

Gut microbiota of Arctic breeding shorebirds

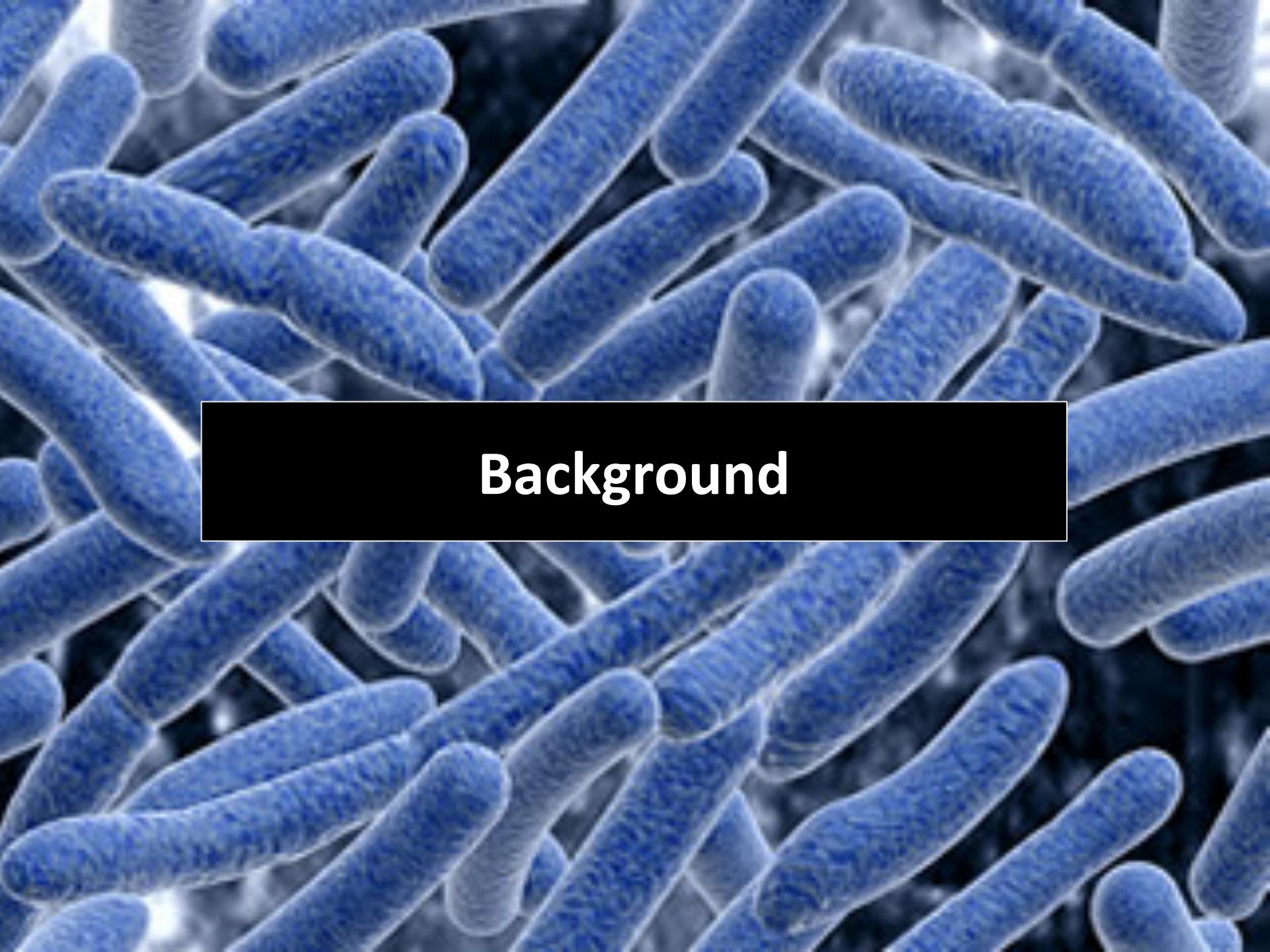


Kirsten Grond, Allison Veach,
Hamida Mahmood, Naomi Ohta, and Denise Case

Outline

- Background
- Research Question
- Experimental Design
- Methods & Analyses pipeline
 - Local Alignment
 - Multiple Sequence Alignment
 - Classification & OTU Clustering
 - OTU Classify
- Data visualization & Results
- Future Work





Background

Why gut microbiota?

Gut microbiota important in health (Kohl 2012)

- Nutritional uptake
- Detoxification
- Interactions with immune system

Humans: Unbalanced gut microbiota → Obesity,
Diabetes, Gut inflammation (IBS), Depression

- (Bonfrate et al. 2013; Dinan & Cryan 2013; Fang & Evans 2013;
Le Chatelier et al. 2013)

Why shorebirds?



Dunlin *Calidris alpina*

- Sensitive to change
- “Living on the edge”
- Many Arctic breeding species declining fast

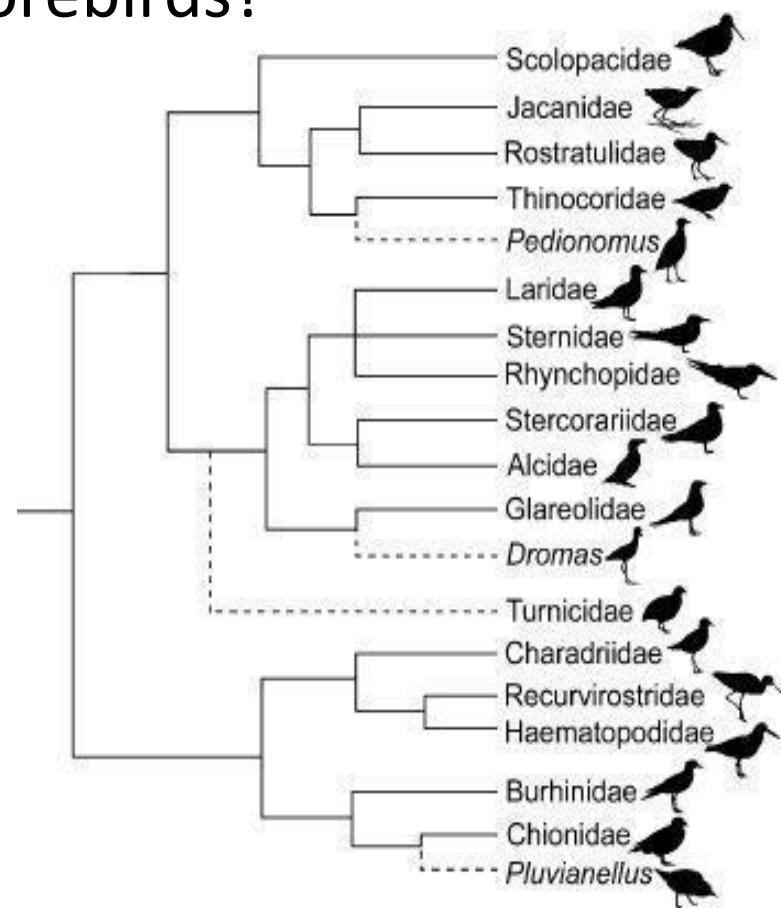


Research Question

Do *intrinsic* or *extrinsic* factors determine gut microbiome composition in shorebirds?

Intrinsic → Host species phylogeny

Extrinsic → Environment
(breeding site, migration route,
interaction with conspecifics etc.)

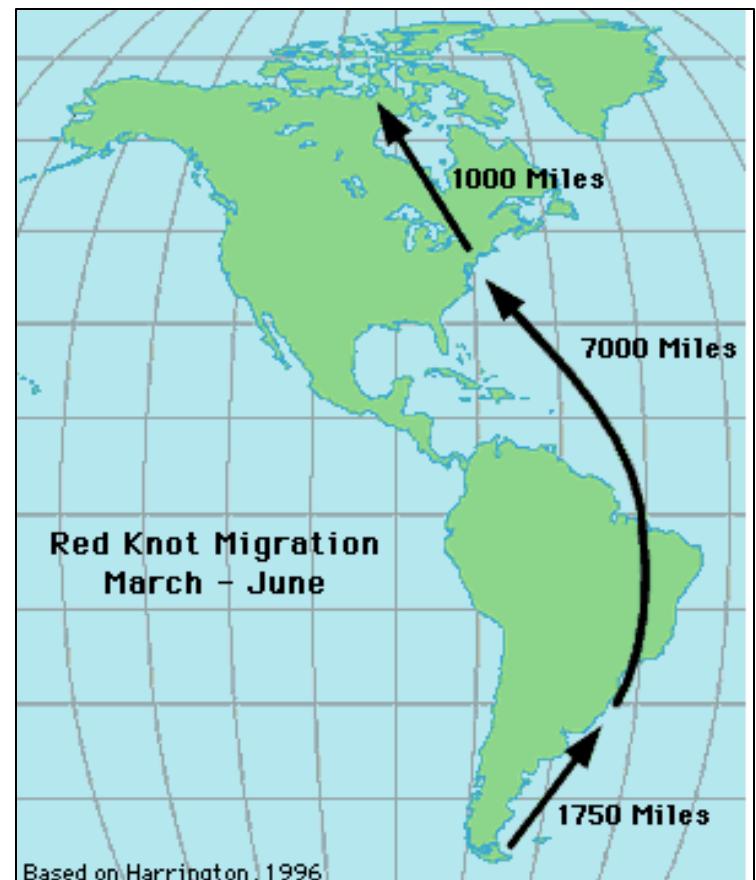


Research Questions

Do *intrinsic* or *extrinsic* factors determine gut microbiome composition in shorebirds?

Intrinsic → Host species phylogeny

Extrinsic → Environment
(breeding site, migration route,
interaction with conspecifics etc.)





