CHAPTER 2

LITERATURE REVIEW

2.1. Genomics

2.1.1. History overview

The story of molecular markers starts way back before the 19th century, when philosophers gazed into the theory of conception. It all started with Hippocrates with his theory of “Pangenesis” (Sturtevant, 2001), however Aristotle, the famous Greek philosopher, was one of the pioneers to sense the existence of an “active element” contributing to an offspring’s exterior characteristics (Lynn K.), although most of his theories were faulty, he kick started what would be a science of its own in current times. Theories of evolution were also proposed by Charles Darwin and Alfred Wallace, but their hypotheses were mostly unwelcomed and strongly opposed (Watson et al., 2004). With the discovery of the microscope in the early 1960s, the recognition of sperm and egg as carriers of nuclear material of inheritance was established. In the mean time, Augustinian monk Gregor Mendel, the “Father of Modern Genetics”, began testing the theory of inheritance on pea plants, yet his principles that laid the mere foundation of hereditary were not discovered until the end of the century. In 1903 Walter S. Sutton produced a paper “The Chromosomes in Heredity”, merged the disciplines of both Breeding and Cytology (Sutton, 1903). Later on in 1910, Thomas Hunt Morgan presented the principle of Linkage demonstrated after a breeding study on Drosophila.

Following a few more advancements in the science genetics and hereditary, the dawn of a new era began in with the fourth decade of the 20th century. Scientists with the likes of Beadle, Tatum, Garrod, Avery and team, Franklin and Wilkins and many more, continued the path of uncovering what was most likely to the discovery of the century,

the DNA double helix, which was revealed to the world in 1953, by Scientists Watson and Crick (Watson & Crick, 1953) . The genetic code was cracked in 1963 (Nirenberg, 1963) and ten years later, Cold Spring Harbor Laboratories enhanced the gel electrophoresis experience by using agarose-ethidium bromide electrophoresis (Sharp, Sugden, & Sambrook, 1973). Sanger broke new grounds with the start of DNA sequencing (Sanger & Coulson, 1975), followed by Maxam and Gilbert (Maxam & Gilbert, 1977). DNA amplification took off when Mullis invented the Polymerase Chain Reaction (PCR) in 1983 (Saiki, et al., 1988). Ever since then advancements rolled in and technologies evolved, handing us fast, economical and highly valuable molecular information, which gives us an insight as to what goes on in a cell.

2.1.2. History of Population Genetics and Evolution

As mentioned earlier Darwin was one of the first to show interest in the theory of evolution. His writings and postulations were always a source of controversy, even to this day. But of course he had his followers; Galton and Weismann are only two of many others that followed (Sturtevant, 2001). In 1908 Hardy and Weinberg both proposed notion that genotype frequencies should remain stable throughout generations of a population, they also came up with a formula that tests their null hypothesis in studied populations, which is now a prerequisite in any population genetics project (Hardy, 1908; Weinberg, 1908).

Nevertheless the brains that set the stepping stones and all the mathematical formulas that makeup the basics of population genetics are Haldane, Fisher and Wright (Li, 1955). In 1937, Dobzhansky One of the first to apply mathematical algorithms on an existing natural population (Drosophila) (Dobzhansky & Queal, 1937). Parallel to that

(in 1937) Nei addressed the drastic effects of the bottleneck phenomena on existing population and the strains it causes on the heterozygosity and genotypes of natural populations (Nei, Maruyama, & Chakraborty, 1975). Kreitman employed genetic variability reflected in polymorphisms in 1983; to study the variation in *Drosophila melanogaster* species (Kreitman, 1983). This was a first, since previous studies looked at the phenotypic characteristics of the bands on the gel electrophoresis and not the genotypes. Many advances followed and still much more is yet to come.

2.1.3. Applications of Genomics

Genomics on its own has a vast meaning that is difficult to summarize in a few words. It includes mapping, sequencing and identifying functional genes. In other words it is a process whereby it takes not only a lot of effort and the latest in scientific technologies. This science needs brain power to analyze the unknown.

The idea started in mid 1980s, when the desire to sequence the human genome arose; however the project did not take off until October 1990 and finally finished in 2003, with a total cost of $3 billion (J. Craig Venter Institute, 2004; Collins, Morgan, & Patrinos, 2003). This was a major advancement in the world of science, medicine and the pharmaceutical industry.

The first disease was mapped in 1983 using DNA polymorphism (Gusella, et al., 1983). Since then complete mapping of genomes in organisms as simple as *E. coli* (Blattner, et al., 1997), to more complexed organisms such as mice (Waterson, et al., 2002), rats (Gibbs, et. al., 2004), chicken (International Chicken Genome Sequencing Consortium, 2004), dogs (Lindblad-Toh, et al., 2005), *etc.* were carried out. One of the first publicized genome drafts after the human genome was the marine pufferfish *Fugu*

rubripes (Aparicio, et al., 2002), and comparative studies to the human genome have already been conducted; this gives a better understanding of the functionality of human genes, when compared to other vertebrates (Hedges & Kumar, 2002).

The utilization of the whole genome can come in handy in several aspects, such as:

- Quantitative trait locus (QTL) and nucleotide (QTN) mapping. QTL mapping proved very useful in plant breeding programs (Brummer, et al, 1997), and live stock improvement (Kim, et al, 2003).
- Single Nucleotide Polymorphism (SNP) genotyping.
- Gene expression and methylation profiling and RNA interference (Sellner, et al, 2007).

Marker assisted selection is studied for the purpose of detecting and improving QTL, and to give us an idea of how linkage and markers are actually related. It has taken off massively, mainly in the livestock industry (Dekkers, 2004), However, Zebrafish, as a model organism for freshwater fish, has also had many QTLs mapped, for the purpose of studying their neurological behavior (Wright, *et al*, 2006). It's worth mentioning that this model organism has had more than 634,608 Expressed Sequence Tags sequenced and therefore microarrays have been established for expression profiling (Dahm & Geisler, 2006).

2.1.4. Molecular Markers

Molecular markers have been around for decades, from allozymes to microsatellites, they have contributed to numerous disciplines such as: population and evolutionary biology (Avise, 1994), forensic science (Szibor, *et al,* 2003), pharmaceutical genetics and medicine (Joshi, *et. al.,* 2004), etc. What almost every researcher is mostly looking

for in marker studies is polymorphism. These give us hints as to changes occurring inside the cell; due to environmental effects, mutations, *etc*.

The search for the perfect molecular marker is ever going and for the past 40 years, biologists have been revolutionizing molecular markers in search for a perfect way to distinguish fish populations, and determine migration patterns (O'connell & Wright, 1997). Out of the massive number of markers, satellite markers are increasingly getting popular.

The first attempt ever to apply molecular on fish for population research purposes was in 1950s, using blood groups as a marker on fish such as salmon, tuna and cod (Kochzius, 2009).

Advances in this science have been so overwhelming that new techniques have taken the place of older ones. For instance allozymes are still a favorable marker, however it is not utilized as much as it used to before, compared to the current status of microsatellite markers. Nevertheless these “older markers” are still playing their part and are used coupled with the novel ones (O'connell & Wright, 1997).

2.1.4.1. Allozymes

The discovery of polymorphism in proteins came by in the early to mid 1960s, when many scientists (Harris being one of them), associated metabolic disorders in human with genetic deficiency of certain enzymes (Hopkinson, Spencer & Harris, 1963). A paper followed describing enzyme polymorphism in man (Harris, 1966). However in fish, polymorphism was first noted in 1961, when hemoglobin proteins were used to distinguish between discreet sub-populations of fish (Sick, 1961).

Frequency of alleles associated with allozymes was studied in 1974, on alcohol dehydrogenase enzymes *D. melangoster* (Bijlsma-Neeles & Van Delden, 1974). This is one of the first markers to be studied on fish species (May, Wright, & Stoneking, 1979) and is still employed to this day (Phillips, Storey, & Johnson, 2009).

Allozymes are Type I molecular markers, since they infer known coding regions, they represent variations in proteins in a single locus and are mostly codominant (Liu & Cordes, 2004). Isolating this marker is rather easy once the basic procedure has been established for a species; however polymorphism is very hard to come across when employing this particular marker. Polymorphism of allozymes in fish is found only 21% of the time, compared to other organisms such as Gymnosperm 58% (Parker, et. al., 2003). Changes or mutations in the DNA are impossible to predict when dealing with proteins. Electrophoretic resolutions are so low that they cannot uncover any minor changes in the amino acids. This resulted in the use of histochemical staining methods to effectively view proteins (Moller, 1970). Also samples for protein analysis have to be fresh and large amounts of tissue are needed which is inconvenient when dealing with endangered species (Kochzius, 2009).

