



NIST Seafood Reference Materials for Authenticating Imported Products

Debra Ellisor¹, Amanda Bayless¹, Clay Davis¹, Tracey Schock¹, Mary Gregg², Angela Folz^{2,3}, Antonio Possolo⁴

¹National Institute of Standards and Technology Chemical Sciences Division, Charleston, South Carolina USA

²National Institute of Standards and Technology Statistical Engineering Division, Boulder, Colorado, USA

³Department of Physics, University of Colorado, Boulder, Colorado, USA

⁴National Institute of Standards and Technology Statistical Engineering Division, Gaithersburg, Maryland USA



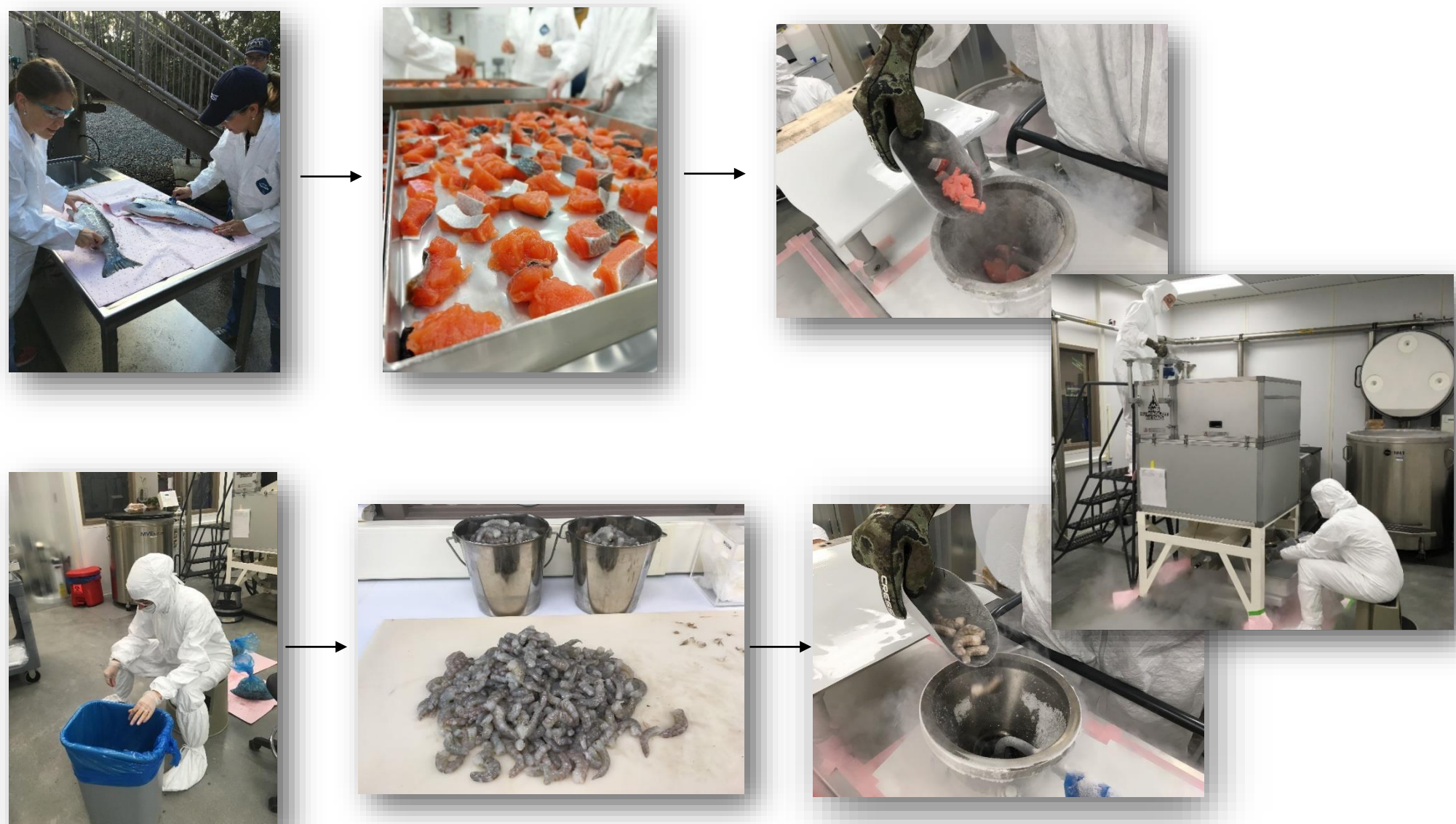
Abstract

Seafood is one of the most highly-traded international commodities and product price is determined by characteristics which can be potentially falsified at import (ie. species, provenance, weight), resulting in inflated payouts, encouraging unsustainable practices, and damaging domestic seafood economies. Additionally, species substitutions and the inclusion of additives such as gels, dyes, and antibiotics commonly found in fish feed can negatively impact consumer health. Food regulatory and customs agencies would benefit from authentic seafood materials to aid in identifying cases of import fraud. To address this need, the National Institute of Standards and Technology (NIST) utilized shrimp and salmon, two of the three most consumed seafood products in the US, to develop authenticity reference materials (RMs). Wild-caught and aquacultured seafood materials were procured from verified sources and edible portions, including fish skin, were cryohomogenized and blended from multiple individuals to produce fresh-frozen homogenates. Values for total fat, fatty acids and total protein were provided, as were genetic assessments using targeted PCR-based methods and Next-Generation Sequencing. Data generated from the latter was used to develop a novel method for quantifying confidence in genome-based identifications. Additional untargeted NMR and HRMS-based analyses were conducted to identify global differences in metabolite and lipid profiles based on material source. RM 8256 Wild-caught Coho Salmon, RM 8257 Aquacultured Coho Salmon, RM 8258 Wild-caught Shrimp, and RM 8259 Aquacultured Shrimp are currently available for authentication and seafood safety studies.

Materials and Processing

All materials were procured from verified/authenticated sources.

- NIST RM 8256 (Wild-caught Coho Salmon): *Oncorhynchus kisutch* hook and line caught ~15 mi off the coast of Alaska
- NIST RM 8257 (Aquacultured Coho Salmon): *O. kisutch* procured from a land-based aquaculture facility in Washington
- NIST RM 8258 (Wild-caught Shrimp): *Farfantepenaeus aztecus* and *Panesus setiferus* trawl caught off the coast of South Carolina
