

Analysis of Group A Data

Andrew Meng, Kathi Munoz, Joseph Zou

Pi for each population

Brutes : 0.021514117647058822
Elites: 0.01679764705882353
Grunts: 0.005002352941176471
Humans: 0.016254117647058825
Jackals: 0.01683764705882353
Prophets: 0.007009411764705882
Spartans: 0.010384705882352942

Figures 1.1-1.7 gives pi in windows of size 5,000 bp for watch population. In every figure, we see that pi values remain consistent for the most part. The figures appear to have horizontal bars, indicating very similar amounts of polymorphism across the genome. There are a few peaks, some wider than others (Figure 1.1). Notably, Humans (Figure 1.4) and Spartans (Figure 1.7) have remarkably similar pi values across the board. This is in keeping with our understanding of the Halo Universe - Spartans and Humans are supposed to be closely related. However, Figure 1.1 shows pi values for the humans that look much more random compared to the Spartans (Figure 1.7) on the parts resembling horizontal lines. Perhaps this is due to more outside influences/restrictions on the Spartan genome since they are genetically modified. Another interesting one is Figure 1.3 for Grunts; it is highly stratified with most of the pi values falling in these horizontal lines. Furthermore, many of the values are very low, most of them at or very close to zero, and they have the lowest average pi. This makes us wonder if there were some big demographic changes to this population that would put these kinds of restrictions on diversity, like a bottleneck.

N_e for each population

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

N_e	Using S	Using Pi
Brutes	703	53785
Elites	609	41994
Grunts	577	12506
Humans	850	40635
Jackals	629	42094
Prophets	506	17524

Spartans	585	25962
----------	-----	-------

Clearly, the values for N_e estimated with π are much larger than those estimated with s . One reason for this is that with the data we had, we only had access to the number of segregating sites for samples of each population, not the whole. So, it makes sense that this value is underestimating N_e for each. In our own analysis, we found that using π in this calculation consistently overshoot the true size. This makes it difficult to glean anything from these values since things like selection are clearly at play. However, both Grunts and Prophets have the smallest estimates using both s and π , so it seems likely that these two are the smallest among these populations. Looking at Figure 1.6, Prophets also exhibit fairly low π values across the board, which is further evidence for a small population.

dN/dS plots for each population

Here, we calculate dN/dS ratios in windows of size 5,000 resulting in 20 windows for each (Figures 2.1-2.7). 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 indicate zero synonymous sites.

Something that sticks out immediately is the wide range of values seen in humans (Figure 2.4). There is a single window with a ratio close to 70, and the next closest to that is 30. This would indicate that this window is experiencing very strong positive selection, both compared to other windows in its own genome as well as compared to the other populations. Most of the populations also have a large number of windows with negative ratios. This indicates that there are not that many neutral mutations - most mutations appear to affect the fitness of the individual.

Location of Likely Sweeps in each population

To do this, we looked at π in windows of size 5,000 bp again, using Group A's CLR (Figures 3.1-3.7) tool to compare the possible sweep locations they determined with our polymorphism data. Looking back at Figures 1.1-1.7 is also useful, as you would expect there to be troughs in diversity around the area of a sweep.

There appear to be very strong, or at least distinguishable sweeps for humans (Figure 3.4) and spartans (3.7). They are in very similar locations as well. This is in keeping with what we saw in Figures 1.4 and 1.7, though instead of very low π values around this region, we see very high values. Perhaps this makes sense since the π values in that area were so much higher than the rest, a composite-likelihood test would probably find that to be unlikely. Additionally, Figures 3.4, 3.6, and 3.7 look strikingly similar. Perhaps this indicates a strong common ancestry between these populations (Humans, Prophets, and Spartans). We're not sure what to make of the graphs indicating a possible sweep in just a single window. From what we understand of sweeps, there should be a more continuous window. Perhaps it is the remains of an old sweep (Figure 3.1-3.3)

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. In this data, the recombination rate was different over certain 'zones' of the genome, so to calculate r for a specific sweep, we calculated a weighted average for the recombination rate over the range of the sweep.

For the sweep occurring in the Human population from bp 45000 - 105000 (Figure 3.4), we estimate the time of the sweep was approximately 1168 generations ago. For the sweep occurring in the Prophets population from bp 11000 - 80000 (Figure 3.6), we estimate the time of the sweep was approximately 3464 generations ago. For the sweep occurring in the Spartan population from bp 40000 - 100000 (Figure 3.7), we estimate the time of the sweep was approximately 7220 generations ago. It is strange that the Spartan sweep is estimated as taking place farther back than the humans. It made sense that they would have a very similar sweep since they should be related. However, we would have thought the timing would be very similar to, if not later than the sweep in humans. Our results indicate the opposite. We think that the location of the sweeps affected this. Though they were in similar locations, the human Spartan sweep fell over into a different recombination zone than the humans. The determined recombination rate for Spartans was much lower than Humans ($1.9e-9$ vs $9.9e-9$). With the equation being used, a smaller recombination rate would lead to a longer estimate of time.

Estimates of F_{st} for each population

Brutes: -0.002
Elites: -0.002
Grunts: -0.002
Humans: -0.001
Jackals: -0.003
Prophets: -0.002
Spartans: -0.002

These values are all very similar, so there isn't much to glean from it. In general, though, this would indicate very similar amounts of heterozygosity across the populations. This is supported by the figures for F_{st} in windows across each population.

F_{st} in windows across each population

Figures 4.1-4.7 give values for F_{st} in each population in 5,000 bp windows and shows very consistent values for F_{st} , many of them falling between 0 and -0.01. This tells us that most locations are not mutations, or at least not non-neutral mutations. This conflicts with what we observed in the dN/dS ratios of section III. Some figures (like figure 4.1, 4.2, 4.6, 4.7) are off in the number of windows. This is because there are no segregating sites in those samples in those windows. This tells us that either these populations (Brutes, Elites, Prophets, Spartans) have no mutations in this area, or the samples of data that we have for these populations do not represent the mutations present in these locations.

Structure analysis of all data

The table to the right describes the distribution of K, viewed as a random variable, as analyzed by Structure. The most obvious interpretation of these results is that Structure really does not know what is going on with this data, as indicated by the high variances across the board.

Figures 5.1-5.4 show the Structure plots for all data. Once it goes beyond K=10, Structure seems to have difficulty differentiating between populations, so that the true number of populations in this data is less than 10 seems to be a reasonable conclusion. Figure 5.1 (K=7) depicts light blue in both the groups apparently corresponding to elites and brutes (the group corresponding to grunts is mostly light blue). However, figures 5.2 and 5.3 do not seem to differentiate all that much between elites and brutes. The highest p-value, according to the table, belongs to the run assuming K=7, as expected; the large value given for variation is concerning, however.

K	Ln P(D)	Var[LnP(D)]
3	-4728655.9	9416587.7
4	-8068528.6	16100880.0
5	-11898505.2	23760913.4
6	-10860854.9	21683218.4
7	-21015921.1	41996664.9
8	-15886189.9	31734302.8
9	-15532452.7	31024662.1
10	-17301379.4	34562435.3
11	-19864213.3	39686499.8
12	-18213823.8	36383763.8
13	-14631475.4	29211147.9

Structure analysis of some individual populations

Figures 5.7 and 5.8, which describe a sample from the population of spartans, identify the same two groups fairly conclusively, as the shapes of the two graphs are almost identical. One can observe this in that the green in figure 5.7 appears in figure 5.8 as red, and vice-versa. There may be some differentiation within the population of spartans that is significant in some way. It seems reasonable to conclude that, within the population of spartans, $k=2$.

For humans, the $k=3$ suggests more mixing within the population of humans in some significant way. It is also likely meaningful that the shape of the two graphs $k=2$ and $k=3$ (figures 5.5 and 5.6, respectively) is noticeably different between the two graphs. That the graph depicted by 5.5 is less succinctly shaped also supports the conclusion that there is more mixing within the human population than there is within the spartans.

Estimates of migration rates

Figure 5.9 is a PCA plot describing the relationships between the seven known populations. As it describes, populations 1, 4, and 5 are all lumped fairly close together. This could be the result of an attempt at displaying very complex data in only two dimensions, but it could also imply that these three populations (prophets, jackals, and grunts) are closely related. There is no overlap between the three groups, however, so they likely are not mixing. Conversely, populations 6 and 7 (humans and spartans)

overlap a lot and exist almost entirely in the same place, so, most likely, humans and spartans are mixing, so they are a candidate for migration between the two populations.

To estimate the potential migration rate between these two, we use the estimates of N_e from section II, using both s and p_i for the equation $m = 1 - F_{st} / F_{st}^* 4N_e$, where m is the migration rate, F_{st} is the F_{st} value between the proposed migrant population and the total of the migrant and 'mainland' population, and N_e is the size of the proposed migrant population.

Humans migrating from spartans using p_i N_e 40,635 m rate = $8.128366509589902e-06$
Humans migrating from spartans using s N_e 850 m rate = 0.000388583733079042

Spartans migrating from humans using p_i N_e 25962 m rate = $1.0958415597417414e-05$
Spartans migrating from humans using s N_e 585 m rate = 0.0004863288645130785

Although our computations estimate that spartans are migrating from humans more than humans are migrating from spartans in each case, this difference could easily be accounted for by what seem to be very poor estimations of N_e , so we can not be very confident in claiming that spartans are migrating from humans more than the other way around. Additionally, we do not have data for a composite population of spartans and humans, so we treated the composite population of prophets, elites, brutes, jackals, grunts, humans, and spartans as the 'mainland' population, which violates an assumption made by our computation. We also violate the assumption that the 'island' population be much smaller than the 'mainland' population in at least one of the two cases, as one of them has to be at least half as large as the total.

