H3ABioNet

Pan African Bioinformatics Network for H3Africa

16SrRNA Intermediate Bioinformatics Online Course: Int_BT

Module 3:

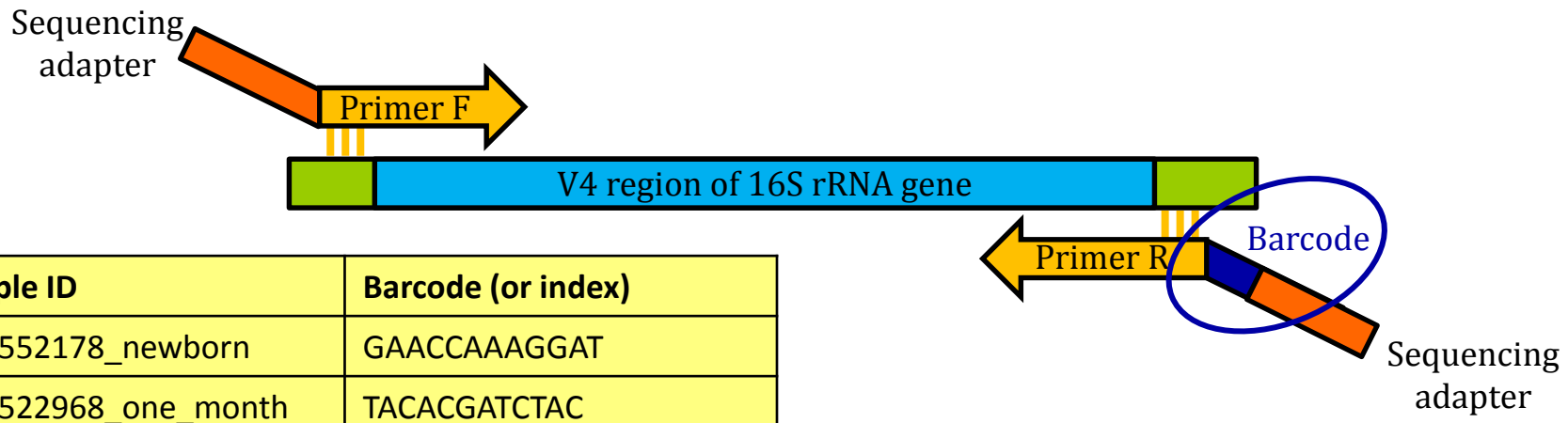
Sample collection, extraction and library prep for 16S NGS analyses

Part 3.4

16S rRNA high throughput sequencing: Library preparation

16S rRNA high throughput sequencing: library preparation

How many samples can be included in a sequencing run?



Sample ID	Barcode (or index)
GITB552178_newborn	GAACCAAAGGAT
GITB522968_one_month	TACACGATCTAC
GITB536987_two_months	GCGATATATCGC
GITM528796_birth	CAGTGCATATGC
GITM564789_birth	TCCAAAGTGTTC

16S rRNA high throughput sequencing: library preparation

How many samples can be included in a sequencing run?

Sample ID	Mother or infant	DNA yield (ng/ul)	Plate position	Run number	Barcode (or index)
GITB552178_newborn	infant	25	P1A01	Run1	GAACCAAAGGAT
GITB522968_one_month	infant	39	P1A02	Run1	TACACGATCTAC
GITB536987_two_months	infant	60	P1A03	Run1	GCGATATATCGC
GITM528796_birth	mother	221	P1A04	Run1	CAGTGCATATGC
GITM564789_birth	mother	157	P1A05	Run1	TCCAAAGTGTTC
GITM566987_birth	mother	195	P1A06	Run1	GGCCACGTAGTA