On the other hand, when financially restricted and when dealing with loci that are highly variable and minimal differences can be ignored, this marker could be represented as the perfect candidate for the study (Parker, et al, 2003).

2.1.4.2. DNA based molecular markers

DNA based molecular markers are quite common in our times. We mainly owe that to its physicality since DNA is quite stable, even in dead tissue specimens. There is no hassle what so ever when it comes to DNA storage and when extracting usually small minute amounts are needed. Different regions or loci are subjected to different mutational pressures and most often, coding regions of important functional proteins are a lot more conserved, with less mutational rates. Therefore much thought has to be put in, as to which molecular marker is more suitable for a population genetic study. Considerations such as individual relations and interactions in a population, and also the selectional mutation on a particular region in the genome, are all very important when initializing a structural study (Parker, et al, 2003).

2.1.4.2.1. Mitochondrial DNA markers (mtDNA markers)

Mitochondria are intracellular organelles that are found in most of the higher organisms such as some plants, fungi and animals. It is considered the powerhouse of the cell, given that its purpose is to produce ATP (Castro, Picornell, & Ramon, 1998). This molecule (of which 93% is coding) encodes 13 proteins, two rRNA and 22 tRNA (Kochzius, 2009). They are almost independent from the other neighboring organelles, having their own DNA and replicating autonomously from the nuclear DNA. On the other hand, we cannot ignore that both nuclear and mitochondrial DNA complement each other, since most of the mitochondrial proteins are encoded by nuclear DNA and certain protein subunits of nuclear origins, such as some metabolic enzymes, are rendered nonfunctional without their mitochondrial subunits (Brown, Evolution of animal mitochondrial DNA, 1983).

Mitochondrial DNA ranges in size between 15-17 Kbps (Grey, 1989) and lacks mutational repair mechanisms when replicating; therefore when a mutation occurs, they tend to accumulate, giving the species a diverse sequence compared to others (Wilson, et al., 1985). This maternally inherited marker evolves ten times faster than nuclear DNA in some organisms (Brown, George, & Wilson, 1979). The most variable region in the mtDNA is the Displacement loop (D-loop), the origin of replication; making it an ideal marker to test the diversity in recent divergent populations (Parker, et al, 2003).

Because mtDNA has a unique uni-parentage quality; it is perfect to trace back parental lines, in other words, founder effects (Wallace, Garrison, & Knowler, 2005). Mitochondrial studies have been conducted on various freshwater species such as brown trout (Bernatchez, Guyomard, & Bonhomme, 1992), river catfish (Dodson, Colombani, & Ng, 1995) and even the fish under study the striped snakehead *Channa striata* (Abol-Munafi, et al, 2007).

The first mitochondrial human population study was back in 1981 and it was based on restriction enzyme analysis, like most studies at that time (Denaro, et al., 1981). Population studies of diverse organisms followed, such as the beetle of the Canary Islands (Juan, Oromi, & Hewitt, 1995).

It has been proven that interactions do exist between mtDNA and nuclear DNA, in fact it is said that polymorphisms in the nuclear DNA have affected those of mtDNA (Fos, et al, 1990). The effective population size of this marker is only one quarter of nuclear DNA because of its clonal and maternal inheritance, therefore variation among population is lower, however divergence is considerably higher (Latorre, et al, 1992).

A lot of risks are involved when dealing with this type of marker, problems such as back mutations, substitutions and mutational hotspots tend to arise (Liu & Cordes,

2004). Another precaution when dealing with markers developed for organisms other than fish is that results are not always identical to the original organism. For instance the mtDNA segment D-Loop is highly variable in human and other organisms such as *Drosophila,* however it was found to be less variable in fish species. Therefore minor studies and modifications are necessary before applying them on fish (Ferguson, 1994). This goes for most markers with emphasis on highly variable regions.

2.1.4.2.2. Nuclear DNA

Sizes of animal nuclear DNA can range anywhere from 0.02pg (*Pratylenchus coffeae*, Plant-parasitic nematode) to 132.83pg (*Protopterus aethiopicus*, Marbled lungfish) (Gregory, 2005). Contents of these genomes can vary between unique to common, repetitive to single genes, coding and non-coding…etc. The most common DNA markers include:

2.1.4.2.2.1. Restriction Fragment Length Polymorphism RFLP

This is one of the first markers to be employed by population geneticists, not to mention the best known marker at the start of human genome mapping (Donis-Keller, et al., 1987). The concept behind this marker is rather simple, restriction enzymes are used to recognize precise sequences that are 4, 5, 6 or 8bps in length and then cut the DNA into fragments that differ in sizes and therefore have different sized bands, characterizing a certain genome or DNA. Usually the smaller the enzyme restriction site the smaller the fragment size, therefore a 4bp restriction site would yield DNA fragments that are relatively shorter than a 5 or 6bp restriction site (Parker, et al, 2003).

When a mutation occurs at an enzyme recognition cut site, it is either lost or gained, leading to a change in the pattern of the viewed bands.

RFLP polymorphism was first described in 1980 (Botstein, et al, 1980), back then the most common way to separate and view the restricted fragments was Southern Blot Hybridization (Southern, 1979). Today PCR has replaced the old method and fragments now can be amplified and viewed on a normal agarose based gel after staining with ethidium bromide, since size ranges of the cut fragments are usually wide; they are relatively easy to view without the need of high resolution gels. Small DNA segments such as mitochondrial DNA are easily amplified (Hofmann, et al., 1997), however "long PCR" is needed for larger size DNA (Takayama, et al, 1996).

A major drawback of this codominant marker when considering this marker is that, indels, reallocations of restriction sites and base mutations occur quite often, but chances of that happening in the locus under study are unlikely. Polymorphism is also hard to come by. Lastly, when performing PCR (to develop RFLP markers) partial sequences should be known; in order to optimize the process, which sometimes might cause some inconvenience (Liu & Cordes, 2004), the procedure overall can be tedious and time consuming. Nonetheless the use of random primers that have been established for other species can come as a quick solution to the problem (Parker, et al, 2003).

2.1.4.2.2.2. Random amplified polymorphic DNA (RAPD)

A method that was first developed in 1990 (Williams, et al, 1990), this PCR based procedure is a product of amplification of an anonymous fragment usually not longer than 2kbps. It uses identical pairs of primers that anneal at a low temperature either perfectly or not. Semi-perfect annealing such as 1bp difference also can result in a PCR

product, if the 3' region is strongly annealed. These primers are typically 10bps long and are commercially available.

The amplified products are usually of different loci and are "inward oriented" due to the flanking region of which the identical primers annealed (Parker, at al, 2003). Products of RAPD are usually non-coding and can be viewed on a gel stained with ethidium bromide. Polymorphism is indicated by the band product size. Mutations which have occurred at the primer annealing site will most probably lead to no resulting product. Indels on the other hand, can give rise to fragments that are smaller or larger than the normal sized band in other individuals of the same species (Bhattramakki, et al., 2002; Liu & Cordes, 2004).

Advantages of this dominant mendelian inherited marker, lies in the fact that only a single primer is needed for the development of the marker and prior knowledge of the sequence of the genome is not necessary (Hadrys, Balick, & Schierwater, 1992). Not much DNA template is needed for RFLP analysis and a few samples can be assessed concurrently (Tamate, et al, 1995). Furthermore this marker has been employed for genetic variation studies (Gwakisa, Kemp, & Teale, 1994), population studies (Lynch & Milligan, 1994) and pedigree assessment (Dweikat, at al, 1993). Furthermore it has been used to scan for other markers in the resulting amplified fragment (Klein-Lankhorst, et al, 1991). It is also applied to target certain regions in the genome that are in linkage disequilibrium, by pooling genetic information from other individuals that have conserved sequences at a particular site (Tingey & del Tufo, 1993).

Application of this marker in the world of aquaculture is widely known, RFLP is used to identify fish species (Bardakci & Skibinski, 1994), asses their genetic structure (Leuzzi, et al, 2004), determine kinship and parentage (Ali, Ahmed, & El-Zaeem, 2004) and genome mapping (Postlethwait, et al., 1994; Liu, et al, 1999).

RAPD has been used widely; however disadvantages to this method make it unfavorable. Determining the right primer size, is just one of the difficulties, since primers that are too short will amplify large numbers of random sequences, which is undesirable, and vice versa. Also PCR reactions do not tolerate slight changes in temperature profiles, Mg^+ concentrations and also type of polymerase chosen for the reaction. In addition to that using gel to separate fragments can be difficult when dealing with fragments of the same size.