Experimental Design & Software

Sample collection

538 fecal samples of 10 shorebird species from 9 Arctic breeding sites

Selected a subset of 138 samples, all from Dunlin to speed up analysis

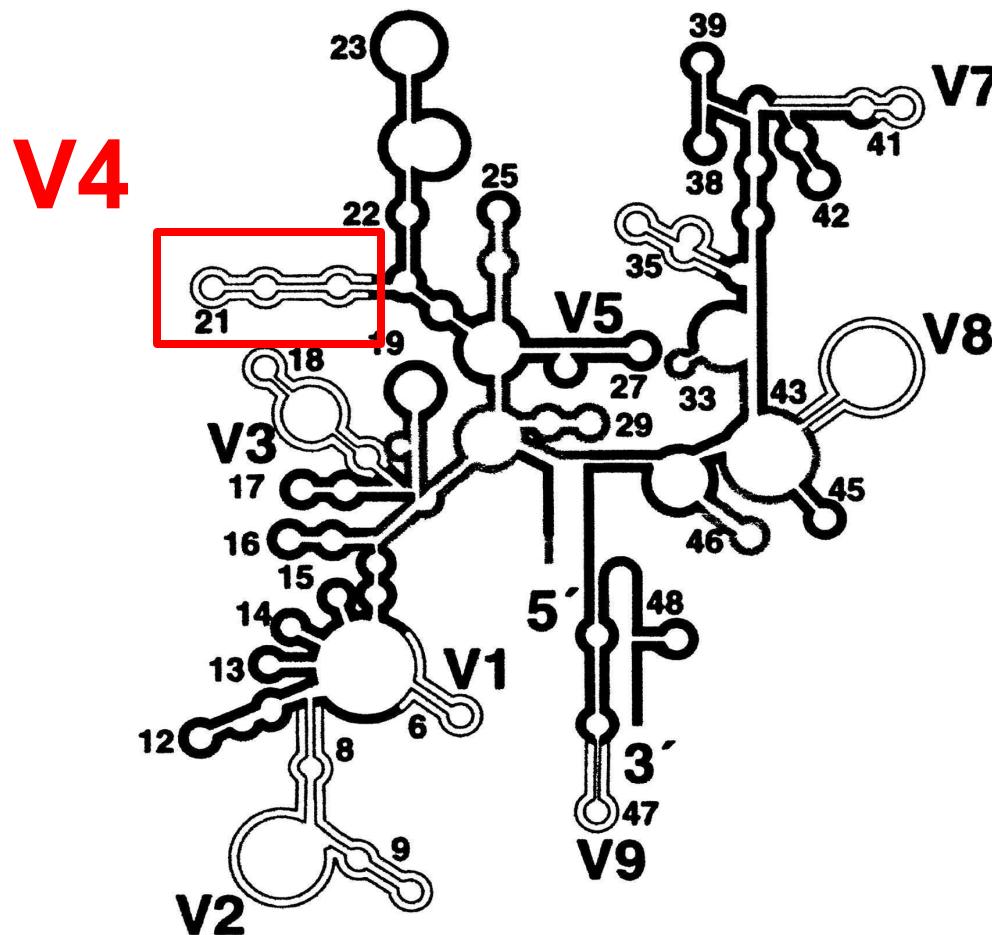
- Making contigs for 538 samples took >200 hrs



Used the Illumina MiSeq platform to sequence the V4 region of the 16S ribosomal RNA



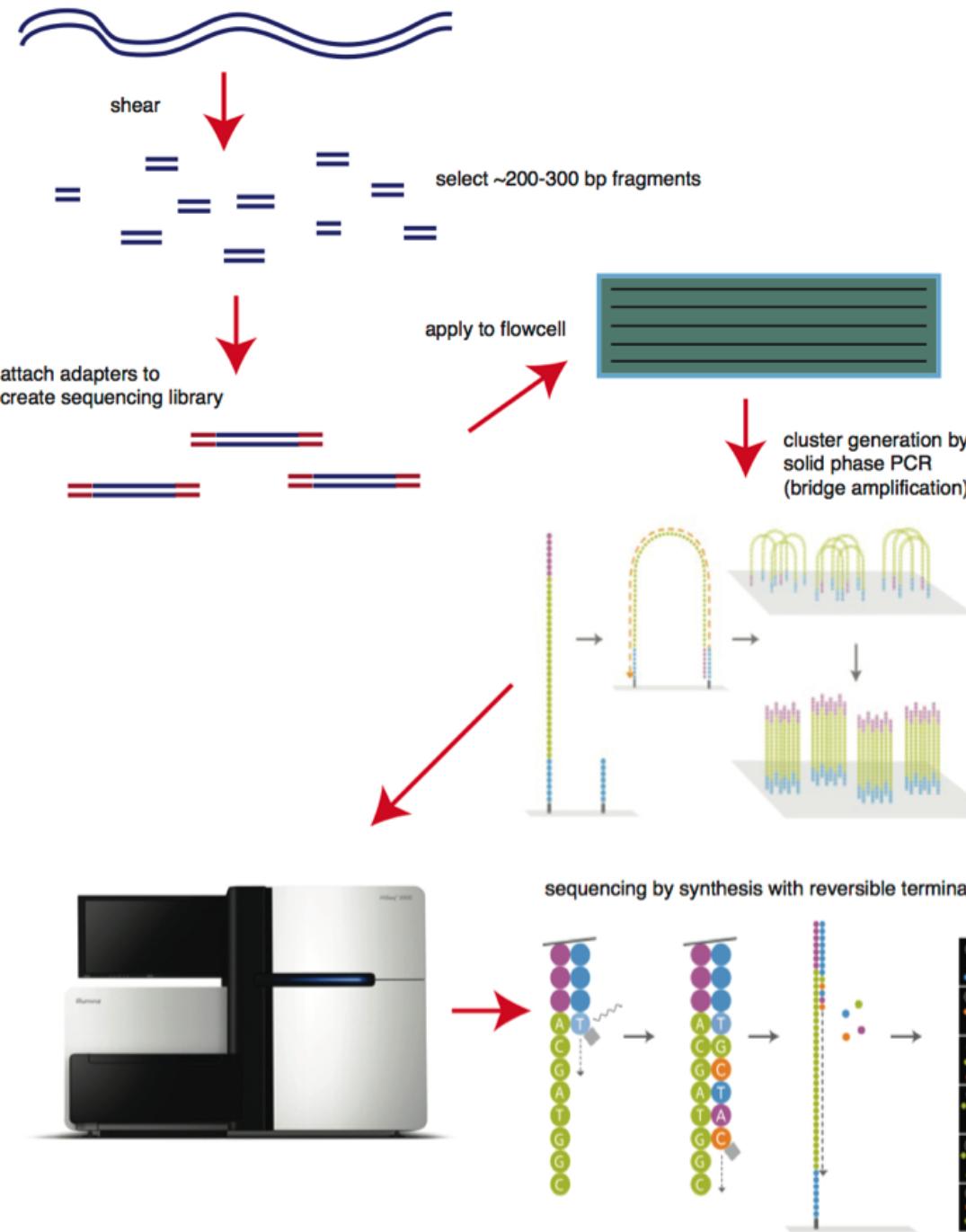
16S rDNA (double lines indicate variable or hypervariable; gray lines indicate highly conserved; V1 to V9 indicate major variable regions).



Tortoli E Clin. Microbiol. Rev. 2003;16:319-354

Clinical Microbiology Reviews

Illumina MiSeq



Genomic DNA

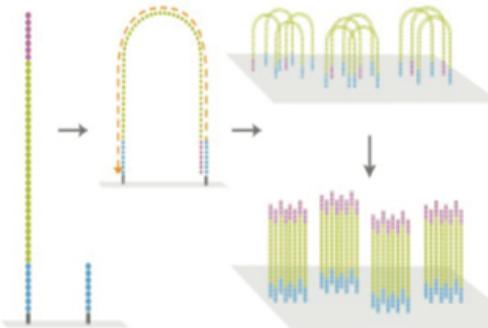
shear



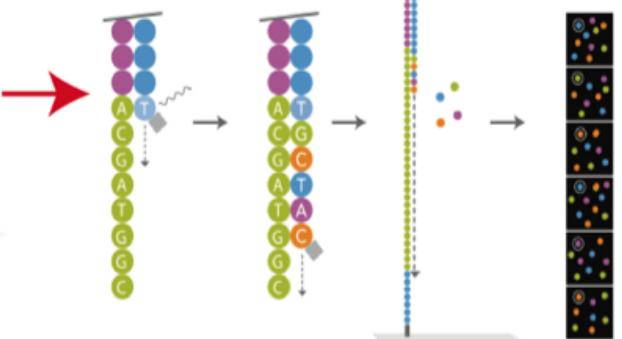
cluster generation by
solid phase PCR
(bridge amplification)