16S rRNA high throughput sequencing: library preparation

Batch extractions based on your sequencing plates layout

Plate 1 sequencing layout

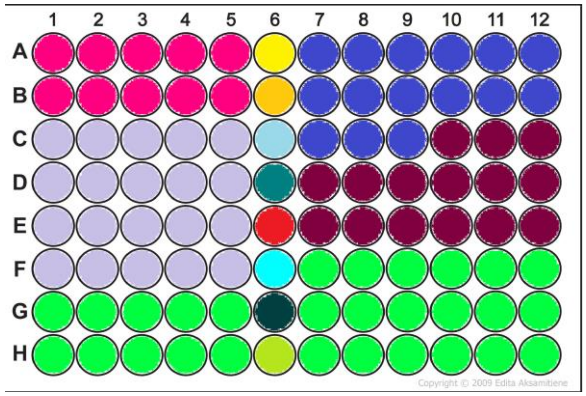


Plate 1 extraction layout

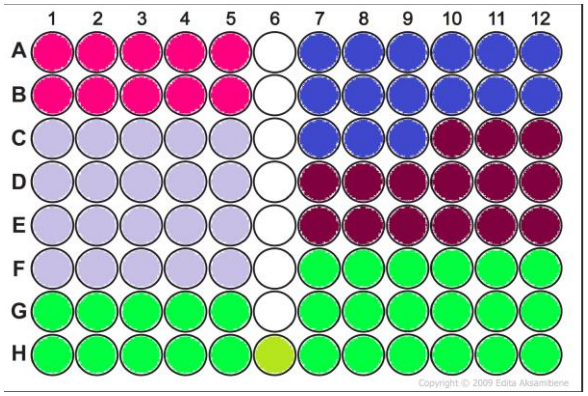


Plate 1 sequencing layout

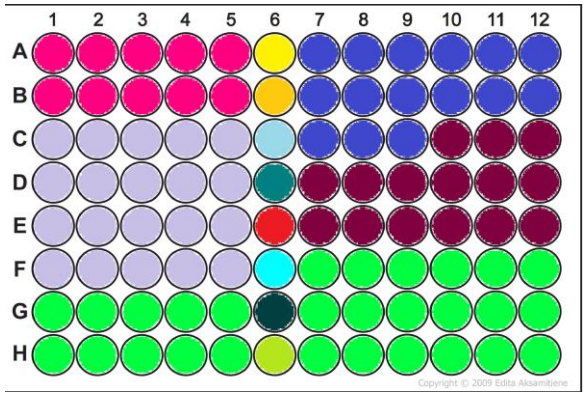
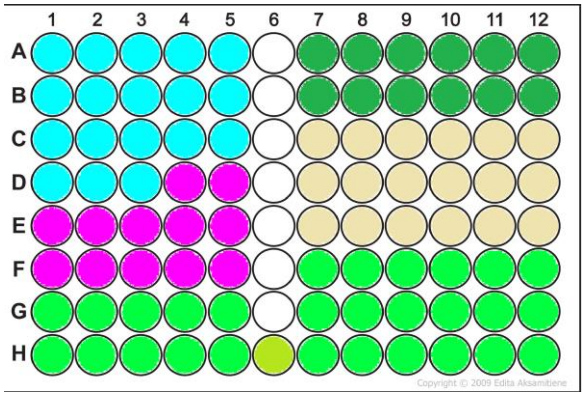


Plate 1 extraction layout



16S rRNA high throughput sequencing: library preparation

Include a set of controls on each 96-well plate

Plate 1

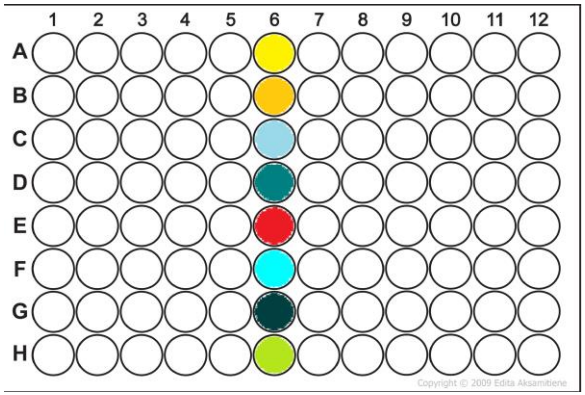
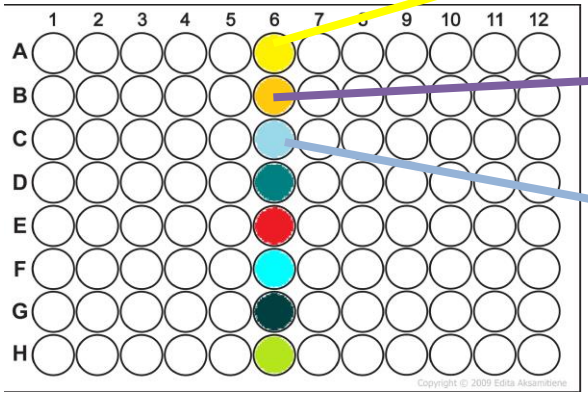


Plate 2



Within-run repeat (process any specimen in duplicate on each of the plates to measure reproducibility)

Between-run repeat (process any specimen in duplicate per run to measure reproducibility across runs)

Water used during PCR (to determine if any contamination was introduced during PCR reactions)

