# Genetics of the whale shark (*Rhincodon typus*)

**Jennifer V. Schmidt, Claudia L. Schmidt, Robin E. Ernst, Fusun Ozer, Kevin A. Feldheim, Mary V. Ashley and Marie Levine**

The University of Illinois at Chicago, Chicago, IL, The Field Museum, Chicago, IL and The Shark Research Institute, Princeton, NJ

UNIVERSITY OF ILLINOIS
CHICAGO • SPRINGFIELD • URBANA-CHAMPAIGN

SHARK RESEARCH INSTITUTE

# What are microsatellites?

Microsatellites are DNA sequences that contain mono, di, tri, or tetranucleotide tandem repeats

GTGTGTGTGTGT would be referred to as $(GT)_6$

ATCATCATCATC would be referred to as $(ATC)_4$

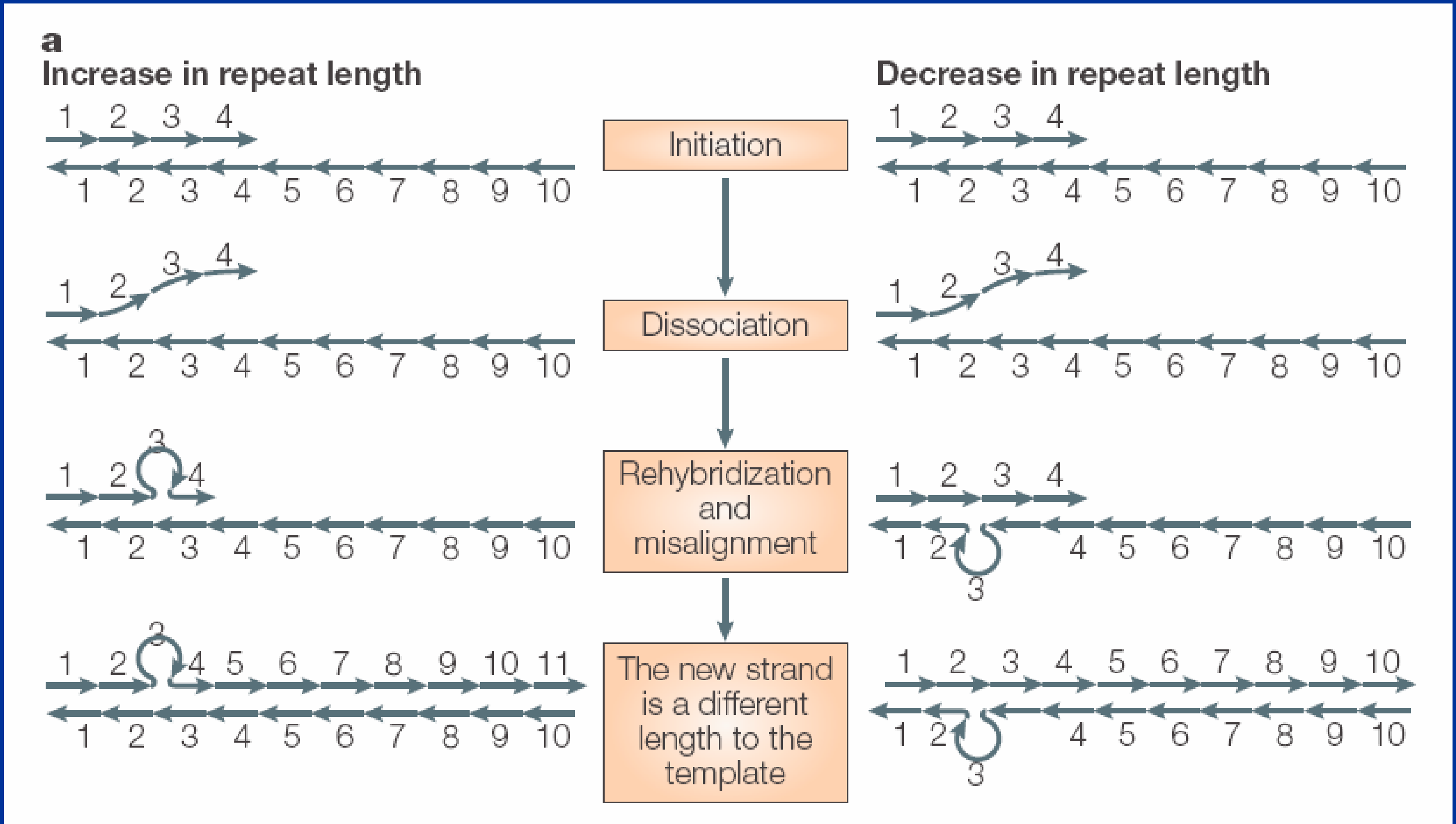
CTGACTGACTGA would be referred to as $(CTGA)_3$

