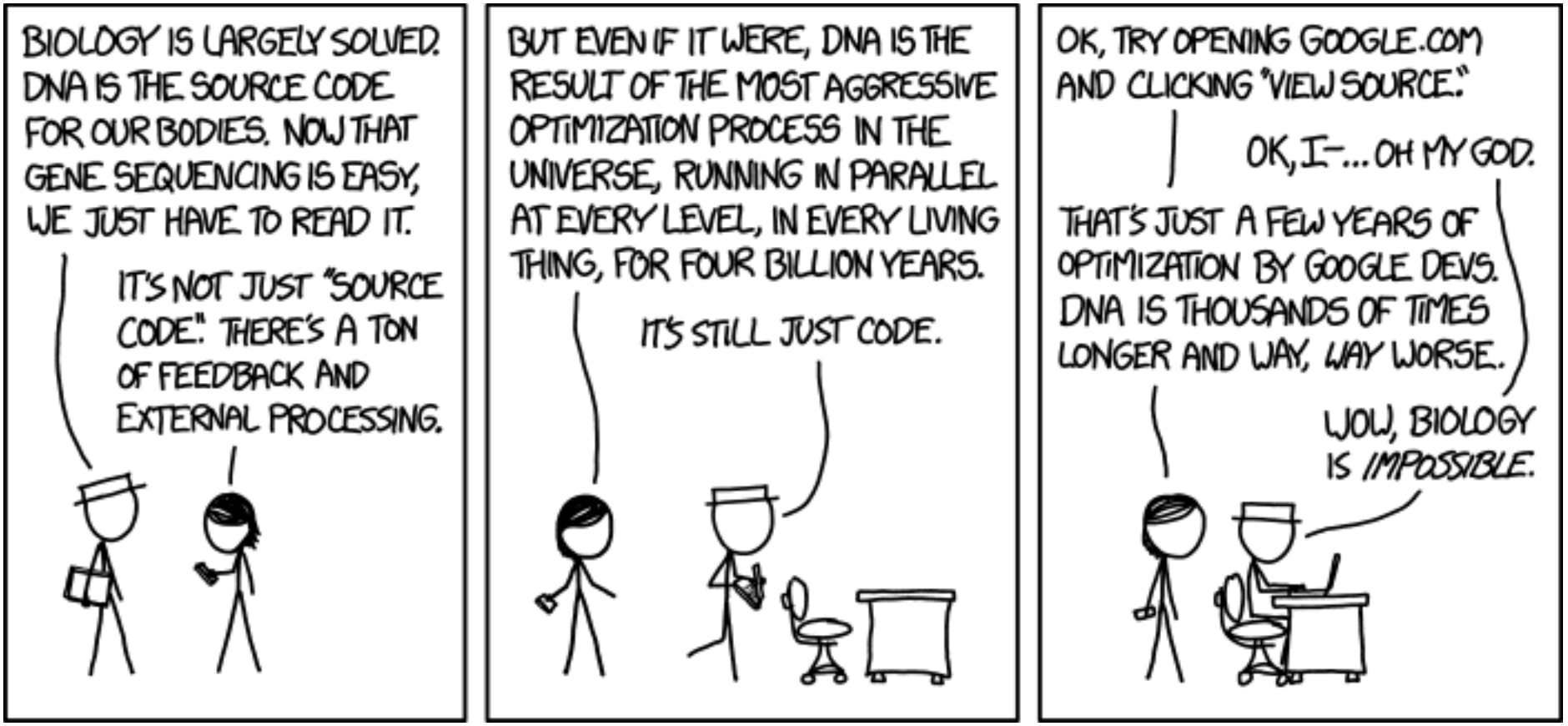
And second


BIOLOGY IS LARGELY SOLVED. DNA IS THE SOURCE CODE FOR OUR BODIES. NOW THAT GENE SEQUENCING IS EASY, WE JUST HAVE TO READ IT.
IT'S NOT JUST "SOURCE CODE." THERE'S A TON OF FEEDBACK AND EXTERNAL PROCESSING.
BUT EVEN IF IT WERE, DNA IS THE RESULT OF THE MOST AGGRESSIVE OPTIMIZATION PROCESS IN THE UNIVERSE, RUNNING IN PARALLEL AT EVERY LEVEL, IN EVERY LIVING THING, FOR FOUR BILLION YEARS.
IT'S STILL JUST CODE.
OK, TRY OPENING GOOGLE.COM AND CLICKING "VIEW SOURCE."
OK, I— ... OH MY GOD.
THAT'S JUST A FEW YEARS OF OPTIMIZATION BY GOOGLE DEVS. DNA IS THOUSANDS OF TIMES LONGER AND WAY, WAY WORSE.
WOW, BIOLOGY IS IMPOSSIBLE.

Motivation

- Metagenomics not just for microbes – can be used to estimate insect diversity

But first…

- I would first like you to group up in groups of 3 to discuss
  - What was your favorite aspects of the Galaxy paper?
  - What weakness, if any, do you see in the Galaxy paper?
  - Brainstorm at least two additional applications of metagenomics sequencing

Great quote #1

- “Because morphological identification is precluded by the destructive nature of the collection procedure, only DNA sequence analysis is feasible making this study de facto metagenomic”

