Indian Mackerel Working Group Meeting: 28-29 May, 2012 in Sri Lanka.

FOR INDIA, INDIAN MACKEREL AND GENETICS: We need to develop a sampling plan for tissue collection along the Bay of Bengal including the Andaman Islands to understand how stocks vary across the regions and whether there is one stock or sub-stock structure in the region that will influence how a management plan will be developed for the region. In addition India should identify the lead geneticist from National Bureau of Fisheries Genetics Research (NBFGR) to send to the meeting in Colombo. The details of the meeting in Colombo are shown below.

Objectives of the May meeting in Colombo

The purpose of the Indian Mackerel WG Meeting in May will be to establish a robust genetic sampling and analysis scheme for assessing the stock structure of Indian Mackerel in the Bay of Bengal Region. The following items will be reviewed at this working group meeting, and it will be integral to have all the genetic labs that will be used to perform the analysis to be present at this meeting, as we will cover the following *items/objectives* to make sure that this study is done correctly in the region:

1. Evaluate the Genetic Markers developed for stock identification for Indian Mackerel in the region.
2. Evaluate the ability of the labs in the region to process the data, and analyse the data using a coordinated standardized technique (this will involve a blind-test across labs to make sure consistency is used in processing the samples).
3. Develop a plan to collect the correct information for stock identification using a statistical valid sampling plan that is both spatial and temporally adequate.
4. Review the results at a later period to assess stock structure in the region.

To this effect the current know-how of the labs and their ability to do stock structure analysis will need to be presented by each country proposing this work. *A Genetic expert, who has much experience in developing sampling plans and studies like this on large scale marine studies will be present to assist the labs and make sure that this work is done correctly.*

Topics that will be covered in detail

In terms of adequacy of the study and the ability to make inference, the following items are discussed further in detail (namely sampling coverage, sampling protocols, lab-coordination, marker selection and stock- structure analysis), and will be the focus of the 2 day meeting to get the genetic study of Indian Mackerel moving in the region.

Sampling

1. Suitable sample collection across strata to capture potential gene flow boundaries.

a) Temporal sampling is important. Marine fish species may not be accessible during spawning, or are broadcast spawners, or have low fidelity to a particular spawning site. Levels of genetic differentiation vary widely for marine teleosts reflecting the amount of gene flow due to different reproductive strategies, spawning behaviour, and larval dispersal (Winans 1980). Risk: Baseline samples taken from a site should represent a single stock but is actually an

admixture from multiple stocks. This would result in the baseline not represent true stock structure. Samples collected can be tested for admixture using a number of techniques. We do not want to get into the situation where we are not properly capturing allelic frequencies for each stock. Assuming stock structure exists this would be a false negative. Understanding life-history characteristics of the Indian Mackerel will be critical here.

b) Spatial sampling is important. Sampling across geographic range of the species should capture different spawning populations assuming population ranges are not overlapping. Indian Mackerel fisheries are likely the easiest way to obtain these samples unless there are opportunities for specific sampling trips. Additional sampling should occur outside the study area (Bay of Bengal), if possible, to provide reference populations we know are genetically distinct. We may be surprised at how much stock structure there is over such a large study area (Bay of Bengal). Menezes et al. (1993) found using allozyme markers that Andaman and peninsular India, Indian Mackerel had genetic distance that were relatively large, within the range that would suggest two local races (stocks) of a species.

c) Any additional information from tagging or genetic studies will be useful for determining how to establish a temporal and spatial sampling. Although Meneze et al. (1993) found genetic differences between sample sites, Jayasankar et al. (2004) did not find significant genetic distances between three coastal Indian sites.

2. Establishing sampling protocols.
   a) Avoiding degraded samples can prevent significant data loss and additional time/expense attempting to re-assay with additional laboratory time. Sample degradation usually occurs from enzymatic break-down of the DNA caused by insufficient desiccation (preservative too dilute) or presence of PCR (polymerase-chain reaction) inhibitors in preservative (e.g. denatured ethanol). A standard procedure is to send out sampling kits because it is worth the extra work initially to provide kits rather than having analytical problems later caused by poor sample quality. Degraded samples shared between laboratories can cause undue stress when one laboratory can get a sample to work and another laboratory cannot get them to work.
   b) Individual vs. bulk sampling. A standard sampling kits has the additional benefit that collections can be sampled into individual vials so that age, sex, and size can be tracked through for each genotype. Tissues from a bulk sampling (all tissues in one bottle) will not have corresponding one-to-one biological data.
   c) Sample sizes. Sample sizes are important for determining the power of the analysis. Where highly polymorphic microsatellite loci are used samples sizes should be >200 individuals (Beacham et al. 2006) while SNP sample sizes can be 100 or less. Sample size determines how well stock specific allelic frequencies of multi-locus genotypes are resolved.

<u>Loci and Databases</u>

3. Multiple laboratory coordination (India, Thailand, Malaysia).
   a) Laboratory analysis: Years of experience dealing with multi-laboratory issues around data standardization, sample sharing, laboratory cost/structure differentials (overhead, wages, technology, and technical skills levels), quality control, platform standardization, and laboratory specialization gives a good idea of potential pitfalls. All parties must have clear expectations and a common set of requirements to allow a successful outcome.
   b) Databases: Experience dictates that all laboratories should start at the beginning of the project with data standards. Flat files can be ideal for database up-loads and allows individual laboratories to make data modifications, additions, and corrections without having a

complete data reload. Joint-agency database structure and web-based access to data is critical for ongoing fisheries management purposes the database should be designed with consideration for potential GSI (Genetic Stock Identification) expansion to incorporate mixed-stock fisheries results once stock structure has been determined. Database shells for handling baseline and mixture may be directly transferable and available saving the cost of database development.