Microsatellite alleles change (mutate) over time, while the flanking sequences generally remain stable

A single microsatellite locus can have many different alleles, which differ in the number of repeats

- -a population of animals can have 10, 20 or more alleles
- -a single animal might carry one allele with 15 repeats and another allele with 20

# Microsatellites are useful for genetics because they change rapidly (on an evolutionary time scale)




**Strand slippage during DNA replication causes in changes in repeat number, and this is not always repaired by the cell**

# Microsatellite analysis in conservation

## Population Studies

Microsatellite variation across different populations provides information about population structure, gene flow between populations (i.e. interbreeding), genetic bottlenecks, etc

In combination with field studies, microsatellite data can help us understand the behavior of the whale shark

## Conservation Biology

Microsatellites can be used to detect changes in populations, effects of population fragmentation and interaction between different populations

Microsatellite data can be used to devise conservation strategies

# A pressing question in whale shark biology and conservation

Do the aggregations of whale sharks seen in different parts of the world comprise a single migratory interbreeding population

or

Are there numerous, *geographically* distinct populations that are reproductively isolated

Determining effective conservation strategies depends on the answer to this question

# Microsatellites are usually species-specific

- Because microsatellites change rapidly, they are usually conserved between only the most closely related species
- There is no closely related species to the whale shark
- New microsatellites must usually be identified for the species being studied

# Isolation of microsatellites

1) Fragment whale shark DNA into small pieces

2) Bind fragmented DNA with short DNA sequences of the repeat you are looking for (i.e. *CACACACACACACACACA*)

3) Use magnetic beads to purify fragments containing repeats

4) Amplify by PCR and clone in bacteria

5) Sequence random clones

6) Generate PCR primers to sequences flanking the repeat

7) Type microsatellites for variability across multiple animals

8) Analyze many animals for a panel of microsatellites

(Our modification of Hamilton, et al)

# Results of the whale shark microsatellite screen

Efficiency of the protocol:

Number of DNA fragments isolated and sequenced -- 98

Number of DNA fragments with a repeat >10 bp -- 22

Range of repeat length -- 11 units – 47 units

Average repeat length -- 20 units

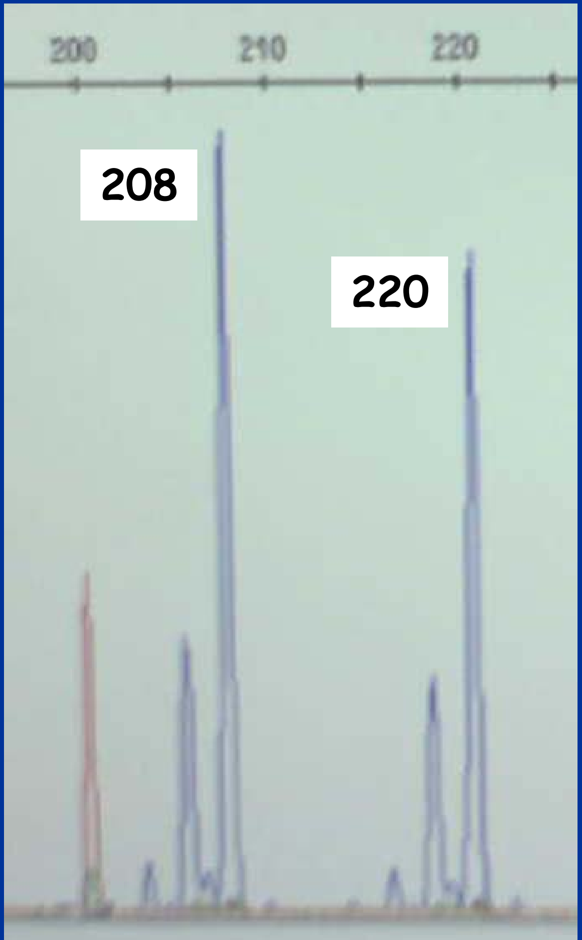
Clones without flanking sequence -- 17

# Microsatellite analysis

Forward primer

```
  1 TCCAAATTGT AGTGAGAATG TTTTTGCAGA CAGAGAGAGA GTTGAGAGAG AGAGAGAGAG
 61 AGAGAGAGAG AGAGAGAGAG AGAGAGAGAA CGCACATCTG CCAGTGCAGC ACTCCCTCTG
121 TGTCACTCAG CCTCAGAGAA AGGAAAC
```

Reverse primer




ABI Sequence analyzer detects fluorescent products, and size standards are used to identify the different alleles

# Whale shark microsatellite loci

| Locus | Repeat | $N_a$ | Observed Heterozygosity ($H_o$) | Expected Heterozygosity ($H_e$) | $F_{IS}$ |
|---|---|---|---|---|---|
| Rt1A | $(TG)_2TC(TG)_6TC(TG)n(TG)_9n(TG)_9TC(TG)_7$ | 5 | 0.687 | 0.681 | -0.005 |
| Rt1L | $(TG)_3TC(TG)_{12}TC(TG)_2TC(TG)_8CG(TG)_2$ | 7 | 0.731 | 0.714 | -0.109 |
| Rt3C | $(CT)_3nTTTTCTGT(CT)_{14}GTCT$ | 4 | 0.472 | 0.402 | -0.18 |
| Rt3G | $(TG)_{14}n(TG)_5n(TG)_4$ | 4 | 0.802 | 0.546 | -0.423 |
| RtG3 | $(CA)_{20}$ | 7 | 0.853 | 0.874 | 0.065 |
| RtG4 | $(GA)_{41}$ | 34 | 0.571 | 1.005 | 0.474 |
| RtG24 | $(CA)_3TG(CA)_3TA(CA)_3(CT)_4CC(CA)_{19}$ | 8 | 0.738 | 0.826 | 0.029 |
| Rt7L | $(CA)_3TGT(GC)_4(CA)_9TACA$ | 3 | 0.446 | 0.455 | 0.02 |

FSTAT and Microchecker were used to analyze the microsatellites for the expected percent of homozygotes and heterozygotes

A negative $F_{IS}$ indicates an excess of heterozygotes, a positive $F_{IS}$ indicates a loss of heterozygotes

Only locus G4 was significantly different from expected

# Microsatellites in other shark species

| Species | Number of Loci | Longest Repeat | Number of Alleles | Heterozygosity | Reference |
|---|---|---|---|---|---|
| Lemon shark (*Negaprion brevirostris)* | 15 | 33 units | 19-43 alleles | 0.69-0.90 | Feldheim et al, 2001 |
| White shark (*Carcharodon carcharias* ) | 5 | 23 units | 2-10 alleles | 0.45-0.95 | Pardini et al, 2000 |
| Nurse shark (*Ginglymostoma cirratum* ) | 9 | 26 units | 2-15 alleles | 0.17-0.90 | Heist et sal, 2003 |
| Zebra shark (*Stegostoma fasciatum* ) | 14 | 30 units | 3-22 alleles | 0.40-0.97 | Dudgeon et al, 2006 |
| Blacktip shark (*Carcharhinus limbatus* ) | 8 | NA | 4-42 alleles | 0.10-0.96 | Keeney et al, 2005 |
| Sandbar shark (*Carcharhinus plumbeus* ) | 5 | 42 units | 12-45 alleles | 0.84-0.97 | Portnoy et al, 2006 |