The main disadvantage is that heterozygote markers are undistinguishable from homozygote markers, because of its dominance, therefore if a null allele exists in a heterozygote, it would be masked by the other allele and the band will be identical to a homozygote marker (Lynch & Milligan, 1994). It is also impossible to tell whether resulting bands are of the same loci, or amplified from another having the same flanking primer annealing site, this could lead to the erroneous count of total number of loci (Liu & Cordes, 2004; Wirgin & Waldman, 1994).

Overall the technique is rather simple, with no need for sequencing or cloning, making it an attractively easy marker to work with (Bardakci, 2001).

2.1.4.2.2.3. Amplified fragment length polymorphism (AFLP)

This is a PCR based fingerprinting technique that was first applied in 1995. The method is rather simple, DNA is fragmented using restriction enzymes and then adapters are attached on both ends of the fragments, with a 3’ extension; allowing the PCR primers to anneal. Resulting fragments are viewed normally using get electrophoresis (Vos, et al., 1995). An added advantage to this method is that unlike RFLP and RAPD, there is the possibility of selecting the amplified fragments, by choosing primers that have three

extra nucleotides which are complementary to the nucleotides flanking the restriction sites. Annealing temperatures are very high, making the marker highly reproducible (Liu & Cordes, 2004). No prior knowledge of the fragment sequence is necessary for this technique.

Mutations can occur either at the restriction sites, indels between the restriction sites or even at the primer binding site. The number of fragments produced depends on the primers used and their specificity. Still high numbers polymorphic fragments arise (Liu & Cordes, 2004) even though it is bi-allelic; this is done by increasing the number of loci assessed.

Overall AFLP is an excellent tool in identifying polymorphic restriction fragment sites, identification of genomic clones (Vos, et al., 1995), assessment of genetic diversity, systematics (Keim, et al., 1997), taxonomy (Liu, et al., 1998), population genetics and evolutionary biology (Yan, et al., 1999; Taylor, et al., 1999), kinship studies (Ellis, et al., 1997) and QTL mapping (Yin, et al, 1999).

Setbacks include inability to detect homologous markers (Mueller & Wolfenbarger, 1999). Dominant by nature, it is always outdone by other powerful co-dominant markers such as microsatellites. However, there are occasions where co-dominant AFLPs have been found (Wong, et al, 2001) in about 4-15% of all polymorphic AFLP markers isolated.

2.1.4.2.2.4. Variable number tandem repeat (VNTR)

This marker is scattered throughout the genome and is overall non-coding. It is composed of repeating units that are extremely variable. At first multilocus markers were used for genetic studies, but then interpreting the banding patterns proved rather

complicated, so a transition was made and single locus VNTRs are now more applicable.

There are two types of VNTRs, minisatellites and microsatellites. Although the later has proven to be a lot more applicable since it can be amplified by PCR and many other advantages (discussed below). Still when applying markers in gene mapping and quantitative trait linkage, both markers can be used coherently (Prodiihl, Taggart, & Ferguson, 1995).

2.1.4.2.2.4.1. Minisatellites

This marker was first discovered by Wyman and White in 1980 (Wyman & White, 1980) and was then used by Jeffreys in 1985 (Jeffreys, Wilson, & Thein, 1985) as a molecular marker. Inherited in a menedelian manner, minisatellites consist of short DNA sequence repeats from 9-100bps in length repeated twice to several hundred times (Prodiihl, Taggart, & Ferguson, 1995).

It initially started off as a multilocus assay. The technique was first applied in fisheries in 1991 on Tilapia (Carter, et al, 1991; Harris, Bieger, & Doyle, 1991). However problems arose when applying multilocus markers in population genetics; as most statistical tests do not comply with anything but single locus assays. Another issue was reproducibility; bands would not show the same pattern consistently (Bentzen, Harris, & Wright, 1991), not to mention the fact that at times over 40 bands can arise in a single run, making it extremely hard to pin point which bands belong to which locus, such as the situation found in Salmon minisatellites (Taggart & Ferguson, 1990).

To overcome most of these complexities, single locus minisatellites were developed using probes (Wong, et al, 1987), these were first applied on freshwater fish species Tilapia in 1991 (Bentzen, Harris, & Wright, 1991).

The drawback of isolating a single locus minisatellite is that cloning is needed (especially if it is done on an unstudied species), which raises the cost and effort of this technique of marker development. When resulting minisatellites are developed, they produce distinct banding patterns that detect polymorphic characteristics. A single band indicates homozygosity while a double band a heterozygous individual. Mutations in allelic size are mainly due to unequal crossovers and slippage during replication (Jeffreys, et al, 1988).

This breakthrough in marker development came in hand in several studies involving aquatic species, such as genetic variation studies in fish (for example rainbow trout) (Taylor E., 1995), bottleneck events (Ferguson, et al., 1995) and population studies (Galvin, et al, 1996). In a comparison study between allozymes, mtDNA and minisatellite done on Atlantic salmon, outcomes proved that the later showed much more variability than the former two markers. Results also showed comparable observed and expected statistical value with regards to population statistical simulations (Galvin, McGregor, & Cross, 1995; O'Connell & Wright, 1997).

The main concerns however include the tedious procedure of isolation that involves cutting using restriction enzymes, gel running and Southern blot hybridization, all of which requires enormous amounts of DNA and effort. A PCR based isolation technique was developed for Atlantic cod isolation of minisatellites (Galvin, Sadusky, McGregor, & Cross, 1995), however large allele dropouts were observed and in some cases where the large alleles did amplify; they would be very hard to distinguish even in higher

resolution gels because the separation of the band is not adequate enough to separate larger sized fragments (O'Connell & Wright, 1997).

2.1.4.2.2.4.2. Microsatellites

Widely known for its technical simplicity, microsatellites have revolutionized the world of molecular markers. These simple sequence repeats (SSRs) consist of 1-6bp tandem arrays, usually less than a 100bps long and flanked by unique sequences and primers are usually designed to anneal to these sequences, amplifying the repeat region in between (Tautz, 1989). Usually found in higher organisms, the repeat sequences can have many hotspots where the marker is commonly found includes around the telomeres (Royle, et al, 1987). Repeats can either be adjacent to each other with no interruptions or with a few interposed sequences.

The marker can also exist as cryptically simple sequences that are scrambled across coding or non-coding regions, these were originally normal tandem microsatellite sequences, however due to point mutation; they have a final intermixed sequence. These cryptic sequences are allocated often in noncoding regions of higher eukaryotes, but they also exist elsewhere (Tautz, Trick, & Dover, 1986). It is also claimed the microsatellites can arise in an opposing manner. Meaning tandem repeats can also arise by point mutation from a cryptic sequence.

Microsatellites are considered the strongest marker yet because of the high number of alleles per locus, giving it the highest PIC number compared to other markers (Liu & Cordes, 2004).

Mutation rates can go anywhere from 0 to $8x10^{-3}$ (Weber & Wong, 1993) and it is thought to be found at least once every 100kbps in eukaryotes (Tautz, 1989) and once

every 10kbps in fish (O'Connell & Wright, 1997). This shows that they are highly dispersed in fish genome and some studies even proved their high degeneracy in certain fish (Wright J. M., 1994).

Mutation is believed to be due to slippage during replication (Slipped strand mispairing, SSM), unequal crossovers (Levinson & Gutman, 1987), or indels and base substitutions (Lopez-Giraldez, Marmi, & Domingo-Roura, 2007). However a recent study has shown that slipped stand mispairing and unequal crossover is in fact not a cause, but sister chromatid exchange might have something to do with it; because length mutations happen within the allele itself and not in the flanking regions (Wright J. M., 1994). It is believed that if a repeat array is interrupted by non-repeat segments chances of slippage will be highly reduced (Petes, Greenwell, & Dominska, 1997).

There has always been an ongoing debate, as to whether microsatellites are neutral or subjected to selection. It is thought that those loci adjacent or within genes that are expressed, most probably would be under selection. Others that occur in introns can vary depending on its location and neighboring genes. Yet they are thought to coincide with their expressed fellow gene in a phenomenon that is better known as genetic hitchhiking. Nevertheless, most microsatellite lie far from coding regions, branding them as selectively neutral (Scribner & Pearce, 2000).

Size homoplasy is also a concern when talking about mutations in microsatellites. Not all alleles sharing the same size are necessarily of a common ancestor. This can complicate statistical data, especially those that are applied for parentage assignments and relatedness studies (Scribner & Pearce, 2000). Some have suggested considering interrupted microsatellite loci instead of perfect tandem repeats. This reduces the chance of homoplasy (Estoup, et al., 1995).

