

Standardized GSI Reporting

Standardization issues

- Interlab Standardization of Allele Calling
 - ✓ GAPS example
- Coastwide standardized GSI database
 - ✓ Now and future options
- GSI and CWTs
 - ✓ Can we learn from 3 decades of experience?

Standardization of allele calling

- Problem - different instruments and chemistries between labs cause variable size “calls” for the same allele

- Steps -
 - ✓ assign “curator lab” for each locus
 - ✓ “reference” set of tissues genotyped by all laboratories
 - ✓ “labels” assigned to each allele in “reference” set
 - ✓ each lab standardized call to “labels”
 - ✓ as new alleles found additional “labels” assigned

Standardization of allele calling

➤ Validation

- ✓ “Blind test” of unknown genotypes
- ✓ True = “modal score” for the all labs (most common call)
- ✓ Concordance = % of True
- ✓ Microsatellites (Seeb et al. in press)
 - Blind 2 - initial 95.5% concordance for 9 labs
 - corrected data handling errors 99.4% concordance
- ✓ SNPS (41 loci)
 - 98.8% concordance for 7 labs (Banks et al. draft report)
 - error attributed to two SNPs

➤ Conclusion

- ✓ Successfully replicate “allele calling” across geographically dispersed labs (Seeb et al. in press)

Standardization issues

- Standardization of allele calling
 - ✓ GAPS example
- Coastwide standardized GSI database
 - ✓ Now and future options
- GSI and CWTs
 - ✓ Standardization of GSI with CWT database?

Coastwide standardized GSI database

- Concept for standardized coastwide database
 - ✓ Standardized format of “common language” for sharing and understanding data collection by multiple entities
 - ✓ Ideally single entity selected to house and manage database
 - ✓ Strict ‘Data Neutrality’ is required by the host to ensure full confidence in coastwide shared data
 - ✓ Warehoused coastwide data must be available to all by web-based data retrieval

Coastwide standardized GSI database

➤ At present:

- ✓ Baseline data -GAPS database –NMFS-Seattle
 - “state of the art” web-based
 - GIS capabilities
 - data upload and retrieval
 - Chinook salmon only, easily extend to other species
- ✓ Other agency-based baselines of SNPs and micros for a number species
 - some “allele calling” standardization between labs
- ✓ Mixture (fisheries) data
 - Agency-based, no standardized datasets or coastwide access to standardized mixed-stock genotypes

Coastwide standardized GSI database

➤ Key Questions for future of a standardized database:

Is there a desire to include different marker types in a coastwide database?

Is there a desire to include additional species in coastwide standardized database?

Is there a desire to include catch data (mixture genotypes) as part of a standardized database and where is this database housed?
-options?

Can future efficiencies be gained from adopting CWT (RMIS) data standards and protocols for GSI data?

Coastwide standardized GSI database

1. Extend GAPS database to upload mixture data – at NMFS site.
2. Extend GAPS database to upload mixture data but move database to “neutral party” – e.g. Pacific States Marine Fisheries Commission (PSMFC)
3. Leave GAPS baseline data at NMFS site, develop catch database at “neutral site”

1 Extend GAPS database to upload mixture data – at NMFS site.

➤ Pros

- ✓ Knowledge of genetic database construction
- ✓ Already extending existing database to other species and markers types

➤ Cons

- ✓ “Non-neutral” site
- ✓ Funding associated with further development at NMFS?
- ✓ Less familiar with RMIS and associated data standards.

2 Extend GAPS database to upload mixture data but move database to “neutral party” – PSMFC

➤ Pros

- ✓ Knowledge of similar web-based data (RMIS)
- ✓ Capable of extending existing GAPS database
- ✓ Transferable data standards from RMIS
- ✓ “Neutral” site

➤ Cons

- ✓ Little experience with genetic data but RMIS data is similar?
- ✓ \$ associated with portability, development and maintenance (e.g. training, software costs)

3 Leave GAPS baseline data at NMFS site, develop mixture database at PSMFC

➤ Pros

- ✓ Baseline left with developers
- ✓ Capable of extending existing baseline database
- ✓ Transferable data standards from CWT for catch data
- ✓ “Neutral” site for catch data

➤ Cons

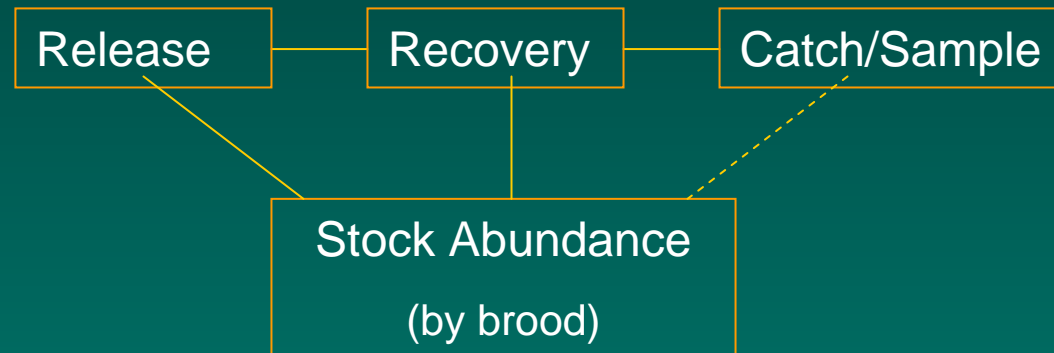
- ✓ Requires close integration between two sites to ensure data compatibility
- ✓ Requires some duplication of effort and expense if at two sites
- ✓ Requires separate funding for both databases

