

Group B Simulation Analysis Report

Andrew Meng, Kathi Munoz, Joseph Zou

A population history tree for our simulation can be found in Figure 1.0. Note that for two of the species in our simulation (Wookies and Humans) there are ‘intermediate’ populations. This results from what could be thought of as the leftover population from a speciation split in our simulation. Since they don’t experience the same mutations as the ‘main’ populations do for their species, they are sort of background noise or more a look at what that species was like farther back in history. We didn’t intend for them to be a focus, but for the sake of all-inclusive analysis, they’re included.

I. Pi for each population:

For each population, both an average pi as well as pi in windows of size 5,000 bp was estimated. For our data (genome size of 100,000) this means 20 windows for each.

- Common Ancestor Population:
Avg Pi: 0.00688
Pi in each window:
[0.0018, 0.0132, 0.005, 0.0086, 0.0074, 0.005, 0.0032, 0.0, 0.0, 0.0074, 0.0148, 0.0092, 0.0032, 0.0124, 0.0054, 0.0148, 0.0032, 0.0152, 0.0042, 0.0036]
- Crolute Population:
Avg Pi: 0.00835
Pi in each window:
[0.0096, 0.0108, 0.0114, 0.0194, 0.0138, 0.0048, 0.0048, 0.008, 0.0066, 0.0, 0.0156, 0.0018, 0.008, 0.0042, 0.0132, 0.0078, 0.0084, 0.006, 0.0018, 0.011]
- Abednedo Population:
Avg Pi: 0.0065
Pi in each window:
[0.0036, 0.0114, 0.0156, 0.0032, 0.0032, 0.0118, 0.005, 0.0082, 0.0032, 0.0032, 0.0128, 0.0074, 0.0114, 0.0, 0.0086, 0.0064, 0.0054, 0.0032, 0.0032, 0.0032]
- Ewok Population:
Avg Pi: 0.02187

Pi in each window:

[0.0118, 0.0486, 0.1102, 0.1118, 0.06, 0.035, 0.0186, 0.01, 0.0036, 0.0018, 0.0, 0.0032, 0.005, 0.0, 0.0042, 0.0, 0.0018, 0.005, 0.0, 0.0068]

- Gungan Population:

Avg Pi: 0.00946

Pi in each window:

[0.0068, 0.0066, 0.005, 0.0098, 0.0068, 0.0236, 0.01, 0.015, 0.0018, 0.0154, 0.0018, 0.01, 0.018, 0.0, 0.005, 0.0218, 0.005, 0.0068, 0.005, 0.015]

- Main Human Population:

Avg Pi: 0.01812

Pi in each window:

[0.0228, 0.0746, 0.026, 0.0146, 0.0096, 0.0092, 0.0164, 0.0054, 0.0126, 0.0296, 0.045, 0.0036, 0.0068, 0.0146, 0.0116, 0.0096, 0.005, 0.0086, 0.0136, 0.0232]

- Wookies Population:

Avg Pi: 0.00182

Pi in each window:

[0.0, 0.0, 0.0074, 0.0, 0.005, 0.0, 0.0018, 0.0032, 0.0, 0.0068, 0.0068, 0.0, 0.0, 0.0, 0.0, 0.0018, 0.0, 0.0036, 0.0, 0.0]

- Intermediate Wookies Population:

Avg Pi: 0.00273

Pi in each window:

[0.0018, 0.006, 0.0042, 0.0, 0.0018, 0.0054, 0.0036, 0.0036, 0.0, 0.0018, 0.0036, 0.0018, 0.0018, 0.0, 0.0048, 0.0, 0.0018, 0.0036, 0.0018, 0.0072]

- Intermediate Humans Population:

Avg Pi: 0.00675

Pi in each window:

[0.0074, 0.0018, 0.0068, 0.0036, 0.0018, 0.0054, 0.005, 0.0086, 0.0198, 0.0142, 0.0064, 0.006, 0.0018, 0.0, 0.0106, 0.013, 0.006, 0.0114, 0.0, 0.0054]

- Twi'leks Population:
Avg P_i : 0.04593
Pi in each window:
[0.0126, 0.0398, 0.0204, 0.0084, 0.0308, 0.0326, 0.0036, 0.0096, 0.0084, 0.0174, 0.0624, 0.0406, 0.005, 0.0284, 0.098, 0.215, 0.0662, 0.111, 0.0602, 0.0482]
- Mon Calamari:
Avg P_i : 0.00674
Pi in each window:
[0.0054, 0.0, 0.0018, 0.0096, 0.006, 0.0, 0.0078, 0.0032, 0.0042, 0.0142, 0.0, 0.005, 0.0188, 0.0068, 0.0082, 0.0092, 0.0092, 0.005, 0.0102, 0.0102]

We find that often locations where we would expect low polymorphism (genes under strong selection, where we would expect a sweep) exhibit high polymorphism instead. The potential reasons for this are discussed at length in Section IV. Ewoks do display a very interesting trend in polymorphism as seen in Figure 1.4. Almost logarithmic, there is a steep peak and slope down starting at around 15,000 bp. This suggests there is a rather specific area where the Ewok genome exhibits diversity. The reason for this is unclear, as there are genes under selection relatively unique to Ewoks in that area, so this does follow the trend we see of high estimates of polymorphism in areas that would be expected to be sweeps.

II. Effective Population Size (N_e) for each Population

To calculate N_e , we use the concept that, in theory, $4N_e\mu$ should be equal to both s (segregating sites) and p_i (polymorphism). Of course, this is not so realistic, so we calculate N_e with both s and p_i .

Using s :

- Common Ancestor Population: 13375
- Crolute Population: 12200
- Abednedo Population: 15100
- Ewok Population: 18800
- Gungan Population: 8625
- Main Human Population: 66525
- Wookies Population: 10225
- Intermediate Wookies Population: 14725
- Intermediate Humans Population: 14150

- Twi'leks: 30550

Using pi:

- Common Ancestor Population: 17200
- Crolute Population: 20875
- Abednedo Population: 16250
- Ewok Population: 54675
- Gungan Population: 23650
- Main Human Population: 45300
- Wookies Population: 4550
- Intermediate Wookies Population: 6825
- Intermediate Humans Population: 16875
- Twi'leks: 114825

Each of these values is clearly an overestimation in the context of our simulation. The greatest population size simulated was 10000, belonging to the humans; the next largest population was the Twi'leks, maxing out at 4000. Only two of the obtained values manage to get beneath this 10,000 mark, so it is likely that something in our simulation has interfered with an implicit assumption of the calculation. It does not seem that either s or π is more or less effective in obtaining the true population size, so the broken assumption is most likely independent of the difference between computations involving s and π . Complicating factors that could reduce the efficacy of our approach include (but are not necessarily limited to): abrupt changes in population size, selection, migration (albeit in only one case), and the length of the simulation.

