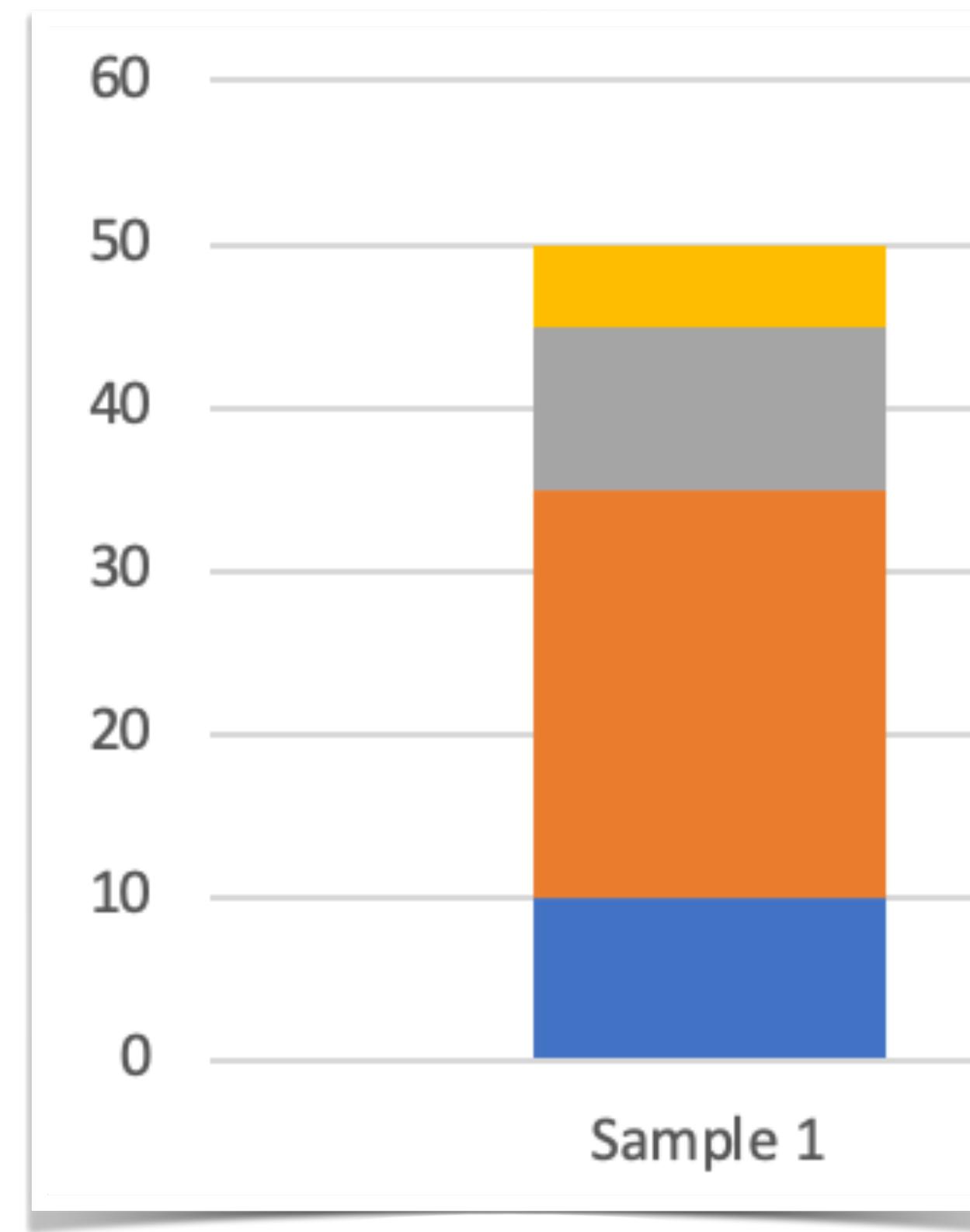


Estimating abundance from DNA-based data

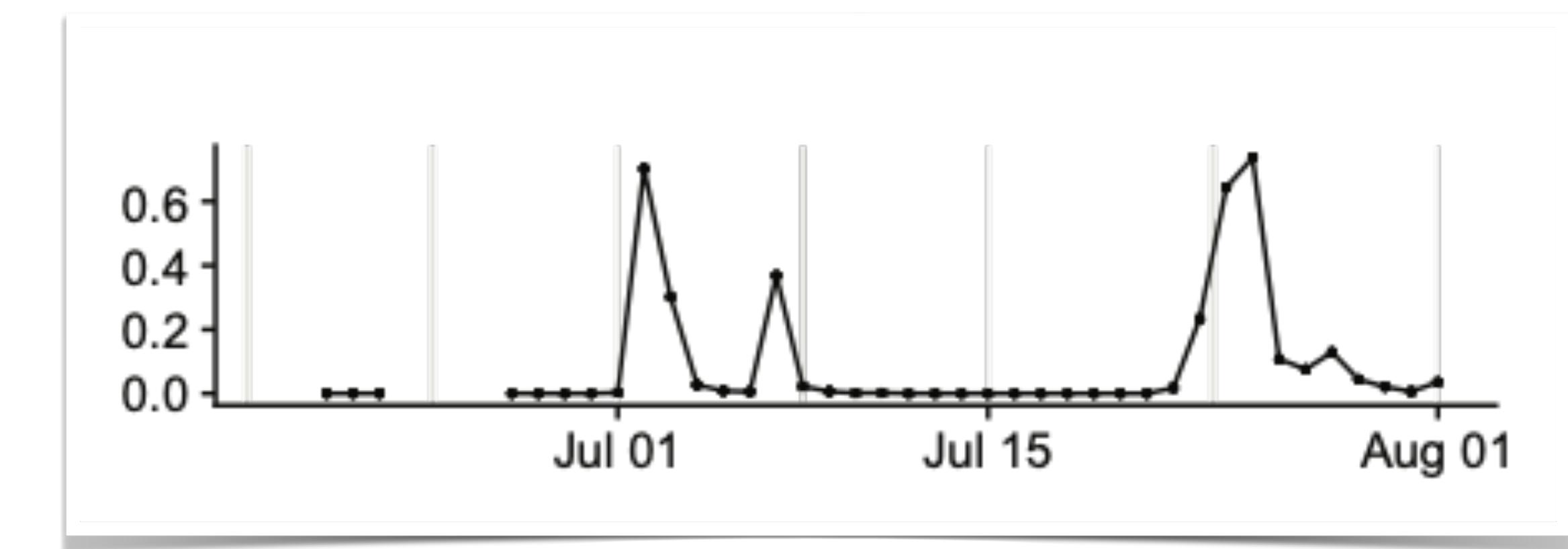
Douglas Yu

Two kinds of abundance information



Across-species (within-sample)

“Species A is more abundant than Species B in this sample” Relative species abundance (RSA).



Within-species (across-sample)

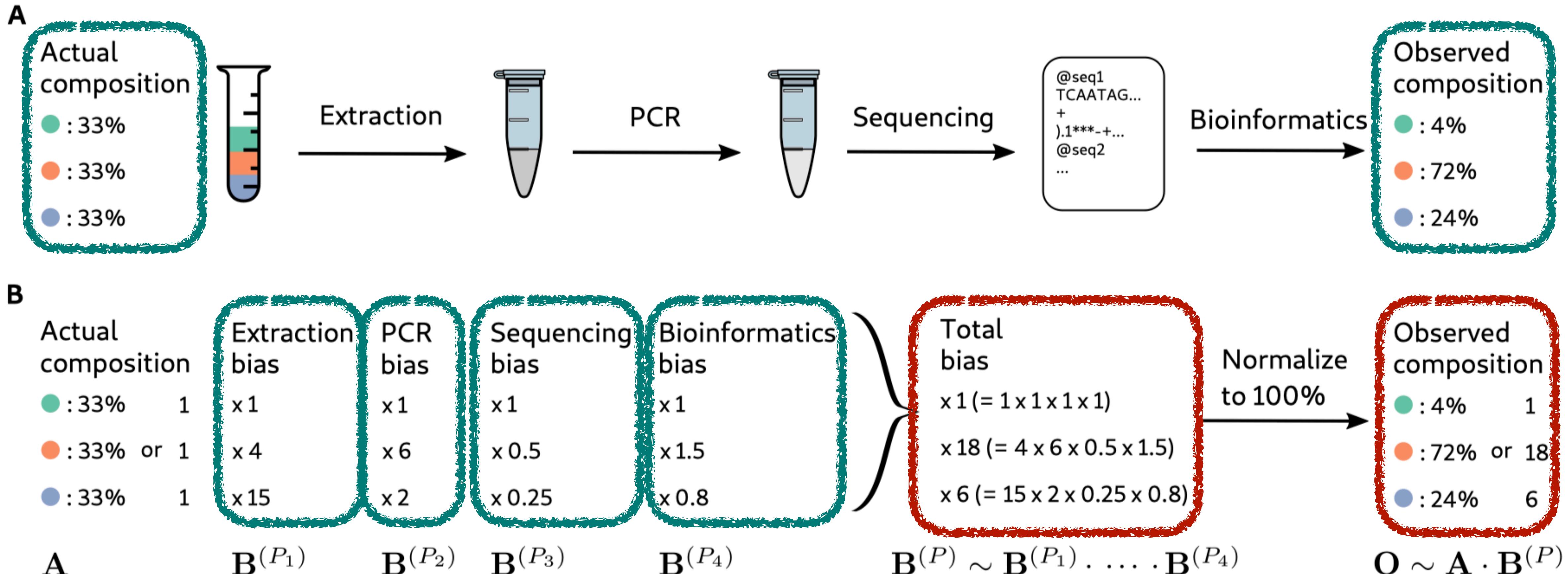
“Species A is more abundant in these samples, less abundant in those samples.” Follow one species over time and space.

Take home

Species

		Within species											
		A	B	C	D	E	F	G	H	I	J		
Taxonomies Time or Space gradient	R4069-		R2420_Insecta_Hy	R2516-	R3842-	R2002-	R4278-	R5123-	R6630-				
	6_Insecta_Diptera_	R7337_Insecta_Dip	menoptera_Vespidae	6_Insecta_Diptera	2_Insecta_Hymenopt	2_Arachnida_Araneae	5_Insecta_Raphidio	57_Insecta_Diptera	2_Insecta_Psocodea_D	R9119_Insecta_Dipt			
	Rhagionidae_NA_NA	era_Syrphidae_Mile	Dolichovespula_ma	Rhagionidae_Symp	era_Vespidae_Eumen	Gnaphosidae_Orodra	ptera_Raphidiidae_A	Rhagionidae_NA_N	asydemellidae_Teliapso	era_Muscidae_Phaon			
	_BOLD_ACX1094_siz	ia_pulchra_BOLD_A	culata_BOLD_AAB63	harmomyia_NA_BOLD	es_consobrinus_BOLD	ssus_canadensis_BOLD	gulla_NA_BOLD ACA	A_BOLD_ACU3161_si	cus_conterminus_BOLD	ia_nigricauda_BOLD_			
	1e=14821	Y9056_size=62348	88_size=239475	1CY3832_size=556	_AAG9053_size=3920	_ACD2419_size=31	6997_size=2491	ze=353	_AAP4627_size=5126	AAP6480_size=1261			
	2	5927	2925	2520	1720	1100	1087	1040	633	547	540		
	3	1638		0	0	0	0	1236	0	910	0		
	4	4950	303	1012	1822	0	0	5940	630	1935	7268		
	5	72550		0	0	0	0	0	0	350	0		
	6	0		115344	0	0	0	0	0	0	0		
This species abundance)	7	1552		12762	0	0	0	1282	0	127	412		
	8	0		82983	0	0	0	0	0	650	0		
	9	0		22231	0	407	0	0	0	0	345		
	10	0		0	0	0	0	0	0	0	336		
	11	0		0	0	0	0	0	0	0	0		
	12	1102		0	2204	2693	0	0	0	0	0	367	
	13	731	658	6888	0	2588	0	0	0	0	0	354	
	14	0		91047	0	0	0	0	0	0	0	0	
	15	325		9825	0	475	0	5763	0	990	0		
	16	0	605	1773	0	1038	260	0	0	0	0	476	
Samples	17	3960	2793	4560	1587	0	0	3600	0	640	0		
	18	9417	1218	4017	0	0	0	1000	0	1317	0		
	19	300		62700	0	0	0	0	0	100	0		
	20	845		24873	0	0	0	0	0	1647	98		
	21	0		58940	0	0	0	0	0	0	760	0	
	22	1944		17032	0	0	0	0	0	0	280	0	
	23	3000	6440	3633	0	867	0	0	0	0	567	233	
	24	0		1298	0	0	0	0	0	0	330	0	
	25	0		75748	0	0	0	0	0	0	0	0	
	26	0		8879	0	0	0	0	0	0	1792	0	
38 samples	27	0		12162	0	4317	0	0	0	0	0	0	
	28	1346		117167	0	0	0	0	0	0	0	1108	
	29	0		73068	0	0	0	0	0	0	0	0	
	30	0		3160	0	0	0	0	0	0	0	160	
	31	64	483	20055	0	0	0	0	0	0	0	0	
	32	0		155440	0	0	0	0	0	0	0	367	
	33	0		0	0	0	0	0	0	0	0	0	
	34	0		84480	0	0	0	0	0	0	0	0	
	35	315		18320	0	0	0	0	0	0	0	0	
	36	0		0	0	0	0	0	0	0	0	0	
	37	0		3335	0	0	0	0	1913	0	0	0	
	38	0		2215	0	65	1051	0	0	0	0	0	

Across-species quantification is difficult



eLife 8:e46923.

Consistent and correctable bias in metagenomic sequencing experiments

Michael R McLaren¹, Amy D Willis², Benjamin J Callahan^{1,3*}

Across-species quantification is really difficult!

Even with the same underlying biases, it is possible for a given species to become more or less abundant in a sample after metabarcoding.

Imagine this is a diet study. You only see the Observed proportions. Your conclusions about the relative abundances of prey items would be incorrect.

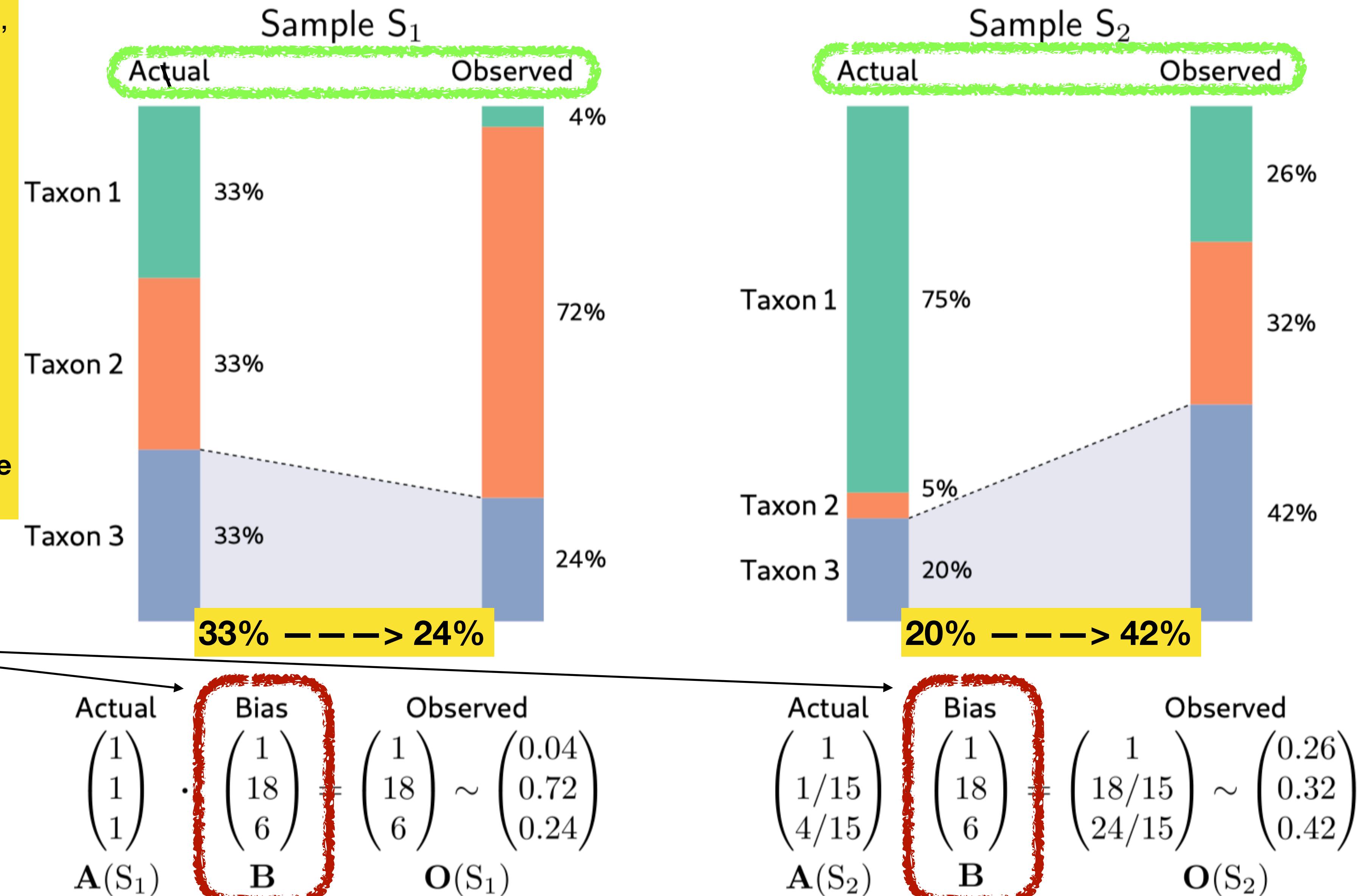
Diet studies try to achieve “across-species” quantification:

“Taxon 3 has higher/lower relative abundance than Taxon 2”

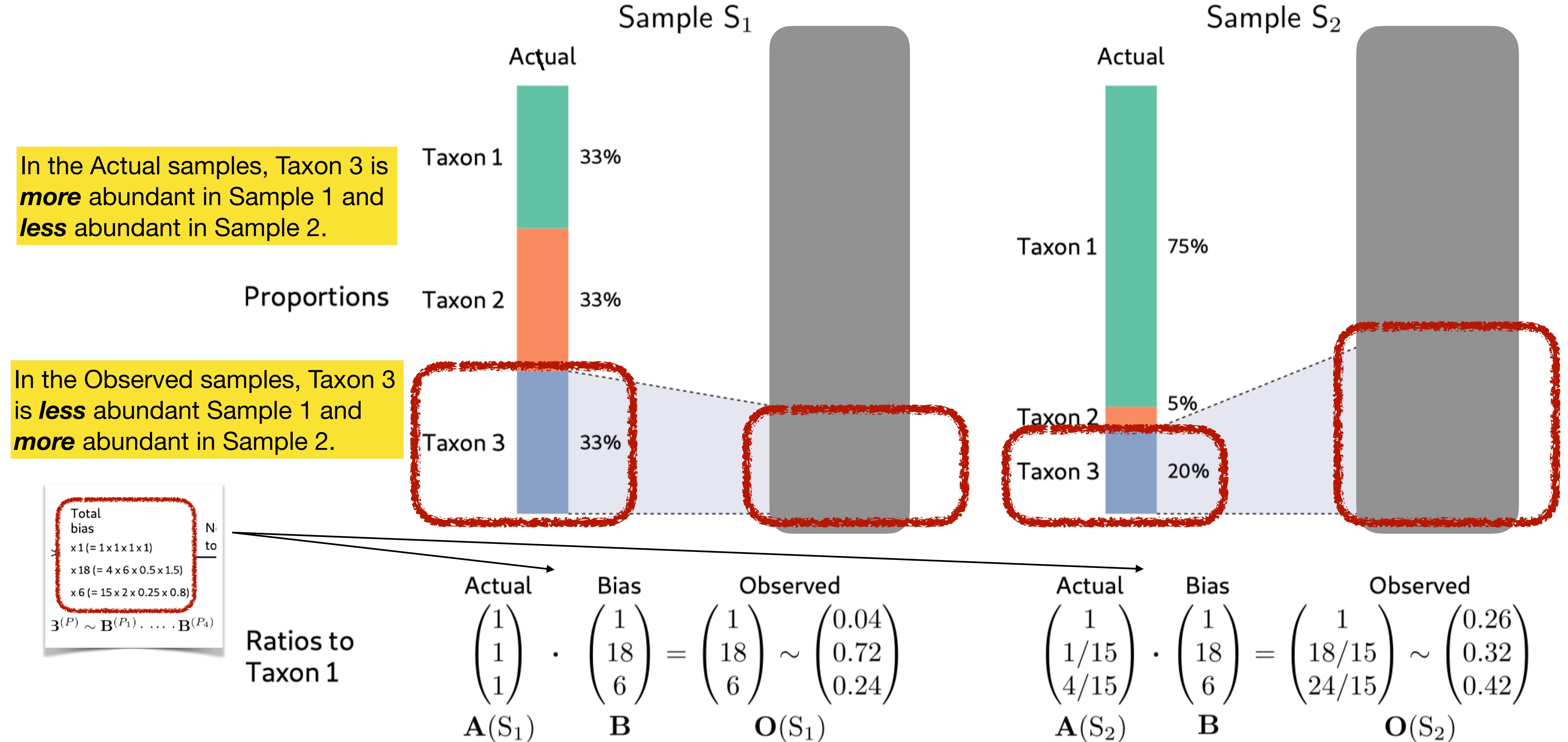
$$\begin{aligned} \text{Total bias} \\ \times 1 (= 1 \times 1 \times 1 \times 1) \\ \times 18 (= 4 \times 6 \times 0.5 \times 1.5) \\ \times 6 (= 15 \times 2 \times 0.25 \times 0.8) \end{aligned}$$

$$B^{(P)} \sim B^{(P_1)} \dots B^{(P_4)}$$

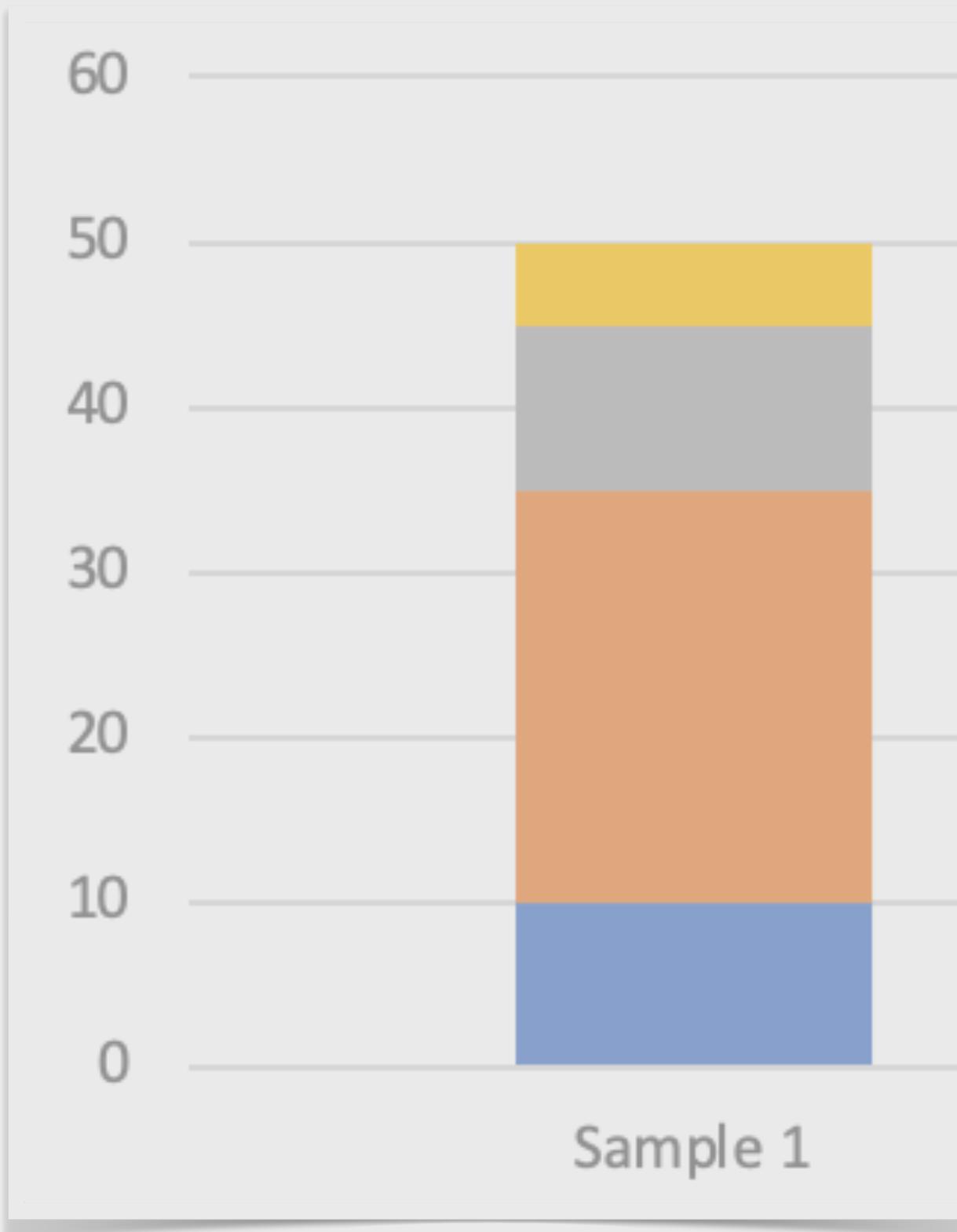
Ratios to Taxon 1



Species biases can reverse your conclusions!

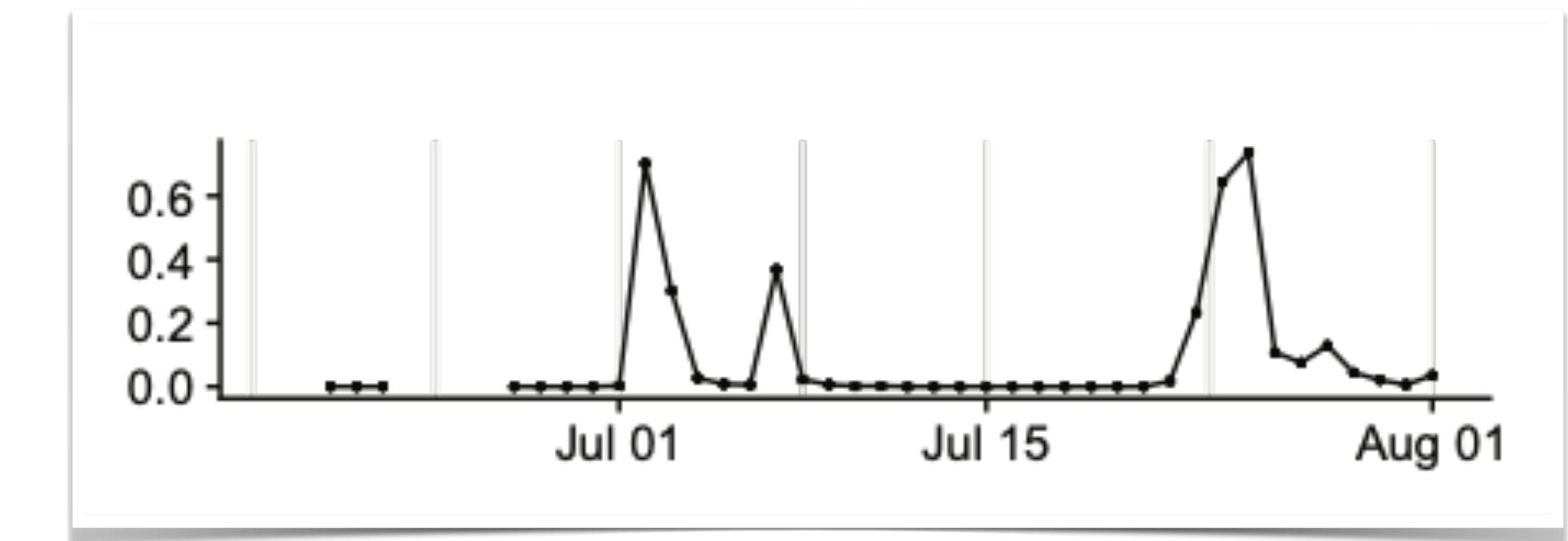


Two kinds of abundance



Across-species (within-sample)

“Species A is more abundant than Species B in this sample” Relative species abundance (RSA).