Standardization issues

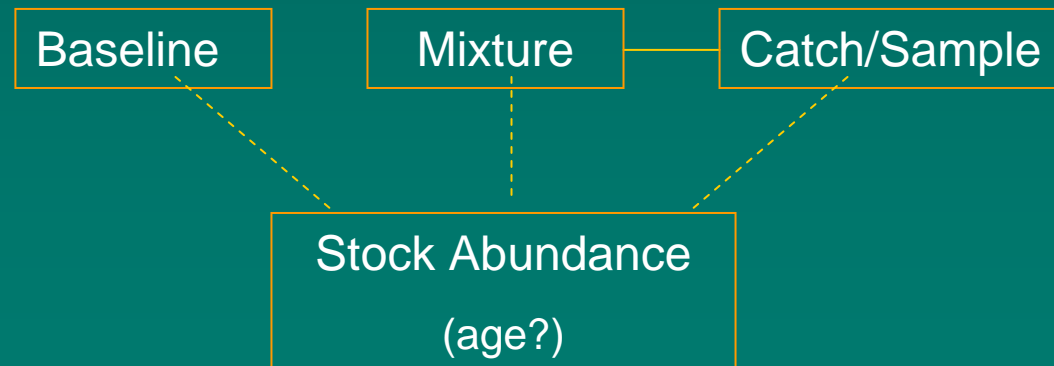
- Standardization of allele calling
 - ✓ GAPS example
- Coastwide standardized GSI database
 - ✓ Now and in the future
- GSI database and CWTs
 - ✓ Expert panel – GSI complement existing CWT programs
 - ✓ Standardization of GSI with CWT to facilitate this?

GSI database and CWT

CWT data structure



GSI data structure?



GSI database and CWT

CWT expansions

$$\text{Estimated CWT Recoveries (Stratum } i,j) = \frac{\text{Total Catch}}{\text{Sampled Catch}} \times \text{No. Observed Recoveries}$$

$$\text{Total Contribution (Stratum } i,j) = \frac{\text{No. Fish Released}}{\text{No. Fish Tagged}} \times \text{Estimated Recoveries}$$

GSI expansions?

All fish tagged with “natural mark” so no release/tagged expansion

$$\text{Total Contribution (Stratum } i,j) = \text{Total Catch} \times \text{Estimated Stock Proportion in Sample}$$

GSI database and CWT

- Important differences between CWT and GSI
 - ✓ no age information for GSI required for cohort analysis
 - ✓ GSI est. stock contrib. probability based (mixture model)
 - Must capture model parameters
 - ✓ GSI estimates contribution for all stocks not just tagged component
 - ✓ GSI cannot separate wild and hatchery contribution if genetically similar
 - ✓ Others?

<u>CWT Field Name</u>	GSI field
Record Code	Y
Format Version	Y
Submission Date	Y
Reporting Agency	Y
Sampling Agency	Y
Catch Sample ID	Y
Species	Y
Catch Year	Y
(Sampling) Period Type	Y
Fishery	Y
Adclip Selective Fishery	Y
Estimation Level	
Catch Location Code	Y
Detection Method	
Sample Type	Y
Sampled Maturity	Y
Sampled Run	Y
Sampled Length Range	Y
Sampled Sex	Y
Sampled Mark	Y
Number Caught	Y
Escapement Estim. Meth.	
Number Sampled	Y
Number Estimated	

Comparison of a subset of CWT catch/sample fields and GSI fields

GSI database and CWT

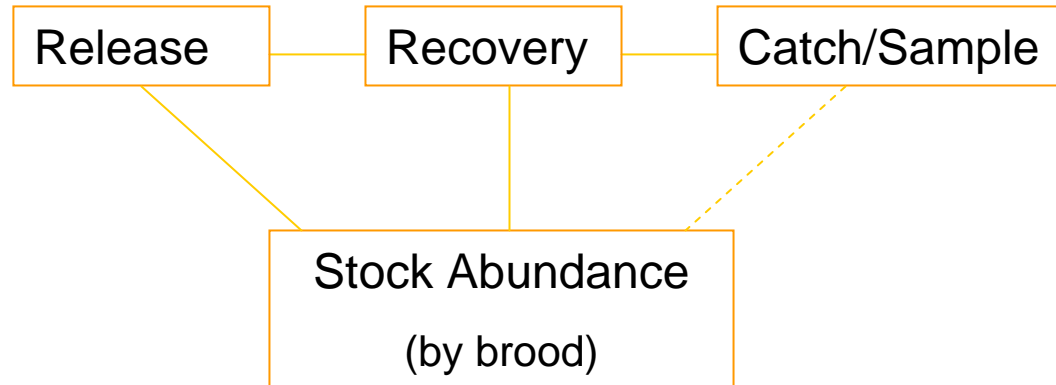
Stock Group	DNA Region	MRP Production Area
Fraser	UPFR	UPFR
	MUFR	UPFR
	LWFR-F	LWFR
	LWFR-Sp	LWFR
	NOTH	TOMF
	SOTH	TOMF
	LWTH	TOMF
Vancouver Island	ECVI	GSMN, JNST
	WCVI	SWVI, NWVI
Mainland Coast	SOMN	JNST, GSMN
	NOMN	RIVR, CCST, NCST
Nass/Skeena	NASS	NASS
	Skeena	SKNA
Queen Charlotte	QCI	QCI
Transboundary	Taku	TRAN
	Stikine	SIK, TRAN
	Alsek	TRAN

Comparison of DFO-CWT Production Areas and GSI Regions

Recomendations

- Continued development of the GAPS baseline to accept additional marker types and species.
- The coastwide standardized GSI database should be located at PSMFC includes GSI data from fisheries.
- Evaluation of standardizing GSI data using RMIS data standards and protocols through the Data Standards Group.

CWT data structure



GSI data structure?

