Welcome to
RADcamp NYC 2019

Eaton Lab

Columbia University

Dr. Sandra Hoffberg

NSF Postdoctoral Research Fellow

Department of Ecology, Evolution, Environmental Biology Department, Columbia University

Sandra.Hoffberg@Columbia.edu @SandraHoffberg

The University of Georgia

COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

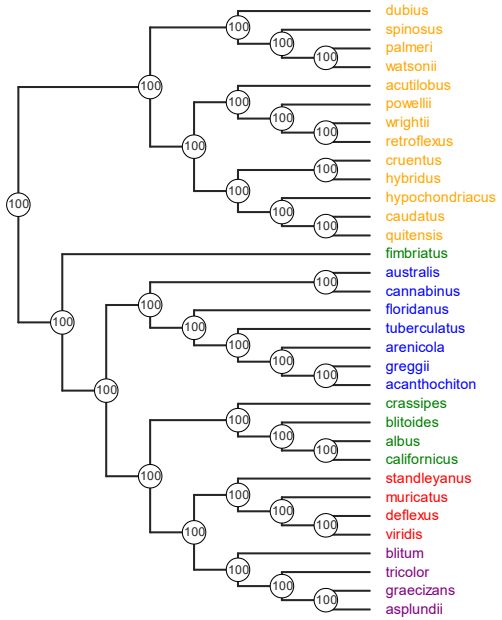
MOLECULAR ECOLOGY RESOURCES

Molecular Ecology Resources (2016) **16**, 1264–1278 doi: 10.1111/1755-0998.12566

SPECIAL ISSUE: SEQUENCE CAPTURE

RADcap: sequence capture of dual-digest RADseq libraries with identifiable duplicates and reduced missing data

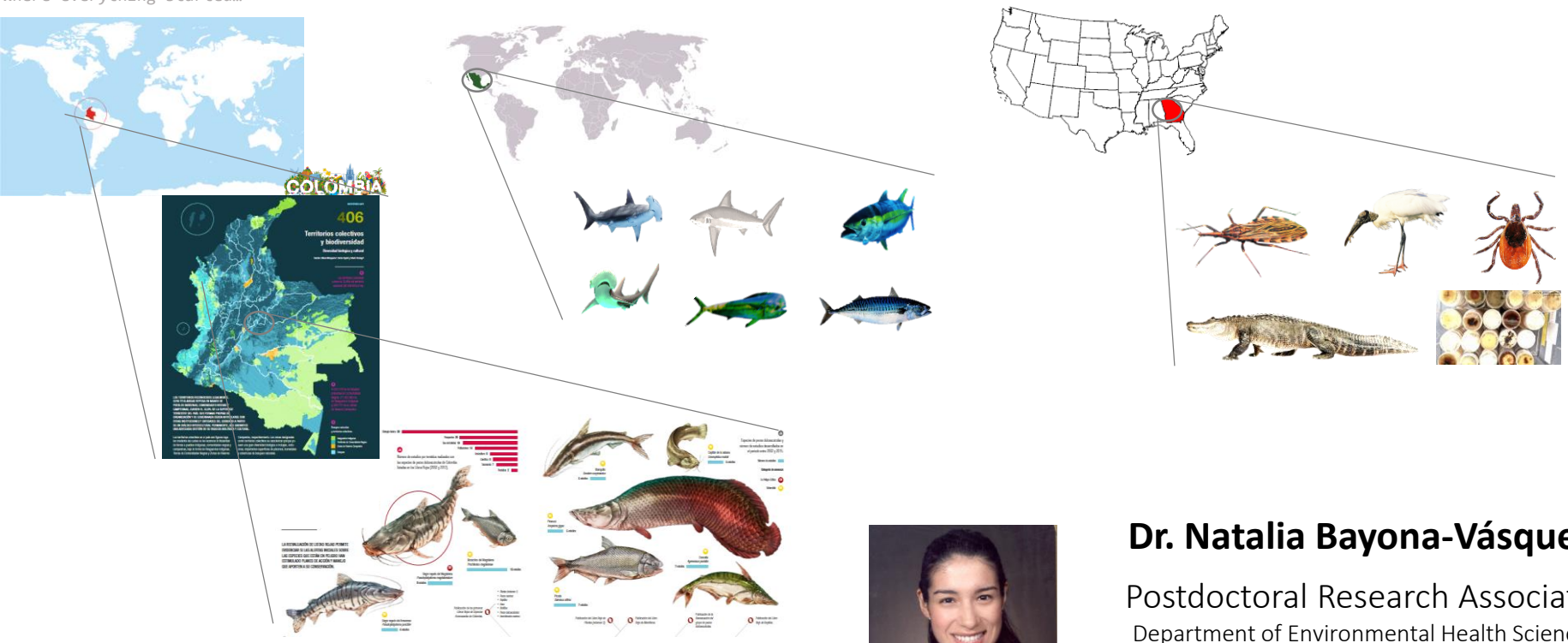
SANDRA L. HOFFBERG,* TROY J. KIERAN,† JULIAN M. CATCHEN,‡ ALISON DEVAULT,§ BRANT C. FAIRCLOTH,¶ RODNEY MAURICIO* and TRAVIS C. GLENN*,†



Sandra Hoffberg

Dr. Natalia Bayona Vásquez

Where everything started…



Dr. Natalia Bayona-Vásquez

Postdoctoral Research Associate

Department of Environmental Health Science

Institute of Bioinformatics

University of Georgia

@Nati_bayona

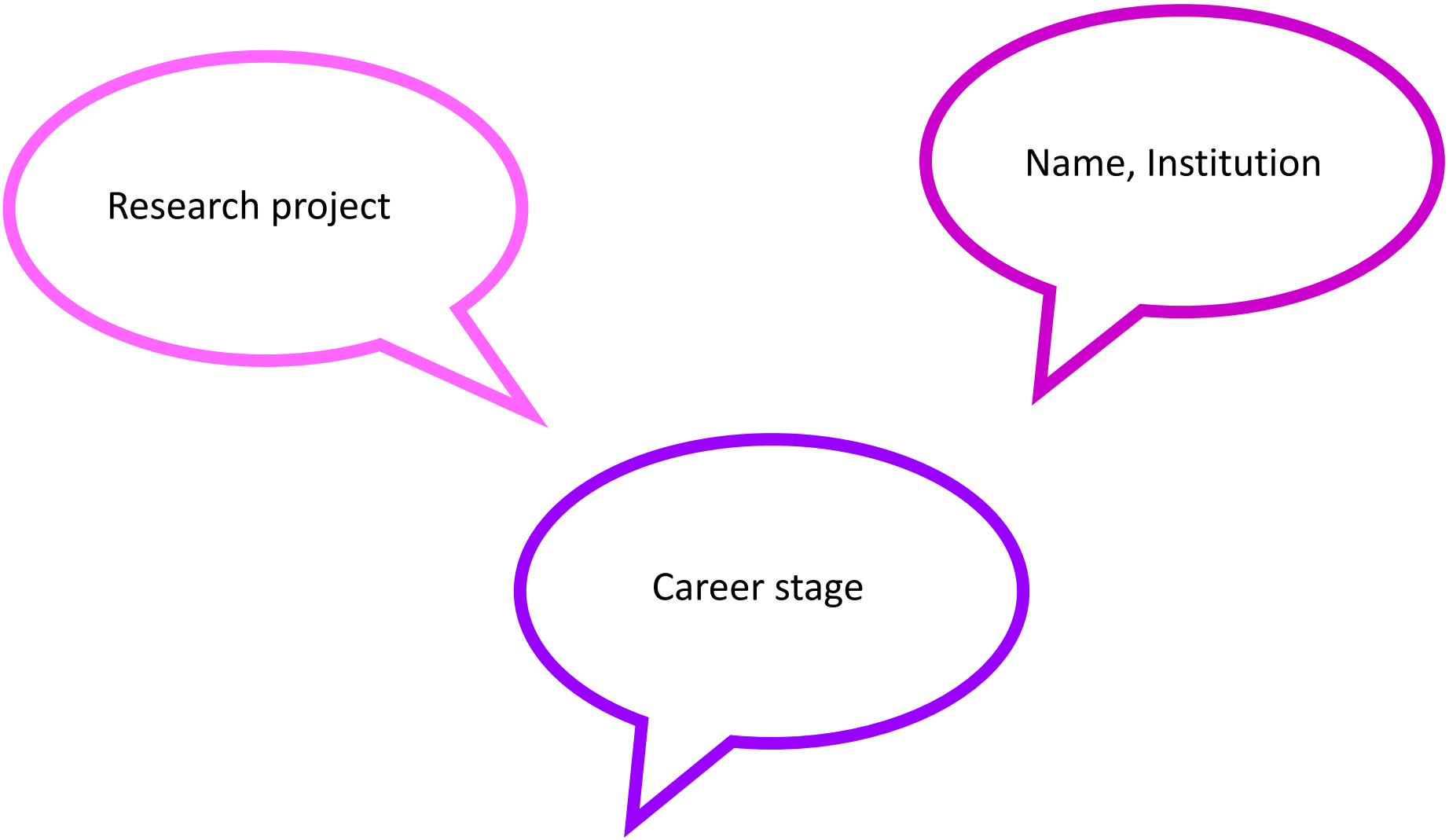
http://njbayona7.wixsite.com/natalia-bayona

njbayonav@gmail.com

Institute of Bioinformatics
UNIVERSITY OF GEORGIA

Environmental Health Science
College of Public Health
UNIVERSITY OF GEORGIA

Participant introductions



RADcamp Schedule

This weekend	Library Prep
Saturday	Lecture, Digestion, Ligation
Sunday	Amplification
During the next 1 month	Send Samples to Illinois for sequencing on Illumina Novaseq
	Demultiplex data and distribute
October 12-13	Assemble reads, call variants in ipyrad
	Phylogenetic and demographic analyses