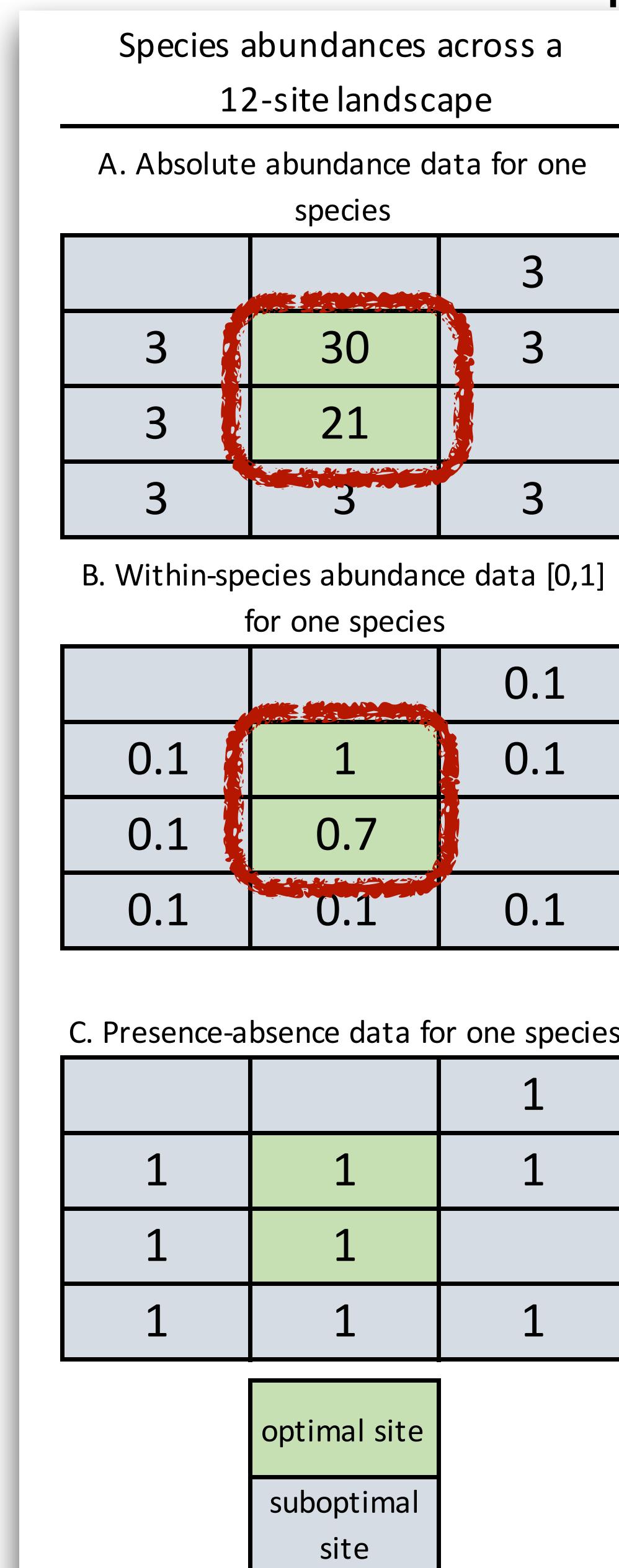


***Within*-species (across-sample)**

“Species A is more abundant in these samples, less abundant in those samples.”

Follow each species separately over time and space.

Within-species abundance information is quite useful: 1

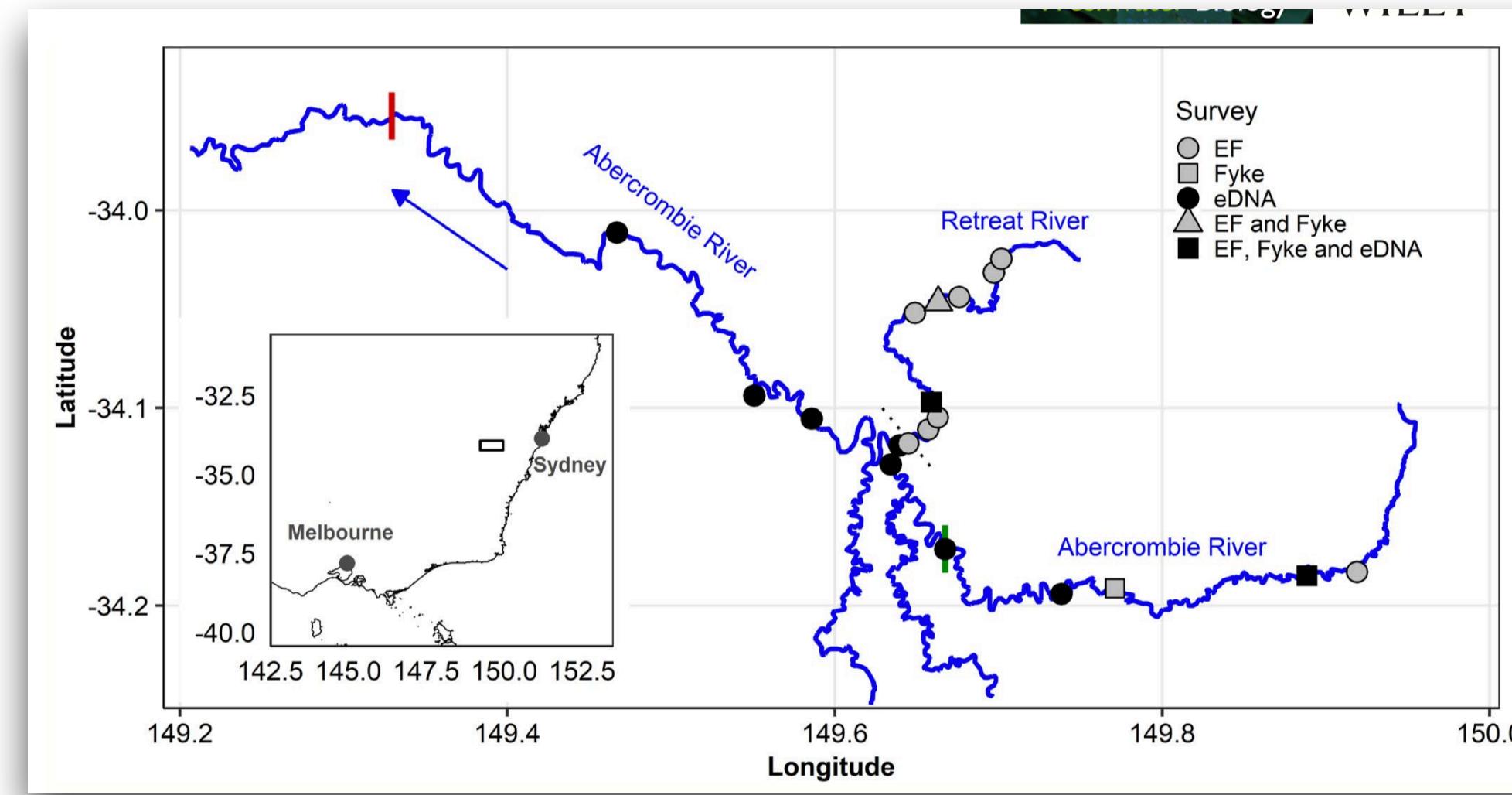


A. Imagine that a species is found in many sites but that **only two sites are optimal (green)** with high abundances, while most sites are suboptimal (grey) with low abundances maintained by immigration.

B. Even if we rescale the numbers to [0,1] (within-species abundance), we still see that this species is more abundant in the green sites, which still lets us hypothesise that the green habitat is optimal.

C. Note that conversion to presence/absence loses the distinction between green and grey habitat.

Within-species abundance information is quite useful: 2



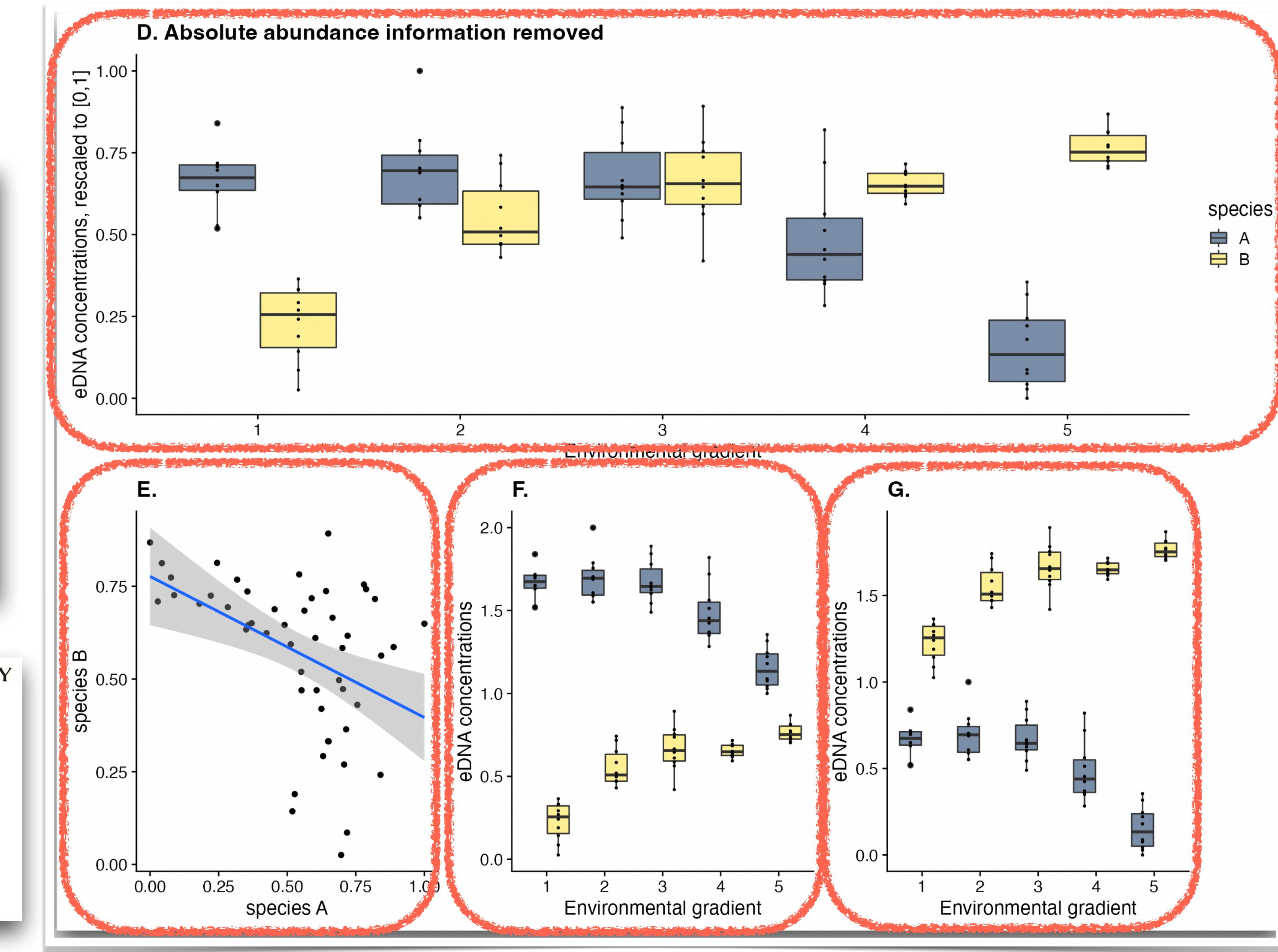
ORIGINAL ARTICLE

Freshwater Biology

WILEY

The value of quantitative environmental DNA analyses for the management of invasive and endangered native fish

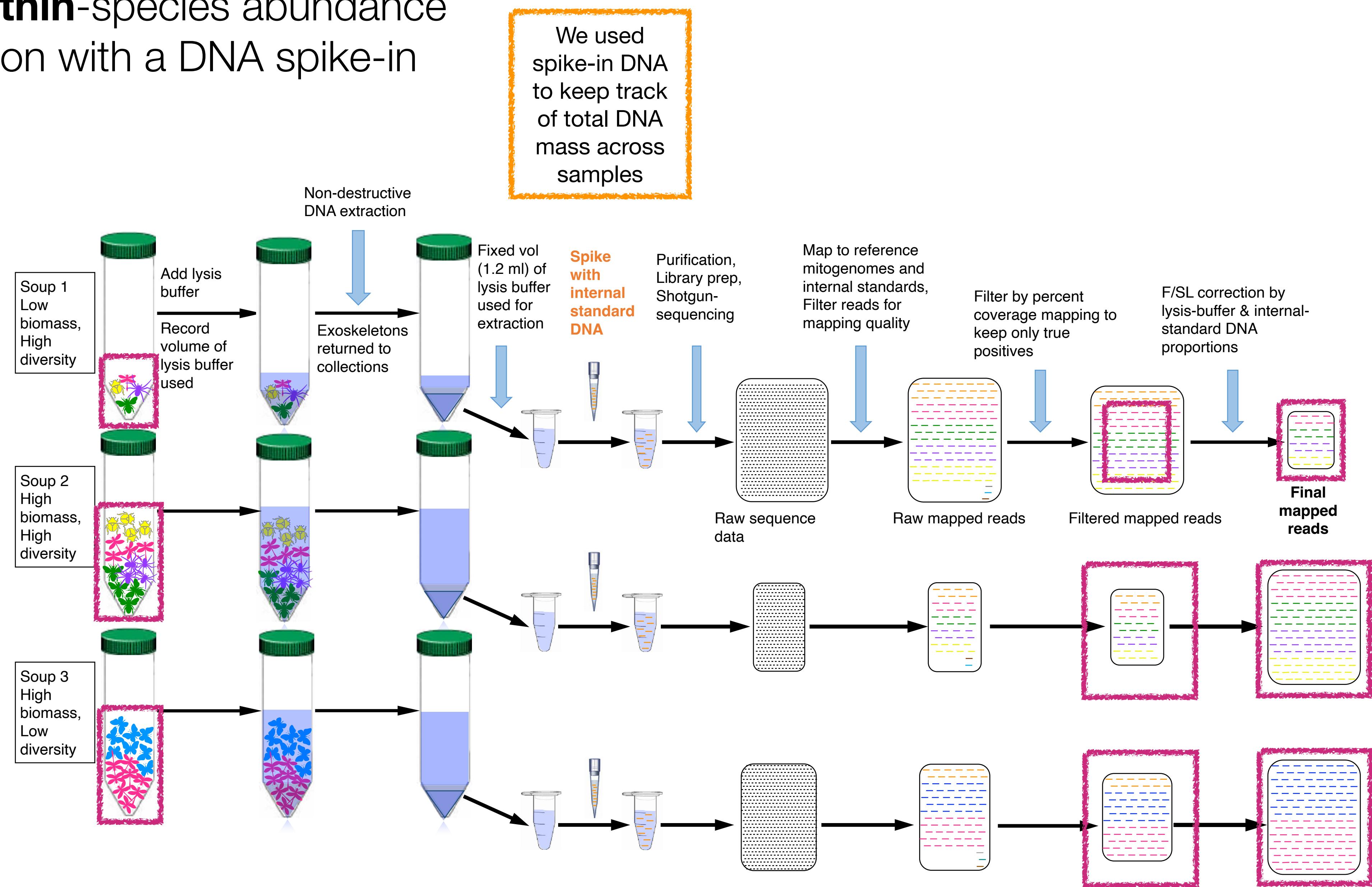
Jack Rojahn^{1,2} | Luke Pearce³ | Dianne M. Gleeson^{1,2} | Richard P. Duncan¹ |
Dean M. Gilligan⁴ | Jonas Bylemans^{1,2,5}



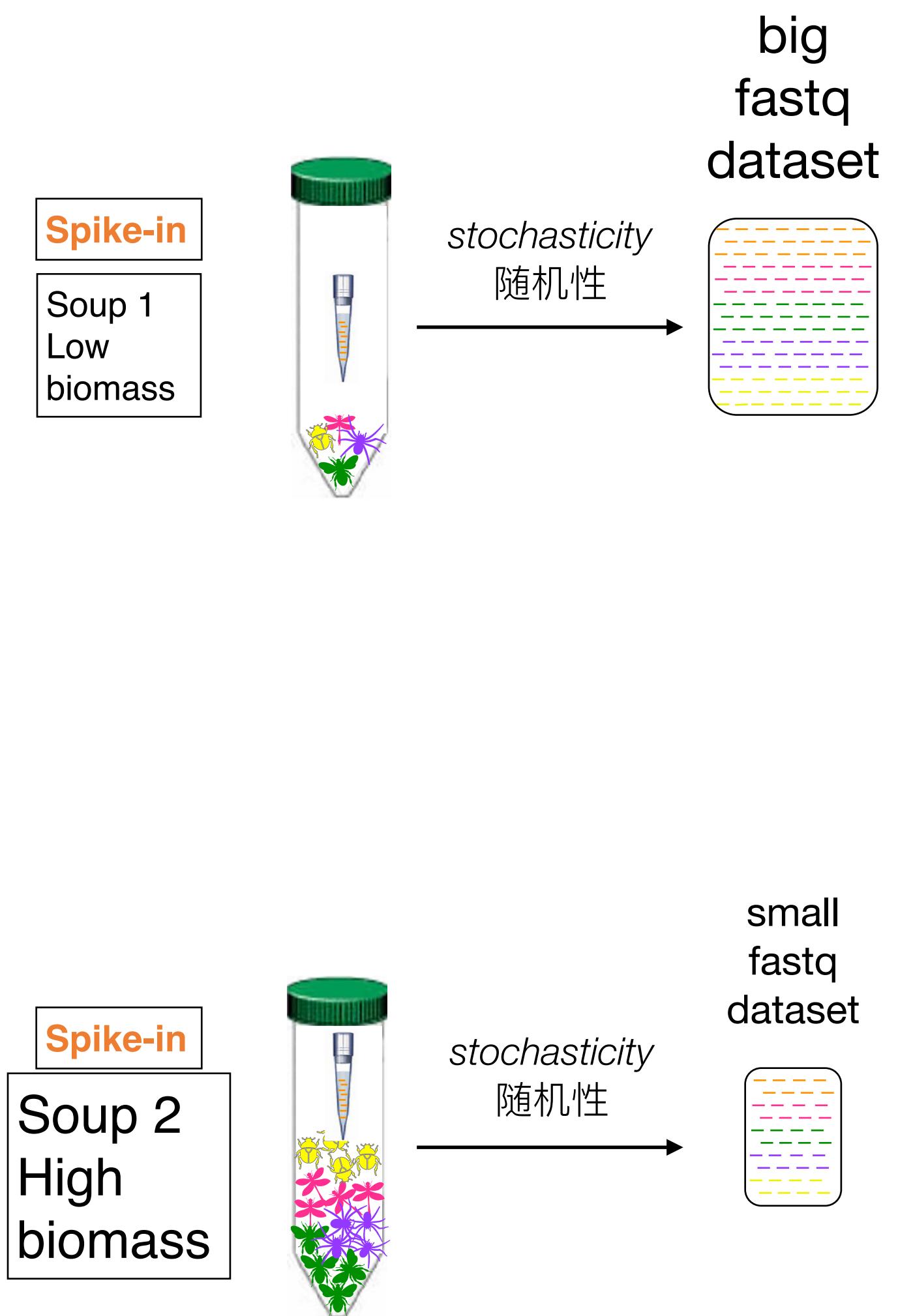
Here, we have 2 species. We do not have across-species abundance information (we do not know which species is more abundant).

However, we can compare within-species abundance changes across an environmental gradient, and we see that the two species are negatively correlated. Rojahn et al. suggest that the invasive species is competitively displacing the native species.

Extract **within**-species abundance information with a DNA spike-in

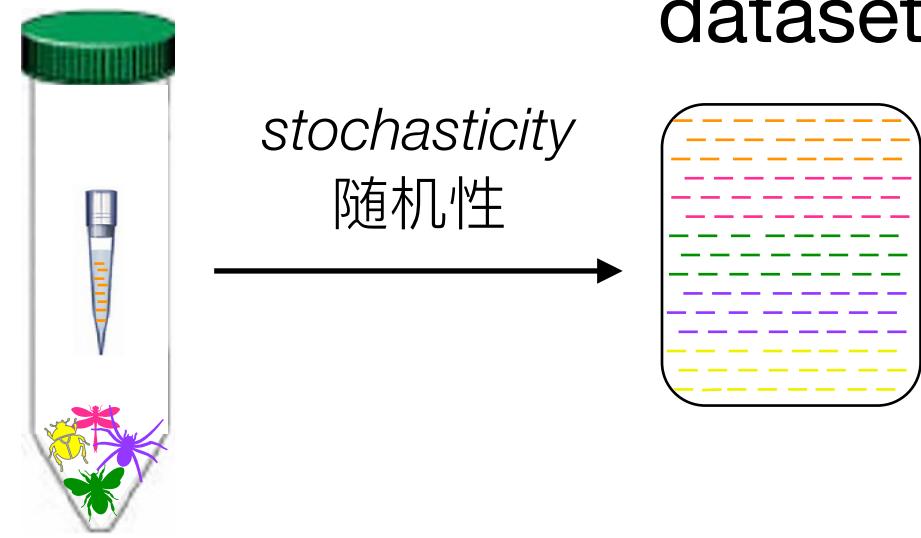


A big problem is “pipeline noise”
We need to get rid of pipeline noise

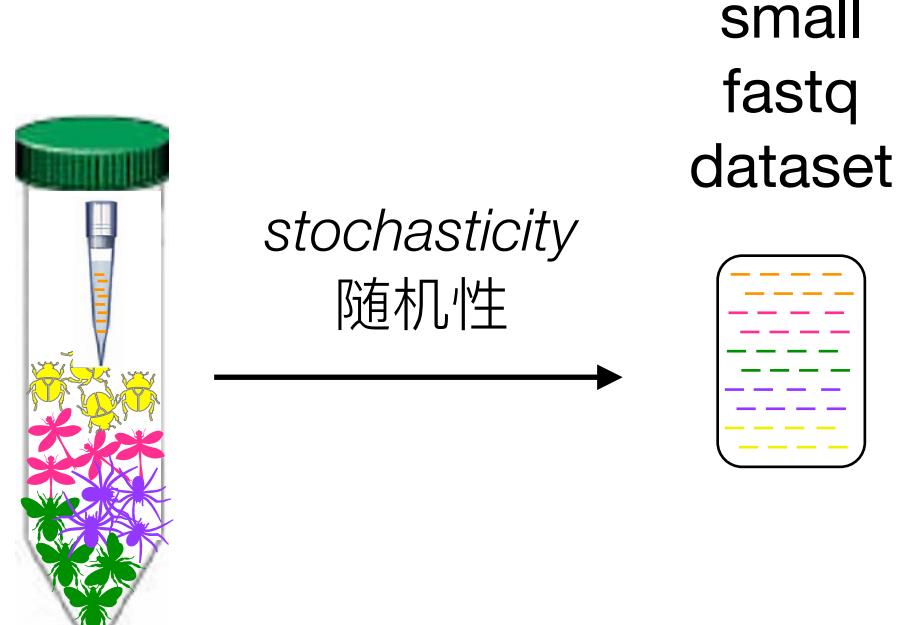


Sample 1	Reads_Raw
Taxon1	1000
Taxon2	1200
Taxon3	1600
Taxon4	2000
Taxon5	400
Taxon6	200
	total = 6400
Sample 2	Reads_Raw
Taxon1	500
Taxon2	600
Taxon3	800
Taxon4	1000
Taxon5	200
Taxon6	100
	total = 3200

Spike-in
Soup 1
Low
biomass

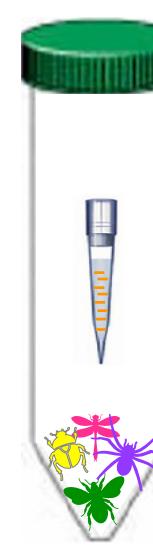


Spike-in
Soup 2
High
biomass



Sample 1	Reads_Raw
OTU1	1000
OTU2	1200
OTU3	1600
OTU4	2000
OTU5	400
OTU6	200
OTU_SPIKE	40
Sample 2	Reads_Raw
OTU1	500
OTU2	600
OTU3	800
OTU4	1000
OTU5	200
OTU6	100
OTU_SPIKE	10

