Laboratory Exercise – Migration / Selection equilibria in Lake Erie Water Snakes (*Nerodia*)

        Natural Selection on allelic variation in a population that receives migrants from another population with different allele frequencies may achieve an equilibrium (Δq = 0), if the rate at which natural selection eliminates an unfavourable allele in the target population is counter-balanced by the influx of alleles from a source population where that same allele is favoured by selection. In the classic model, an Island population exists where the '**AA**' phenotype has an adaptive superiority to the '**BB**' phenotype, such that selection will reduce f(**B**) → 0.0. The Island receives migrants from a Mainland population at a rate **m**, where **m** is defined as the fraction of individuals in the island population newly arrived from the mainland each generation. The selection scheme on the Mainland is the reverse of that on the Island, such that '**BB**' is favoured over '**AA**' and f(**B**) ≈ 1.0. Where **m < s**, f(**B**) remains **> 0.0** as the gain of f(**B**) from migration (**m**) is over-balanced its loss by selection (**s**). Where **m ≈ s**, f(**B**) may reach an 'equilibrium' where Δq ≈ 0. Where **m > s**, f($\mathbf{B_I}$) → 0.0 as the island population becomes identical to that on the Mainland, despite the difference in adaptive environments.

        In a classic study of ecological genetics, Ehrlich & Raven (1958) studied differences in the relatives frequencies of banded and unbanded phenotypes of the Lake Erie Water Snake (*Nerodia* [*Natrix*] *sipedon insularum*). Islands in the middle of the lake consist of dark grey slate rock, on which unbanded snakes are cryptic to their avian predators. The Ohio and Ontario mainlands on the southern and northern shores of the lake are sandy, on which banded snakes are more cryptic. The mainland populations are essentially fixed for the banded phenotype, which may be attributed to a homozygous **BB** genotype. The island populations show high proportions of the unbanded phenotype, attributed to a homozygous **AA** genotype. Intermediate, partially banded phenotypes are attributed to the **AB** genotype. Of the three major island groups, Kelly's Island is closest to the Ohio shore, the Bass Island complex more distant, and Pelee Island rather distant from both the Bass Islands and the Ontario shores. Ehrlich & Raven (1958) argued that the fraction of unbanded **AA** phenotypes on any island was proportional to its distance from the mainland and hence the rate of migration **m** of **BB** phenotypes, moderated by a '*stepping stone*' effect from other, nearer islands with differing f(**B**) (**Figure 1**).

Model

        In this lab, we will examine the Lake Erie ecosystem with a modified Stepping-Stone Model (**Figure 2**), so as to determine how patterns of genetic variation **q** are dependent on **m** and **s**. In a two-allele model, **A** is genetically dominant to **B**; this may be modified to a **semi-dominant** model by choice of $\mathbf{t_1}$ & $\mathbf{t_2}$ selection coefficients (below). A **Source** population **S** has an initial f(**B**) = $\mathbf{q_s}$ ≈ 1.0 linked to four successive **islands 1**, **2**, **3**, & **4** with initial f(**B**) $\mathbf{q_1}$, $\mathbf{q_2}$, $\mathbf{q_3}$, & $\mathbf{q_4}$. Migration rates **m** between populations are $\mathbf{m_s}$ → $\mathbf{m_1}$ → $\mathbf{m_2}$ → $\mathbf{m_3}$ → $\mathbf{m_4}$, where $\mathbf{m_X}$ is the rate *from* population **X** *to* the next in the chain. $\mathbf{m_4}$ is the return rate to $\mathbf{m_1}$, such that the series is *linear* or *circular*, depending on $\mathbf{m_4}$ = 0 or **> 0**. Selection operates on f(**B**) on $\mathbf{q_s}$ in the **Source** population, as calculated in a separate GSM worksheet for **s** = 0.000, 0.001, 0.010, and 0.050. Typically, $\mathbf{q_s}$ = 0.95 , and $\mathbf{q_1}$ = $\mathbf{q_2}$ = $\mathbf{q_3}$ = $\mathbf{q_4}$ = 0.0, which simulates the classic model of an allele initially fixed on the mainland and absent from the islands. Alternatively, values of **q** may be set to those calculated from the data in Camin & Ehrlich (1958). Thus in the model, the expected island **q** = 0 is disrupted by long-term migration from the mainland $\mathbf{q_s}$ ~ 1, which declines over time due to selection against **B**. Island **q** varies despite any selective advantage. We ask, what are the short-, middle-, and long-term consequences?
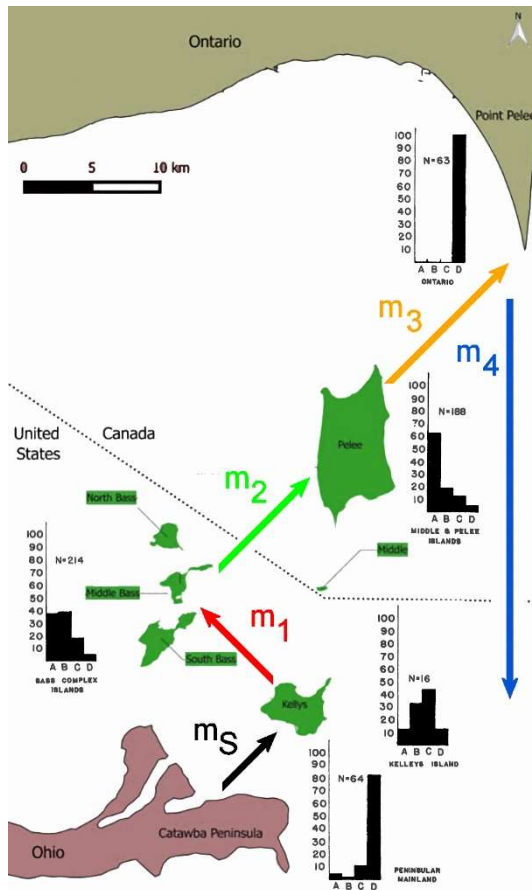
Figure 1 – **Distribution of *Nerodia sipedon insularum* in Lake Erie**. Minor islands not shown. Observed phenotype frequencies from Camin & Ehrlich (1958): Category "**A**" unbanded, Category "**D**" banded, Categories **B** & **C** intemediate. Unidirectional migration rates **m** between mainlands and islands occurs as shown, color coded as in Figure 2.