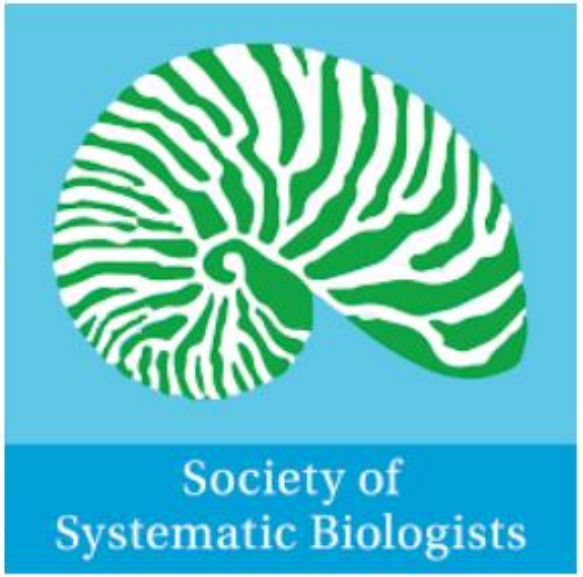
Learning objectives

At the end of this workshop, participants should be able to…

- Perform protocol and teach it to their lab-mates.
- Identify which reagents and supplies are necessary for 3RAD library prep and where to obtain them.
- Generate useable pilot data.
- Use and explain scripts to demultiplex their 3RAD data in ipyrad.
- Analyze data in ipyrad analysis toolkit and interpret results.
- More efficiently use tools for reproducible coding (unix, jupyter, ipyrad, etc)

RADCamp NYC 2019 co-sponsored by:

SOCIETY *for the* STUDY *of* EVOLUTION



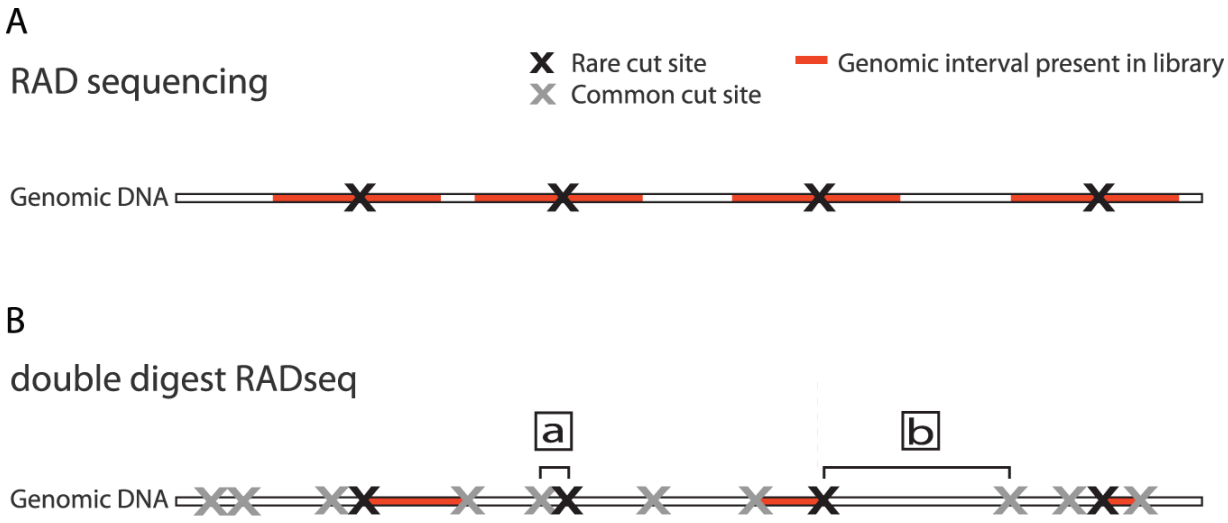
THE GRADUATE CENTER
CITY UNIVERSITY OF NEW YORK

Introduction to RADseq and 3RAD

Slides that contain this symbol in the bottom right corner, indicate that the material is available in our RADcamp 2019 materials webpage

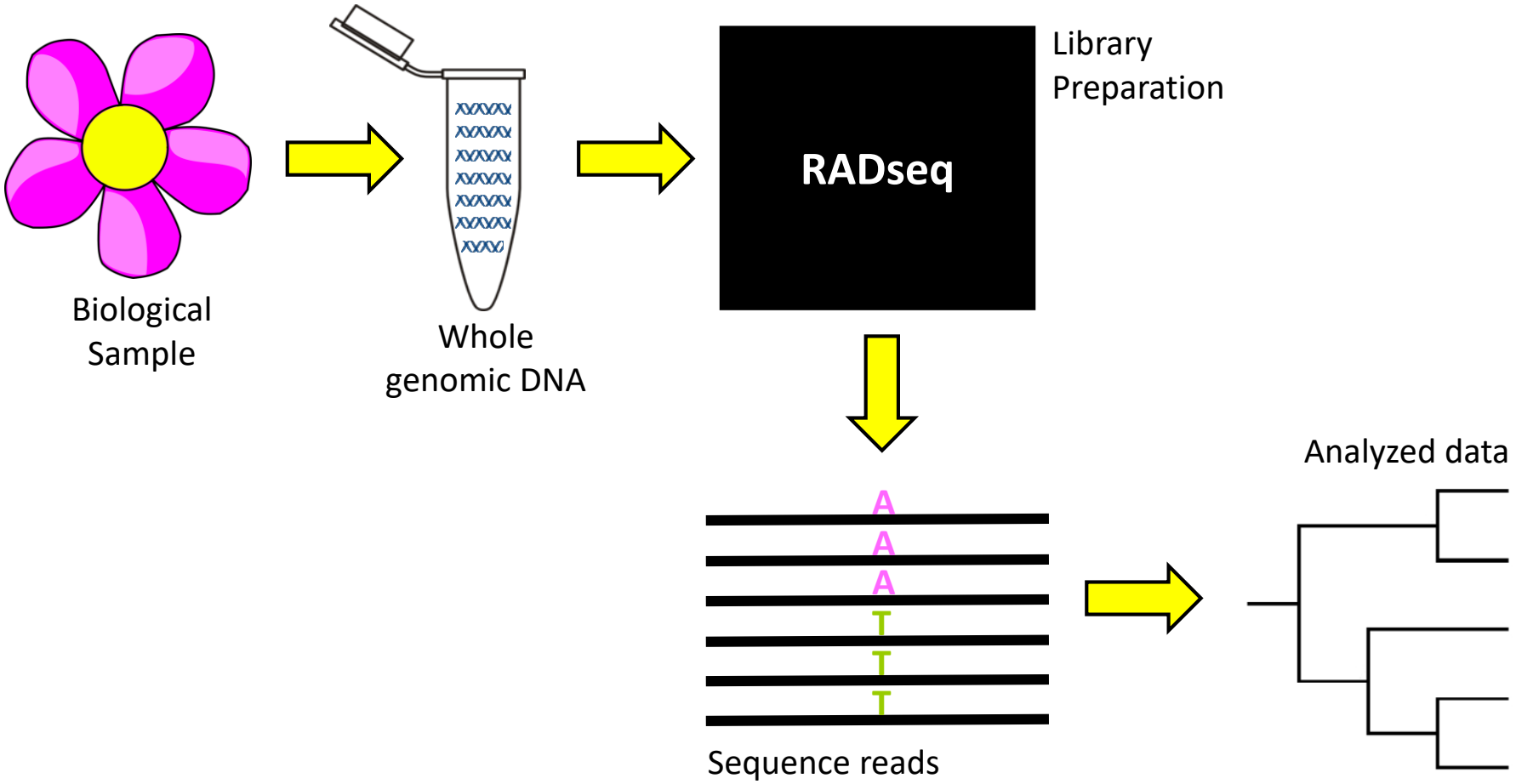
What is RADseq?

- Restriction-enzyme associated DNA – sequencing
- Reduced representation library preparation method

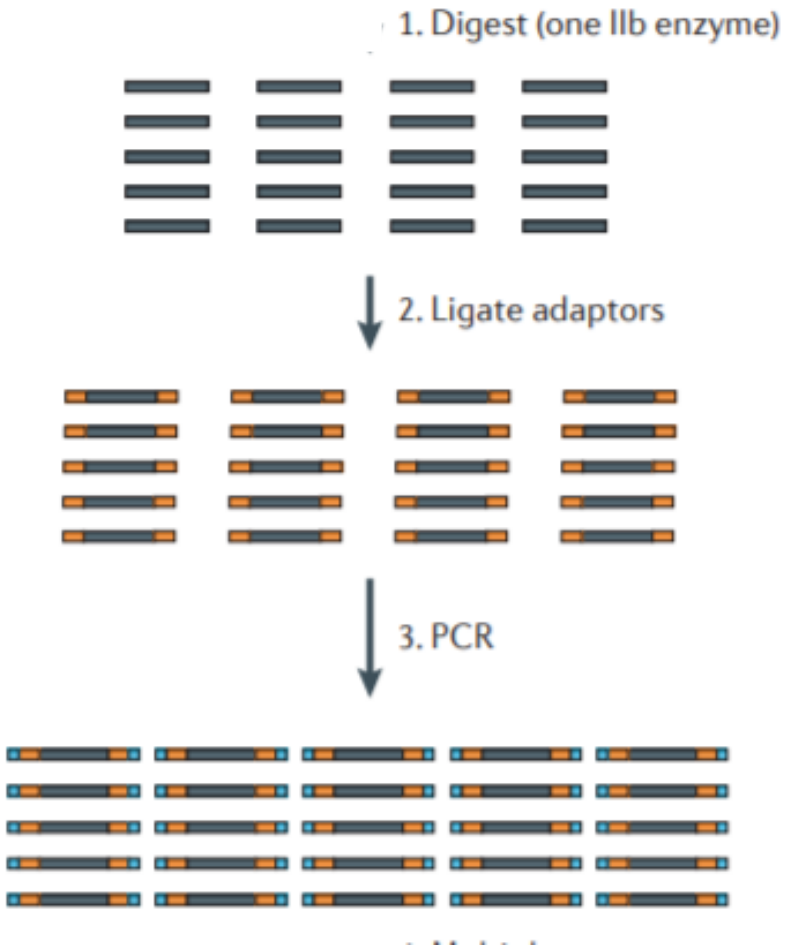


Modified from Peterson et al. 2012

RADseq Project Workflow



What is RADseq, technically?



