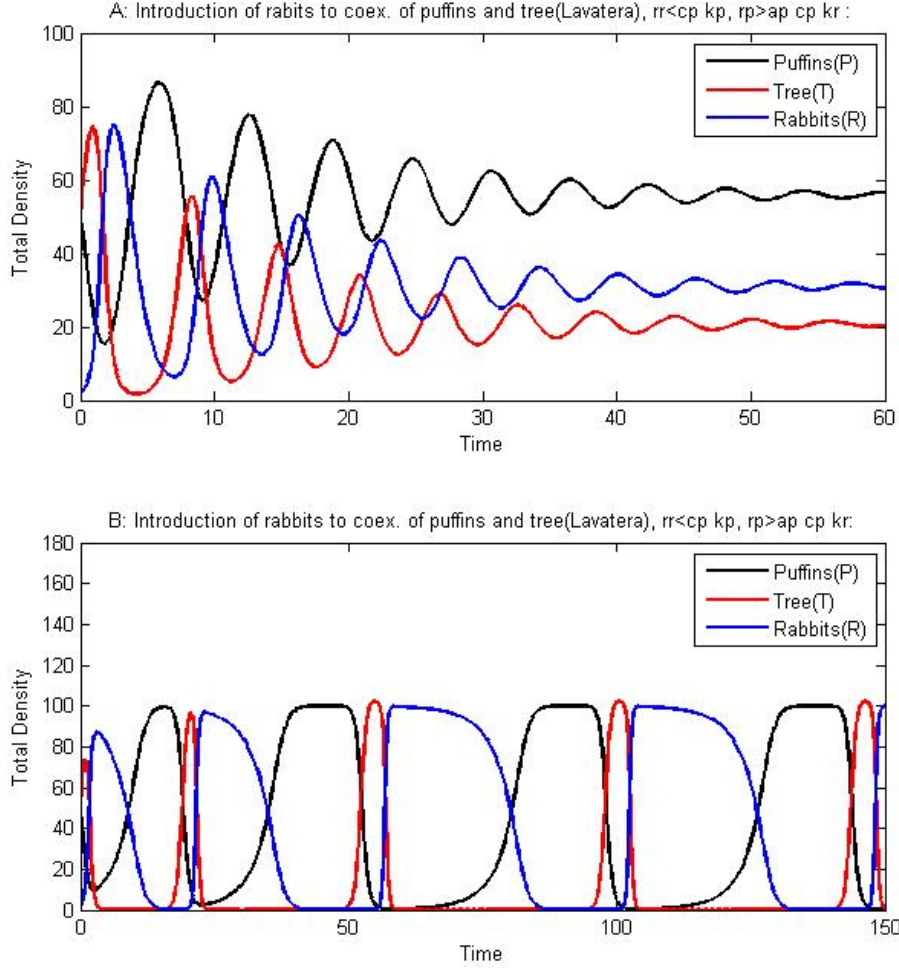


Figure 9: The temporal dynamics of the puffins, *L. arborea*, rabbits system. It is assumed that puffins could be totally excluded by *L. arborea* ($r_P < c_T k_T$) and that *L. arborea* can be totally excluded by rabbits ($r_T < c_R k_R$), and $r_P > a_P c_P k_R$ and $r_R < c_P k_P$. Parameter values in A: $r_P = r_R = 3, r_T = 1, k_P = k_T = k_R = 100, c_P = c_T = c_R = 0.04, a_P = 0.4, a_T = a_R = 0.2$. Referring to the *R-H conditions* here $a_1 \approx 2.81 > 0, a_3 \approx 3.11 > 0, a_1 a_2 - a_3 \approx 1.29 > 0$. In B: all parameters remain the same as in the A except a_P : $a_P = 0.7$. Referring to the *R-H conditions* here $a_1 \approx 2.69 > 0, a_3 \approx 1.77 > 0, a_1 a_2 - a_3 \approx -1.02 < 0$.

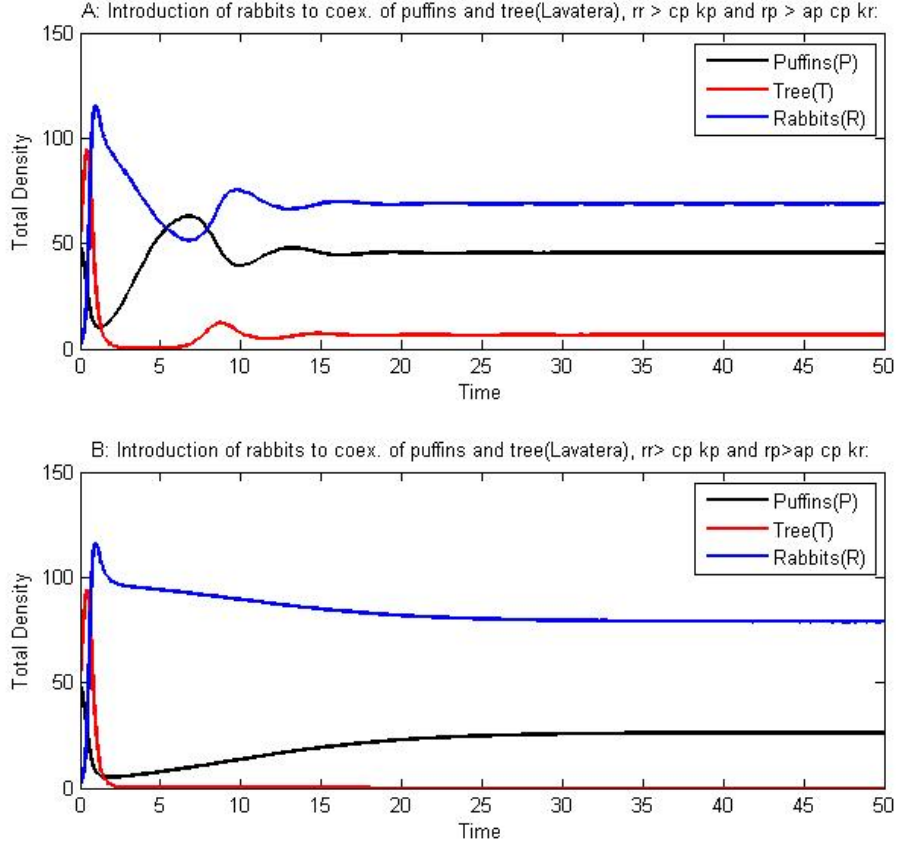


Figure 10: As in Figure 9 puffins could be totally excluded by *L. arborea* ($r_P < c_T k_T$) and that *L. arborea* can be totally excluded by rabbits ($r_T < c_R k_R$), and two more conditions hold: $r_P > a_P c_P k_R$ and $r_R > c_P k_P$ then possible outcomes include an oscillatory stable steady state and coexistence of puffins and rabbits without *L. arborea*. Parameter values in A: $r_P = 3, r_T = 1, r_R = 5, c_P = c_T = c_R = 0.04, k_P = k_T = k_R = 100, a_P = 0.5, a_T = a_R = 1$, referring to the *R-H conditions* here $a_1 \approx 4.87 > 0, a_3 \approx 4.69 > 0, a_1 a_2 - a_3 \approx 13.28 > 0$. In B: all parameters are the same as in A except a_P , here $a_P = 0.7$, referring to the *R-H conditions* here $a_1 \approx 4.51 > 0, a_3 \approx -1.13 < 0, a_1 a_2 - a_3 \approx 4.03 > 0$.