Figure 2 – **Schematic of migration model**. Initial frequencies for f(**B**) ['Banded' allele] in the Source ['Mainland'] and four island populations as shown. Note than m_x is the migration rate <u>from</u> population '**X**' <u>to</u> population '**X+1**' with initial q_{x+1}. m_4 is the return rate from the fourth to the first population: if $m_4 > 0$, the model is circular, if $m_4 = 0$ the model is linear terminating in Population 4.
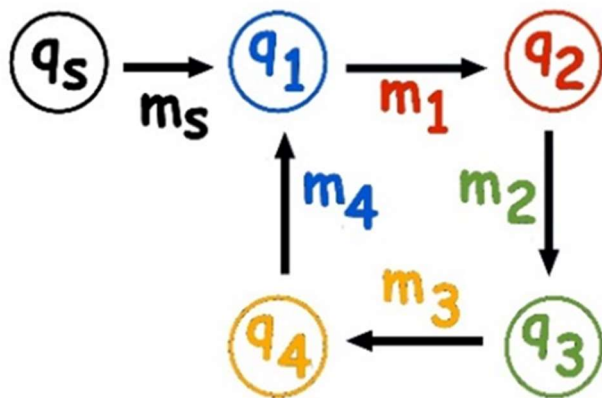
The Migration / Selection Excel worksheet contains three sections. The left-hand section calculates **Δq** across populations accoring to the parameters in the center section. This sections also shows the graphical results (see below), reference schematics of the migration model, and a series of migration scenarios.

s	qS	mS	q1	m1	q2	m2	q3	m3	q4	m4
	Ohio		Kelly's		Bass		Pelee		Ontario	[Kelly]
E&R Data	0.95	0.010000	0.75	0.001000	0.50	0.001000	0.25	0.000100	0.95	0.000010

Set up **Migration Scenarios** from **4 Models** below; Copy & Paste **VALUES ONLY** into grey cells above										
Model1	0.95	1.000000	0.95	0.000100	0.00	0.000000	0.00		0.95	
Model2	0.95		0.00		0.00		0.00		0.95	
Model3	0.95		0.00		0.00		0.00		0.95	0.000000
E&R Data	0.95		0.75		0.50		0.25		0.95	