Illumina MiSeq

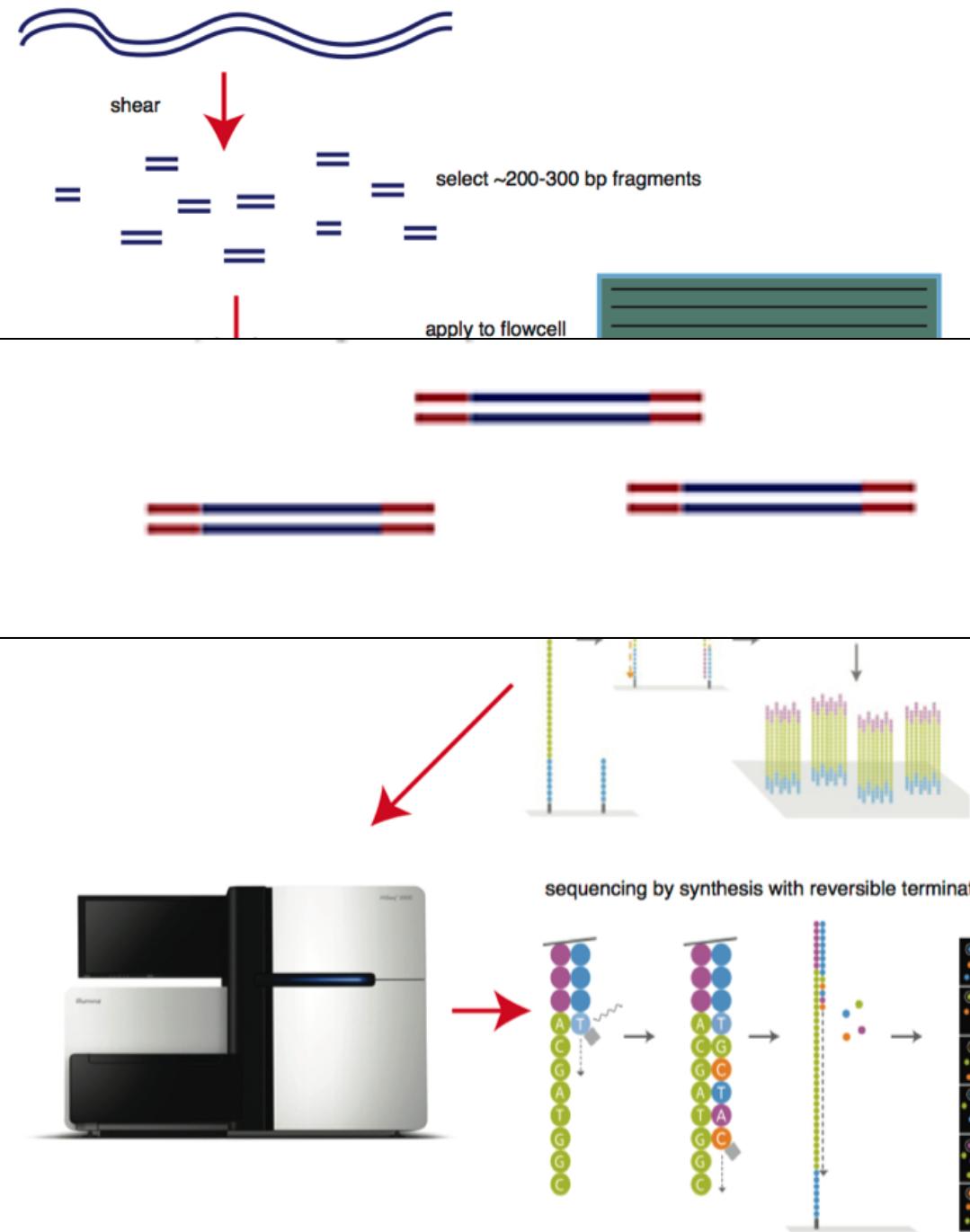
Short fragments (V4)
291 bp



sequencing by synthesis with reversible terminators

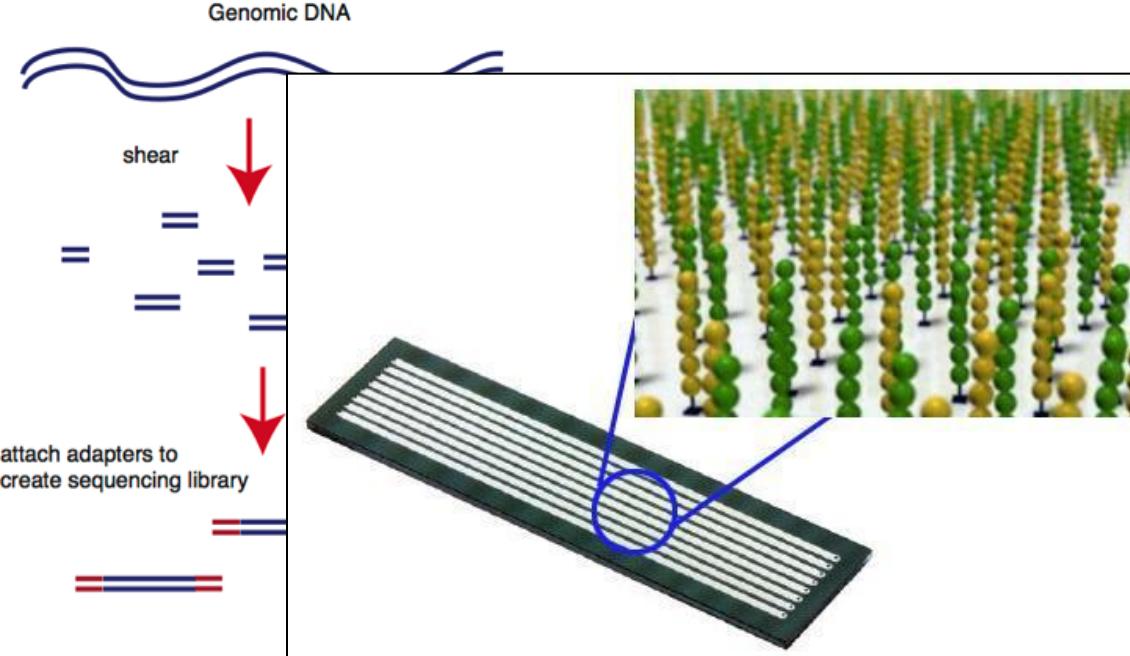


Illumina MiSeq



Short fragments (V4)
~250 bp

Attach adapters for sequencing

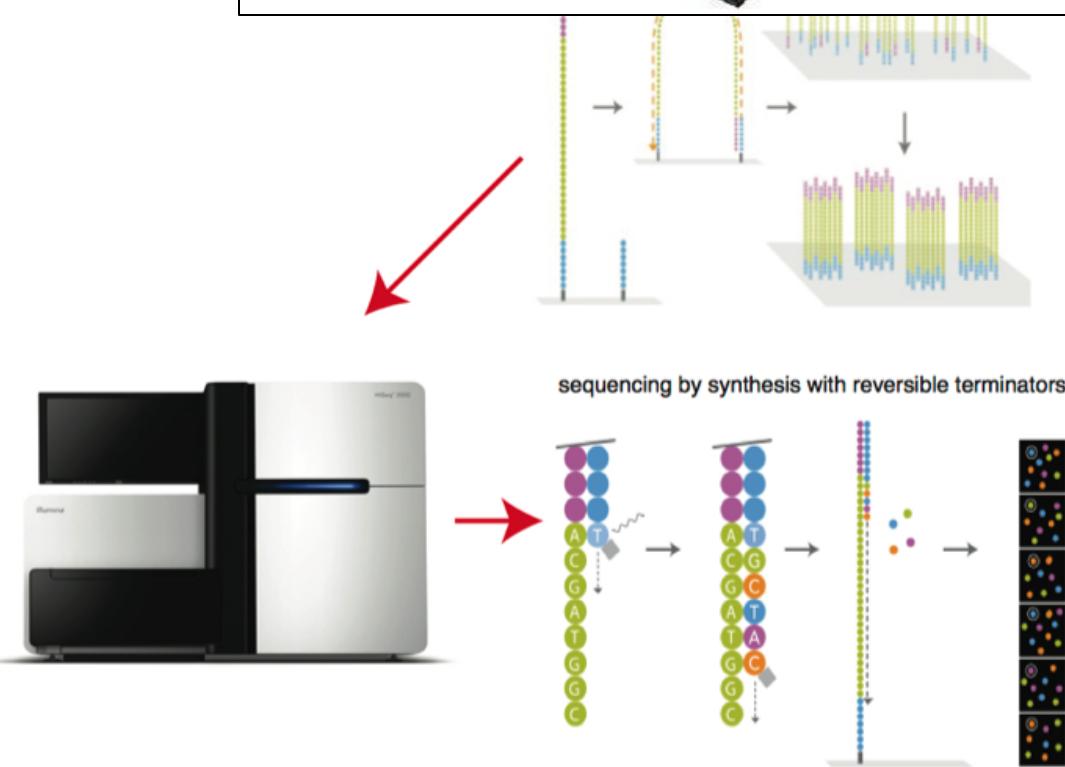


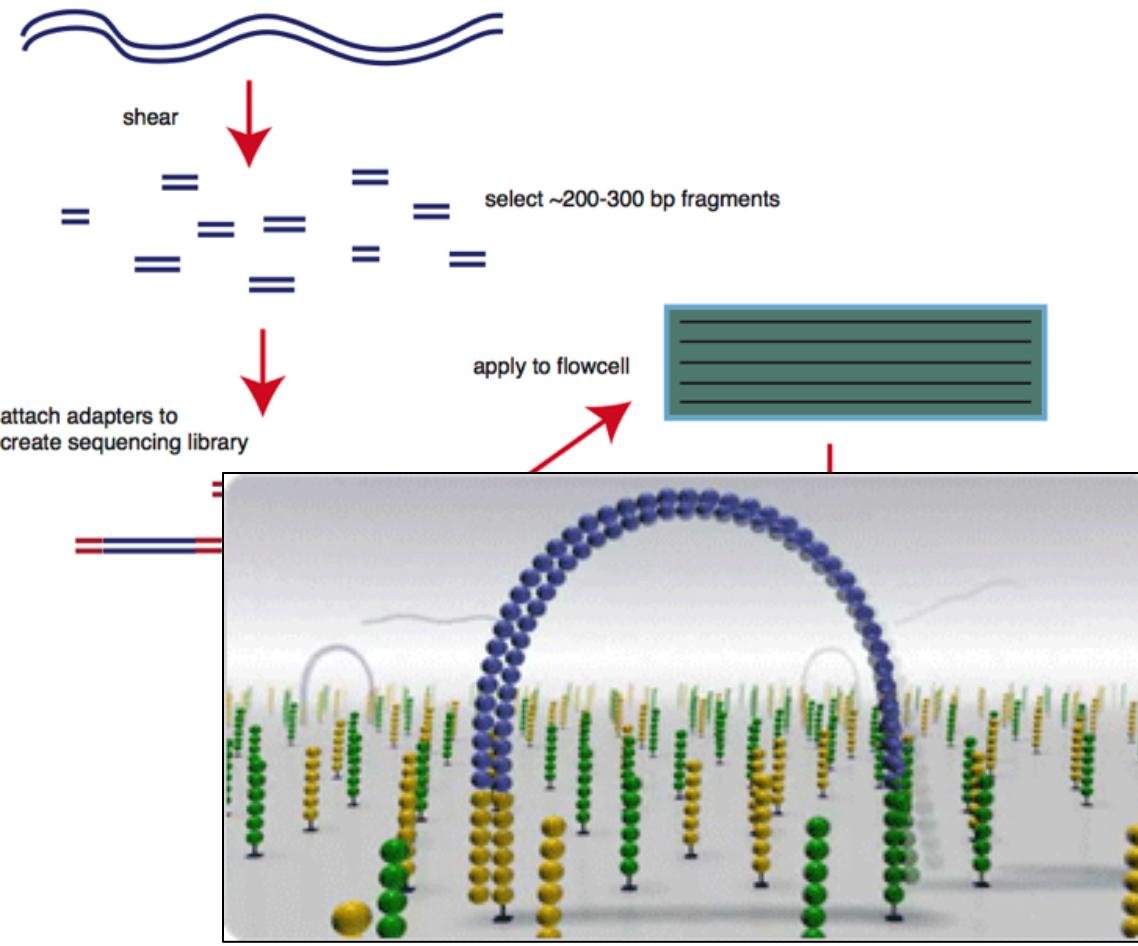
Illumina MiSeq

Short fragments (V4)
~250 bp

Attach adapters for sequencing

Apply to flow cell with attached primer





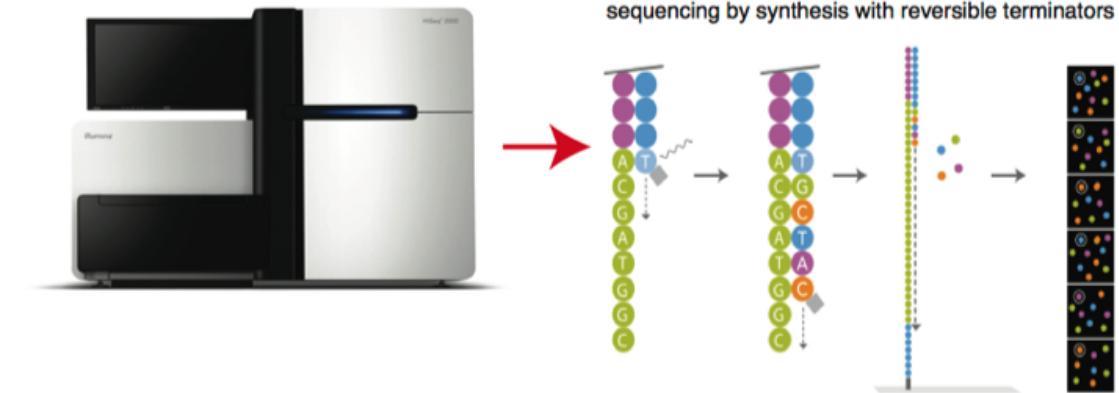
Illumina MiSeq

Short fragments (V4)

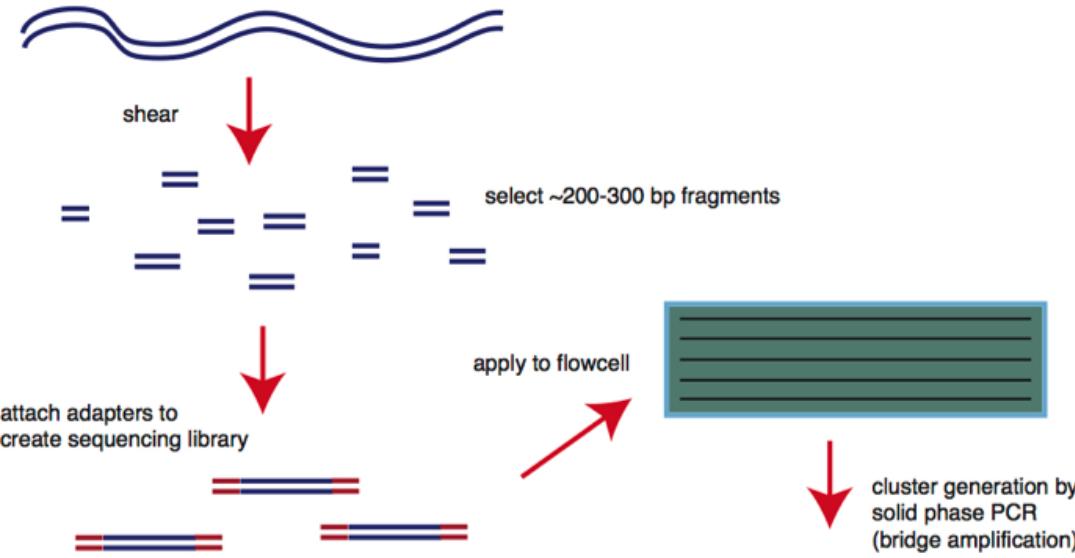
~250 bp

Attach adapters for sequencing

Apply to flow cell with primers



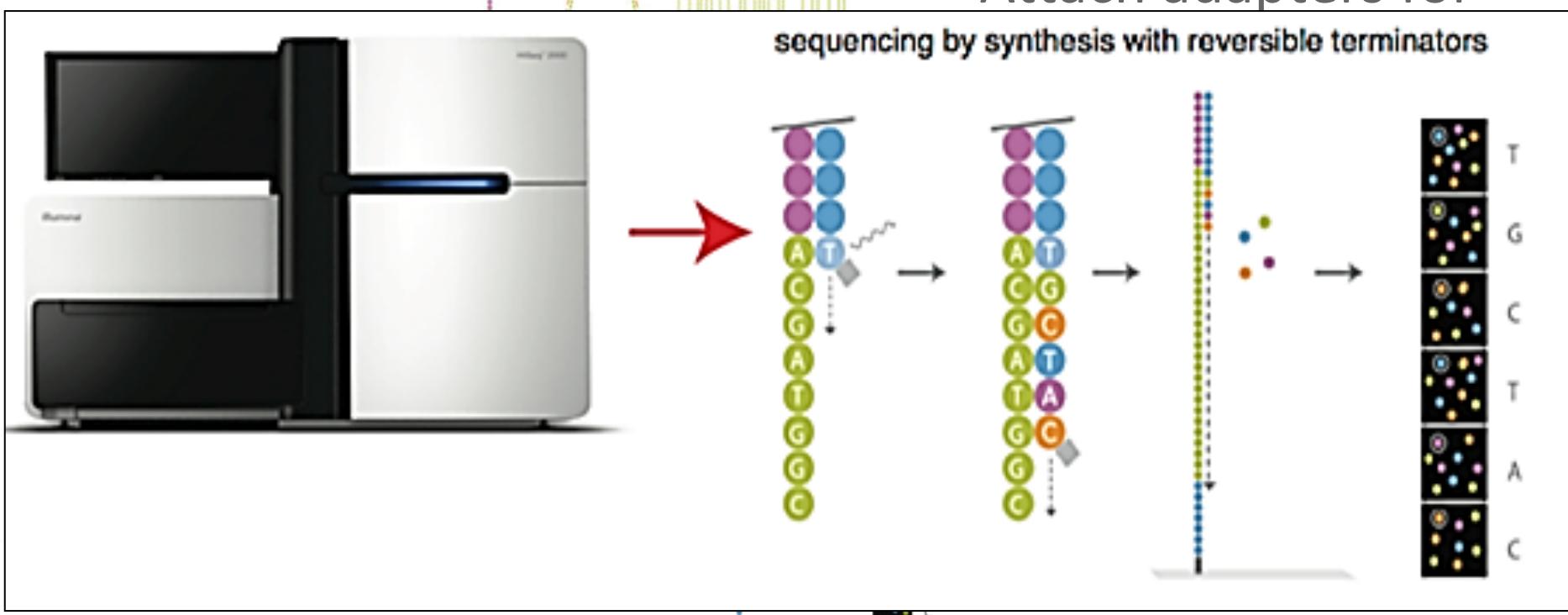
Illumina MiSeq



Short fragments (V4)
~250 bp

Attach adapters for

sequencing by synthesis with reversible terminators



Welcome to the website for the mothur project, initiated by [Dr. Patrick Schloss](#) and his software development team in the [Department of Microbiology & Immunology](#) at [The University of Michigan](#). This project seeks to develop a single piece of open-source, expandable software to fill the bioinformatics needs of the microbial ecology community. In February 2009 we released the first version of mothur, which had accelerated versions of the popular DOTUR and SONS programs. Since then we have added the functionality of a number of other popular tools including s-libshuff, TreeClimber (i.e. the parsimony test), UniFrac, distance calculation, visualization tools, like to contribute or make your own imports, the programming environment would like to add, monthly.



