PERSONALIZED GENOMICS
NOTES FOR ASIC200 (PARTS 1 and 2)

(Get students to pre-read "Breakfast of Champions for Replication" before 1st class. Note that this piece is also reprinted here within the notes when brought up during lecture)

Note that text in red not required for exam purposes, although hopefully interesting and pertinent to provide background in how certain things are done. Plus, some of it is just cool to understand (just maybe not to memorize).

CONTEXT

What is personalized or personal genomics? Take definition adapted from Wikipedia:

Personal genomics is a branch of science where individual genomes are analyzed and characterized using computer tools.

Essentially, this suggests that the ability to have information on our own individual genomes, allows us to better understand our own personal biological makeup. This has important ramifications in characterizing, predicting, and possibly *controlling* anything that is influenced by our genetics.

Silly: "Geneticscope"

DNA (BASICS)

First: a slew of definitions: (all adapted from wiki)

DEOXYRIBONUCLEIC ACID or DNA is a molecule that contains the genetic instructions used in the development and functioning of (almost) all known living organisms.

NUCLEOTIDES are molecules that, when joined together, make up the structural units of DNA.

A GENE is a unit of heredity in a living organism. It normally resides on some stretches of DNA and RNA that codes for a type PROTEIN that has a FUNCTION in the organism.

The FUNCTION of a GENE PRODUCT can (by itself or in tandem with other GENE PRODUCTS) result in an observable PHENOTYPE or TRAIT.

- - -

So what exactly are the key features of an organisms DNA? Well, central to this is the idea that the DNA contained within an organism (the genome) is analogous to a blueprint for the construction and operation of that organism. In other words, the DNA is very much like an instruction manual – a very detailed and voluminous instruction manual.

No offense to all the wonderfully talented individuals in the world, but Mother Nature has really outdone herself here with a rather superb job of getting this genome business to work. It is nothing short of amazing.

This instruction manual is basically a code that is written in the language of a molecule called *deoxyribonucleic acid*, (this here is our *DNA)*. DNA is this rather pretty looking molecule that is composed of four different building blocks. Together, these building blocks are known as *nucleotides*, and individually they each have a chemical name which is often abbreviated with a single letter - these letters being A for *adenine*, T for *thymine*, C for *cytosine*, and G for *guanine*. In effect, your DNA code is much like a language, a script of sorts, with the principle difference being that it is composed of only four letters instead of the full twenty six.

But what exactly does it code for? – PLAY TRUE/FALSE GAME.
Re: genetics of MATE SELECTION | LONGEVITY | INTELLECT | BRUSSEL SPROUTS

As alluded to earlier, a classic example of what your DNA code is capable of doing is the textbook case of natural eye colour. Your eyes are a certain colour because of the instructions within your DNA. The same is also true for your natural hair colour, and in other school examples such as whether you are able to roll your tongue or not. However, it's important to realize that virtually *every* physical attribute you have is determined by your genetic makeup. In other words, this also includes subtle nuances like the fact that some of your acquaintances are more prone to farting when they ingest dairy products or that a few of your friends may get drunk more quickly than others. Taken together, this means that your DNA is responsible for an awful lot of information, which at first glimpse is difficult to fully appreciate.

To put this all in perspective, it's important to try and visualize the enormity of the task at hand. One good way of doing this is to concentrate and focus on one of your thumbs. Ask yourself a few simple questions. How does your thumb know that it's a thumb? How does its cells distinguish themselves from the cells of other fingers? How does it know to come out of a certain place on your hand – next to your forefinger, not next to the pinkie finger? For that matter, how does it even know that it should be protruding from your hand and not from your foot? In truth, these somewhat bizarre thoughts centre round a field of research known as developmental biology. These sorts of scientific questions are constantly asked in this dynamic field, although not necessarily always for the thumbs - rather for the architecture of the entire body, or even possibly some other creature's body. In essence, these biologists continually think about the following question. How do we go from a single cell entity, created from a marriage of a sperm and an egg, to a being of very set features, full of many different types of cells and many different types of tissues? If you look around you, distinct though we are from each other, we are all basically the same. What I mean here is that generally speaking (and I hope I don't offend anyone), we all have heads, we all have torsos, and so on and so on. Furthermore, all of these bits and pieces are usually in the right sorts of places.

You must remember that it is your genome that is providing and directing all of this information. Imagine doing this yourself with pen and paper. Think of all the countless notes and scribbles you would need, so that something as basic as your body shape is done properly. For example, you may need to devote a few pages to your eyebrows. You would need to ensure that your eyebrows are in the right place. Not anywhere unsettling like your nipples, but somewhere on your face. Over your eyes and not under your eyes. You would need to describe their thickness, their length, and their colour. Hopefully, you get the point - the details are endless. Simply put, the amount of physical information in your DNA code is mind boggling.

And it doesn't even stop there. Although a bit more controversial, it is becoming clear that an individual's general behaviour and personality is, in part, predetermined by your DNA. The game we just played which essentially covered the genetics of things like LOVE and INTELLIGENCE easily demonstrates this. Obviously in this case, a person's environment and experience plays a vastly more dominant role, but there is nevertheless plenty of evidence to suggest the importance of genetic factors in these types of traits.

SOME NUMBERS

A fairly good estimate of the size of your genome is a total of 3.0 billion letters of code.

It's worth noting that this is actually a huge number, the scale of which I find is often lost to the casual listener. You get habituated, I think, by references to the country being in debt so many billion dollars, or by certain athletes signing billion dollar contracts. This is, matter of factly, a *very* big number and many other analogies abound that are much more eloquent than my shit example. If we were, for instance, to take 3.0 billion nucleotides and translate them into text, letter for letter, the genome in its entirety would be equivalent to about 8000 copies of the first Harry Potter book. Another one is to take 3.0 billion grains of sugar and pile them up in one spot. Apart from wasting a lot of sugar, you would discover that you would've formed a mound about the size of three cars. And we could even say that piling 3.0 billion cars on top of each other would likely resemble a mountain of Everest proportions.

Regardless of the analogy you use, the shear size of your genome does make sense. It is, after all, responsible for so many things, and you would assume that you would need that much information to get all the details and all the nuances sorted out.

ALSO note that actual “letters” of code per person can be defined by other ways. For instance, the code is double stranded, so in effect, you actually have 6 billion letters.

Also, we’re organisms that are DIPLOID (meaning that our code technically is a dual set. Almost all of our cells have two sets of double stranded DNA code. One from your Mother and one from your Father). This pushes the number up to 12 billion!

CODE = PHENOTYPE

How exactly does DNA code translate itself into an observable characteristic, a phenotype?