Plate 3

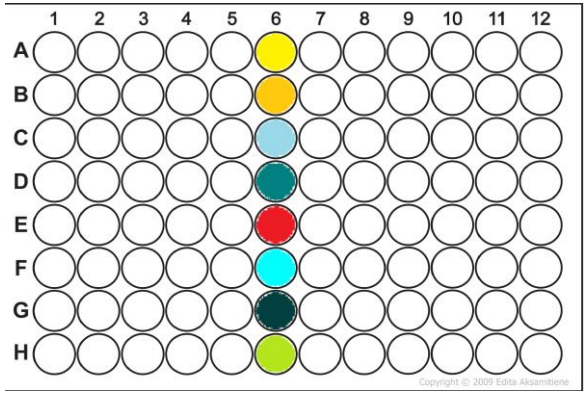
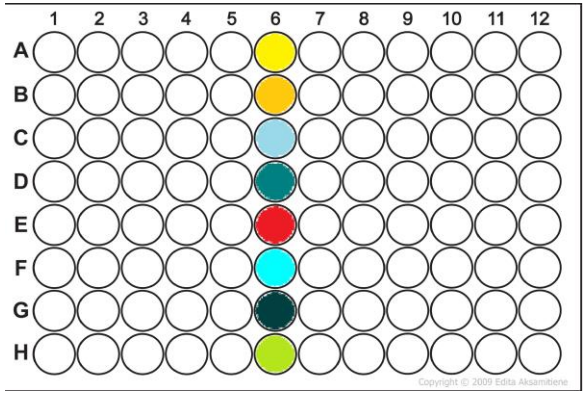


Plate4



Shantelle Classen-Weitz

16S rRNA high throughput sequencing: library preparation

Include a set of controls on each 96-well plate

Plate 1

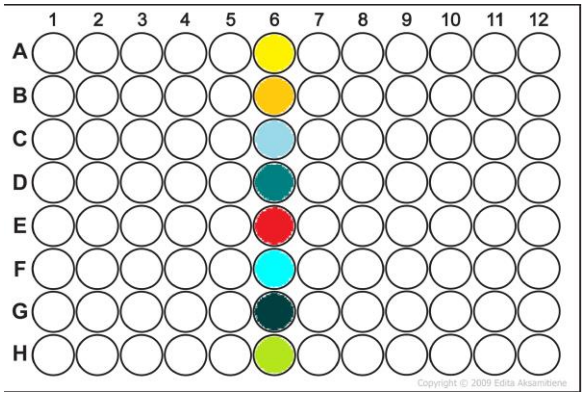


Plate 2

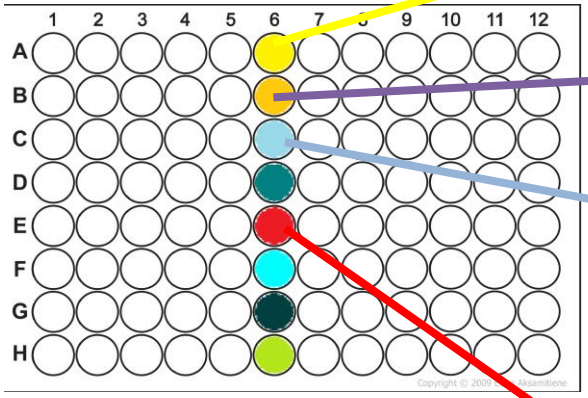


Plate 3

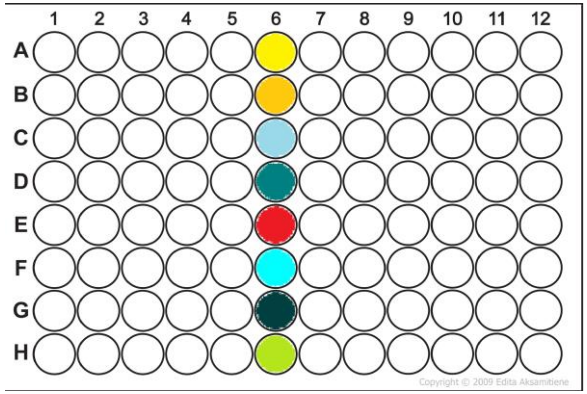
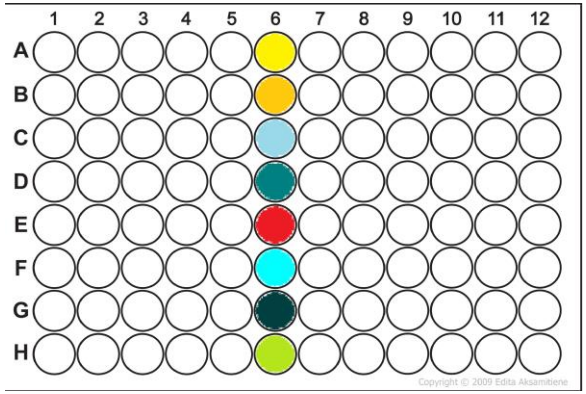


Plate4



Within-run repeat (process any specimen in duplicate on each of the plates to measure reproducibility)

Between-run repeat (process any specimen in duplicate per run to measure reproducibility across runs)