Sequence processing

mothur provides a number of **sequence processing** commands to go from Sanger sequences or pyrosequences to a distance matrix. It also includes a fast and flexible sequence analysis pipeline to enable you to carry out OTU-based approaches and hypothesis testing approaches. There are many pages, but several users have provided several analysis examples, which use these commands. An exhaustive list of the commands can be found in the index.

General commands

- `reverse.seqs` - output the reverse complement of a sequence collection
- `summary.seqs` - summarize a collection of sequences
- `summary.qual` - summarize a collection of sequences quality data
- `merge.files` - merge two or more sets of files
- `merge.sfffiles` - merge two or more sets of sfffiles
- `merge.taxsummary` - combines tax summary files.
- `list.seqs` - write the sequence names contained within a file to a new file
- `get.seqs` - write the data for a sequence name to a new file
- `remove.seqs` - remove the data for a sequence name from a new file
- `remove.groups` - remove sequence from a specific group or set of groups
- `get.groups` - select sequence from a specific group or set of groups

Mothur in Beocat

Created shell files with Mothur commands in batch

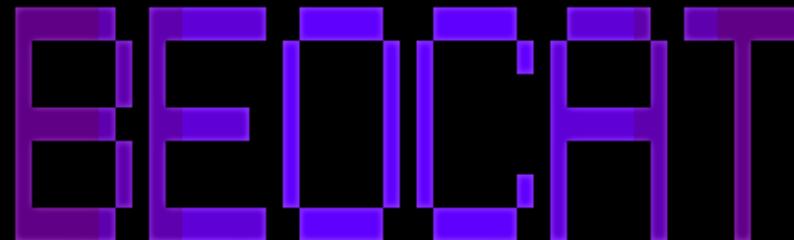
```
1 #!/bin/bash
2
3 /homes/kgrond/Shorebird_seqs/mothur/mothur "#make.contigs(file=/homes/kgrond/
summary.seqs(fasta=/homes/kgrond/Shorebird_seqs/mothur/dunltrim.contigs.fast
group=/homes/kgrond/Shorebird_seqs/mothur/dunlcontigs.groups, minlength=250,
summary.seqs(fasta=/homes/kgrond/Shorebird_seqs/mothur/dunltrim.contigs.good.
unique.seqs(fasta=/homes/kgrond/Shorebird_seqs/mothur/dunltrim.contigs.good.
summary.seqs(fasta=/homes/kgrond/Shorebird_seqs/mothur/dunltrim.contigs.good.
name=/homes/kgrond/Shorebird_seqs/mothur/dunltrim.contigs.good.trim.names);
align.seqs(fasta=/homes/kgrond/Shorebird_seqs/mothur/dunltrim.contigs.good.t
reference=/homes/kgrond/Shorebird_seqs/mothur/silva.bacteria.fasta, flip=t,
summary.seqs(fasta=/homes/kgrond/Shorebird_seqs/mothur/dunltrim.contigs.good.
```



kgrond@athena:~/Shorebird_seqs/mothur



```
Using username "Kgrond".  
Using keyboard-interactive authentication.  
Password:
```



Welcome to Beocat. Documentation is at <http://support.beocat.cis.ksu.edu>

If you require assistance, please e-mail beocat@cis.ksu.edu

The headnodes are for **light** prep work only. All resource intensive applications should be submitted to the job scheduler. Your use of this system indicates your agreement to comply with all technology usage policies related to Beocat and KSU. <http://beocat.cis.ksu.edu/beocat/documentation/policy>

Beocat is **not** to be used with classified or PII/HIPAA data.

```
kgrond@athena ~ $ cd Shorebird_seqs/  
kgrond@athena ~/Shorebird_seqs $ cd mothur/  
kgrond@athena ~/Shorebird_seqs/mothur $ qsub -l mem=3G -l h_rt=100:00:00 dun11.sh
```

```
kgrond@athena ~ $ cd Shorebird_seqs/  
kgrond@athena ~/Shorebird_seqs $ cd mothur/  
kgrond@athena ~/Shorebird_seqs/mothur $ qsub -l mem=3G -l h_rt=100:00:00 dun11.sh
```

Pipeline

Local alignment

make.contigs, summary.seqs, screen.seqs, trim.seqs



Multiple alignment

unique.seqs, summary.seqs, align.seqs



Classification & OTU clustering

pre.cluster, chimera.uchime, classify.seqs, remove.lineage, cluster



OTU Classify

remove.rare, classify.otus, remove.lineage



Methods:
Local Alignment & Cleanup

Local Alignment

- Making contigs -

```
make.contigs(file=/homes/kgrond/Shorebird_seqs/mothur/allfecal)
```

- Combine forward and reverse fastq files and make contigs to reduce PCR errors.

avian3AB1	/homes/kgrond/Shorebird_seqs/mothur/avian3AB1_S1_L001_R1_001.fastq
avian3AB2	/homes/kgrond/Shorebird_seqs/mothur/avian3AB2_S2_L001_R1_001.fastq
avian3AB3	/homes/kgrond/Shorebird_seqs/mothur/avian3AB3_S3_L001_R1_001.fastq
avian3AB4	/homes/kgrond/Shorebird_seqs/mothur/avian3AB4_S4_L001_R1_001.fastq
avian3AB5	/homes/kgrond/Shorebird_seqs/mothur/avian3AB5_S5_L001_R1_001.fastq

Sample ID



Forward fastq

/homes/kgrond/Shorebird_seqs/mothur/avian3AB1_S1_L001_R2_001.fastq
/homes/kgrond/Shorebird_seqs/mothur/avian3AB2_S2_L001_R2_001.fastq
/homes/kgrond/Shorebird_seqs/mothur/avian3AB3_S3_L001_R2_001.fastq
/homes/kgrond/Shorebird_seqs/mothur/avian3AB4_S4_L001_R2_001.fastq
/homes/kgrond/Shorebird_seqs/mothur/avian3AB5_S5_L001_R2_001.fastq

Reverse fastq



Local Alignment

- Make contigs. algorithm -

`make.contigs(file=/homes/kgrond/Shorebird_seqs/mothur/allfecal)`

- *Uses the Needleman algorithm for local alignment of paired end sequences*

In case of...

Gap: the quality score of the base must be over 25 to be considered real.

Mismatch: require one of the bases to have a quality score 6 or more points better than the other, otherwise sequence will shown as “N”.

Local Alignment

- summary seqs -

Summary.seqs(fasta=/homes/kgrodn/Shorebird_seqs/mothur/allfecal)

	Start	End	NBases	Ambigs	Polymer	NumSeqs
Minimum:	1	228	228	0	2	1
2.5%-tile:	1	251	251	0	4	458,899
25%-tile:	1	252	252	0	4	4,588,981
Median:	1	253	253	0	4	9,177,962
75%-tile:	1	253	253	2	5	13,766,942
97.5%-tile:	1	373	373	17	6	17,897,024
Maximum:	1	503	502	292	251	18,355,922
Mean:	1	258.67	258.67	2.2006	4.46568	
# of Seqs:	18,355,922					

Most reads vary between 228 and 373 bases.

Cleanup the sequences

- Screen.seqs -

```
Screen.seqs(fasta=/homes/kgrond/Shorebird_seqs/mothur/allfecal.trim.contigs.  
fasta,group=/homes/kgrond/Shorebird_seqs/mothur/allfecal.contigs.groups,  
minlength=250, maxlenlength=350, maxambig=0)
```

- Length range between **250 – 350**
- Sequences with ambiguous based are not retained



Summary.seqs

After each step, we run “Summary.seqs” to check if the previous command has worked.

Cleanup the sequences

- Trim.seqs -

```
Trim.seqs(fasta=/homes/kgrond/Shorebird_seqs/mothur/allfecal,  
maxhomop=8, flip=T)
```

- Any sequences with homopolymers (-ATCCCCCCCC) are removed.
- Take the reverse complement of PCR primer.
“flip=T(rue)



**Multiple Alignment, Clean up,
Sequence Classification**

Multiple Alignment, Sequence Cleanup and Classification

Unique.seqs(fasta=/homes/kgrond/Shorebird_seqs/mothur/allfecal.trim.contigs.good.fasta)

- Analysis uses only representative sequences to reduce computing time

Summary.seqs

**(fasta=/homes/kgrond/Shorebird_seqs/mothur/allfecal.trim.contigs.good.trim.unique.fasta,
name=/homes/kgrond/Shorebird_seqs/mothur/allfecal.trim.contigs.good.trim.names)**

Multiple Alignment, Sequence Cleanup and Classification

```
Align.seqs(fasta=homes/kgrond/Shorebird_seqs/mothur/allfecal.  
trim.contigs.good.trim.unique.fasta ,  
reference=/homes/kgrond/Shorebird_seqs/mothur/ silva.bacteria.  
fasta)
```

- Template DB is SILVA containing aligned 16S rRNA sequences



Multiple Alignment, Sequence Cleanup and Classification

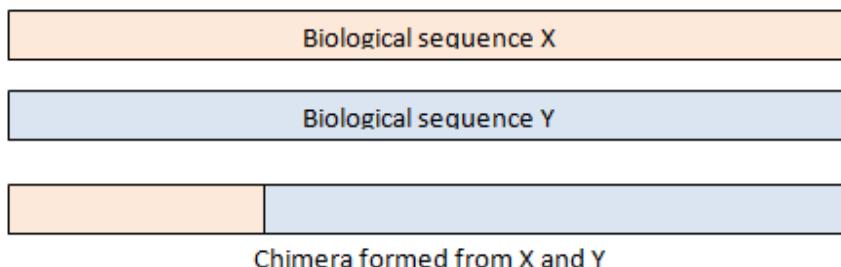
```
Pre.cluster(fasta=/homes/kgrond/Shorebird_seqs/mothur/allfecal.  
trim.contigs.good.trim.unique.good.align,  
name=/homes/kgrond/Shorebird_seqs/mothur/allfecal.trim.contigs.  
good.trim.good.names,  
group=/homes/kgrond/Shorebird_seqs/mothur/allfecal.contigs.good.  
good.groups, diffs=2)
```