Here, we need to go over the CENTRAL DOGMA. Essentially a mechanism of going from “instruction booklet” to an actual “product.”

For now, ignore the bit about RNA

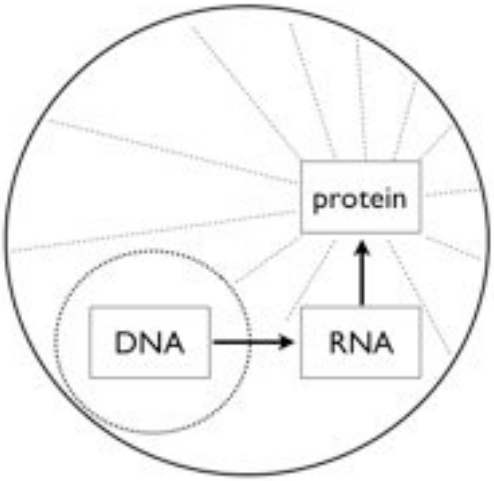
...Which is very much about proteins: in fact Martha Stewart would say, "proteins are a good thing." In life, they are the true movers and shakers of any organism, and are the molecules that actually go through the daily business of living. In short, these are the building blocks that give your various tissues their shape and their function. I bring this up now, because all of this talk about genomes and DNA is only illuminating if you recognize the fact that your genetic code is simply the instructional package for making all of the different types of proteins needed for life.

And things you need for life include: proteins that regulate chemical reactions (for instance, in the conversion of the food we eat into energy); proteins that transport key molecules from one place to another (like the pump implicated in cystic fibrosis); proteins that become the basis of cell structure (like how the architecture of certain tissues is achieved); proteins that facilitate cellular communication (how all the different bits and pieces of your body can work together). In truth, the diversity in protein makeup is responsible for all the diversity in life itself. In other words, bring on the bacon.

In itself, how proteins come about from your DNA code is quite clever. Proteins are built by piecing together molecules that are collectively called *amino acids*. It's a bit analogous to DNA in that if you recall, your DNA code consists of specific combinations of nucleotides. However, whereas our DNA is composed of an alphabet of only four different letters (A, T, C, and G), proteins are built with a much larger alphabet of 20 different letters or 20 different *amino acids.* Again, for any given protein, the determination of which of the 20 amino acids to use and in what order they are to be pieced together is dependant on the nucleotide code itself. This might sound a little confusing but in essence the production of proteins is dependent on dealing with two types of code. More specifically, each combination of *three* nucleotides (often referred to as a *codon*) will signify a particular amino acid. For example, the nucleotide T, followed by a G and another G (or TGG) codes for the amino acid Tryptophan (abbreviated ‘W’), the sequence ATG codes for the amino acid Methionine (abbreviated ‘M’), and so on. The sequence TGGATG would therefore code for two adjacent amino acids, Tryptophan and Methionine. Altogether, there are three letter codons for all 20 different amino acids. In this manner, a long sequence of nucleotides can potentially and theoretically be translated into a long chain of amino acids - i.e. a protein molecule.

To illustrate this two code system, the best example that comes to mind, is the use of Morse code to send messages overseas. In this situation, you essentially have a binary code (dot or dash, two options), which when rearranged into units of three, can translate into one of the 26 letters. For example, *dot dot dot* is the same as the letter 'S', and *dash dash dash* is the same as the letter 'O'. This is a two code system. Your first part being the Morse code element, and the second part being the formation of words from letters. In our biological example, the first code involves the use of nucleotides to provide information for which amino acids to use, whereas the second code dictates the length and combination of amino acids to form a functionally relevant protein.

(Now let’s take back the bit about ignoring the RNA)
In truth, the relationship between proteins and DNA is a little bit more complicated. First, it turns out that the overwhelming majority of the human genome doesn't do anything, and is basically considered to be garbage, junk, filler or if you want to be particularly nasty, crap. This accounts for an astounding 97% of your genome having absolutely no function or significance. This introduces an interesting logistical problem in that humans are using what is essentially a polluted genetic code. In other words, there has to be a system that allows the deciphering of the good stuff from the bad. You don't want to waste your time decoding your junk regions in that it could translate into some random, useless or potentially harmful protein.



Even your freakin' earlobe is doing this right now!

Secondly, the location of your DNA and the location of protein synthesis are different. This of course, makes no sense because how can you translate your DNA into proteins if the two molecules reside in geographically distinct places? Here, we find that your DNA is found within a small physically enclosed area of the cell called a *nucleus*, and proteins are awkwardly made *outside the nucleus*. Although this nucleus could be viewed as simply a mechanism to "house" and protect your genomic DNA, it does create a rather unfortunate conundrum in that the all important DNA code is not accessible to the machinery necessary for its translation into proteins.

In a rather crafty way, biology has managed to solve these problems through the use of a middle-man known as the *messenger RNA* molecule or *mRNA* for short. For the sake of clarity, mRNA is fundamentally similar in structure to DNA having nucleotides. There is a slight difference but it's visually quite minor - it could actually be the basis of a challenging 'spot the difference' comic. However, thinking conceptually, mRNA is comparable to a no-nonsense piece of genetic code that is constructed from only the useful parts of your DNA. This is similar to having study notes for a particular subject where only the crucial parts are highlighted and regurgitated. Consequently, problem one is solved. Here we have a strategy that can weed out the good from the bad and hence no crap.

Additionally, mRNA is special in that it is a string of nucleotides with the ability to move and ultimately leave the nucleus. You have to remember that your genomic DNA living inside the nucleus of a cell is akin to an elephant stuck in the upstairs toilet. It is simply too big to pass through doors that might otherwise be situated along the walls of the nucleus. mRNA molecules do not need to be so big. They are much more manageable in size because for each molecule, they contain only the sequence of one protein (not all of them), and more importantly they contain only the *necessary* sequence of that one protein (no junk). This means that problem two is also solved, as mRNA acts as a mobile representative of the genetic code that can now get out and come into contact with components required for protein production.

NOW also consider that only specific DNA sequences, (specific GENES) get copied into RNA pieces (this process is called TRANSCRIPTION – The RNA is "transcribe" from DNA), which in turn are TRANSLATED into proteins. i.e. an earlobe cell will only follow the central dogma path of information for things relevant to "being the earlobe cell" for specific protein necessary

for the “earlobe to BE.” You can imagine that different DNA GENE sequences are transcribed to RNA then TRANSLATED to different proteins in eye cells, or heart cells, or liver cells, etc.