Appendix

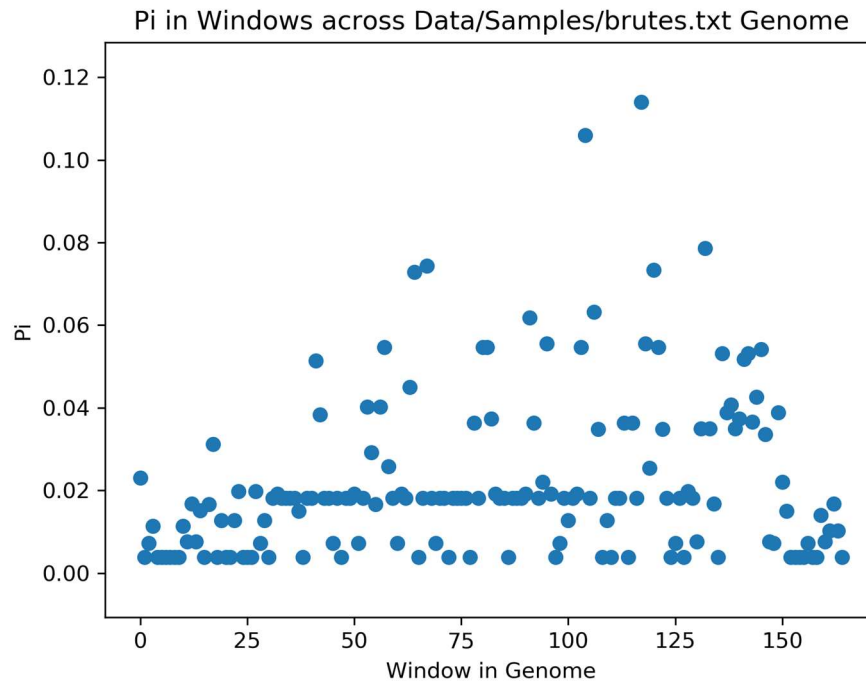


Figure 1.1 π_i value in windows for brutes

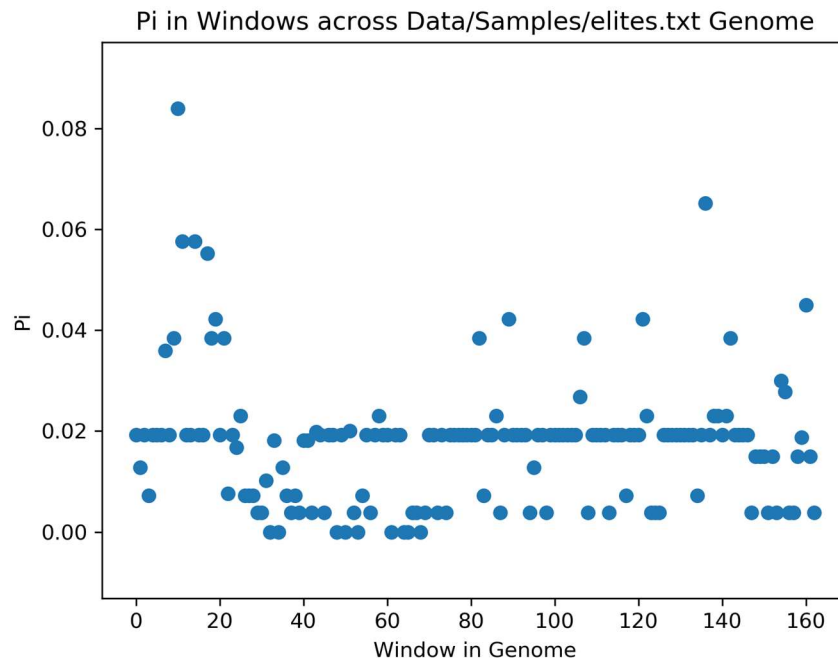


Figure 1.2 π_i value in windows for elites

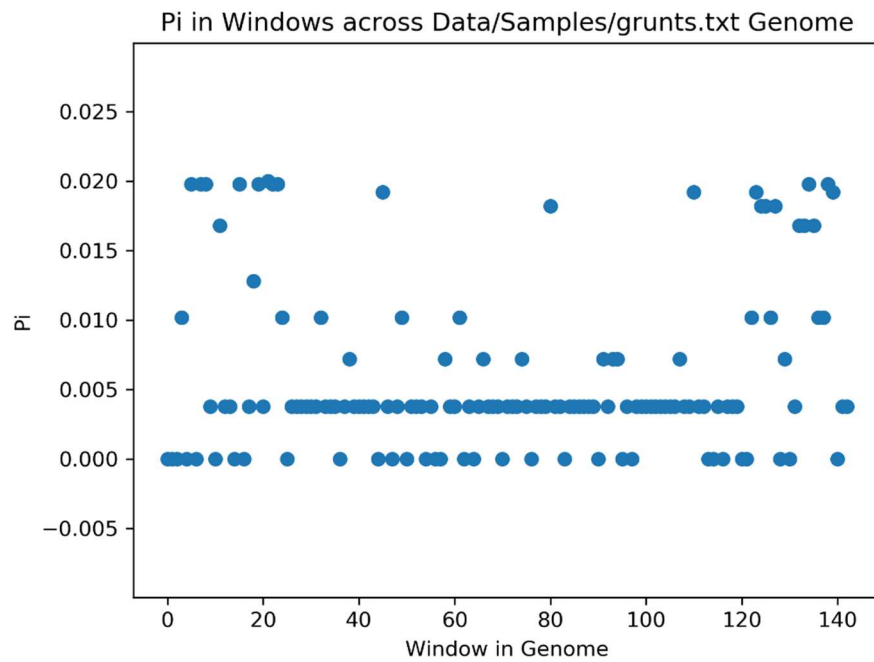


Figure 1.3 Pi value in window for grunts

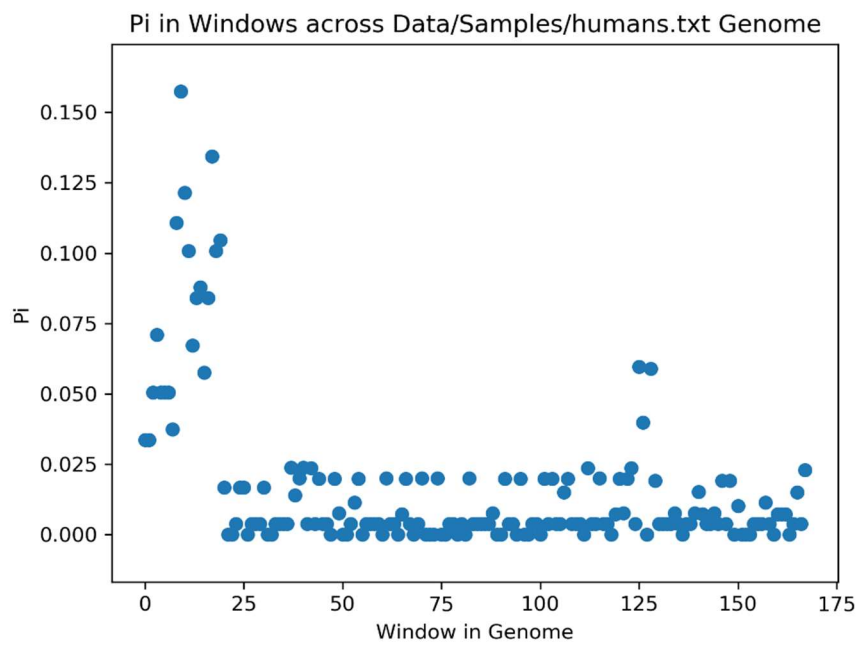


Figure 1.4 Pi value in windows for humans

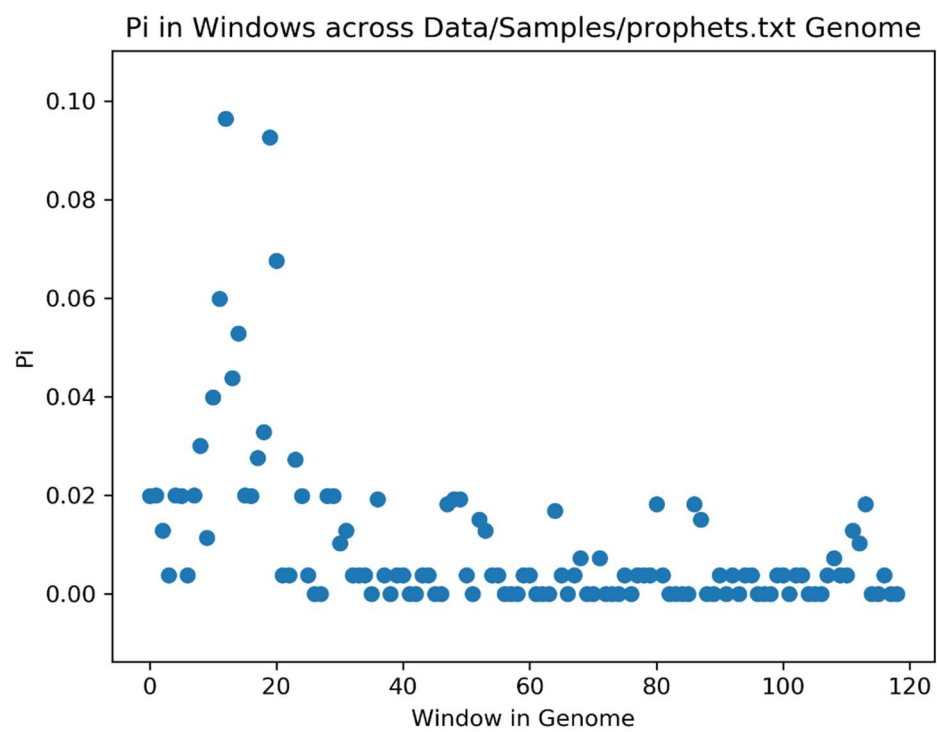
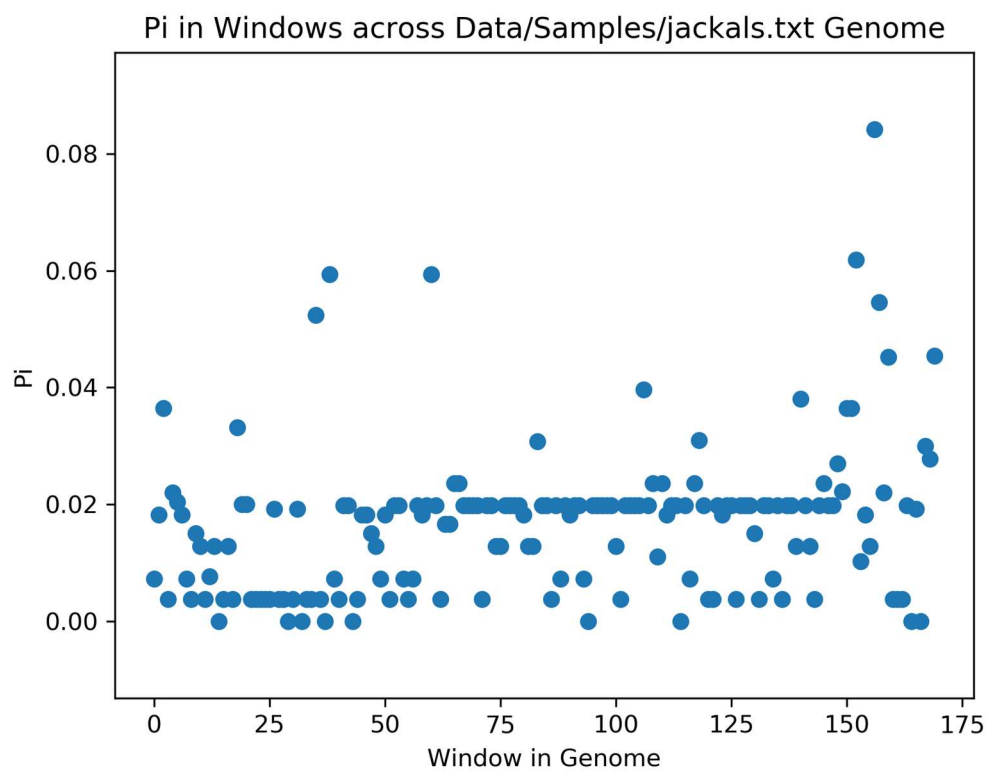