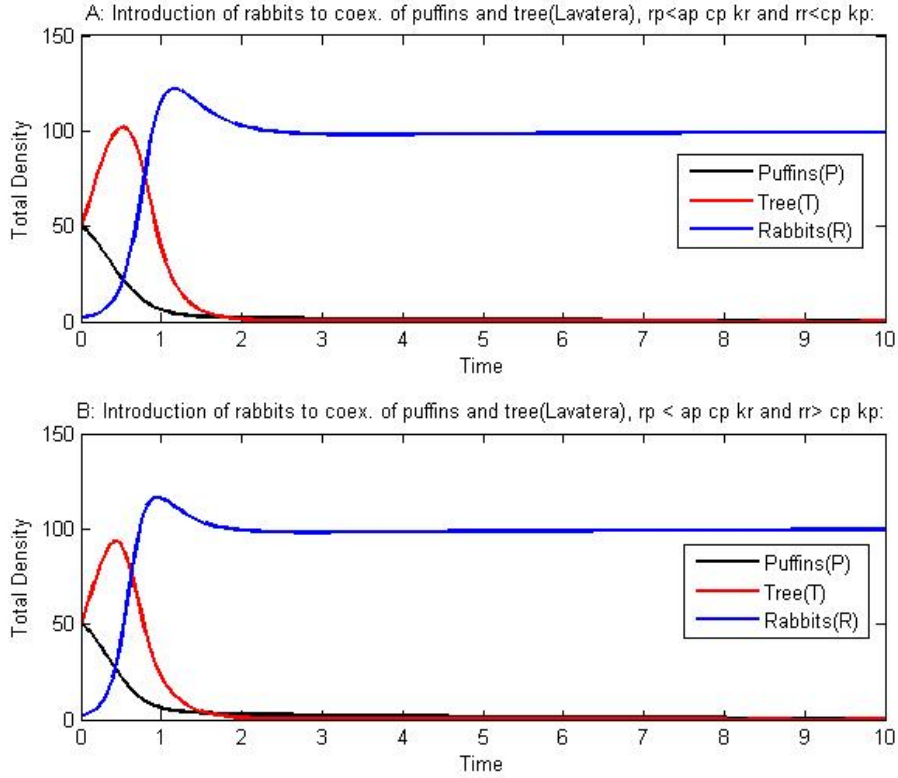


Figure 11: The temporal dynamics of the puffins, *L. arborea*, rabbits system. In A and B we chose parameters that again satisfy that puffins are excluded by *L. arborea* ($r_P < c_T k_T$) and *L. arborea* is excluded by rabbits ($r_T < c_R k_R$). In A: $r_P < a_P c_P k_R$. In addition we assume $r_R < c_P k_P$. We always end up with no puffins. Parameter values $r_P = r_R = 3, r_T = 1, k_P = k_T = k_R = 100, c_P = c_T = c_R = 0.04, a_P = 0.8, a_T = a_R = 1$. Referring to the *R-H conditions* here $a_1 \approx 2.80 > 0, a_3 \approx 1.15 > 0, a_1 a_2 - a_3 \approx -1.89 < 0$. In B: $r_R > c_P k_P$ (as opposed to the condition in case A), but here again $r_P < a_P c_P k_R$. Only one outcome is possible: puffins die out. Parameter values are the same as in A, except r_R , which was changed to $r_R = 5$. Referring to the *R-H conditions* here $a_1 \approx 4.33 > 0, a_3 \approx -3.43 < 0, a_1 a_2 - a_3 \approx 0.71 > 0$.

3 Experimental methods to examine the genetic diversity of *L. arborea*

In this section we will discuss in general the techniques used to manipulate and record genetic (DNA) data. We will then describe specific details of how we apply these techniques to samples of *L. arborea* collected from 4 islands in the Firth of Forth.

3.1 Background into experimental methods

We will briefly describe basic techniques used to manipulate DNA. They include Polymerase Chain Reaction (PCR), Amplified Fragment Length Polymorphism (AFLP) and gel electrophoresis.

3.1.1 Polymerase Chain Reaction

PCR is a quick, easy and cheap method to extract and amplify a specific region on a DNA strand. To run PCR three basic ingredients are needed: two primers² that flank the region to be amplified, a synthesizing enzyme that creates a complementary strand to a single DNA strand, and finally the building blocks - nucleotides (dNTPs, namely T,A,G,C).

PCR takes place in cycles. In the first cycle, the DNA molecule is denatured by heating, so that from a double-stranded DNA molecule two single-stranded molecules are obtained. The reaction is then cooled to a specific ‘annealing temperature’ and each primer anneals to one end of the DNA region to be amplified. In the next step the mixture is heated to 72°C and the extension of primers takes place. This produces two double-stranded molecules. The cycle of denaturation, annealing, extension is then repeated approximately 30 times. The number of copies grows exponentially with each cycle (2^{n-1} copies in n th cycle) [4, 10].

²Primer is a short single stranded DNA sequence that anneals to the sequence to be amplified. Two primers fully specify the region. Primers are also used by a synthesizing enzyme for chain elongation.

3.1.2 Amplified Fragment Length Polymorphism

In comparison with other methods, AFLP is a fast and reliable method to generate large number of genetic markers³ particularly where little is known about the species in advance. AFLP technique was first published in 1995 and has been widely used since then [11]. The main feature of AFLP is its capacity for simultaneous screening of many different DNA regions distributed randomly across the genome.

The technique involves three steps - firstly DNA digestion and ligation, secondly pre-amplification of digested DNA fragments with ligated adaptors, and finally selective amplification of pre-amplified fragments [11, 13, 22].

Digestion and ligation: First the double stranded DNA is digested by restriction enzymes. A restriction enzyme can recognize a specific sequence on the DNA strand and cut at that specific site to produce DNA fragments. If, by introducing a restriction enzyme we do not obtain any fragments, this indicates that the specific restriction sequence is not present on the DNA strand. Depending on the presence or absence of the restriction sequences along the DNA molecule, digestion with restriction enzymes can produce different numbers of fragments for different DNA molecules. There exists a variety of restriction enzymes, some recognize a short sequence (e.g. restriction enzyme *MseI* recognizes a sequence consisting of four nucleotides), some recognize a longer sequence (e.g. restriction enzyme *EcoR I* cuts at a sequence consisting of six nucleotides). As the first group recognises short sequences, it cuts frequently, whilst the second group cuts more rarely, due to a longer recognition site. The number of fragments obtained by applying a frequent cutter is generally much higher than the number when a rare cutter is used. This is because the short sequence occurs more often in the DNA strand than the specific long base sequence the rare cutter requires. Most often, fragments of interest have been cut by a rare cutter on one end and by a frequent cutter on the other end. Subsequently, short double stranded fragments called adaptors are ligated to the ends of restriction fragments.