- NIST RM 8259 (Aquacultured Shrimp): *Litopenaeus vannamei* procured from an aquaculture facility in Alabama



- Salmon were scaled, filleted, chopped (including skin) and frozen. Shrimp were beheaded, peeled, chopped and frozen.

- Materials were cryohomogenized three times using a Palla VM-KT vibrating cryomill to create a homogenous fresh frozen powder and material was bottled.

Fat and Protein Value Assignment

Fat Measurements

- Fat measurements were made across 10 jars randomly selected from the production batch for each seafood RM.
- Crude fat content was determined using AOAC Official Method 948.15.

Table 1	
Analyte	% ^(a)
RM 8256, Crude Fat	3.73 ± 0.11
RM 8257, Crude Fat	9.87 ± 0.24
RM 8258, Crude Fat	0.721 ± 0.050
RM 8259, Crude Fat	0.643 ± 0.086

^(a) The assigned value is the average of the values measured for the analyte and the corresponding expanded uncertainty listed is a suitable multiple of the standard uncertainty to achieve 95% coverage

- Additional values, including fatty acid concentrations determined using AOAC Official Method 996.06 and genetic verification using Sanger sequencing and subsequent phylogeny reconstructions, can be found on the respective material certificates and in the material production publications [1,2].

Protein Measurements

- Protein measurements were made across 3 jars randomly selected from the production batch for each seafood RM.
- Crude protein was determined using AOAC Official Method 968.06.

Table 2	
Analyte	% ^(a)
RM 8256, Total Protein	23.77 ± 0.29
RM 8257, Total Protein	20.60 ± 0.86
RM 8258, Total Protein	21.63 ± 0.14
RM 8259, Total Protein	19.30 ± 0.43

^(a) The assigned value is the average of the values measured for the analyte and the corresponding expanded uncertainty listed is a suitable multiple of the standard uncertainty to achieve 95% coverage

Uncertainty in Genomic Verification of Species Identity

- RMs were sent to Authen Technologies for Next Generation Sequencing using an Ion Torrent Personal Genome Machine.
- Fragmented DNA was amplified, and resultant sequences were provided to NIST in FASTQ files.

Data from RM 8256 is shown below for demonstration of methods [3]:

- 100 fragments of 100-200 bp in length were chosen at random from the 52115 fragments provided in the FASTQ file for RM 8256.

- Phred* quality scores were evaluated based on ASCII characters corresponding to each base in the FASTQ files. This provided the uncertainty derived from errors in identification of individual nucleotides.
GGAACAGGATGAACAGTTTCCCACC
6) -) -5; 37>=73- . ---) --) --6

- 25 “perturbed” replicates of each fragment were generated. The error probability at a site is (*p*), the probability it is retained is (1 - *p*), and the probability it is replaced with another letter is (*p*/3). Differences are shown in red below.