III. dN/dS Plots for each Population:

Here, we calculate dN/dS ratios in windows of size 5,000 resulting in 20 windows for each. It's worth noting that some negative values are given for dn/ds; this is because ds was found to be zero in some cases, and substituting a negative value was deemed more meaningful. These values indicate a lack of synonymous mutations.

The majority of the information presented in figures 3.1-3.11 is as expected, exhibiting relatively neutral selection. An exception is the dN/dS values at window 10 in several populations (Twi'leks, Humans, Wookies, Ewoks, etc); each of these is exceptionally high and displays a population under heavy positive selection. This makes sense in the context of our simulation, as the section of the genome examined in each figure's tenth window is that in which the PAX3 gene lies. Coding for eye shape and placement, PAX3

has a positive effect on the fitness of individuals in each of these mutations, so it makes sense that they would be under positive selection at this locale.

IV. Location of Likely Sweeps in each Population:

To do this, we looked at π in windows of size 5,000 bp again, comparing them by magnitude because, intuitively, you would expect a sweep location to have lower diversity. In addition, we compared this to Group A's CLR tool to compare the possible sweep locations they determined with our polymorphism data.

- Common Ancestor Population Figure 4.1:
 π 's ordered by magnitude
 [(7, 0.0), (8, 0.0), (0, 0.0018), (6, 0.0032), (12, 0.0032), (16, 0.0032), (19, 0.0036), (18, 0.0042), (2, 0.005), (5, 0.005), (14, 0.0054), (4, 0.0074), (9, 0.0074), (3, 0.0086), (11, 0.0092), (13, 0.0124), (1, 0.0132), (10, 0.0148), (15, 0.0148), (17, 0.0152)]
- Crolute Population Figure 4.2:
 π 's ordered by magnitude
 [(9, 0.0), (11, 0.0018), (18, 0.0018), (13, 0.0042), (5, 0.0048), (6, 0.0048), (17, 0.006), (8, 0.0066), (15, 0.0078), (7, 0.008), (12, 0.008), (16, 0.0084), (0, 0.0096), (1, 0.0108), (19, 0.011), (2, 0.0114), (14, 0.0132), (4, 0.0138), (10, 0.0156), (3, 0.0194)]
- Abednedo Population Figure 4.3:
 π 's ordered by magnitude
 [(13, 0.0), (3, 0.0032), (4, 0.0032), (8, 0.0032), (9, 0.0032), (17, 0.0032), (18, 0.0032), (19, 0.0032), (0, 0.0036), (6, 0.005), (16, 0.0054), (15, 0.0064), (11, 0.0074), (7, 0.0082), (14, 0.0086), (1, 0.0114), (12, 0.0114), (5, 0.0118), (10, 0.0128), (2, 0.0156)]
- Ewok Population Figure 4.4:
 π 's ordered by magnitude
 [(10, 0.0), (13, 0.0), (15, 0.0), (18, 0.0), (9, 0.0018), (16, 0.0018), (11, 0.0032), (8, 0.0036), (14, 0.0042), (12, 0.005),

(17, 0.005), (19, 0.0068), (7, 0.01), (0, 0.0118), (6, 0.0186), (5, 0.035), (1, 0.0486), (4, 0.06), (2, 0.1102), (3, 0.1118)]

- Gungan Population Figure 4.5:

Pi's ordered by magnitude

[(13, 0.0), (8, 0.0018), (10, 0.0018), (2, 0.005), (14, 0.005), (16, 0.005), (18, 0.005), (1, 0.0066), (0, 0.0068), (4, 0.0068), (17, 0.0068), (3, 0.0098), (6, 0.01), (11, 0.01), (7, 0.015), (19, 0.015), (9, 0.0154), (12, 0.018), (15, 0.0218), (5, 0.0236)]

- Main Human Population Figure 4.6:

Pi's ordered by magnitude

[(11, 0.0036), (16, 0.005), (7, 0.0054), (12, 0.0068), (17, 0.0086), (5, 0.0092), (4, 0.0096), (15, 0.0096), (14, 0.0116), (8, 0.0126), (18, 0.0136), (3, 0.0146), (13, 0.0146), (6, 0.0164), (0, 0.0228), (19, 0.0232), (2, 0.026), (9, 0.0296), (10, 0.045), (1, 0.0746)]

- Wookies Population Figure 4.7:

Pi's ordered by magnitude

[(0, 0.0), (1, 0.0), (3, 0.0), (5, 0.0), (8, 0.0), (11, 0.0), (12, 0.0), (13, 0.0), (14, 0.0), (16, 0.0), (18, 0.0), (19, 0.0), (6, 0.0018), (15, 0.0018), (7, 0.0032), (17, 0.0036), (4, 0.005), (9, 0.0068), (10, 0.0068), (2, 0.0074)]

- Intermediate Wookies Population Figure 4.8:

Pi's ordered by magnitude

[(3, 0.0), (8, 0.0), (13, 0.0), (15, 0.0), (0, 0.0018), (4, 0.0018), (9, 0.0018), (11, 0.0018), (12, 0.0018), (16, 0.0018), (1, 0.0018), (6, 0.0036), (7, 0.0036), (10, 0.0036), (17, 0.0036), (2, 0.0042), (14, 0.0048), (5, 0.0054), (1, 0.006), (19, 0.0072)]

- Intermediate Humans Population Figure 4.9:

Pi's ordered by magnitude

[(13, 0.0), (18, 0.0), (1, 0.0018), (4, 0.0018), (12, 0.0018), (3, 0.0036), (6, 0.005), (5, 0.0054), (19, 0.0054), (11, 0.006)]

, (16, 0.006), (10, 0.0064), (2, 0.0068), (0, 0.0074), (7, 0.0086), (14, 0.0106), (17, 0.0114), (15, 0.013), (9, 0.0142), (8, 0.0198)]

- Twi'leks Population Figure 4.10:

Pi's ordered by magnitude

[(6, 0.0036), (12, 0.005), (3, 0.0084), (8, 0.0084), (7, 0.0096), (0, 0.0126), (9, 0.0174), (2, 0.0204), (13, 0.0284), (4, 0.0308), (5, 0.0326), (1, 0.0398), (11, 0.0406), (19, 0.0482), (18, 0.0602), (10, 0.0624), (16, 0.0662), (14, 0.098), (17, 0.111), (15, 0.215)]

- Mon Calamari:

Pi's ordered by magnitude

[(1, 0.0), (5, 0.0), (10, 0.0), (2, 0.0018), (7, 0.0032), (8, 0.0042), (11, 0.005), (17, 0.005), (0, 0.0054), (4, 0.006), (13, 0.0068), (6, 0.0078), (14, 0.0082), (15, 0.0092), (16, 0.0092), (3, 0.0096), (18, 0.0102), (19, 0.0102), (9, 0.0142), (12, 0.0188)]

For the most part, we find that the CLR graphs are indicating sweeps that make sense, in areas that correspond to genes that are in fact under selection, where we would expect to find a sweep. Sometimes, however, our polymorphism data in the same area/window indicates the opposite. For example, the Twi'leks have two likely sweeps in Figure 4.10. However, the estimated polymorphism in the same windows is significantly higher than most. This could indicate an error in computation of polymorphism, but we found that we could detect sweeps using our polymorphism calculation on very simple sweep cases from SLiM recipes. We think a contributing factor to these conflicting statistics is the age of the sweeps (see section V for more discussion) as well abrupt changes in population size, particularly in the Twi'lek population which experiences a decrease in size 200 generations prior to simulation output. This in addition to the complex nature/inheritance of our simulation could be affecting the estimated polymorphism here.