Pre-amplification: The next stage is a process of PCR, usually called pre-amplification, and uses digested - ligated DNA fragments obtained in the

³A genetic marker is a specific DNA sequence in a specific region on a DNA strand that can vary within a population under examination. We consider genetic maker to be a sequence at a specific DNA locus/region that enables to identify genetically distinct individuals.

previous stage as a template. From digested-ligated fragments a subset of fragments is amplified using pre-amplification primers.

Selective amplification: Pre-amplification primers used in pre-amplification are still not discriminating enough⁴. Pre-amplification is performed to amplify those fragments that represent a set from which further selection can be done. From the pre-amplified set a subset of fragments can be obtained by constructing selective primers. Selective primers are based on the primers that were used in the pre-amplification stage but extended further by attaching additional nucleotides to them. A primer extension of one, two or three nucleotides reduces the number of amplified fragments by factors of 4, 16 and 64, respectively [6]. PCR using selective primers is referred to as a selective amplification.

The outcome of AFLP is often scored using gel electrophoresis. We should mention that an amplified DNA fragment (one DNA fragment in many copies) forms what we call a band. Therefore what we see on the gels are bands of different lengths, each band is a multiple copy of a specific DNA fragment.

3.1.3 Gel electrophoresis

Gel electrophoresis is a technique that is used to compare genetic differences among individuals. Gel electrophoresis sorts fragments of DNA that differ in length. If we obtain DNA from different individuals, apply AFLP-method to obtain fragments, and then sort these fragments using gel electrophoresis, we obtain different pattern for each individual⁵.

The technique is based on the fact that DNA fragments are negatively charged and therefore are attracted to a positively charged anode. After the gel is prepared we insert fragments into the gel in a lane on one end of the cradle filled with the gel. We place a positive anode at the opposite end of the cradle and apply electric current. Subsequently, all fragments start to migrate across the gel towards the positive anode. Shorter fragments move through the gel more easily than the longer fragments. Therefore, after a while, the fragments are sorted according to their lengths, shorter fragments

⁴Our aim is to obtain only limited number of fragments, that will provide easily readable pattern when scored on the gel. With too many fragments, the pattern on the gel would become hardly readable and less useful.

⁵Only clones or twins would have the same pattern.

ahead of the longer ones.

Gels can be made with different media and different concentrations, depending on the resolution required. Agarose is used for low resolution electrophoresis, whilst polyacrylamide is used for higher resolution analysis.

3.2 Specific techniques applied to *L. arborea*

General techniques for determining the presence or absence of genetic markers for samples of *L. arborea* was applied in the project. The sample collection and a DNA extraction was undertaken prior to the project, but the experimental procedures of PCR, AFLP and gel electrophoresis we conducted in this project, leading to a data set for the presence or absence of 51 genetic markers for samples of *L. arborea*.

3.2.1 Sample collection

Samples of plant *Lavatera arborea* were collected in June 2005 on the Islands of the Firth of Forth:

1. Bass Rock (43 samples),
2. Inchcolm (41 samples),
3. Craigleith (40 samples),
4. Fidra (16 samples).

3.2.2 DNA extraction

All DNA samples were extracted from leaf tissue of *Lavatera arborea*. This was conducted prior to this project and it is these samples which have been used in this project.

3.2.3 AFLP

Digestion

Digestion of DNA fragments was performed using 5 μ l sample DNA, 2 μ l of 10 \times Reaction Buffer, 1 μ l of restriction enzyme *EcoRI*, 1 μ l of restriction enzyme *MseI*, and Distilled Water was added to a final volume of 25 μ l. Ingredients were gently mixed and centrifuged. The mixture was then incubated for 2 hours at 37°C and then for another 15 min at 70°C to inactivate the

restriction enzymes.

Ligation

Ligation of Adaptors to DNA fragments was performed using 24 μ l of Adapter/Ligation Solution (provided in the IRDye Fluorescent AFLP Template Preparation Kit for Large Plant Genome Analysis (LI-COR Biosciences)) and 1 μ l of T4 DNA Ligase. These two components were added to the mixture, mixed and briefly centrifuged, then incubated at 20°C for 2 hours. 1:10 dilution of the mixture was performed. Diluted ligation mixture was then used for pre-amplification.

Pre-amplification

Pre-amplification of diluted mixture was performed using 1 μ l of primer combination *EcoRI*+0 and 1 μ l of primer combination *MseI*+TC, 1 μ l of dNTP, 2 μ l of 10 \times Buffer, 0.5 μ l of Taq and 9.5 μ l of Distilled Water.

The PCR was performed on the ThermoHybaid MultiBlock System. The Cycle started by denaturing the samples by heating the samples to 94°C for 5 minutes and then 21 cycles of {94°C for 1 minute, 54°C for 30 seconds, 72°C for 1 minute} were performed.

The resulting mixture was looked at using agarose gel (1g of agarose, 100ml of 1 \times TBE, 50 μ l of Ethidium Bromide).

All samples were diluted by adding 180 μ l of Distilled Water.

Selective PCR

Selective PCR was performed using 5 μ l of DNA pre-amplified mixture, 1 μ l of primer combination *EcoRI*+AT, 1 μ l of primer combination *EcoRI*+AC, 1 μ l of primer combination *MseI*+TCGA, 2 μ l of dNTP, 2.4 μ l of 10 \times Buffer, 11.1 μ l of Distilled Water and 0.5 μ l of Taq. PCR was performed using ThermoHybaid MultiBlock System. The cycle started by denaturing the samples by heating the samples to 94°C and then 36 cycles of {94°C for 1 minute, 54°C for 30 seconds, 72°C for 1 minute} were performed. Obtained mixture was diluted in ratio 1 (DNA mixture) : 9 (Distilled Water).

Prepared sample mixture was combined with a formamide loading dye, (1 μ l of sample mixture and 2 μ l of dye), denatured at 94°C for 5 minutes and electrophoresed on denaturing polyacrylamide gels (8g of Urea, 2ml of 10 \times TBE, 3.2ml of Long Ranger, 14.8ml of Deionised Water, 10ml of Temed and

100 μ l of APS) using a LI-COR Long ReadIR 4200 DNA sequencer. Inserted in the gel was 0.3 μ l of each sample mixture.