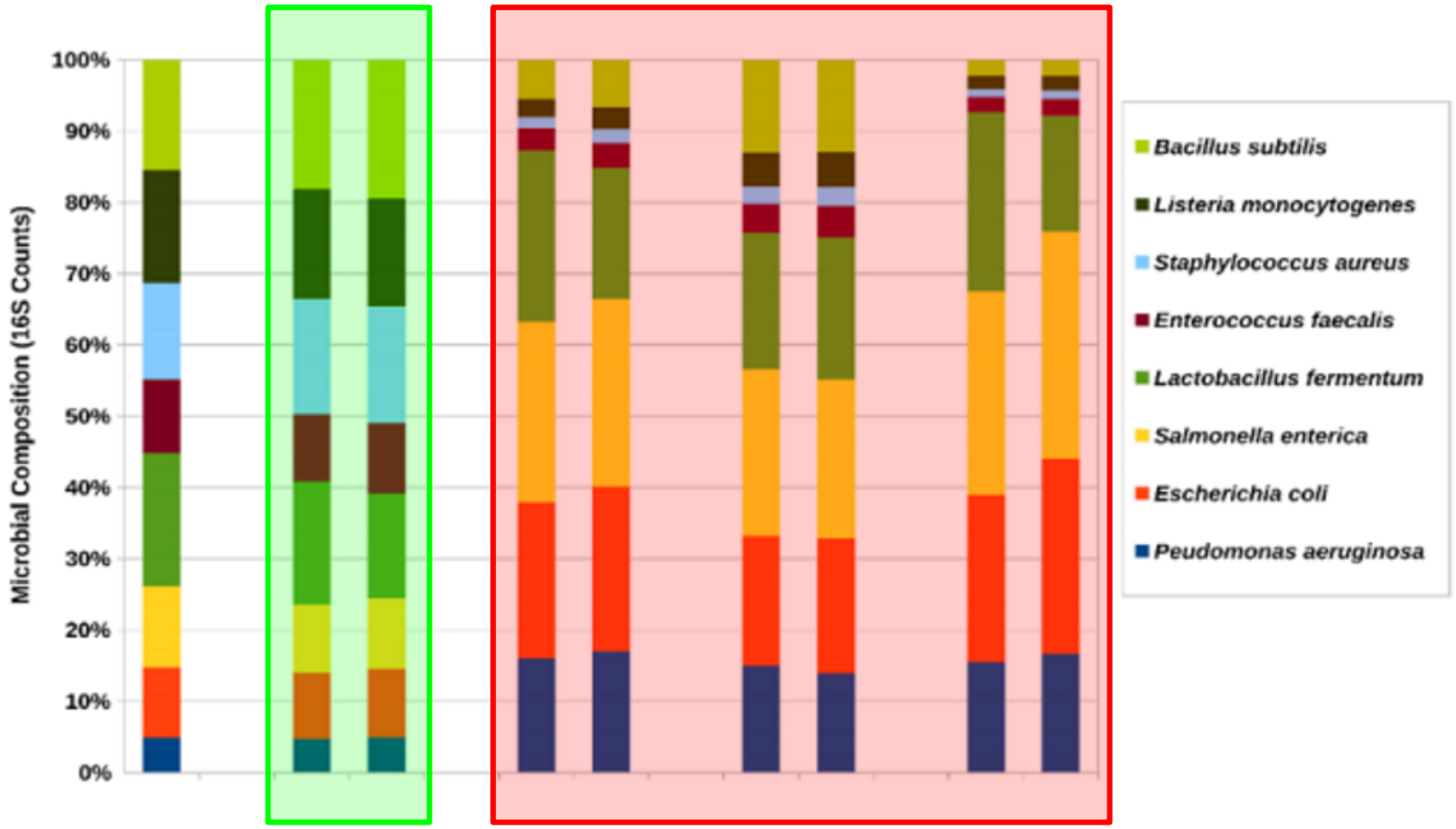
Water used during PCR (to determine if any contamination was introduced during PCR reactions)

Bacterial mock community cells (mix of bacterial cells serves as extraction control)

Bacterial mock community DNA (mix of bacterial DNA serves as sequencing control)