- ✓Digest genomic DNA with a restriction enzyme
- ✓Incorporate a molecular ID to associate sequence reads to individuals.
- ✓Add sequencing primers
- ❑Multiplex samples
- ❑Size selection to further reduce genome

Andrews et al. 2016

Adapterama I: Universal stubs and primers for 384 unique dual-indexed or 147,456 combinatorially-indexed Illumina libraries (iTru & iNext)

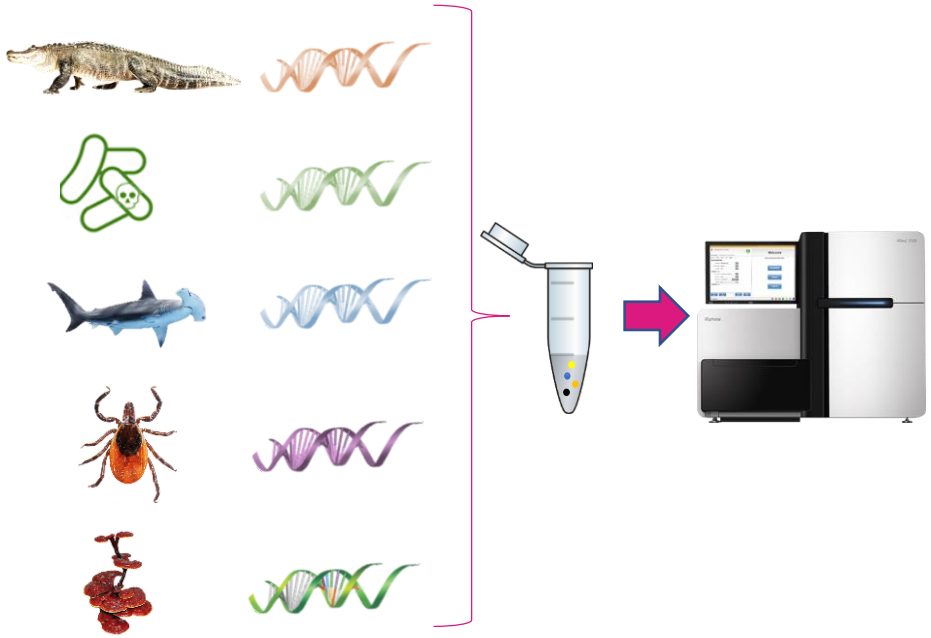
Travis C. Glenn, Roger A. Nilsen, Troy J. Kieran, Jon G. Sanders, Natalia J. Bayona-Vásquez, John W. Finger Jr., Todd W. Pierson, Kerin E. Bentley, Sandra L. Hoffberg, Swarnali Louha, Francisco J. García-De León, Miguel Angel Del Río-Portilla, Kurt D. Reed, Jennifer L. Anderson, Jennifer K. Meece, Samuel E. Aggrey, Romdhane Rekaya, Magdy Alabady, Myriam Bélanger, Kevin Winker, Brant C. Faircloth

doi: https://doi.org/10.1101/049114

Adapterama II: Universal amplicon sequencing on Illumina platforms (TaggiMatrix)

Travis C. Glenn, Todd W. Pierson, Natalia J. Bayona-Vásquez, Troy J. Kieran, Sandra L. Hoffberg, Jesse C. Thomas IV, Daniel E. Lefever, John W. Finger Jr., Bei Gao, Xiaoming Bian, Swarnali Louha, Ramya T. Kolli, Kerin Bentley, Julie Rushmore, Kelvin Wong, Timothy I. Shaw, Michael J. Rothrock Jr., Anna M. McKee, Tai L. Guo, Rodney Mauricio, Marirosa Molina, Brian S. Cummings, Lawrence H. Lash, Kun Lu, Gregory S. Gilbert, Stephen P. Hubbell, Brant C. Faircloth

doi: https://doi.org/10.1101/619544



Adapterama III: Quadruple-indexed, double/triple-enzyme RADseq libraries (2RAD/3RAD)

Natalia J Bayona-Vásquez, Travis C Glenn, Troy J Kieran, Todd W Pierson, Sandra L Hoffberg, Peter A Scott, Kerin E Bentley, John W Finger Jr., Swarnali Louha, Nicholas Troendle, Píndaro Díaz-Jaimes, Rodney Mauricio, Brant C Faircloth

doi: https://doi.org/10.1101/205799

Highlights of 3RAD

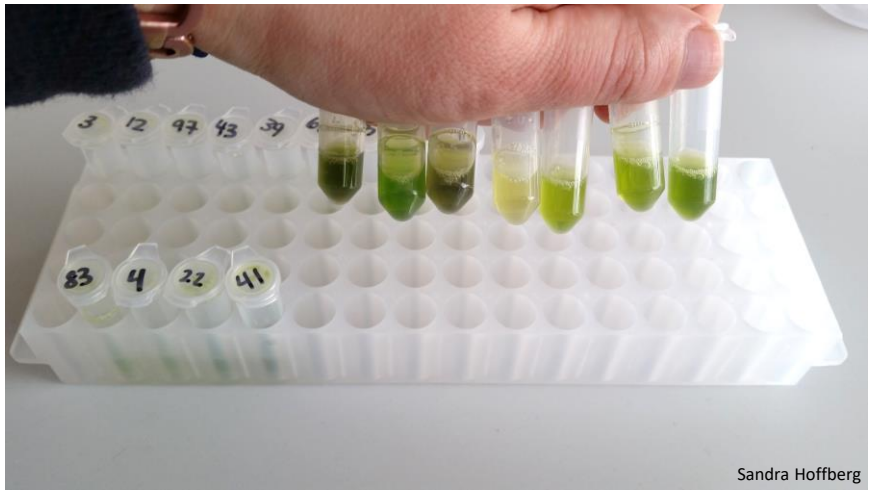
1. Minimize the number of processing steps.
2. Ligate adapters in the presence of active restriction enzymes, thereby reducing chimeras.
3. Includes an optional third restriction enzyme to cut apart adapter-dimers.
4. Interchangeable adapter designs.
5. Variable-length internal indexes to increase sequence diversity.
6. Compatible with Illumina sequencing reagents and libraries.
7. Easy modification for the identification of PCR duplicates.
8. Cheap.

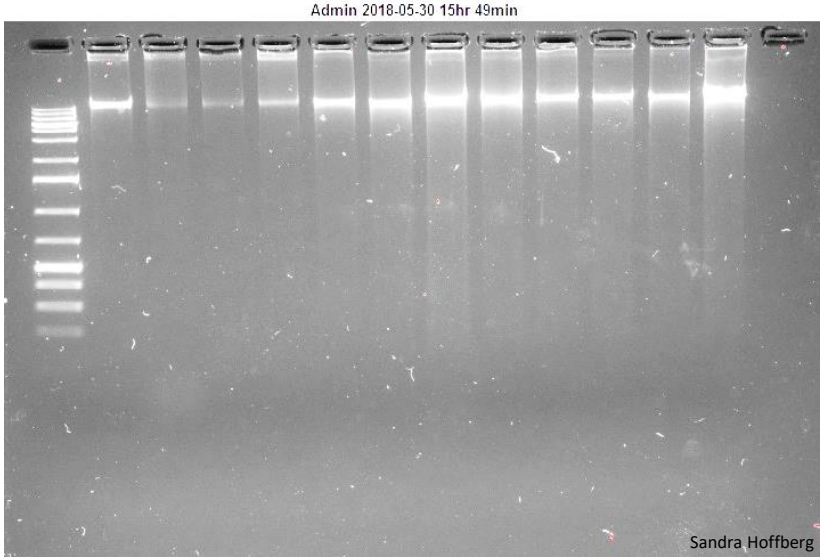
What are the steps of 3RAD?

Collect Samples

©2007 United States Department of Agriculture

Extract DNA





Many ways to do this

- Spin column (Qiagen kit)
- Phenol Chloroform
- Magnetic Bead based

Quantify and Normalize DNA

BioTek Synergy


Agilent TapeStation

Invitrogen Qubit

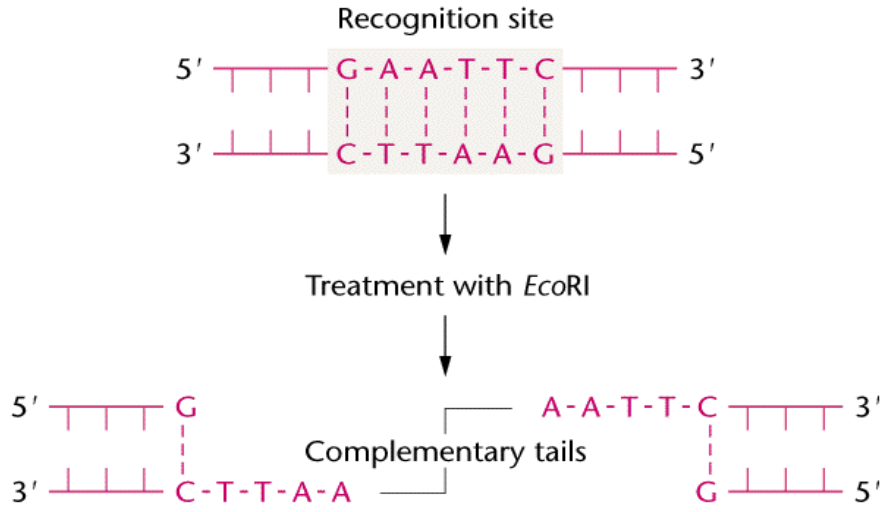

ThermoFisher Nanodrop

- Necessary to ensure even # reads across samples.
- Many options, more sensitive method will be better.
- Which machine is available to you?