There are two mutation models thought to have given rise to microsatellites. First is the infinite allele mutation model (IAM) (Kimura & Crow, 1964), which states that new alleles arise regardless of the numbers of repeats different. While the second model, stepwise mutation model (SMM) (Ohta & Kimura, 1973), infers that a single repeat unit is added or deleted for every single mutation, making it plausible that the same alleles might arise again (O'Connell & Wright, 1997). Initial evidence were all for SMM, since most differences parentage analysis were 1-2 repeat units different from one generation to another, this was mainly studied in human (Valdes, Slatkin, & Freimer, 1993). Although this model did in fact fit the observed results of human loci perfectly, it did not apply to other organisms with higher allele variability. Finally factors that have to be considered regarding mutation include: the chromosomal location of the loci and the flanking regions and its surroundings (Bachtrog, *et al*, 2000).

The alleles of polymorphic loci in fish tend to have rather large size differences or repeat number. For instance some fish have allele numbers of over 50, such as Atlantic Salmon which reportedly has more than 52 alleles (Norris, Bradley, & Cunningham, 1999), making it most likely an infinite allele model (Liu & Cordes, 2004; Balloux & Lugon-Moulin, 2002), this theory was supported by studies on rainbow trout (O'Connell, et al, 1997) and bees (Estoup, et al, 1995). The variability in the number of alleles can vary among loci even in the same species. It would be worth looking into the objectives of the study and decide the allelic diversity most appropriate for the study. For instance population studies would require a lower allelic range than genetic aquaculture studies (Wright & Bentzen, 1994).

Isolation of microsatellites include amplification by PCR and separation either using acrylamide gel or a genetic analyzer, for a better resolution and discrimination of up to 2bps. In the case of the second method, one of the primers at least should be

fluorescently labeled, in order to be detected by the machine. The good thing about microsatellites is that multiple loci can be analyzed and separated simultaneously on the same gel. Also rather small amounts of DNA are needed since no Southern blot is involved.

Once these markers are developed, they can be of high value over the long run, because cross species amplification is rather feasible. This has been reported in studies of haddock (*Melanogrammus aeglefinus*) and pollock (*Pollachius virens*), both of which have used microsatellite primers of Atlantic cod (Wright & Bentzen, 1994).

Microsatellites have already been developed for freshwater fish such as, brown trout *(Salmo trutta)* (Estoup, et al, 1993), sticklebacks (*Gasterosteus acuelatus*) (Rico, et al, 1993), tilapia (*Oreochromis* shiranus) (Ambali, 1996) and zebrafish (Goff, et al, 1992), just to name a few.

These sequences were previously thought to be of no use in other words “junk DNA” (Tautz, 1989), however they also have regulatory purposes (Kashi & Soller, 1999). They work as enhancer elements in promoter regions such as the intronic promoter region of the gene thought to be associated with type I diabetes in the Japanese population FOXP3/Scurfin. The promoter contains a microsatellite polymorphism that might be a cause of this disease (Bassuny, et al., 2003). Also they are known to bind to proteins. In the case of fragile X syndrome a trinucleotide of an intronic region that binds HeLA nuclear proteins was discovered to be associated with the disease (Oudet, et al., 1993; Richards, Holman, Yu, & Sutherland, 1993). Lastly they are thought to be associated with developmental factors in higher organisms associated with repeat numbers (Kashi, King, & Soller, 1997). Such as the stretch of trinucleotide repeat thought to be associated with sex reversal trait in mice (Coward, et al, 1994).

These hypervariable markers have prevailed over the setbacks of many markers such as allozymes and mtDNA, especially seen in highly studied fish such as Atlantic salmon and Atlantic cod (Wright & Bentzen, 1994).

Studies that this co-dominant marker has contributed in include: 1) Identifying organisms and their origins (Hansen, Kenchington, & Nielsen, 2001) and observing changes in genetic variability, especially due to participations in breeding programs (Norris, Bradley, & Cunningham, 1999). Since SSRs are single locus markers they are perfect for: 2) population studies such as those done Atlantic cod (Bentzen, et al., 1996), however the first intra and inter population study was done on rainbow trout (Nielsen, et al., 1994). 3) Also linkage analysis, as in the study on tilapia and channel catfish (Kochera, et al., 1998; Waldbieser, et al., 2001). 4) Genome mapping is extremely handy in the world of genetic improvement; this is an overall scan of genes and loci that are expressed.

Therefore QTL mapping for instance, has taken the genomic world by storm since most traits of good nature are polygenic, i.e. controlled by different loci. Hence the complete understanding of how these loci work together, is of upmost importance. Marker assisted selection has helped extremely with this and microsatellite polymorphism has also assisted in allocating such QTLs (Poompuang & Hallerman, 1997), however fish genome is quite large and requires even spaced markers to map the trait loci (O'Connell & Wright, 1997). Linkage mapping started off using allozyme markers such as the one done on salmonoid fish (May & Johnson, 1990). Then because of the limited variability that allozymes had to offer; RAPD was used instead on species such as zebrafish (Postlethwait, et al., 1994), a model freshwater fish. Nevertheless RAPD also came with its complications especially because the primers were too specific, which led to

limitations in cross species studies and comparative mapping (O'Connell & Wright, 1997).

When the era of microsatellites dawned, it was utilized in full. Their polymorphic markers gave us a higher resolution of the genome. Mapping of fish genomes, including tilapia (Lee & Kocher, 1996) and zebrafish (Shimoda, et al., 1999) were conducted using this marker. Nowadays most of the mapping projects take advantage of several markers such as AFLP, microsatellites and others, an example of that is a genetic linkage map on Tilapia (Kocher, et al, 1998).

Other studies applying this marker include: 5) Detection of bottlenecks in certain populations such as a hatchery strain of Japanese flounder (Sekino, Hara, & Taniguchi, 2002). 6) Pedigree analysis and parentage assays, mainly done to deduce mating behavior and stock management, this proves very helpful since most new born fish cannot be tagged. One of the first to study and apply this field on fish was Herbinger, who estimated kinship relations among Atlantic cod individuals of the Western Bank of the Scotian Shelf (Herbinger, et al., 1997). Another example would be again the Japanese flounder (Sekino, et al, 2003).

One of the major draw backs with this marker is that slippage can occur most often during PCR, this causes “stuttering”, especially when amplifying di- repeats; there are quite a number of theories as to why this phenomenon happens. The first is misannealing after the strands segregate in replication; due to the repetitive nature of the sequence. Therefore when this happens a few times, a domino effect occurs, resulting in a few repeats added or removed from the segment, depending on the manner reannealing takes place (Hancock, 1999). The more stringent the repeat array the more likely stutter is to occur on that strand after replication, due to the conserved energy profile that restricts the polymerase’s flexibility (Hummel, 2003). This may come as a

nuisance when conducting population studies on species, since it might mask the availability of certain alleles and increase homozygote access. Stuttering has also negative effects on kinship studies. However in mapping studies this would be much of a concern, since most loci used already established and developed; therefore sequence repeats and their numbers are all known.

This is why many choose to work with larger repeat units. This not only will reduce stuttering, but also will eliminate the need to use high resolution gels and automated sequencer machines, if funding is limited. If at all necessary to use di- repeats, it is recommended to have an overall length of <120 bp; in order to effectively increase the run time of the gel, which results in higher scoring resolution (O'Connell & Wright, 1997).

Null alleles, which amplify poorly or not at all, are also considerably a problem in proteins, RFLP and microsatellites. When only one allele is amplified and the other (in the homologous chromosome) is not, it underestimates heterozygosities and genotypes (Callen, et al., 1993).

Reasons for this can vary, one of which is mutations in the primer annealing site, especially in the 3' end; because that is where the amplification starts (Kwok, et al., 1990). Also when PCR is involved in amplification, there is a favorance of shorter alleles than larger ones, a phenomena that is called "large allele dropout", some of these alleles can be salvaged by adjusting contrast and DNA concentrations (Dakin & Avise, 2004).

When a null allele is heterozygote or in other words "Partial null allele" is present only one copy is amplified while the other is not. This results in misinterpretation of results and is mainly seen in cold-blooded fish (O'Connell & Wright, 1997). Poor template

quality can also erroneously declare a heterozygote to be mono-, or show inconsistent amplification, appearing in some runs and not appearing in others (Gagneux, Boesch, & Woodruff, 1997). Also large insertions and deletions in the repeat motif can cause null alleles (O'Connell & Wright, 1997).