The basic **Model 2** has the initial **f(B) = q** for the Banded allele almost fixed in the mainland populations [if **q = 1**, selection could not act] and absent in the four island populations, in a *circular* model with back migration from Ontario to Kelly's Island [set this as a very low rate]. **Model 3** is the same as Model 2, except there is no migration from Ontario (m_4 **= 0**), so the model is *linear* ending on the Ontario mainland. **Model 1** is the Classical model in a slightly different form. **Kelly's Is** is identical to the **Mainland** because **m = 1** and identical initial $q_S = q_1$ = **0.95**, in contrast with **Bass Is** with q_2 **= 0.0**. Selection on the mainland and **Kelly's Is** drives $q_S = q_1$ → **0.0**, whereas migration to **Bass Is** raises q_2 → **1.0** depending on m_1, despite absence of a selective advantage. Because m_2 **= 0.0** terminates migration, **Model 1** comprises only two populations, with different selection regimes in each. The **E&R data** set initial values of **q** to their observed values, so as to show how modifying **m** causes them to depart from initial values.
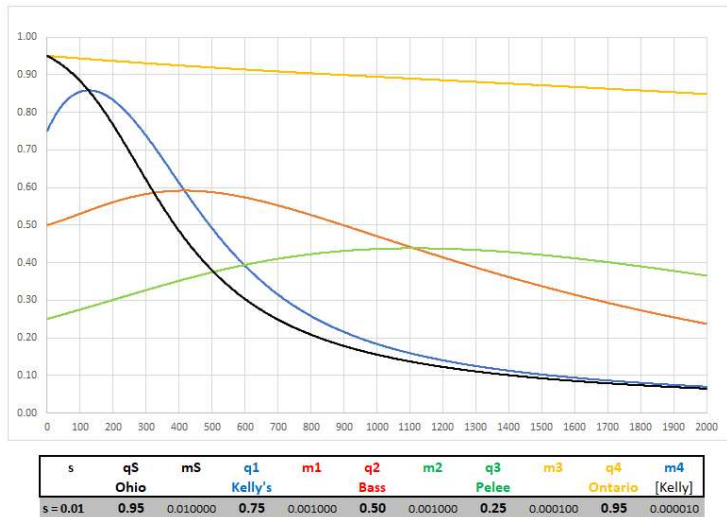
Set values of **m** according to an *a priori* migration model [examples in Excel worksheet], and in combination with the table of **Δq** under selection **s** (right-hand section). Consider (1) **m decreases** by orders of magnitude as **relative distances** increase among populations [an **inverse-log₁₀** relationship], (2) **m** is **constant**, and results depend on choice of the selection scheme **s**, (3) **combinations** of (1) & (2) where both **s** and the migration gradient values of **m** to vary by orders of magnitude.

Choose a column of **Δq** under selection **s**, in the third section. Try **s = 0.01** and **s = 0.1** first, then **s = 0.001** and **s = 0.5**. Recall from the theoretical discussion that a migration / selection equilibrium occurs when **s ≈ m**. *A good starting point is* **s = 0.01** *and a uniform* **m = 0.01** *among all populations*.

Set-up of Model Scenarios

1) Compose values of **q** and **m** for the scenario of interest in **one line** of the work area. Include the value of **s** used, in the first column. [Retain all scenarios run, so they can be re-run as necessary].
2) Copy & Paste **Values Only** into boxes **O33 – Y33**: these are then automatically copied into boxes **C2 – M2** in the first section.
3) Choose the column of **Δq** for the chosen value of **s**. Copy & paste **Values Only** into **Column AA**: these are then automatically copied into **Column L** starting at **Row 6** in the first section. [**Be careful**: overwriting the first five rows of Column L by simple pasting will destroy the worksheet!].

4) The values in the first section will then be re-calculated for the input values of **s**, **q**, and **m**. The graphical result of **q** for each population over time will change to something like this:

s	qS	mS	q1	m1	q2	m2	q3	m3	q4	m4
	Ohio		Kelly's		Bass		Pelee		Ontario	[Kelly]
s = 0.01	0.95	0.010000	0.75	0.001000	0.50	0.001000	0.25	0.000100	0.95	0.000010

This scenario uses the initial data from Camin & Ehrlich, with **s = 0.01** and migration rates starting at **m = s = 0.01** and decreasing an order of magnitude with each doubling of distance [an **inverse-log$_{10}$** realtionship]. Kelly's Is, the closest to Ohio, quickly approximates the mainland. Both Bass and Pelee remain at relatively low f(**B**), with **Pelee** crossing Bass ~1,100 generations to a higher f(**B**) (**Why**?). The Ontario mainland retains a high f(**B**). Notes patterns at 100~500, 1,000, & 2,000 generations.

5) **Copy and paste the Excel block N36 – Z34** into a drawing program to retain the input parameters and the result. Repeat for all scenarios of interest. Use these figures for the Discussion.

Discussion & Interpretation

Consider the following scenarios for interpretation and discussion. Broadly, combine **s = 0.00**, **0.01**, & **0.001** *versus* **m = 0.0001**, **0.0010**, & **0.0100** [except consider **m < 0.00001** as necessary]. Consider short-, middle-, and long-term results (ca. 100, 500, and 2,000 generations, respectively).

1. **Five-population model**:
    a. For a **constant m** among populations, how does varying **s** by plus-or-minus an order of magnitude modify the short-, middle, and long-term results?
    b. Same question, for a **declining m proportional** [inverse – log$_{10}$] to distance among population?
2. **Mainland / Island model**: For the [**Mainland** + Kellys] → Bass **model**, compare **s = 0.00**, **0.01**, & **0.001**
    a. **m > s**
    b. **m ~ s**
    c. **m < or << s**
3. Under what circumstances do island **q** values **converge** on the mainland **q**? Why?
4. What values of **m** & **s** best **approximate** the observed data of **Ehrlich & Raven (1958)** [Figure 1]
    a. Calculate f(**B**) counts for each population, approximating frequencies to the closest 10%.
    b. Calculate **Chi-Square** for *observed* and *expected* genotype counts. Which populations show the greatest departure from **Hardy-Weinberg expectations**? Why do you think so?
5. Under what (if any) combinations of **m** & **s** can an approximate migration / selection **equilibrium** be reached and (or) maintained? Why?