Additionally, we see in the CLR graphs that the PAX3 gene has likely swept through multiple populations. These graphs indicate possible sweeps in Crolutes, Humans, Wookies, and the MRCA of humans and Twi'leks (labeled as intermediate humans). PAX3 codes for eye location and size, and it has a positive impact on fitness in all of these populations, making it a good candidate for a sweep.

V. Time of some Strong Sweeps:

To calculate this, we use the equation $0.5 = 1 - e^{-rT}$ where r is the chance of recombination between sites over a certain range and T is how many generations ago the sweep occurred. We decided to estimate the time of the two sweeps that seem evident in the Twi'lek population, as seen in Figure 4.10

The sweep indicated at around roughly 75,000~80,000 bp range is estimated to have occurred around 17,300 generations ago. This is about the range of the gene *OCA2*, which was under positive selection in this population as it codes for skin color, a trait particularly important in Twi'leks. However, in the simulation, this gene was made to be under positive selection for only Twi'leks at generation 15,000. By our estimate, the gene is predicted to have begun sweeping around generation 7,700. Clearly, we're overshooting the time of the sweep.

The sweep to the right of it at roughly 87,000~90,000 bp range is estimated to have occurred around 23,000 generations ago. This is roughly the location of the gene *UBR4*, which also was simulated under positive selection in Twi'leks. According to our computation, the sweep began around generation 2,000, which is about 13,000 generations before the gene came under positive selection (in generation 15,000). In both cases, we're very much overshooting the estimated time of the sweep. This is likely because we violate the assumption inherent in this calculation that $T \ll 2N$. In this population, there is a reduction in size around generation 24,800. The recentness of this change may well be affecting the estimate due to the impacts it had on polymorphism around the genes in question.

VI. Estimates of Fst for each Population:

- Common Ancestor Population: -0.006
- Crolute Population: -0.006
- Abednedo Population: -0.007
- Ewok Population: -0.009
- Gungan Population: -0.005
- Main Human Population: -0.009
- Wookies Population: -0.003
- Intermediate Wookies Population: -0.006
- Intermediate Humans Population: -0.004
- Twi'leks Population: -0.021
- Mon Calamari Population: -0.008

These Fst estimates show most populations in the simulation exhibiting relatively similar estimates of heterozygosity, all being in the same order of magnitude. There is one distinct exception in the Twi'lek population: it's value indicates a much higher amount of heterozygosity compared to the other populations. This is likely because they had more mutations impacting their populations than the others. They had four mutations under selection, whereas most had around two.

VII. Fst in Windows across each Population:

- Common Ancestor Population Figure 7.1:
[-0.006, -0.007, -0.006, -0.003, -0.003, -0.004, -0.002, -0.006, -0.002, -0.008, -0.005, -0.008, -0.012, -0.012, -0.012, -0.003, -0.002, -0.015, -0.002, -0.002]
- Crolute Population Figure 7.2:
[-0.005, -0.008, -0.006, -0.008, -0.007, -0.002, -0.001, -0.003, -0.005, -0.006, -0.009, -0.002, -0.005, -0.001, -0.012, -0.01, -0.007, -0.005, -0.002, -0.011]
- Abednedo Population Figure 7.3:
[-0.004, -0.007, -0.007, -0.006, -0.006, -0.004, -0.008, -0.008, -0.006, -0.004, -0.011, -0.006, -0.008, -0.002, -0.012, -0.006, -0.006, -0.009, -0.004, -0.008]
- Ewok Population Figure 7.4:
[-0.006, -0.012, -0.025, -0.025, -0.018, -0.014, -0.008, -0.006, -0.004, -0.006, -0.004, -0.004, -0.004, -0.008, -0.008, -0.004, -0.008, -0.006, -0.002, -0.004]
- Gungan Population Figure 7.5:
[-0.003, -0.004, -0.002, -0.005, -0.001, -0.014, -0.003, -0.007, -0.002, -0.008, -0.002, -0.004, -0.008, -0.005, -0.006, -0.009, -0.003, -0.001, -0.006, -0.008]
- Main Human Population Figure 7.6:
[-0.004, -0.015, -0.008, -0.011, -0.011, -0.003, -0.005, -0.009, -0.006, -0.008, -0.011, -0.008, -0.007, -0.009, -0.005, -0.016, -0.011, -0.012, -0.013, -0.011]
- Wookies Population Figure 7.7:
[-0.002, -0.007, -0.004, -0.002, -0.002, -0.002, -0.0, -0.002, -0.006, -0.004, -0.004, -0.002, -0.002, -0.002, -0.002, -0.002, -0.002, -0.002, -0.002, -0.004]
- Intermediate Wookies Population Figure 7.8:

[-0.006, -0.005, -0.008, -0.002, -0.004, -0.004, -0.004, -0.006, -0.002, -0.01, -0.004, -0.002, -0.006, -0.004, -0.002, -0.006, -0.008, -0.008, -0.004, -0.017]

- Intermediate Humans Population Figure 7.9:
[-0.003, -0.004, -0.002, -0.006, -0.008, -0.002, -0.004, -0.004, -0.012, -0.012, -0.002, -0.004, -0.002, -0.002, -0.004, -0.006, -0.003, -0.004, -0.002, -0.002]
- Twi'leks Population Figure 7.10:
[-0.011, -0.019, -0.008, -0.003, -0.01, -0.01, -0.012, -0.008, -0.008, -0.009, -0.025, -0.025, -0.007, -0.011, -0.041, -0.086, -0.034, -0.056, -0.018, -0.019]
- Mon Calamari Population:
[-0.006, -0.004, -0.004, -0.009, -0.009, -0.004, -0.005, -0.004, -0.005, -0.013, -0.006, -0.008, -0.015, -0.011, -0.008, -0.01
5, -0.005, -0.005, -0.007, -0.011]

For the most part, F_{st} across the genome appears mostly consistent for the majority of the populations. However, in the Ewok population in Figure 7.4 we see a rather sharp dip around the 5,000-30,000 bp range. This encompasses two of the genes under selection in this population: *ANGPT2* and *HR*, coding for respiration and fur, respectively. While *ANGPT2* affects most of the populations in the simulation, *HR* uniquely affects Ewoks and Wookies. Interestingly, we don't see the same dip in F_{st} for Wookies in that area. This suggests that it had a greater impact on the heterozygosity of the Ewok population than the Wookies. A reason for this might lie in the next gene over, *FBN1*, which also impacts both Wookies and Ewoks, but under different kinds of selection. This could be affecting the heterozygosity in this area for the two populations as well. However, there is a fairly long intergenic region between these two genes (20,000 bp of a 100,000 bp total) making this seem less likely.