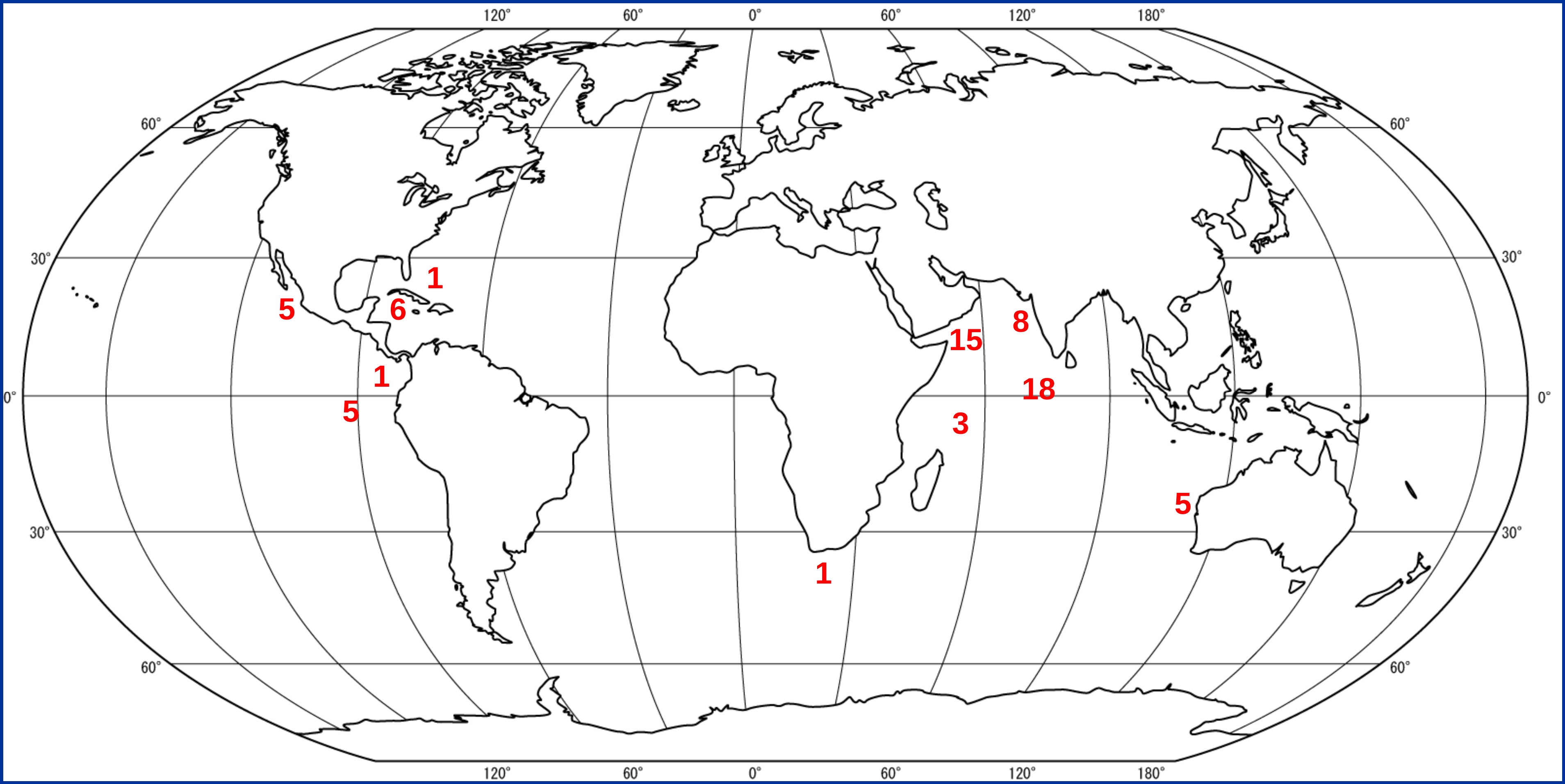
# Private alleles can indicate genetically unique populations

All loci combined show 26 private alleles, with the majority of these found at the highly polymorphic G4 locus

Excluding the G4 locus, there are 7 private alleles distributed across 4 populations (Veraval, Utila, La Paz and Maldives) and across most loci