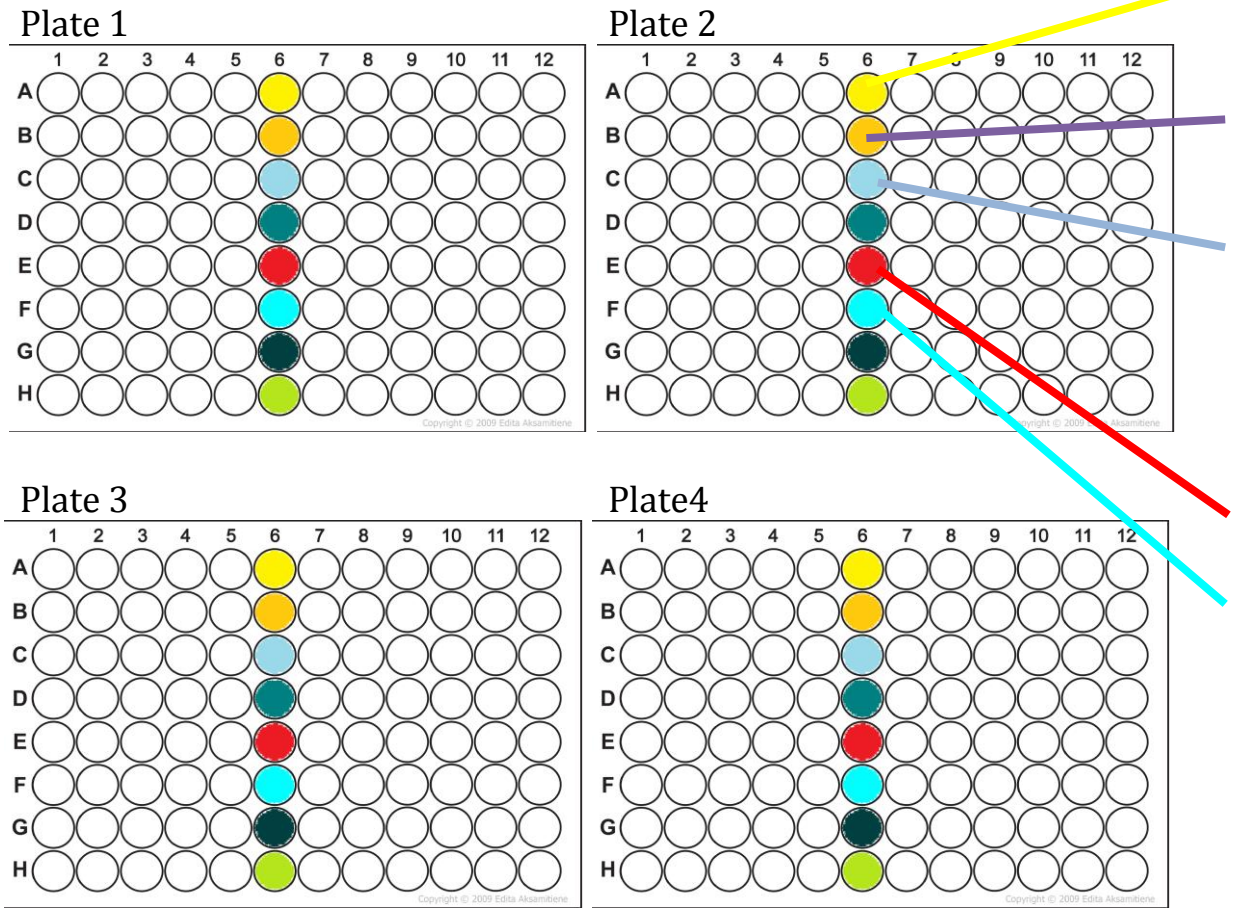
16S rRNA high throughput sequencing: library preparation

Include a set of controls on each 96-well plate



16S rRNA high throughput sequencing: library preparation

Include a set of controls on each 96-well plate



Between-run repeat (process any specimen in duplicate per run to measure reproducibility across runs)

Within-run repeat (process any specimen in duplicate on each of the plates to measure reproducibility)

Water used during PCR (to determine if any contamination was introduced during PCR reactions)

Bacterial mock community DNA (mix of bacterial DNA serves as sequencing control)

Sample storage medium or lysis buffer (to determine if contaminants was introduced during sample collection)

16S rRNA high throughput sequencing: library preparation

We perform a 2-step PCR to amplify the V4 region of the 16S rRNA gene

Claassen-Weitz et al. 2018. *Scientific Reports* volume 8, Article number: 5078
Caporaso et al., *PNAS*., 2011; 108:4516-4522

Short PCR (10 cycles)

We use 4ul of template for stool samples and 7ul of template for nasopharyngeal and breast milk samples

515F: GTGCCAGCHGCYGCGGT
806R: GGACTACNNGGGTWTCTAAT

Long PCR

We use 4ul of amplicon from the short PCR for stool samples (and 7ul for nasopharyngeal and breast milk samples) as template in the long PCR

515Fmod4_std_12-16N

5' AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNNNNN(12-16)GTGCCAGCHGCYGCGGT 3' →