Additionally, Figure 7.10 reveals a sharp decrease in F_{st} around the 65,000-85,000 bp range. This is in keeping with our findings from Section VI where we saw Twi'leks exhibited a much lower F_{st} on average. The location of the dip supports our postulation that the *OCA2* gene is heavily impacting the heterozygosity of this population as it includes the location of this gene.

VIII. Structure Analysis of all Data:

Likelihood of each quantity of populations is given in the table to the right, where K represents the number of populations and a smaller value of $\text{Ln } P(D)$ represents a more likely outcome. Despite being reasonably confident in all of its estimations based in all of its estimations, Structure clearly does not have an idea of which number of populations is most likely in a union of samples of size 10 from each of our 12 subpopulations. A few bar graphs, with each sorted according to the original ordering in the file (so subpopulations are grouped together but not necessarily colored consistently), are given in the appendix of figures (Section XI).

K	$\text{Ln } P(D)$	$\text{Var}[\text{Ln}P(D)]$
8	-37110.1	1030.6
9	-37682.0	2419.7
10	-37325.9	1520.3
11	-32096.7	3896.4
12	-32370.6	1180.7
13	-35810.4	1412.4
14	-29775.3	756.6
15	-30815.7	1182.7

Our conclusion is that the volume of complications in the simulation, such as a large, complex tree, migration, selection, mutations with variable effects, and rapid or instantaneous shrinkages and expansions, has inundated Structure with more background noise than it can handle. Therefore, we reason that the graphs output (figures 8.3, 8.4, and 8.5) are almost entirely without significance.

IX. Structure Analysis of Some Individual Populations:

In our simulation, one group of humans left the population of humans, and there was a small amount of migration between the two subpopulations. We used Structure to analyze the composition of the two subpopulations. The values of $\text{Ln } P(D)$ are displayed to the right:

K	$\text{Ln } P(D)$	$\text{Var}[\text{Ln}P(D)]$
1	-1965.3	123.5
2	-7901.1	12241.9
3	-11103.7	18653.6
4	-1987.4	184.7
5	-1965.7	135.2

Structure concludes that K is probably not 1, 4, or 5, but it clearly is not confident in giving K=2 or K=3. It seems as though Structure is tempted to falsely recognize an intermediate between subpopulations as a true subpopulation.

It seems as though the 20 individuals sampled (10 from each subpopulation) have some ancestry from each group, based on the K=2 graph (Figure 9.1). The K=3 graph (Figure 9.2) seems to be less meaningful than the former, but it is clear that some individuals have more ancestry from one perceived 'group' than others in this case. Finally, as observed in the K=5 bar graph (Figure 9.3), the only reasonable conclusion is that K does not in fact equal 5.

Ewoks: Our simulation considered only one population of Ewoks, involving subpopulations or migration. As a result, one may be tempted to conclude that a Structure analysis of a sample of Ewoks may be uninteresting. However, Structure seems to be very confident that there is more than one subpopulation, as indicated by the probability distribution of K, described to the right. Structure is very confident that K does not equal 1, 4, or 5, but, as was the case with humans, is not confident whether K=2 (Figure 9.4) or K=3 (Figure 9.5).

K	Ln P(D)	Var[LnP(D)]
1	-1216.4	89.9
2	-24047.4	46442.2
3	-21750.9	41735.7
4	-1246.0	157.0
5	-1223.4	121.5

Using the K=2 graph (Figure 9.4), individuals can be fairly easily categorized into two groups with almost exactly the same ratios of ancestry. This may indicate that there was some sort of unforeseen split within the Ewok population.

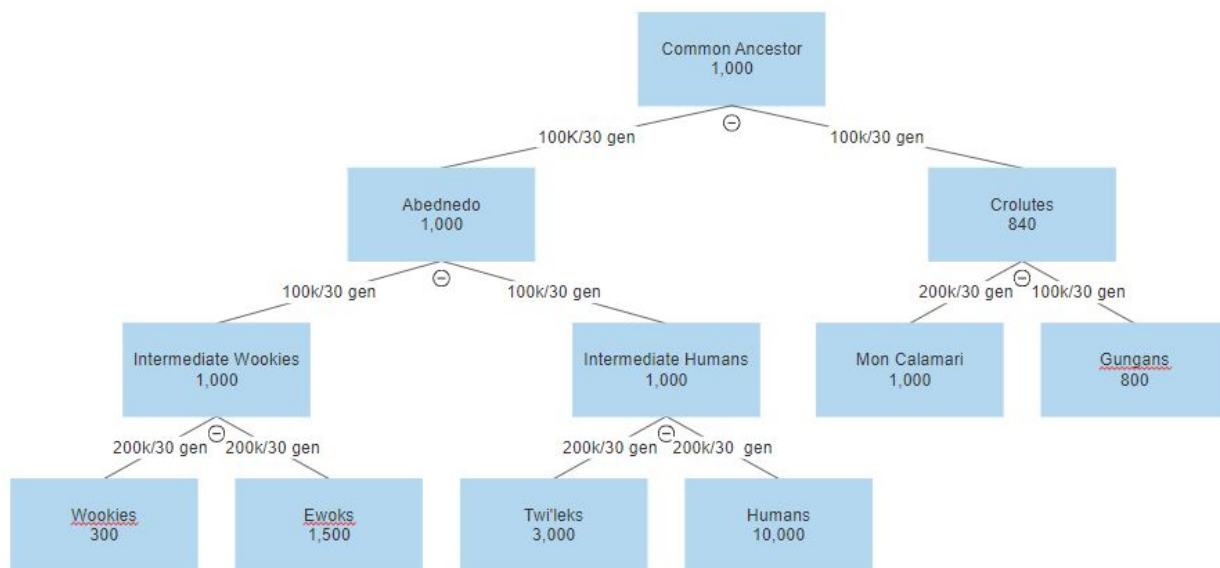
X. Estimates of Migration Rates for Migrant Populations

Migration rate for p60: 0.002114706978943864

We simulated only one population, humans, which experienced migration. We found the rate of migration to be $m=0.00211$, which is off from the anticipated value (0.01) by a factor of about 5. This computation involves numerous inherent assumptions, almost none of which are fulfilled. These include the size of both populations (both are fairly small but are supposed to be large), that there is no mutation, and that, since there is no mutation, there is no selection. All of these assumptions are violated in our model, which has led us to be pleasantly surprised by how close our estimation came to the true value.