- - -

Confused? Don't worry, it's alright if it seems a little perplexing right now. I know many people who have had nightmares over this stuff. If you do find yourself waking up in the middle of the night screaming nonsense about RNA and elephants, try thinking of the following analogy.

Because you are such a wonderful person, you wish to prepare a nice chocolate cake for your friend, and to do this, you visit the library to look for a good cake recipe. For some unexplained reason, you are also a huge Martha Stewart fan, which is why you decide to look for a cake recipe in one of her many 'Martha Stewart Living' magazines. After searching for several hours, lo and behold, you find a promising recipe in her 'Weddings Issue', but notice that the magazine itself has a sticker on it that says 'for reference only.' This is a bit of a bother because it means that you won't be able to take the magazine out of the library, and hence, into your kitchen where you had plan to spend most of your time being a wonderful person. Furthermore, despite your best efforts, you can't seem to find any semblance of a photocopy machine anywhere, since this is the sort of library this is, and since the analogy wouldn't work otherwise. Begrudgingly, this small nuisance forces you to look for a pen and a piece of paper so that you can manually scribble down the recipe to take home. As you do this, you quietly think to yourself that your friend had better appreciate all of this effort.

No offence to Ms. Stewart, but I find her magazines are always full of extraneous and in my opinion useless information. Do you really need to know the history of the chocolate cake? Do you really need to know about the appropriate cutlery used for serving cake? Do you really need to see and evaluate 15 different colour schemes for acceptable presentation? I don't think so. All you really need to concern yourself with is the ability to make the cake taste good. This is why, when you go to the bother of copying down the recipe, you don't include all of the nonsense - you just copy down what you need. In short, this turns out to be just a few lines of ingredients and directions scrawled neatly on your piece of paper. The crucial point is that you can now freely walk out of the library with the recipe in hand.

Next, of course, is a trip to the local grocery store where you would get all the necessary cake ingredients and maybe indulge yourself with the smutty magazine about child actors gone bad. After which, you would head home and bake a wonderful chocolate cake which is met with such praise, that you are glad you didn't waste your time using table setting number six for the occasion.

A strange story indeed but here is how the analogy works. First, you need to envision the entire library with all of its resources as the genome, and also envision the building itself as the nucleus. The complete recipe found in the magazine actually represents the genomic sequence for one particular protein. As mentioned before, Martha publications tend to have a lot of useless information, some of which is not even directly related to the production of the cake (advertisements and historic footnotes). This is identical in premise to the crap in your genomic material, and the concise notes you scribbled down symbolize the messenger RNA molecule. This, as mentioned before, is twofold important because, (1) it represents the minimal amount of information needed and, (2) it represents the ability to leave the nucleus (in this case, the library) and the ability to go to places where protein production can take place (in this case, the rest of the world, but more importantly the grocery store and your kitchen). Finally, the cake itself represents the protein. Remember I said that in living systems, it's really the proteins that are the

real movers and shakers? They are certainly the most interesting parts of the big biological picture, and wouldn't you say that the cake itself is the most interesting part of this process?

ANYWAY, this is some of the basics of genetics (replication, central dogman, DNA, RNA, protein, genes, genome). This stuff is really powerful, and clocking along at a phenomenal speed.

(BREAK)

MOLECULAR BIOLOGY

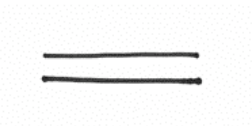
This is a word that essentially covers a science discipline concerned with the act of looking at the molecules involved in biological processes! i.e. How to study DNA, RNA, and proteins!

Let's look a little more closely at DNA, by first going over the act of DNA replication. This, here, is the process that allows DNA to make copies of itself. Important because as cells grow and divide (aka multiply), the code needs to be maintained, and therefore new copies need to be produced for newly produced cells.

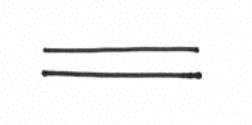
- - -

AND NOW... without further ado... ASIC 200 presents: "REPLICATION."

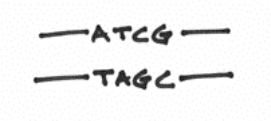
(Use magnetic board)



To begin with, we'll start with a chicken scratch drawing of a DNA molecule, which you know is double stranded. My poor pathetic attempt at illustration is therefore going to look like this:

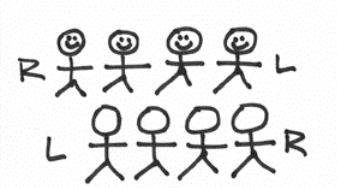


You also know that each strand of DNA is composed of building blocks called nucleotides, and that these nucleotides are always interacting in a complementary manner. For example, A's are always with T's, C's are always with G's, Beavis is always with Butthead, etc etc etc. Let's draw them in like so:



What you haven't been told at this point is that chemically speaking, the two strands are going in opposite directions. The correct term for this is actually known as *anti-parallelism*. To denote this, I'll draw some arrowheads on the DNA strands:

Although, this may seem a little confusing at first, try to picture two lines of square dancers facing each other. In this circumstance, you notice that when focusing on the left or right hands of the row of dancers, the two lines are going in opposite directions. This picture should help:

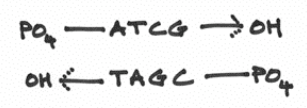


Your DNA strands are doing something very similar in a chemical sense. The difference, of course, is that instead of dancers, you have your choice of four nucleotides. Furthermore, like the situation of left hands versus right hands, the ends of the DNA strands are also different. One end is known as the *3'* (pronounced 3 prime) end and the other is known as the *5'* end. To the layman, these rather stoic terms are an unfortunate consequence of chemical labeling. So now, our picture should look like this:

I should reemphasis that the 3' and 5' ends are very different from each other. To be more specific, we say that they are chemically distinct from each other. They are as different from each other as apples and oranges. In fact the 3' end is composed of a *hydroxide group* and the 5' end is composed of something known as a *phosphate group*. These groups look a little like this:

Hopefully, it's easy to see that they are indeed distinct from each other —even more so than apples and oranges. The hydroxide group being comparatively small and meek, whereas the phosphate group is prominent, overbearing even. This turns out to be a crucial factor because replication is carried out by the activities of a variety of different enzymes which *all* function by focusing on one DNA end or another or both.