Figure 1.6 π value in windows for prophets

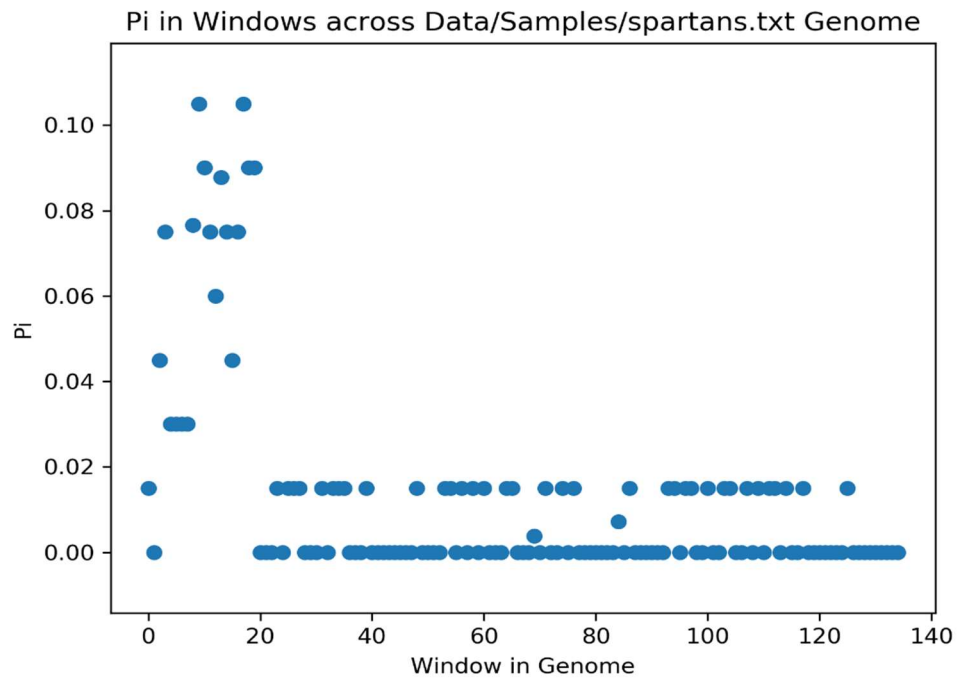


Figure 1.7 Pi value in windows for spartans

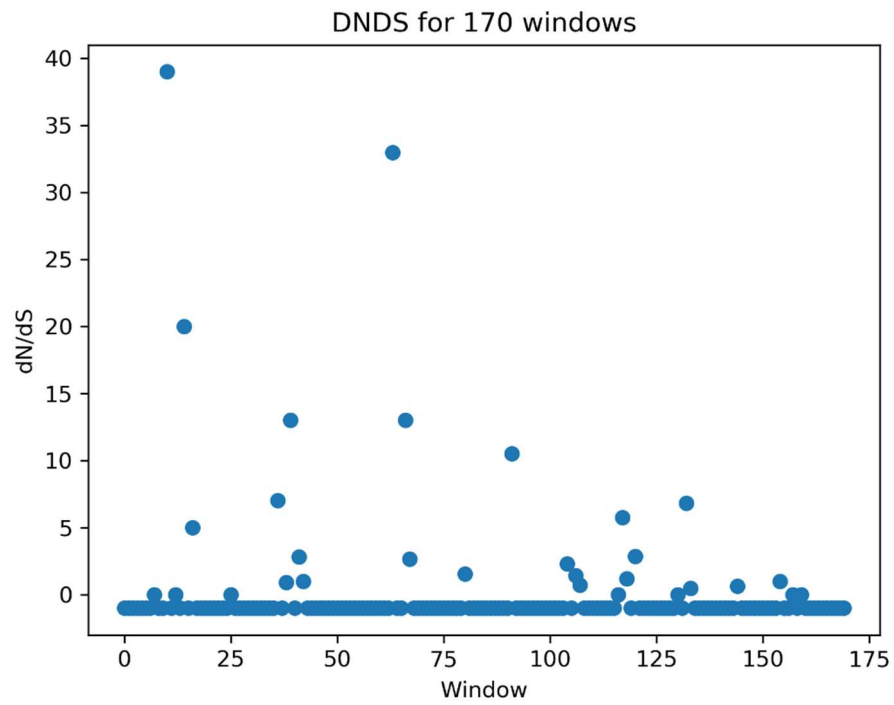


Figure 2.1 DNDS value in windows for brutes

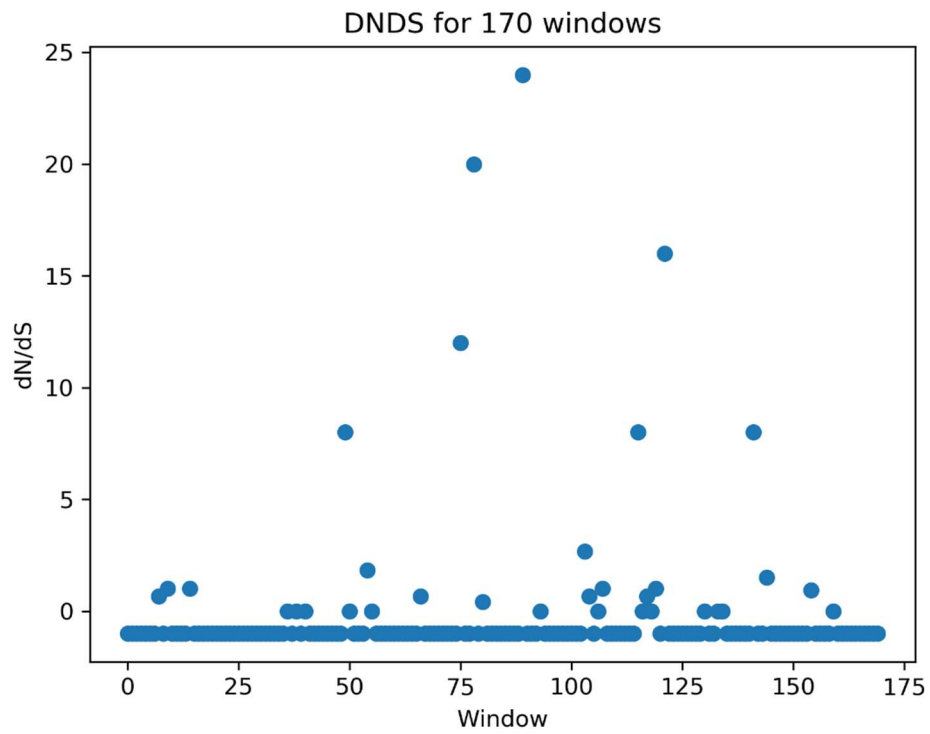


Figure 2.2 DNDS value in windows for elites

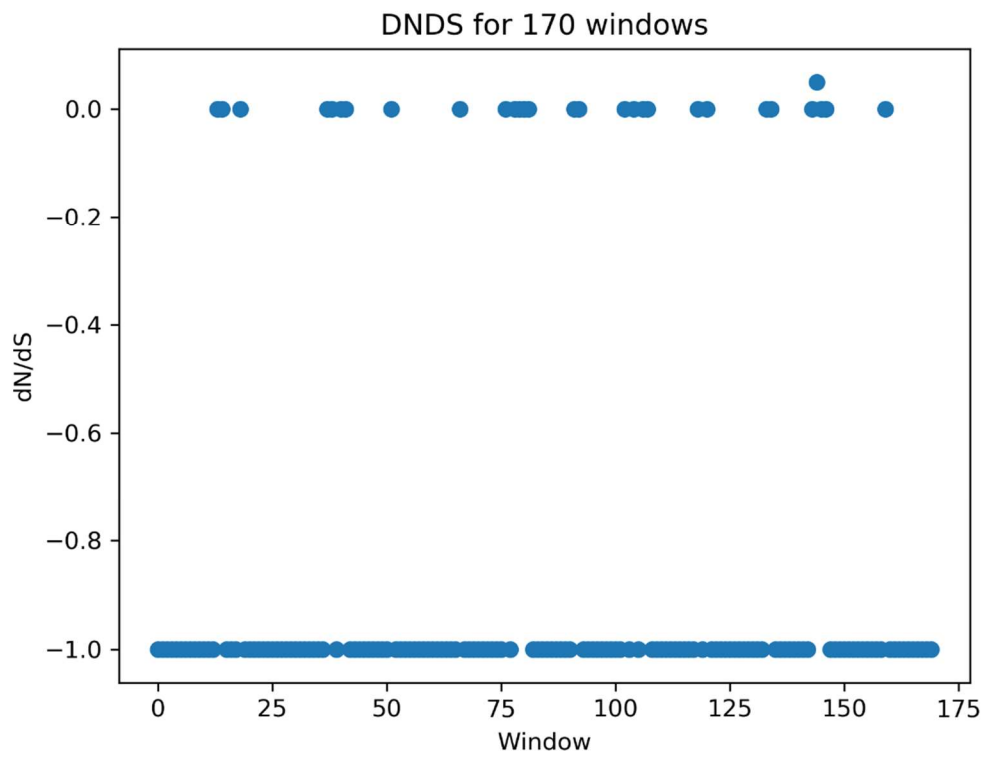


Figure 2.3 DNDS value in windows for grunts

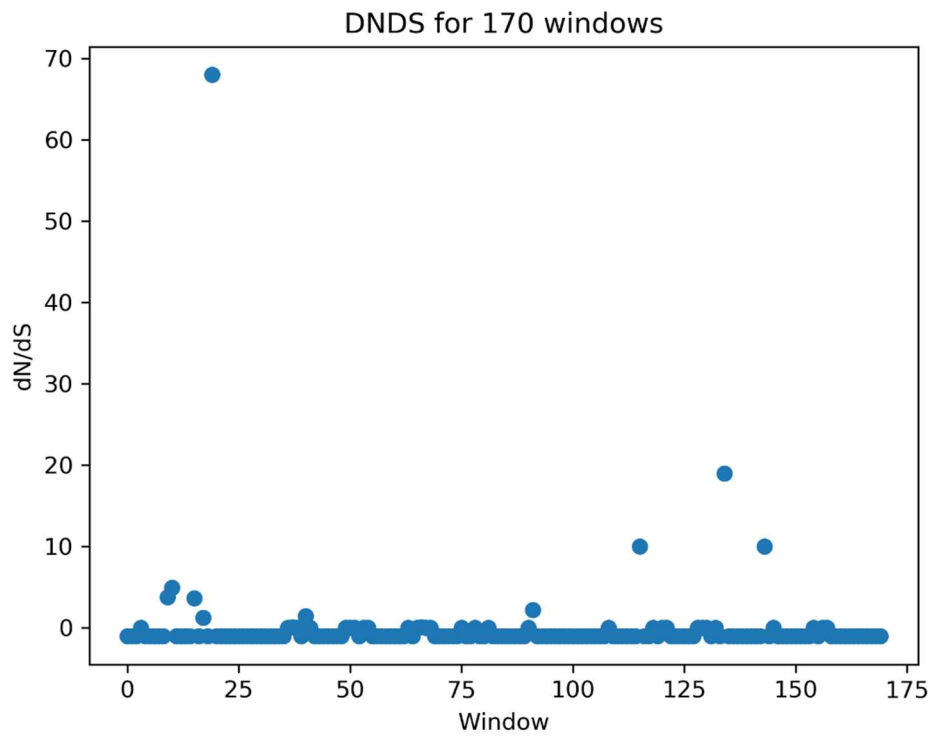