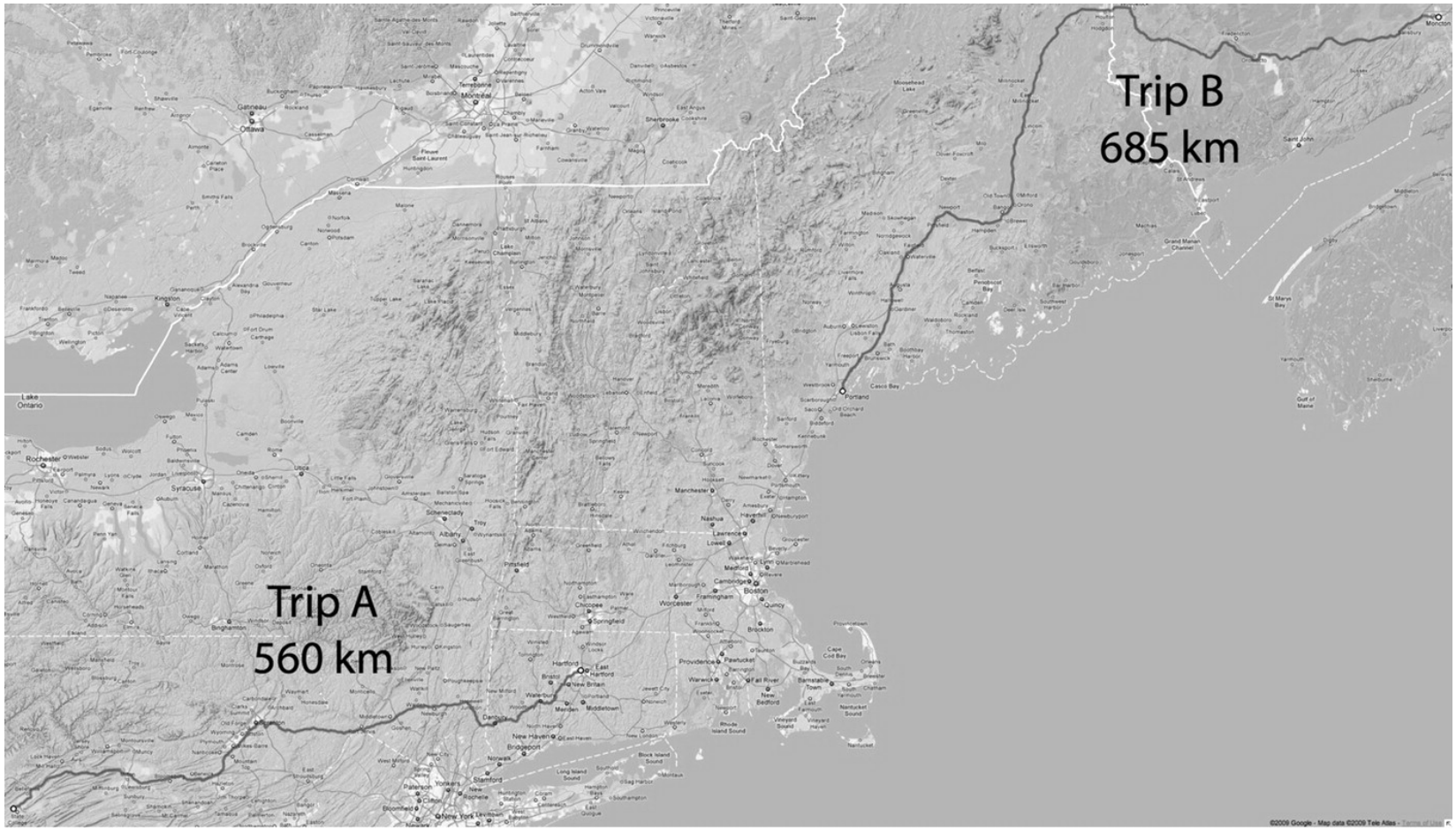
Bioinformatics methods

- Profile-based (aka binning)
  - GC content
  - K-mer content
  - Codon usage bias
- Homology-based
  - **BLAST against known databases**
  - Use of known markers (MetaPhyLan)

Online resources

- CAMERA
- MG-RAST
- BLAST-parsers such as MEGAN
- **Online galaxy portal system**

Map of the collection routes.