Another setback is the unspecific amplification of loci, however this has been overcome by using either a third labeled primer, that falls anywhere (preferably in the centre) in the amplified fragment or direct sequencing of each of the segments, this is not only pricey, but also troublesome (Tautz, 1989). The initial stages of isolation and characterization of microsatellites are costly and time consuming; however results over years have proven that once developed, they are of high value. Extremely high numbers of alleles is not always a good thing, in cases such as genetic stock identification and even genetic variation studies, binning closely sized fragments are recommended (O'Connell & Wright, 1997). Also shifts in mean allele distribution and sizes can come as a nuisance. They are mainly due to insertions/deletions in the flanking regions of closely related individuals (O'Connell & Wright, 1997).

Many comparative studies have been done between different data sets of both mtDNA and microsatellites. Although there is no doubt that the later proved a lot more variable, it has to said that population studies solely done on microsatellites are mainly mutational, however there is still a need for a solid biogeographical basis to the study that can give clearance as to the demography of the population (O'Connell & Wright, 1997).

2.1.4.2.2.5. Single nucleotide polymorphism (SNP)

This marker has been around since the late 1970s, when sequencing revolutionized the molecular world after a decade or more of full utilization of microsatellites. It is basically a single base pair point mutation in a specific position on a loci, the frequency of this mutation (single nucleotide substitutions) is between $1\text{-}5\times10^{-9}$. Although in theory any of the nucleotides can replace a single base, this rarely happens, in fact there are only two allele variants. This is because of the low mutation rate (as mentioned earlier) and the bias towards transition mutations rather than transversions. The allele with the least frequency should account for at least 1% among all other alleles (Vignal, et al, 2002; Liu & Cordes, 2004). This co-dominant marker does not have a PIC value as high as microsatellites, however it is more abundant in the genome (Liu & Cordes, 2004), making it the cause of most of the variability among individuals of the same species (Altshuler, et al., 2000).

SNPs are not as easy to develop a one would think. Since the basic developmental technique of sequencing a loci in multiple individuals always ends up with sequence artifacts.

Another way to approach this issue is to look into EST libraries of a locus from different individuals; since this not only will save time, but also insure that the SNP lies in a coding region (Vignal, et al, 2002).

A different method developed in 2000 was reduced representation shotgun (RRS), whereby partial segments of a genome from multiple individuals are aligned and scanned for SNPs. This is done by mixing in DNA from random individuals and cutting them up with restriction enzymes, running them on gel and then re-extracting these segments and cloning them into plasmids. After that the common regions (overlapping)

are aligned and scanned for SNP using software programs (Altshuler, et al., 2000). This approach was also used for the human genome project (Collins, Brooks, & Chakravarti, 1998). Other approaches such as: single strand conformation polymorphism (Orita, et al., 1989), heteroduplex analysis (Lichten & Fox, 1983), variant detector arrays (DNA chips) (Wang, et al., 1998) and many others were used for isolation. The latter technique has rapidly picked up in recent years compared to others. It relies on a basic concept of hybridization of PCR products to oligonucleotides that are secured to a tiny chip, measuring the annealing number of bases and identifying the odd one out.

More extravagant methods that do not suite small budget laboratories include: Matrix-assisted laser desorption ionization time. This is a multiplex PCR based assay that uses a laser to measure amplified fragments extended by a dideoxynucleotide (Karas & Hillenkamp, 1988; Griffin & Smith, 2000). Another technique is pyrosequencing (Ahmadian, et al., 2000), a method based on sequencing, emitting lights as each nucleotide is added. Also real-time mixed stock analysis can be done on fish such as Alaskan salmon (Smith, et al, 2005).

The method that proved most popular is microarray mapping, the technique has been used to identify SNP in zebrafish (Stickney, et al., 2002). However the channel catfish was the only aquaculture species to have as many numbers of SNPs (He, et al, 2003) as livestock (Jungerius, et al., 2003). This then led to another study on Atlantic salmon, in which 2507 putative SNPs were detected (Hayes, et al., 2007).

Because they are easily automated and do not require electrophoresis, SNPs are quickly becoming as popular as microsatellites, if not more popular. This is because it is not as prone to error as microsatellites. However it is still a very expensive marker to work with even though it is expected to get more financially suitable for even the lowest of project budgets. It also does not offer as much variability as microsatellites, limiting its

use in population structure studies (Hauser & Seeb, 2008), although some have successfully applied it in the field (Smith, et al, 2005).

2.1.4.2.2.6. Expressed Sequence Tags (ESTs)

These tags are a result of sequencing cDNA, which itself is reverse transcribed from mRNA. These markers are best known for physical mapping of genomes (Adams, et al., 1991), gene expression and identification (Liu & Cordes, 2004). This gives an insight as to what goes on inside each tissue and expression variation according to the developmental stage and changes caused by diseases. cDNA libraries have been used to identify ESTs in Atlantic salmon (Davey, et al., 2001) and flatfish (Douglas, et al., 1999). Polymorphic ESTs can be used in aquaculture mapping; an example was linkage mapping of catfish (Liu, Karsi, & Dunham, 1999).

Microarrays have been developed using ESTs and is used to quantitatively assess gene expression, this has been used widely in plants (Schena, et al., 1996; Southern, Mir, & Shchepinov, 1999; Richmond & Somerville, 2000). In recent years this has been applied on fish such as salmonid (Rise, et al., 2004).

Overall this type 1 marker has been used to analyze gene expression of medeka (Hirono & Aoki, 1997), Atlantic halibut (Park, et al., 2005), rainbow trout (Kono, Sakai, & LaPatra, 2000), Japanese flounder (Inoue, et al., 1997), etc. Statistical studies indicate that by the end of 2007 the number of fish sequences exceeded 5.5 million, of which 64% were ESTs (Kochzius, 2009).

Summary

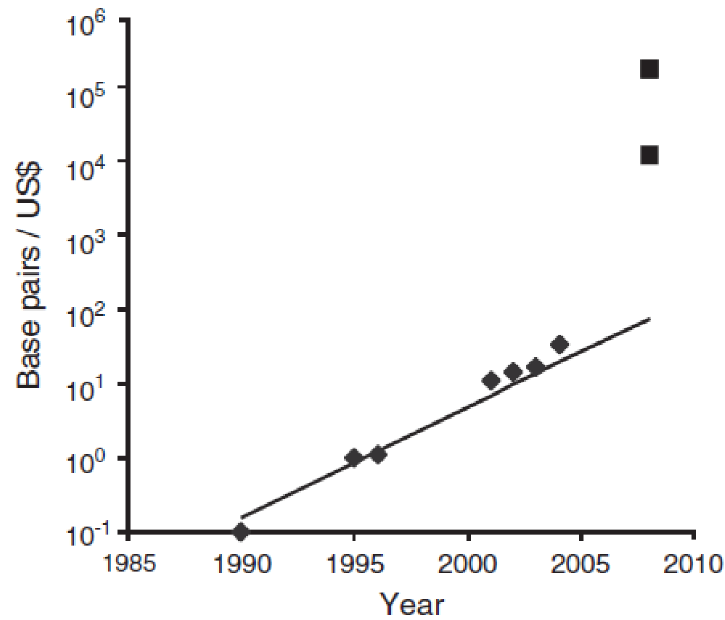

Figure 1 Increase in the number of basepairs sequenced per 1$US. Diamonds show sequencing costs every 2 years, while squares show the efficiency (Hauser & Seeb, 2008).

It has to be said that sequencing techniques are picking up quite rapidly and becoming increasingly cheaper on a daily basis (Figure 1). Some fish transcriptome can cost as low as US$ 30,000 (Hauser & Seeb, 2008). The study of functionality of genes can help reveal scientific ambiguities such as tolerance to cold, hypoxia and viral diseases in fishes such as common carp (Williams,et al., 2008). Also pollutants effects on expression of hepatic genes in European flounder were examined (Diab, et al., 2008). A chart showing the research publications done on fisheries can be seen in Figure 2 and a table of comparison between all popular markers and the studies they are most applicably utilized, is shown in Table 1.