Fragment GGAACAGGATGAACAGTTTCCCACC
MC Replicate GGAGCAGGATGATCAGGAAGCCTCC

- For each “perturbed” replicate, 16 high-accuracy alignments to each reference sequence (Table 3) were generated in MUSCLE 5.2 [4].

- Demerau-Levenshtein distances were calculated between each of the 16 high-accuracy alignments (*a*₁ through *a*₁₆) and the reference sequences from vouchered specimens of 10 closely related species available from public databases.

	Table 3						
	<i>a</i> ₁	<i>a</i> ₂	<i>a</i> ₃	...	<i>a</i> ₁₄	<i>a</i> ₁₅	<i>a</i> ₁₆
<i>Oncorhynchus gorbusha</i>	10	12	12	...	12	11	10
<i>Oncorhynchus keta</i>	8	11	10	...	9	10	8
<i>Oncorhynchus kisutch</i>	6	7	7	...	7	9	5
<i>Oncorhynchus mykiss</i>	9	12	8	...	9	8	10
<i>Oncorhynchus nerka</i>	7	13	8	...	8	9	10
<i>Oncorhynchus tshawytscha</i>	8	13	9	...	9	9	10
<i>Parahucho perryi</i>	10	12	10	...	10	11	10
<i>Salmo salar</i>	11	12	11	...	10	11	10
<i>Salmo trutta</i>	10	10	9	...	10	9	9
<i>Salvelinus alpinus</i>	9	13	9	...	12	11	11

- To determine the best alignment in each column, *j* (1 through 16), we found the reference sequence for which the DL distance between *a*_{*j*} and each reference sequence is the smallest of all 10 entries in the column, denoted by red numbers in Table 3.

- The species corresponding to each reference sequence with the smallest DL distance received 1 vote per column. A similar table was generated for all 25 “perturbed” replicates of each sequence.

- The total votes for each species (10) were compiled for all 100 fragments (matrix *V*) and the entropy was calculated, where if all 400 votes per fragment were cast for *O. kisutch*, the probability for *O. kisutch* would be 1 and for all other species would be 0:

$$-(1 \times \log(1) + 9 \times 0 \times \log(0)) = 0 \text{ nat}$$

- The votes cast were weighted based on their informativeness, gauged by a decreasing function of increasing entropy (*S*). A weight proportional to $\exp(-2S)$ was assigned to a fragment producing an identification probability distribution with entropy (*S*), motivated by the relation between entropy and variance for Gaussian distributions.

- Where *w* denotes a vector of non-negative weights adding to 1 in matrix *V*:

$$w_i = \exp(-2S_i) / (\exp(-2S_1) + \dots + \exp(-2S_{100}))$$

for *i* = 1....100, assigned to selected fragments.

- The vote counts in column *j* (1...10, denoting the species) in matrix *V* were pooled into a weighted average and converted to the probability of each species being the source of RM 8256:

$$p_j = \sum_{i=1}^{100} w_i V_{i,j} / T$$

$$\text{where } T = \sum_{j=1}^{10} \sum_{i=1}^{100} w_i V_{i,j}$$

- The probabilities are shown in Table 4, with upper (UPR) and lower (LWR) endpoints of 95% confidence intervals obtained by application of the non-parametric bootstrap. The results verify the species of RM 8256 as *Oncorhynchus kisutch* with 97.7% confidence.

	Table 4		
Species	Probability (%)	LWR (%)	UPR (%)
<i>Oncorhynchus gorbusha</i>	0	0	0
<i>Oncorhynchus keta</i>	0	0	0
<i>Oncorhynchus kisutch</i>	97.7	94.7	99.7
<i>Oncorhynchus mykiss</i>	0.151	0.0396	0.308
<i>Oncorhynchus nerka</i>	0.00237	0	0.00721
<i>Oncorhynchus tshawytscha</i>	0.267	0.0738	0.543
<i>Parahucho perryi</i>	0.00473	0	0.0144
<i>Salmo salar</i>	0	0	0
<i>Salmo trutta</i>	1.89	0	4.75
<i>Salvelinus alpinus</i>	0	0	0