The final results are represented by photo images of gels containing fragments with primer combination *EcoRI*+AT and *MseI*+TCGA and fragments with primer combination *EcoRI*+AC and *MseI*+TCGA. An example of the gel image with primer combination *EcoRI*+AT and *MseI*+TCGA is given in Figure 12.

Following the same steps as described above, another AFLP was performed but using different primer combinations: *EcoRI*+AA with *MseI* + TCGA and *EcoRI*+CC with *MseI* + TCGA.

3.2.4 Gel scoring

Obtained fragments were sorted according to their lengths. We score a gel by looking for the presence or absence of a marker across all samples. For example we pick up one line on the gel, see Figure 12, and look across all the samples (in Figure 12 there are 36 samples, i.e. 36 columns, there are also 3 additional columns - the first, twentieth and the last - that were filled with standard DNA fragments that serve as a measure which all real samples can be compared to) and then make a record of presence or absence of that band. We use the term marker for a line on the gel. Each line is equivalent to one marker. That is because each line represents a specific region/locus on the DNA where the cutting sequence is either present (band on the gel is present) or absent (band on the gel is absent).

Out of four gel images only two gel images were clear enough to score: gels with primer combinations *EcoRI*+AT with *MseI* + TCGA and *EcoRI*+AA with *MseI* + TCGA, the former providing a much better image than the latter. In total 49 markers were scored for the primer combination *EcoRI*+AC with *MseI*+TCGA and only 2 markers were scored for the primer combination *EcoRI*+AA with *MseI*+TCGA. Polyacrylamide gels were scored manually. Only polymorphic bands were scored, i.e. those that across all samples contained at least one sample with a band missing, or at least one with a band present.

From the initial number of 140 samples only 123 samples provided a readable pattern on the gels. Therefore 51 genetic markers (called 51 loci) were

scored for 123 *L. arborea* samples. A data matrix of dimension 51×123 was obtained and entered on an Excel spreadsheet as a binary matrix (1 was used for recording a band presence, 0 was used to record a band absence).

4 Analysis of genetic data for *L. arborea*: Techniques and results

We choose two methods to analyse the genetic variation in our data. In section 4.1 we outline the technique and results for conducting an Analysis of Variance (ANOVA) test on the proportion of heterozygotes in the samples. In section 4.2 we outline technique and results for an Analysis of Molecular Variance (AMOVA), a method which analyses the genetic distance between samples directly from the presence/absence data.

4.1 ANOVA test on heterozygosity

We are interested in testing the genetic variation among *L. arborea* populations on the different islands of the Firth of Forth. This will help us to infer information about the spread of the population between islands. One way to test genetic variation is to convert the presence/absence data into a table which represents the proportions of heterozygotes⁶ at each loci. This is a measure of the variation of alleles at each loci. We can then compare the variation in the proportion of heterozygotes between the population samples for the different islands.

4.1.1 Determining the proportion of Heterozygotes

To determine the number of heterozygotes for each island population we assume that individuals reproduce according to the assumptions of the Hardy-Weinberg equilibrium [10]. These assumptions are:

- The organism is diploid.

⁶Every diploid organism has two copies of DNA molecule. We are often interested whether, at a specific region, these two copies are completely the same or whether they differ. An individual whose DNA copies differ regarding to a specific DNA region is called a heterozygote, whereas an individual whose both DNA copies are the same at that region is called a homozygote (relating to that specific region). An individual can be a homozygote for some DNA regions and a heterozygote for another DNA regions.

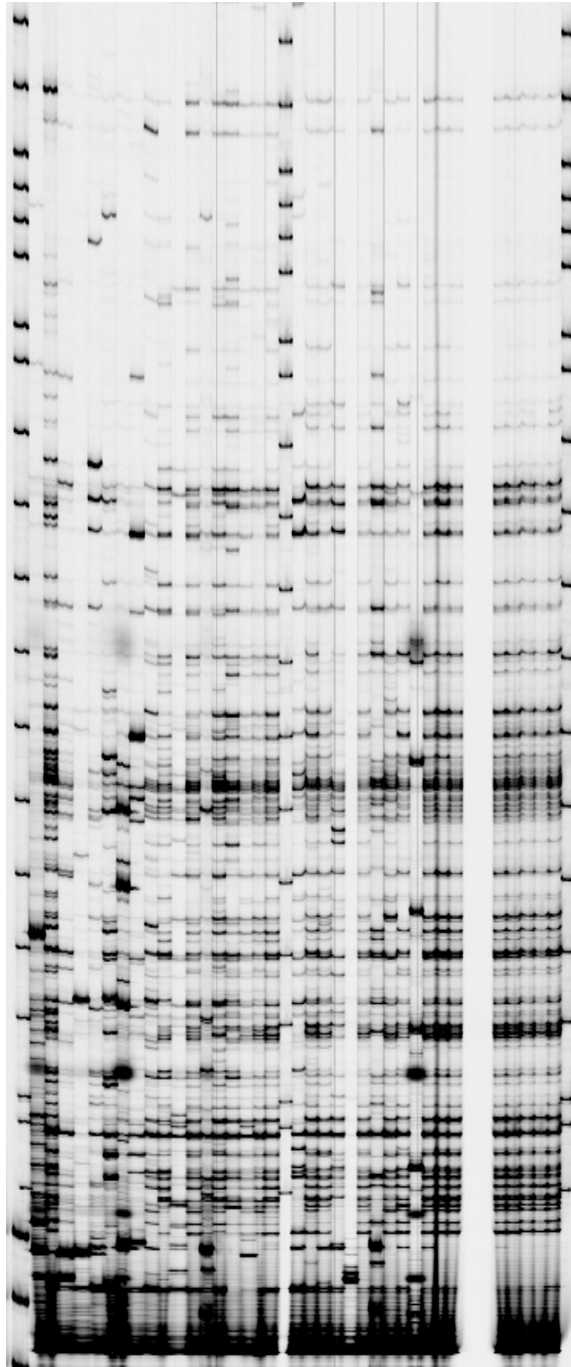


Figure 12: The image of the polyacrylamide gel electrophoresis of the first 36 *L. arborea* samples. Black bands are DNA fragments with primer combination *EcoRI*+AT and *MseI* + *TCGA*.

- Reproduction is sexual.
- Generations are nonoverlapping.
- Each marker has two alleles.
- Mating is random.
- Mutation is ignored.
- Migration is negligible.

We will assume that *L. arborea* is diploid, i.e. possesses two copies of a DNA molecule. (For the purposes of this study, we will assume that the species behaves as a simple diploid (Cavers, personal correspondence)).