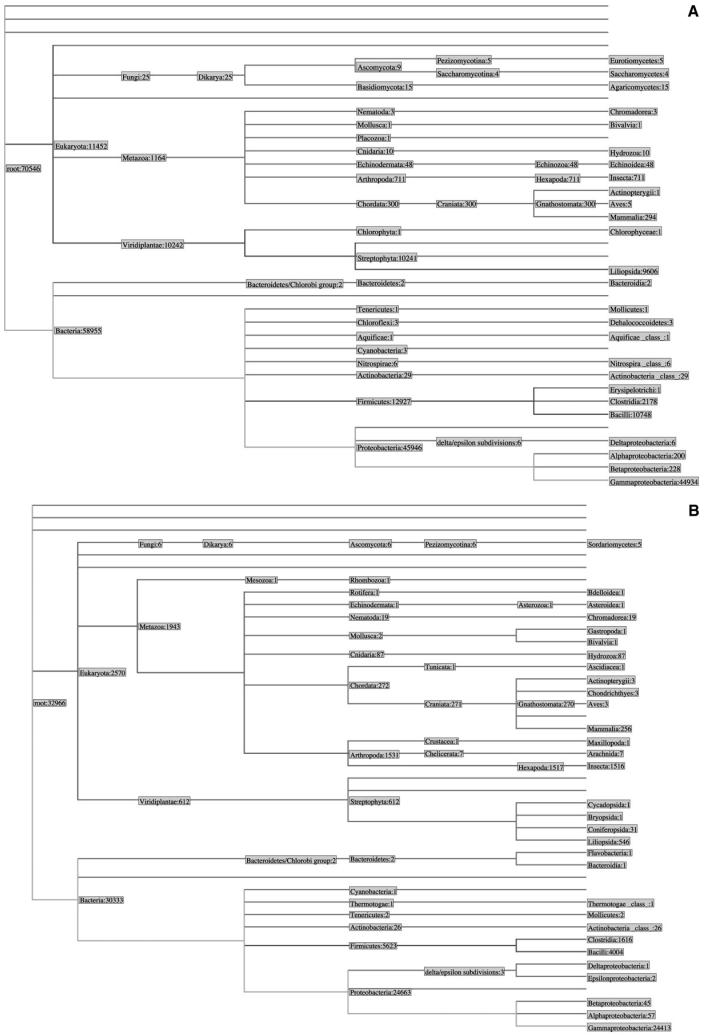


**Sergei Kosakovsky Pond et al. Genome Res.
2009;19:2144-2153**

Steps of Galaxy pipeline

- Assess sequence quality (FastQC)
- Sequence trimming (Trimmomatic)
- Sequence comparisons (megablast; will be Metaphylan or TIPP in future)
- Phylogenetic assessment

A class-level phylogenetic profile of windshield splatter material.



Sergei Kosakovsky Pond et al. Genome Res. 2009;19:2144-2153

Great quote #2

- The most prominent difference between the two trips is in the number of reads identified with green plants (Viridiplantae): 10,242 in trip A versus 612 in trip B…. Because during each trip we collected two samples (left and right sides of the vehicle; see Methods) we were able to trace the majority (9317) of Viridiplantae reads to the left subsample. The most likely explanation for this overabundance is that a piece of plant material (e.g., a leaf or stem fragment) adhered to the collection surface

Great quote #3

- *The list included unexpected entries such as the genus Homo even though the two trips were uneventful. Such matches are likely caused by road debris (which often includes roadkill) adhering to the collecting tape.*

Goals

- A complete, easy-to-use and reproducible workflow for the entire analysis
- Overcome “all vs all “ comparison on very large databases
- Collecting environmental DNA and comparing locations – very similar to YY’s project