Untargeted Analysis

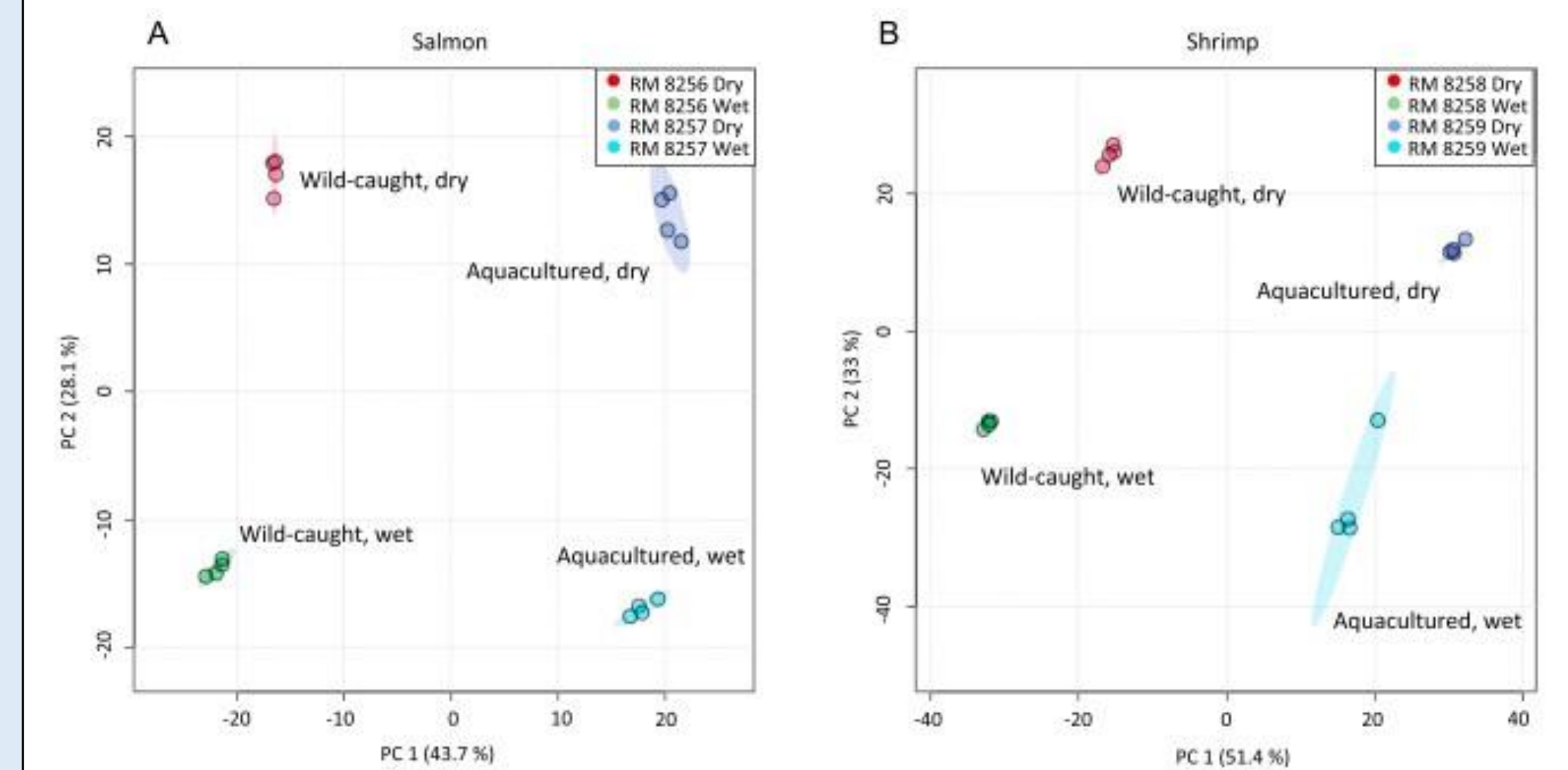
- Metabolites and lipids were extracted from subsamples of each fresh frozen RM using a 2:2:1.8 methanol:chloroform:water solvent system. Freeze dried preparations of each RM were similarly extracted to evaluate processing effects.

- The polar and non-polar fractions were separated and retained for metabolite (polar) and lipid (non-polar) analyses using LC-HRMS/MS.

- Raw data files were processed and searched against multiple local and online mass spectral libraries for feature identification and signal intensities were used to infer relative abundances.