Figure 2.4 DNDS value in windows for humans

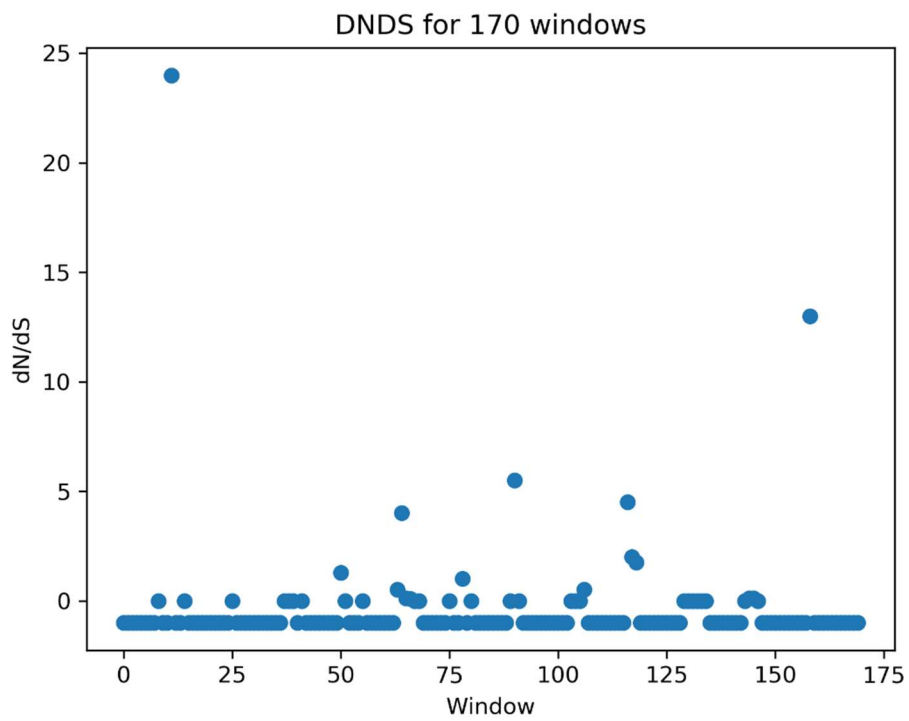


Figure 2.5 DNDS value in windows for jackals

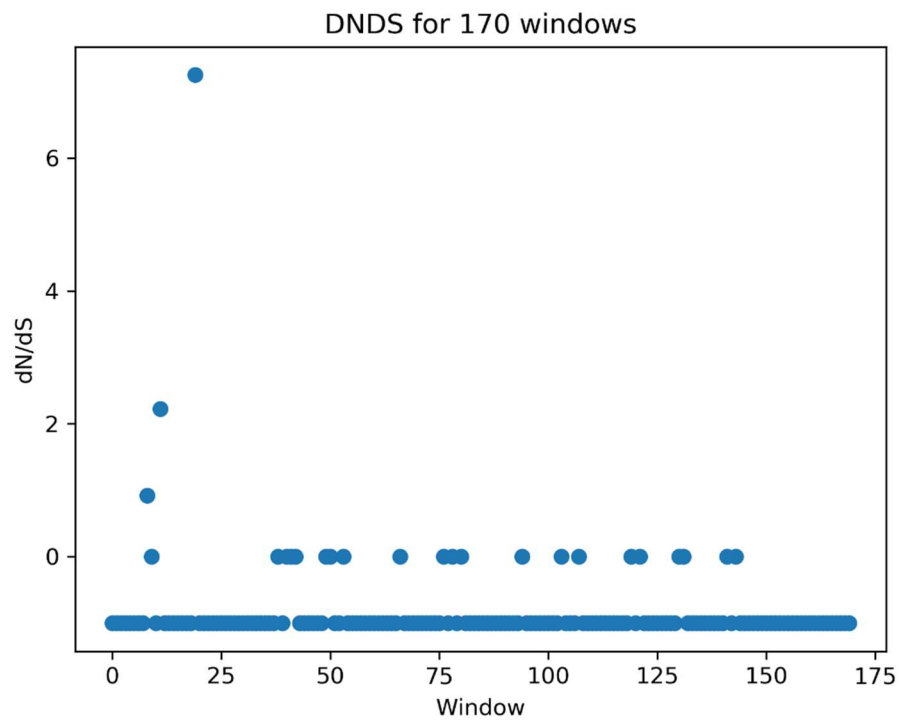


Figure 2.6 DNDS value in windows for prophets

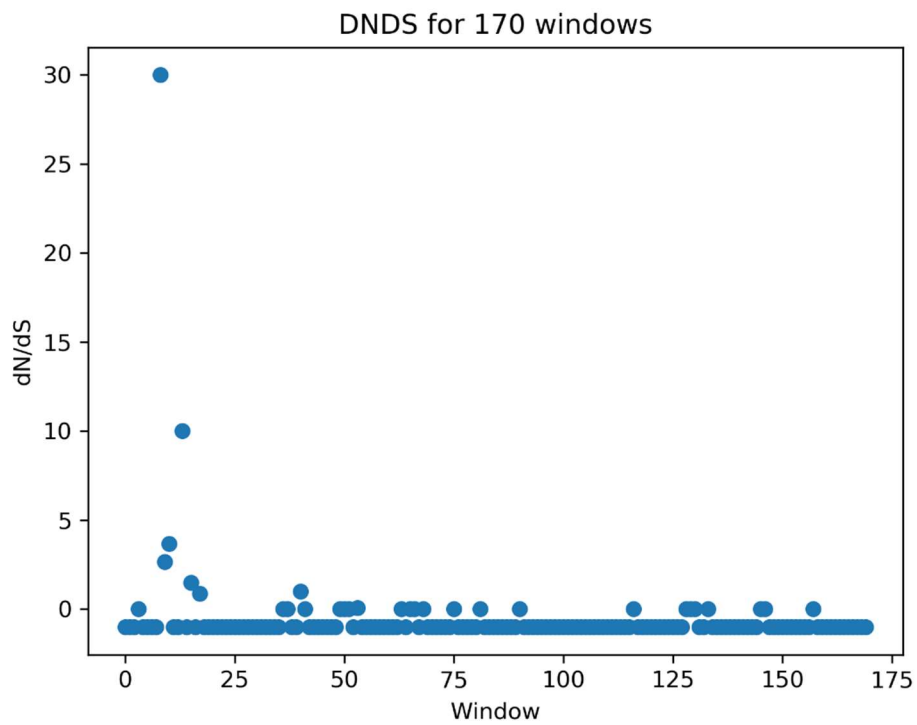


Figure 2.7 DNDS value in windows for spartans

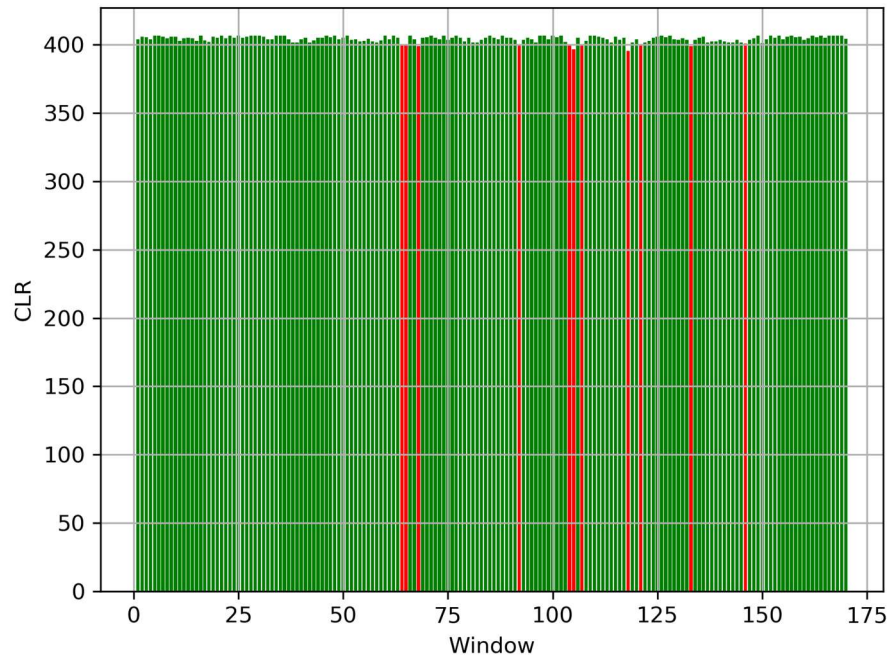


Figure 3.1 CLR in windows for brutes

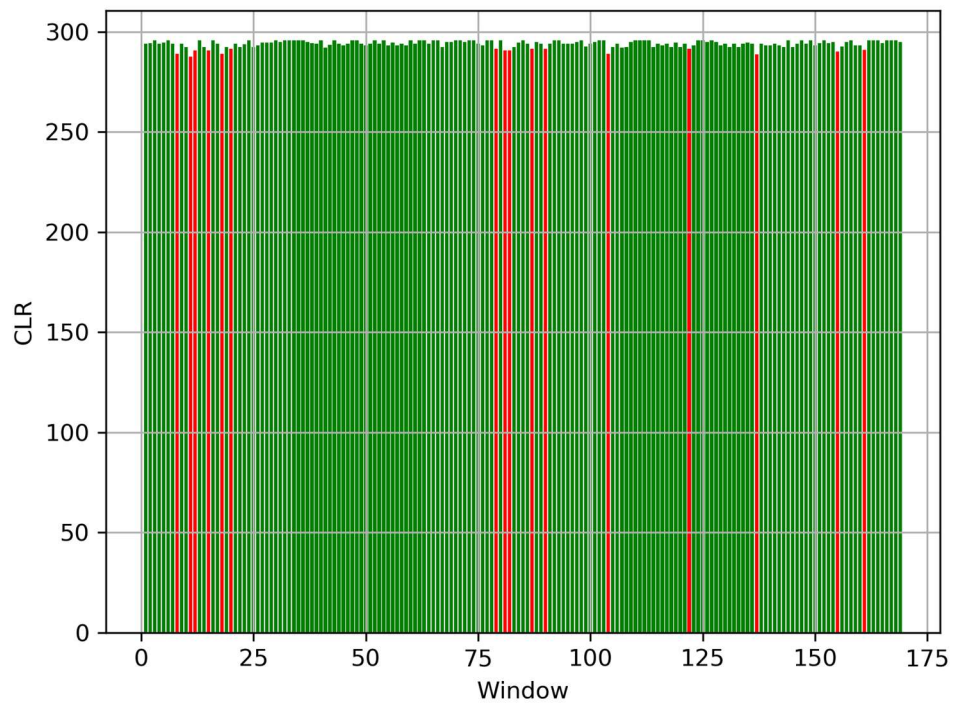