XI. Appendix of Figures

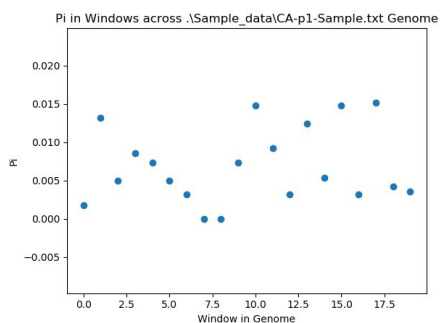
Population History Tree **Figure 1.0:**



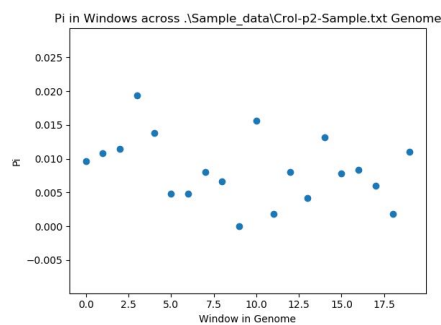
Section I - Pi for each population

Here, figures depict estimates of Pi in 5,000 bp windows across the genome for a given population.

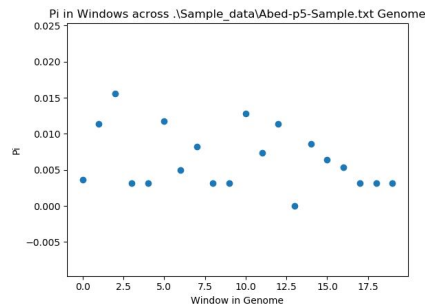
Common Ancestor **Figure 1.1:**



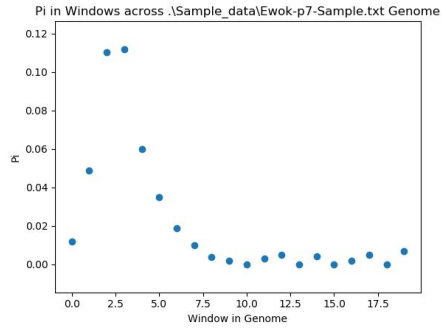
Crolutes **Figure 1.2:**



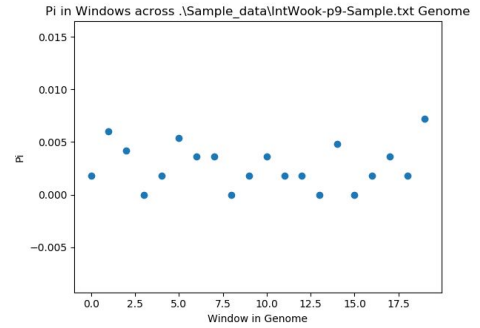
Abednedo **Figure 1.3:**



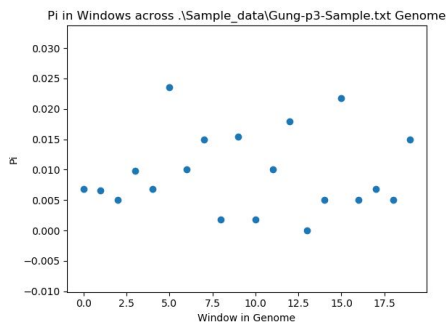
Ewoks **Figure 1.4:**



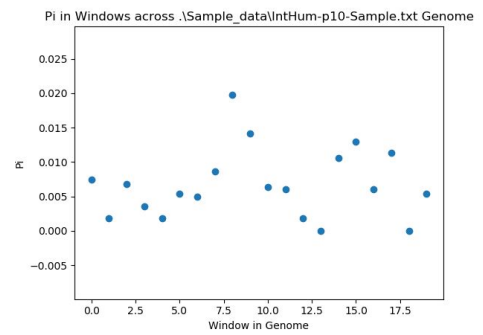
Intermediate Wookies **Figure 1.8:**



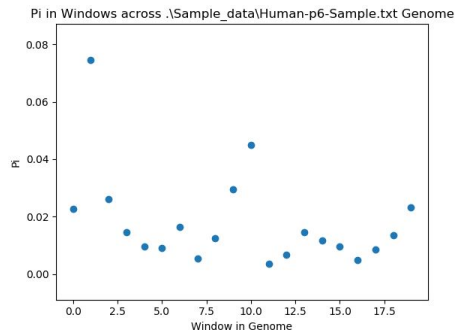
Gungans **Figure 1.5:**



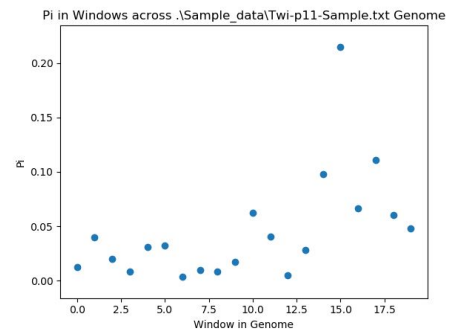
Intermediate Humans **Figure 1.9:**



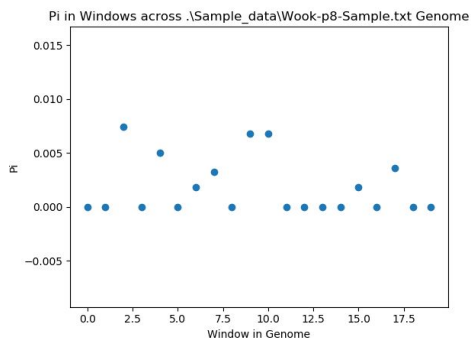
Humans **Figure 1.6:**



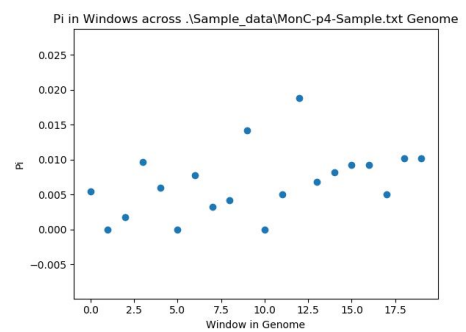
Twileks **Figure 1.10:**



Wookies **Figure 1.7:**



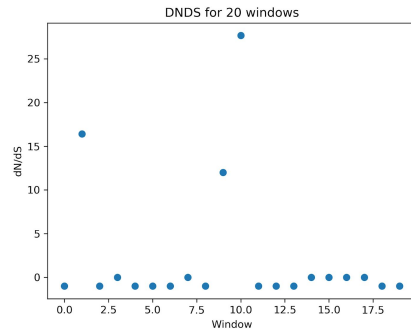
Mon Calamari **Figure 1.11:**



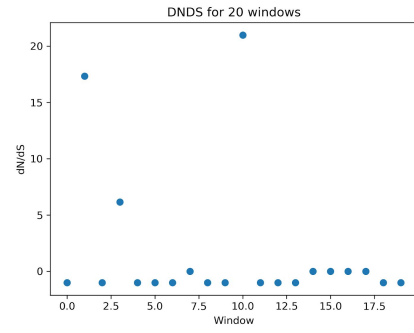