Sequence adapter | Sequencing priming region | Random nucleotides | 16S primer

806Rmod1_std_12-16N_BC001

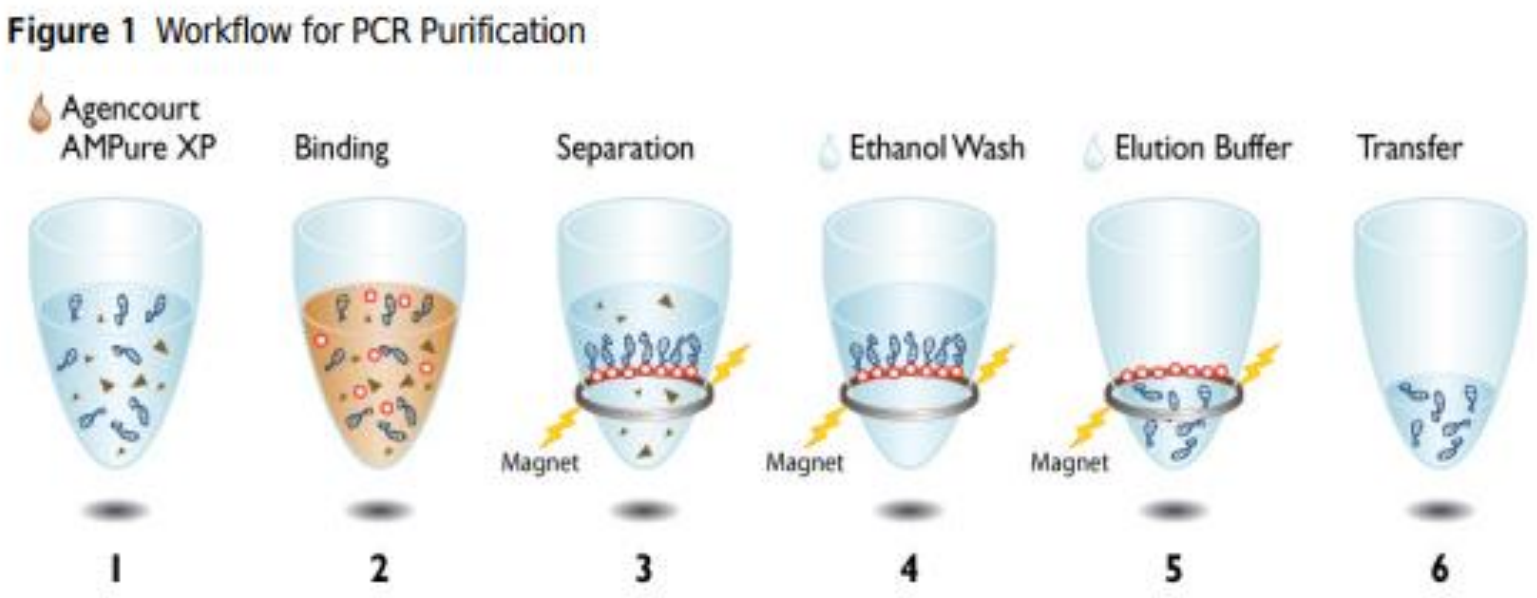
← 3' TAATCTWTGGGNNCATCAGG(12-16)NNNNNNNNNNNNTCTAGCCTTCTCGTGTGCACAGTTGAGGTCAGTGTTAGTCAGAGCATAGAGCATAGAGCATACGGCAGAAGACGAAC 5'

16s primer | Random nucleotides | Sequencing priming region | Barcode | Sequence adapter

16S rRNA high throughput sequencing: library preparation

The Agencourt AMPure XP PCR Purification System

Agencourt AMPure XP utilizes an optimized buffer to selectively bind DNA fragments 100 bp and larger to paramagnetic beads.
Excess primers, nucleotides, salts, and enzymes can be removed using a simple washing procedure.
The result is a more purified PCR product.



16S rRNA high throughput sequencing: library preparation

Quantification of individual libraries and pooling

Plate 1	1	2	3	4	5	6	7	8	9	10	11	12
A	1.23	9.43	5.32	2.45	6.77	4.07	6.18	4.00	4.71	1.76	2.49	1.87
B	4.10	2.77	2.32	4.33	2.75	6.02	4.51	3.08	4.96	2.08	5.54	1.69
C	4.31	1.81	2.14	8.33	7.91	5.48	4.35	2.67	6.37	2.86	3.40	1.72
D	3.68	3.06	4.33	6.39	3.27	5.31	7.01	1.95	7.37	6.56	2.35	2.45
E	5.78	1.34	2.93	4.84	2.58	4.00	4.02	3.34	7.99	4.76	2.58	4.77
F	1.36	2.30	3.71	4.40	1.82	37.55	9.65	4.87	7.01	3.23	2.49	1.52
G	4.25	1.50	2.48	1.90	2.20	5.88	31.28	2.70	6.60	5.78	1.65	1.33
H	1.18	2.35	63.67	1.72	3.93	7.11	4.23	1.89	3.52	1.82	2.55	1.57