Spike-in
Soup 1
Low
biomass



stochasticity
随机性

big
fastq
dataset

total = 6400

Spike-in
Soup 2
High
biomass

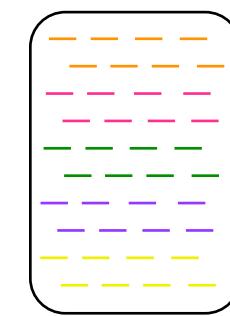


stochasticity
随机性

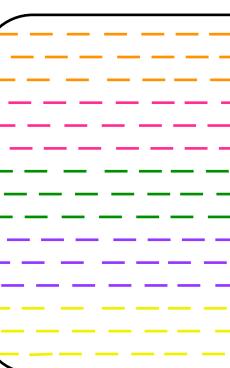
small
fastq
dataset

total = 3200

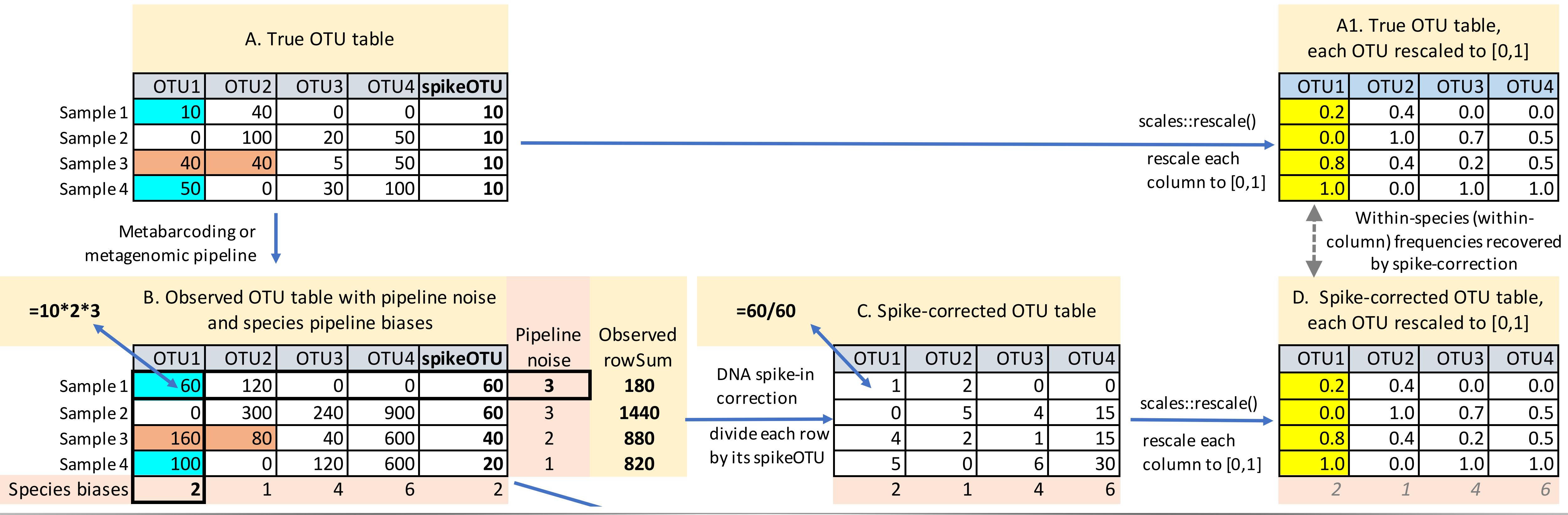
Sample 1	Reads_Raw	Reads_Corr	
OTU1	1000	25	$1000/40 = 25$
OTU2	1200	30	$1200/40 = 30$
OTU3	1600	40	$1600/40 = 40$
OTU4	2000	50	$2000/40 = 50$
OTU5	400	10	$400/40 = 10$
OTU6	200	5	$200/40 = 5$
OTU_SPIKE	40		
Sample 2	Reads_Raw	Reads_Corr	
OTU1	500	50	$500/10 = 50$
OTU2	600	60	$600/10 = 60$
OTU3	800	80	$800/10 = 80$
OTU4	1000	100	$1000/10 = 100$
OTU5	200	20	$200/10 = 20$
OTU6	100	10	$100/10 = 10$
OTU_SPIKE	10		



total = 160



total = 320



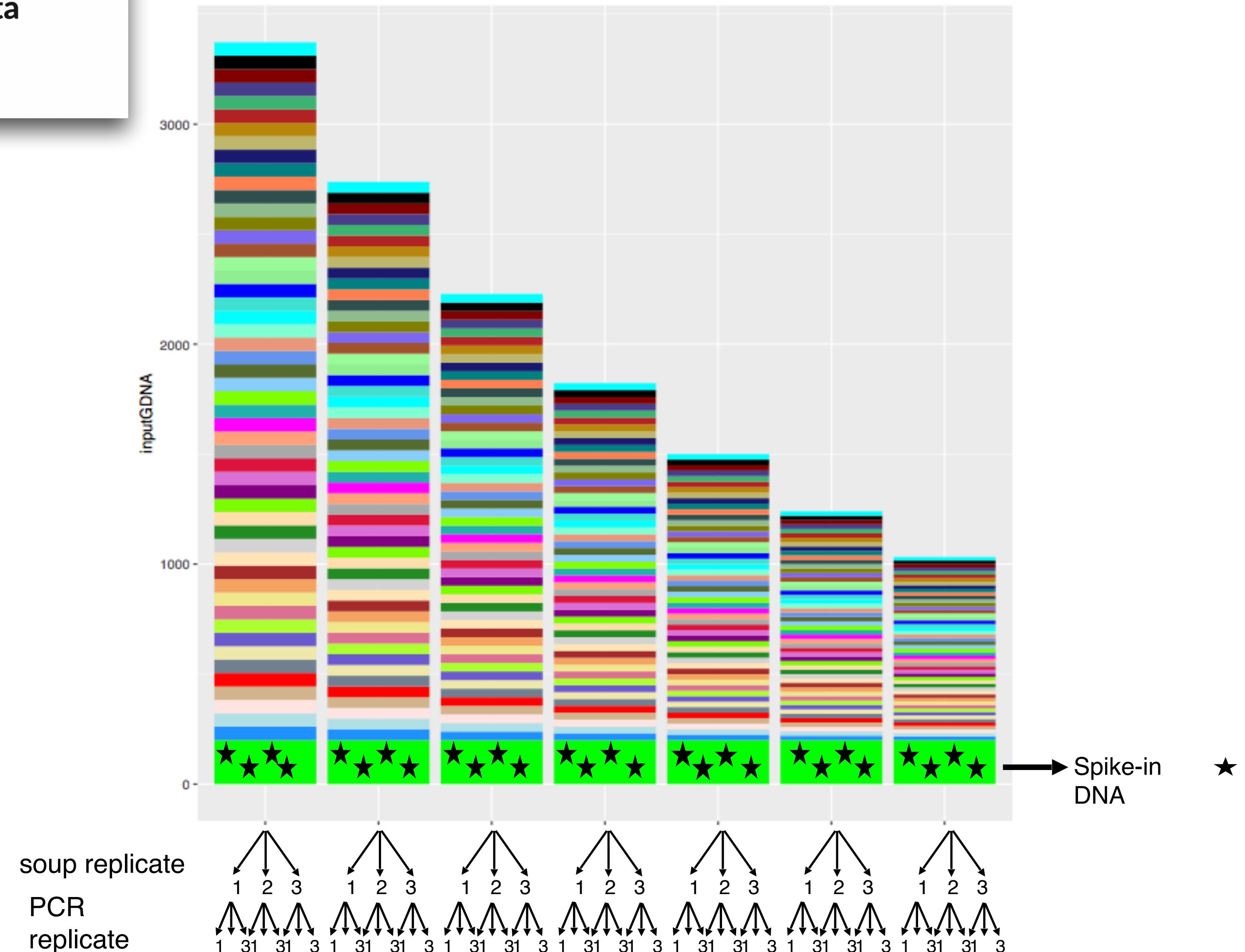
Excel spreadsheet showing how to do the calc

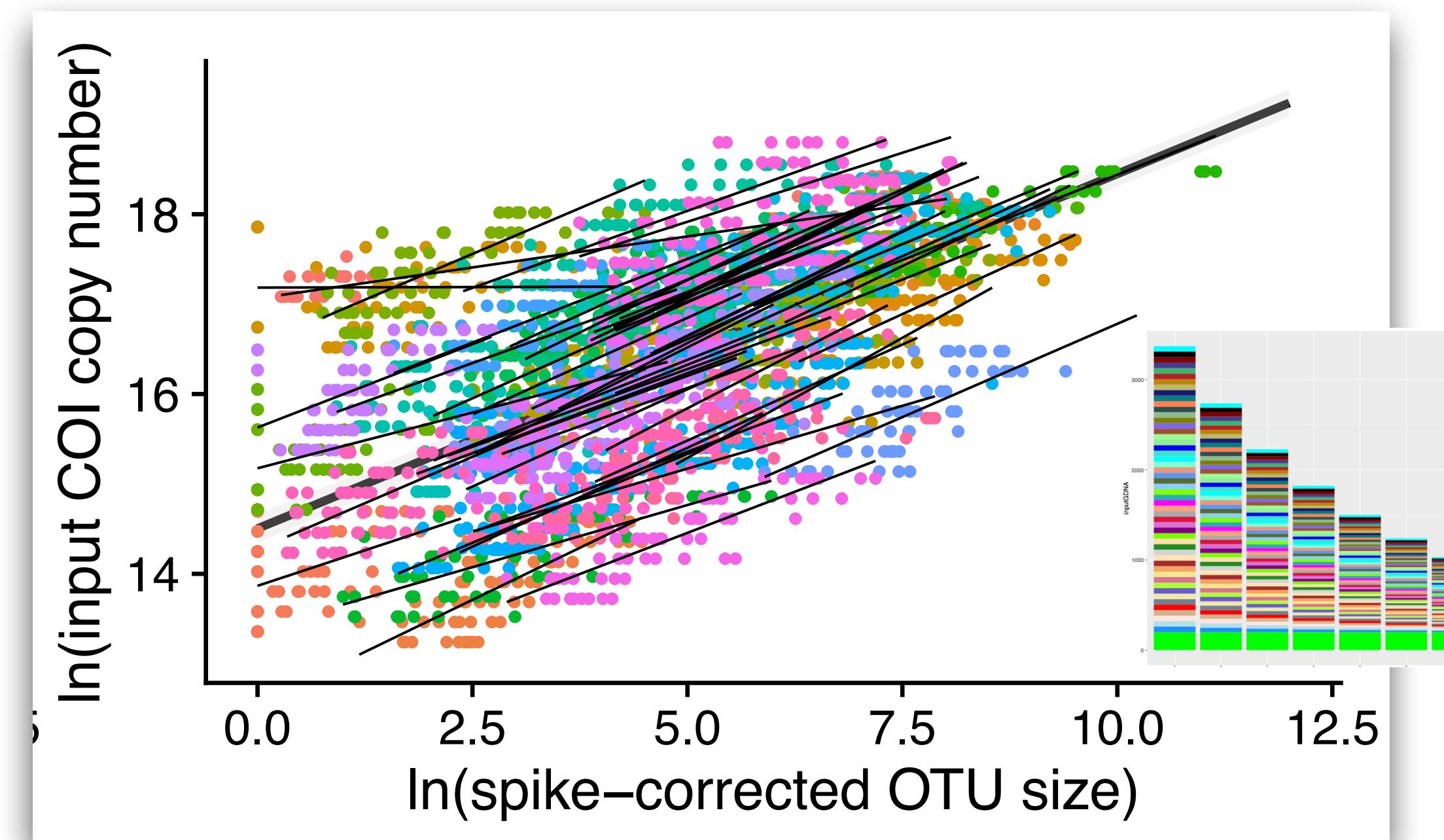
qSeq Experiment

Extracting abundance information from DNA-based data

Mingjie Luo^{1,2} | Yinqiu Ji¹ | David Warton^{3,4} | Douglas W. Yu^{1,5,6} 

- 7 mock soups, 52 insect OTUs.
- Soups serially diluted by 0.8X
- Equal genomic DNA concentration per OTU (PicoGreen)
- 3 soup replicates, 3 PCR replicates.
- Spike-in DNA (3 DNA barcodes in plasmids)
- qPCR to quantify COI copy number per OTU (SYBR green)





It works! For each species, the OTU size predicts the input DNA (= within-species abundance information)

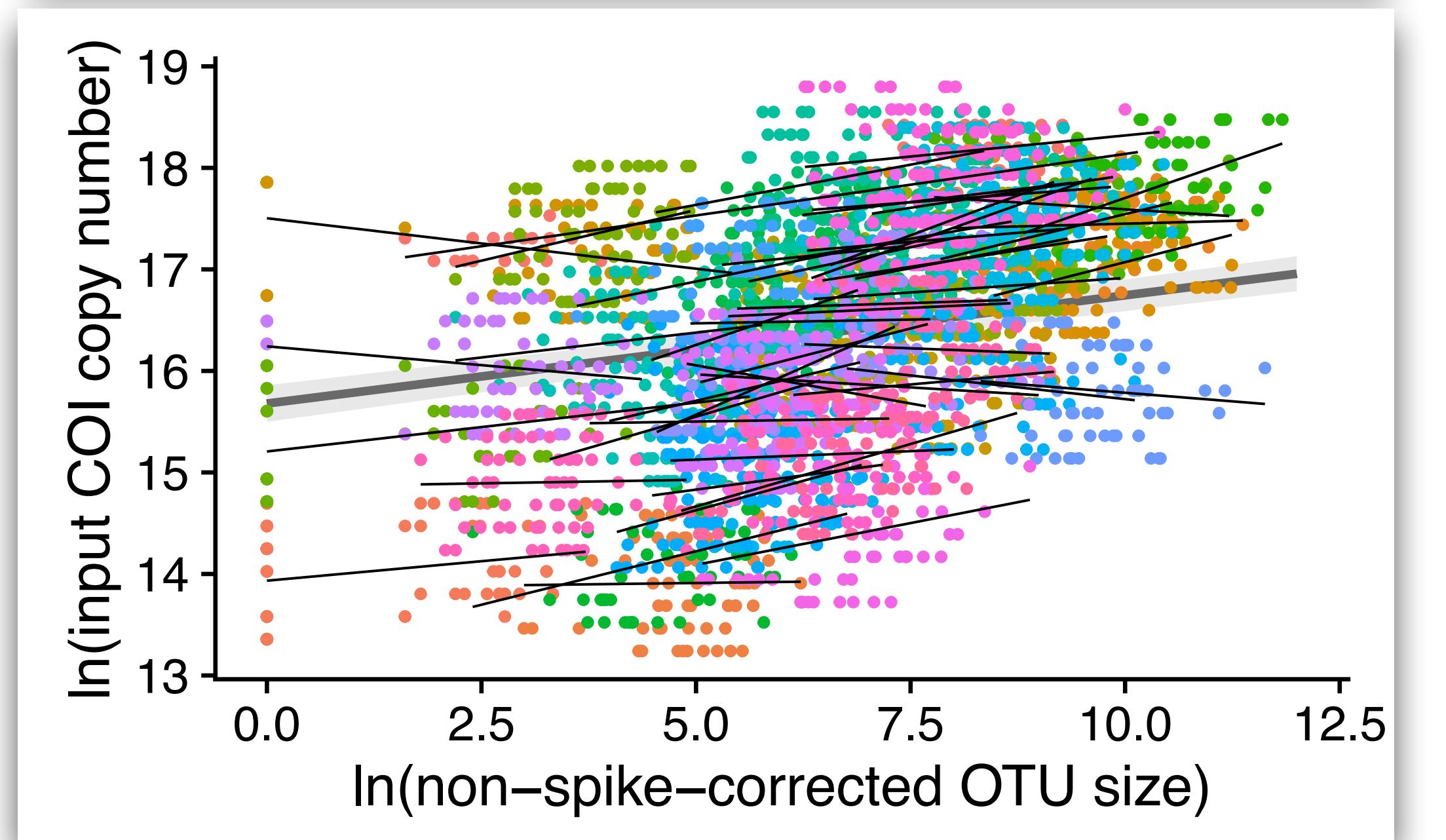
```
lme4::lmer(log.inputCoiCopy ~ log.readSpikeCorr +  
(log.readSpikeCorr | OTUID) + (1 |soupRep/pcrRep))
```

marginal $R^2=0.40$ (soup and PCR replicates averaged)

We can estimate **within-species** change in abundance (COI copy number)

But we cannot estimate relative species abundances

(the intercepts are all different, but they should be the same because all species used the same amounts of DNA)



Without spike-in: huge amounts of pipeline noise, the OTU size does not predict the input DNA (= within-species abundance information)

```
lme4::lmer(log.inputCoiCopy ~ log.read +  
(log.read | OTUID) + (1 |soupRep/pcrRep))
```

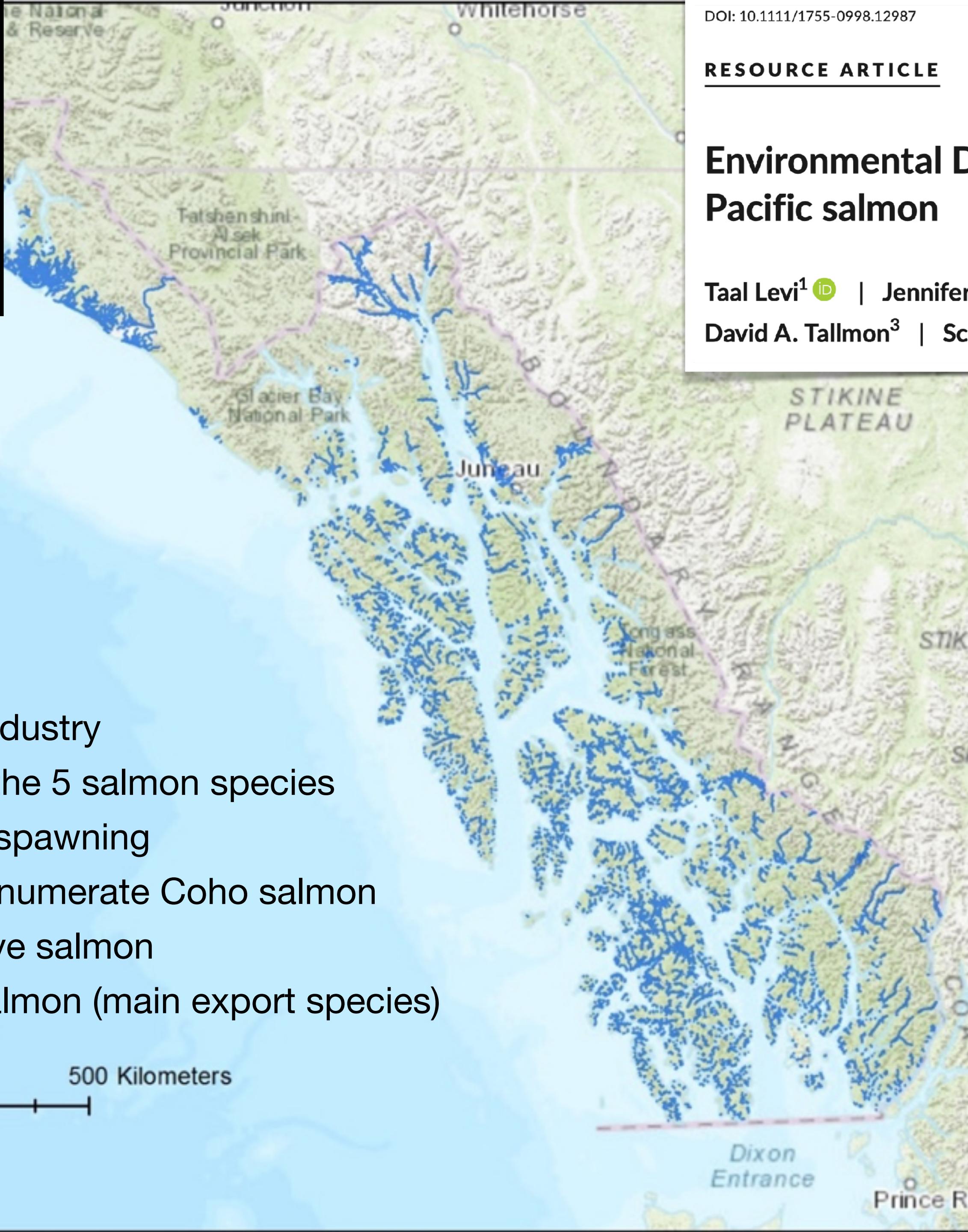
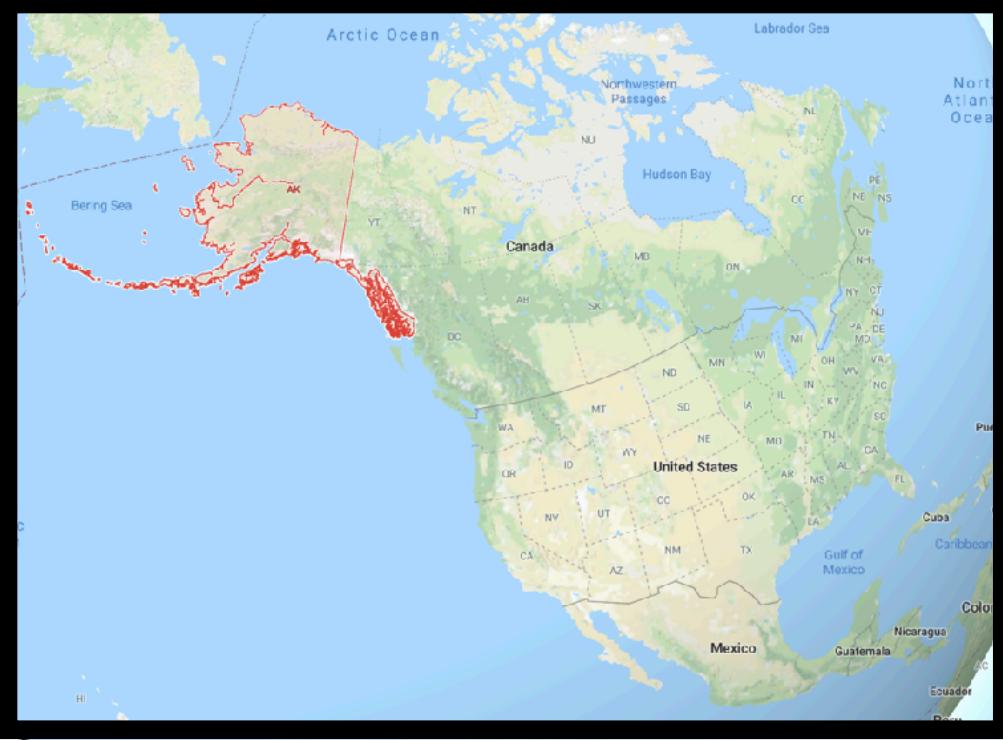
marginal $R^2=0.04$ (soup and PCR replicates averaged)

Take home

Species

Methods to extract abundance information from DNA data

- **Single-species quantitative PCR (qPCR): within-species quantification**
- Multiplexed specimen barcoding (mBRAVE): within- and across-species quantification
- Mitogenomics and DNA spike-in (SPIKEPIPE): within-species quantification
- Metabarcoding and DNA spike-in (qSeq): within-species quantification



>>\$500 million / year industry
6000 streams used by the 5 salmon species
1000 streams used for spawning
4 - 9 streams used to enumerate Coho salmon
~14 streams for Sockeye salmon
0 for Pink and Chum salmon (main export species)

0 125 250 500 Kilometers

DOI: 10.1111/1755-0998.12987

RESOURCE ARTICLE

2019

WILEY MOLECULAR ECOLOGY RESOURCES

Environmental DNA for the enumeration and management of Pacific salmon

Taal Levi¹ | Jennifer M. Allen¹ | Donovan Bell² | John Joyce² | Joshua R. Russell² | David A. Tallmon³ | Scott C. Vulstek² | Chunyan Yang⁴ | Douglas W. Yu^{4,5,6}

Alaskan salmon fishery



The ‘escapement’
is the breeding
population



Is the fishery allowing enough salmon to escape?

 Fish & Game State of Alaska[Home](#) [Fishing](#) [Hunting](#) [Subsistence](#) [Viewing](#) [Education](#) [Species](#) [Lands & Waters](#) [Regulations](#)[Licenses & Permits](#) [Commercial](#) [Sport](#) [Subsistence](#) [Personal Use](#) [Aquatic Farming](#) [Hatcheries](#) [Research](#)

Commercial Fishing

[Commercial Fishing Home](#)[News Releases & Announcements](#)[Information By Area](#)

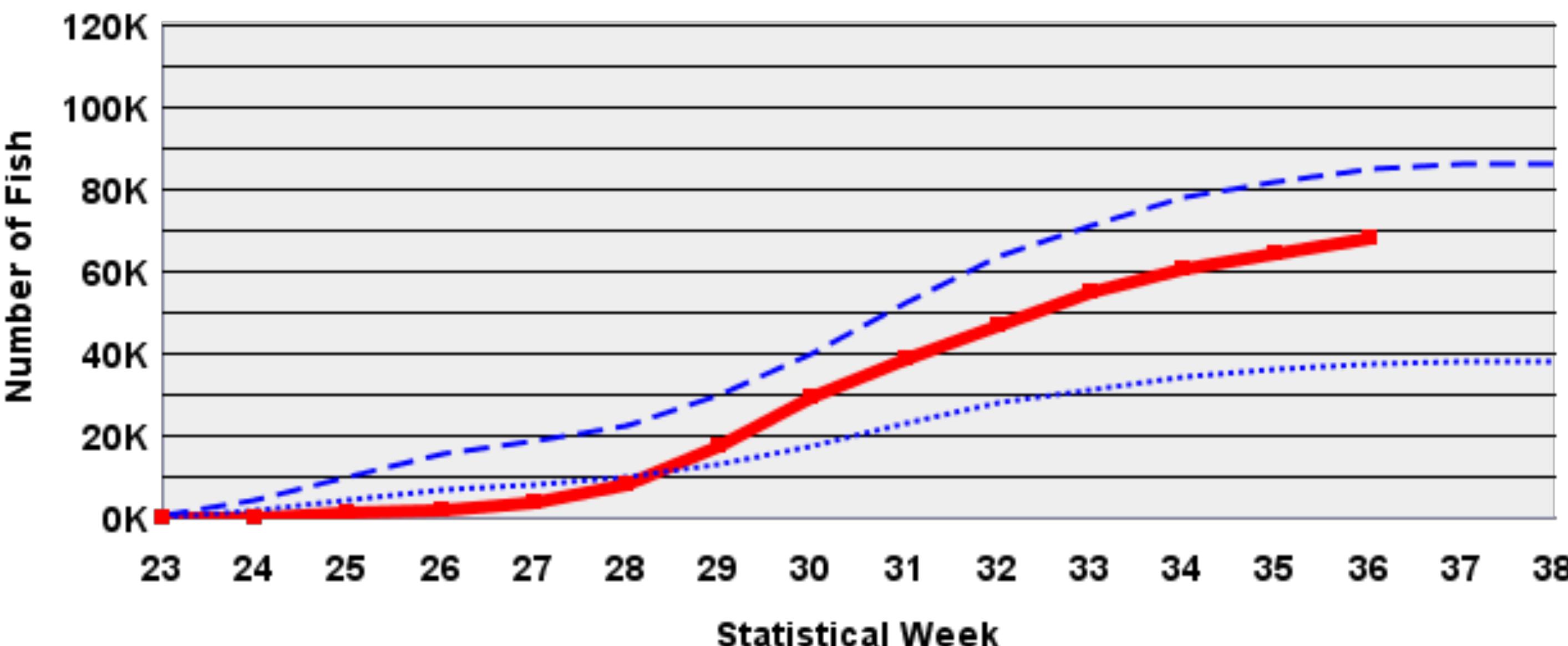
- Arctic-Yukon-Kuskokwim
 - Kuskokwim
 - Arctic Area
 - Norton Sound & Kotzebue
 - Yukon
- Central Region
 - Bristol Bay
 - Copper River
 - Cook Inlet
 - Lower Cook Inlet
 - Upper Cook Inlet
 - Prince William Sound

[Southeast Region](#)

- Westward Region
 - Alaska Peninsula
 - Bering Sea & Aleutian Islands
 - Chignik
 - Kodiak Island

[ADF&G Home](#) » [Fishing](#) » [Commercial](#) » [Information By Area](#) » [Southeast](#)

Commercial Salmon Fisheries Chilkoot Lake Weir - Sockeye Counts

Cumulative Counts Compared to Escapement Goals**2015****...Lower Goal — Upper Goal**

ALASKA DEPARTMENT OF FISH AND GAME

DIVISION OF COMMERCIAL FISHERIES

NEWS RELEASE



*Sam Cotten, Commissioner
Jeff Regnart, Director*



Contact:

Gordie Woods

Phone: (907) 784-3255

Fax: (907) 784-3254

Yakutat Area Office

P.O. Box 49

Yakutat, AK, 99689

Date: September 4, 2015

Time: 8:30 a.m.

YAKUTAT COMMERCIAL SET GILLNET OPENING ANNOUNCEMENT

Dangerous River: will be open from 12:01 p.m., Sunday, September 6 through 12:00 noon, Wednesday, September 9.

Akwe River: will be open from 12:01 p.m., Sunday, September 6 through 12:00 noon, Wednesday, September 9.

Lost River: will remain closed until further notice.

Situk-Ahrnklin Inlet: will be open from 12:01 p.m., Sunday, September 6 through 12:00 noon, Wednesday, September 9 with the following restrictions:

North Bank of Situk-Ahrnklin Inlet: commercial set gillnet fishing will be prohibited along the north bank of the Situk-Ahrnklin Inlet between two ADF&G regulation markers located 500 yards above and 500 yards below the confluence of the Situk-Ahrnklin Inlet and the Lost River;



Validation data from Auke Creek Station in Juneau, Alaska

Daily counts of

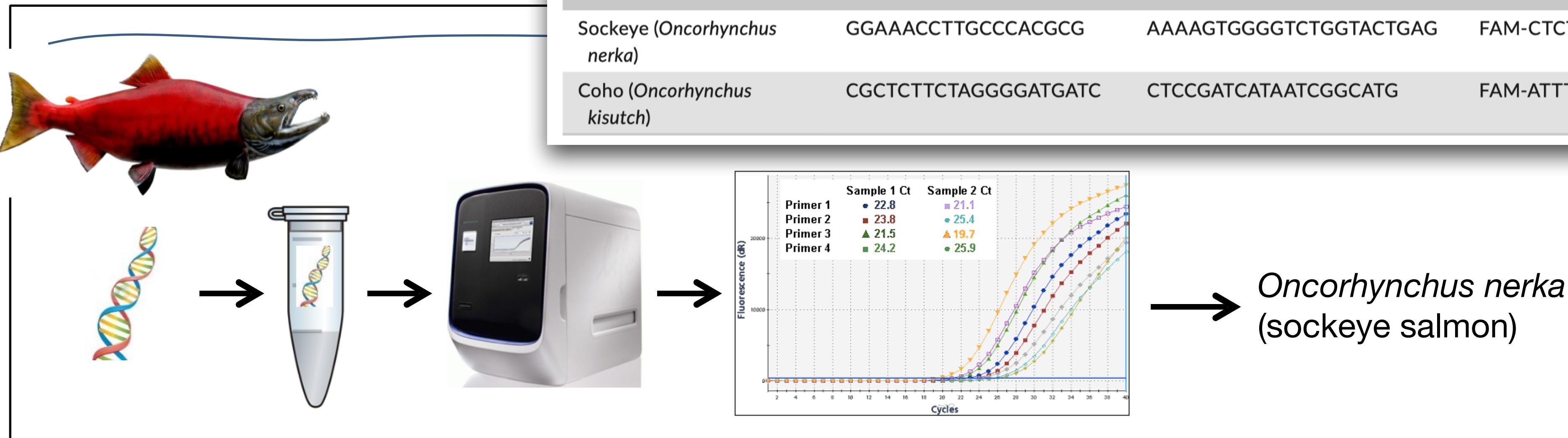
- in-migrating Coho and Sockeye **adults**
- out-migrating Sockeye **smolts (juveniles)**



Does environmental DNA (eDNA) contain enough information to estimate salmon escapement sizes?