- Sequences abund ranked then sequences within ≤ 2 mismatches are merged
- Reduces computing time and accounts for sequenced generated artifacts

Multiple Alignment, Sequence Cleanup and Classification

chimera.uchime

(fasta=/homes/kgrond/Shorebird_seqs/mothur/allfecal.trim.contigs.
good.trim.unique.good.precluster.align,
name=/homes/kgrond/Shorebird_seqs/mothur/allfecal.trim.contigs.
good.trim.unique.good.precluster.names)
Remove chimeric sequences

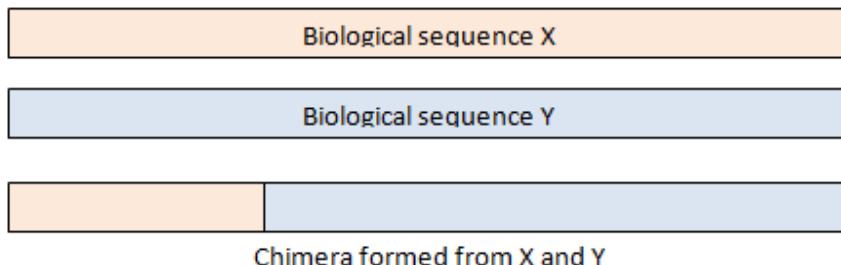


Picture credit: http://drive5.com/usearch/manual/chimera_formation.html

Multiple Alignment, Sequence Cleanup and Classification

remove.seqs

```
(accnos=/homes/kgrond/Shorebird_seqs/mothur/allfecal.trim.  
contigs.good.trim.unique.good.precluster.uchime.accnos,  
fasta=/homes/kgrond/Shorebird_seqs/mothur/allfecal.trim.contigs.  
good.trim.unique.good.precluster.align,  
name=/homes/kgrond/Shorebird_seqs/mothur/allfecal.trim.contigs.  
good.trim.unique.good.precluster.names, dups=T)
```



Picture credit: http://drive5.com/usearch/manual/chimera_formation.html

Multiple Alignment, Sequence Cleanup and Classification

```
classify.seqs(fasta=/homes/kgrond/Shorebird_seqs/mothur/allfecal.  
trim.contigs.good.trim.unique.good.precluster.pick.align,  
name=/homes/kgrond/Shorebird_seqs/mothur/allfecal.trim.contigs.  
good.trim.unique.good.precluster.pick.names,  
template=/homes/kgrond/Shorebird_seqs/mothur/trainset9_032012  
.rdp.fasta,  
taxonomy=/homes/kgrond/Shorebird_seqs/mothur/trainset9_03201  
2.rdp.tax, cutoff=80)
```

- Template database used is the RDP training set 9
- Taxonomy is also derived from RDP DB
- Cutoff retains only taxonomic affiliations with bootstrap support at 80%

Multiple Alignment, Sequence Cleanup and Classification

```
remove.lineage(fasta=/homes/kgrond/Shorebird_seqs/mothur/allfecal.trim.  
contigs.good.trim.unique.good.precluster.pick.align,  
name=/homes/kgrond/Shorebird_seqs/mothur/allfecal.trim.contigs.good.trim.  
unique.good.precluster.pick.names,  
taxonomy=/homes/kgrond/Shorebird_seqs/mothur/allfecal.trim.contigs.good.trim.  
unique.good.precluster.pick.rdp.wang.taxonomy, taxon=unknown-Bacteria;  
unclassified-Bacteria;Cyanobacteria_Chloroplast;unclassified-Bacteria;  
Cyanobacteria_Chloroplast;Chloroplast-Archaea-Mitochondria-Eukaryota)
```

- Template database used is the RDP training set 9
- Taxonomy is also derived from RDP DB
- Cutoff retains only taxonomic affiliations with bootstrap support at 80%

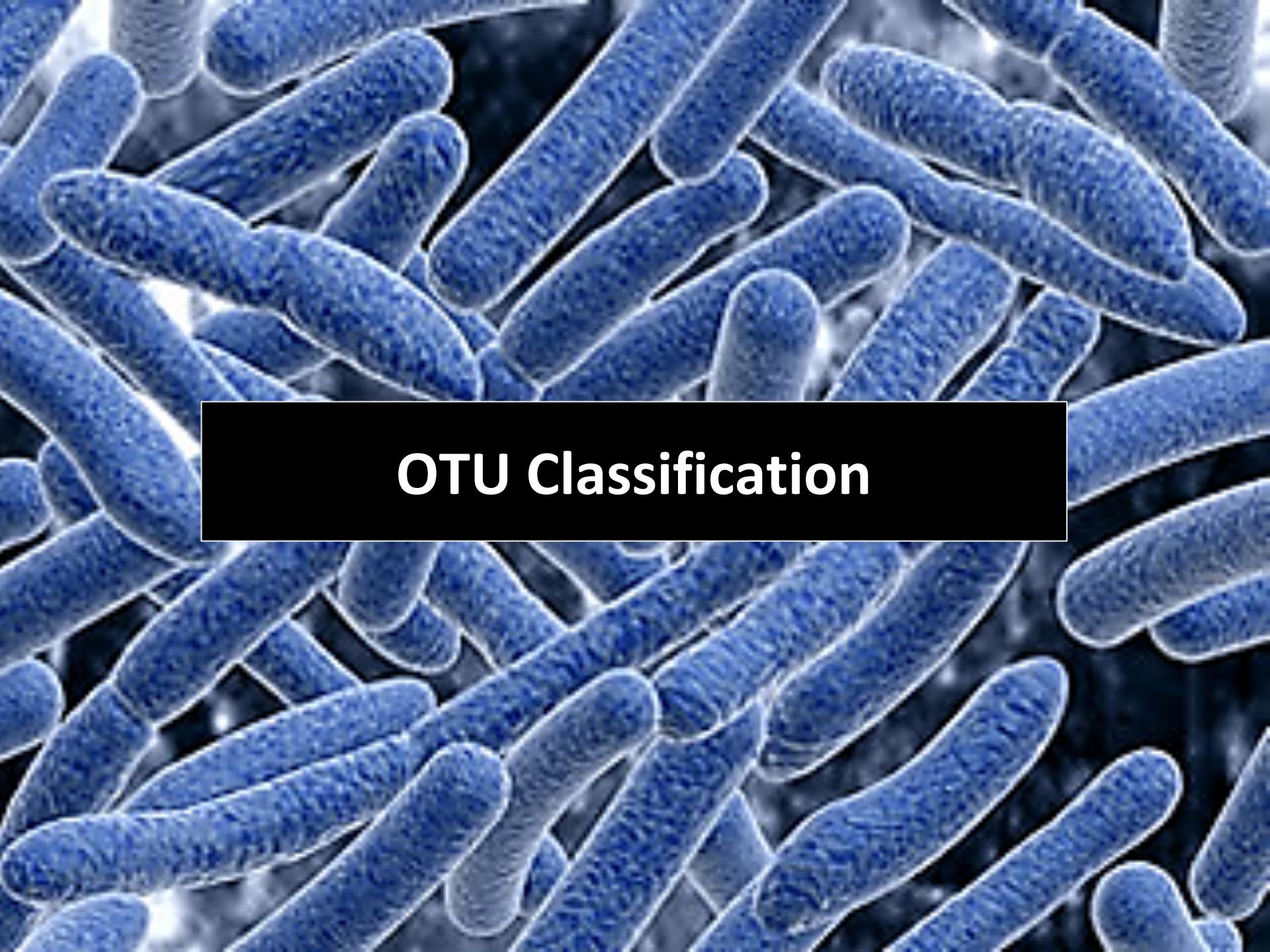
Multiple Alignment, Sequence Cleanup and Classification

Sequence name



```
AY457915 Bacteria(100);Firmicutes(100);Clostridiales(100);Johnsonella_et_rel.  
(100);Johnsonella_et_rel.(100);Johnsonella_et_rel.(100);  
Eubacterium_eligens_et_rel.(100);Lachnospira_pectinoschiza(100);unclassified;  
unclassified;unclassified;unclassified;
```

100% bootstrap support for Domain Bacteria, Phylum Firmicutes, Order Clostridiales... Genus



A scanning electron micrograph showing a dense, overlapping cluster of numerous blue-colored, rod-shaped bacteria. The bacteria have a slightly textured surface and are set against a dark, almost black, background which provides a strong contrast.

OTU Classification

OTU Clustering and Classification

Dist.seqs(fasta=final.fasta, cutoff=0.10)

Cluster

**(column=/homes/kgrond/Shorebird_seqs/mothur/name=final.dist,
name= /homes/kgrond/Shorebird_seqs/mothur/)**

- Dist.seqs calculated uncorrected pairwise distances between aligned and classified sequences
- Using a nearest neighbor clustering method, distances are used to create OTU's at 97% similarity threshold

OTU Clustering and Classification

```
Classify.otu(list= /homes/kgrond/Shorebird_seqs/mothur/ final.an.  
list, name /homes/kgrond/Shorebird_seqs/mothur/ final.names,  
taxonomy /homes/kgrond/Shorebird_seqs/mothur/ final.  
taxonomy, label=0.03)
```

- Using taxonomy after removing unwanted sequences and rare OTU's, the updated RDP reference is used to classify each OTU to lowest taxonomic affiliation with acceptable bootstrap support



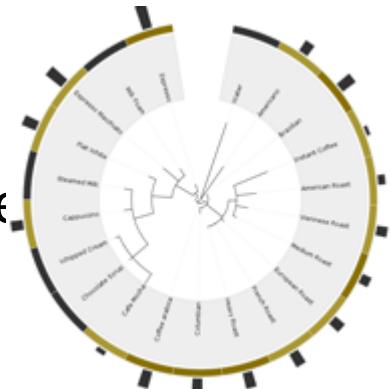
Data Visualization

Data Visualization

Ecological studies involve hundreds of samples and large result sets. The ability to visualize large quantities of information helps scientists evaluate and understand results.