Step 1 in the lab: Normalize DNA

Digestion

- Use restriction enzymes to make ds breaks in DNA at enzyme-specific cut sites.
- Most enzymes leave ss DNA overhang



DNA digestion reaction will include

- Enzymes
- CutSmart buffer
- up to 100ng DNA
- Water
- R1 and R2 adapters

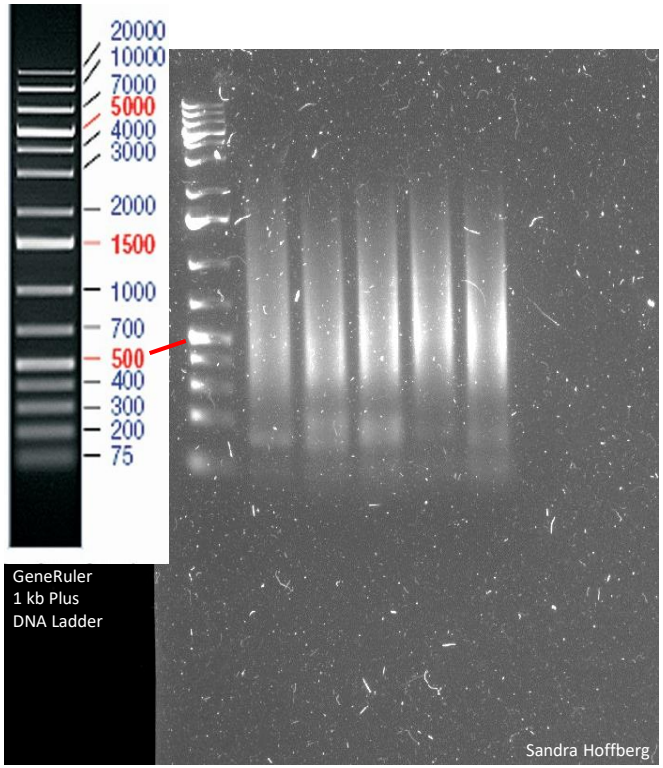
- Incubate 37°C for 1 hour

Enzyme selection

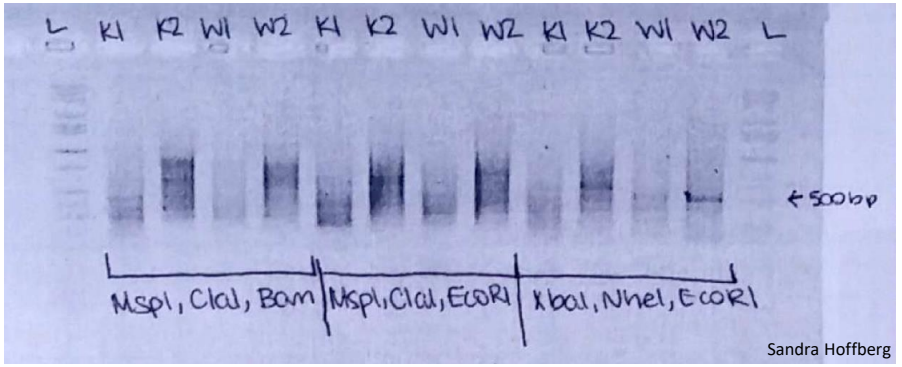
- 3RAD compatible enzyme sets were designed based on price and recognition site length
- You have choices for which sets of enzymes to use
- How to pick?
    1. Do a 3RAD prep in few samples (digestion, ligation PCR)
    2. Look at size distribution on an agarose gel
    3. Pick best enzymes or continue with multiple
    4. (Sequence)
    5. (Evaluate # loci and coverage)

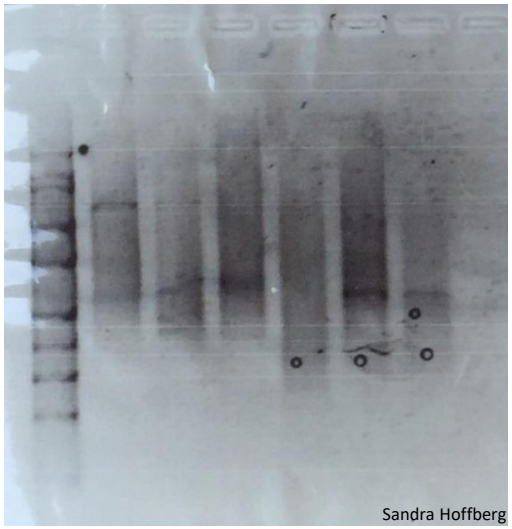
Enzyme selection

Smear in ~500 bp range is good



High molecular weight DNA is bad
Bands in ~500bp range is bad

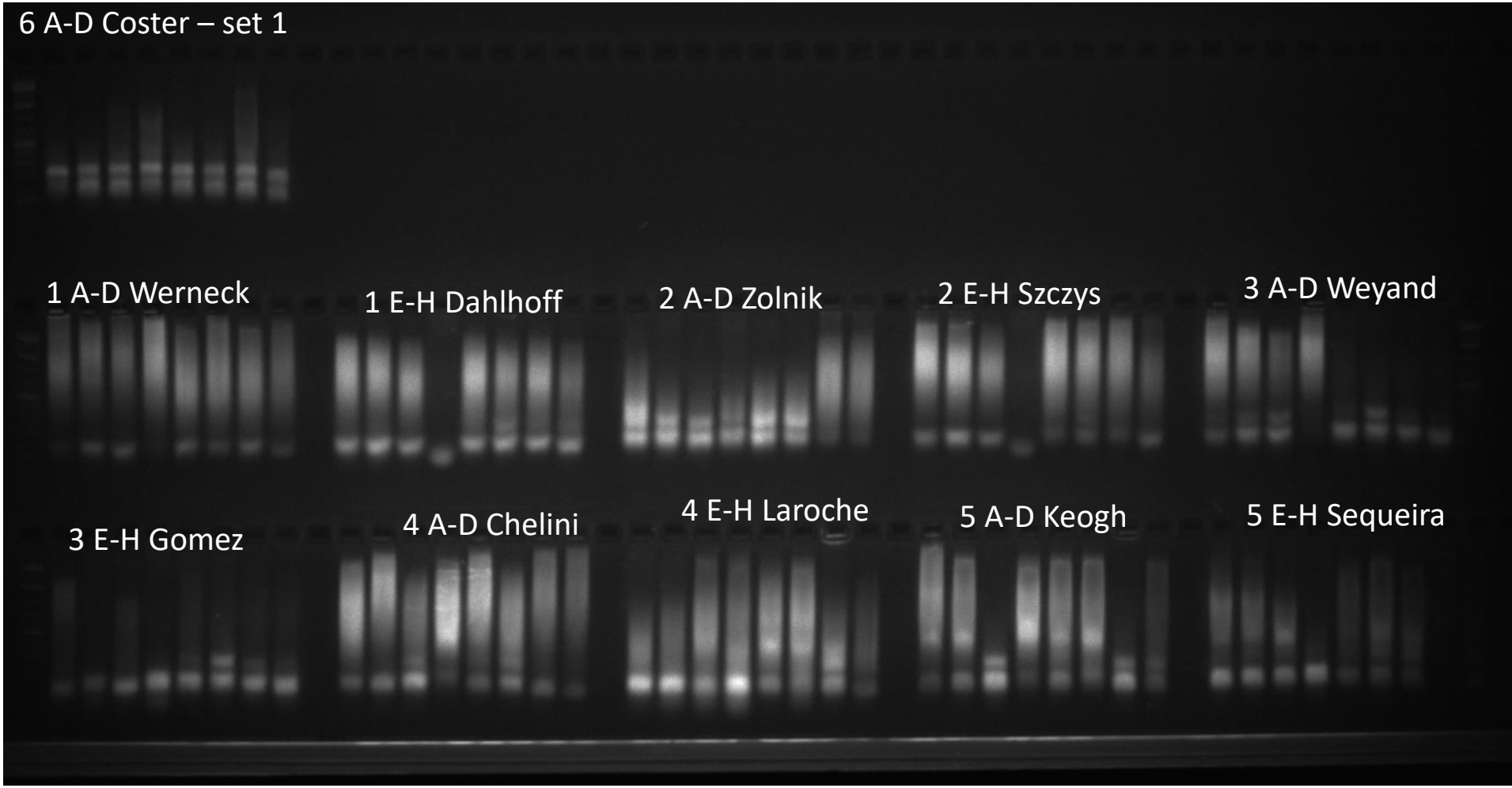




We did this for you!
This is why you sent samples in advance.