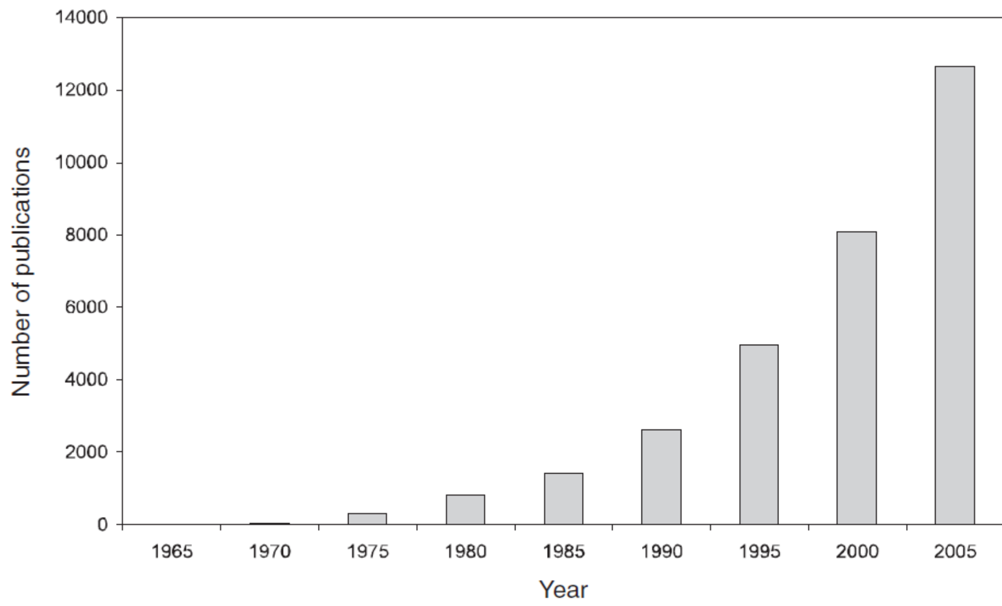

Figure 2 Fishery genetics publications over the decades, in literature database aquatic sciences and fisheries abstracts (Kochzius, 2009).

Table 1 Comparison of different markers, (O'Connell & Wright, 1997; Vignal, et al, 2002; Liu & Cordes, 2004).

Marker/ technique	Technicality	Cost	Studies undertaken using this marker		
			Genome mapping	Parentage assessment	Population genetics
Allozyme	Low	Low	Low	Low	Moderate/high
MtDNA	Low	Low	_	Low/moderate	Moderate/high
RAPD	Low	Low	High	Moderate	Low
RFLP	High	High	High	Low	Moderate
AFLP	Low	Low	High	Moderate	High
Single locus minisatellites	High	High	High	High	High
Microsatellites	High	High	High	High	High
SNP	High	High	High	Low	Moderate

2.2. Striped Snakehead (*Channa striata*)

Channa striata is a freshwater fish most commonly known as snakehead murrel or striped snakehead. Its commercial and local name in Malaysia is Haruan Figure 3. The fish was first described by Bloch in 1793, who found the fish in Malabar, India (Lee & Ng, 1991) and was originally named *Ophicephalus striatus* (Bloch, 1793). The mud fish falls under the family Channidae and has two genera: 21 species of *Channa* in Asia and three *Parachanna* in Africa (Berra, 2007). However a total of 31 species (including subspecies) have been known so far (Phen, et al., 2005) and webpages such as Zipcodezoo have recorded 54 (including subspecies) (zipcodezoo, 2009). Out of the many species out there, at least 11 of them originate from South-East Asia (Lee & Ng, 1991). The exact taxonomy of the species is shown in Figure 4. The family used to be placed under Gasterosteiformes and Synbranchiformes (Greenwood, et al, 1966).

Figure 3 Images of C*hanna striata* (Top images: Jean-Francois Helias, Thailand: www.snakeheads.org) (Bottom image: Courtenay W. J., et. al., 2004)

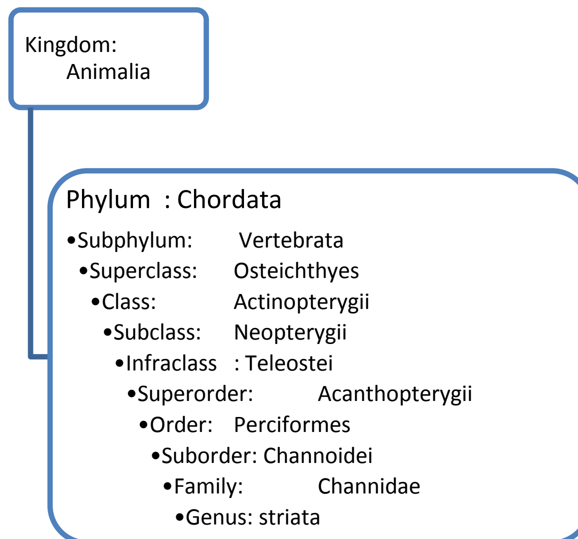

Figure 4 Taxonomy of Channa striata (Scopoli, 2004).

2.2.1. Physical characteristics

It can grow to 28cm in length in its first year of life (Bhatt, 1970) and up to 100cm in a lifetime, giving it an average to large size. Its maximum weight is about 6Kgs. The heaviest was caught on a reserve in Hawaii, where the fish weighed up to 20Kgs (World Records Freshwater Fishing). Also in the same region another fish that was 150cm in length was caught (Yamamoto & Tagawa, 2000).

It has a flattened head shaped like a snake, hence its name and a body shaped like an eel (Husin, 2001) and a wide elastic mouth with sharp teeth, that helps this carnivore penetrate into a prey's flesh (Qin & Fast, 1996) (Figure 3). The dorsal part is darker (Between brown and black) when compared to the abdomen, making the stripes hard to point out (Courtenay & Williams, 2004). Juveniles of the fish tend to have spots at the posterior end of the body; these tend to disappear as soon as the fish matures. The body

is covered with large scales, except for under the jaw (Yamamoto & Tagawa, 2000; Grzimek, 2003). It is also covered with a slimy, mucus substance that helps keep the body moist (Lee & Ng, 1991). The murrel has pectoral, dorsal, anal, caudal and pelvic fins and rays (Courtenay & Williams, 2004).

2.2.2. Genetic content

On the genetic part, the fish has 40 diploid chromosomes (Banerjee, et al., 1988), however another conference proceeding stated that the chevron fish had 44 chromosomes instead (Donsakul & Magtoon, 1991). Haploid DNA content is 0.75 picograms (Phen, et al, 2005). Not many markers have been established or published for this species so far, except for mitochondrial studies (Li, Musikasinthorn, & Kumazawa, 2006; Ambok, et. al., 2007).

2.2.3. Geographical distribution

The mud fish is found across Asia, from Pakistan to South China (Grzimek, 2003) and was first discovered in the oriental region of Southeast Asia (Figure 5) (Weber & de Beaufort, 1922). It is thought to have originated around the Malaysian peninsula half a million years ago (Jais, 2007) and is mainly found on the coast of either side of peninsular, close to human populations (Figure 6) (Lee & Ng, 1994).

Since then the fish has been introduced to many neighboring countries such as the Philippines in 1908 (Cagauan, 2007). It also can be seen as a native of the African fresh waters and swamps (Ambak, et al, 2006). Since its introduction in the United States, the tropical fish more than a century ago (Courtenay W. J, et. al., 2004), it has not been

widely accepted, considering them "predatory killers" (Jamaludin, 2002) and media has even sparked wild rumors that these fish attack humans referring to it as "Frankenfish" (the_junglist). The reason behind this scare in the US, is because this nonindigenous fish is thought to inbreed with native species, changing the energy flow of the ecosystems, altering biodiversity (Herborg, et al, 2007) and causing extinction of other weaker species if they compete for resources (Hilton, 2002).

In August 2002 a mission was executed to eliminate snakeheads from a Crofton pond in Maryland, USA, using herbicides (Huslin, 2002). Many *C. maculate* were wrongly identified as *C. striata* in Hawaii at museums such as Bernice P. Bishop Museum in Honolulu (Courtenay & Williams, 2004).

