Devan Bicher

Project 2: Implementation Report

4/5/13

Section 1

Reproductive fitness is the ability of an organism or a population of organisms to replicate and survive. Needless to say analyzing genomes is a very useful way to look at reproductive fitness. Furthermore comparing genomes or portions of genome sequences between organisms in populations is a very useful way to look at reproductive fitness between different organisms. These sorts of analyses can tell us many things about the genomes of the population. One of the things that alignments can tell us are conserved regions of a genome that are crucial to reproduction. If a certain region of a genome is the same between many different organisms in a population that this is probably a very good indication that this gene or sequence within the genome is vital for the reproduction of the organism. Regions that vary more however may be an indication of regions that are either not as crucial or perhaps regions that may vary to give diversity to a population. But just because a region varies more may not mean that it is any less important to the survival of the organism. If the genomes amongst a population were completely the same then the organisms would probably not have much reproductive fitness. This is true because variation in a genome is necessary so that if some strange environmental change were to occur there would be differences that would allow the population to adapt to the pressure. Highly variable regions of the genome mostly likely represent regions where crossing over may have occurred, where variability is unimportant, such as non-coding regions, or perhaps where a gene no longer functions. These regions may still be important for the genetic diversity of the population to produce and survive because these highly variable regions represent a significant source of genetic diversity for adapting to extremely variable settings.

The other case where the set of sequences is not representative and is just a set of sequences instead still has some significance, but is less likely to be as important. Highly similar regions might be indicative of conserved regions, assuming of course that these sequences came from different organisms. In fact if these sequences are from vastly different organisms than very similar sequences would be an interesting phenomenon and it would represent a gene or sequence that has been extremely well preserved across species. But the significance of similar sequences decreases rapidly form there; slightly variable sequences may be different proteins with similar functions just as well as they might be the same proteins with genetic diversity, as might be the case within a population. Or it may be the case that similar sequences are just that, similar and they may not code for similar proteins at all. Lastly where regions of sequences are highly different it probably has very little significance and in fact it is much more likely that these different sequences just represent much different proteins than they do similar proteins with high mutational variability. If this is the case there is no real importance in these sequences as there is not much more to deduce other than the fact that these are simply dissimilar sequences.

Section 2

Genes tell us about the proteins that they code for and aligning genes tells us about the difference between the genes of the respective organisms that the genes have come from. However highly similar, or completely the same for that matter, genes may have drastically different affects in the organism that they reside in. This is because genes and their sequences are only the start to the story about how they influence the whole of the phenotype of the organism. Many other factors influence how a gene functions in an organism. One of these is the transcription factors that control activation of the gene, whether it be through promotion or blocking, however the transcription factors themselves are still only part of this story because the factors themselves are only the proteins that directly do the controlling. Another part of this is the influences outside of the actual gene that directly affect the transcription factor, and therefore indirectly affect the controlling of the genes. Transcription factors are influenced by other molecules, coactivators or corepressors, that are present in the organism at the time and depending on the conditions promote or block, respectively, the function of that factor. This is important biologically because more of one substance might promote the factor thereby amplifying the transcription and function of the gene, while more of another substance might block the factor and decrease the function of the gene. These are examples of outside factors that contribute to the functioning of the gene. There are other factors that are detectable on the genome near and around the gene that will also affect its function.

Typically these other influences are near the gene and they usually work by acting on the transcription factors as well. One such example is the regulating DNA regions, again promoting or blocking, near the target gene; these are the regions that the transcription factors actually bind to. Once again these regions may allow the factors to enhance or decrease, respectively, the functioning of the gene a certain. These regions of genomes may not even necessarily be near the gene, and in fact they may be far away, but by the constantly twisting and bending of DNA they can affect the gene’s function through the coupled binding of the regulating region and actual coding region of the gene. Other influences on the gene are the frequency of the gene, because some genes repeat themselves in a genome so that they produce more of a crucial protein. Another factor may be the location of a gene in a genome, if the gene is near a region that is bound up so as to not produce much of that other gene than often the secondary gene will not be transcribed as frequently as a gene that is farther away from such a region.