Figure 3.2 CLR in windows for elites

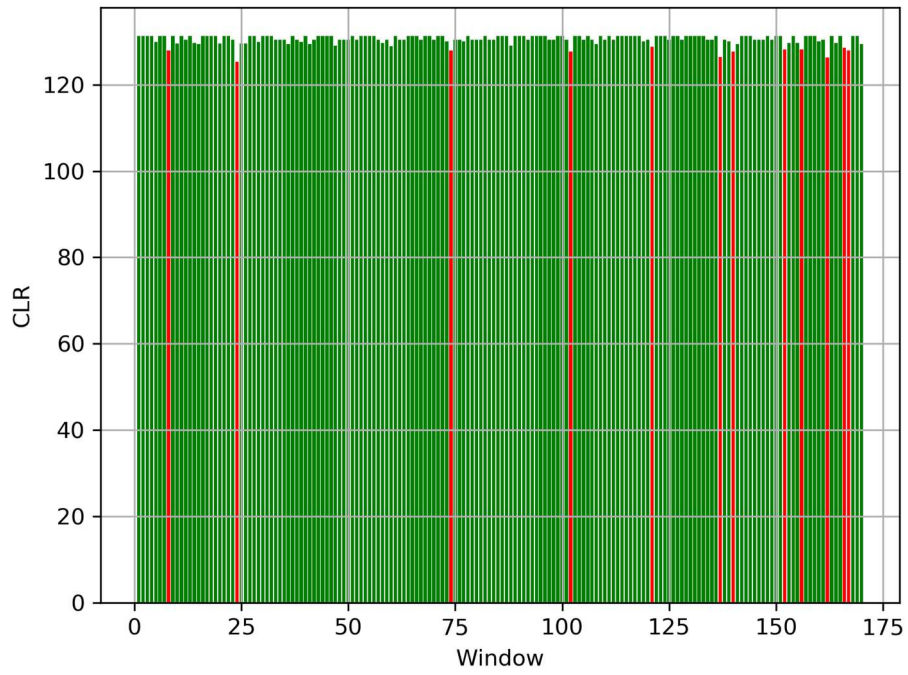


Figure 3.3 CLR in windows for grunts

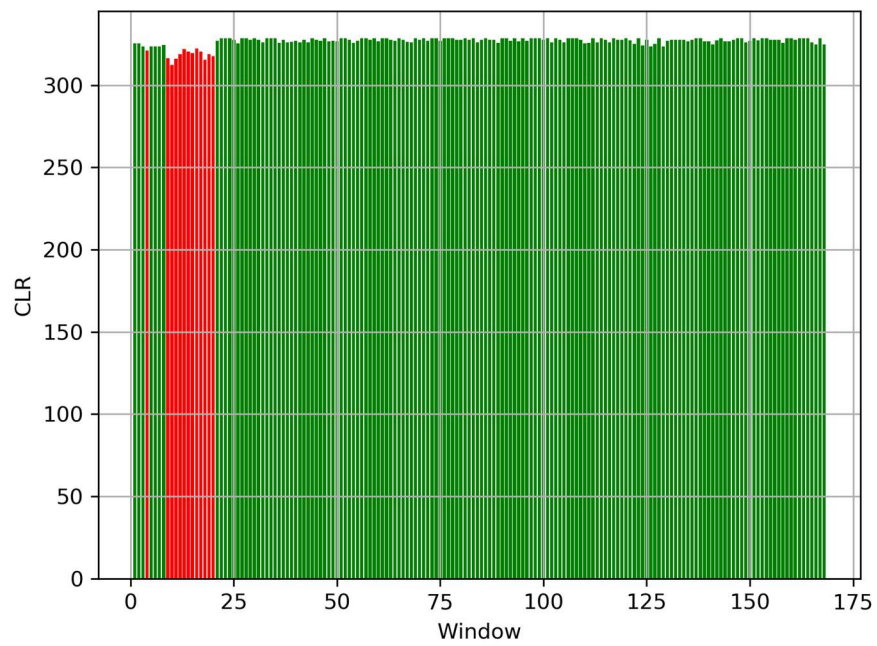


Figure 3.4 CLR in windows for humans

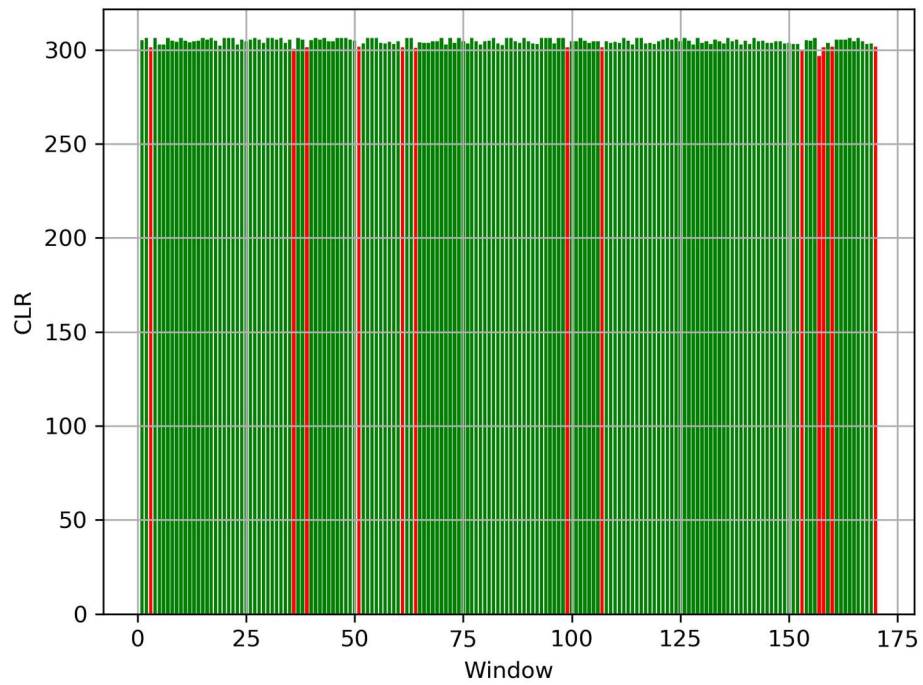


Figure 3.5 CLR in windows for jackals

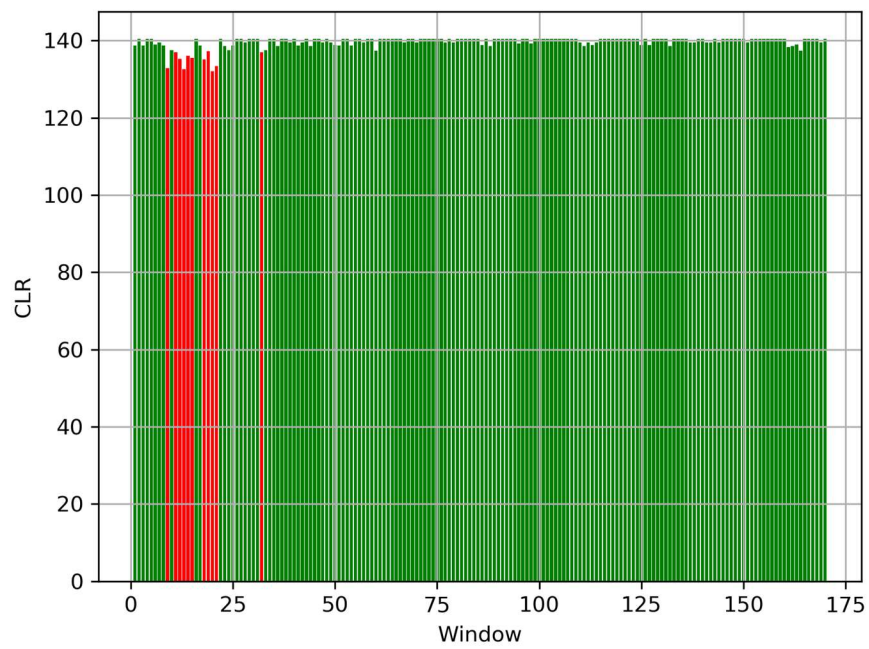


Figure 3.6 CLR in windows for prophets

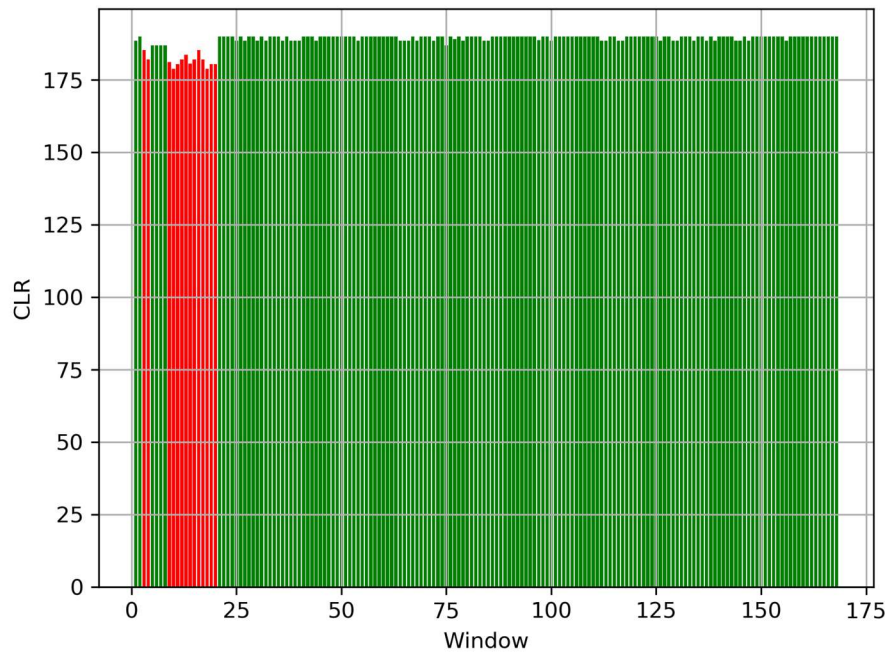


Figure 3.7 CLR in windows for Spartans