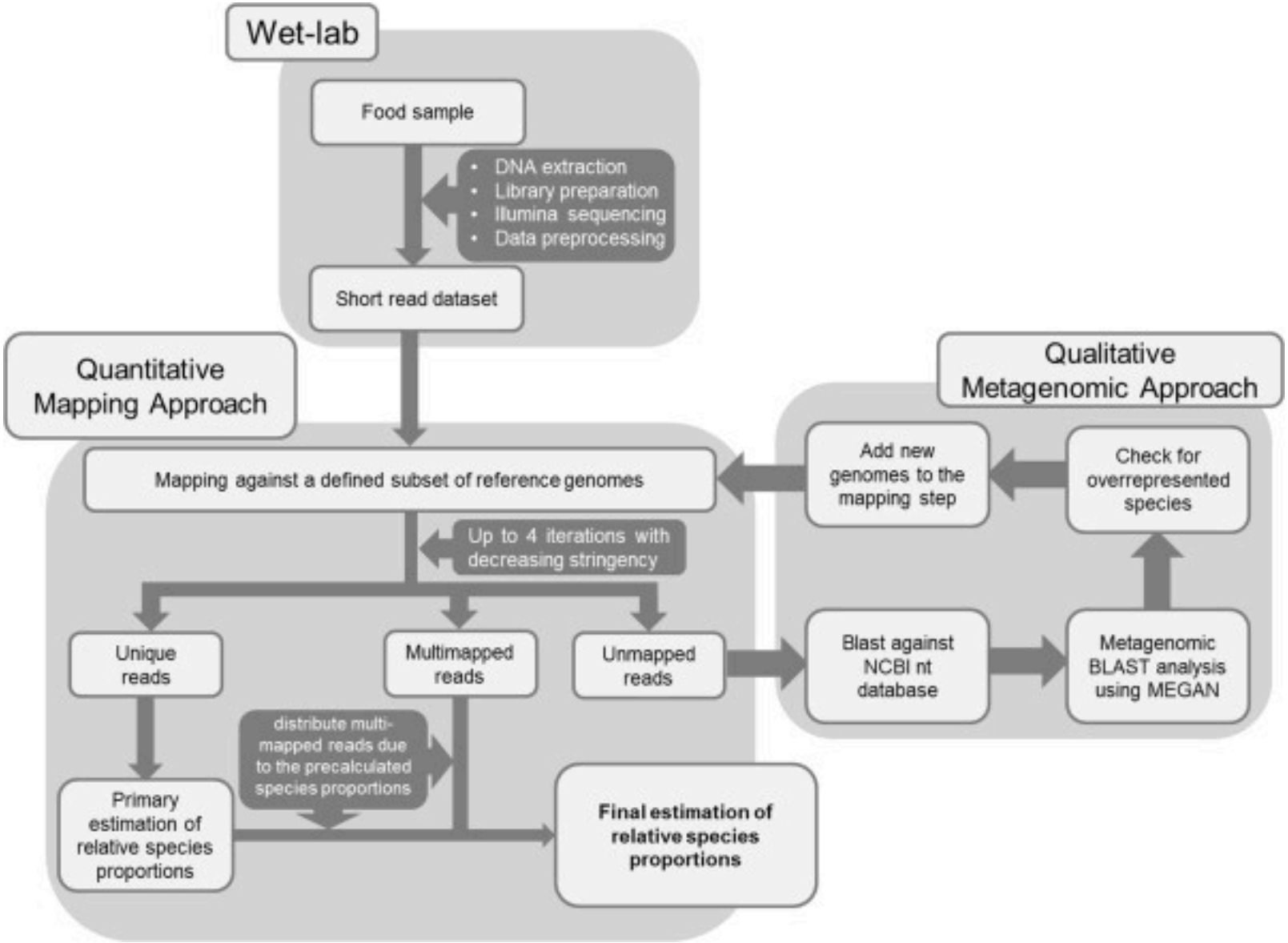
Other reasons this is cool

- You can do it yourself. See:
  - https://usegalaxy.org/u/aun1/p/windshield-splatter
- They named their Dodge Caravan “The Wanderer”
- An excuse to visit friends!

Other applications

- National Biodefense and Analysis and Countermeasures Center (NBACC)
  - Genomics arm of Dept. of Homeland Security

- "Food"seq


Wet-lab
Food sample
DNA extraction
Library preparation
Illumina sequencing
Data preprocessing
Short read dataset
Quantitative Mapping Approach
Qualitative Metagenomic Approach
Mapping against a defined subset of reference genomes
Up to 4 iterations with decreasing stringency
Add new genomes to the mapping step
Check for overrepresented species
Unique reads
Multimapped reads
Unmapped reads
Blast against NCBI nt database
Metagenomic BLAST analysis using MEGAN
distribute multi-mapped reads due to the precalculated species proportions
Primary estimation of relative species proportions
Final estimation of relative species proportions

Great quote

- “Total genomic DNA was extracted from 200 mg of the homogenized calibration sausages “KalD” (boiled sausage) and “KLyoA” (Lyoner sausage)