Section III - dN/dS plots for each population

Here, figures depict dN/dS ratios in 5,000 bp windows across the genome for a given population.

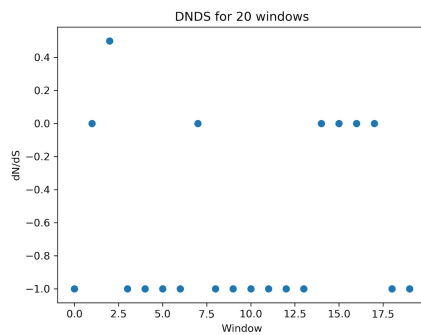
Common Ancestor **Figure 3.1:**



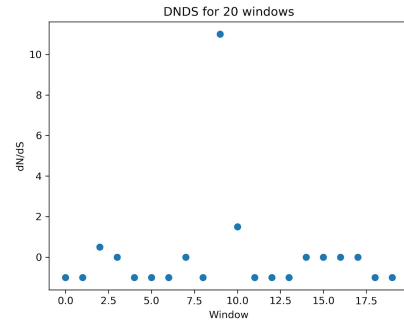
Ewoks **Figure 3.4:**



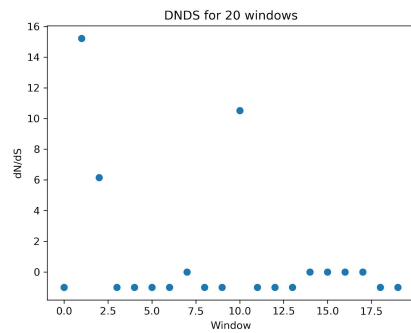
Crolutes **Figure 3.2:**



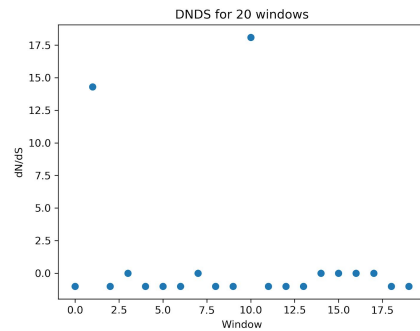
Gungans **Figure 3.5:**



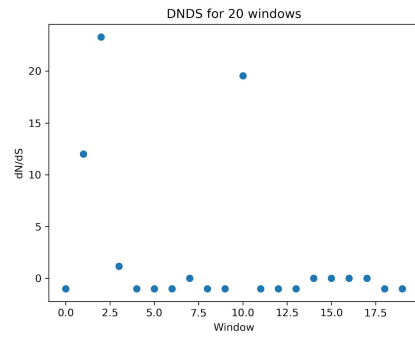
Abednedo **Figure 3.3:**



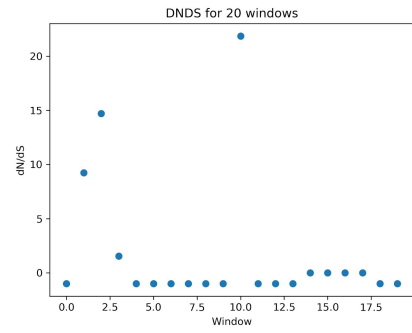
Humans **Figure 3.6:**



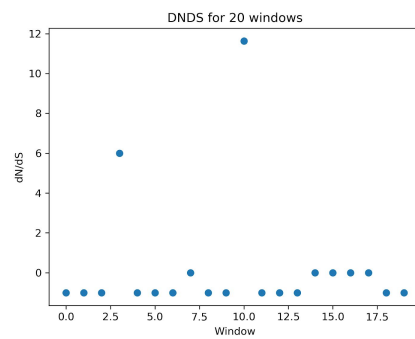
Wookies **Figure 3.7:**



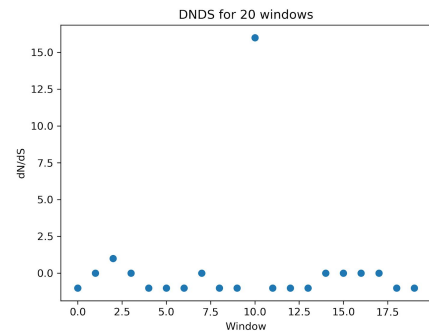
Twileks **Figure 3.10:**



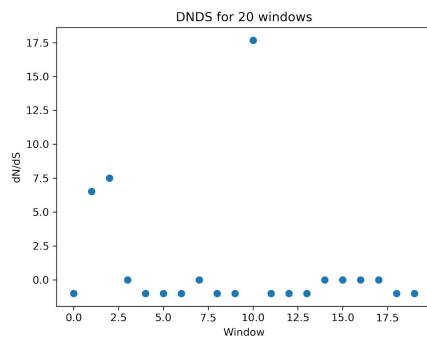
Intermediate Wookies **Figure 3.8:**



Mon Calamari **Figure 3.11:**



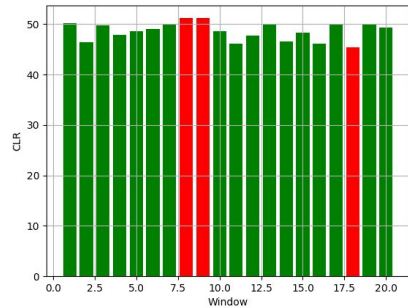
Intermediate Humans **Figure 3.9:**



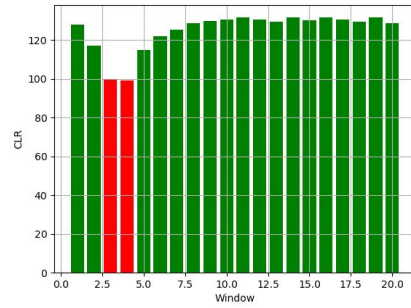
Section IV:

Here figures depict CLR graphs (using Group A's tool) with windows of 5,000 bp over the genome of a given population.