So now, the picture looks like this:



It should also be pointed out that DNA is not really like this flat goofy looking cartoon. As mentioned in a previous chapter, the two DNA strands are actually intertwined around each other in a rather pretty helical fashion. This is where the two strands are wound around each other, sort of like two elastic strings twisted and coiled together. Sort of like this:

Now that the stage is set, it's time to introduce the proteins or the *enzymes*, which are responsible for the actual process of replication. Enzyme is just a fancy word for a protein that is able to facilitate a chemical process. What I'll do here is to focus on terminology associated with a simple organism like the bacteria, *e. coli*. However, all organisms, even those as complicated as humans, do more or less the same thing when it comes to doubling their DNA — the principle difference being that unfortunately, the enzymes have difference names and labels.

That aside, the first enzyme for replication in *e. coli* that we should introduce is, of course, the most important enzyme in the entire process. In *e. coli*, this enzyme is called *DNA polymerase III* (or DNA pol III for short), and is essentially the one that is responsible for the actual business of making more DNA. If this entire exercise was analogous to a movie, then this enzyme is the marquee player. It is the Tom Hanks, the Julia Roberts, the proverbial bread and butter of replication. It is, quite simply, the star of the entire process. Instead of drawing a picture of Tom Hanks or a picture of Julia Roberts, I think a picture like this should suffice:

Problem is, if we were to draw this enzyme to scale with a helical DNA molecule (like this),

you'll notice that the DNA pol III is actually too big to get inside the DNA strands. It can't go about its business of copying the DNA, because the strands are all coiled up in the helical structure. In other words, there is a serious issue of accessibility. Even our star enzyme, despite its importance, can't do its job without access to the molecules of DNA it wants to copy. Consequently, the enzyme that inevitably has to act first is one that is responsible for opening up the DNA strand. This enzyme is known as a *helicase*, and its role is to essentially unwind the DNA molecule, which would look like this:

The net effect being the production of a "bubble" of opening where the two DNA strands are pried apart and are subsequently accessible to the whims of the replication machinary.

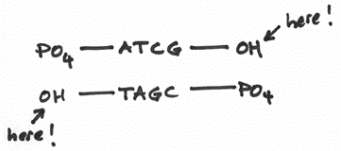
Curiously, the DNA pol III, which after the unwinding event, can now interact with the DNA molecules, does so whilst attached to a bunch of other enzymes. This attachment is a little like a bunch of buddies hanging out together. The complex actually looks a little like this:

You'll notice it has the following... (i) two DNA polymerase III's: which kind of makes sense given the fact that there are two strands of DNA that need to be copied; (ii) one helicase molecule: which also sort of makes sense, because as this replication complex is doing its thing along the DNA molecule, wouldn't it be handy to have the built-in ability of opening up the DNA molecule as it moves along; and (iii) one new enzyme which is known in *e. coli* as the *primase*. However, the purpose of the primase molecule is a little complicated and so to fully comprehend the role of this enzyme, we need to switch gears a little and tell you a bit more about the DNA pol III molecule.

What actually needs to be done, is for us to go over a few mechanisms that all DNA polymerases seem to use. In fact, it's apparent that every DNA polymerase that has been discovered on this planet:

In fact they all (without exception) seem to follow a two basic rules.

Rule number one states that all DNA polymerases function by adding nucleotides to the 3' end of the DNA strand. What this means exactly is that a DNA strand can be extended by the addition of new A's, T's, C's or G's. However, the new nucleotides can only be added to one particular end, namely the 3' hydroxide group. This is a molecular restraint in that the DNA polymerase can only join nucleotides via this smallish chemical group. This rule can be drawn out like this:



Rule number two states that all DNA polymerases require a *primer* to function properly. This is probably the most challenging concept that needs to be addressed. If you get through this, then you consider yourself home free.

To simplify the notion of a primer, let's look at a single strand of DNA, complete with its 5' and 3' ends. It should look a bit like this:

Now according to rule number one, a DNA polymerase can extend this single strand chain but only by adding nucleotides to the 3' end. In effect, you can argue that all of the relevant chemical groups are present for making more DNA. However, the problem lies in the fact that under these circumstances, the DNA polymerase doesn't actually know what to add. How does it know, whether to add an A, a T, a C or a G? It can't exactly be a random event, because replication is all about making sure cells receives an identical copy of the DNA code.

Take the following picture:

Under this layout, it should be clear that now, the DNA polymerase has the required 3' group, AND it also has a *template* to read and ascertain what those nucleotides should be. For instance, if the nucleotide in the opposite strand is a G, then the DNA polymerase knows it should add a C. If the nucleotide in the opposite strand is a T, then the DNA polymerase knows it should add an A. Hopefully, at this point, you'll at least agree with the following statement. A DNA polymerase can not do anything with a single strand of DNA. True, it has the right chemistry, but in effect, it does not have the template or instructions needed to define how the chain is extended.

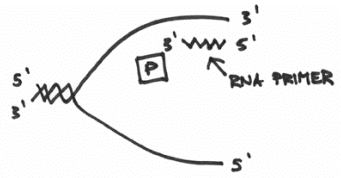
If we redraw the picture. Say like this:

What you'll notice are two strands of DNA, one long and one short. You'll also notice that the strands are anti-parallel as discussed earlier. If you focus on the arrowhead, you'll find yourself focusing on a perfectly situated 3' group. Here is the end of a DNA strand that is chemically ready to have nucleotides attached. Furthermore, it is also a 3' end that is located where a template is present on the opposite strand. In other words, everything is in place. The right chemistry, and a means for instructing which nucleotides to add. Again, taken at the simplest level, we can conclude that in order for a DNA polymerase to do its thing, it needs an area of double strandedness.

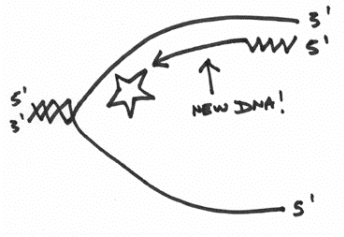
So,.. the small sequence of nucleotides that has been circled here...

... which makes an area of double strandedness is technically known as a primer. With this all sorted out, hopefully the rule about requiring this primer makes a little more sense, and you can probably guess that the enzyme called the primase may have something to do with this nuance.