TABLE 1 Species-specific primers and probes used in this study (Rasmussen Hellberg et al., 2010)

Target species	Forward Primer (5'-3')	Reverse Primer (5'-3')	Probe (5'-3')
Sockeye (<i>Oncorhynchus nerka</i>)	GGAAACCTGCCACGCG	AAAAGTGGGTCTGGTACTGAG	FAM-CTCTGTTGACTAACCATC-MGB
Coho (<i>Oncorhynchus kisutch</i>)	CGCTCTCTAGGGGATGATC	CTCCGATCATAATCGGCATG	FAM-ATTTACAACGTAATCGTC-MGB



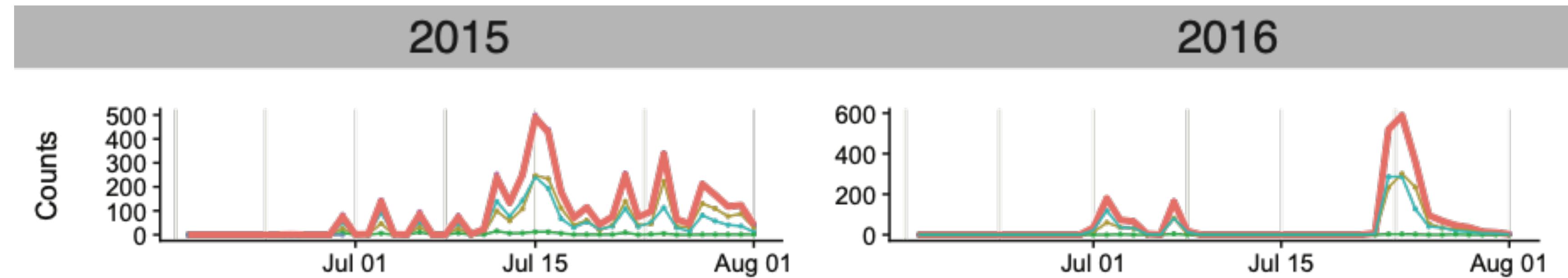
=



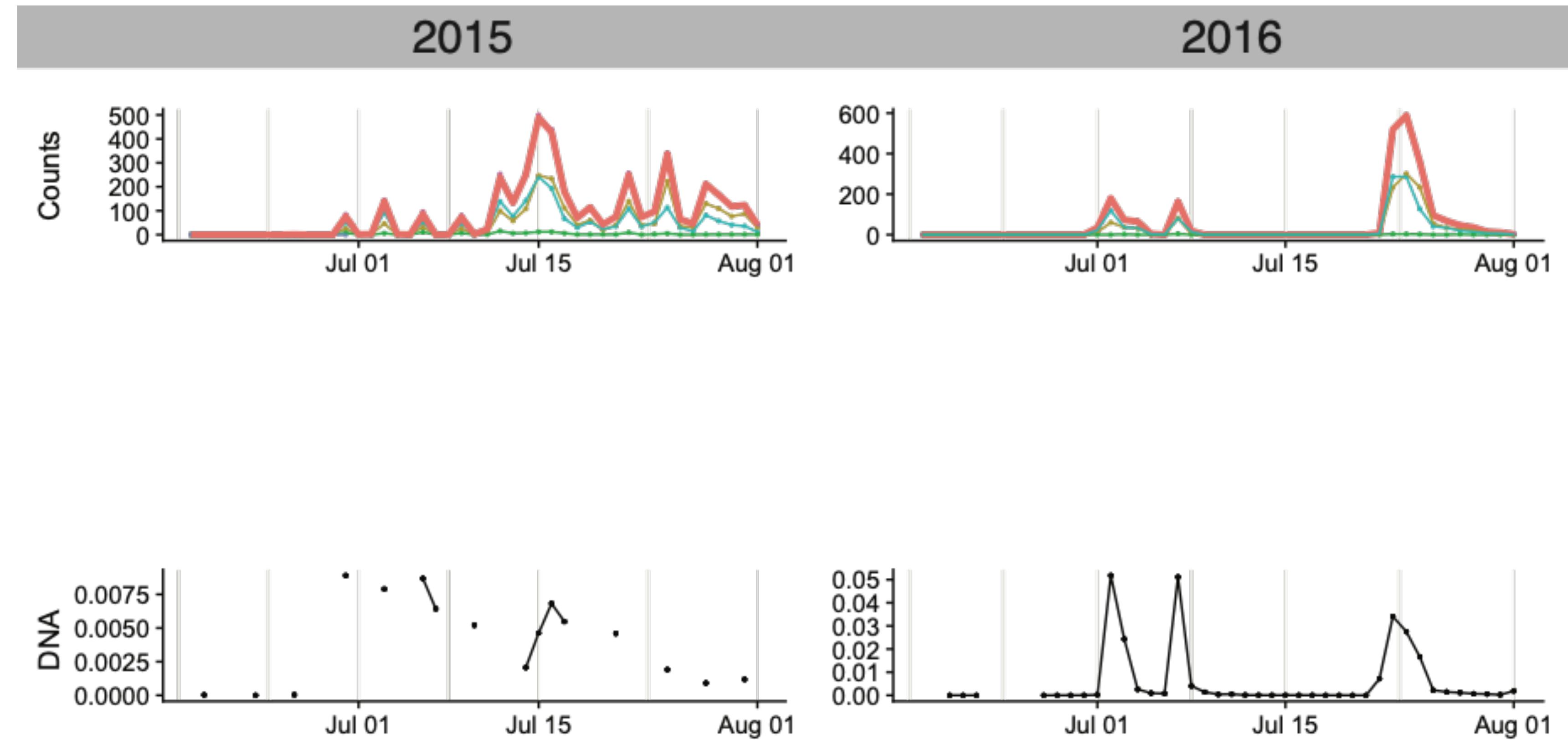
One person, one sample /day

Many people, all day

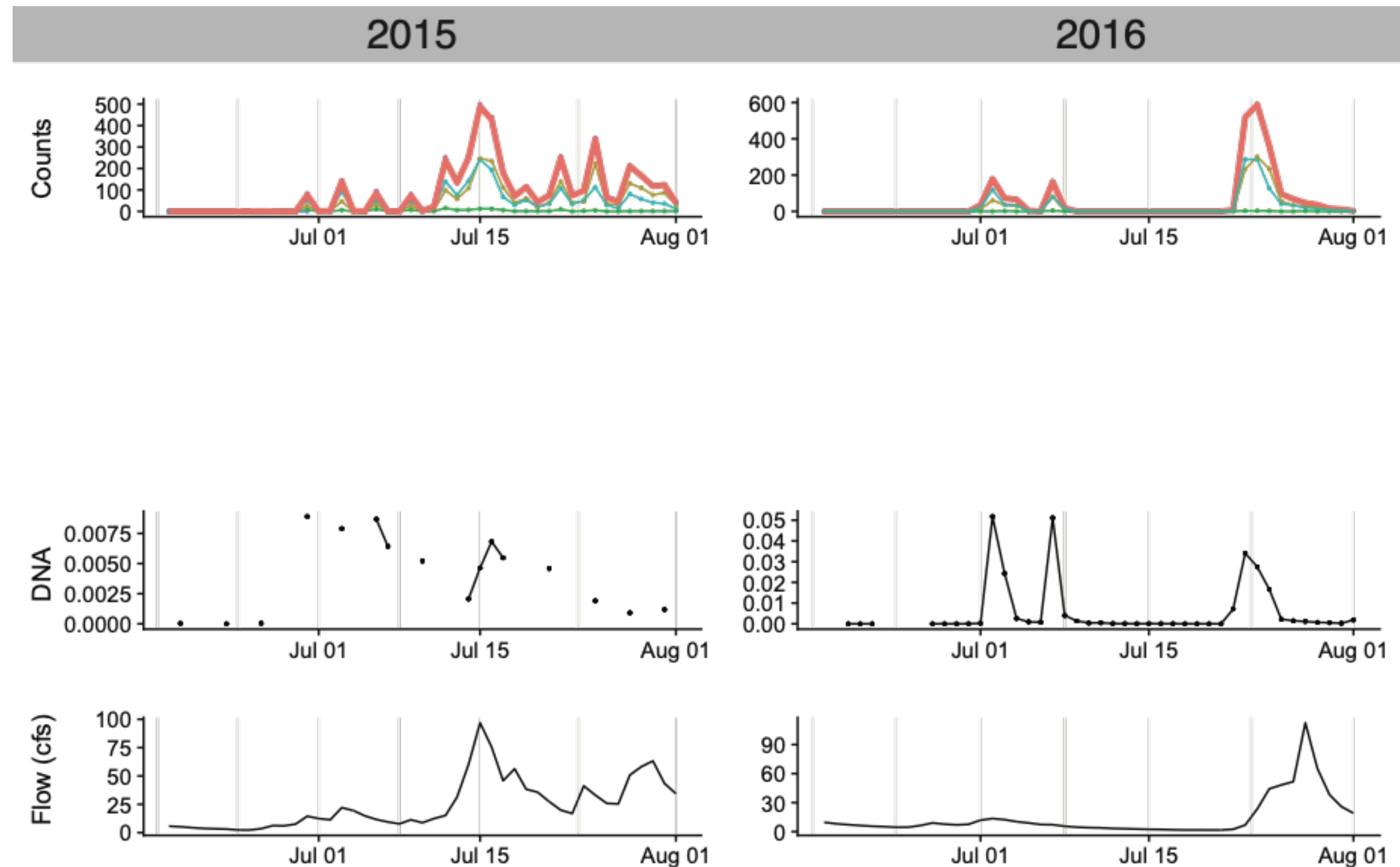
Sockeye in-migrating adults



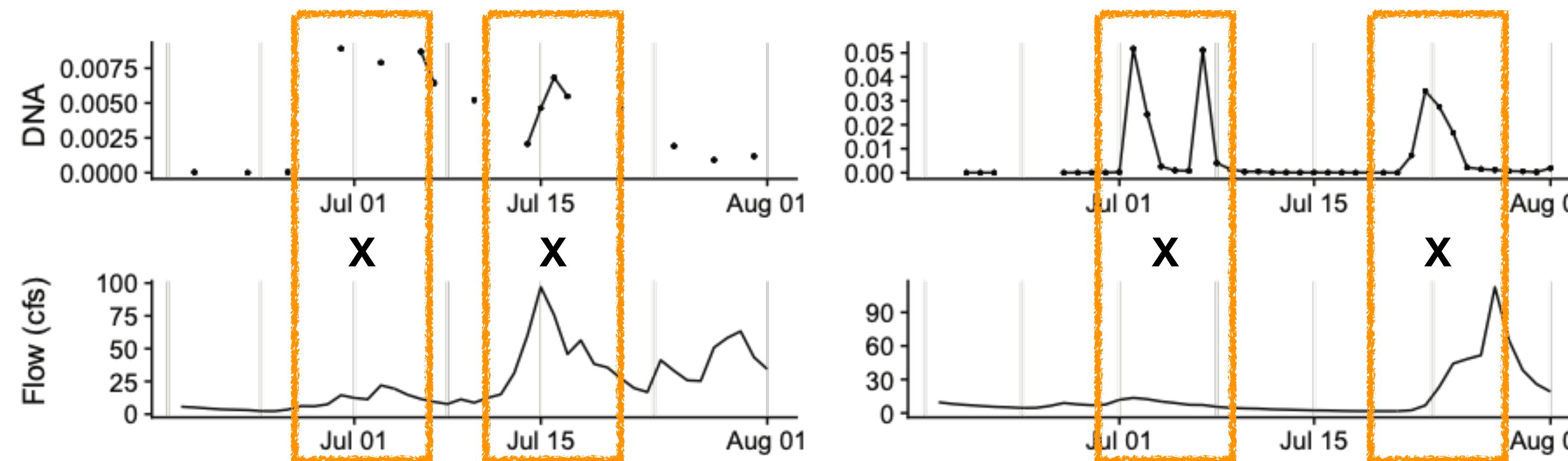
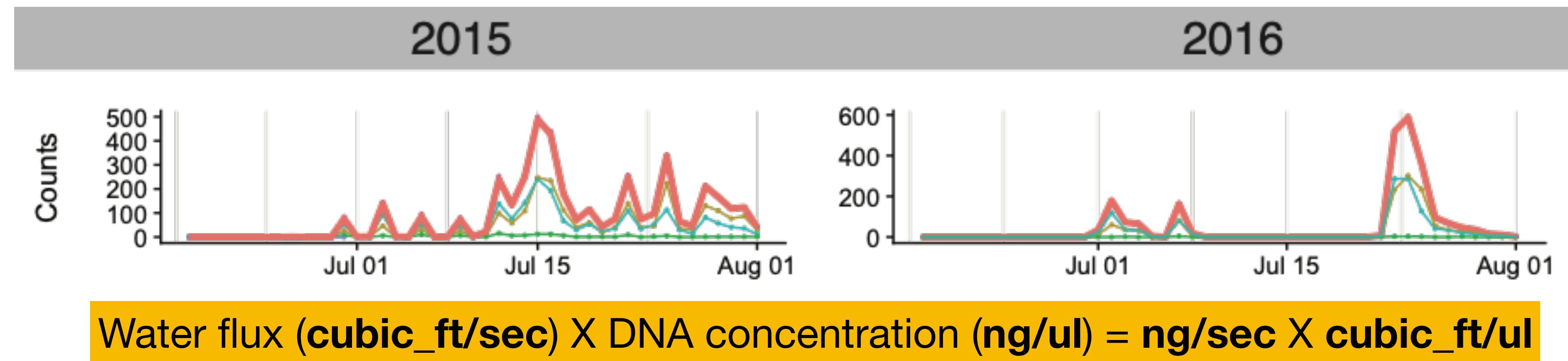
Sockeye in-migrating adults