The gel image (Figure 12) reveals differences between individual samples. Let us consider locus 1 for the Bass Rock population and calculate how many times the band was present and how many times that band is absent. To convert presence/absence proportions into information about alleles we need to clarify the way AFLP method works. The band is not present on the gel image if and only if the cutting sequence at locus 1 is not present on either of the two DNA copies of a diploid organism. On the other hand, the band is present on the gel image if the cutting sequence is present on at least one of the DNA copies a diploid organism possesses. We know that each individual is one of the following types (genotype) at locus 1:

- Type 1 (homozygote) : cutting sequence present on both DNA copies
- Type 2 (heterozygote): cutting sequence present on one copy but absent on the other copy of the DNA (in any order)
- Type 3 (homozygote) : cutting sequence absent on both DNA copies

We now relate the cutting sequence being present as equivalent to the individual possessing allele A at a particular loci. Using this notation individual types (genotypes) are:

- Type 1 (homozygote) : AA
- Type 2 (heterozygote): Aa or aA
- Type 3 (homozygote) : aa

To enable distinction between the $j = 1, \dots, 51$ loci we add the subscripts j as follows: A_j and a_j . To distinguish k island populations $k \in \{BR, IC, F, CL\}$,

we add superscripts as follows: A_j^k and a_j^k . The frequency of ‘cutting sequence presence’ (A) at locus 1 for the Bass Rock population is denoted p_1^{BR} . Frequency of ‘cutting sequence absence’ (a) at locus 1 across the Bass Rock population is denoted q_1^{BR} . If we know the proportions of type A and type a allele in the population we can use the Hardy-Weinberg model to determine the proportion of heterozygotes (type Aa or aA) in the population. For locus 1 at the Bass Rock this is determined in the Table 2.

Table 2: Hardy-Weinberg model for the genotype of the next generation

		mother		
		Allele	Cutting Sequence	Cutting sequence
		Allele	present - locus 1	absent - locus 1
		Freq.	(p_1^k)	(q_1^k)
father	Cutting sequence present - locus 1	(p_1^k)	$A_1^k A_1^k$ ($(p_1^k)^2$)	$A_1^k a_1^k$ ($p_1^k q_1^k$)
	Cutting sequence absent - locus 1	(q_1^k)	$a_1^k A_1^k$ ($q_1^k p_1^k$)	$a_1^k a_1^k$ ($(q_1^k)^2$)

The proportion (frequency) of homozygotes of the type ‘presence, presence’ at locus 1 (i.e. $A_1^k A_1^k$) in the Bass Rock population ($k = BR$) is equal to $(p_1^k)^2$, because an offspring gets randomly one DNA strand with ‘presence’ from one parent (which occurs with frequency p_1^k in the Bass Rock population) and independently gets one DNA strand with ‘presence’ from the other parent (which occurs with frequency p_1^k in the Bass Rock population). Similarly, the proportion of homozygotes of the type ‘absence, absence’ (i.e. $a_1^k a_1^k$) is equal to $(q_1^k)^2$ (allele a_1^k occurs with frequency q_1^k in the Bass Rock population). Finally, the proportion of heterozygotes of the type ‘presence, absence’ (i.e. $A_1^k a_1^k$) in the Bass Rock population is equal to $p_1^k q_1^k$, because an offspring gets randomly one DNA strand with ‘presence’ from one parent and independently gets one DNA strand with ‘absence’ from the other parent. Heterozygotes of the type ‘absence, presence’ (i.e. $a_1^k A_1^k$) is equal to $q_1^k p_1^k$. Therefore proportion of heterozygotes (of the type $A_1^k a_1^k$ or $a_1^k A_1^k$) is equal to $2p_1^k q_1^k$.

The proportion of individuals with the band absent at a particular locus is equivalent to the proportion of genotype aa in the population for that locus. The proportion of individuals with the band present is equivalent to the

proportion of genotypes AA, Aa and aA in the population.

The frequency of the genotype $a_j^k a_j^k$ at DNA location $j = 1$ in a subpopulation of Bass Rock ($k = BR$) is $\frac{4}{40} = 0.1$, because among 40 individuals from Bass Rock, 4 individuals were AFLP-scored as 0. Therefore 0.9 represents the frequency of individuals with genotype $A_1^k A_1^k$, $A_1^k a_1^k$ or $a_1^k A_1^k$.

If assumptions of Hardy-Weinberg model were all satisfied, then we could estimate allele A frequencies and allele a frequencies by solving the trivial system

$$\begin{aligned}(p_j^k)^2 + 2p_j^k q_j^k &= 0.9 \\ (q_j^k)^2 &= 0.1\end{aligned}$$

The estimated Bass Rock frequencies would therefore be $p_j^k = 0.684$ and $q_j^k = 0.316$. In the *L. arborea* population on Bass Rock, the frequency of individuals with genotype $A_j^k A_j^k$ would be $(p_j^k)^2 = 0.468$, individual with genotype $a_j^k a_j^k$ would be $(q_j^k)^2 = 0.1$ and individual with genotype $A_j^k a_j^k$ would be $2p_j^k q_j^k = 0.432$. We define heterozygosity to be $h_j^k = 2p_j^k q_j^k = 0.432$, for $j = 1$, $k = BR$ (i.e. heterozygosity in the Bass Rock population for locus $j = 1$).

By examining 51 random loci for each island population, we obtain 51 random samples of heterozygosity in each of the 4 island population.

Remark: As we do not have the information about real genotypes, we lack the full observation. We have observed only genotypes $a_j a_j$, from which frequencies p_j and q_j are estimated. In our approach we can only assume that the Hardy-Weinberg model holds and that estimated values of p_j and q_j are so close to real values that we can use them as true values.

In Table 3 we calculated a mean value of heterozygosity for each island population and standard error of each mean. We also present the percentage of polymorphic bands (defined earlier) within samples from each island. The number of samples over which each band is examined has an influence on proclaiming the band as polymorphic, therefore (having available substantially less samples from Fidra than from other islands) we treat that information with caution.

Table 3: General statistics relating to *L.arborea* populations

	Bass Rock	Craigleith	Fidra	Inchcolm
Population size	40	37	13	33
Mean Heterozygosity (Standard error of the mean)	0.408 (0.019)	0.364 (0.024)	0.335 (0.024)	0.251 (0.024)
Percentage of Polymorphic Bands	98%	86%	86%	75%

4.1.2 ANOVA and hypothesis testing

The F-statistic is often used to test a hypothesis that the means of all samples are equal (H_0), versus the alternative hypothesis that at least one of them is different from the others (H_1)⁷. We would like to test whether the means of heterozygosities of all four islands are equal, versus the alternative hypothesis that at least one of them is different from the others.