Which turns out to be exactly what this primase enzyme is all about. In a nutshell, it is an enzyme capable of making a short sequence of nucleic acids which functions as a primer. A key point that needs to be emphasized, however, is that this primer is made up of RNA, which if you recall, is a molecule that is very similar to DNA in that it is also composed of the representative four nucleotide code. This is actually due to a biological technicality whereby it is possible to make a complementary strand of RNA without the use of a primer (hmmm, think about this for a second). Taken together, the function of the primase should end up looking like this:

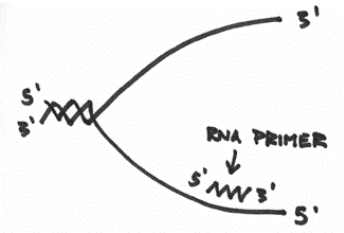


If you've been following along, then hopefully you can see that replication from this RNA primer can proceed in a manner that can be drawn like this:



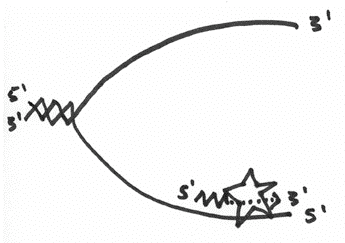
However, it's wise to pause here for a second, because you have to understand that whilst this top strand is being replicated, the lower strand is also being worked on simultaneously (There are two DNA polymerase III's attached together afterall). The lower strand is actually a bit messier for reasons that will become clearer as we proceed in this discussion.

Basically, the primase enzyme will also go about preparing a primer for the lower strands. However, if we draw this primer and label the ends in the anti-parallel manner, you can hopefully see a logistical problem in this set-up. Take a look at the following picture, and see if you can find the problem (remember, the DNA polymerases, the helicase and the primase all move as a single unit in one direction, and remember that all DNA polymerases must add to the 3' end):

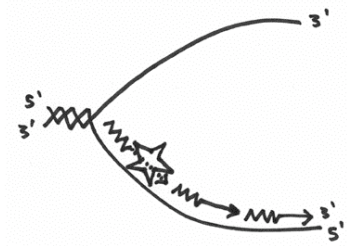


Do you see the problem? Do you see a problem with the direction of the primer? Do you see that the 3' end of the lower primer is facing the wrong direction?

This is obviously a problem, and it turns out that in order to overcome it, the DNA polymerase will still add nucleotides to the 3' end, but can only do so for a short distance. To keep it simple, think of it as being able to replicate as far as the enzyme is big, which should look a little like this:



Unfortunately, this doesn't inherently solve the direction problem, so what ends up happening, is that with this lower strand, the primase has to continually make a primer, and the DNA polymerase III has to continually replicate a little bit. In the end, it should look like this:



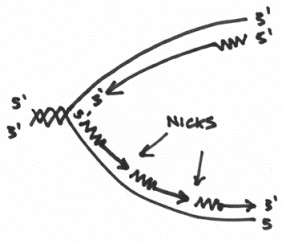
The difference in how each strand gets copied is reflected in why some people call them the *leading* and *lagging* strands of replication. One strand is obviously fairly straight forward whereas the other is quite labour intensive.

Anyhow, after this is all said and done, hopefully, you'll agree with the following statement. That is, we have finally doubled or copied our genetic sequence. However, it should also be clear that the whole thing is a bit messy. For instance, there are bits of RNA everywhere, and the lagging strand is composed of pieces. To address these problems, we have to introduce a few more enzymes.

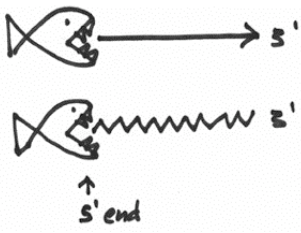
The first of which is *DNA polymerase I*, which I will draw as a fish with sharp teeth. This enzyme is special in that, in a nutshell, it is responsible for dealing with the RNA. In a nutshell, its job is to somehow replace it with DNA. In a nutshell, I'll draw it like this:

DNA polymerase actually has two distinct functions. Firstly, as its name implies, it is a DNA polymerase, meaning that it is capable of extending the DNA chain, but in doing so must follow the same two rules that govern these enzymes. In other words, it must add nucleotides to the 3' end and it must use a primer as a springboard. Ironically, it is a shitty DNA polymerase. Whereas DNA polymerase III can replicate for several hundred nucleotides, DNA polymerase I has difficulty getting past a few dozen.

Secondly, DNA polymerase I is also an *exonuclease*. This means it's capable of degrading or chewing up nucleotides. Which is another reason why I drew a fish with teeth. And not only does it chew stuff up, it does so in a fairly specific manner. To begin with, it likes to start at areas, which are termed as *nicks* in the DNA. In our picture, this is where the nicks would be:

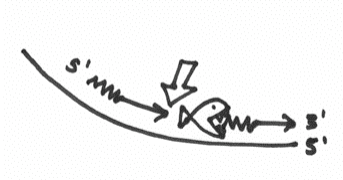


Furthermore, this exonuclease is picky in that it always chews from the 5' end. Basically it is gunning for that big phosphate group. So that you don't forget this, I've drawn this picture to help you visualize this:



Now, if you take all of this into consideration, you come up with the following mechanism. DNA polymerase I will come in on our replication picture, and zone in on a nick in the strands. Once there, it will begin chewing on the 5' end, which should look a bit like this:

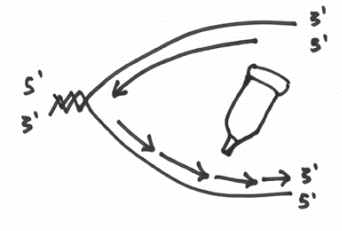
Don't forget that this enzyme is also a DNA polymerase, and if you look at the other side of the nick, you will hopefully realize that there is this beautiful 3' end ready for action. This beautiful 3' end is right here:



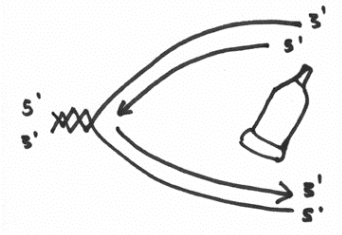
Let's say that the fish's ass happens to contain the DNA polymerase function. What therefore happens is that DNA polymerase I will start replicating from that 3' end, which incidentally fills up the gap that was created by the exonuclease activity. This should nicely demonstrate how DNA pol I achieves its function of replacing the RNA with DNA. This whole step should kind of look like this:

Hopefully, this puts the shittiness of this DNA pol I in perspective. It's quite biologically pretty because, I hope you can appreciate that DNA pol I doesn't need to be very good. It's only responsible for replicating the small region encompassed by that RNA primer.

So,.. after this enzyme has done its thing, you should now agree with the following statement — that you have now doubled your DNA. Of course, it's still a bit untidy because the strands (especially the lagging strand) are still in bits and pieces. Enter the next and final enzyme, which is called the *ligase*. This enzyme has only one job and that is to seal all of the bits and pieces together. It fairly analogous to a glue job and essentially your picture will go from something like this:



To something like this:



And (drum roll please) VIOLA! You have *doubled* your DNA. You have made two copies of the same genetic code - which during the process of cell division, will enable each of the two new cells to receive a copy of the genome.