Regulation of genes plays such an important role that it is being widely investigated now that we are so proficient at revealing the sequence of a gene. But because of the complexity of biological systems and the numerous factors that affect these systems it is much more difficult to describe the affects at any given bodily condition. Reading the sequence of a gene and deducing the sequence of its encoded protein is relatively straightforward but revealing how all of the molecular interactions affect these genes is far more complicated. The most logical place to start when analyzing a genes function outside from its actual sequence is to see what protein it encodes for. Knowing this it is possible to deduce experimentally what outside factors may control transcription.

A hypothetical experimental procedure to deduce these factors would be to gather a large testing set of molecules that may or not affect this gene. Then in the presence of only the most basic ensemble of transcription molecules, like the raw transcription factor, transcriptase, and ribosomes to produce the protein (just to name a few of these molecules), that will produce this gene’s protein we will wait for the molecular machinery to do its work and produce lots of protein. After some specific and predetermined time the concentration of protein will be measured. This is essentially the control level of protein made from this gene and with this knowledge we can systematically apply the other hypothetical molecules both individually and in different combinations at varying concentrations. After the same preset time we will collect the level of protein encoded by this gene in the presence of each molecule or molecular combination and then we can determine which molecules affect the functioning of the gene and in what ways. This type of experiment would be significantly aided by the use of computer technology because we could tell the computer which molecules to add, in what concentrations, in what combinations and at what time to collect the protein levels. The computer would take in all of this data and then we could integrate it and use it in a meaningful way to determine the biological impacts of these other molecular factors. An extension of this experiment would, with the knowledge of how these outside molecules might be present in the body at certain conditions, show how this gene is affected by different bodily stresses and conditions.

A short example of these types of mechanisms is illustrated by lactase. Increased levels of lactose in the body increase transcription of lactase, in order to breakdown lactose into sugars useable by the body. However if glucose is present, regardless of how much lactose is present, the body will suppress the transcription of lactase because it costs much less metabolic energy to breakdown glucose for use by the body than it does to breakdown lactose.

In a similar experiment to the one described above we could again take a baseline reading of the production of protein by a certain gene. Then we could look at the affects of other regulating regions of DNA near the gene of choice by gathering a testing set of these regions and looking at protein production for some ‘wild-type’ sequences of these regions. Then we could systematically mutate these regions, or implant mutant genes, and look at the effects on protein production due to the mutations of these influential regions. Similar to the experiment above it would be easy to see how computational techniques could easily regulate the mutation of regulating regions and systematically document the effects of these mutations as they relate to protein production from the target gene.

Section 3

The most basic explanation of why virii use the same regulatory factors is that in a biological context: “if it isn’t broken don’t fix it.” What I mean by this is that evolutionary pressure tends to push organisms to produce and utilize the most energetically and metabolically efficient versions of genes and proteins. So since virii have been around far longer than man has they have very well developed mechanisms for replication and function. Virii also have the ability to mutate much faster than eukaryotic organism as so the combinatorial space for the functioning of their genomes has probably been through far more changes and combinations then ours. From this it is easy to see how out of the incredible amount of possibilities that could arise from the mutations in their genomes the most efficient one would be environmentally pressured into becoming the most prevalent one.

Another explanation that falls under the same idea is that the most common virii inhabit and infect relatively similar organisms, in this case I am talking about mammals. With that in mind, virii function by hijacking our own metabolic processes in order to proliferate. Since they infiltrate organisms with similar molecular machinery then it only makes sense that virii would have in their genes regulatory regions that can be recognized and used by our own transcription factors. This is also the most evolutionary efficient way for a virus to replicate its genes and its own viral machinery because it doesn’t need to code for or come packaged with its own transcription factors, instead it can simply use ours.

Lastly since virii are all, on the most basic level, just mutated versions of one another or of some distant ancestor, albeit of course vastly different versions, then it would make sense that they would inherit the same genes for regulation and recognition of transcription factors. Virii that inhabit bizarre environments make up only a small fraction of the total viral population and so those will most likely have within their genome different transcription regulatory regions, because they are far mutated form virii that inhabit more common environments. But since a virus that inhabits our body, regardless of how it affects our body came from some relatively similar ancestor then of course it would inherit the mechanisms of that ancestor. As described above environmental pressure pushes molecular machinery to be the most efficient, with that in mind then there is no reason for a new generation or population to change the mechanisms of regulation if they already work well enough to allow them to thrive in the same organism.