Table 3

Mapping results for the reference sausage KalD

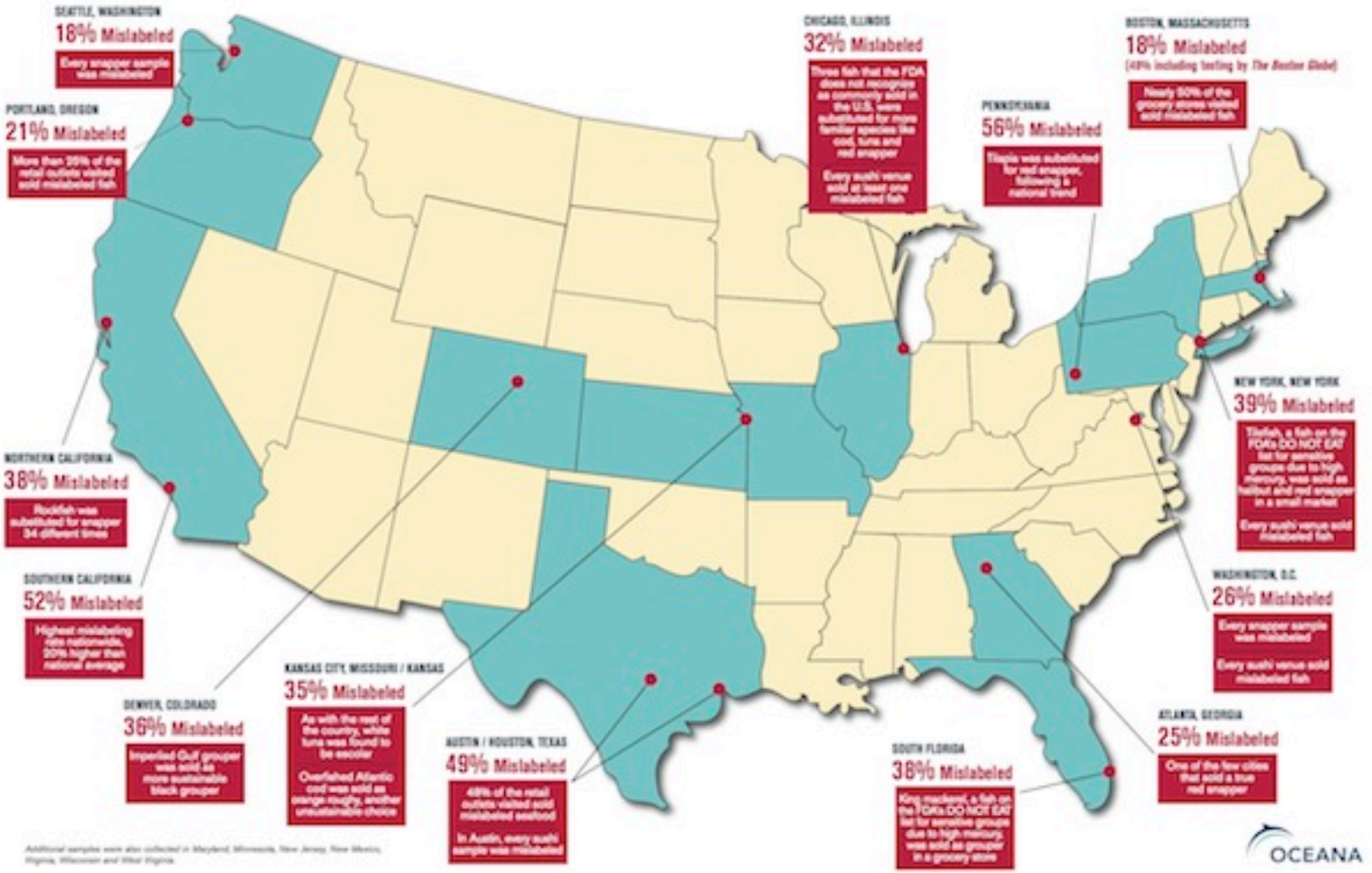
Species	Target value [%]	Proportion [%]		Difference abs. [%]		Difference rel. [%]	
		AFS-quant	AFS-spec	AFS-quant	AFS-spec	AFS-quant	AFS-spec
Cattle	35	36.05 ± 0.04	41.16 ± 0.02	1.05 ± 0.04	6.16 ± 0.02	3 ± 0.11	17.6 ± 0.03
Horse	1	1.27 ± 0.01	1.45 ± 0.01	0.28 ± 0.01	0.45 ± 0.01	27.67 ± 0.67	45 ± 1
Pig	9	7.22 ± 0.05	7.59 ± 0.09	1.79 ± 0.05	1.41 ± 0.09	19.85 ± 0.48	15.67 ± 1
Sheep	55	54.76 ± 0.09	49.71 ± 0.08	0.24 ± 0.09	5.29 ± 0.08	0.44 ± 0.17	9.62 ± 0.15
Waterbuffalo	0	0.64 ± 0.03	0.07 ± 0	0.64 ± 0.03	0.07 ± 0	n.a.	n.a.
Total	100			4 ± 0.1	13.38 ± 0.04		