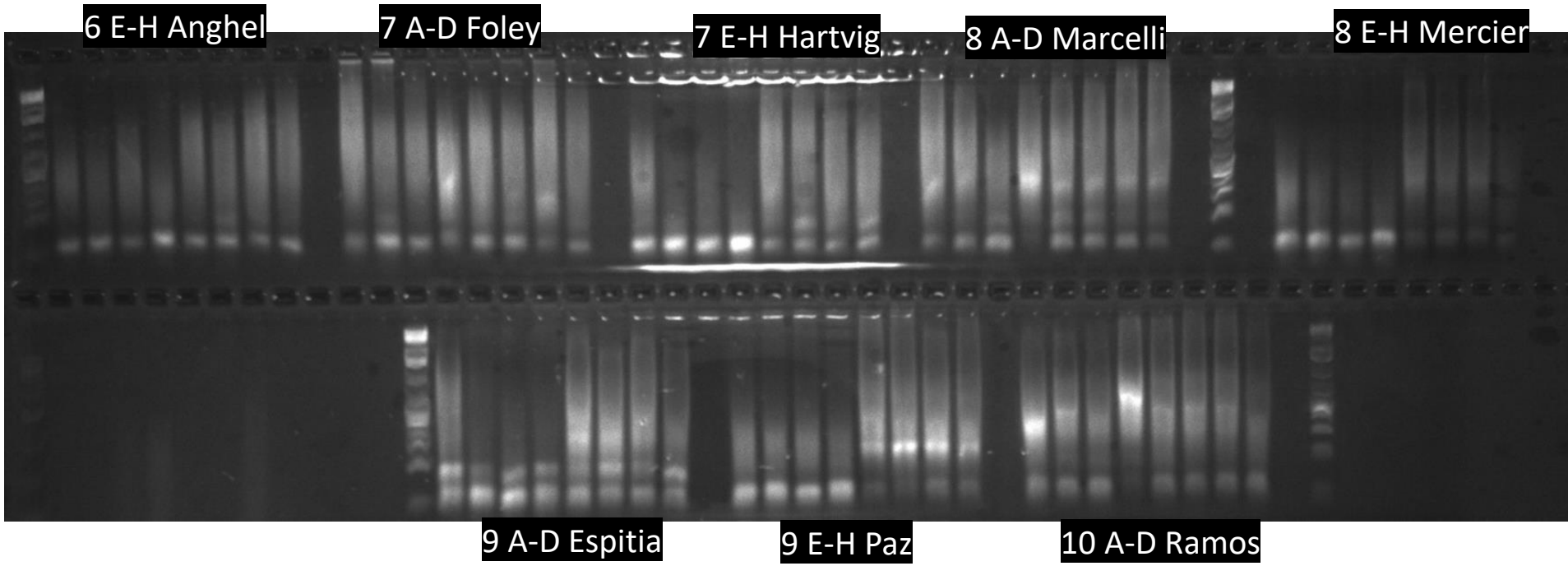
- Design 1 (XbaI, NheI, EcoRI-HF)
- Design 2 (MspI, ClaI, BamHI-HF)
- Compared 4 samples for each project with each design.
- Design 1 always on left, Design 2 always on right