Plate 2	1	2	3	4	5	6	7	8	9	10	11	12
A	1.27	1.02	1.68	1.51	1.23	9.78	5.70	4.02	2.47	2.25	1.85	2.37
B	1.81	2.43	2.36	1.82	1.50	2.55	109.21	7.64	4.06	6.29	2.45	1.46
C	1.18	1.36	2.40	5.93	4.31	2.96	3.66	9.45	2.15	2.30	3.32	1.95
D	1.83	1.16	1.57	6.41	14.95	5.28	4.90	11.21	2.96	2.37	1.61	21.58
E	0.99	1.42	1.93	1.83	3.77	2.27	2.78	17.45	2.72	9.26	1.93	2.53
F	3.46	1.70	1.77	4.81	5.01	-191.47	10.34	4.92	-669.00	6.45	1.52	3.86
G	3.70	2.11	2.68	5.84	2.07	3.01	7.84	3.15	2.95	1.64	1.47	1.89
H	7.17	1.48	1.68	1.39	5.64	1.79	2.91	2.38	1.68	1.88	1.59	2.36

Plate 3	1	2	3	4	5	6	7	8	9	10	11	12
A	2.51	0.55	0.80	0.95	0.81	6.37	0.57	0.58	0.70	1.13	75.70	2.47
B	0.63	0.59	0.54	0.94	0.73	1.07	0.73	0.84	0.73	0.69	1.92	1.45
C	4.40	0.71	0.80	0.77	0.81	0.64	1.97	1.32	0.84	1.06	1.94	2.45
D	0.75	1.14	0.74	0.70	0.88	0.73	1.10	1.24	1.40	1.19	2.29	1.72
E	2.95	1.00	0.61	3.08	3.19	0.74	1.11	3.06	1.12	1.15	2.79	2.20
F	0.97	0.55	0.58	0.64	0.73	3.48	26.25	2.73	1.37	1.49	1.57	1.34
G	0.96	0.64	0.96	0.64	0.96	0.67	3.37	1.09	0.90	1.91	1.73	2.24
H	2.35	1.10	0.68	0.60	0.57	0.76	2.09	1.21	2.32	1.92	1.13	1.73

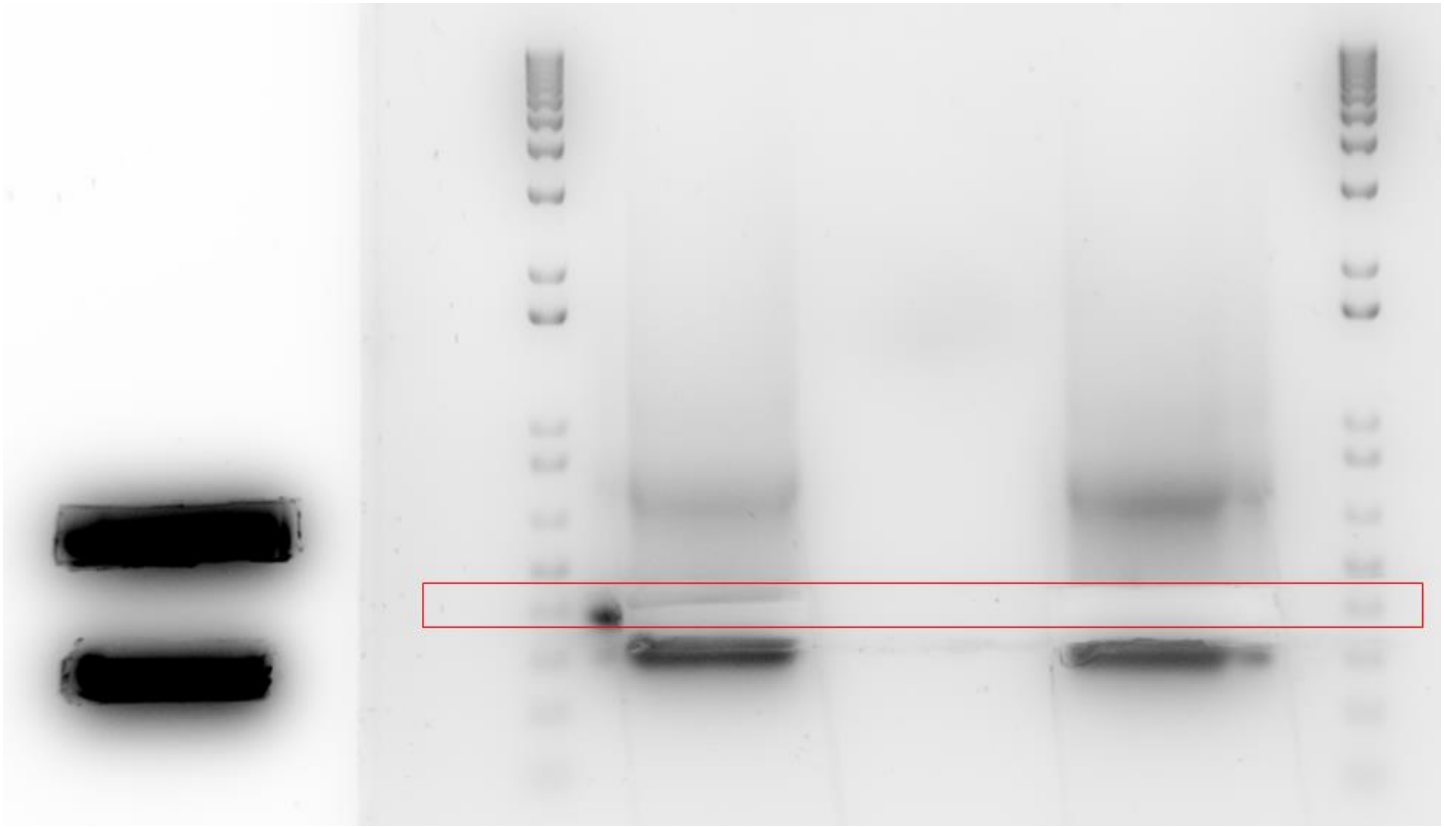
Plate 4	1	2	3	4	5	6	7	8	9	10	11	12
A	1.09	1.10	1.93	1.87	1.80	2.15	1.69	1.73	2.19	2.69	1.54	1.42
B	3.40	3.13	1.54	1.25	1.37	2.72	3.86	2.30	6.13	2.26	2.77	2.78
C	1.38	1.75	1.37	1.37	1.83	1.61	1.75	3.01	2.20	2.30	2.77	2.52
D	1.90	4.34	1.20	4.37	2.03	1.33	1.57	2.36	1.51	2.76	3.18	1.96
E	2.90	2.29	1.66	1.40	1.89	2.08	2.88	2.92	1.93	1.75	3.32	2.20
F	1.65	1.43	1.84	1.97	2.66	45.62	28.16	2.69	2.89	2.19	3.99	2.15
G	1.34	1.92	1.66	1.88	1.72	1.69	2.05	1.84	3.29	2.35	2.06	2.09
H	1.58	1.96	1.43	1.32	1.83	1.43	3.68	2.20	2.14	2.23	2.52	2.21