Quantitative species analysis obtained by Illumina sequencing of DNA from the "KalD" reference sausage [37]. The AFS-quant and AFS-spec approaches (see text for details) were compared. Each dataset tested contained 1 mio of paired-end sequence reads, randomly selected from a larger dataset. Three different sub-datasets (1 mio reads each) were analyzed and mean values plus standard deviations are displayed. "Difference abs." shows the difference between the proportion of reads as determined by AFS ("proportion") relative to the expected amounts existing in the sample ("target value"). "Difference rel." is calculated by dividing "Difference abs." by the expected proportion value.

Photograph by Getty Images

Housed at -112F in a Bern freezer is the Swiss cheesemakers' secret weapon against forgeries. That's where government scientists keep 10,000 strains of milk bacteria that can be used to identify copycats. The country's cheese industry, with 604 million Swiss francs ($658 million) in exports last year, has turned to DNA fingerprinting to fight counterfeits, which Emmental producers estimate have cost them as much as 20 million Swiss francs annually.

Forgeries contribute to declining revenue for an industry already beleaguered by high production costs; and the rising Swiss franc also makes the country's cheese more costly abroad. Exports of Emmental, for example, sank 18 percent in terms of volume in the first half of 2014. Even worse, the Switzerland Cheese Marketing industry association estimates that about 10 percent of cheese labeled as Swiss-made Emmental on supermarket shelves worldwide isn't real.


SEATTLE, WASHINGTON
18% Mislabeled
Every snapper sample was mislabeled
PORTLAND, OREGON
21% Mislabeled
More than 20% of the retail outlets visited sold mislabeled fish
NORTHERN CALIFORNIA
38% Mislabeled
Rockfish was substituted for snapper 34 different times
SOUTHERN CALIFORNIA
52% Mislabeled
Highest mislabeling rate nationwide, 20% higher than national average
DENVER, COLORADO
36% Mislabeled
Imperiled Gulf grouper was sold as more sustainable black grouper
KANSAS CITY, MISSOURI / KANSAS
35% Mislabeled
As with the rest of the country, white tuna was found to be escolar
Overfished Atlantic cod was sold as swai—roughly another unsustainable choice
AUSTIN / HOUSTON, TEXAS
49% Mislabeled
48% of the retail outlets visited sold mislabeled seafood
In Austin, every sushi sample was mislabeled
CHICAGO, ILLINOIS
32% Mislabeled
Three fish that the FDA does not recognize as commonly sold in the U.S. were substituted for more familiar species like cod, tuna and red snapper
Every sushi venue sold at least one mislabeled fish
PENNSYLVANIA
56% Mislabeled
Tilapia was substituted for red snapper, following a national trend
SOUTH FLORIDA
38% Mislabeled
King mackerel, a fish on the FDA's DO NOT EAT list for sensitive groups due to high mercury, was sold as grouper in a grocery store
BOSTON, MASSACHUSETTS
18% Mislabeled
(48% including testing by The Boston Globe)
Nearly 50% of the grocery stores visited sold mislabeled fish
NEW YORK, NEW YORK
39% Mislabeled
Tilefish, a fish on the FDA's DO NOT EAT list for sensitive groups due to high mercury, was sold as halibut and red snapper in a small market
Every sushi venue sold mislabeled fish
WASHINGTON, D.C.
26% Mislabeled
Every snapper sample was mislabeled
Every sushi venue sold mislabeled fish
ATLANTA, GEORGIA
25% Mislabeled
One of the few cities that sold a true red snapper
Additional samples were also collected in Maryland, Minnesota, New Jersey, New Mexico, Virginia, Wisconsin and West Virginia.
OCEANA