Gel 1



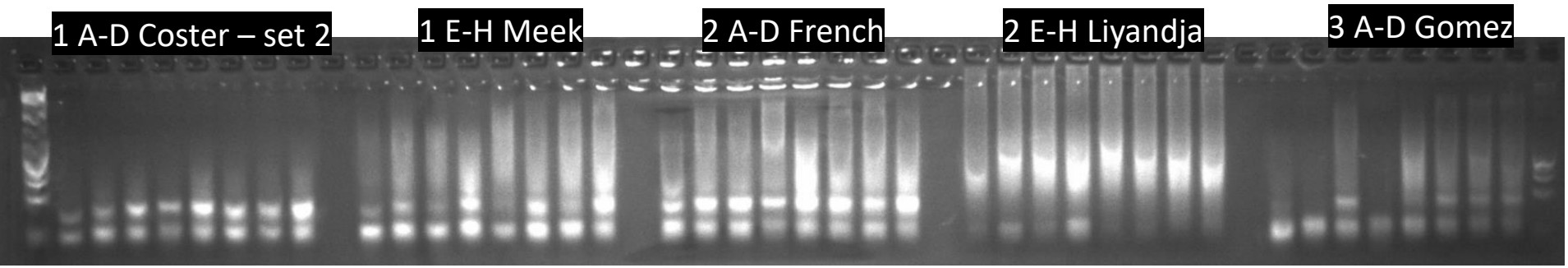
Sandra Hoffberg

Gel 2

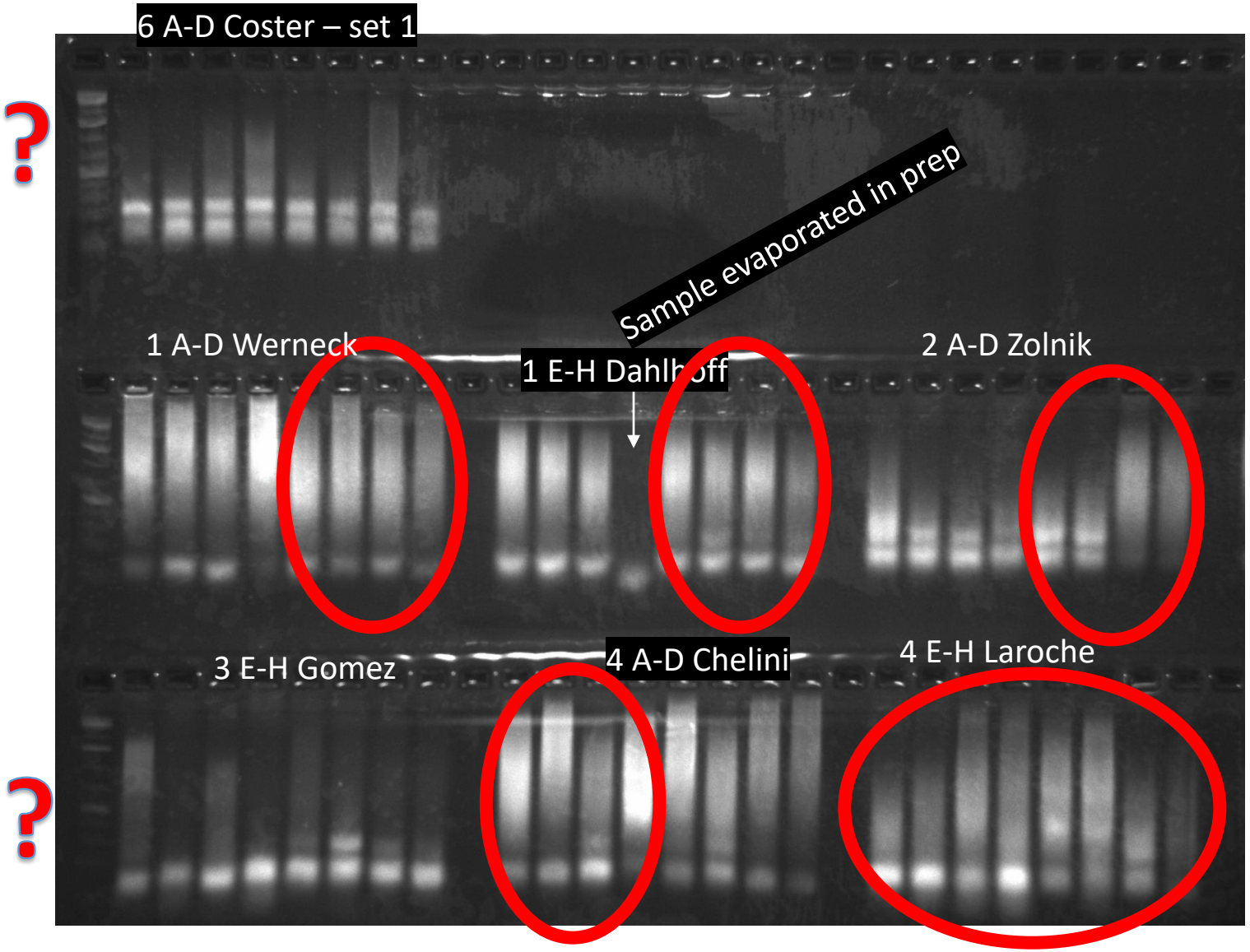


Sandra Hoffberg

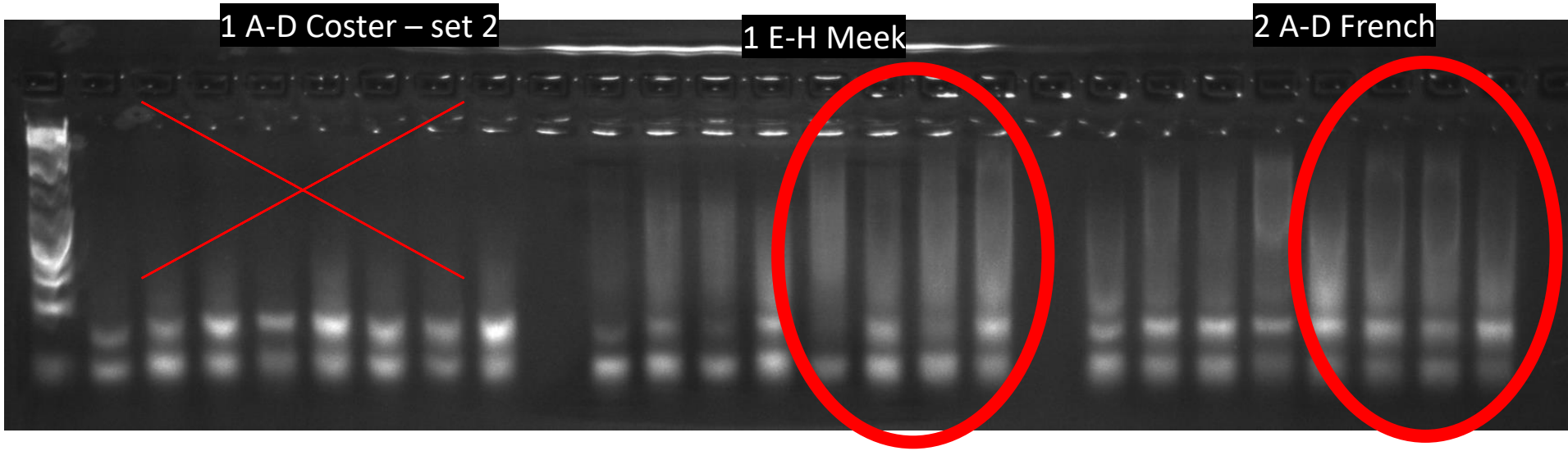
Gel 3



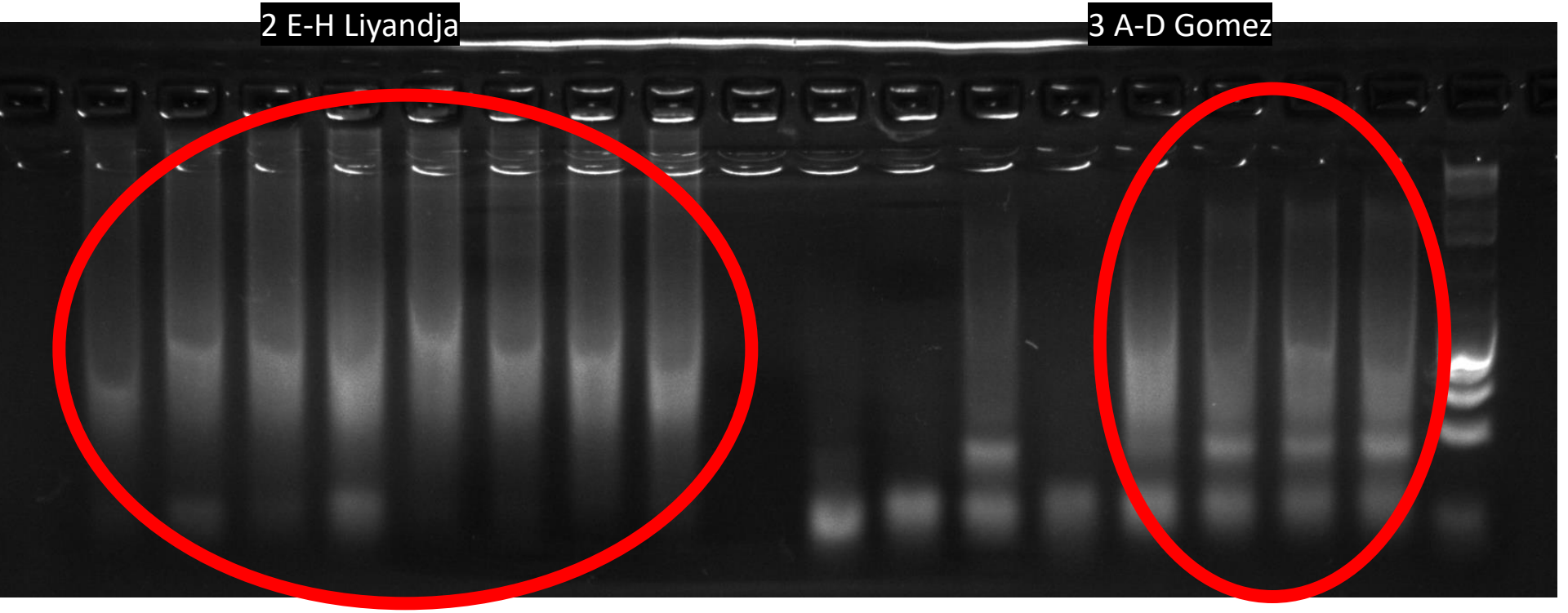
Sandra Hoffberg



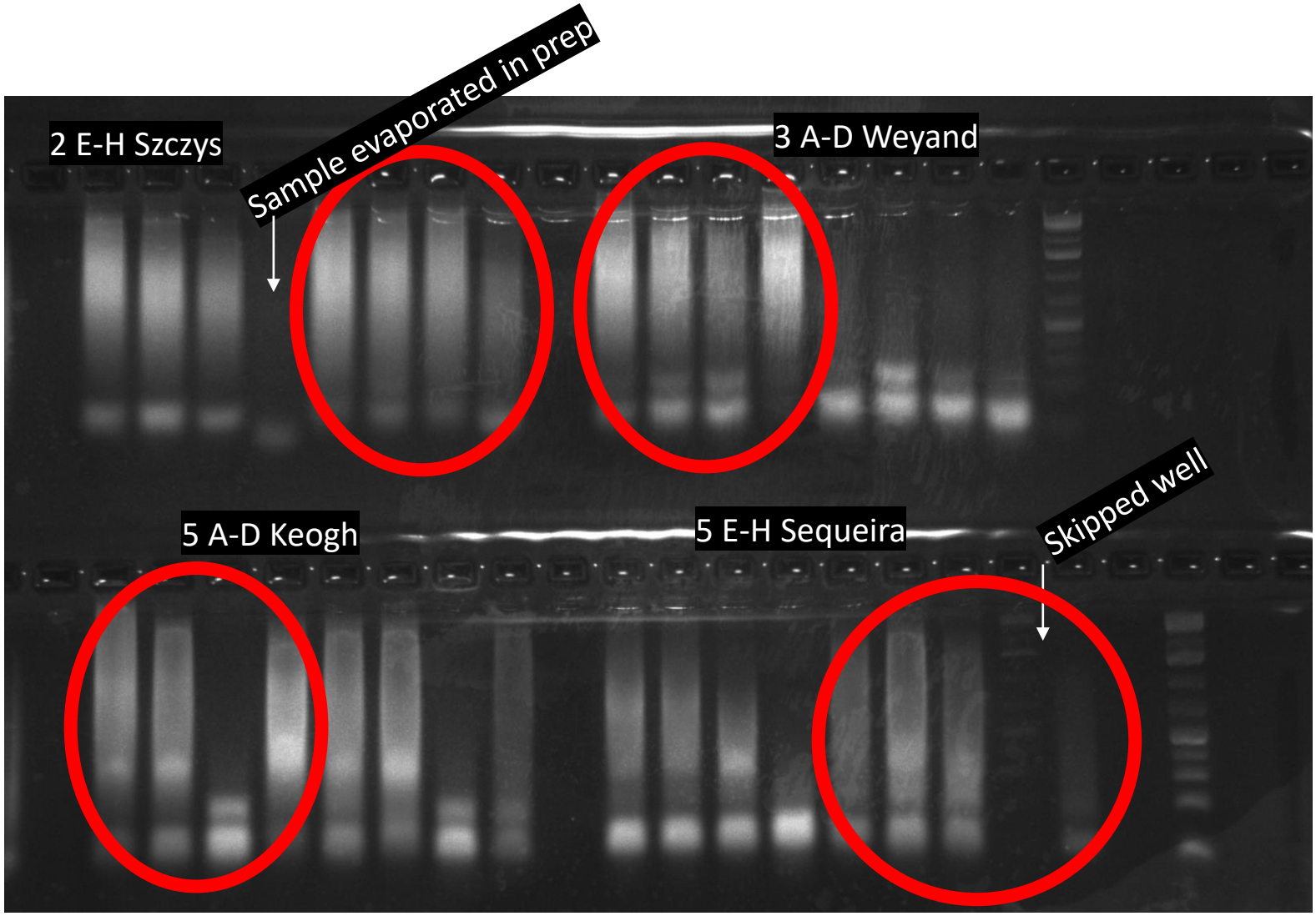
Sandra Hoffberg



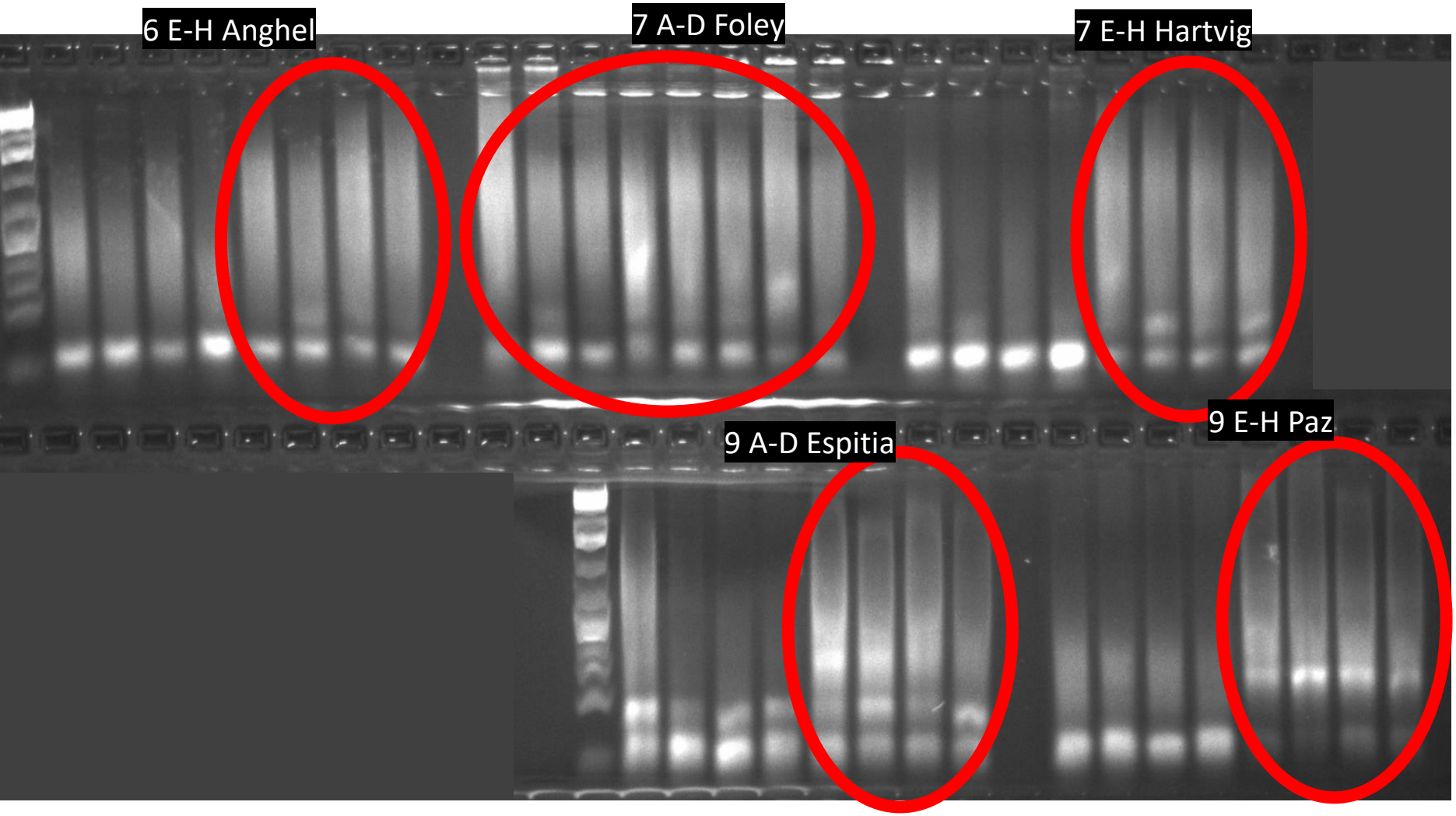
Sandra Hoffberg



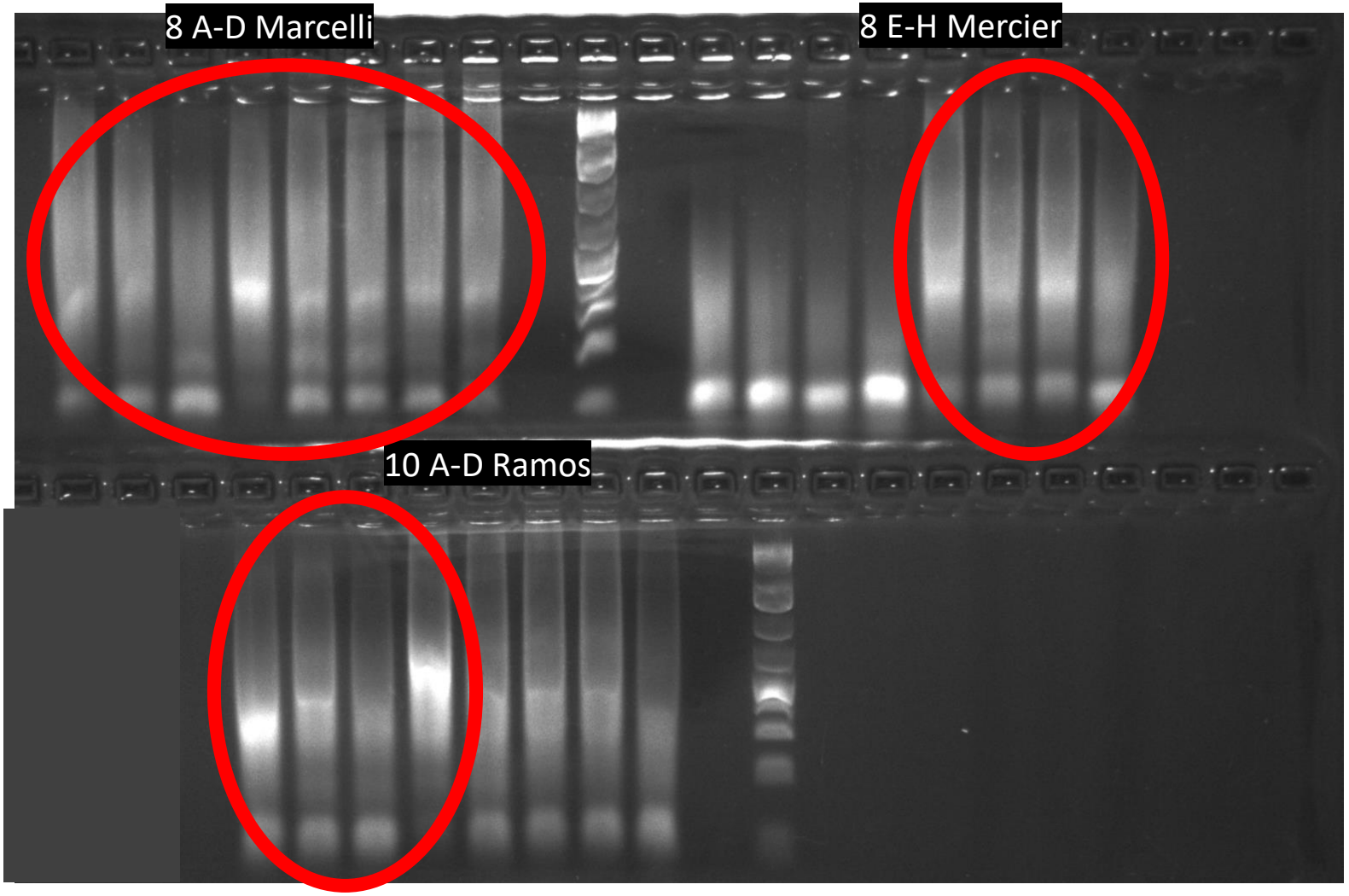
Sandra Hoffberg



Sandra Hoffberg



Sandra Hoffberg



Sandra Hoffberg

Groups

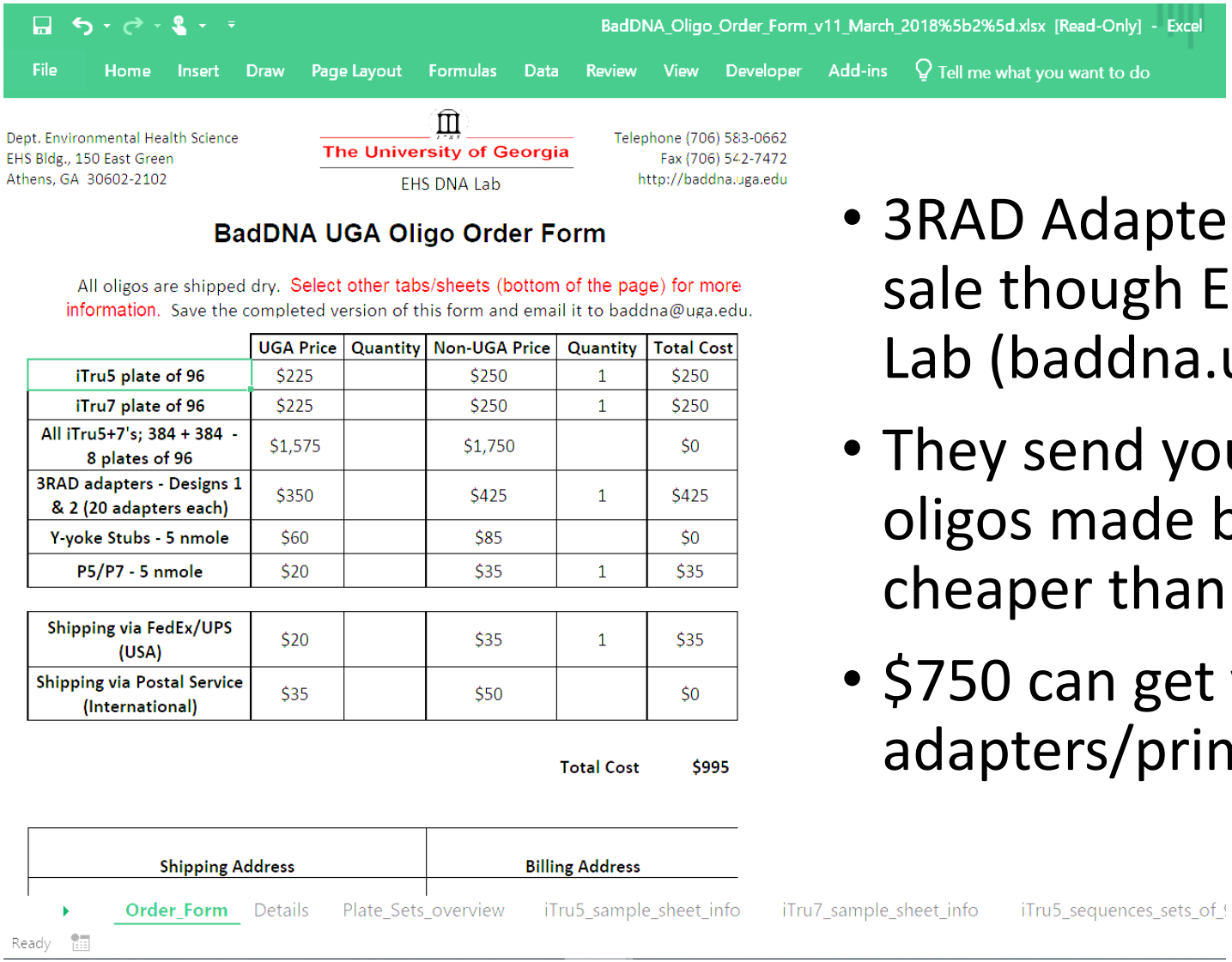
Group	Enzyme combo	Name	Number
1	1	Courtney Weyand	5
1	1	Marie Claire Chelini	7
1	1	Tobit Liyandja	21
2	1	Giovanni Tolentino Ramos	22
2	1	Katie Foley	11
2	1	Stephanie Coster	13
3	1	Sean Keogh	9
3	1	Monica Marcelli	15
3	1	Robert Laroche	8
4	2	Camilo Calderon Acevedo	24
4	2	Elizabeth Dahlhoff	2
4	2	Andrea Paz	18
5	2	Ioana Anghel	12
5	2	Juan Diego Gaitan Espitia	20
5	2	Andrea Sequeira	10
6	2	Ida Hartvig	14
6	2	Connor French	19
6	2	Matias Gomez	6
7	2	Patricia Szczys	4
7	2	Jared Meek	17
7	2	Christine Zolnik	3
8	2	Kathryn Mercier	16
8	2	Fernanda Werneck	1
8	2	Rilquer Mascarenhas	25
8	2	Alexis Earl	26