The Veraval population showed the largest number of private alleles, 3 alleles at 3 different loci

# Whale shark populations studied

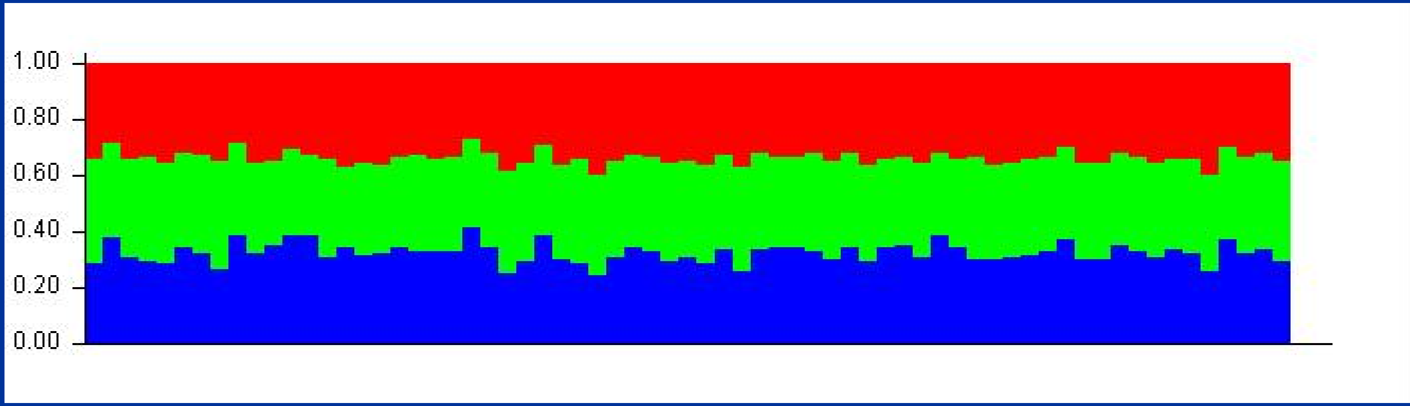


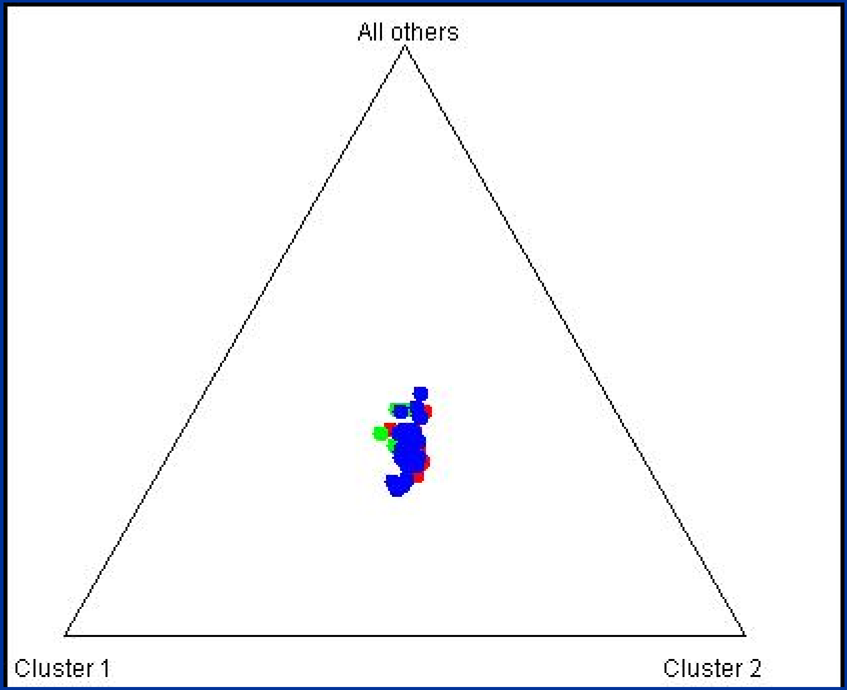

68 whale shark samples were collected during the period 2000-2007

# Male/female sample disparity

- For the 49 samples where gender information was available – 38 sharks were males, and 11 are females
- Female shark samples were collected primarily in the Galapagos and La Paz, with a few collected in Djibouti
- Female sharks ranged in size from 3.5m – 15m, and the 5 largest animals overall were females from the La Paz and Galapagos populations