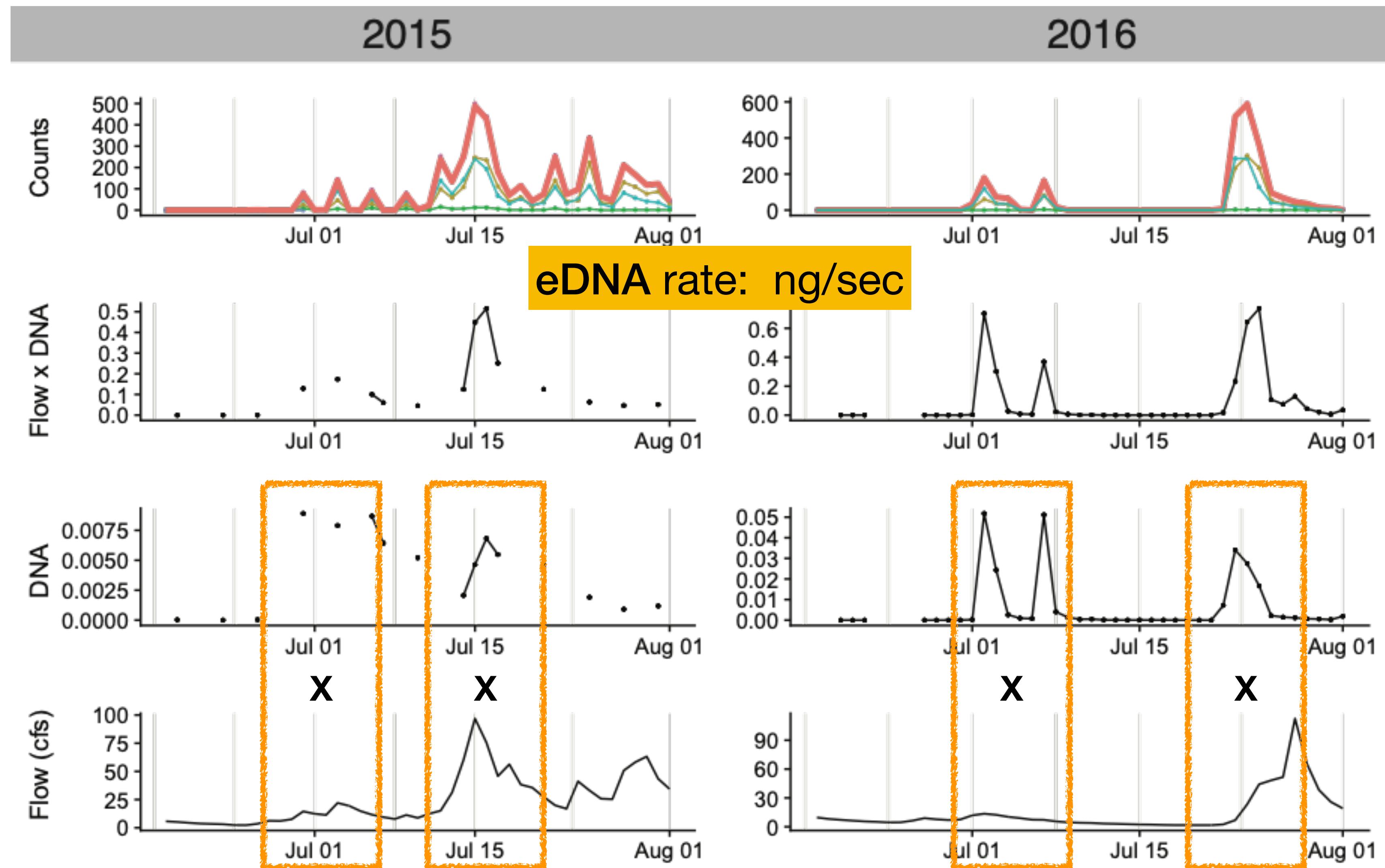
Sockeye in-migrating adults



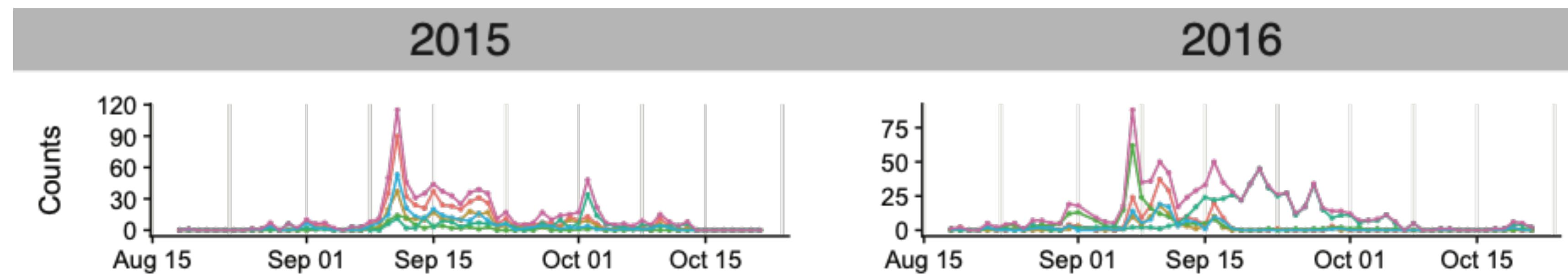
Sockeye in-migrating adults



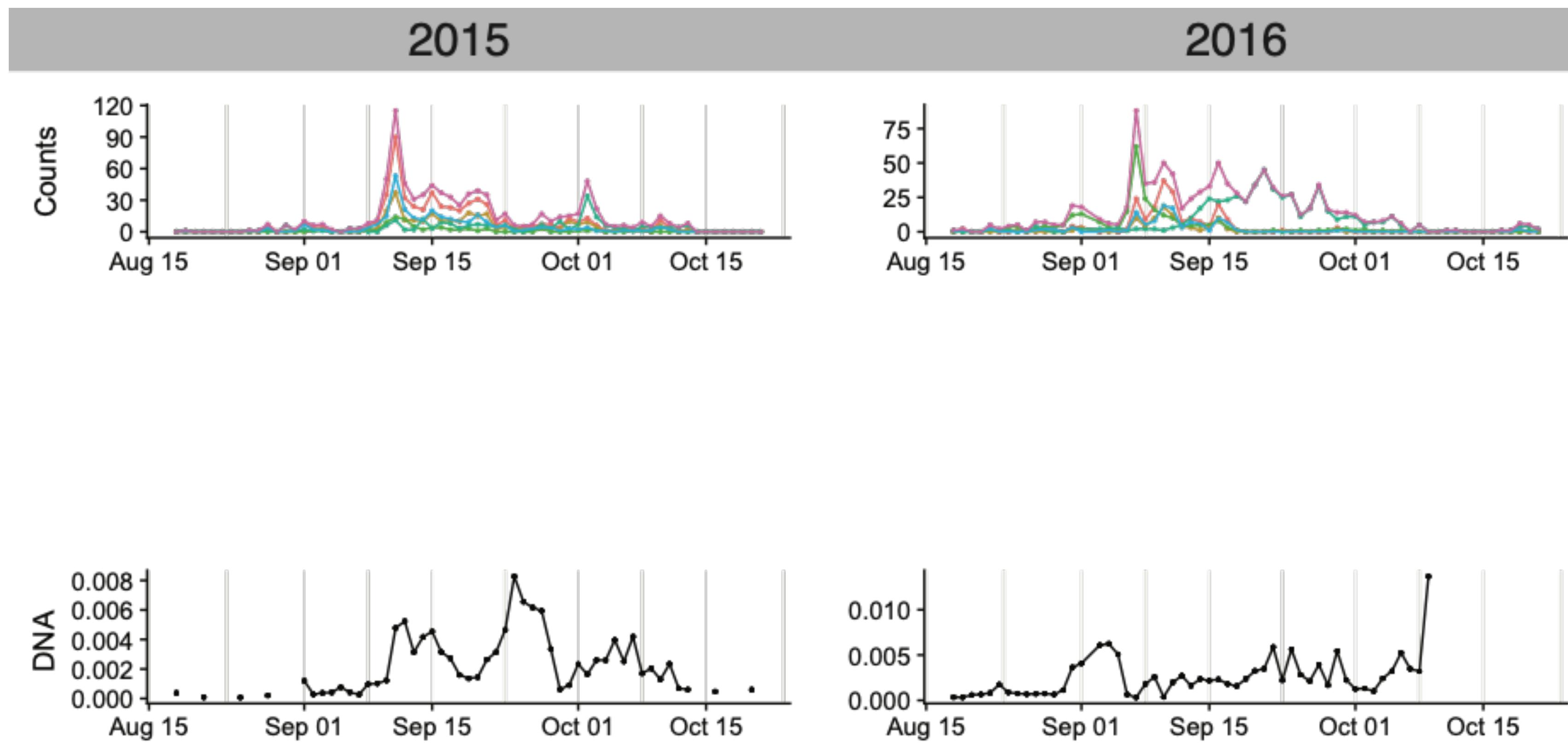
Sockeye in-migrating adults



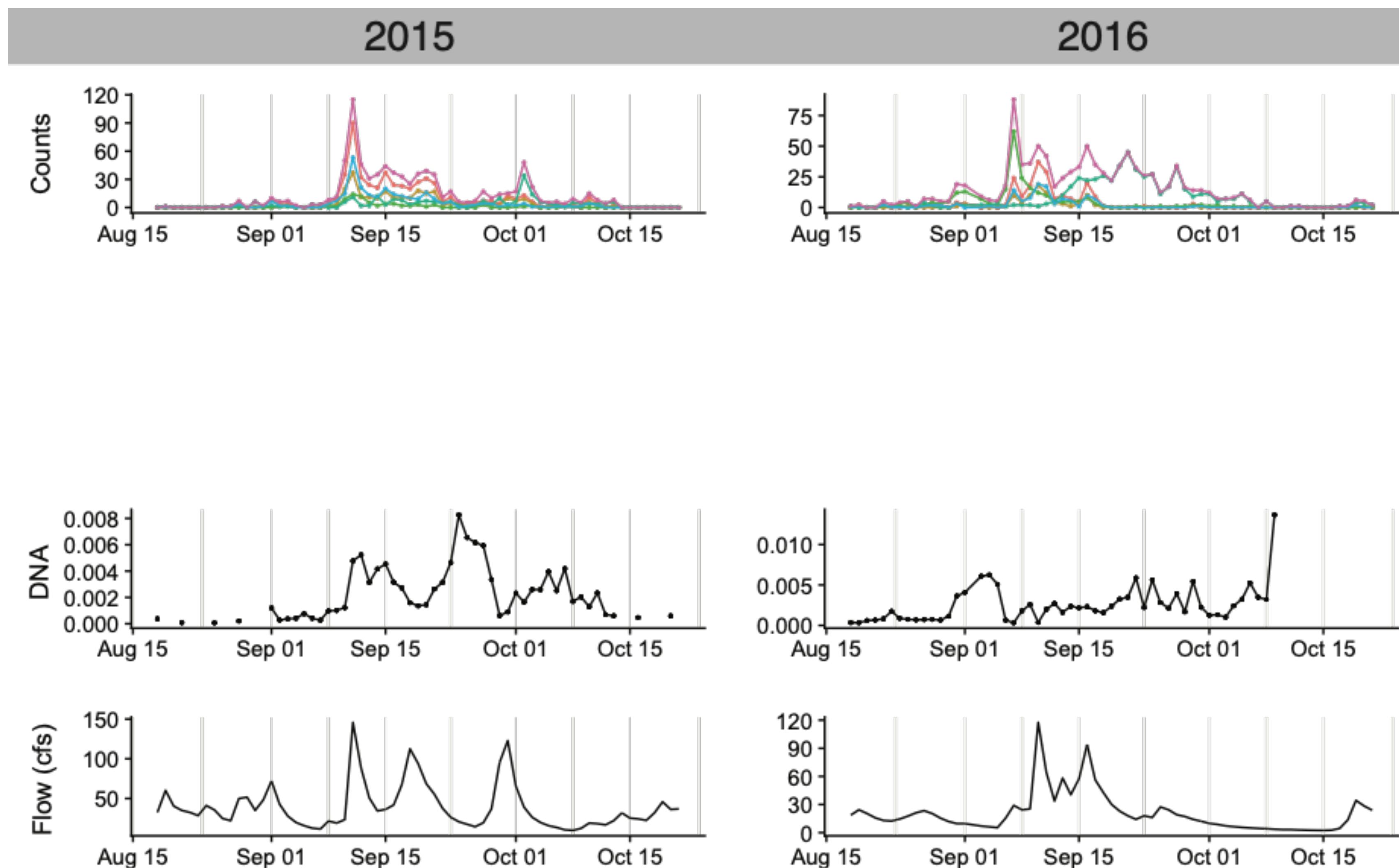
Coho in-migrating adults



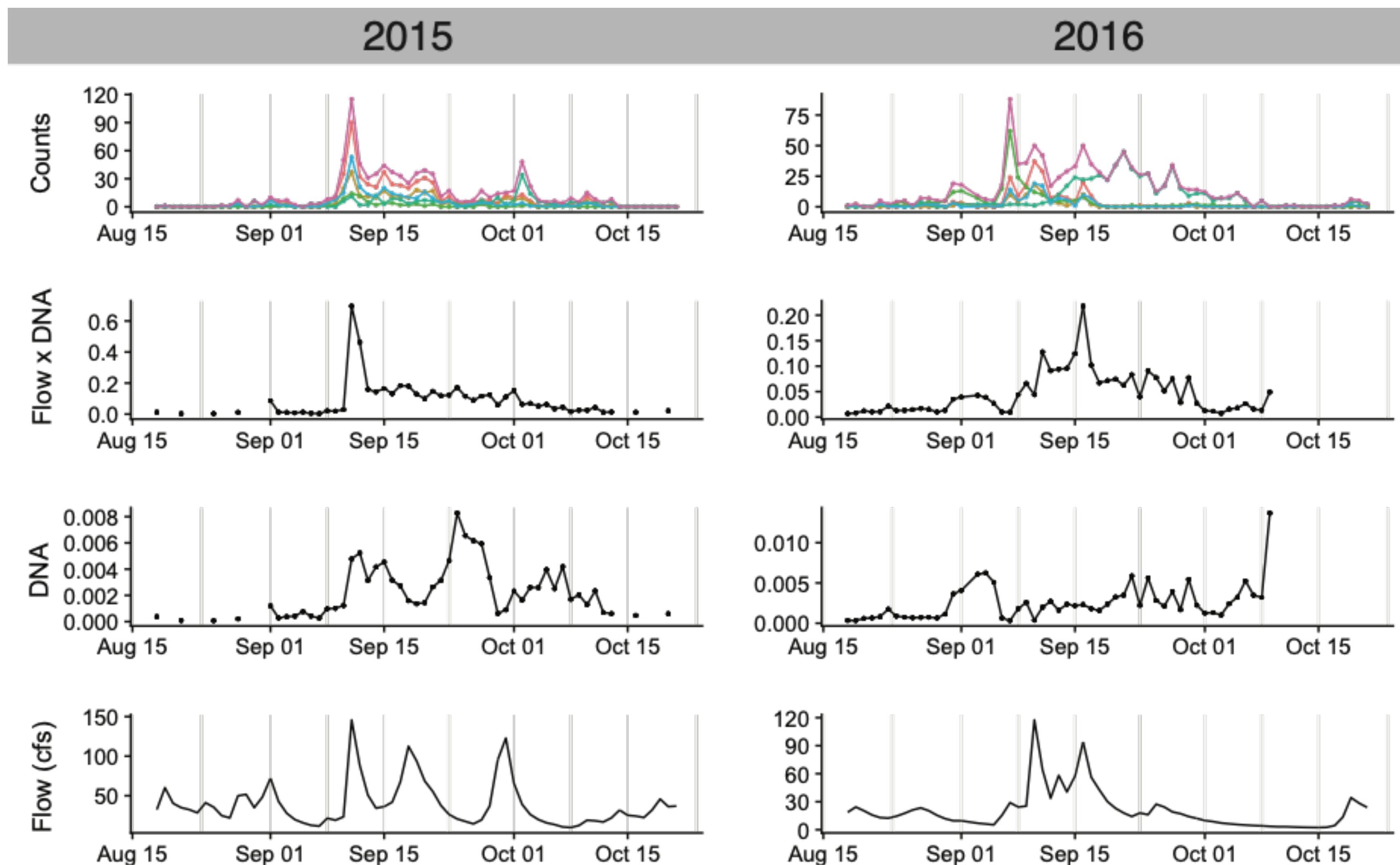
Coho in-migrating adults



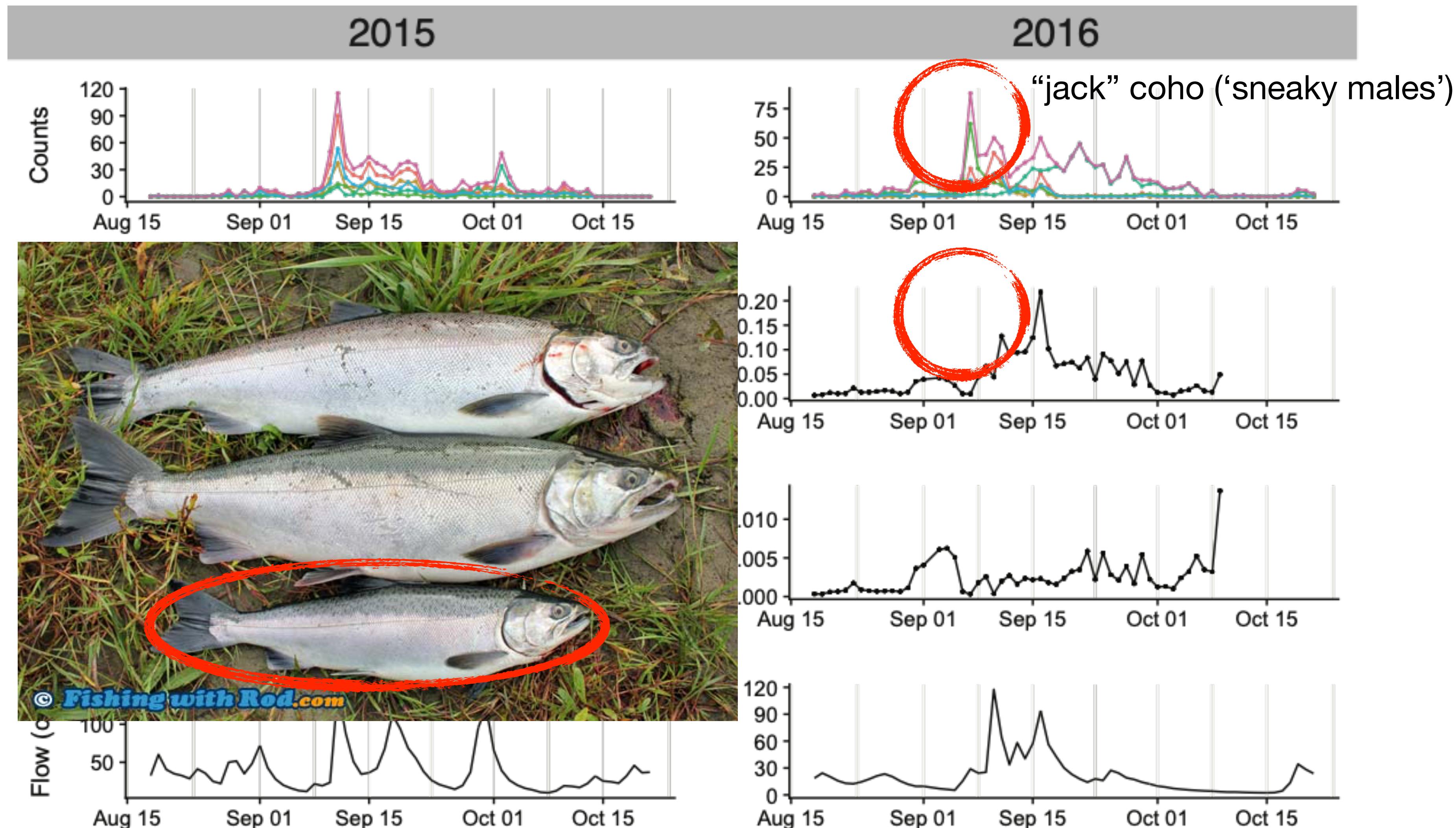
Coho in-migrating adults



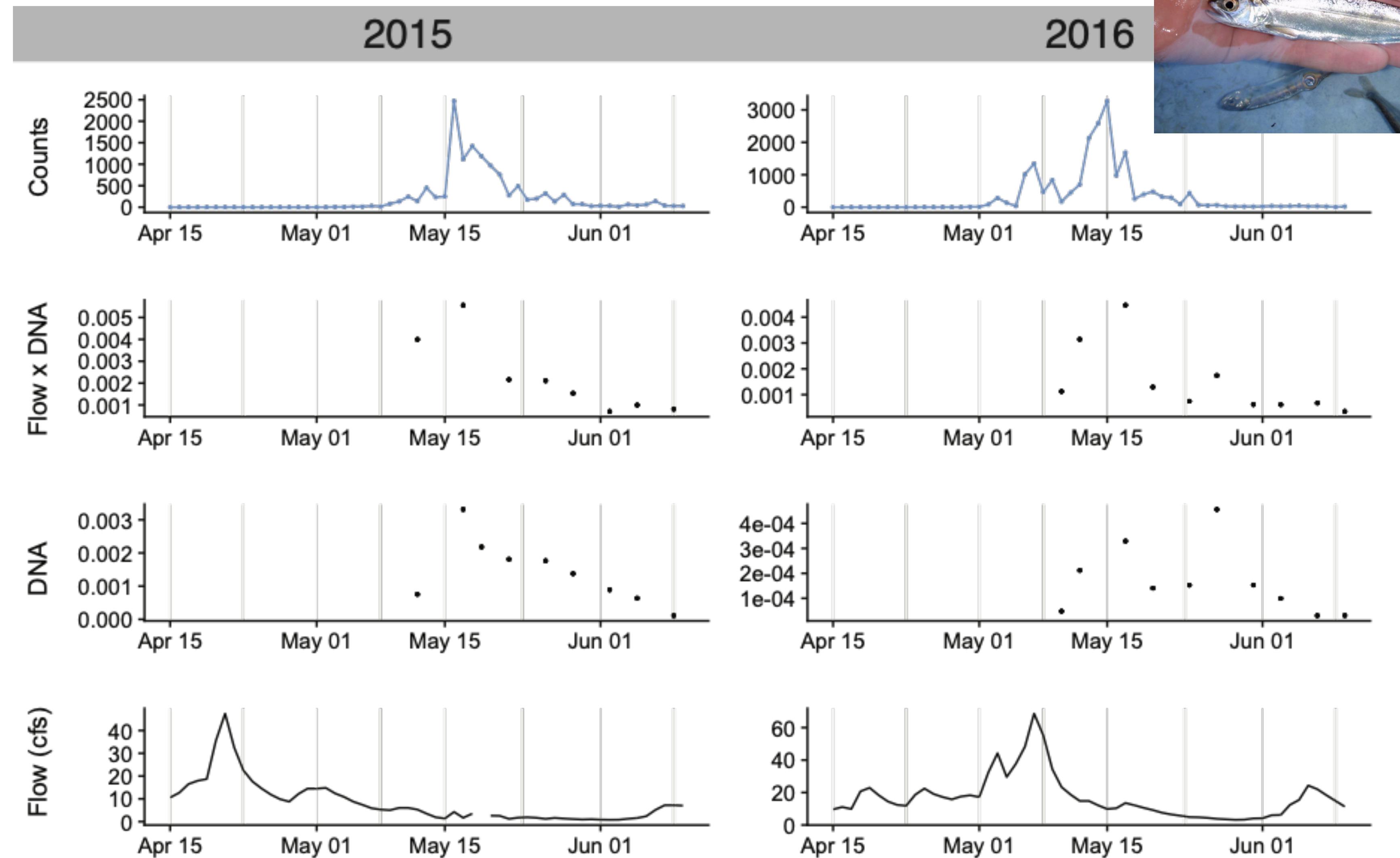
Coho in-migrating adults



Coho in-migrating adults



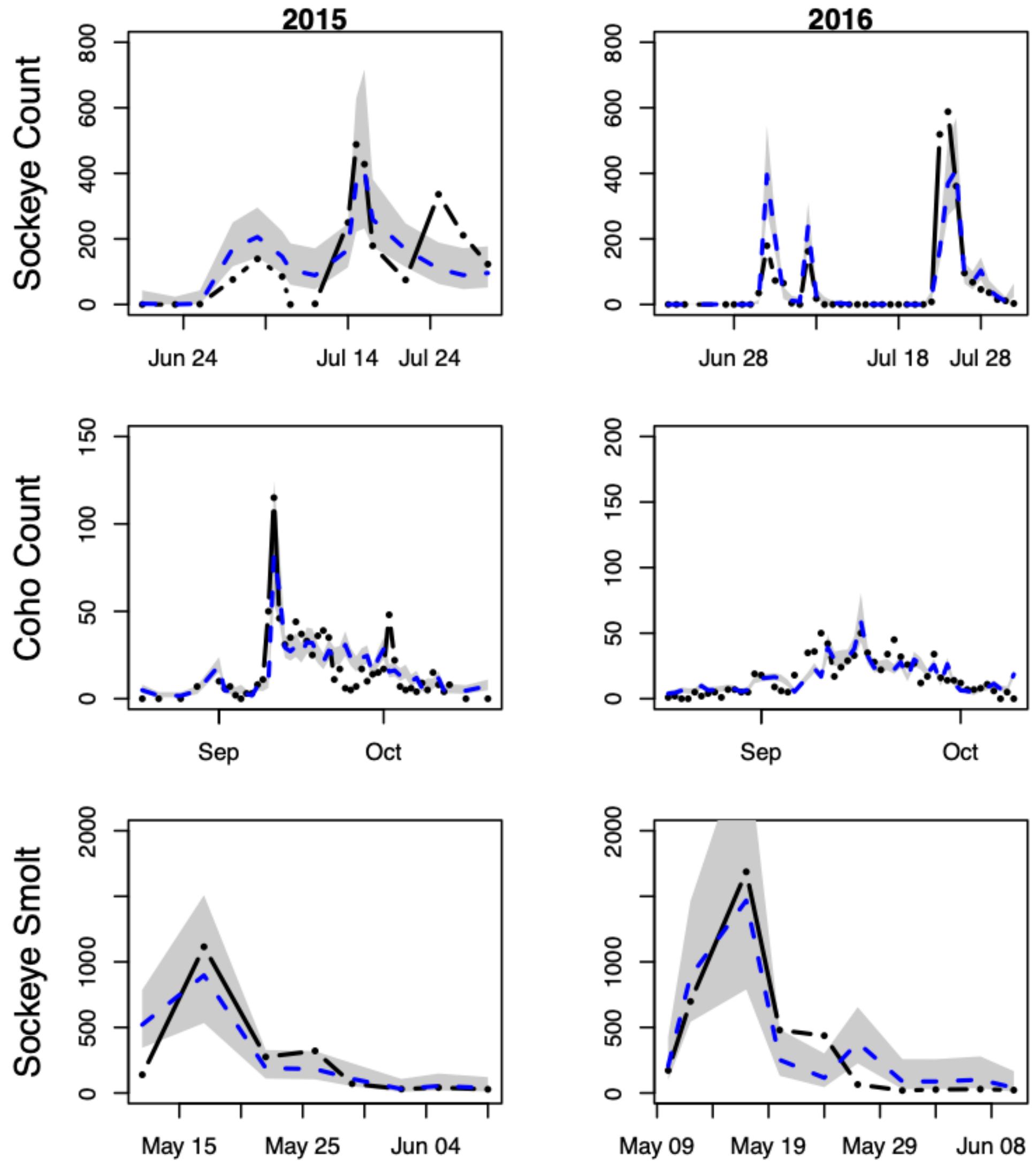
Sockeye out-migrating smolts (juveniles)





Fitted models (blue) vs. Data (black)

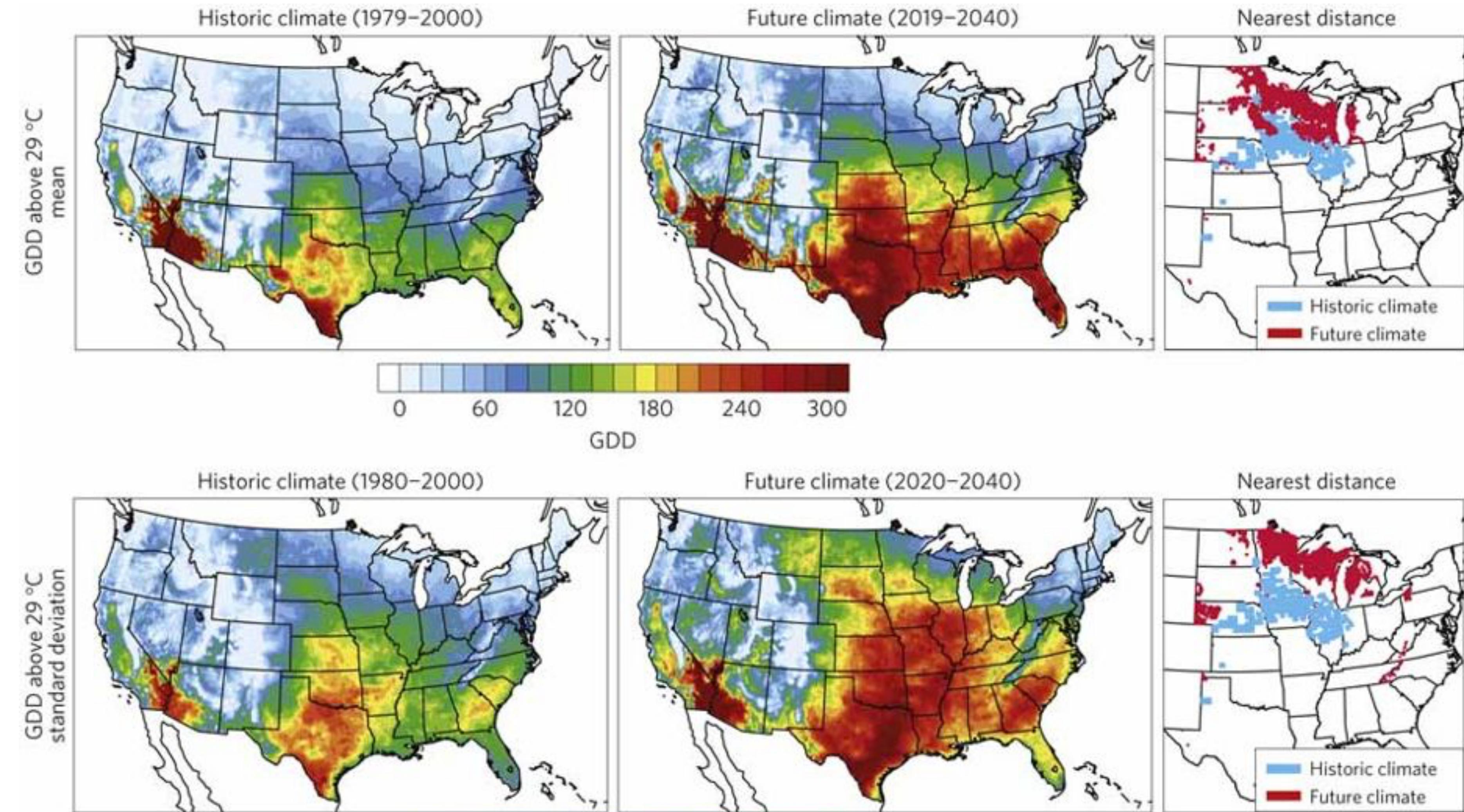
- `glm(log(count) ~ log(eDNA_rate), quasipoisson)`
- All highly significant
- Models not significantly different **between years**
- Small signal of salmon one day ago,
No effect of salmon two days ago



Methods to extract abundance information from DNA data

- Single-species quantitative PCR (qPCR): *control for eDNA transport dynamics*
- **Multiplexed *individual* barcoding (mBRAVE)**
- Mitogenomics and DNA spike-in (SPIKEPIPE)
- Metabarcoding and DNA spike-in (qSeq)

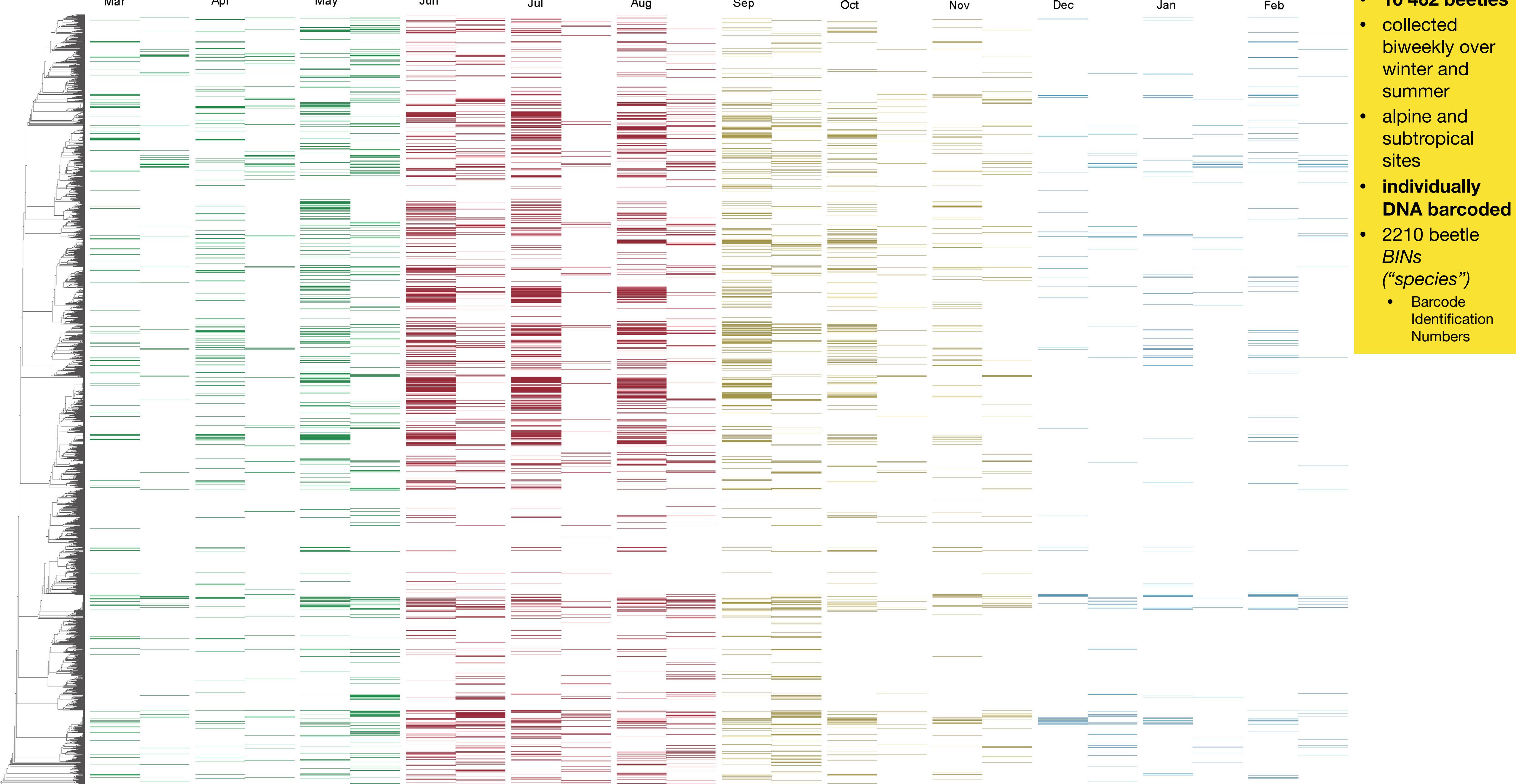
How much biodiversity will we lose because of climate change?



Changes in distribution due to changes in climate envelopes



South North

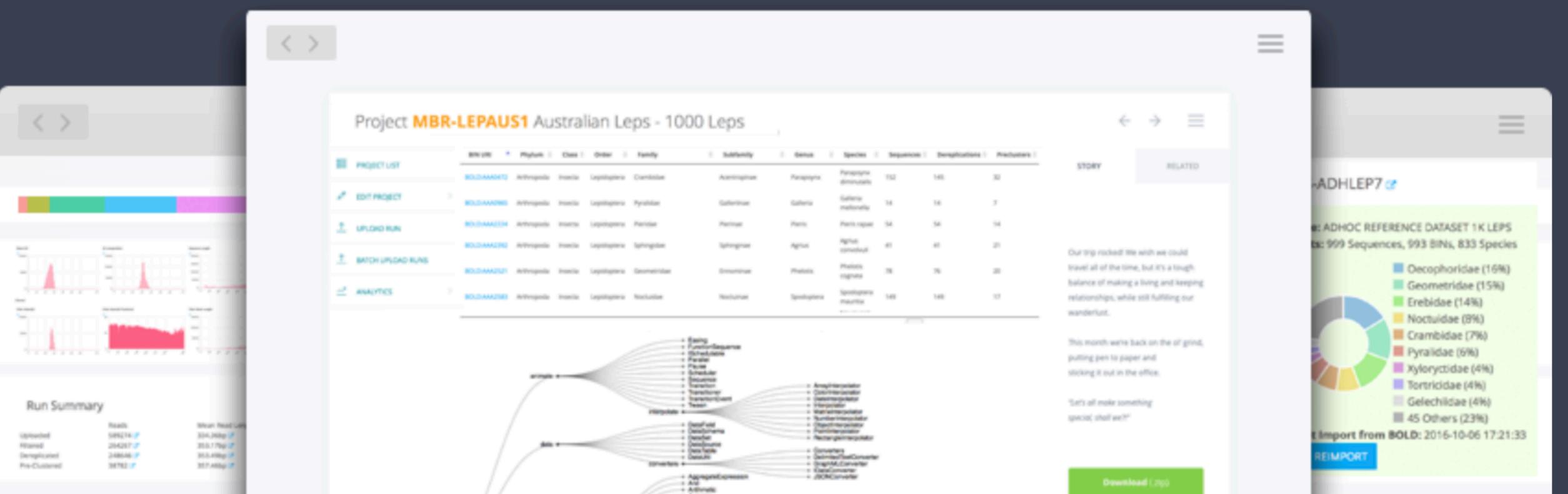


- 10 462 beetles
- collected biweekly over winter and summer
- alpine and subtropical sites
- individually DNA barcoded
- 2210 beetle BINS (“species”)
 - Barcode Identification Numbers



Multiplex Barcode Research And Visualization Environment

mBRAVE is a multi-user platform supporting the storage, validation, analysis, and publication of highly multiplexed projects based on high-throughput sequencing (HTS) instruments. This system builds on the [BOLD Platform](#) to support species identification and discovery for HTS data.



Login

Register

mBRAVE is currently accepting a limited number of new users on a weekly basis.