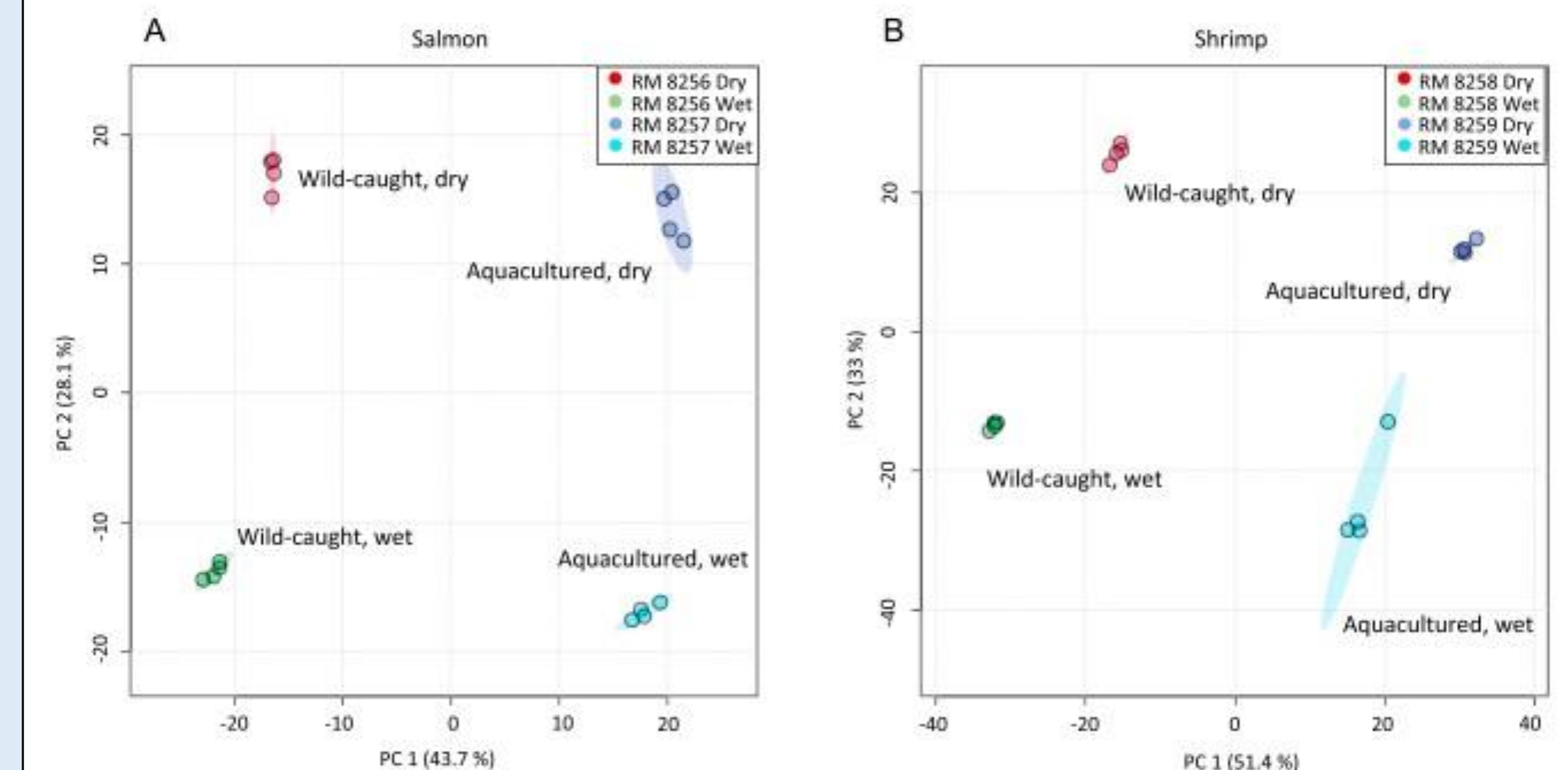
- The seafood RM profiles were compared based on source and preparation method using Principle Components Analysis. LC-HRMS/MS PCA results in positive ionization mode are shown.

Metabolites



- Wild-caught and aquacultured salmon (A) and shrimp (B) showed significant differences in metabolite profiles by LC-HRMS/MS. Processing method (freeze-dried or fresh frozen) also significantly impacted the metabolite profiles of each seafood RM. Ellipses represent 95% confidence intervals.

Lipids



- Lipid profiles were also significantly different between materials based on product source and preparation method by LC-HRMS/MS. Ellipses represent 95% confidence intervals.

Conclusions

NIST seafood safety/authenticity RMs 8256, 8257, 8258, and 8259 can be used to help discriminate between wild-caught and aquacultured salmon and shrimp products. The confidence surrounding genetic verifications can be quantified and future work will focus on mixed-species materials. RM source and preparation can also be distinguished by metabolite/lipid profiles.

References

- <https://doi.org/10.6028/NIST.SP.260-214>
- <https://doi.org/10.1007/s00216-023-04928-9>
- <https://doi.org/10.1007/s00216-024-05724-9>
- <https://doi.org/10.1038/s41467-022-34630-w>



Untargeted Analysis



Genomic Verification