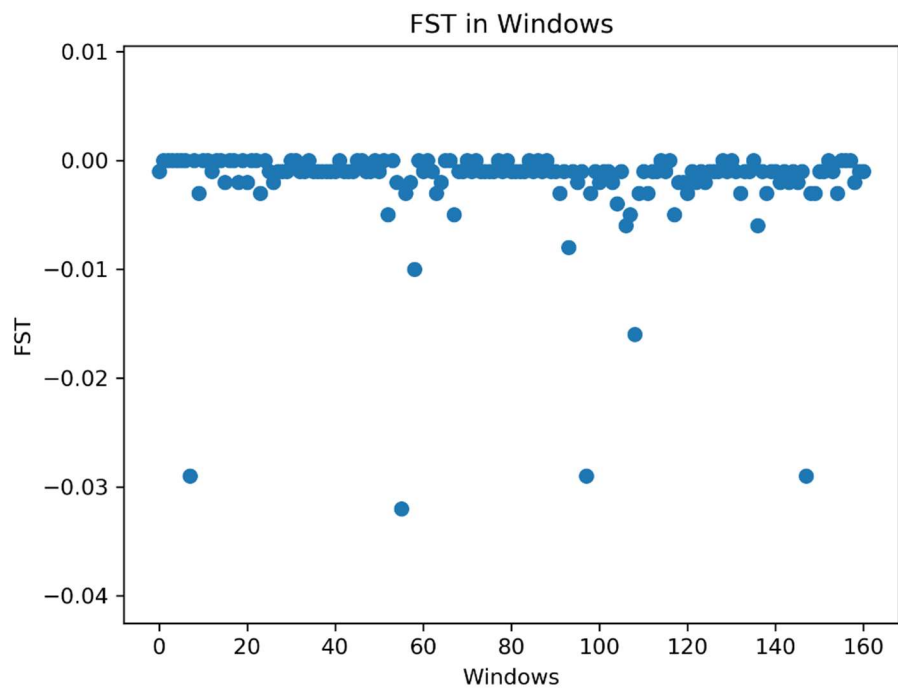


Figure 4.1 FST value in windows for brutes

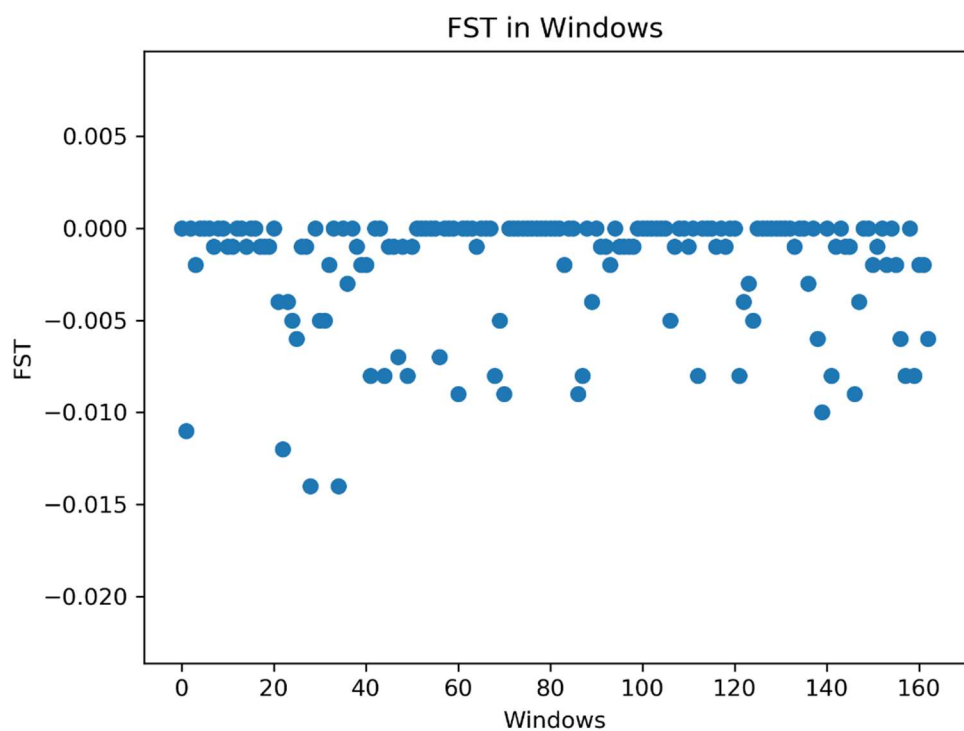


Figure 4.2 FST value in windows for elites

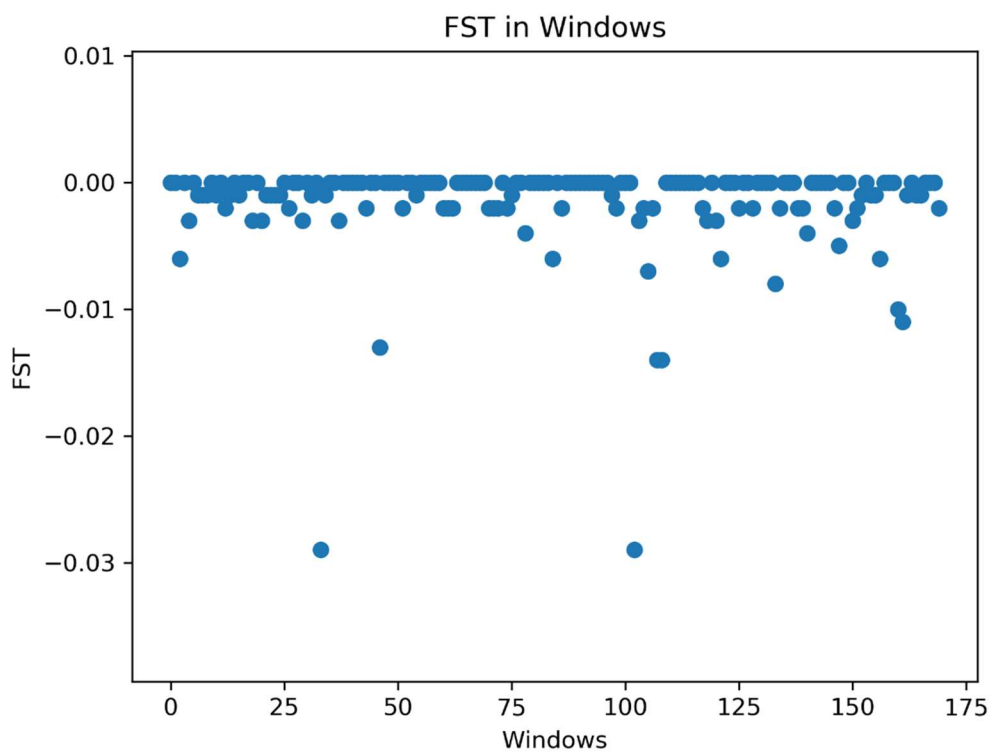


Figure 4.3 FST value in windows for grunts

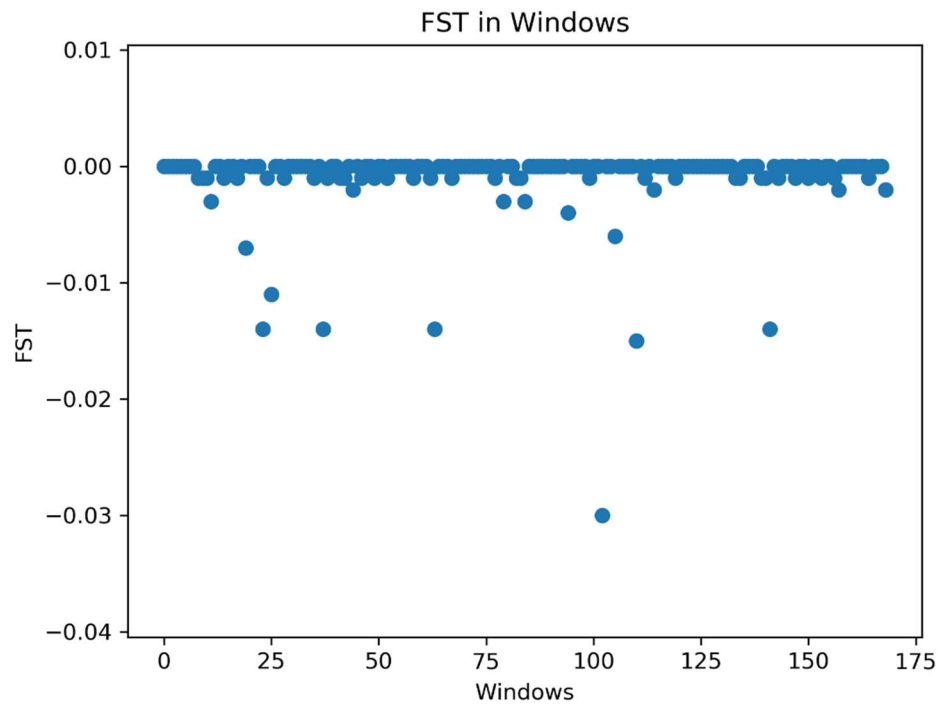


Figure 4.4 FST value in windows for humans

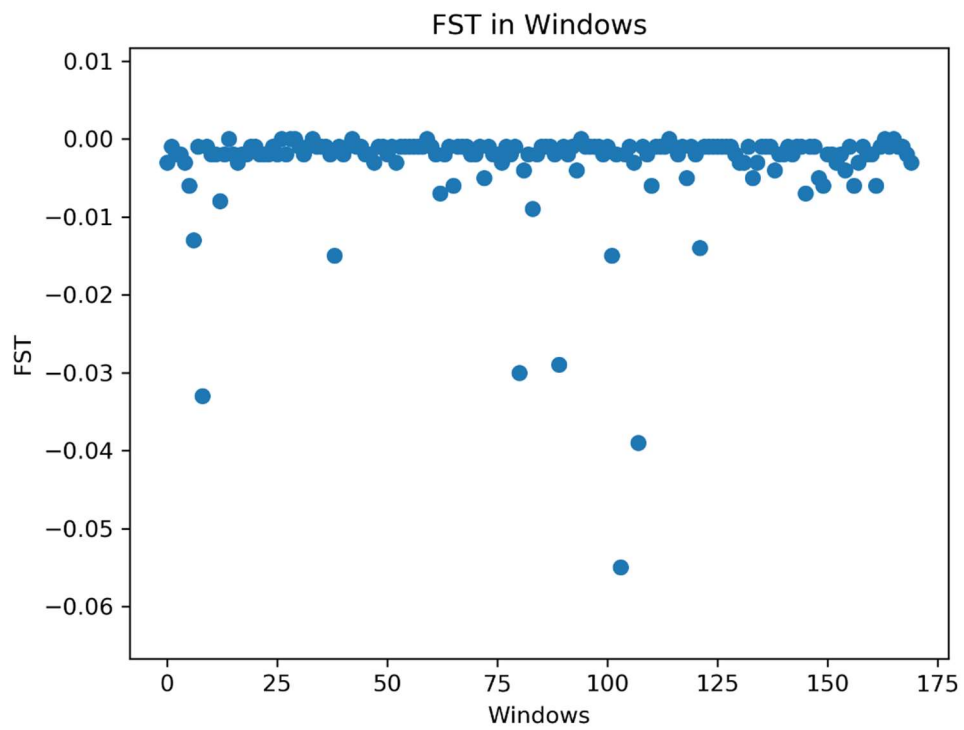


Figure 4.5 FST value in windows for jackals