4. Genetic marker selection.
   - **a)** Type of genetic marker (SNP or microsatellite) used will depend on available technologies in each country. North American salmon genetics laboratories have moved from microsatellite markers to SNP (single nucleotide polymorphisms) markers with the development of new technologies such as the Fluidigm array (www.fluidigm.com) reducing costs per fish. However progress has been slow because of the difficulty in finding enough species specific SNPs useful for population discrimination. SNP discovery remains an expensive and slow, it is unlikely enough would be available for the Indian Mackerel project. Those that are available could be readily used, possibly in conjunction with microsatellite markers. Microsatellites markers are either available for Indian Mackerel or from a closely related species which could be evaluated and potentially used (cross-specific amplification). Other marker types such as mitochondrial DNA may be available to be used as well, but not considered to have the resolving power that highly polymorphic microsatellites or custom SNP panels.
   - **b)** Once a set of good markers (SNPs or microsatellites) have been identified, a decision must be made as to how the work is shared among laboratories. Since this is multinational project, laboratories in each country would want to participate in the analysis. Wide participation builds capacity for future use of the baseline applying GSI analysis in local mixed-stock fisheries. Some of the basic criteria for marker selection for either SNPs or microsatellites would be easy amplification (robustness). Microsatellite markers are chosen with medium to high polymorphism (>10 to <100 alleles), no upper allele drop-out, no null alleles, easy to score, and the ability to multiplex. Since each SNP has only two alleles. We would want them to be polymorphic over the range of populations. Roughly, one to two microsatellite alleles provide equivalent information as a SNP locus (Kalinowski 2002, Hess et al. 2011), so a lot of SNPs are required for equivalency with a single highly polymorphic microsatellite marker with 50-60 alleles. Data from either SNPs or microsatellite markers can be treated the same.
   - **c)** Loci standardization requires that each laboratory analyze the same set of tissues on their equipment. SNPs comparison between laboratories will be either nucleotides A, C, G, or T, whereas microsatellites will be sized fragments of nucleotide repeat sequences (usually di- or tetra nucleotide repeats). Post-standardization consistency in laboratory analysis is critical when multiple agencies are building a common baseline. We would want to have probably 10-15 good highly polymorphic microsatellite loci and as many useful SNPs as possible.

<u>Stock Structure</u>

5. Stock structure evaluation.

   Initial steps in determining stock structure will be to provide descriptive statistics for the loci set (Lewis and Zaykin 1996). This would include allele frequencies, heterozygosity, private alleles, allelic richness, observed and estimated Hardy-Weinberg equilibrium. Stock structure analysis can take a number of approaches from this point and multiple analyses are useful to confirm and validate the stock structure seen. I Programs like STRUCTURE (Prichard et al. 2000) can be used cluster the multi-locus genotypic data, with or without making assumptions on the number of groupings present. Alternatively, genetic distance such as FST is a commonly used distance measure (Waples and Gagiotti 2006) that can be visualized using trees (Kalinowski 2009) or clustering algorithms (Candy et al. 2011). Other methods

such as MDS (multi-dimensional scaling) appear to be less in favour recently for stock structure analysis, this may be partly due to poor lower dimensional representation of the data than other newer methods (Roth et al. 2003).

Product Outcome of Indian Mackerel May Meeting

At the end of the 2 day meeting the following will be produced:

- *A sampling plan for implementation of the stock structure study on Indian mackerel in the Bay of Bengal Region.*
- *Establishing standardized procedures for marker use, identification and analysis*.

References

Beacham, TD, JR Candy, KL Jonsen, J Supernault, M Wetklo, L Deng, KM Miller, and RE Withler. (2006) Estimation of stock composition and individual identification of Chinook salmon across the Pacific Rim using microsatellite variation. Transactions of the American Fisheries Society. : 861-888.

Candy JR, CG Wallace, and TD Beacham (2011) Distance-based population structure software using mean-field annealing. Molecular Ecology Resources 11:116-125

Hess JE, AP Matala and SR Narum (2011) Comparison of SNPs and microsatellites for fine-scale application of genetic stock identification of Chinook salmon in the Columbia River Basin. Molecular Ecology Resources. 11 (Suppl 1) 1-13

Jayasankar, P. PC Thomas, MP Paulton, and J. Mathew (2004) Morphometic and genetic analyzes of Indian Mackerek (*Rastrelliger kanagurta*) from peninsular India. Asian Fisheries Science 17:201-215.

Kalinowski S. (2002) How may alleles per loci should be used to estimate distance? Heredity 88; 66-65.

Kalinowski S. (2009) How well do evolutionary trees describe genetic relationships between populations? Heredity 102:506-513.

Lewis PO, and D Zaykin designed to accompany the book "Genetic Data Analysis" by Bruce S. Weir (1996, Sinaur Associates).

Menezes MR, S Naik and M Martins (1993) Genetic divergence in the Indian Mackerel *Rastrelliger kanagurta* (Cuv) from the coastal water of peninsular India and Andaman Sea. Indian Journal of Fisheries 40:3 135-141.

Prichard JK, M Stephens, and P. Donnelly (2000) Inference of population structure using multilocus genotype data Genetics 155:945—959.

Roth V, J Laub, M Kawanabe, JM Buhmann (2003) Optimal cluster preserving embedding of nonmetric proximity data. IEEE Transactions on Pattern Analysis and Machine Intelligence 25:1540-1551.

Seeb, LW and 19 coauthors (2007) Development of a standardized DNA baseline for Chinook salmon. Fisheries 32:11 540-552.

Waples RS and O Gaggiotti (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. Molecular Ecology 15:1419-1439.
Winans, GA (1980) Geographic variation in the milkfish *Chanos Chanos*. I. Biochemical evidence. Evolution 34:3 558-574.