One of the nuances that should be mentioned is that if you examine the entire process, you will notice that each of the DNA sequences is derived from one old strand and one newly synthesized strand. Because of this, replication is often termed *semi-conservative*, whereby each of the original two strands is read individually to synthesize a new and complementary strand.

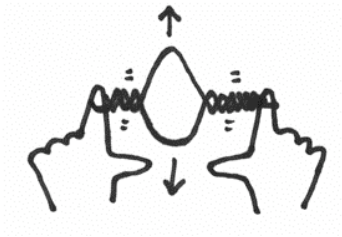
* * *

Actually, I lied. It's not quite over. Before, I finally put this whole replication thing to rest, I think it's also worth talking about one other enzyme, or a family of enzymes, known to scientists as *topoisomerases*. I like mentioning these enzymes, because I think they do a wonderful job of illustrating just how complicated and elegant nature is, when confronted with a specific job.

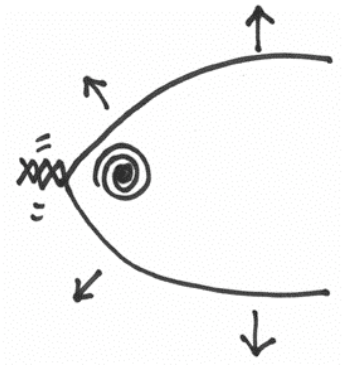
What we'll need to do here is undergo a visual exercise. Let's say I tell you to hold two fingers up like this:

And let's say that I have an elastic band. With this elastic band, I will twist and coil it and then place it around both of your fingers. Essentially, this will represent the double helix and will look a bit like this:

If you recall, the first thing that had to happen was for a helicase enzyme to come in and open up that helix structure. Let's say that I am the helicase, and I come in and grab hold of the two strands of your elastic band and pry them open. It should make a little bubble and should look a little like this:



Can you see that under these circumstances, the helix on either side of the opening will be actually twisted even more. It would be like taking your replication fork, grabbing hold of each strand, and like the helicase forcing an opening like this:



Do you see that this will cause a further tightening of the coil along the helix?

This is actually very bad for the DNA molecule, as this twisting can cause a lot of structural stress. So much so, that the DNA molecule is in very real danger of snapping - which you can imagine would be a very bad thing to happen during replication.

Topoisomerases are enzymes that are designed to take care of this problem. These enzymes can actually detect these areas of high structural stress, and zone in on them. Not only that, but whilst they are at these areas, they will then cut both strands in the DNA complex. Remarkably, they will then hold on to all four ends of the cut, and in a very controlled fashion, unwind to alleviate the stress. Finally, they will also behave like ligases and stick back the correct ends together again.

This is nothing short of amazing, but hopefully you can see that these enzymes play an important role. As the DNA is opening up for replication, there will always be an issue of structural stress, which is always addressed by the actions of these remarkable enzymes.

- - -

There are a variety of different powerful techniques that allow a person to do MOLECULAR BIOLOGY. We'll cover 3, two in lecture class and one in the lab. NOTE that none of the 3 will be tested for on the exam.

In class, we'll cover GEL ELECTROPHORESIS and SEQUENCING! (The third that will be covered in the lab will be the POLYMERASE CHAIN REACTION)

*GEL ELECTROPHORESIS is a methodology that allows the characterization of molecules by their size. In short, it involves to creation of a gel, which (if you were to shrink to microscope man/woman) would look a little like a sieve. Basically, if you're a molecule working your way through this sieve, the smaller you are, the easier time you'll have of moving through it (i.e. the smaller molecule can move faster and further, than say a larger molecule which gets impeded by the sieve).

To make this sieve, you can use a chemical like agarose, which in pictionary land would look a lot like a long string. NOTE: in class, we'll use the 'training a hamster as my Ph.D.' analogy with my 'big ball of string' analogy.

*SEQUENCING: (will go over – goal is that students whilst listening will "get it" and understand how an experiment can elucidate a genetic sequence)
Sanger method: Also known as chain termination, because it relies on the process of replication but with the addition of special nucleotides known as DIDEOXYNUCLEOTIDES. The key here is that this type of nucleotide is basically the same as normal nucleotides (the A, T, C, and G's), except that it is also missing its 3'OH group. If you remember from the DNA replication stuff, this 3'OH was actually very important for replication to occur. i.e. when you're extending the chain, and a dideoxynucleotide gets in, you force the termination of replication, because the polymerase has no OH group to work from.

(see the following URL:

http://www.bio.davidson.edu/Courses/Molbio/MolStudents/spring2003/Obenrader/sanger_method_page.htm

for a full explanation of Sanger technique. You know - if you're also interested in memorizing it for cocktail party conversation or something)

(END)

1. CONTEXT

PERSONALIZED GENOMICS

Personal genomics is a branch of science where individual genomes are analyzed and characterized using computer tools.

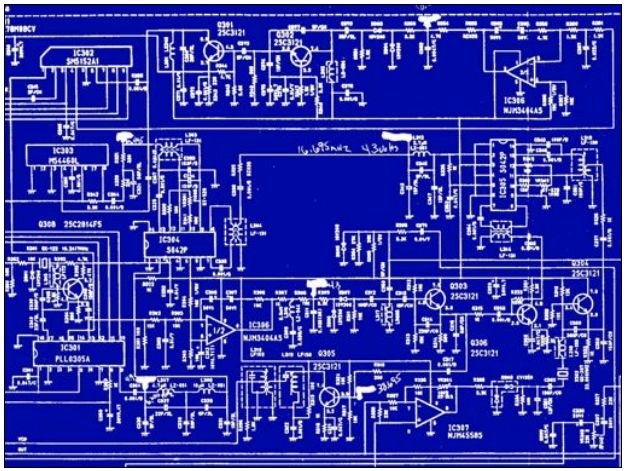
YOUR GENETISCOPE:

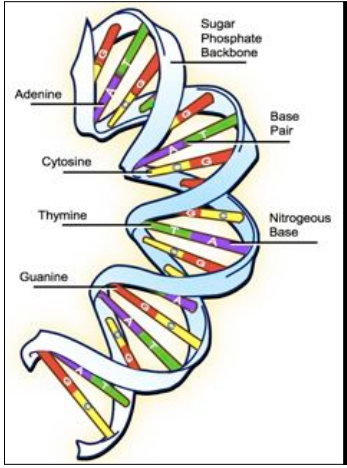
YOUR DNA SEQUENCES ALONG WITH ENVIRONMENTAL PARAMETERS INDICATE THAT YOU WILL HAVE A GENERALLY PLEASANT DAY. WATCH OUT FOR THAT INTERSECTION BETWEEN 4TH AND MAIN, AND NOTE THAT YOU WOULD DO WELL TO AVOID DRIED MANGOES TODAY. ONE OF YOUR COWORKERS WILL PASS ON A WHITE LIE, PROBABLY DURING THE AFTERNOON.