Tools for working with large data sets included:

- **Excel** for the generation of pivot tables and charts that automatically group data into sets for display. (Sample data only as we hit the limit at 1 million rows x 16k columns - not enough for our results.)
- **D3** for creating dynamic displays in the browser using only data and text files. (Powerful SVG graphics in the new HTML5 canvas. No development environment required.)



Content displayed with various GIS options.

Shore Bird Microbiota Res × people.cis.ksu.edu/~dmcase/bird_microbiota/sites.html#

Shore Bird Microbiota

Analysis Charts About Search

Go to Mothur

Samples obtained:
192/542 or 4%
from
Site 7: Mackenzie River Delta, Canada

GINA Best Data Layer (BDL)
Panchromatic Landsat
NOAA Nautical Charts
SDMI Orthomosaic (RGB)
Open Street Maps
Shaded Relief
USGS Topographic Charts

Sites Bird Samples

Bylot Island

0 1 5 female male u (blank)

female 21.00 40% 11%

Interactive charts.

Side bars with site data

Shore Bird Microbiota Res

← → C ⌂ people.cis.ksu.edu/~dmcase/bird_microbiota/sites.html#

_apps

G

G

C

25

g

E

dz

BEO
CAT

mothur

g+

g+

A

g+ P

KSOL

DC

ksuLib

H

DC

AstroBio

»