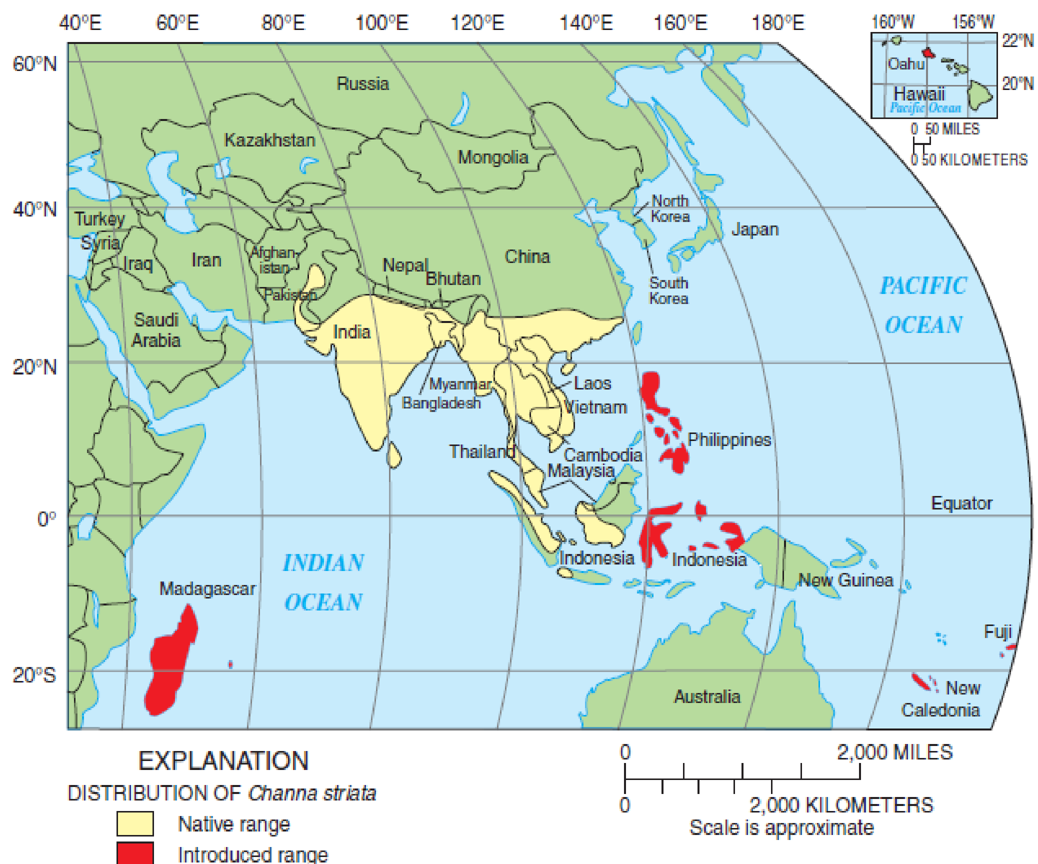

Figure 5 Distribution of *C. striata* across the globe. Reports indicate that there was some misidentification of *C. maculata* in Hawaii) (Courtenay & Williams, 2004)**.**

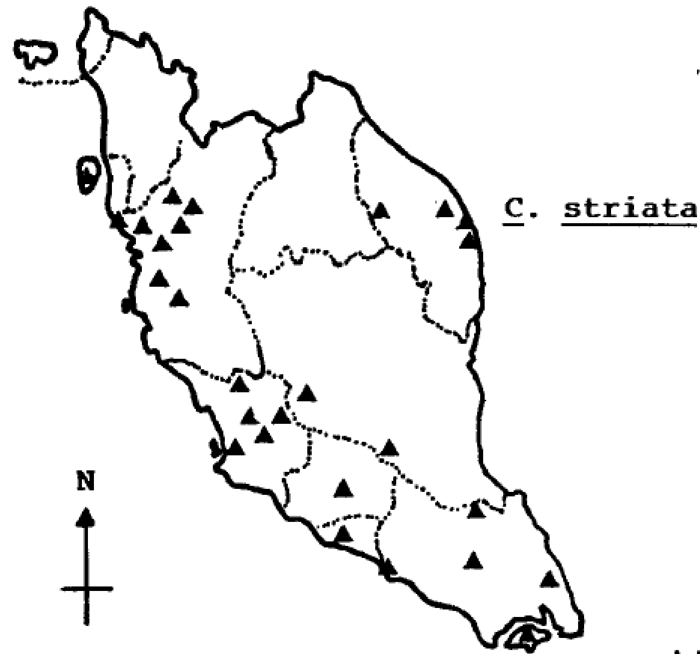

Figure 6 Distribution of *C. striata* in peninsular Malaysia. The dots indicate state boundaries (Lee & Ng, 1994).

2.2.4. Ecological habitat

It is a native of fresh and sometimes brackish water. Therefore ponds, swamps and plains are places where this species could be found. Traditionally the fish was mainly found in the rice fields of the Malaysian peninsular and one of the most cultured fish in Asia (Ali A. B., 1999). However it does prefer stagnant, muddy plains and lives in waters that are 1-2m deep and densely vegetated, nevertheless they have been found in the harshest of environments such as hard mud crusts (Chandra & Banerjee, 2004; Grzimek, 2003).

Although the species prefers stagnant waters, it has been reported that it can withstand river flows of 0.47m/s (Lee & Ng, 1994). Migration to flooded areas during the flood seasons is not unusual. This is obvious in the Mekong river, where *C. striata* is known

to migrate during the flood, however it comes back to its natural habitat eventually and is not considered a migratory fish (Poulsen, et al, 2002).

C. striata is found in water temperatures of 11-40^oC and can tolerate pH values of 4.25-9.40 (Lee & Ng, 1994). It can also withstand the harshest of environments, such as waters high in ammonia concentrations. As reported by Qin and his team, *C.striata* can with stand up to 15.7mg of unionized ammonia per L, at pH 10 (Qin, Fast, & Kal, 1997). Alternatively channel catfish *Ictulurus punctatus* can withstand a maximum of 3.1mg/L of unionized ammonia (Qin, Fast, & Kal, 1997).

2.2.5. Reproduction

This monogamous fish reaches its sexual maturity in two years at a length of approximately 30cm (Courtenay & Williams, 2004; Ali A. B., 1999). They breed externally and spawn all year round, in ponds and ditches. Studies have shown that they tend to be aggressive during mating season (Grzimek, 2003). They build nests in shallow waters and when the time comes the male fish releases its milt as soon as the female lays her yellow eggs in the nest. The egg contains oil, which help them float as soon as they are produced (Lee & Ng, 1991) (Figure 7) (Diameter of mature egg is reportedly 0.9702-1.1672 mm) (Parameswaran & Murugesan, 1976; Lee & Ng, 1991; Ali A. B., 1999).

Reports have showed that parents in some regions tend to bight off the vegetation in the spot where the female lays her eggs, or look for spots that are clear from any vegetation (Lee & Ng, 1991; Courtenay & Williams, 2004). It is said that a female produces between 4326 to 9017 oocytes in Malaysia and between 10.5 to 36.3 oocytes/g of their

body weight (Ali A. B., 1999). Female chevron snakeheads are ripe all year round (Ali A. B., 1999).

When eggs hatch, the fingerlings are guarded by the parent fish just below the surface of the water (Phen, et al., 2005). The male guards the fry (Figure 8) even after they have hatched (Grzimek, 2003). This parent offspring relation is even more obvious when danger is lurking around the waters, as the young get closer and swim in approximation to the parents (Mohsin & Ambak, 1983).

Although fish spawning can be induced using hormones (Parameswaran & Murugesan, 1976), farmers in Malaysia still prefer fishing it from rice paddies and fields due to high demands (Ali A. B., 1999).

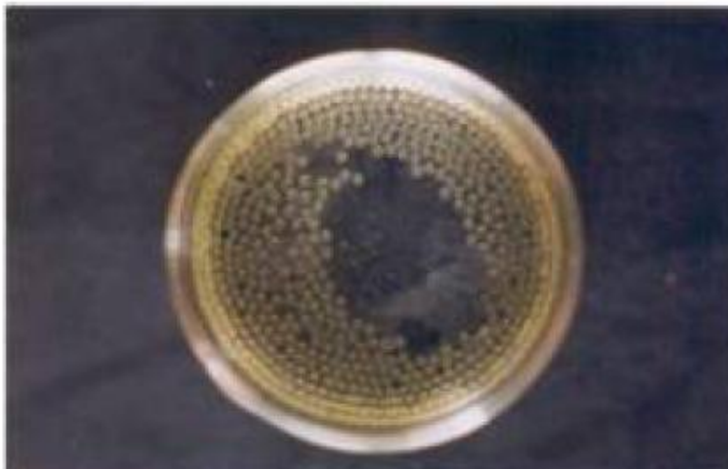

Figure 7 Fertilized egg of *C. striata* (Marimuthu & Haniffa, 2007)

Figure 8 *C. striata* fry (Marimuthu & Haniffa, 2007)

2.2.6. Breathing and organs associated

This particular fish breaths air, meaning they would actually drown if they stayed in the water for too long without coming up for air. This gives Haruan a unique feature of withstanding conditions such as low dissolved O_2 and dry seasons, in which they have to live in shallow waters and move in semi-watery areas (Husin, 2001; Chandra & Banerjee, 2004). The organs that conduct this peculiar breathing mechanism are the suprabranchial chamber, which allows oxygen intake in both water and air (Chandra & Banerjee, 2004). An interesting aspect is the fact that even as 2-3cm long juvenile fish; they are already air breathers (Mohsin & Ambak, 1983). Experimental results have shown that the fish's breathing organs are developed in about 60 days, this is within the period in which the fish is between 7 and 750 mg in weight (1 and 4.5 cm respectively) (Vivekanandan, 1977). The species can withstand up to 8hrs outside water, but this is not nearly as close as other species such as *C.batrachus,* which can stay alive up to 27-31 hrs after being taken out from the water (Chandra & Banerjee, 2004).