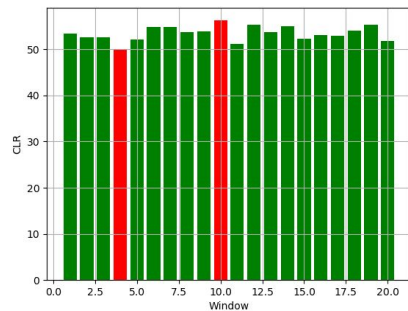
Common Ancestor **Figure 4.1:**



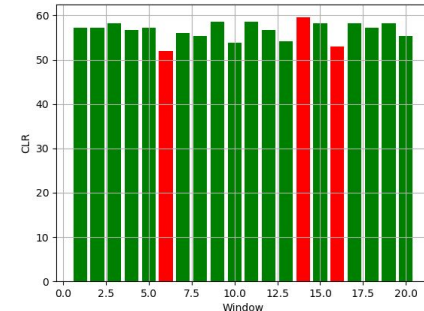
Ewoks **Figure 4.4:**



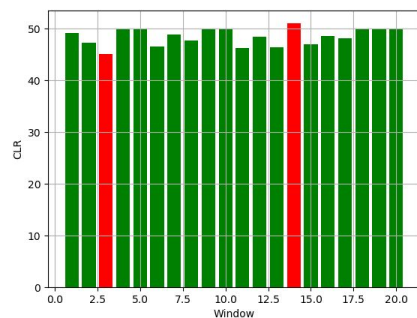
Crolutes **Figure 4.2:**



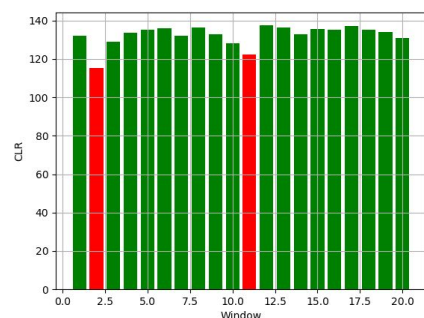
Gungans **Figure 4.5:**



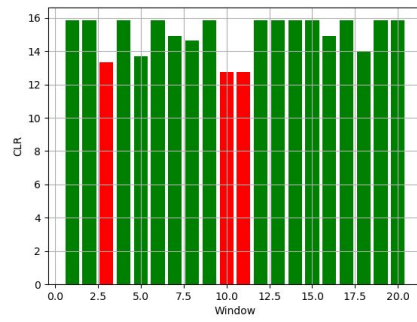
Abednedo **Figure 4.3:**



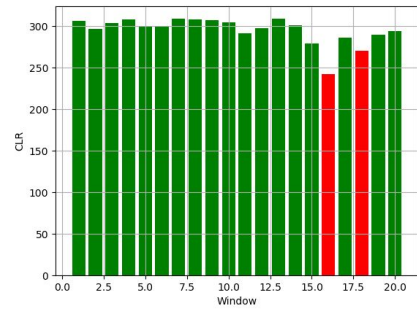
Humans **Figure 4.6:**



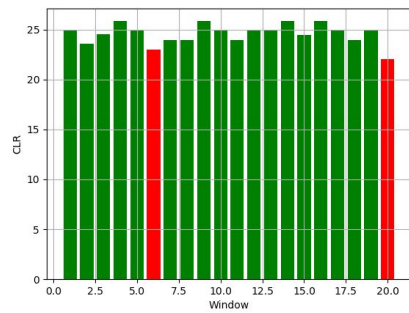
Wookies **Figure 4.7:**



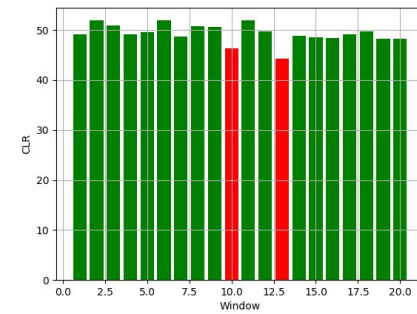
Twileks **Figure 4.10:**



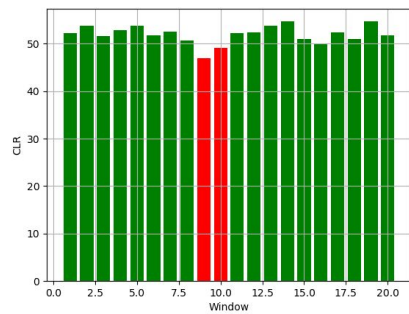
Intermediate Wookies **Figure 4.8:**



Mon Calamari **Figure 4.11:**



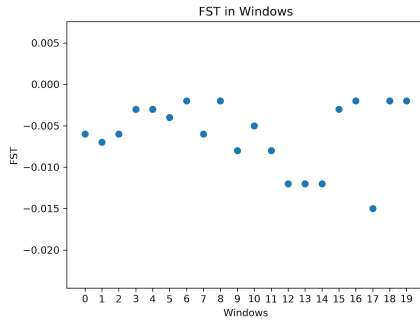
Intermediate Humans **Figure 4.9:**



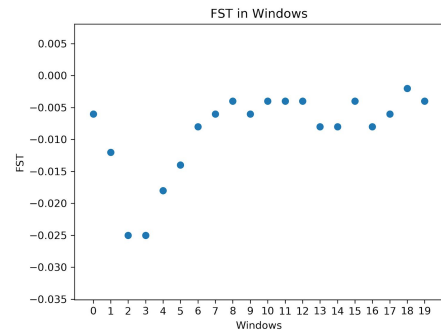
Section VII - F_{st} in windows across each population

Here, figures depict F_{st} values for 5,000 bp windows over the genome of a given population.