The assumption underlying this analysis is that samples, h_j^k with $j = 1, \dots, 51$, $k \in \{BR, CL, F, IC\}$, are drawn from normal distributions which have equal variances:

$$h_j^{BR} \sim N(\mu_{(BR)}, \sigma), \quad j = 1, \dots, 51$$

$$h_j^{CL} \sim N(\mu_{(CL)}, \sigma), \quad j = 1, \dots, 51$$

$$h_j^F \sim N(\mu_{(F)}, \sigma), \quad j = 1, \dots, 51$$

$$h_j^{IC} \sim N(\mu_{(IC)}, \sigma), \quad j = 1, \dots, 51$$

If the means are not equal, the test will give us an evidence that some islands have lower heterozygosity. This will indicate that there are genetic differences between the islands. Apart from that, we may then infer that the population with the highest heterozygosities was the original population. Dispersal from the island has produced subpopulations on the other islands with lower diversity.

Let us outline the idea of testing: We would like to break down the total variance of all observations into variance due to differences between islands, and the variance due to differences within islands.

⁷Note: In population genetics the F-statistics often refers to a quantity F_{ST} which relates to the proportion of genetic structure in a subpopulation - similar to the ϕ -statistic we introduce later.

Total sum of squares = Between sum of squares + Within sum of squares

The total sum of squares, i.e. the total sum of deviations from the overall mean, is given by

$$\sum_{k \in \{BR, CL, F, IC\}} \sum_{j=1}^{51} (h_j^k - \bar{h})^2,$$

where \bar{h} is a mean over the whole *L. arborea* population, so called grand average [2].

The between sum of squares is given by

$$\sum_{k \in \{BR, CL, F, IC\}} \sum_{j=1}^{51} (\bar{h}^k - \bar{h})^2,$$

where each term in the sum measures deviations of the island averages from the grand average.

The within sum of squares is given by

$$\sum_{k \in \{BR, CL, F, IC\}} \sum_{j=1}^{51} (h_j^k - \bar{h}^k)^2,$$

where each term in the sum measures deviations of the observations from the island mean and so the within sum of squares measures dispersion within the islands.

The F-statistic is then calculated [12] as

$$F = \frac{\frac{1}{k-1} \text{Between sum of squares}}{\frac{1}{n-k} \text{Within sum of squares}},$$

where k now denotes the number of categories, here $k = 4$ (corresponding to 4 islands) and n denotes number of all samples, here $n = 51 \times 4$.

A table for the F distribution [20] is used to look up a significance level for F (so called *p-value*). A *p-value* is the proportion of the tail area under the F distribution density function with degrees of freedom $k - 1$ and $n - k$ that is cut-off by the obtained F-statistic. The smaller the *p-value*, the more

likely it is that the null hypothesis H_0 does not hold. We say the null hypothesis, H_0 , is rejected with $(100-p)\%$ confidence.

Alternatively, we can look up the critical value F_{crit} of the F distribution which cuts off 0.1% of the tail area. We find that $F_{\text{crit}}(k - 1, n - k) = 5.63$. Therefore if the obtained F-statistic > 5.63 , we could conclude that the null hypothesis H_0 is rejected with 99.9% confidence.

4.1.3 Results for ANOVA

The following values can be computed for the heterozygosity data.

Total sum of squares	=	6.0199
Between sum of squares	=	0.6759
Within sum of squares	=	5.3440

Degrees of freedom: $k - 1 = 4 - 1 = 3$ and $n - k = 204 - 4 = 200$.

$$F = \frac{\frac{1}{3}0.6759}{\frac{1}{200}5.3440} = 8.4317$$

The result provides a strong evidence against H_0 . We reject H_0 with 99.9% confidence.

4.1.4 Discussion of ANOVA

The ANOVA test shows that mean heterozygosities differ at an extremely high level of significance. Therefore we assume that the island populations of *L. arborea* are genetically distinct. Table 3 also indicates that Bass Rock contains the highest diversity. Craigleith and Fidra contain less diversity. This may indicate that Bass Rock was the source of the populations on Fidra and Craigleith and that Inchcolm may have been populated by Fidra or Craigleith. It may also indicate that Fidra and Craigleith have experienced more dispersal events from Bass Rock than Inchcolm.

4.2 AMOVA test of population diversity

The Analysis of Molecular Variance (AMOVA) method can be used to test a hypothesis about the genetic difference in molecular data directly from the record of the presence or absence of bands. It is based on the ‘genetic distances’ between individuals.

4.2.1 Calculating distances

The presence/absence data produce vectors of zeros and ones. For the simplified gel as demonstrated in Figure 13 we have 3 samples (\mathbf{p}_1 , \mathbf{p}_2 , \mathbf{p}_3) scored at 10 regions/loci:

$$\mathbf{p}_1 = [1, 1, 1, 0, 1, 1, 1, 1, 1, 1],$$

$$\mathbf{p}_2 = [1, 0, 0, 1, 1, 1, 1, 1, 1, 1],$$

$$\mathbf{p}_3 = [0, 1, 1, 0, 1, 1, 1, 1, 1, 1].$$

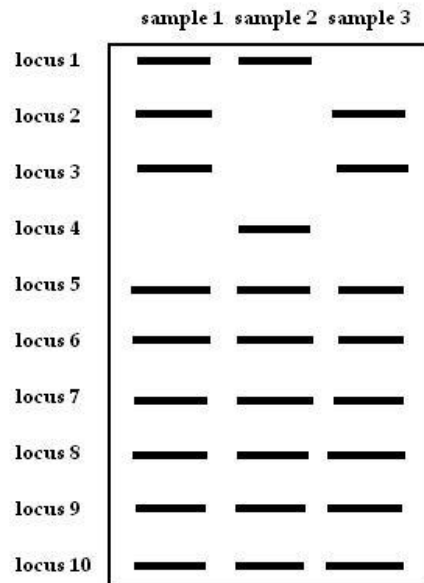


Figure 13: Demonstrative example of a gel

Based on these vectors we can calculate the distances between each couple of samples. The distance between two samples is defined as the number of loci in which they differ. Therefore, the distance between \mathbf{p}_1 and \mathbf{p}_2 is 3, because they differ at 3 loci (second, third and fourth). Distance between \mathbf{p}_1 and \mathbf{p}_3 is 1 (they differ at the first locus), and the distance between \mathbf{p}_2 and \mathbf{p}_3 is 4.

Table 4: Demonstrative distance table

	sample 1	sample 2	sample 3
sample 1	0	3	1
sample 2	3	0	4
sample 3	1	4	0

We will denote the distance between samples j and l as δ_{jl} . Our genetic distance matrix, containing δ_{jl} , $j = 1, \dots, 123$, $l = 1, \dots, 123$, is therefore a symmetric matrix with zeros on its diagonal, size 123×123 (we have 123 samples altogether) as in the example below:

Table 5: Distance matrix for *L.arborea* data

	<i>BassRock</i> ₁	<i>BassRock</i> ₂	...	<i>Craigleith</i> ₁	<i>Craigleith</i> ₂	...
<i>BassRock</i> ₁	0	23		23	22	...
<i>BassRock</i> ₂	23	0		10	11	...
⋮						...
<i>Craigleith</i> ₁	23	10		0	1	...
<i>Craigleith</i> ₂	22	11		1	0	...
⋮	⋮	⋮		⋮	⋮	...