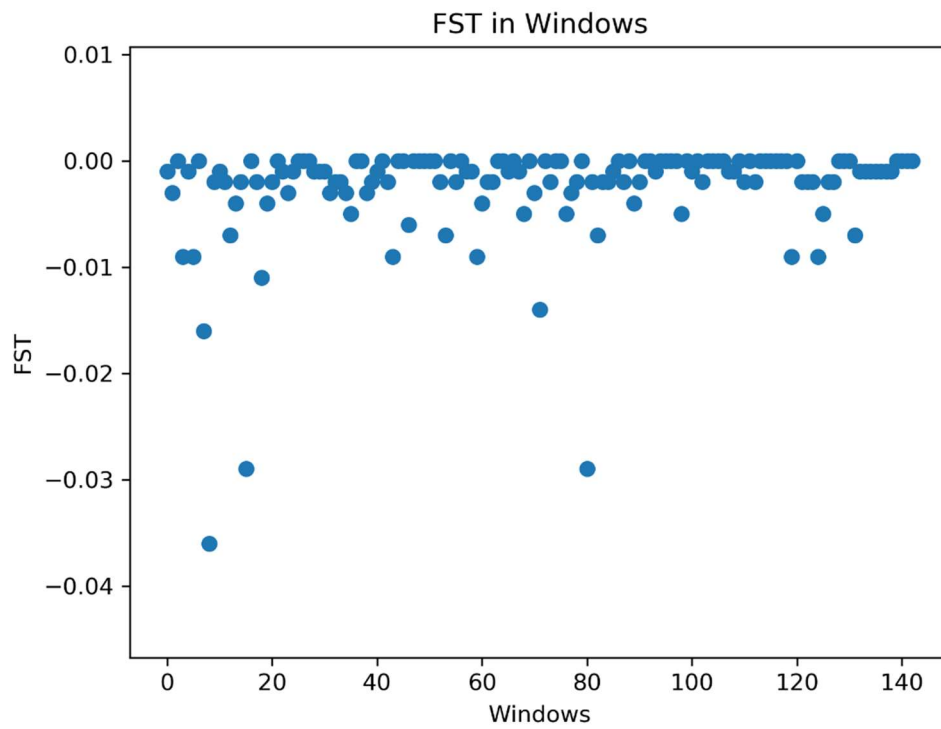


Figure 4.6 FST value in windows for prophets

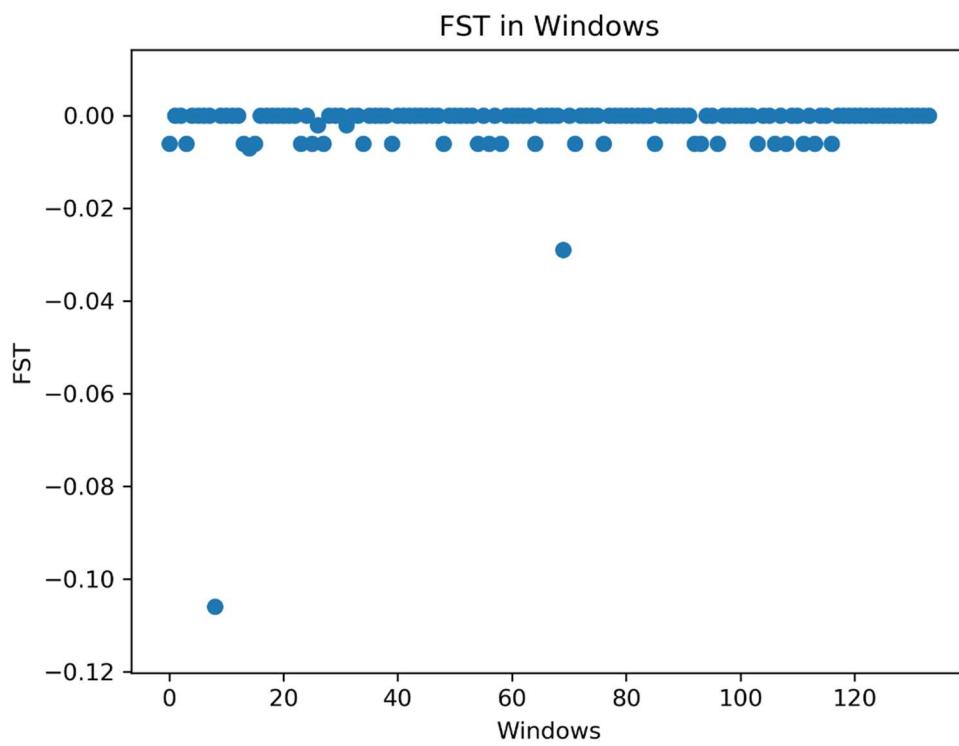


Figure 4.7 FST value in windows for spartans

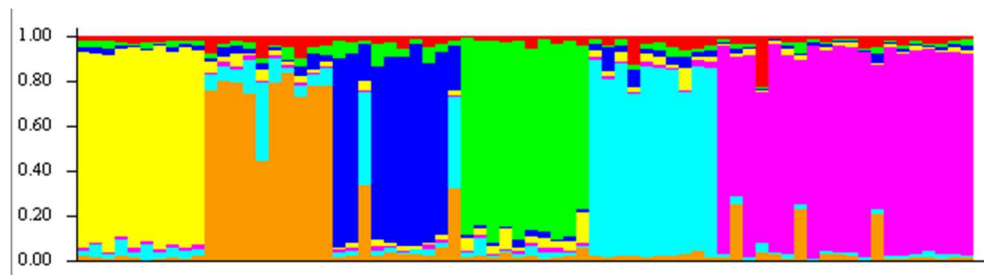


Figure 5.1 Structure for all when $K=7$

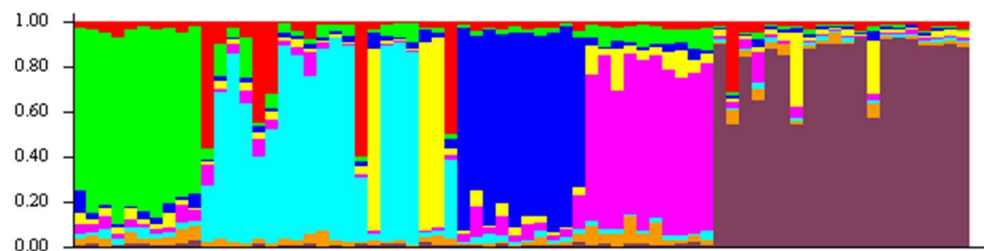


Figure 5.2 Structure for all when $K=8$

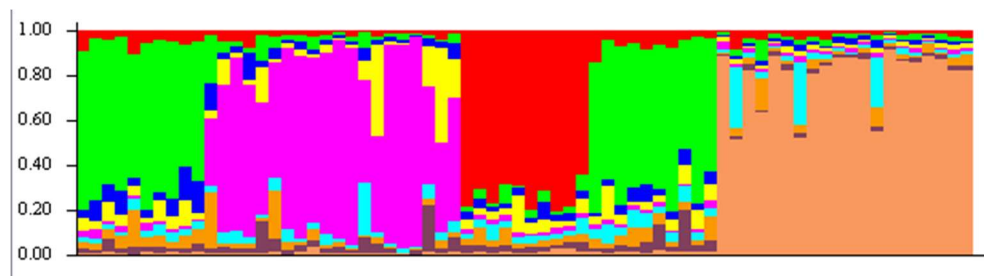


Figure 5.3 Structure for all when $K=9$