Storage

Analysis

Indexing

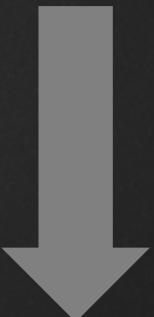
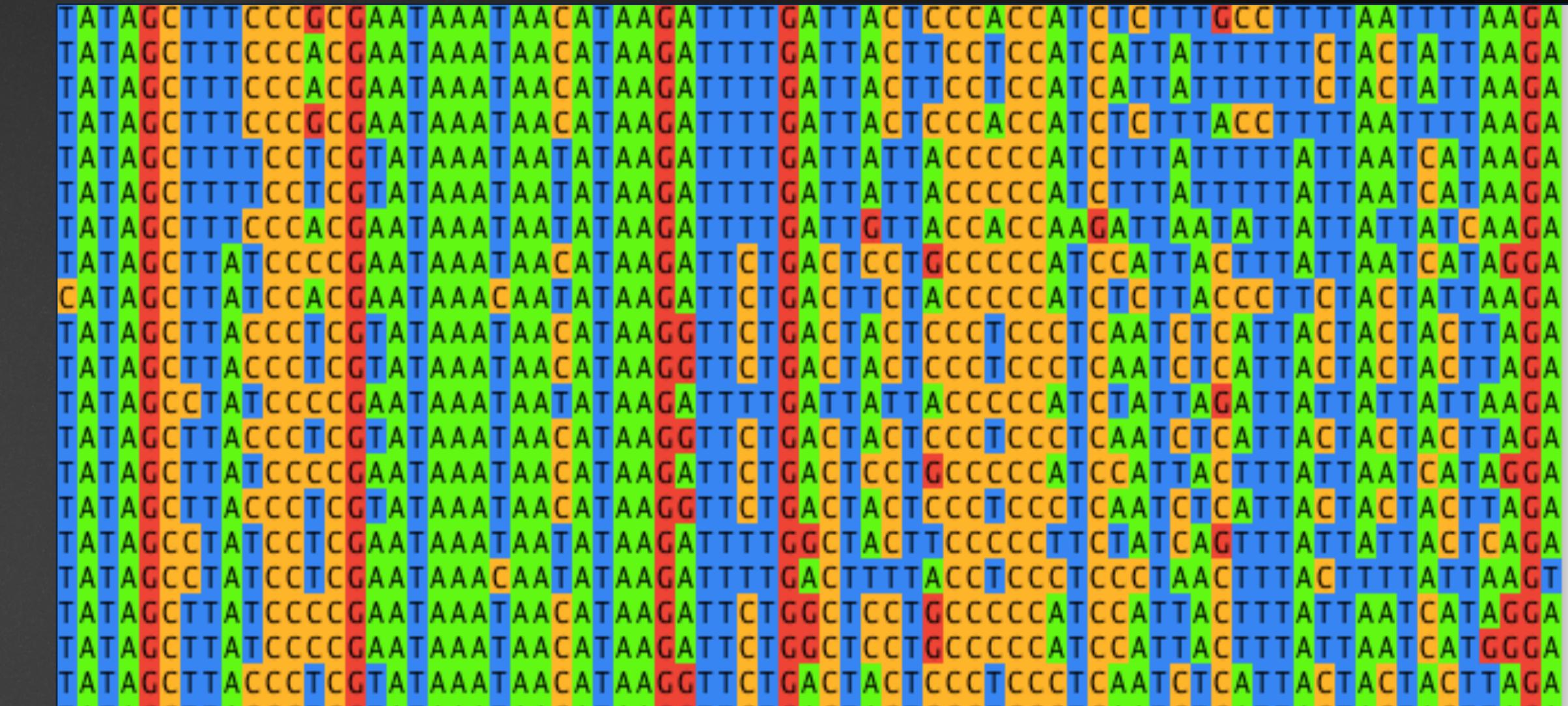
Discovery

Standards

mbraive.net

Massively parallel individual barcoding

Multiplexed individual barcoding

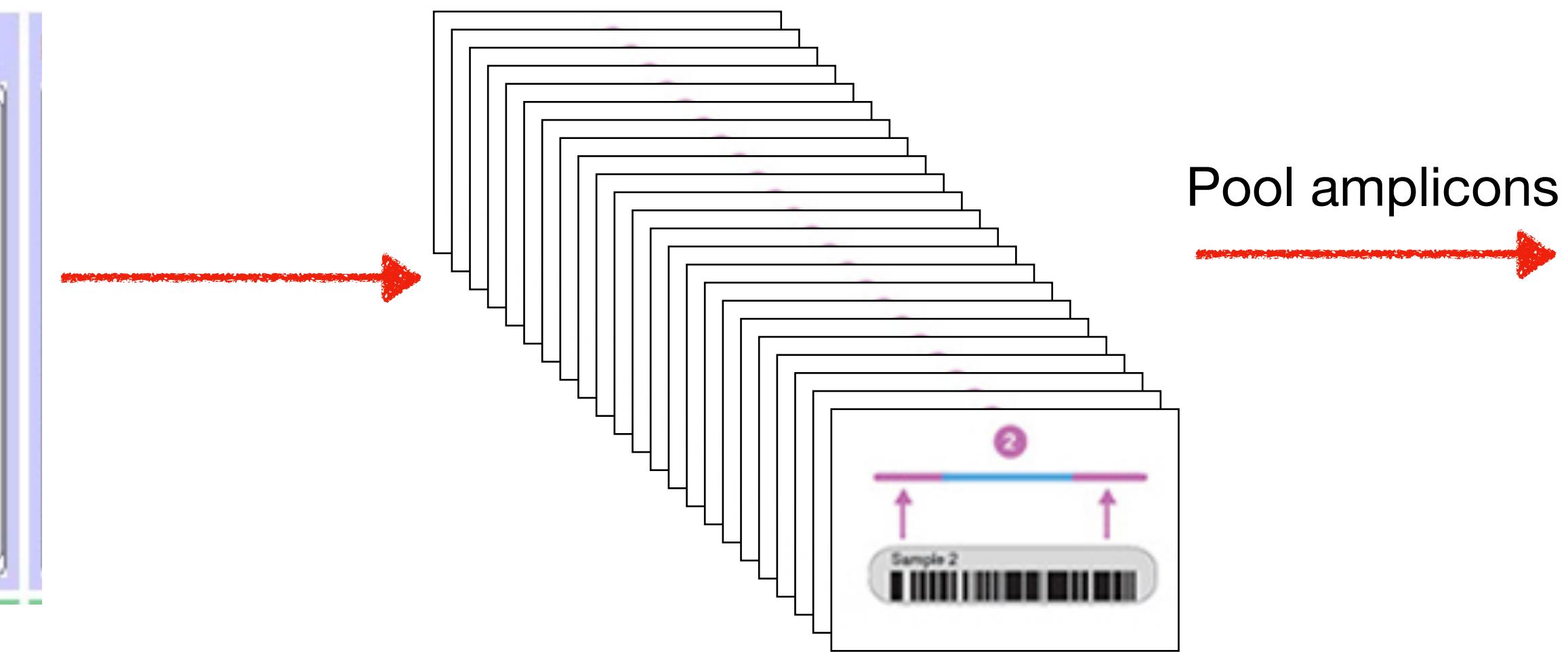
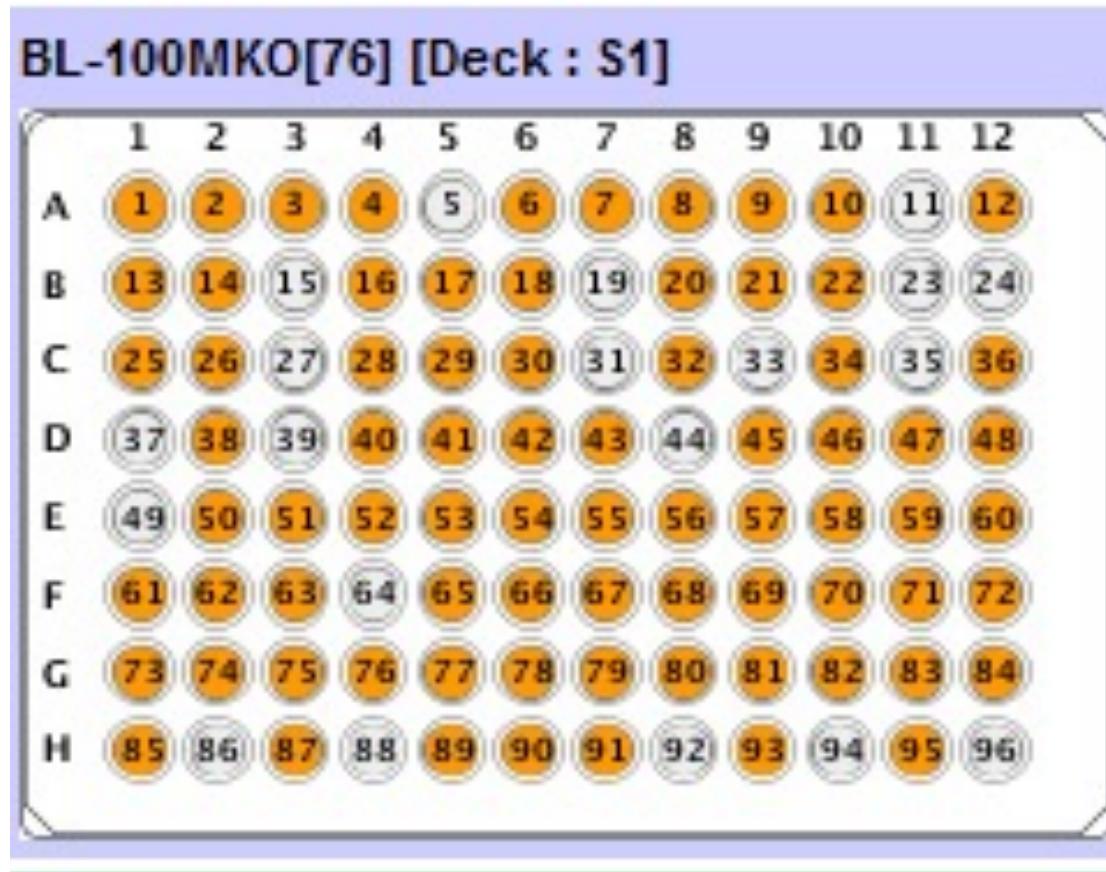


TATACTTATCCCCGAATAAAATAACATAAGATTGACTCCCCCATCATCAAATTATAAT

DNA from 96 beetles in 96 wells

\$\$ Add 96 COI primer pairs
with 96 different twin tags \$\$

Tag F Primer barcode sequence R Primer Tag



Pool amplicons

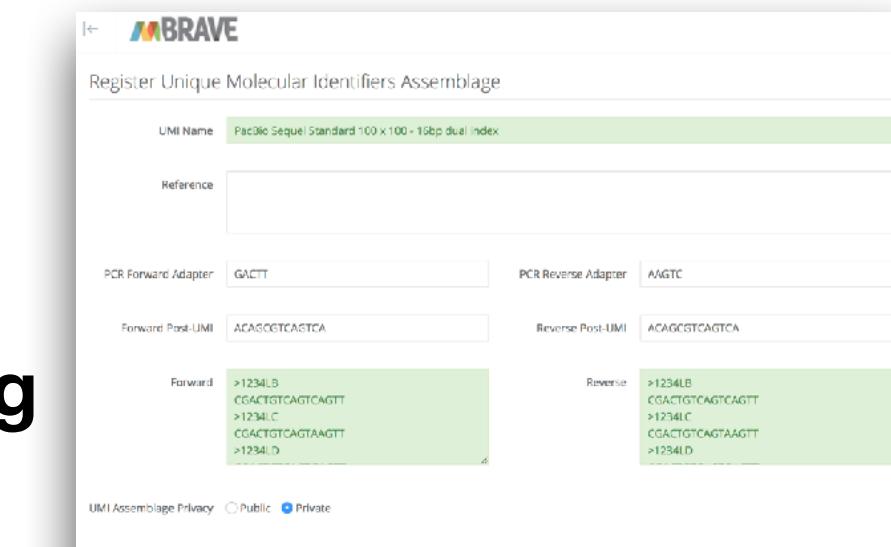


A consensus COI sequence for each beetle,
plus metadata, in fasta file: **count data!!**

```
>ANML_DLJ06_E12|run:ANML_DLJ06_dtd_0|contig_id:1|rep_count:1169|id_similarity:98.  
8950276243|c_count:1|cmdx:1.1|cmnd:0.1116338751|cnnd:None|p:None|c:None|o:None|f:None|g:  
None|s:None|otu:BOlD:AAB7930|date:2019-08-13  
AACTTTTATTTATTGGTGGCTGGGAATAGTTGGTATATCCCTTGATACTGATTGAGTTGAATAAGAACACCTGGTA  
CTCTACTAGGTGACGACCAATATAATGTAATCGTACAGCTCATGATTTATCATAATTCTTGTAGTTACCTATTGTATT  
GGT  
>ANML_DLJ04_C04|run:ANML_DLJ04_dtd_0|contig_id:1|rep_count:419|id_similarity:91.  
1602209945|c_count:3|cmdx:1.7|cmnd:0.2088385489|cnnd:3.3|p:Arthropoda|c:Insecta|o:None|f:  
None|g:None|s:None|otu:OTU14|date:2019-08-13  
AACACTATTTTGTGGAGCTTGAGCAGGTATAGTTGGGACATCACTGAGAGTTACTAATTGAGCAGAAATTAGGAACCCGGAT  
CTCTAATTGGTGTGACCAAATCTATAACGTAATTGTAACAGCCCAGTCATTGTAATAATTCTTGTAGTTACCTATTGTATT  
GGT  
>ANML_2909_D06|run:ANML_2909_dtd_0|contig_id:1|rep_count:20|id_similarity:nan|c_count:2|  
cmdx:1.1|cmnd:0.255|cnnd:2.2|p:None|c:None|o:None|f:None|g:None|s:None|otu:OTU37|date:  
2019-08-01  
TACTTGTATTTCTTTGGTGGCTGGAGCTGGAAATAGTAGGAACCTTAAAGCTCTAATTGAGCAGAAATTAGGAACCTGGTT  
CCCTAATTGGTGTGATCAGATTATAATGTAATTGTAACAGCACATGCTTCAATTATAATTCTTGTAGTTACCAATTGTATT  
GGA  
>ANML_DLJ05_H08|run:ANML_DLJ05_dtd_0|contig_id:1|rep_count:6196|id_similarity:nan|c_count:  
6|cmdx:1.7|cmnd:0.103518399|cnnd:2.2|p:None|c:None|o:None|f:None|g:None|s:None|otu:OTU2|  
date:2019-08-14  
CACTTTTATTTATCTTGGAGCATGGCTGGAATACTTGGAACATCCTTAAGACTCCTCATCCGGCAGAACCTGGAAACCCGGTT  
CTTAATCGGAAACATCAATTACAACGTAATTGTTACTGCCAGCTTCTATAATTCTTGTAGTAAACCATCATAATT  
GGT  
>ANML_2922_F09|run:ANML_2922_dtd_0|contig_id:1|rep_count:1342|id_similarity:nan|c_count:2|  
cmdx:1.7|cmnd:0.1236214605|cnnd:2.2|p:None|c:None|o:None|f:None|g:None|s:None|otu:OTU7|  
date:2019-08-02  
ATCCCTATCTTGTGGTATGATCAGGAATAGTTGGAACCTCATTAAGATTACTAATTGGCTGAATTAGGAATCTGGGT  
CTTTAATTGGAGCACCAATTATAATGTAATTGTAACAGCCCAGTCATTATAATTCTTGTAGTTACCTGGTAATAATT  
GGA  
>ANML_DLJ01_F05|run:ANML_DLJ01_dtd_0|contig_id:3|rep_count:10|id_similarity:nan|c_count:3|  
cmdx:0.6|cmnd:0.24|cnnd:2.6|p:None|c:None|o:None|f:None|g:None|s:None|otu:OTU15|date:  
2019-08-13  
AACCTTTACTTCTGGAGCATGAAGAGGTATAATCGAACCTCCATAAGATTGATGATTGAAACAGAATTAGGAACAGCTGGT  
CTCTAATTGGAGATGATAATTATAATGTAATTGTAACAGCACATGCTTCTATAATTCTTGTAGTAAACCCATCATGATA  
GGA
```

Demultiplex by tag

Keep the most
abundant read per tag
(= well = beetle), filter
out the rest, and
Assign a taxonomy to
using BOLD



A	B	C	D
Forward Labels	Reverse Labels	Label	Group
0048_Asf_Ibc269	0096_AsR_Ibc239_rc	CONTROL_H12_BIOUG39506	BL-106R12
0048_Asf_Ibc269	0095_AsR_Ibc172_rc	CONTROL_H12_BIOUG39372	BL-106R12
0048_Asf_Ibc269	0094_AsR_Ibc151_rc	GMPAT206-18	BL-106R12
0048_Asf_Ibc269	0093_AsR_Ibc180_rc	GMPAT206-18	BL-106R12
0048_Asf_Ibc269	0092_AsR_Ibc311_rc	GMPAT211-18	BL-106R12
0048_Asf_Ibc269	0091_AsR_Ibc354_rc	GMPAT205-18	BL-106R12
0048_Asf_Ibc269	0090_AsR_Ibc122_rc	GMPAU210-18	BL-106R12
0048_Asf_Ibc269	0089_AsR_Ibc283_rc	GMPAT204-18	BL-106R12
0048_Asf_Ibc269	0088_AsR_Ibc266_rc	GMPAU209-18	BL-106R12
0048_Asf_Ibc269	0087_AsR_Ibc203_rc	GMPAT203-18	BL-106R12
0048_Asf_Ibc269	0086_AsR_Ibc_rc	GMPAU208-18	BL-106R12
0048_Asf_Ibc269	0085_AsR_Ibc261_rc	GMPAT202-18	BL-106R12
0048_Asf_Ibc269	0084_AsR_Ibc305_rc	GMPAU207-18	BL-106R12
0048_Asf_Ibc269	0083_AsR_Ibc310_rc	GMPAT201-18	BL-106R12

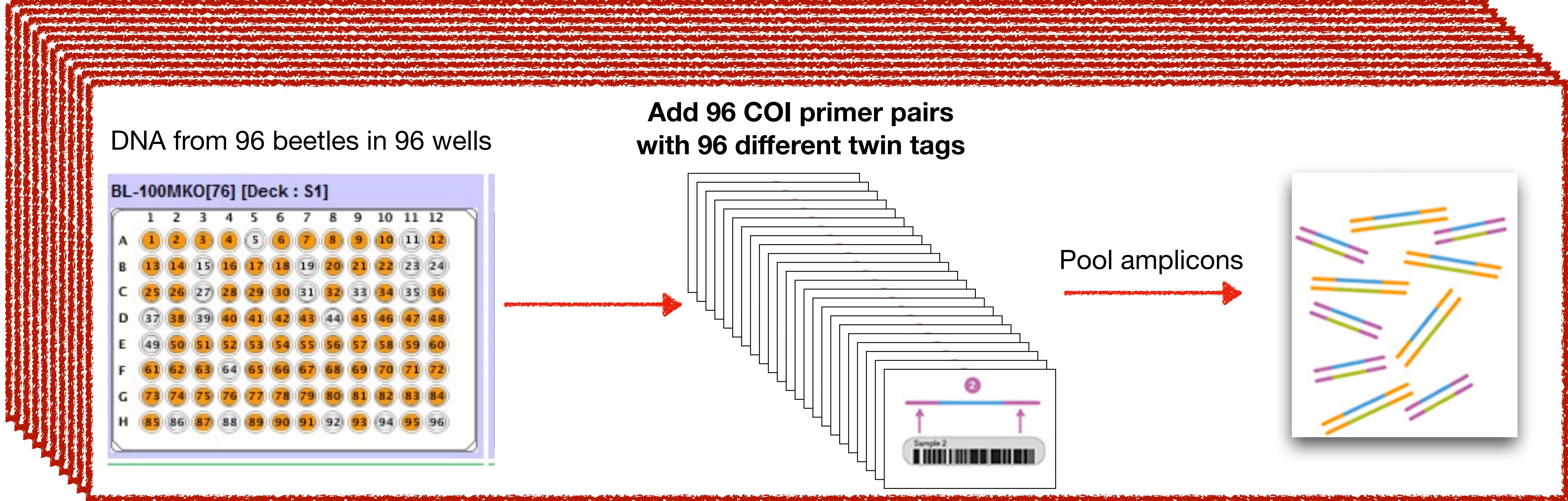
Upload fastq files
to mBRAVE

Upload tag
information to
mBRAVE



Illumina
sequence

Illumina
library
prep



A consensus COI sequence for each beetle,
plus metadata, in fasta file: **count data!!**

>ANML_DLJ06_E12|run:ANML_DLJ06_dtd_0|contig_id:1|rep_count:1169|id_similarity:98.
8950276243|c_count:1|cmxd:1.1|cmnd:0.1116338751|cnnd:None|p:None|c:None|o:None|f:None|g:
None|s:None|otu:BOLD:AAB7930|date:2019-08-13
AACTCTTATTTATTGGTCTGGGCTGGAAAGTGTGATATCCCTAGTATACTGATTGAGTTGAATTAAAGAACACCTGGT
CTCTACTAGGTGACGACCAATATAATGTAATCGTTACAGCTCATGCATTATCATAATTCTTATAGTTACCTATTGTAATT
GGT
>ANML_DLJ04_C04|run:ANML_DLJ04_dtd_0|contig_id:1|rep_count:419|id_similarity:91.
1602209945|c_count:3|cmxd:1.7|cmnd:0.2088305489|cnnd:3.3|p:Arthropoda|c:Insecta|o:None|f:
None|g:None|s:None|otu:OTU14|date:2019-08-13
AAACACTATTTCATTTGGAGCTTGAGCAGGTAGTGTGGGACATCACTGAGATTACTAATTCGAGCAGAATTAGGAAACCCGGAT
CTCTAATTGGTGTGACCAAATCTATAACGTAATTGTAACAGCCATGCATTGTAATAATTCTTATAGTTACCTATTGTAATT
GGT
>ANML_2909_D06|run:ANML_2909_dtd_0|contig_id:1|rep_count:20|id_similarity:nan|c_count:21
cmxd:1.1|cmnd:0.255|cnnd:2.2|p:None|c:None|o:None|f:None|g:None|s:None|otu:OTU37|date:
2019-08-01
TACTTTGTATTCATTTGGTCTTGAGCTGGAAAGTAGTAGGAACTCTTAAGCCTCTAATTGAGCAGAATTAGGAACTCTGGT
CCCTAATTGGTGTGATCAGATTATAATGTAATTGTAACGTCATGCTTCATTATAATTCTTATAGTTACCAATTGTTATT
GGA
>ANML_DLJ05_H08|run:ANML_DLJ05_dtd_0|contig_id:1|rep_count:6196|id_similarity:nan|c_count:
6|cmxd:1.7|cmnd:0.103518399|cnnd:2.2|p:None|c:None|o:None|f:None|g:None|s:None|otu:OTU2|
date:2019-08-14
CACTCTTATTTATCTGGAGCATGGCTGGAAACTTGGAACATCCTTAAGACTCCTCATCCGAGCAGAACTGGAAACCCGGTT
CTTAATCGGAAACGATCAAATTACAACGTAATTGTTACTGCCACGCTTCATTATAATTCTCATAGTAATACCAATCATAATT
GGT
>ANML_2922_F09|run:ANML_2922_dtd_0|contig_id:1|rep_count:1342|id_similarity:nan|c_count:2
|cmxd:1.7|cmnd:0.1236214605|cnnd:2.2|p:None|c:None|o:None|f:None|g:None|s:None|otu:OTU7|
date:2019-08-02
ATCCTTATCTTTGGTATATGTCAGGAATAGTGGAACCTCATTAAAGATTACTAATTCGCTAGAATTAGGTAATCCTGGT
CTTAATTGGAGACGACCAATTATAATGTAATTGTAACGTCATGCTTATTATAATTCTTATAGTTACCTGTAATAATA
GGA
>ANML_DLJ01_F05|run:ANML_DLJ01_dtd_0|contig_id:3|rep_count:10|id_similarity:nan|c_count:3
|cmxd:0.6|cmnd:0.24|cnnd:26.1|p:None|c:None|o:None|f:None|g:None|s:None|otu:OTU115|date:
2019-08-13
AACCTCTTACTCATTGGAGCATGAAGAGGTATAATCGGAACCTCCATAAGATTGATGATTGAGAACAGAATTAGGAAACAGCTGGT
CTCTAATTGGAGATGATCAAATTATAATGTAATTGTAACAGCACATGCTTATCATAATTCTTATAGTAATACCCATCATGATA
GGA

Demultiplex by library and tag

Keep the most abundant read per taxon (= well = beetle), filter out the rest, and Assign a taxonomy to it using BOLD

← **BRAVE**

Register Unique Molecular Identifiers Assemblage

UMI Name	Pacific Sequel Standard 100 x 100 - 16bp dual index		
Reference			
PCR Forward Adapter	GACTT	PCR Reverse Adapter	AAGTC
Forward Post-UMI	ACAGCCTCACTCA	Reverse Post-UMI	ACAGCCTCAGTC
Forward	<pre>>1234LB CGACTGTCACTCAGTT >1234LC CGACTGTCACTAAGTT >1234LD</pre>	Reverse	<pre>>1234LB CGACTGTCACTCAGTT >1234LC CGACTGTCACTAAGTT >1234LD</pre>

UMI Assemblage Privacy: Public Private

Upload fastq files to mBRAVE

Upload tag
information to
mBRAVE

	A	B	C	D
1	Forward Labels	Reverse Labels	Label	Group
2	0048_AsF_lbc269	0096_AsR_lbc239_rc	CONTROL_H12_BIOUG39506	BL-106R12
3	0048_AsF_lbc269	0095_AsR_lbc172_rc	CONTROL_H12_BIOUG39372	BL-106R12
4	0048_AsF_lbc269	0094_AsR_lbc151_rc	GMPAU212-18	BL-106R12
5	0048_AsF_lbc269	0093_AsR_lbc180_rc	GMPAT206-18	BL-106R12
6	0048_AsF_lbc269	0092_AsR_lbc311_rc	GMPAU211-18	BL-106R12
7	0048_AsF_lbc269	0091_AsR_lbc354_rc	GMPAT205-18	BL-106R12
8	0048_AsF_lbc269	0090_AsR_lbc122_rc	GMPAU210-18	BL-106R12
9	0048_AsF_lbc269	0089_AsR_lbc283_rc	GMPAT204-18	BL-106R12
10	0048_AsF_lbc269	0088_AsR_lbc266_rc	GMPAU209-18	BL-106R12
11	0048_AsF_lbc269	0087_AsR_lbc8_rc	GMPAT203-18	BL-106R12
12	0048_AsF_lbc269	0086_AsR_lbc1_rc	GMPAU208-18	BL-106R12
13	0048_AsF_lbc269	0085_AsR_lbc261_rc	GMPAT202-18	BL-106R12
14	0048_AsF_lbc269	0084_AsR_lbc305_rc	GMPAU207-18	BL-106R12
15	0048_AsF_lbc269	0083_AsR_lbc310_rc	GMPAT201-18	BL-106R12



The logo consists of a vertical stack of ten red, textured bars of varying heights, resembling a barcode or a stylized mountain range. Below this stack is a horizontal row of five white, jagged peaks, also resembling a mountain range. The entire graphic is set against a white background.

Illumina sequence

Methods to extract abundance information from DNA data

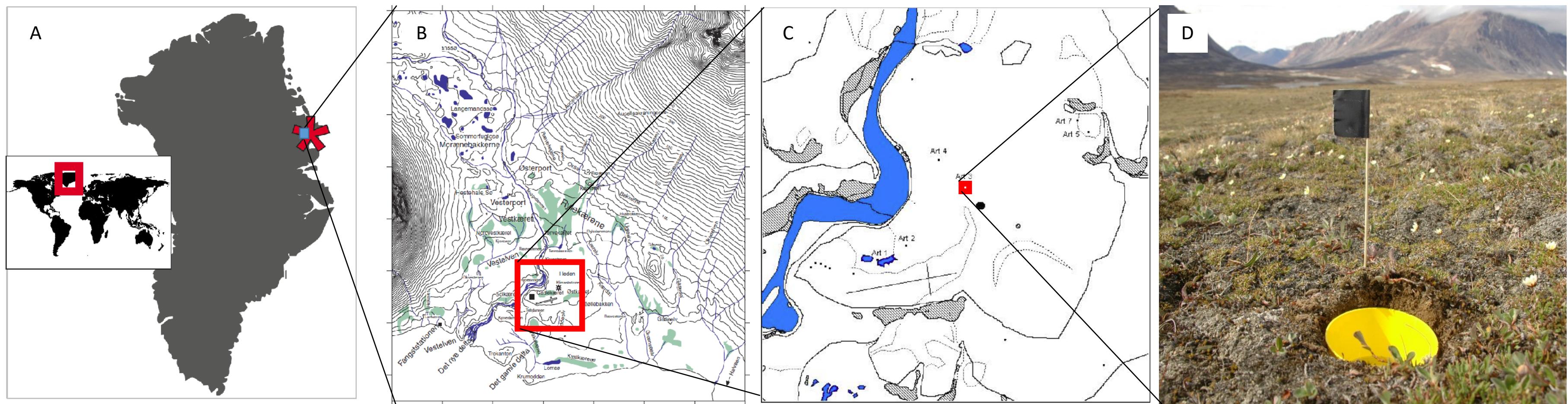
- Single-species quantitative PCR (qPCR)
- Multiplexed individual barcoding (mBRAVE): *individual count data in the 1000s*
- **Mitogenomics and DNA spike-in (SPIKEPIPE)**
- Metabarcoding and DNA spike-in (qSeq)

Zackenberg Research Station, Greenland



SPIKEPIPE: A metagenomic pipeline for the accurate quantification of eukaryotic species occurrences and intraspecific abundance change using DNA barcodes or mitogenomes

Yinqui Ji^{1*} | Tea Huotari^{2*} | Tomas Roslin^{2,3} | Niels Martin Schmidt^{4,5} |
Jiaxin Wang¹ | Douglas W. Yu^{1,6,7} | Otso Ovaskainen^{8,9}