# Analysis of population differentiation using STRUCTURE




**This analysis indicates all sharks studied fall most easily into a single population grouping**

# F-Statistics predict the genetic similarity or difference between populations

$F_{ST}$ is a measure of genetic population differences

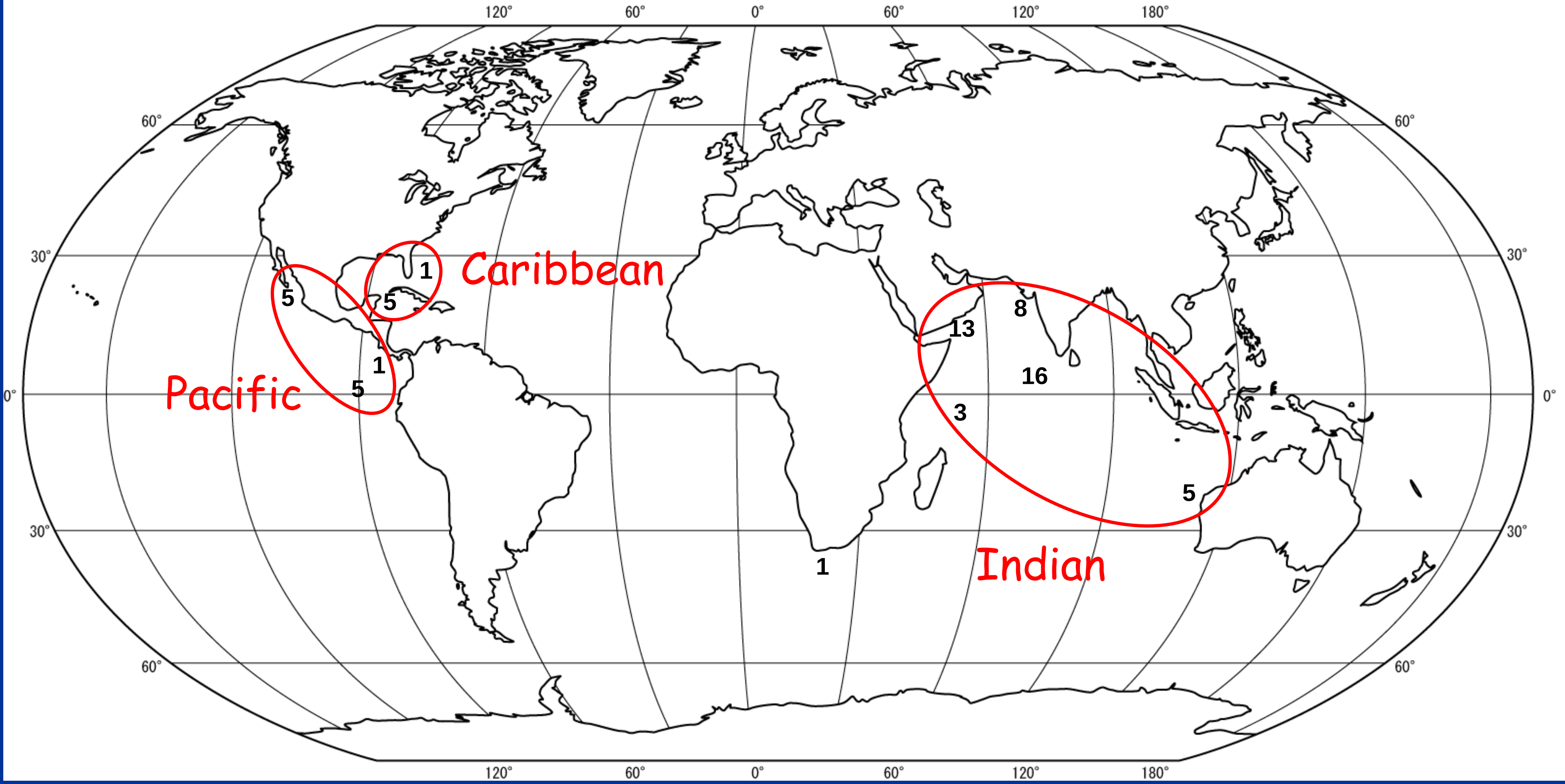
-An $F_{ST}$ value of 1 means two populations are completely different