We can pinpoint 4 submatrices along the main diagonal, which will contain distances between samples only within Bass Rock, Craigleith, etc. The remaining values are distances between samples from different islands.

We would like to calculate the total sum of squares of the deviations of the samples (in our demonstrative example $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3$) from the mean $\bar{\mathbf{p}} = (\mathbf{p}_1 + \mathbf{p}_2 + \mathbf{p}_3)/3$. The sum of squared deviations, SS , from the mean $\bar{\mathbf{p}}$ can be written as a sum of distances δ_{jl} between samples, barring a given constant [1, 3, 7, 23]. In the demonstrative example $SS = (\mathbf{p}_1 - \bar{\mathbf{p}})'(\mathbf{p}_1 - \bar{\mathbf{p}}) + (\mathbf{p}_2 - \bar{\mathbf{p}})'(\mathbf{p}_2 - \bar{\mathbf{p}}) + (\mathbf{p}_3 - \bar{\mathbf{p}})'(\mathbf{p}_3 - \bar{\mathbf{p}}) = 2.667$ which is equal to calculating $SS = (\delta_{12} + \delta_{13} + \delta_{23})/3 = (3 + 1 + 4)/3 = 2.667$.

Therefore the total sum of squares, SS_{total} , for our *L. arborea* data is defined as follows⁸: a mean distance vector is $\bar{\mathbf{p}} = \sum_{j=1}^n \mathbf{p}_j$ and a sum of squared deviations of each \mathbf{p}_j , $j = 1, \dots, 123$ from the mean distance vector $\bar{\mathbf{p}}$ is

$$SS_{\text{total}} = \sum_{j=1}^n (\mathbf{p}_j - \bar{\mathbf{p}})' (\mathbf{p}_j - \bar{\mathbf{p}}) = \frac{1}{n} \sum_{j=1}^{n-1} \sum_{l>j}^n \delta_{jl}.$$

Similarly we can obtain the sum of squares for each island individually and obtain SS_{BR} , SS_{CL} , SS_F and SS_{IC} .

E.g.

$$SS_{BR} = \sum_{j=1}^{n_{BR}} (\mathbf{p}_j^{BR} - \bar{\mathbf{p}}^{BR})' (\mathbf{p}_j^{BR} - \bar{\mathbf{p}}^{BR}) = \frac{1}{n_{BR}} \sum_{j=1}^{n_{BR}-1} \sum_{l>j}^{n_{BR}} \delta_{jl}^{BR}.$$

The sum of squares within islands, SS_{WI} , is defined as

$$SS_{WI} = SS_{BR} + SS_{CL} + SS_F + SS_{IC}.$$

The sum of squares among islands, SS_{AI} , is defined as

$$SS_{AI} = SS_{\text{total}} - SS_{WI}.$$

The variance σ_w^2 for individuals within islands is estimated as [1, 3, 7, 23]:

$$SS_{WI}/(n - k) \tag{19}$$

and the variance σ_a^2 for among regions variance is estimated as [1, 3, 7, 23]:

$$(SS_{AI}/(k - 1) - SS_{WI}/(n - k)) / n^* \tag{20}$$

where n^* is an average island sample size calculated using a formula [1, 3, 7, 23]:

$$n^* = \frac{n - \frac{n_{BR}^2 + n_{CL}^2 + n_F^2 + n_{IC}^2}{n}}{k - 1}.$$

The total variance σ^2 in the whole population is equal to $\sigma^2 = \sigma_w^2 + \sigma_a^2$.

Next we introduce a ϕ -statistic which compares the variance among islands and overall variance in the whole population. The ϕ -statistic is defined as $\phi = \frac{\sigma_a^2}{\sigma^2} = \frac{\sigma_a^2}{\sigma_w^2 + \sigma_a^2}$ and is calculated using (19) and (20).

⁸Note that $n = 123$, $n_{BR} = 40$, $n_{CL} = 37$, $n_F = 13$ and $n_{IC} = 33$; here $k = 4$ denotes 4 island populations.

4.2.2 Results of AMOVA

Software GenALEX was used to calculate the AMOVA [16]. The results are as follows:

Within sum of squares for Bass Rock	= 358.950
Within sum of squares for Inchcolm	= 126.757
Within sum of squares for Craigeleith	= 187.818
Within sum of squares for Fidra	= 109.231

Altogether the within sum of squares SS_{WI} is therefore equal to 782.756. Total sum of squares SS_{total} is equal to 907.252. Using SS_{total} and SS_{WI} we can calculate $SS_{AI} = 124.496$ and $n^* = \frac{123 - (40^2 + 37^2 + 33^2 + 13^2)/123}{3} = 29.54$. Results are summarised in Table 6:

Table 6: Summary statistics from the AMOVA test on genetic distance. Here df stands for degree of freedom, SS stands for sum of squares and MS is calculated as $\frac{SS}{df}$.

	df	SS	MS	Estim. σ^2	ϕ	p
Among Pops.	3	124.496	41.499	1.182		
Within Pops.	119	782.756	6.578	6.578	0.15	0.01

The ϕ -statistic indicates that 85% of the total genetic variation can be explained within island populations. This indicates that the islands are genetically similar and therefore that they have originated from a common source. The fact that 15% of the total variation occur between populations indicates that the dispersal of *L. arborea* allows differentiation between islands. It could be that a single island is acting as the source and dispersal to other islands is ‘rare’ to the extent that not all variation is transmitted to these islands.

It is important to test whether $\phi=0.15$ is significant. To do this we employ the method of permutational analysis [7]. This is achieved by randomly permuting components of the distance matrix. The procedure outlined above is then used to determine the value of ϕ for the random data. We denote ϕ obtained for random data as ϕ_{random} . We do N permutations and then determine the proportion p_{random} exceeding the value achieved for the initial data ϕ_{data} . We did $N = 100$ permutations. In Table 5 we see $p = 0.01$ indicating that ϕ_{data} was only achieved or exceeded once in 100 permutations⁹. This

⁹In these calculations our initial data set is included as one of the conducted permutations. (Only 99 new permutations are created, as our data set is one of 100 permutations.)

indicates that the value of $\phi = 0.15$ is significant at a 1% significance level.

A way to test the relatedness of different island populations is to work out the ϕ -statistic for subsets of the data for combinations of pairs of islands (Table 7).

Table 7: Pairwise Genetic Distances based on ϕ calculations

	BR	IC	CL	F
BR	0.000			
IC	0.075	0.000		
CL	0.058	0.024	0.000	
F	0.077	0.181	0.127	0.000

In terms of genetic distance, the data show that Bass Rock is very similar to all islands. Craigleith and Fidra have more among populations variability. In particular Craigleith and Fidra are more similar to Bass Rock than to each other. When we look at the variation between Inchcolm and Fidra with Craigleith, the data show that Inchcolm is more similar to Craigleith than to Fidra.