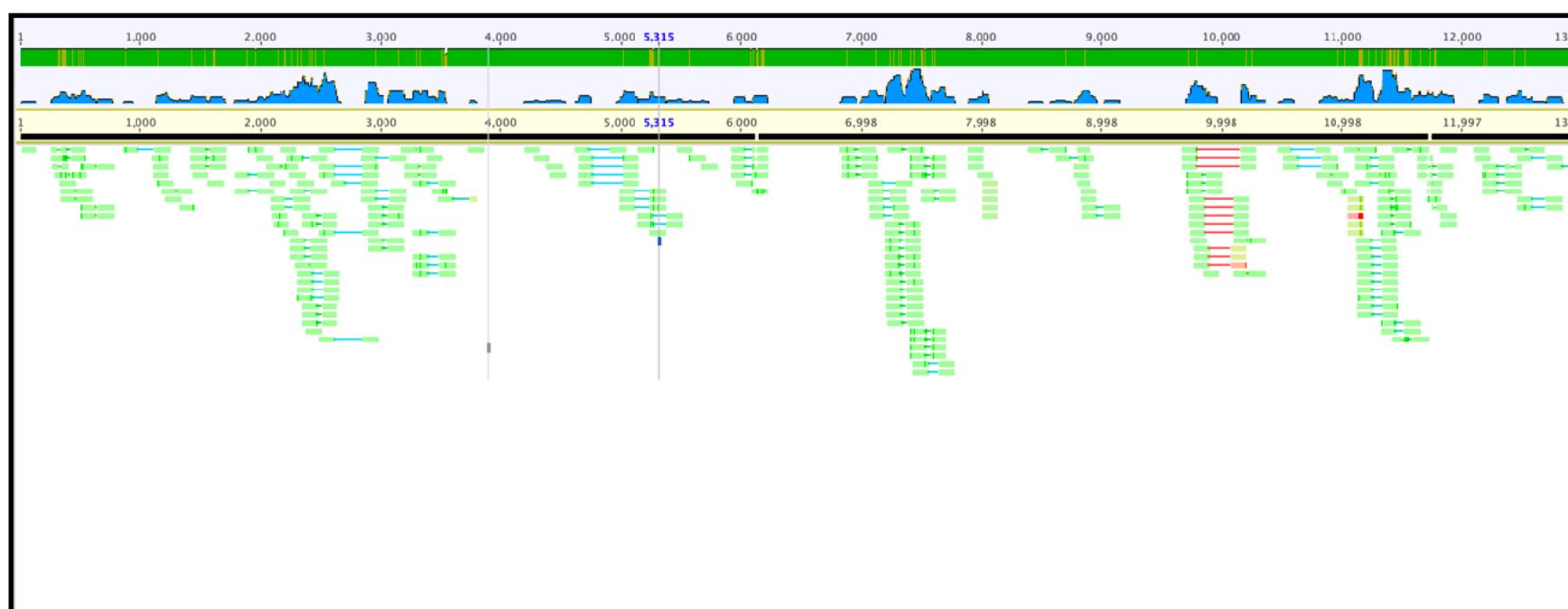


- entire aboveground arthropod community **~375 species**
- **>760,000 arthropods** collected in weekly samples and multiple pan traps from **1996-2013** (and ongoing)
- We assembled **308 mitogenome** sequences
- We shotgun-sequenced **~750 samples: 3 samples per week from 1997-2013** + technical replicates

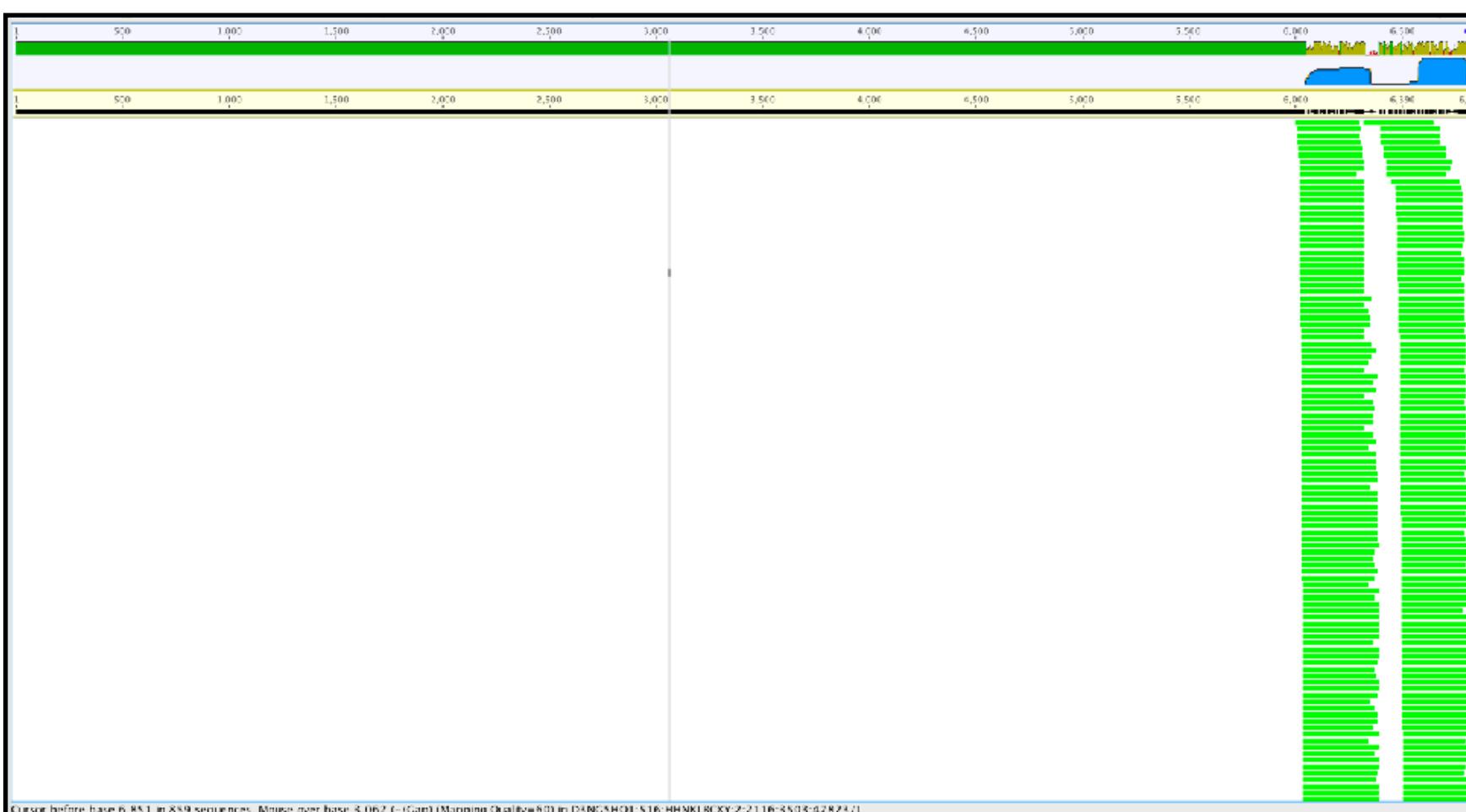
We mapped each sample's short reads to 308 mitogenomes



This mitogenome (a species) is confidently present in high biomass

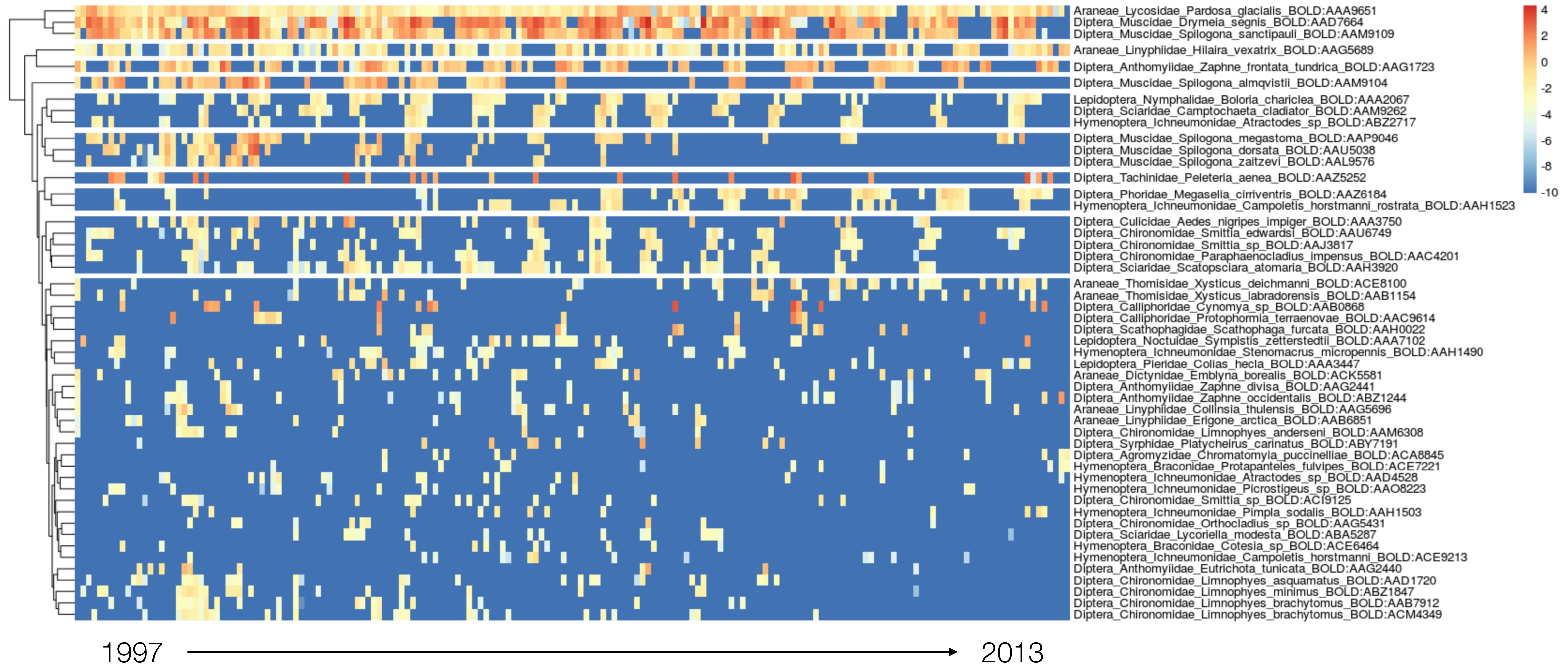


This species is confidently present in low biomass



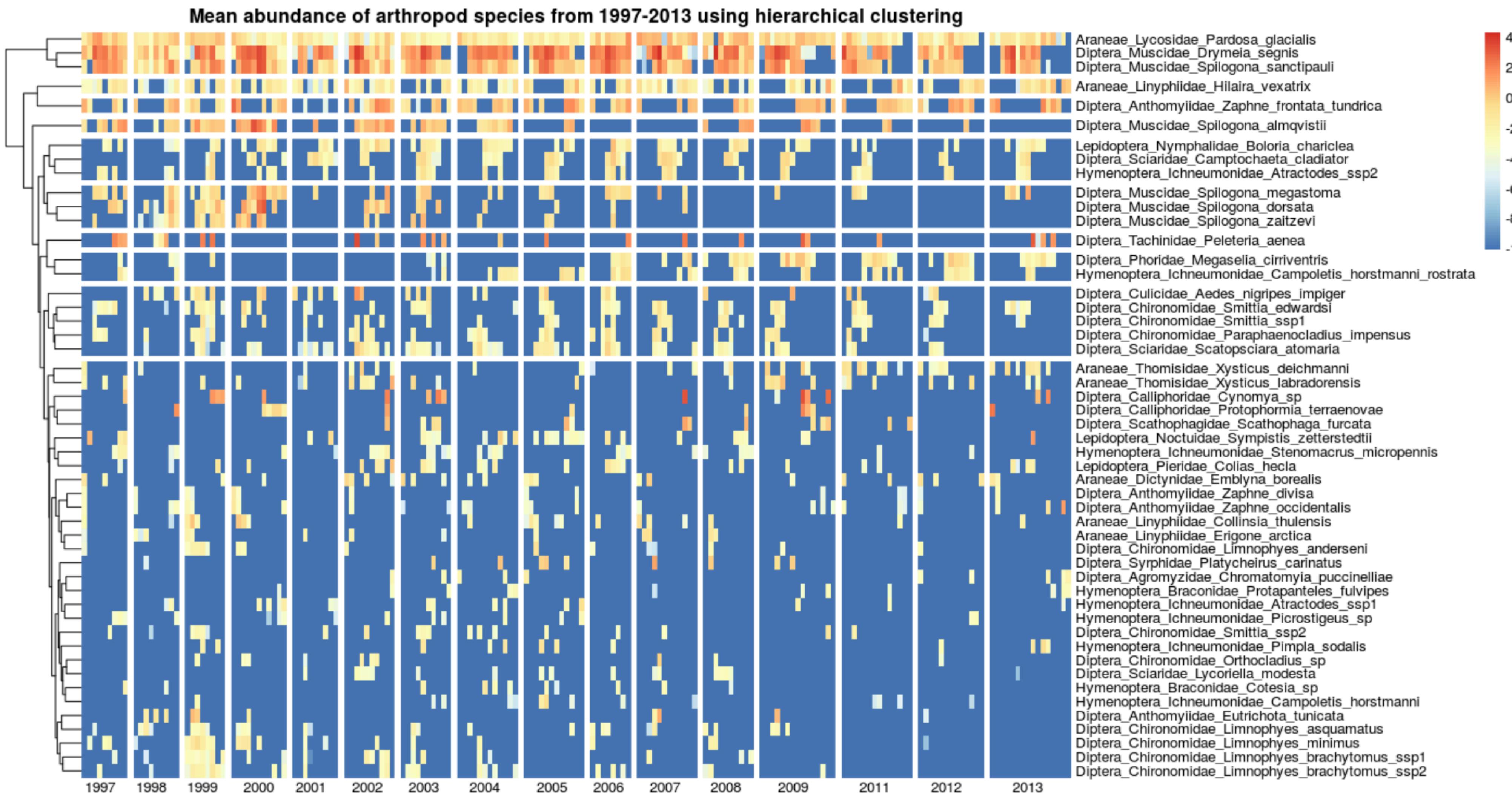
This species is not present, despite the many mapped reads (they mapped only to a conserved sequence)

What we get is a ***community time series***
 (red = highest **within-species** abundance, blue = absence)



more blue cells (more absences) toward 2013

What we get is a ***community time series***
 (red = highest **within-species** abundance, blue = absence)

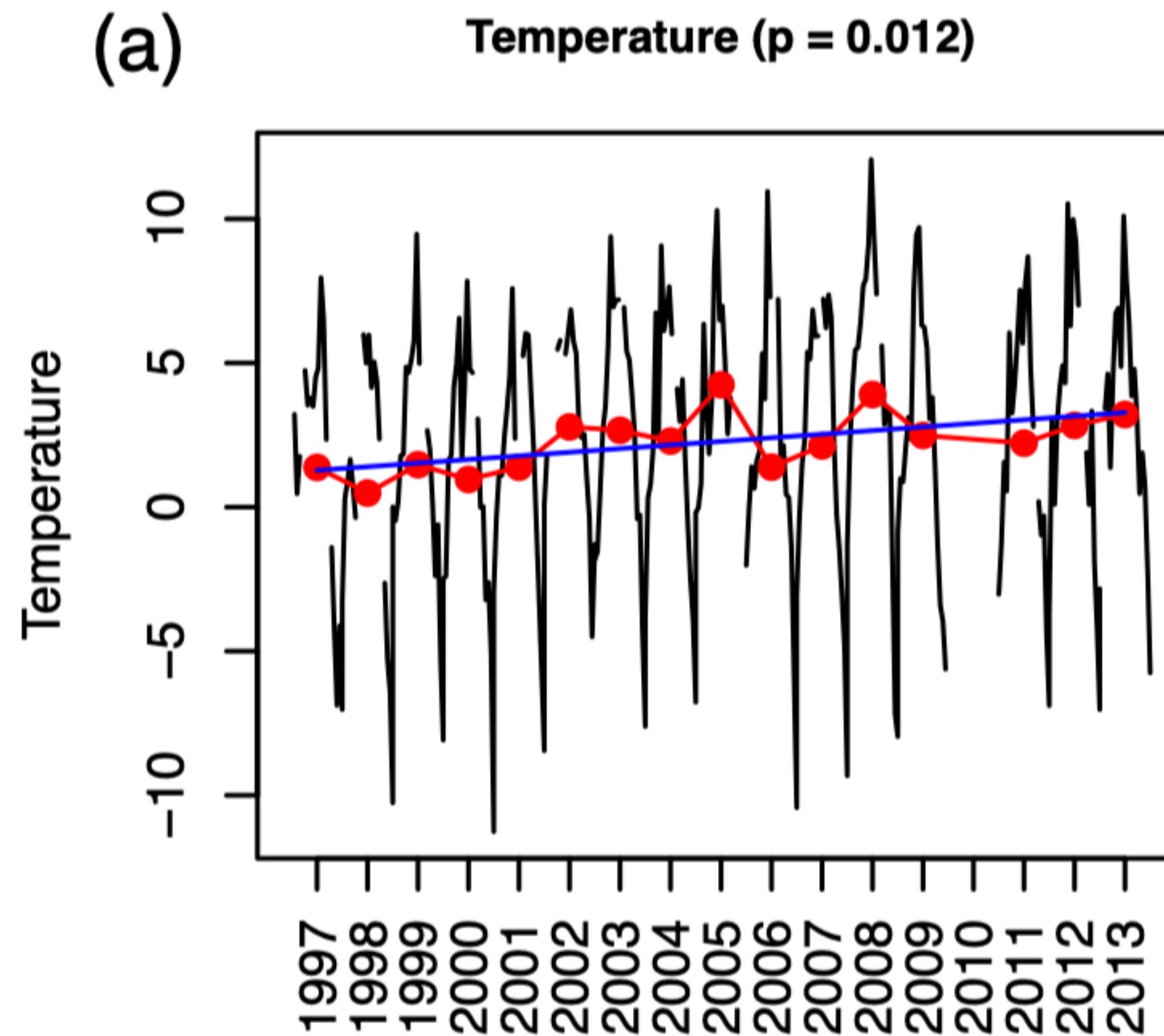


Summer has approx doubled in length in 16 years

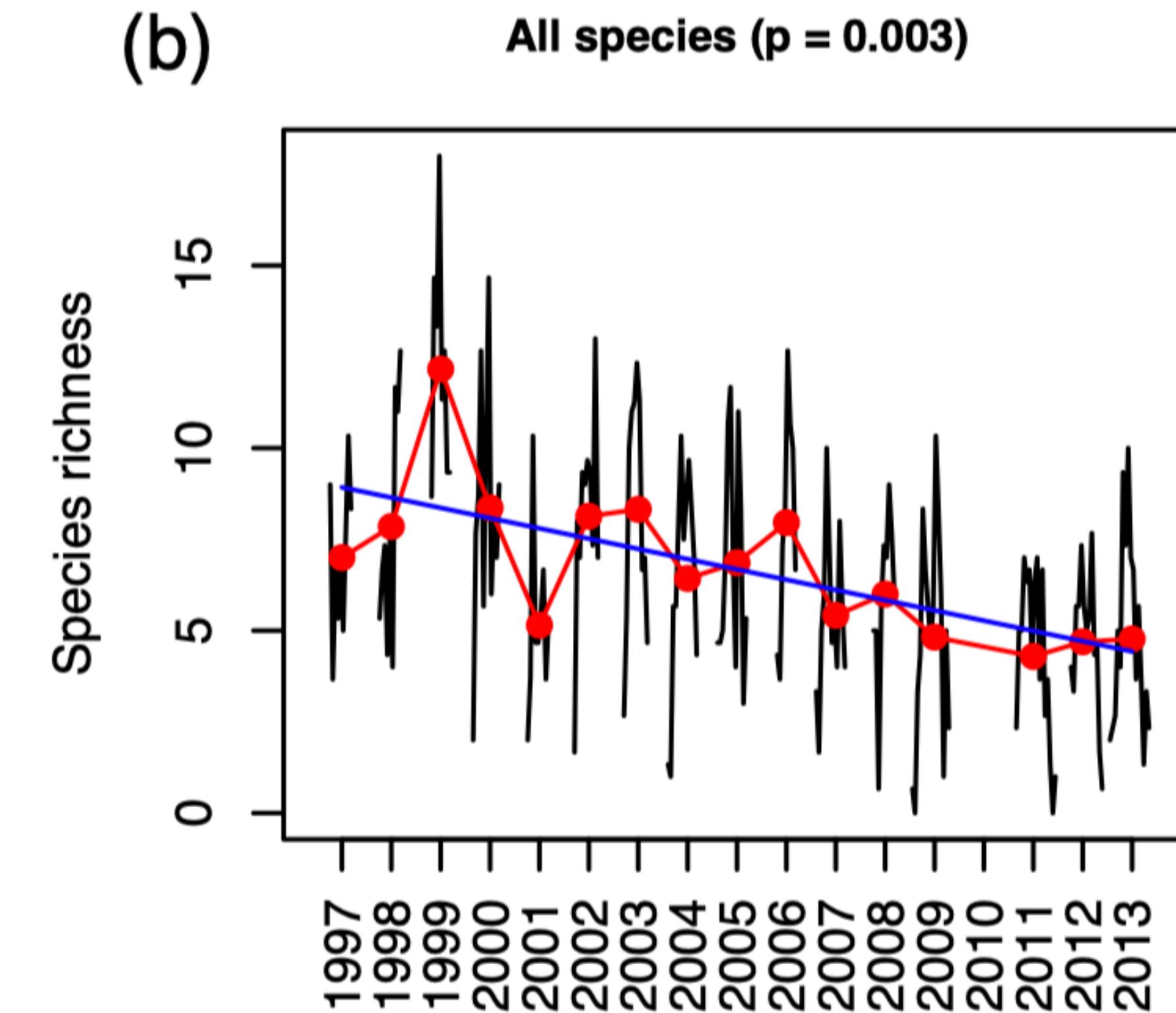
Accounting for species interactions is necessary for predicting how arctic arthropod communities respond to climate change

Nerea Abrego, Tomas Roslin, Tea Huotari, Yinqiu Ji, Niels Martin Schmidt, Jiaxin Wang, Douglas W. Yu and Otso Ovaskainen

Ecography
44: 1–12, 2021
doi: 10.1111/ecog.05547



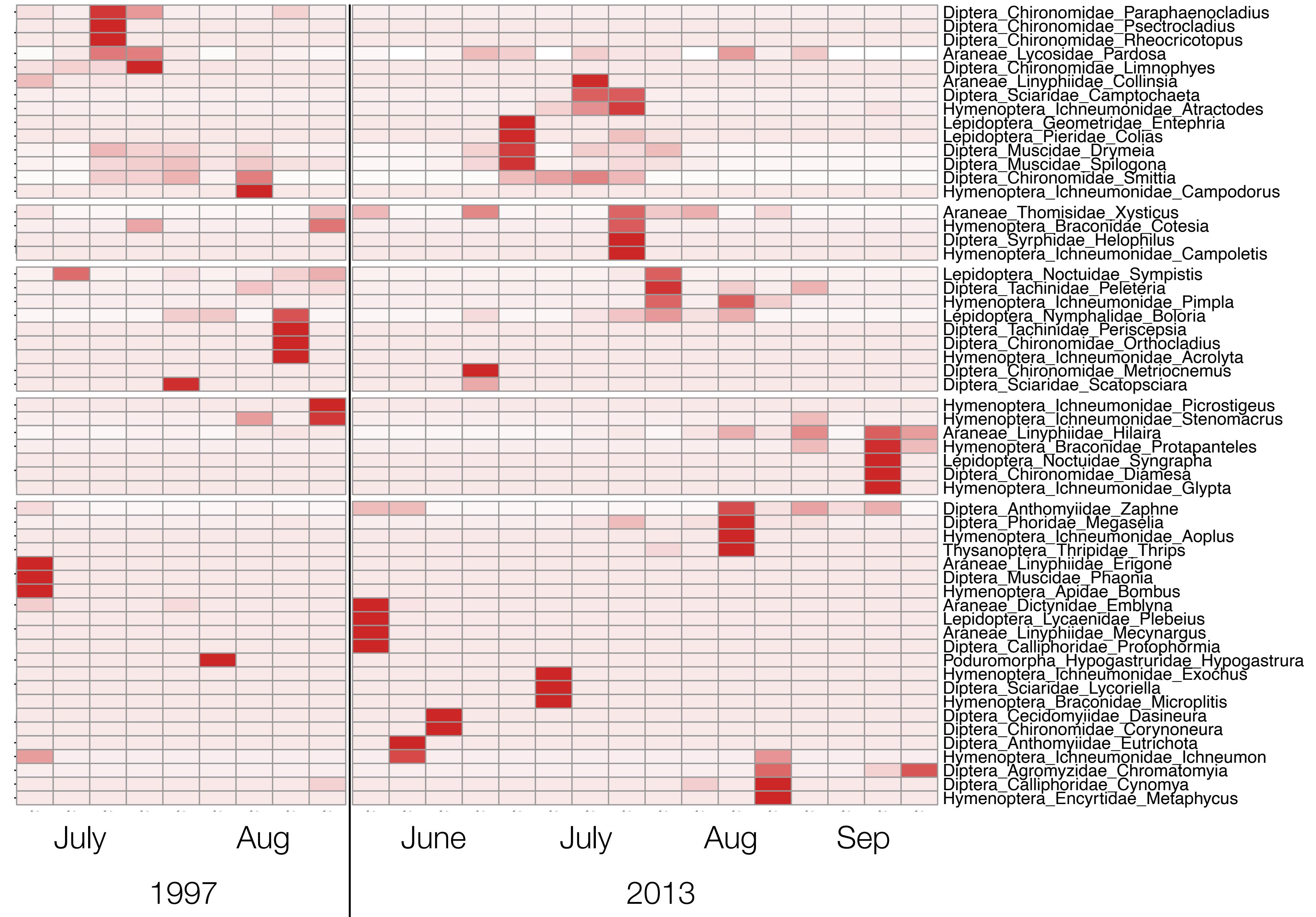
average summer temperature has increased by 2.0°C (from 1.28°C to 3.28°C)



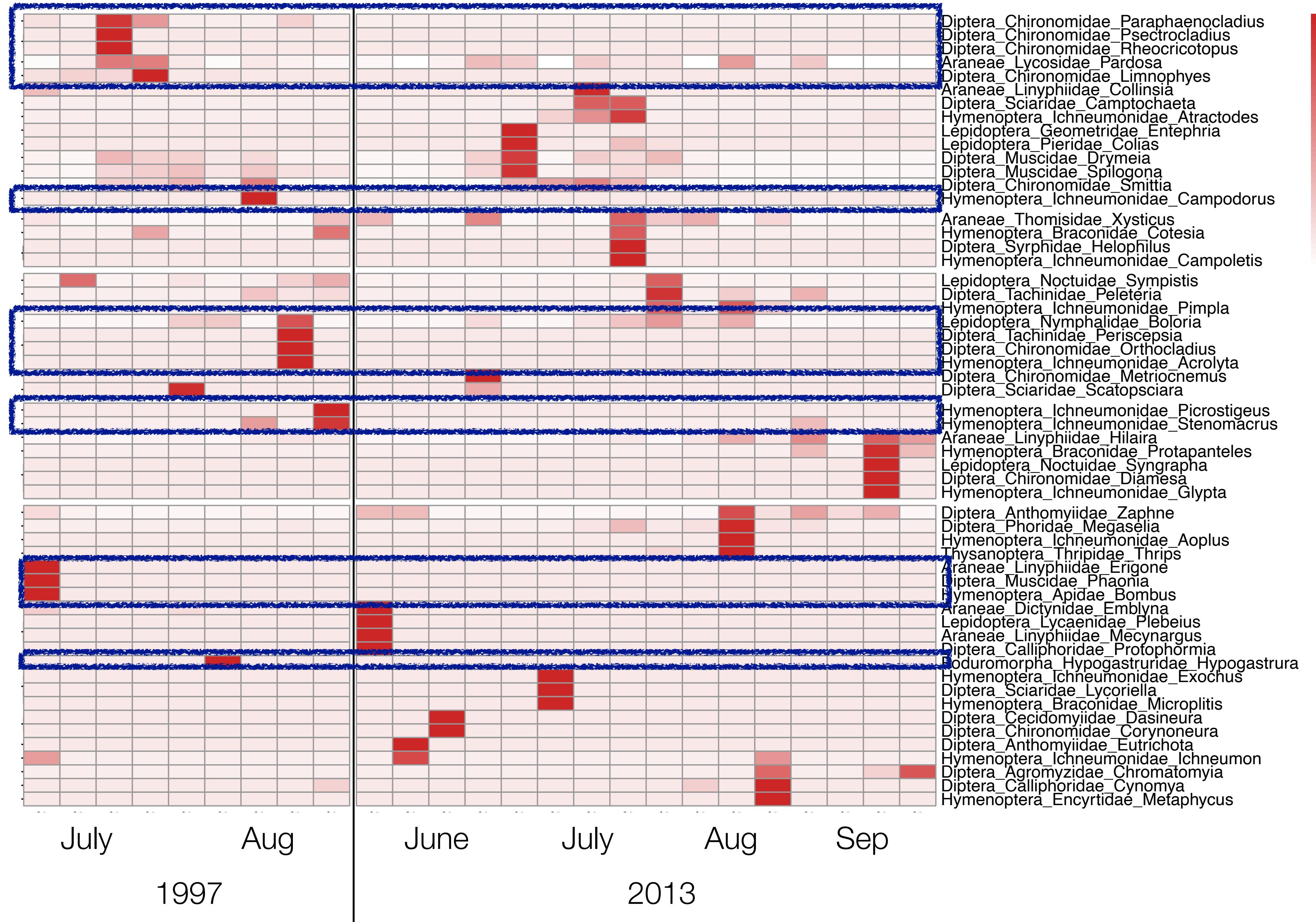
52 species in 542 trap-week samples

(observed) species richness has declined by half (8.9 to 4.4 species per trap-week, decline is mostly in the predator species)