-An $F_{ST}$ value of 0 means two populations are identical

Interpreting $F_{ST}$ values:

| | |
|---|---|
| 0-0.05 | little or no genetic difference |
| 0.05-0.15 | moderate genetic difference |
| 0.15-0.25 | large genetic difference |
| >0.25 | very large genetic difference |

# Whale shark samples were pooled geographically to overcome small sample sizes



# Population differentiation for geographically pooled whale sharks

| | Pacific | Caribbean | Indian |
|---|---|---|---|
| Pacific | 0 | 0.0387 | -0.0022 |
| Caribbean | | 0 | 0.0296 |
| Indian | | | 0 |

$F_{ST}$ values for comparison of the three populations indicate little or no genetic difference

None of these values are statistically significant ($p<0.05$)

# Differentiation among individual whale shark populations

| | Veraval | Utila | Galapagos | Seychelles | Ningaloo | Djibouti | La Paz | Maldives |
|---|---|---|---|---|---|---|---|---|
| Veraval | 0.000 | 0.044 | 0.024 | 0.087 | 0.012 | 0.000 | -0.001 | -0.004 |
| Utila | | 0.000 | 0.069 | 0.170 | 0.095 | 0.048 | 0.063 | -0.012 |
| Galapagos | | | 0.000 | 0.101 | 0.029 | 0.016 | 0.033 | 0.038 |
| Seychelles | | | | 0.000 | 0.025 | 0.085 | -0.012 | 0.122 |
| Ningaloo | | | | | 0.000 | -0.023 | -0.019 | 0.023 |
| Djibouti | | | | | | 0.000 | 0.000 | -0.002 |
| La Paz | | | | | | | 0.000 | 0.005 |
| Maldives | | | | | | | | 0.000 |

Individual comparisons indicate potential differentiation between Utila animals and Pacific/Indian populations, and between Seychelles animals and some Indian Ocean populations

These values are not statistically significant for Utila, and significance cannot be calculated for Seychelles due to the small number of samples

Analysis of a larger number of samples is needed to prove actual differences

# Conclusions

The same microsatellite alleles were found in whale sharks from geographically distinct populations

No genetic differentiation was found among groups of whale sharks inhabiting different oceans

Few alleles were unique to any one population

These data suggest significant levels of gene flow, by migration and interbreeding, between geographically distinct populations

The data are supported by satellite tracking experiments, and they support the need for a global approach to whale shark conservation

# Future approaches to whale shark genetics

It is likely that an appropriate microsatellite panel can be assembled to accurately identify individual whale sharks

Such data would be most appropriately complied and utilized by the creation of a genetic database

This would allow data sharing by different groups, and would provide data for fine-scale analysis of individual populations, breeding studies, etc.

# Genetic databases are commonly used for model organisms

- Among wild populations, such databases exist for cotton, rice, salmon?
- These databases are curated, but data can be freely uploaded and downloaded by individuals
- Databases provide general experimental protocols, specific procedures for individual loci, ,and should attempt to standardize the analysis across groups and across platforms
- http://www.silksatdb.org

# Points for discussion

- Is there support for building such an effort? Would it be a useful resource?
- Where will the database live, and who will curate it? Can it link to one or more of the photo ID databases?
- How to standardize genetic analysis and interpretation across groups?
- How to protect publishing rights?
  - Unpublished data is offered by some databases, with publishing rights protected, or data can be uploaded but embargoed until publication

We acknowledge the contributions of whale shark tissue samples and other field assistance from: Alex Antoniou; James Hancock; Ingo Lange; Roland Mau; Mark Meekan; Kim Onton; Bruno Pardigon; Matt Potenski; Suzy Quasnichka; Morgan Riley; David Rowat and Michele Vely.