BOWEL MOVEMENTS AT APPROXIMATELY 07:00 AND 18:00 HOURS.

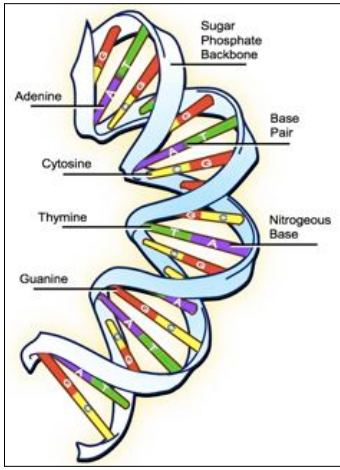
YOUR GENOMIC PROFILE ALSO INDICATES THAT A GREY OR WHITE CAT WILL PISS YOU OFF TODAY.

2. DNA (BASICS)



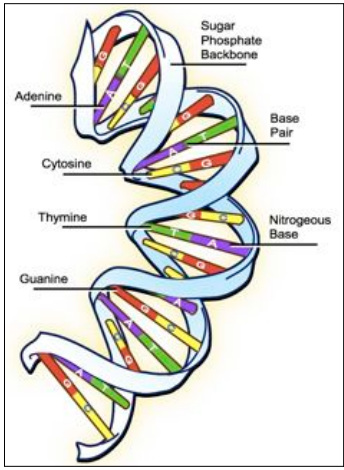


DEOXYRIBONUCLEIC ACID or DNA is a molecule that contains the genetic instructions used in the development and functioning of (almost) all known living organisms.

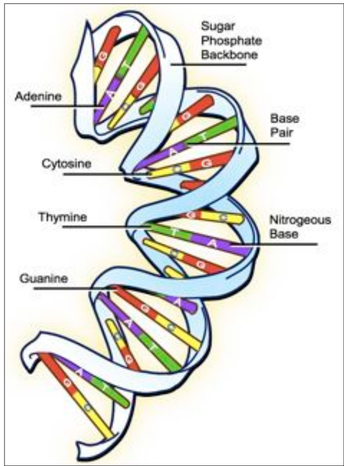


NUCLEOTIDES are molecules that, when joined together, make up the structural units of DNA.

ATCG



A GENE is a unit of heredity in a living organism. It normally resides on some stretches of DNA and RNA that codes for a type PROTEIN that has a FUNCTION in the organism.



The FUNCTION of a GENE PRODUCT can (by itself or in tandem with other GENE PRODUCTS) result in an observable PHENOTYPE or TRAIT.

TRUE OR FALSE?

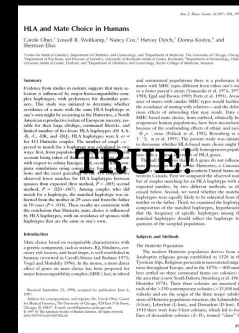
1. Evidence shows that mice are attracted to their mates based on genetic diversity. This they can somehow tell from the smell of their urine. There is also currently some weak evidence that humans indirectly do the same thing.

2. According to established data within the senecense field, people who take longer to "sh*t" arguably should live longer.

3. A transgenic mouse, affectionately known as the "Doogie" mouse has been produced with superior intellect and mental prowess.

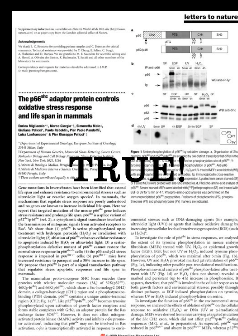
4. Research based on psychogenomic analysis have determined genetic sequences involved in your love/hate relationship with brussel sprouts.

Evidence shows that mice are attracted to their mates based on genetic diversity. This they can somehow tell from the smell of their urine. There is also currently some weak evidence that humans indirectly do the same thing.

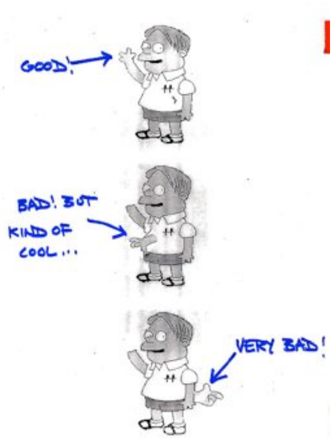


A transgenic mouse, affectionately known as the "Doogie" mouse has been produced with superior intellect and mental prowess.

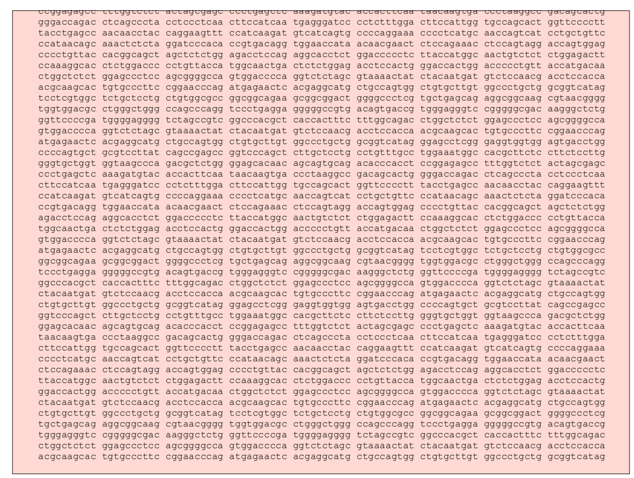
According to established data within the senecense field, people who take longer to "sh*t" arguably should live longer.

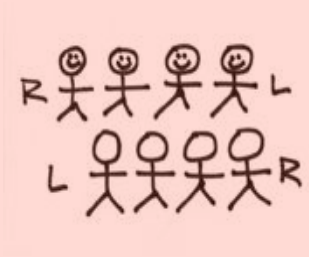


4. Research based on psychogenomic analysis have determined genetic sequences involved in your love/hate relationship with brussel sprouts.



3,000,000,000!