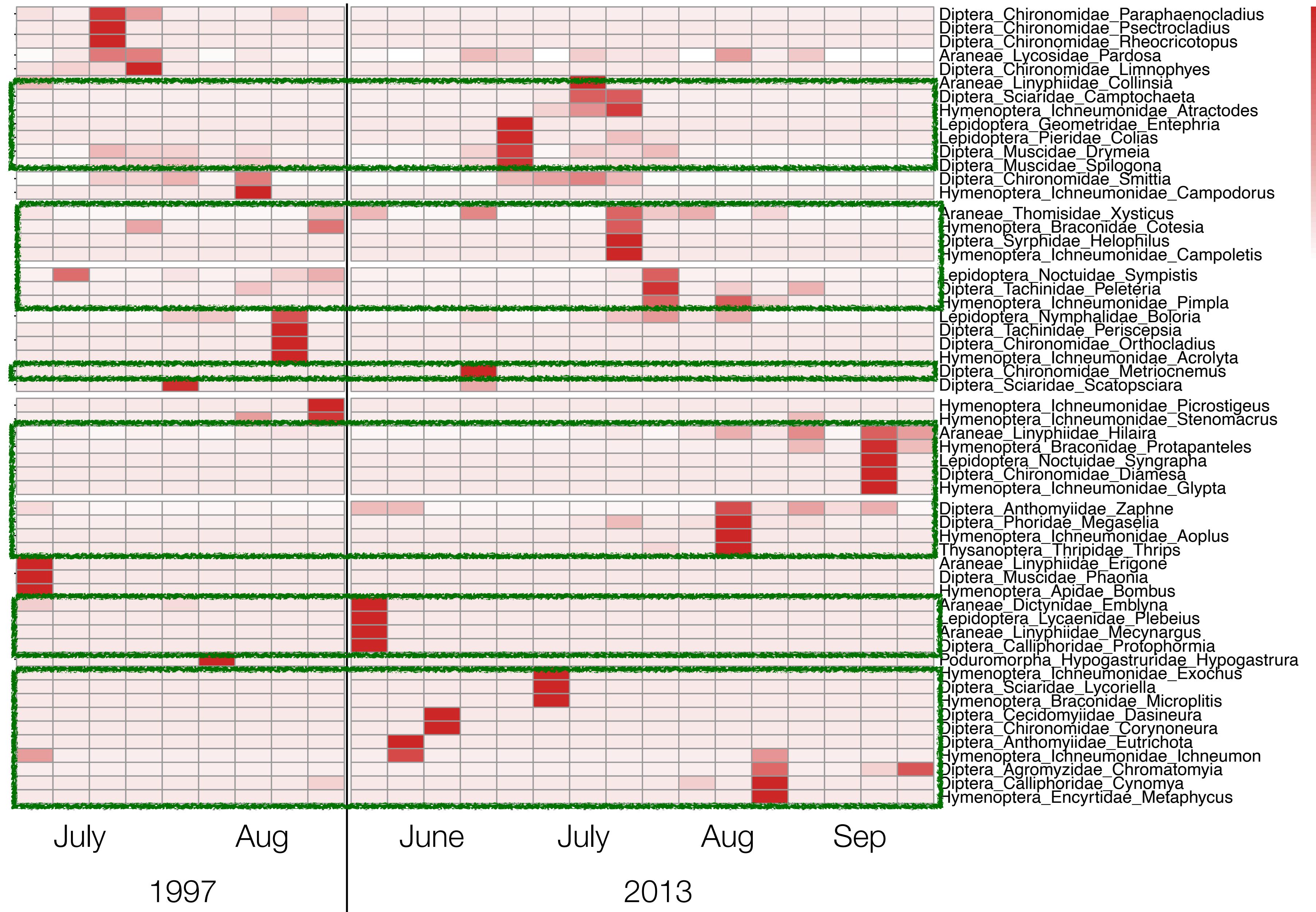
Let's look at 2 years of data (darker red = more biomass of that species <- within-species quantification)



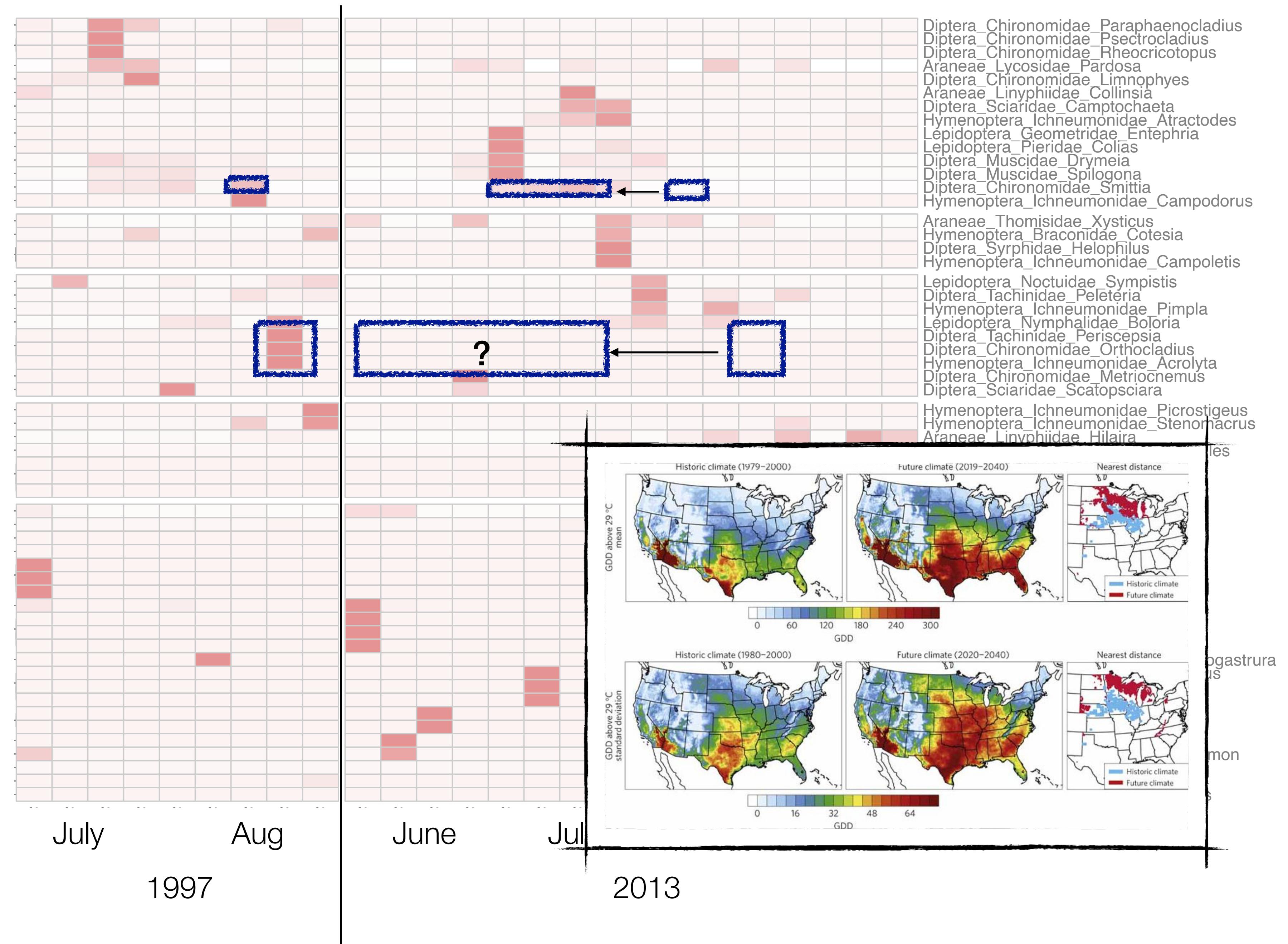
Many species that exhibited biomass peaks in 1997, now no longer do



Many species that showed no biomass peaks in 1997, now do have activity peaks!

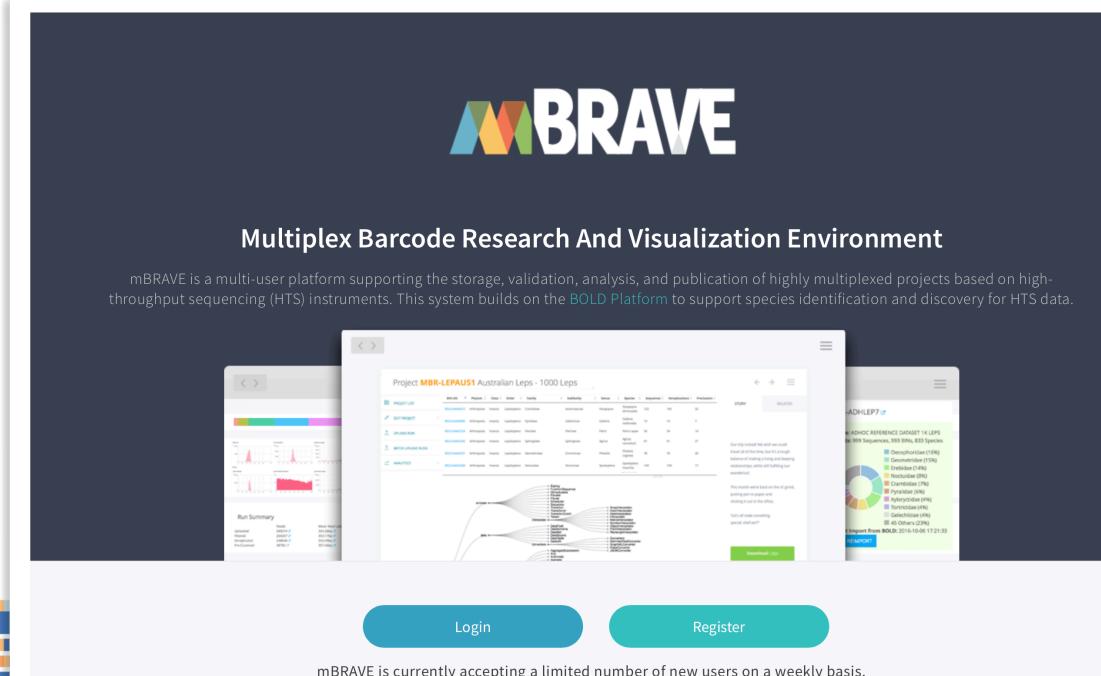
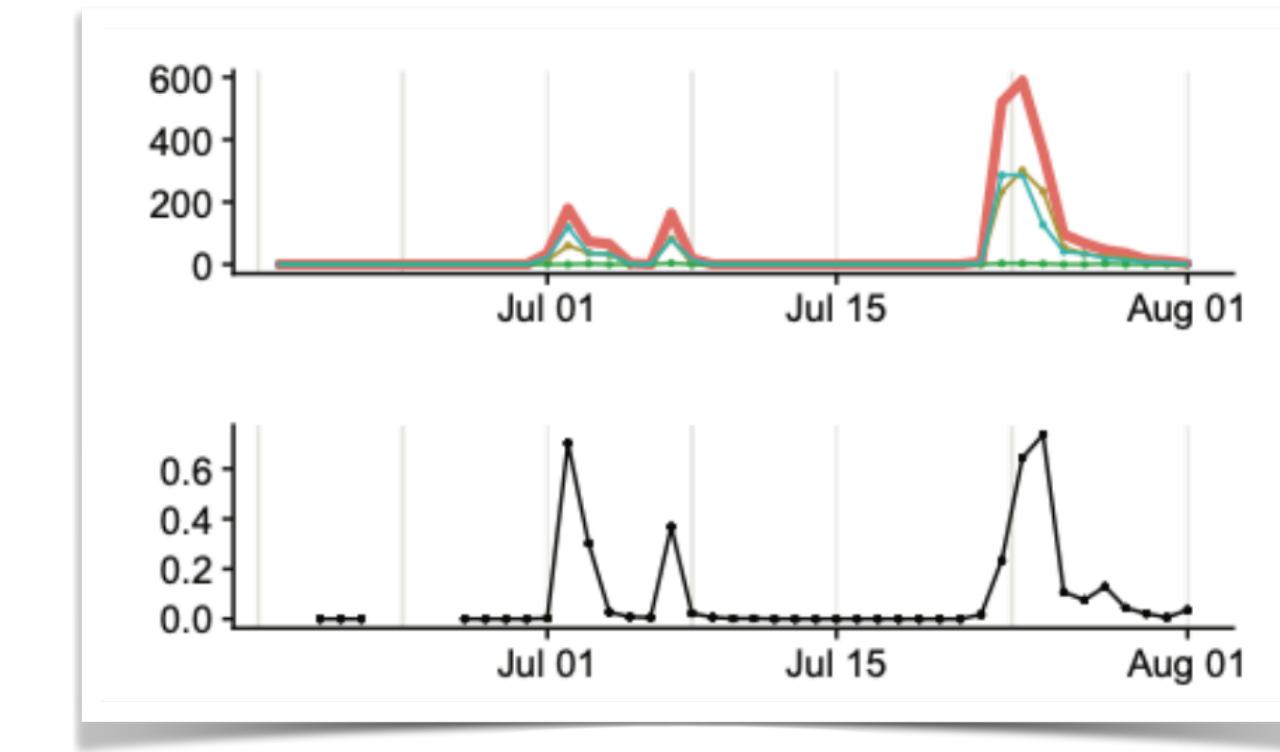


We rarely see simple shifts in phenology (timing of peaks): the climate envelope idea



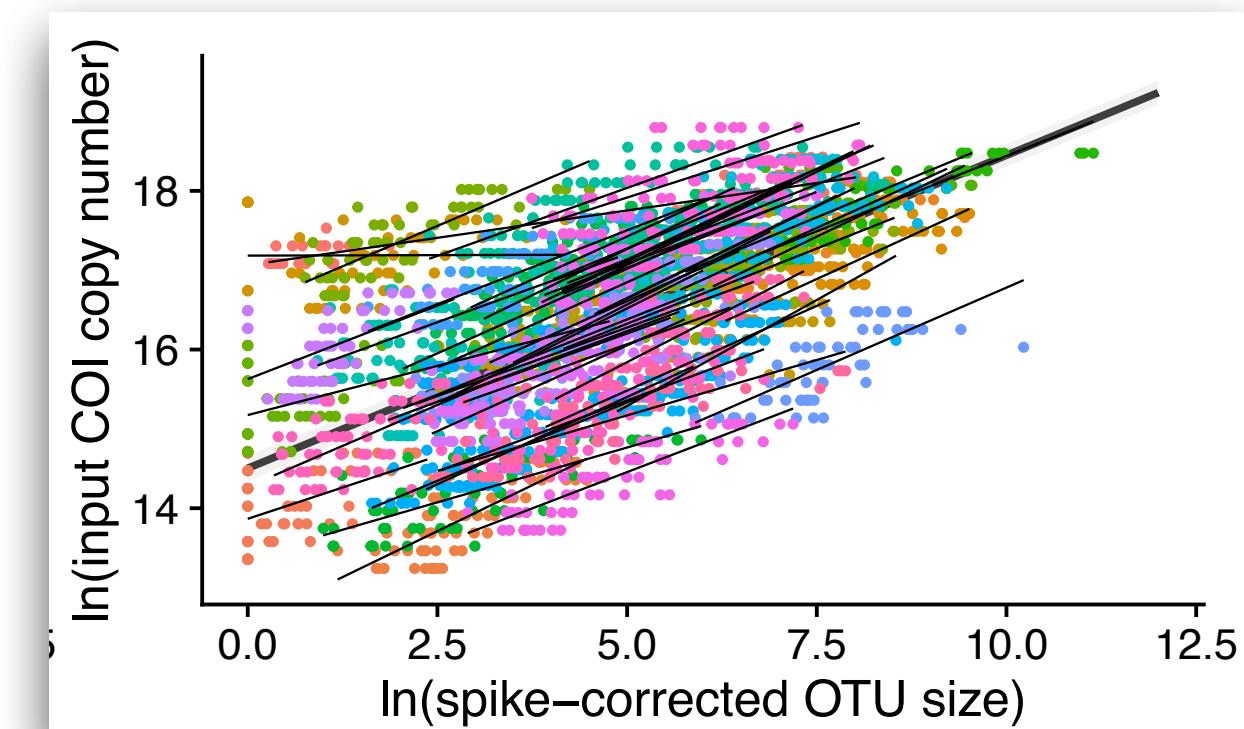
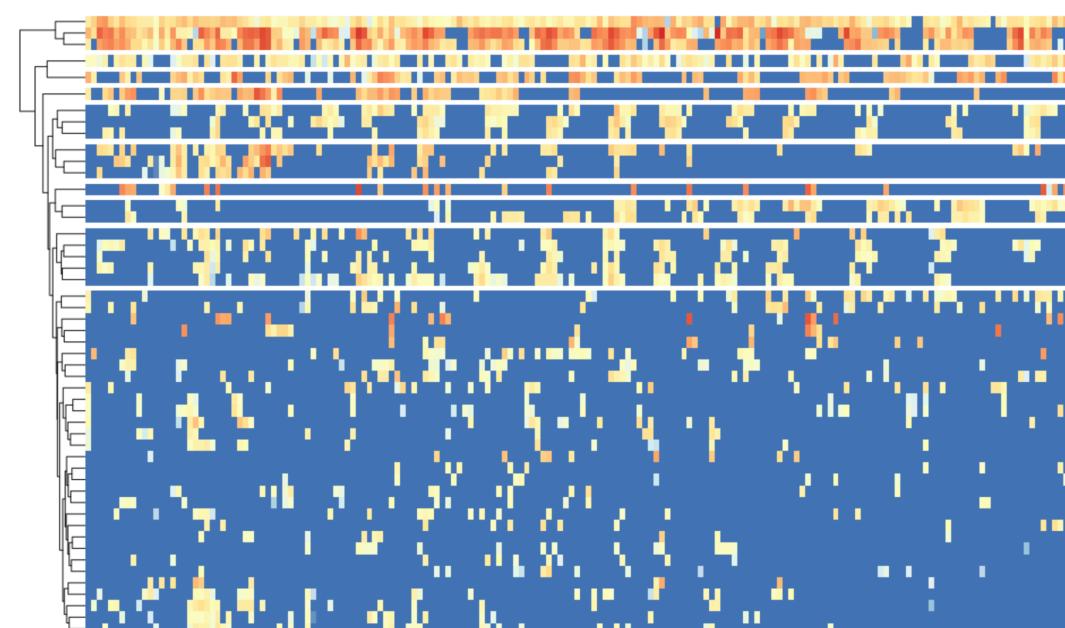
Methods to extract abundance information from DNA data

- Single-species quantitative PCR (qPCR)
- Multiplexed individual barcoding (mBRAVE)
- Mitogenomics and DNA spike-in (SPIKEPIPE)
- Metabarcoding and DNA spike-in (qSeq)



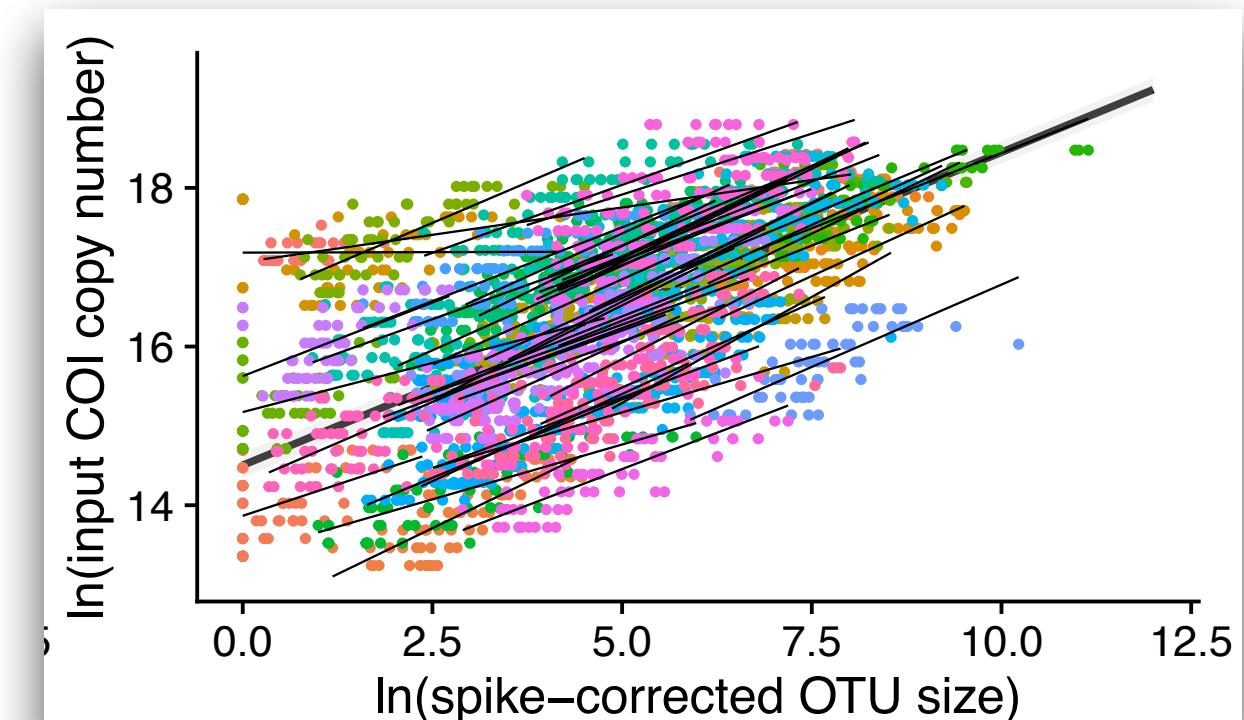
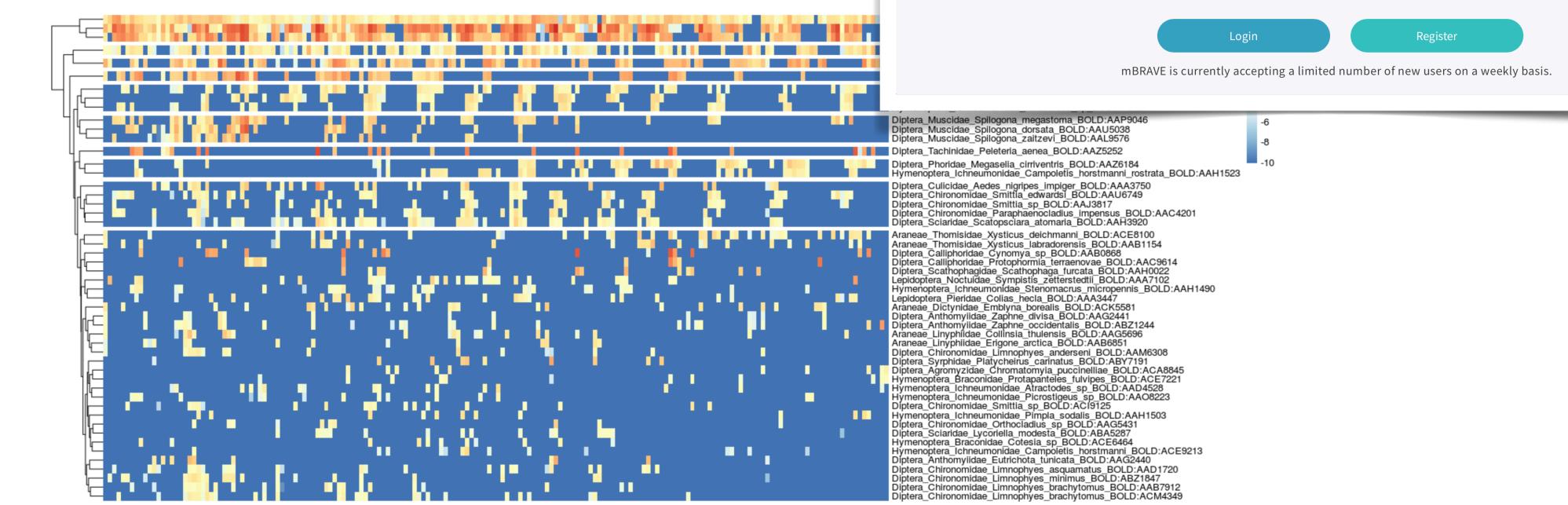
mBRAVE is currently accepting a limited number of new users on a weekly basis.

[Login](#) [Register](#)



Methods to extract abundance information from DNA data

- Single-species quantitative PCR (qPCR)
 - **use eDNA rate** not concentration! ng/sec not ng/ μ l
- Multiplexed individual barcoding (mBRAVE)
 - **count data feasible for many 1000s of individuals**
- Mitogenomics/Metabarcoding + DNA spike-in (SPIKEPIPE/qSeq)
 - **For within-species differences across samples** (e.g. time series)
 - (difficult to get across-species diffs)



Useful papers (and included references)

Within-species

- Kelly, R. P., A. O. Shelton, and R. Gallego. 2019. Understanding PCR Processes to Draw Meaningful Conclusions from Environmental DNA Studies. *Scientific Reports* 9:12133.
- Luo, M., Y. Ji, D. Warton, and D. W. Yu. 2023. Extracting abundance information from DNA -based data. *Molecular Ecology Resources* 23:174–189.
- Tkacz, A., M. Hortala, and P. S. Poole. 2018. Absolute quantitation of microbiota abundance in environmental samples. *Microbiome* 6:110.
- Diana, A., E. Matechou, J. Griffin, D. W. Yu, M. Luo, M. Tosa, A. Bush, et al. 2022. eDNAPlus: A unifying modelling framework for DNA-based biodiversity monitoring. <http://arxiv.org/abs/2211.12213>

Across-species

- Fukaya, K., H. Murakami, S. Yoon, K. Minami, Y. Osada, S. Yamamoto, R. Masuda, et al. 2021. Estimating fish population abundance by integrating quantitative data on environmental DNA and hydrodynamic modelling. *Molecular Ecology* 30:3057–3067.
- McLaren, M. R., A. D. Willis, and B. J. Callahan. 2019. Consistent and correctable bias in metagenomic sequencing experiments. *eLife* 8:e46923.
- Shelton, A. O., Z. J. Gold, A. J. Jensen, E. D'Agnese, E. Andruszkiewicz Allan, A. Van Cise, R. Gallego, et al. 2023. Toward quantitative metabarcoding. *Ecology* 104:e3906.
- Williamson, B. D., J. P. Hughes, and A. D. Willis. 2021. A multiview model for relative and absolute microbial abundances. *Biometrics* 78:1181–1194.

Useful papers (and included references)

Diana, A., E. Matechou, J. Griffin, D. W. Yu, M. Luo, M. Tosa, A. Bush, et al. 2022. eDNAPlus: **A unifying modelling framework for DNA-based biodiversity monitoring.** <http://arxiv.org/abs/2211.12213>

Stage 1 - biomass collection	
<i>Species effect</i>	Every sample contains a certain amount of DNA biomass of each species, with the amount proportional to the biomass available at the site. However, the proportionality constant is unknown and species-specific, since the DNA of different species can be collected at different rates.
<i>Noise</i>	The amount of biomass collected for each species varies stochastically between samples collected at the same site.
<i>Error</i>	It is possible for the DNA of a target species that is present at a site not to be sampled (false negative error), or traces of DNA from one sample to contaminate another sample (false positive error).
Stage 2 - biomass analysis	
<i>Species effect</i>	As a result of differences in gene copy number, DNA extraction efficiency, and PCR amplification efficiency, the correspondence between the source sample biomass and the number of amplicon reads is species-specific (column of the OTU table).
<i>Pipeline effect</i>	PCR stochasticity and the passing of small aliquots of liquid along the laboratory pipeline affects the total number of reads per technical replicate for all species (row of the OTU table).
<i>Noise</i>	In addition to the species and pipeline effect, there is added noise in the number of reads per OTU and PCR (each cell of the OTU table).
<i>Error</i>	It is possible for the DNA of a target species that is present in the sample not to be amplified in the lab (false negative error), or traces of DNA of one sample to contaminate and be detected in other samples (false positive error), due to the high species-detection power of amplicon sequencing.