Other bookmarks

Shore Bird Microbiota

Analysis

Charts

About

Search



Bird Species by Site

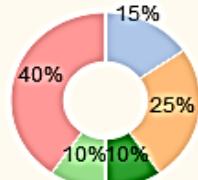
1. Barrow

Legend:
amgp duml lbdo pesa repb
ruth sesa wesa wrsa

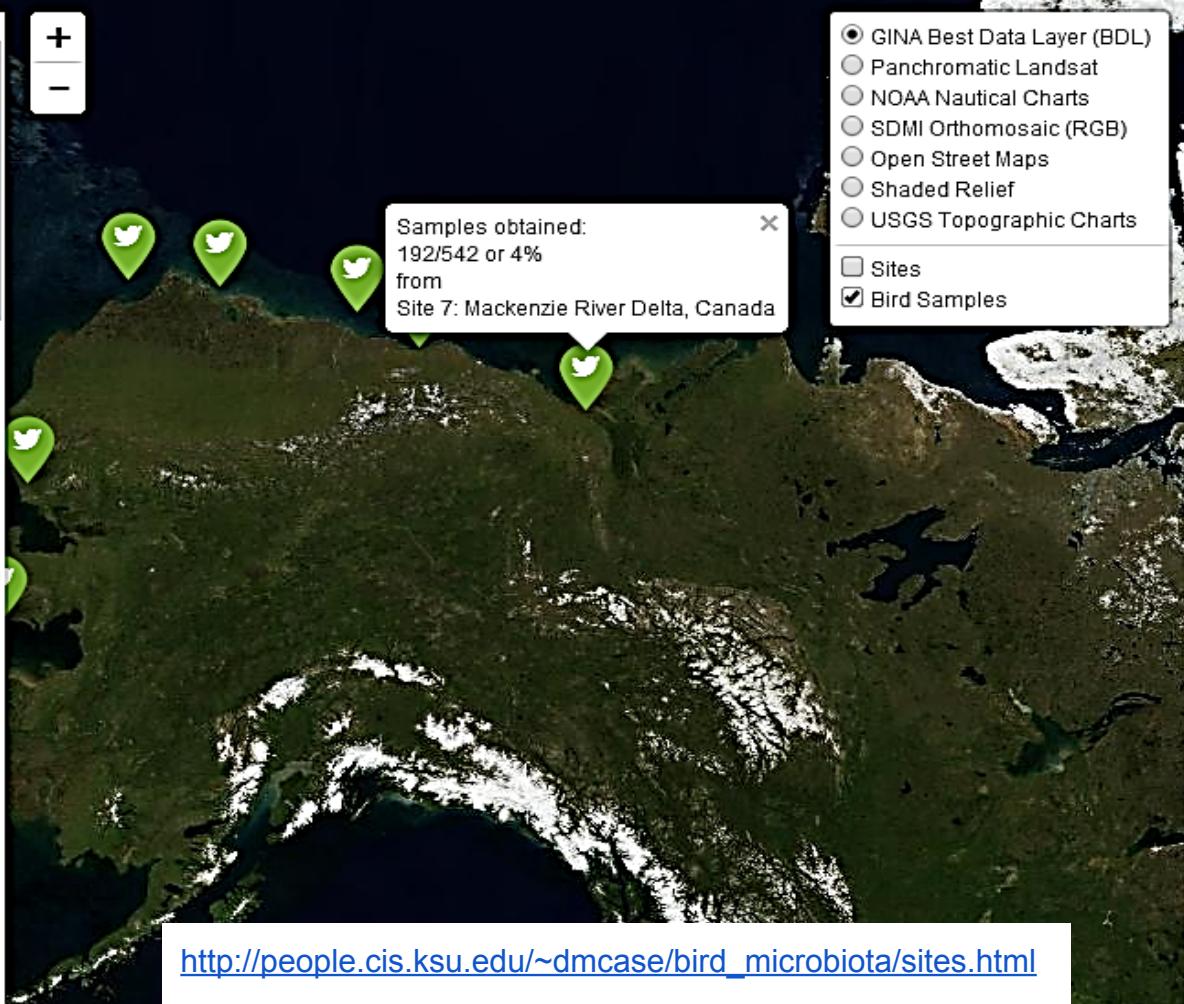


2. Nome

Legend:
amgp duml lbdo pesa repb
ruth sesa wesa wrsa



3. Cape Krusenstern



http://people.cis.ksu.edu/~dmcase/bird_microbiota/sites.html

https://github.com/ksucase/bird_microbiota

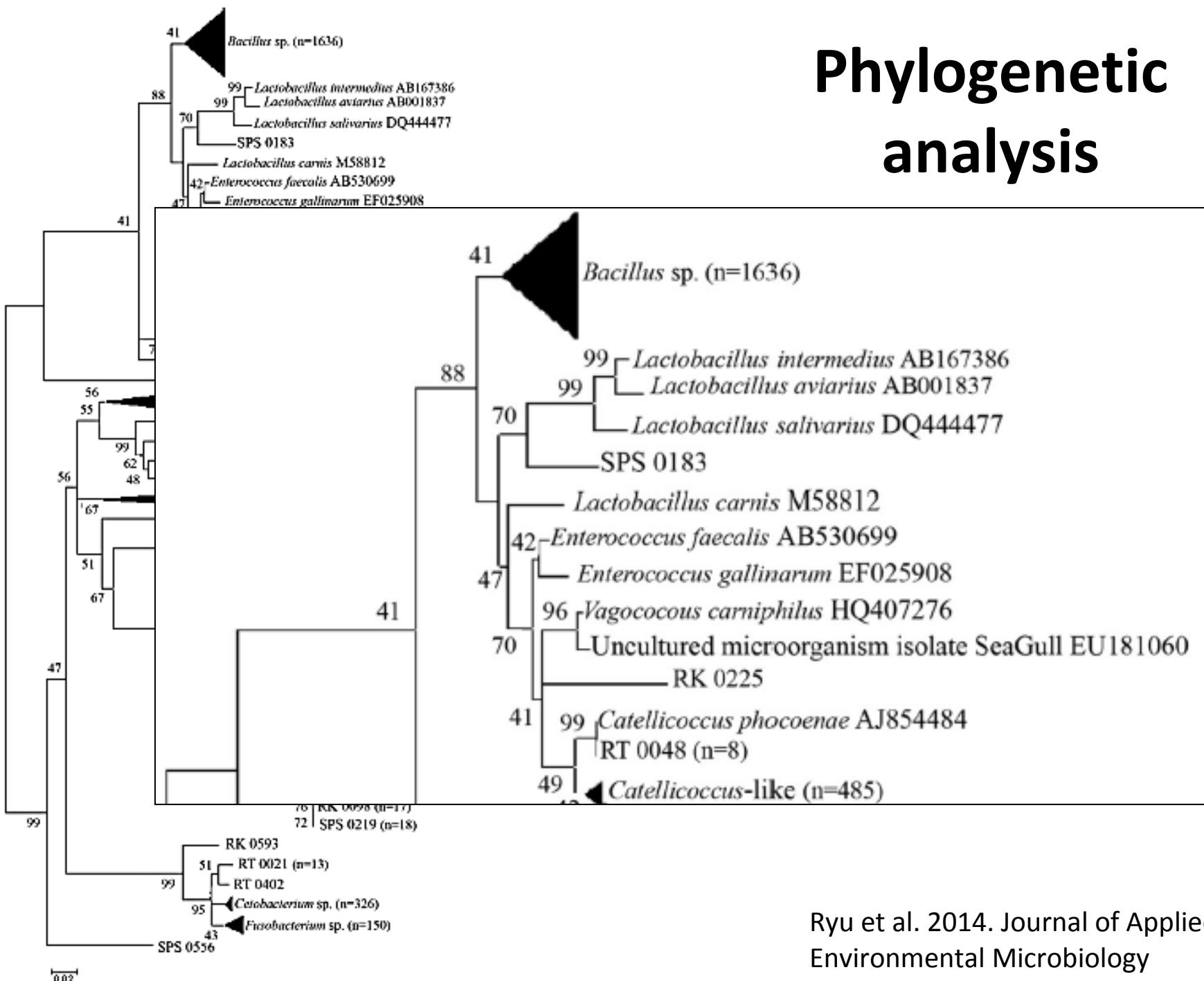
Leaflet

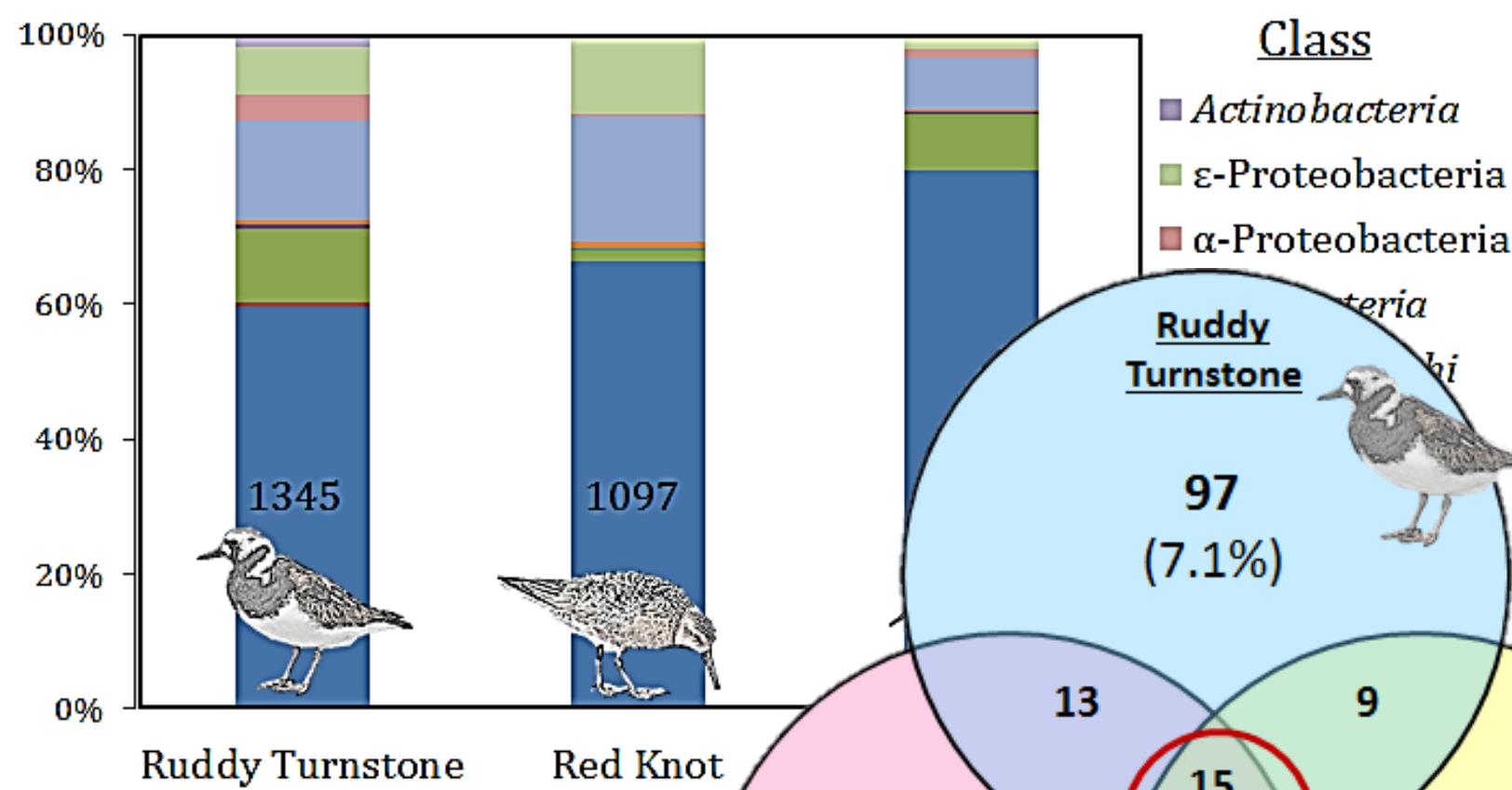
Results

Group	Otu000001	Otu000002	Otu000003	Otu000004
D16_S1_R1	957	390	361	592
D16_S1_R2	877	305	330	653
D16_S1_R3	1242	341	376	697
D16_S2_R1	656	457	286	417
D16_S2_R2	552	406	171	265
D16_S3_R1	1649	741	302	580
D16_S3_R2	2226	775	353	602

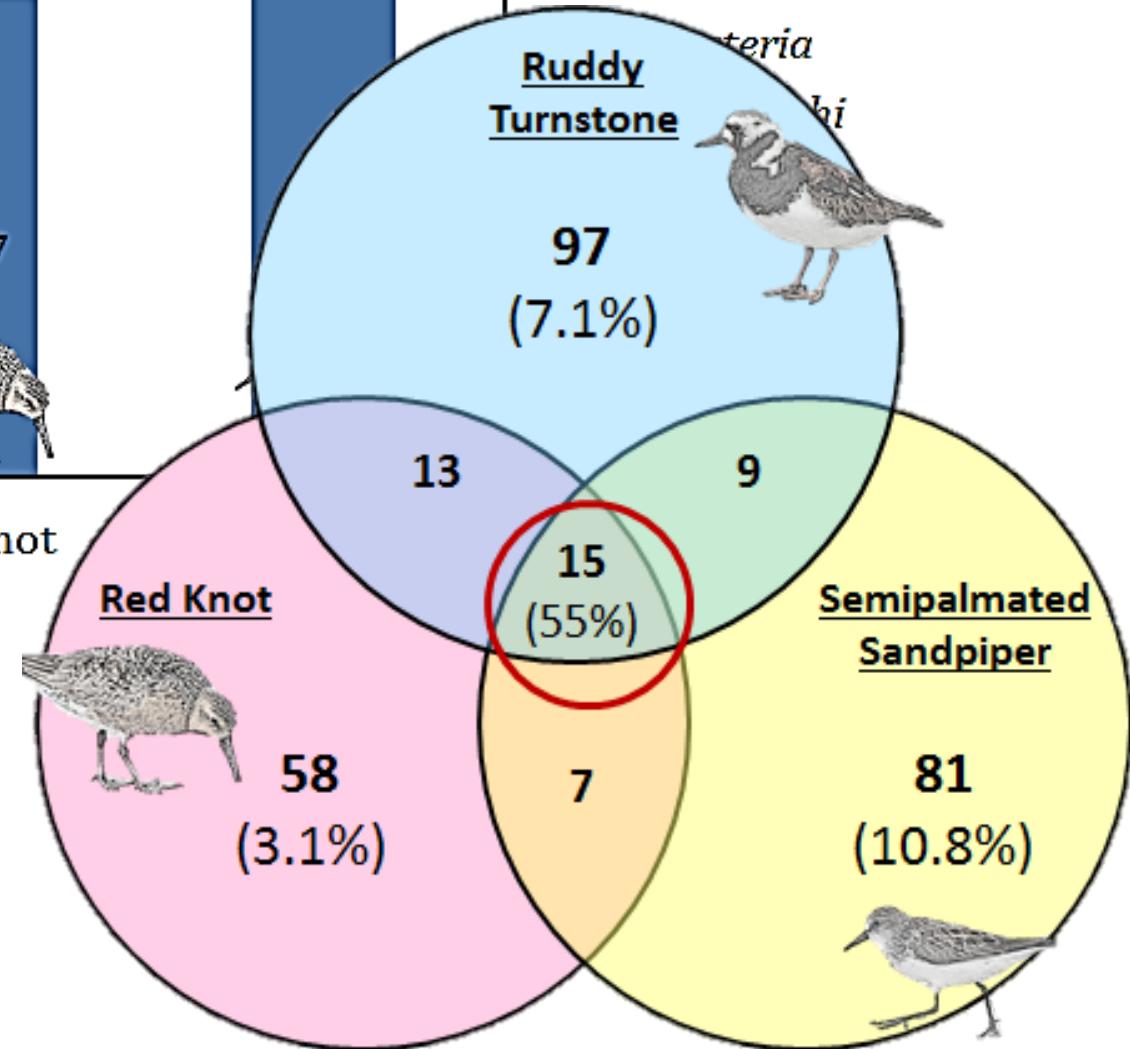
OTU	Size	Taxonomy
Otu000001	48622	Bacteria(100);Verrucomicrobiales(100);Verrucomicrobiaceae(100);Luteolibacter(91);
Otu000002	23757	Bacteria(100);Flavobacteriales"(100);Flavobacteriaceae(100);Flavobacterium(100);
Otu000003	21239	Bacteria(100);Burkholderiales(100);Comamonadaceae(100);unclassified(100);
Otu000004	20940	Bacteria(100);Burkholderiales(100);Comamonadaceae(100);unclassified(100);
Otu000005	20737	Bacteria(100);Burkholderiales(100);Comamonadaceae(100);unclassified(100);

Phylogenetic analysis





Venn diagram



Future work

1) Finish analysis of all fecal samples in Mothur/Beocat

2) Analyze results to reflect:

- inter and intraspecific differences in gut microbiota
- Site and migration route related differences in gut microbiota
- Life-time change in gut microbiota → Embryo, chick, adult
- Inter-annual variation in gut microbiota