2.2.7. Skin

Although the fish can maneuver in mud and breathe air, it still needs to have its skin and air breathing organs wet. It also has a mucus layer on its skin that prevents excessive water loss (Wee, 1982). The skin like amphibians needs to retain certain moisture because it acts as a breathing organ and helps release CO_2 from the fish's body, because it contains sulphated mucopolysaccharides in the subepithelial connective tissues (Chandra & Banerjee, 2004).

The skin is a major defense organ in fishes, it protects it from toxic material in the water, that thrusts against the wound increasing chances of endangering the organism; therefore quick response to cuts and scabs are crucial to its survival.

Today however more studies have been conducted on the species and as it turns out, histopathological studies showed that the regeneration of the fish are different than mammals. After an injury it is clearly seen that the edges of the cut retract closer to one another within an hour and areas surrounding the cut become darker in shade compared to the rest of the fish at first, but then it somehow retrieves its natural color within 24 hours and the wound is completely healed by day 35, with no signs of scars left behind like mammals (Banerjee & Mittal, 1999).

2.2.8. Diet

Channa striata is just one of the mostly commonly found species, of the Channidae family. The Piscivores' diet basically consist of anything live that can fit into their mouth, such insects, crustaceans, little fish, baby birds, frogs and even baby turtles. Diets changed according to regional habitat (Lee & Ng, 1994; Dasgupta, 2000; Grzimek, 2003; Courtenay & Williams, 2004).

A study done on their feeding habits in Bukit Merah in Malaysia is shown in Figure 9 (Siaw-Yang, 1988). They require 40-50% protein in their diet to maintain growth and functionality (Hashim, 1994). Not much is known about their fatty acid and vitamin dietary needs, nevertheless it was found that pantothenate vitamin, is necessary for growth and survival (Paripatananont, 2002).

When this fish is cultured it is necessary to main all the necessities of their diet without causing malnourishment, which can in some cases cause cannibalism, if fish of the right size are placed in the same tank or pond (Qin & Fast, 1996).

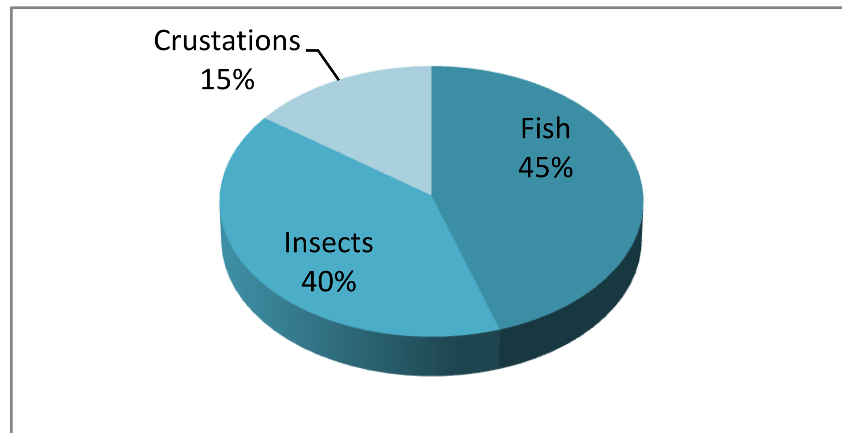

Figure 9 Pie chart of *C. striata* diet in Bukit Merah Reservoir, Malaysia (Siaw-Yang, 1988).

Over feeding also has its consequences, this is not only wasteful and expensive, but also tends to contaminate the water (Qin, He, & Fast, 1997). In captivity these fish are fed rice bran, fish paste or anything of the sort. It is necessary to establish proper culture techniques of *C. striata*, since environmental changes and seasonal reproduction are reducing the numbers of fry (Qin, et al., 1997).

2.2.9. Dietary, pharmaceutical utilizations and more

Channa striata is widely known in South East Asia as a remedy for healing wounds. Consumption of this fish has been going on for years. They are either sold alive in wet markets or sundried across Asia (Berra, 2007; Grzimek, 2003). The fish is not only a boneless delicacy, but also a medical remedy since it improves dermal wound healing properties and reduces post operational discomfort. Women have been consuming them after childbirth and believe that its meat held super healing properties (Gam, Leow, & Baie, 2006). The fish is sacrificed just before its cooking and usually added in a soup with herbs (Lee & Ng, 1991).

A cream was even formulated from Haruan. It is thought to increase the tensile strength of connective tissue and fibers, because it contains the active amino acids and polyunsaturated fatty acids that are needed for the healing process. This includes glycine that combines with aspartic and glutamic acid for wound healing and contracting. Another amino acid is arachidonic acid that serves as a blood clotting agent (Baie & Sheikh, The wound healing properties of Channa striatus-cetrimide cream — tensile strength measurement, 2000). Further studies showed an addition of cetrimide to the *C. striata* cream enhances the affectivity of the wound healing property (Baie & Sheikh, The wound healing properties of Channa striatus-cetrimide cream-wound contraction and glycosaminoglycan measurement, 2000).

A publication has shown that Professor Saringat Baie and his team are working on a spray from Haruan extract. This aerosol acts as a film or membrane when sprayed on a wound. It will not only work as a protection layer, but also wound healing accelerator. This comes as an advantage when abrasions are spread over a large surface of the body and putting and removing casts can be an uncomfortable process. An added advantage is

that the mist coming out of the nozzle is sterile and takes on the shape of the injured body part (Pokar, *et. al*., 2009).

Also Haruan is known to have sufficient amounts of minerals such as magnesium, calcium and iron (Jais, 2007). Another benefit worth mentioning is its antinociceptive/anaesthetic activities. Minor antimicrobial and antifungal properties have also been observed in animals; however plants are still the dominating source (Jais, 2007). Even the predatory characteristics of *C. striata* are considered favorable when it comes to controlling Nile- tilapia population and densities; this is practiced in countries such as the Philippines (Milstein & Prein, 1993; Grzimek, 2003; Yi, et al., 2004). In US there was a time when snakeheads were sold as pets, they were favored over other fish species because of their tough nature that can live and withstand harsh and intolerable conditions and do not require special attention (Hilton, 2002). Lastly, the mucus layer is said to be used as a paste or adhesive between the building blocks in India (Lee & Ng, 1991).

2.2.10. Infections and diseases

Striped snakehead can be host to several parasites, such as *Gnathostoma* (jaw worms), that can actually harm humans if consumed (Grzimek, 2003). Other infestations include *Isoparorchis hypselobagri*, in the winter and dry season. *Pallisentis ophicephali* tend to invade the intestine and *Neocamallanus ophicephali,* targets the digestive system too (Phen, et al., 2005). The species is also susceptible to epizootic ulcerative syndrome, a seasonal infection that causes ulcers and necrosis of tissue, which is thought to be caused by *Aeromonas hydrophila* and pseudofungi *Aphanomyces* (Lio-Po & Lim, 2002; Lio-Po, et al., 2003).

2.2.11. Wild and cultured

Wild striped snakehead is caught using nets, traps, normal fishing rods with bait on its hook and seines (Rainboth & FAO, 1996). Lately though numbers of fish in the population have decreased between 30-60%, compared to populations existing 30 years ago (Cong, Phuong, & Bayley, 2009). This could be a result of environmental changes, over-fishing, pesticide toxification (Pandian & Bhaskaran, 1983), etc (Cong, Phuong, & Bayley, 2009). Due to high demands of the fish, aquaculture industries naturally thought of breeding it outside its natural freshwater environment, in other words culturing it. Farming this fish started in Thailand in the 1950s and spread across Southeast Asia and India (Paripatananont, 2002). They can be cultured in pond or cages set in running streams, depending on the farming capabilities (Grzimek, 2003). In 1991 the 4200 tons of this fish was produced and 97% of it was from Thailand. Attempts were even made to culture the fish in Hawaii, where the fish is considered predatory (Qin, Fast, DeAnda, & Weidenbach, 1997).

When the fish is domesticated farmers and consumers have come with the realization that the fish for some reason loses its healing abilities when consumed by individuals. The reason so far is not known, although studies have been conducted on the proteins and fatty acids of different sizes of that species (Gam, Leow, & Baie, 2006) and full description of amino acids have been established. Yet the crucial question has still not been answered, what exactly causes the loss of healing property in the fish under study? Since this causes a loss of value to the cultured fish.