Double stranded DNA is quantified using the QuantiFluor™ dsDNA System

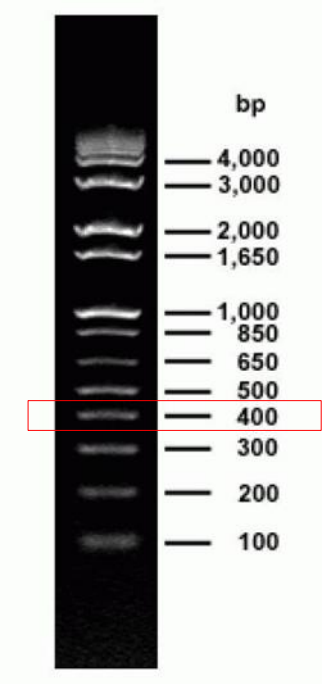
Equimolar amounts of each amplicon are pooled to achieve appropriate numbers of sequences per sample

16S rRNA high throughput sequencing: library preparation

Gel extraction of pooled library (removal of primer dimers and non-specific binding)



TrackIt™ 1 Kb Plus
DNA Ladder



16S rRNA high throughput sequencing: library preparation

- ***The amplicon product is cleaned using the QIAquick Gel Extraction Kit (Qiagen) to remove agarose gel and impurities.***
- ***The fragment size is confirmed using an automated electrophoresis platform***
- ***The KAPA Library Quantification Kit for Illumina® platforms is then used for final quantification prior to denaturation and dilution of the final library.***

16S rRNA high throughput sequencing: library preparation

- ***Sequencing preparation***

1. Denature final pool of DNA library using 0.2N NaOH
2. Dilute quantified final pool of DNA to the desired concentration using Hybridization Buffer (Illumina)
3. PhiX control is denatured and diluted
   Control libraries generated from the PhiX virus.
   Characteristics of the PhiX genome provide several benefits:
   Small—PhiX is a small genome, which enables quick alignment and estimation of error rates.
   Diverse—The PhiX genome contains approximately 45% GC and 55% AT.
   Well-Defined—PhiX has a well-defined genome sequence. Illumina cluster generation algorithms are optimized around a balanced representation of A, T, G, and C nucleotides.
4. Combine sample library and PhiX control:
   For most libraries PhiX is added at 1%. However, for low diversity libraries ≥5% is required.
5. Prepare flow cell and kit for the sequencing run, sequence, and obtain your data (ready for Bio-informatics steps).

16S rRNA high throughput sequencing: library preparation

In summary:

- Carefully plan your experiment!
- Think about things such as extraction controls and sequencing controls. Also think about controls to measure reproducibility within a run and between runs (if you are working with a large number of samples).
- Design your plate layouts to from the start of your experiment to avoid re-aliquoting into different positions on your 96-well plates at any step of your library workflow.
- The library preparation steps outline in this set of slides is an in-house workflow and doesn't mean that workflows other than this should not be followed.
- Finally, you will obtain your sequencing data and you are now ready for the bio-informatics processes.