This suggests that the Bass Rock population was the founder population. It also indicates that the population on Inchcolm may have originated from the Craigleith population.

4.2.3 Discussion of AMOVA

The AMOVA test indicates that 85% of the total genetic variation can be explained within island populations and 15% between populations. This produces evidence that the populations were originally from a common source. The pairwise analysis indicates that the Bass Rock population is the founder population and was the origin of populations on Craigleith and Fidra and that Craigleith may be the source for the population on Inchcolm.

5 Discussion

The objective of this project was to understand the spread of an invasive species, *L. arborea*, on the Islands of the Firth of Forth and its interaction with native species on these islands. To achieve this we have used mathematical modelling techniques and genetic analysis on samples of *L. arborea*.

The mathematical modelling framework was designed to understand the interaction between *L. arborea*, puffins and a herbivore of *L. arborea*. The key objectives were to understand the mechanisms that lead to *L. arborea* establishment and puffin decline/extinction. In the absence of the herbivore, *L. arborea* would force the puffin population to extinction. Introducing the herbivore on the islands where puffins are suppressed by expanding *L. arborea* would help to reduce *L. arborea* density and, providing the herbivore itself would not outcompete puffins, would lead to species coexistence with puffins preserved. This suggests a possible conservation strategy to preserve the declining number of puffins on islands where *L. arborea* has invaded: introducing a suitable herbivore to control *L. arborea* expansion.

The empirical work and analysis of genetic data was undertaken on samples of *L. arborea* collected from Islands of the Firth of Forth. The objective was to use the results from analysis of DNA variation to understand the dispersal route of *L. arborea*. This analysis indicated that the majority of the total genetic variation could be explained by variation within island populations and that a smaller percentage (15%) resulted from difference between the populations. Results from examining the proportion of heterozygosity within island populations also indicated that although the differences between populations were small, they were significant. In terms of genetical distance, the fact that different island populations are similar indicates that *L. arborea* must be able to disperse sufficiently successfully. By analysing the data which highlight the amount of variation within and between pairs of island populations it is possible to infer that the Bass Rock population is the source population for *L. arborea* dispersal. Dispersal events from Bass Rock are likely to have established separate populations on Craigleith and Fidra, and also Craigleith is the likely source of the Inchcolm population. This is clearly what we would have expected from historical information which indicates population of *L. arborea* on Bass Rock well in advance of populations on other islands. It also is what would have been expected due to the geographical location of the islands (Figure 2).

The genetic analysis allows us to speculate on the likely success of *L. ar-*

borea in invading other islands of the Firth of Forth and its ability to survive on those islands which it currently occupies. The fact that populations are genetically similar indicates that dispersal from established populations is not a 'rare' event. If there were few dispersal events we would expect the genetic diversity on the recently colonised islands to be low and therefore the variation between populations to be high. Variation does diminish with distance from the source population on Bass Rock but dispersal is sufficient to maintain most of the diversity. This indicates that even if the populations on certain islands were eradicated it is likely that these islands would be recolonised. The small (but significant) loss of diversity with distance from the source population may however confer a fitness loss for the recently colonised populations. A lack of genetic diversity has been linked with the inability of an individual to survive harsh environmental conditions. This means that the less diverse populations are more susceptible to harsh winters, which is known to be a key factor in the survival of *L. arborea*. To summarise, even if the newly established populations were eradicated or reduced by hard winter in a particular year, it is likely that the islands will be recolonised by direct or secondary dispersal from the source population on Bass Rock.

References

- [1] Balding D. J., Bishop M., Cannings C., (2001) *Handbook of Statistical Genetics*, John Wiley & Sons, Ltd.
- [2] Barrow M., (1996) *Statistics for Economics, Accounting and Business Studies*, Longman Group UK Limited.
- [3] Bishop G. R., (1996) Statistics Department, University of Adelaide, the article on the website
<http://www.bioss.ac.uk/smart/unix/mamova/slides/frames.htm>
- [4] Brown T. A., (2002) *Genomes*, BIOS Scientific Publishers Ltd.
- [5] CEH (2005) Relationship between tree mallow (*Lavatera arborea*) and Atlantic puffin (*Fratercula arctica*) on the island of Craigleith, Firth of Forth (Forth Islands Special Protection Area). *Scottish Natural Heritage*, Commissioned Report No. 106.
- [6] CEH web site
<http://www.ceh.ac.uk/treemallow/what/index.html>
- [7] Excoffier L., Smouse P. E., Quattro J. M., (1992) Analysis of Molecular Variance Inferred From Metric Distances Among DNA Haplotypes: Application to Human Mitochondrial DNA Restriction Data. *Genetics*, **131**, 479-491.
- [8] Harley E. H., (2005) DNA Fingerprinting in Ecology. *Encyclopedia of Life Sciences*, John Wiley & Sons, doi: 10.1038/npg.els.0005454.
- [9] Harrison R. M., (2005) Variation, Within Species: Introduction. *Encyclopedia of Life Sciences*, John Wiley & Sons, doi: 10.1038/npg.els.0004161.
- [10] Hartl D. L., Clark A. G., (1997) *Principles of Population Genetics*, Sinauer Associates, Inc.
- [11] Meudt H. M., Clarke A. C., (2007) Almost Forgotten or Latest Practice? AFLP applications, analyses and advances. *Trends in Plant Science*, **12**, 106-117.
- [12] Miller I., Miller M., (2004) *John E. Freud's Mathematical Statistics with Applications*, Pearson.

- [13] Mueller U. G., Wolfenbarger L. L., (1999) AFLP genotyping and fingerprinting. *Trends in Ecology and Evolution*, **14**, 389-394.
- [14] Murray J.D., (1989) *Mathematical Biology*, Spriner-Verlag Berlin.
- [15] Nasa World Wind software available on the website
<http://worldwind.arc.nasa.gov>
- [16] Peakall R., Smouse P.E., (2006) GENALEX6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, **6**, 228-295.
- [17] Photo gallery of *Lavatera arborea* on the website
<http://www.dipbot.unict.it/orto/0259-1.html>
- [18] Photo gallery of puffins on the website
<http://puffin.gallery.sytes.org>
- [19] United States National Arboretum web site
<http://www.usna.usda.gov/Gardens/invasives.html>
- [20] University of Cambridge, the F-statistic table on the website
<http://thesaurus.maths.org/mmkb/entry.html?action=entryById&id=1446>
- [21] Van Der Wal R., (2006) The management of tree mallow and puffin habitat on Craigleith: a first proposal. *Scottish Natural Heritage*, CEH.
- [22] Vos P., Hogers R., Bleeker M. *et al*, (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research*, **23**, 4407-4414.
- [23] Wener P., article on the website
<http://userwww.sfsu.edu/efc/classes/biol710/amova/amova.htm>