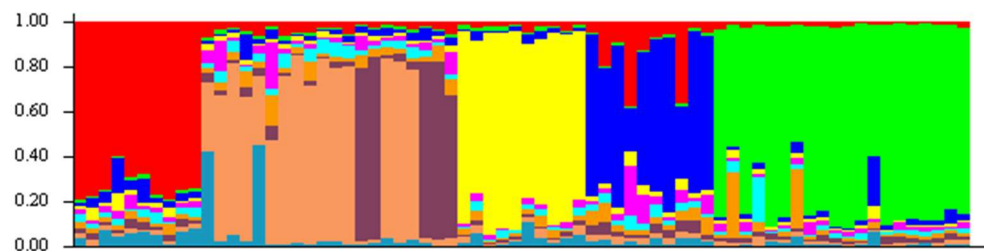


Figure 5.4 Structure for all when $K=10$



Figure 5.5 Structure for humans when $K=2$

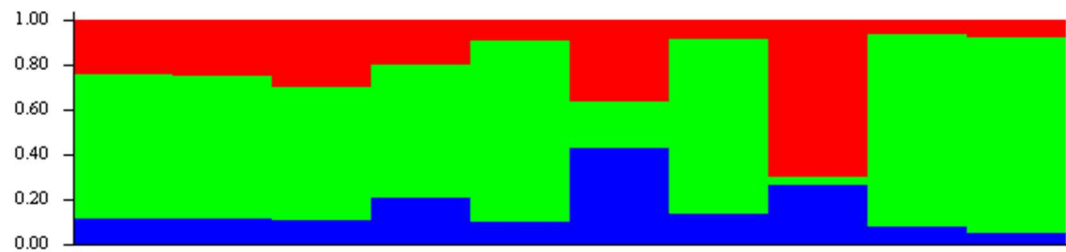


Figure 5.6 Structure for humans when $K=3$



Figure 5.7 Structure for spartans when $K=2$

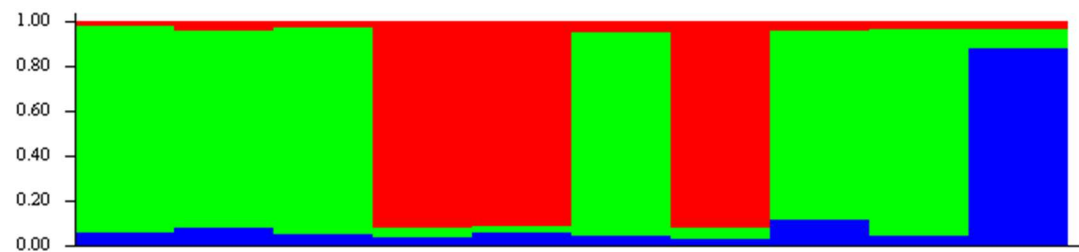


Figure 5.8 Structure for spartans when $K=3$

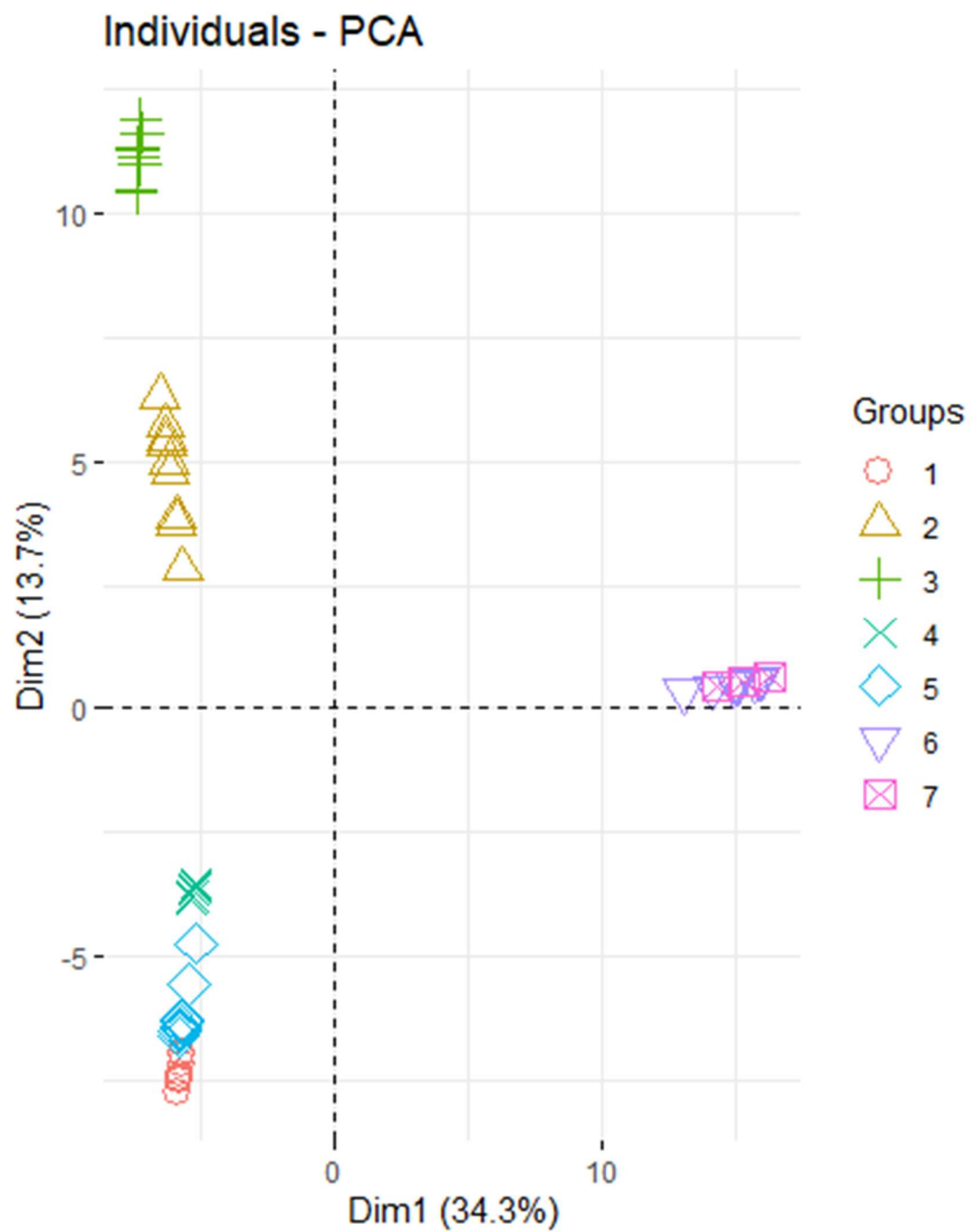


Figure 5.9 PCA diagram for all populations