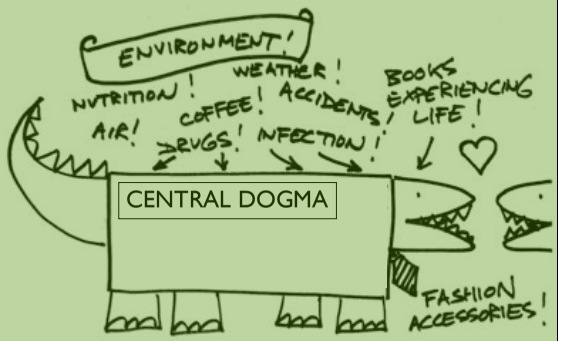


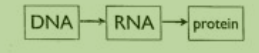
3,000,000,000! (code)

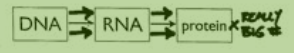
6,000,000,000! (double stranded)

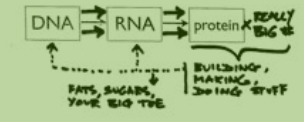
12,000,000,000! (diploid - two parents)

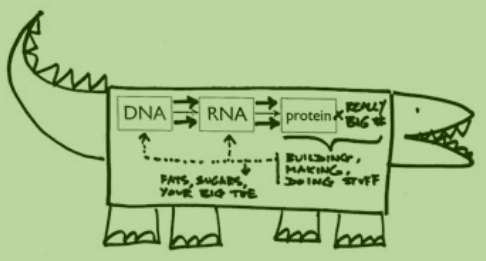
3. CODE = PHENOTYPE?

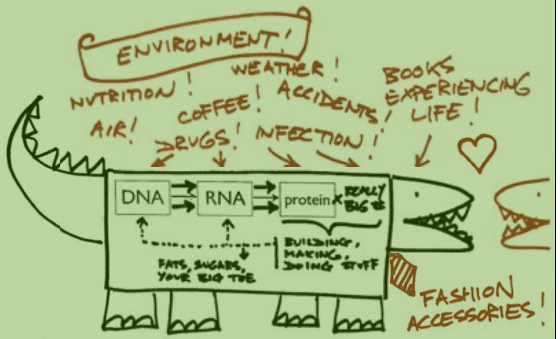


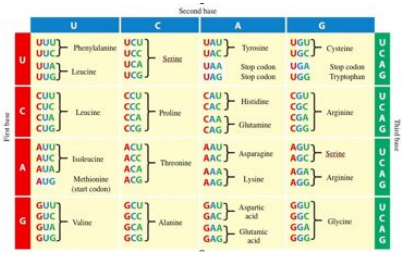










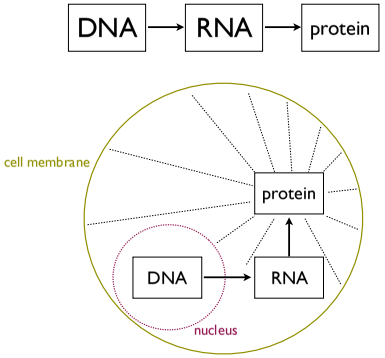


codon

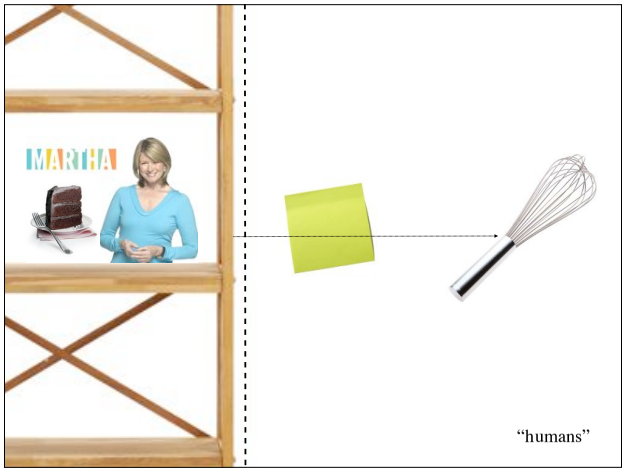
DNA GCC AGC AGT

RNA GCC AGC AGU

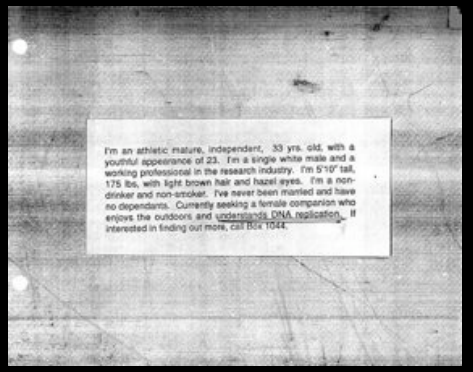
PROTEIN ALA SER SER

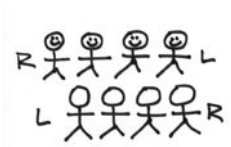


A cell in your earlobe (because earlobes never get the limelight in biology lectures)



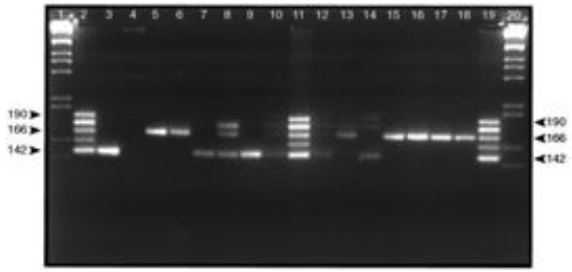
4. MOLECULAR BIOLOGY

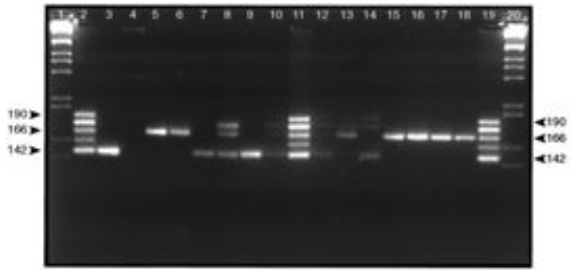




#MEGAFACEPALM

GEL ELECTROPHORESIS





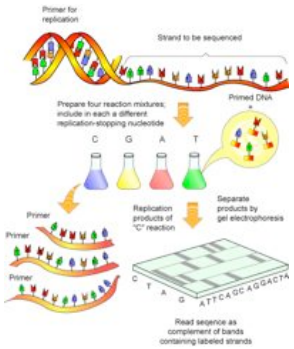


Figure 1. *The Sanger sequencing reaction. Single stranded DNA is amplified in the presence of fluorescently labelled ddNTPs that serve to terminate the reaction and label all the fragments of DNA produced.The fragments of DNA are then separated via polyacrylamide gel electrophoresis and the sequence read using a laser beam and computer.*

source: http://www.scq.ubc.ca/genome-projects-uncovering-the-blueprints-of-biology/