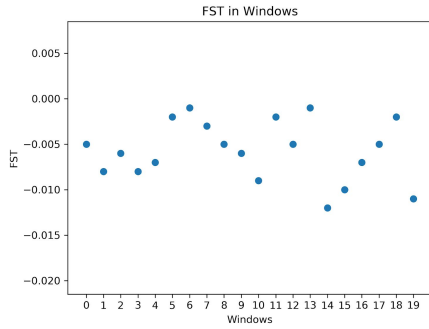
Common Ancestor **Figure 7.1:**



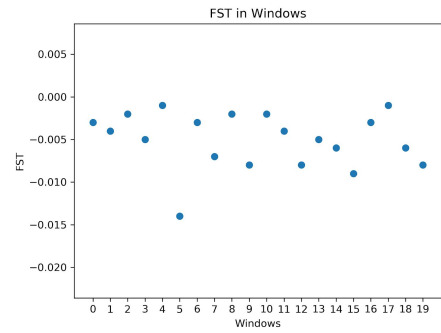
Ewoks **Figure 7.4:**



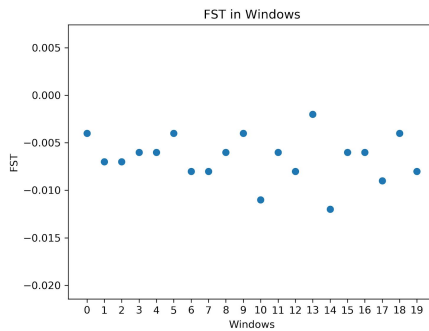
Crolutes **Figure 7.2:**



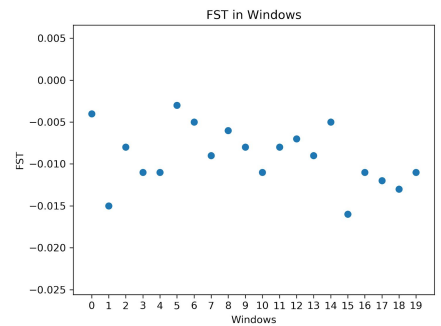
Gungans **Figure 7.5:**



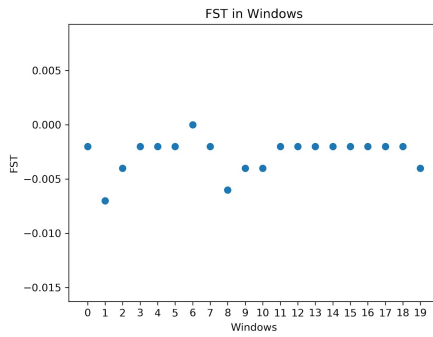
Abednedo **Figure 7.3:**



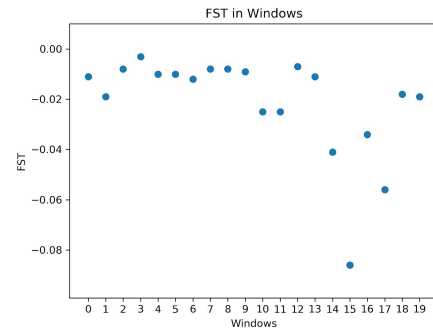
Humans **Figure 7.6:**



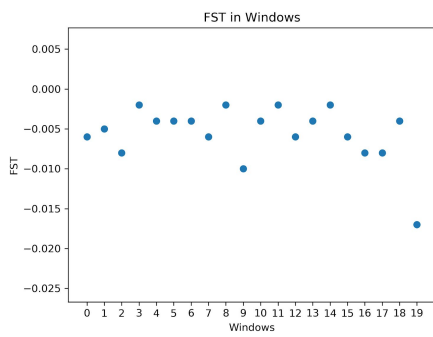
Wookies **Figure 7.7:**



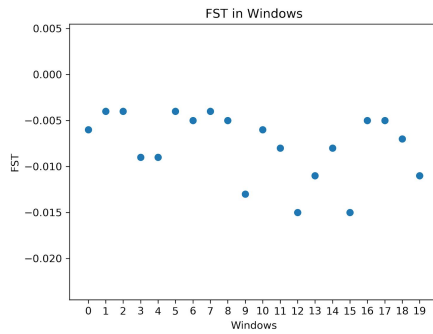
Twileks **Figure 7.10:**



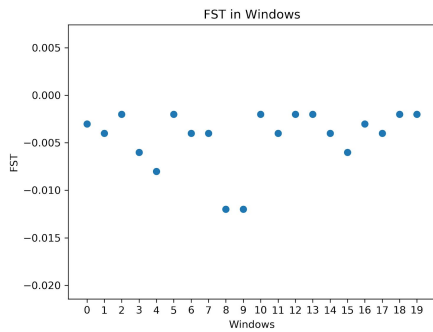
Intermediate Wookies **Figure 7.8:**



Mon Calamari **Figure 7.11:**



Intermediate Humans **Figure 7.9:**



Section VIII - Structure Analysis of all data

Here figures represent figure output from the program Structure over various K values for data for the whole simulation.

Figure 8.1:



Figure 8.2:

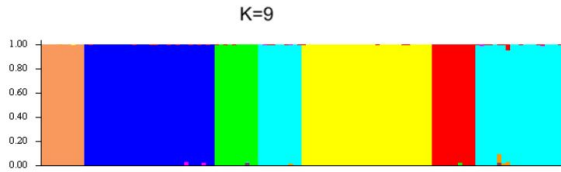


Figure 8.3:

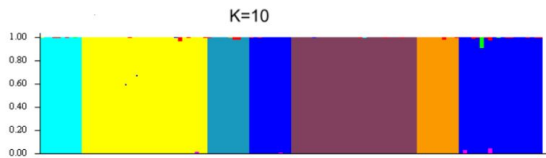


Figure 8.4:

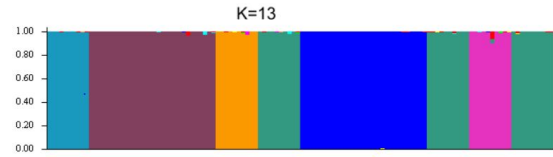


Figure 8.5

Note: K=8 and K=9 (Figures 8.1 and 8.2), which describe the total population simulated, look exactly the same. This is likely due to a bug in structure which seems to cause one simulation to be run in place of another, so it is run twice and another isn't run.

Section IX - Structure Analysis of some individual populations

Here figures represent figure output from the program Structure over various K values for a particular population.

Figure 9.1 K=2 Humans



Figure 9.4 K=2 Ewoks

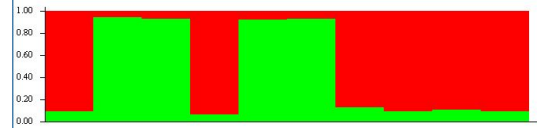


Figure 9.2 K=3 Humans

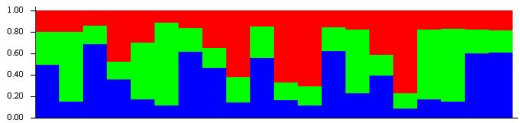


Figure 9.5 K=3 Ewoks

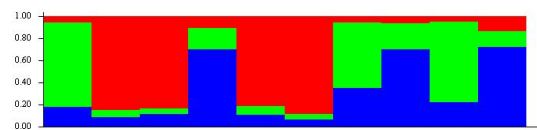


Figure 9.3 K=5 Ewoks



XII. Appendix of Logs

In terms of changing our script, only one change was made during the course of our analysis. We needed to get some extra output with files for our calculations. We needed to get files in the sample format for each individual population as well as a file in MS format for the whole simulation.