DNA digestion reaction will include

- Enzymes
- CutSmart buffer
- up to 100ng DNA
- Water
- R1 and R2 adapters

- Incubate 37°C for 1 hour

Easily purchase 3RAD adapters



BadDNA_Oligo_Order_Form_v11_March_2018%5b2%5d.xlsx [Read-Only] - Excel

File Home Insert Draw Page Layout Formulas Data Review View Developer Add-ins Tell me what you want to do

Dept. Environmental Health Science
EHS Bldg., 150 East Green
Athens, GA 30602-2102

The University of Georgia
EHS DNA Lab

Telephone (706) 583-0662
Fax (706) 542-7472
http://baddna.uga.edu

BadDNA UGA Oligo Order Form

All oligos are shipped dry. Select other tabs/sheets (bottom of the page) for more information. Save the completed version of this form and email it to baddna@uga.edu.

	UGA Price	Quantity	Non-UGA Price	Quantity	Total Cost
iTru5 plate of 96	$225		$250	1	$250
iTru7 plate of 96	$225		$250	1	$250
All iTru5+7's; 384 + 384 - 8 plates of 96	$1,575		$1,750		$0
3RAD adapters - Designs 1 & 2 (20 adapters each)	$350		$425	1	$425
Y-yoke Stubs - 5 nmole	$60		$85		$0
P5/P7 - 5 nmole	$20		$35	1	$35

Shipping via FedEx/UPS (USA)	$20		$35	1	$35
Shipping via Postal Service (International)	$35		$50		$0

Total Cost $995

Shipping Address	Billing Address

Order_Form Details Plate_Sets_overview iTru5_sample_sheet_info iTru7_sample_sheet_info iTru5_sequences_sets_of_

Ready

- 3RAD Adapters are for sale though EHS DNA Lab (baddna.uga.edu)
- They send you dried oligos made by IDT but cheaper than IDT prices
- $750 can get you all adapters/primers

Easily purchase 3RAD adapters

Dept. Environmental Health Science
EHS Bldg., 150 East Green
Athens, GA 30602-2102

The University of Georgia
EHS DNA Lab

Telephone (706) 583-0662
Fax (706) 542-7472
http://baddna.uga.edu

How to Handle Plates with 3RAD v2 Adapter Aliquots

When you receive the adapters, there is 1.625 nmol of each oligo pair dried in each well, but they are **NOT** annealed. You will need to reconstitute them to the appropriate volume (64 µL -> 25 µM), anneal them, then dilute again (to 5 µM) & aliquot them.

Liquid for reconstitution & annealing (10 mM Tris pH 8, 0.1 mM EDTA, 100 mM NaCl):
For 50 mL of salty TLE, add the following to a 50mL conical:
40 mL dH_2O
500 µL 1M Tris pH 7.5 to 8
20 µL 0.5M EDTA pH 8
1 mL of 5 M NaCl (or 5 mL of 1M NaCl)
Fill with distilled water to 50 mL mark.

Protocol:

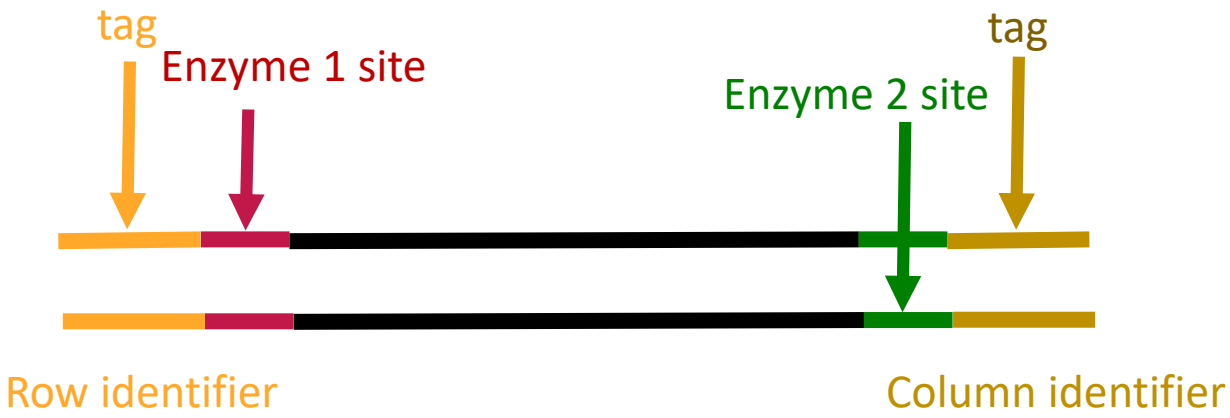
1) Centrifuge the dry plates to get all the adapter to the bottom of the wells.

2) To limit contamination, peel back the foil cover from the plate one row at a time to reconstitute.

3) Add 64 µL of the liquid from above to each well.
-Use the pipet tip to help scrap the bottom of the well to dislodge any of the

Ligation

- Use DNA ligase to add ds adaptors with known sequence to DNA
- Must have a different adapter on each end to form complete library molecule
- Use Y-shaped adaptors to prevent adapters binding to each other

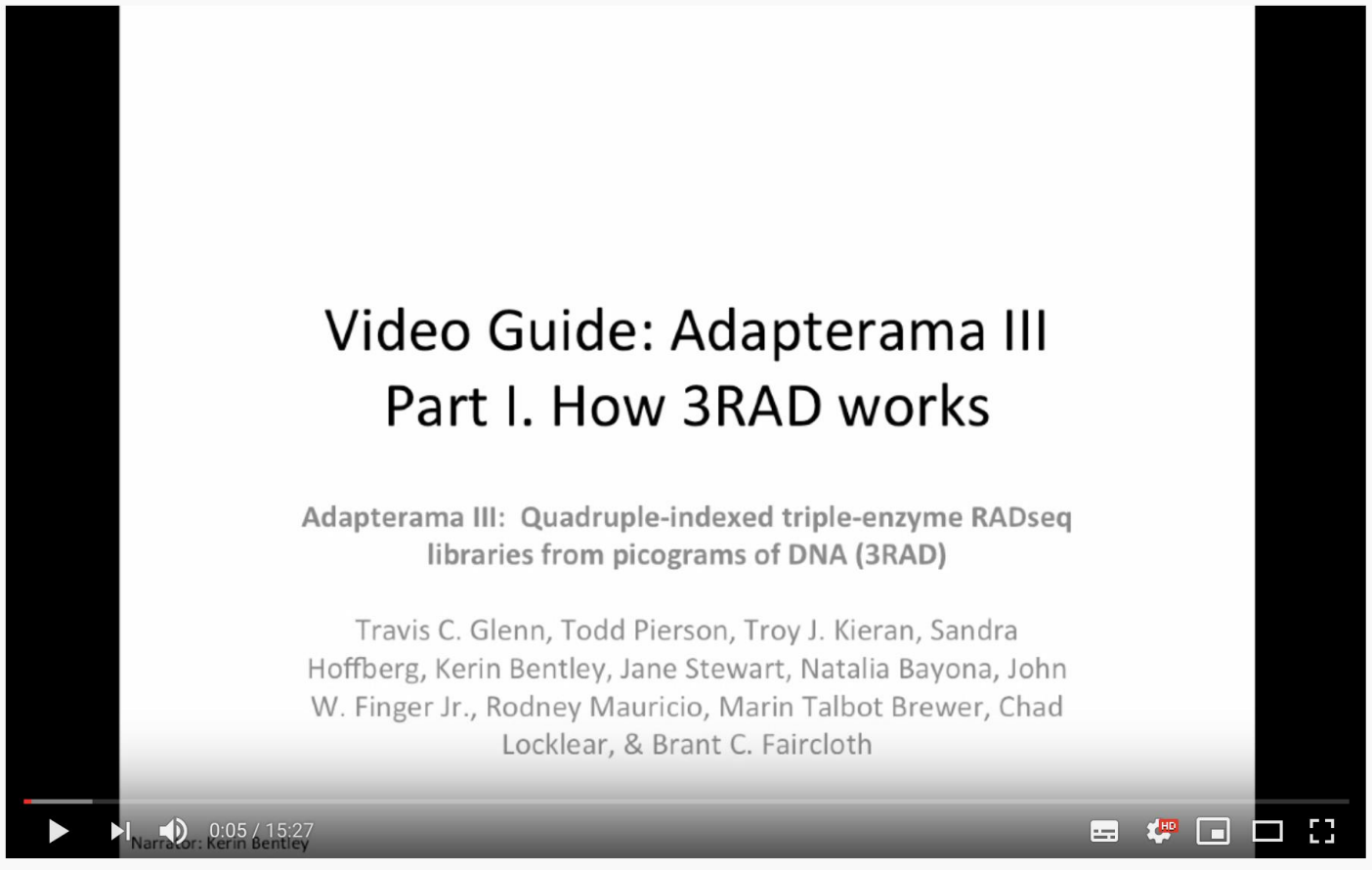


Ligation reaction will include

- Digested DNA and all previous reagents
- rATP (energy)
- Ligase
- Ligase buffer

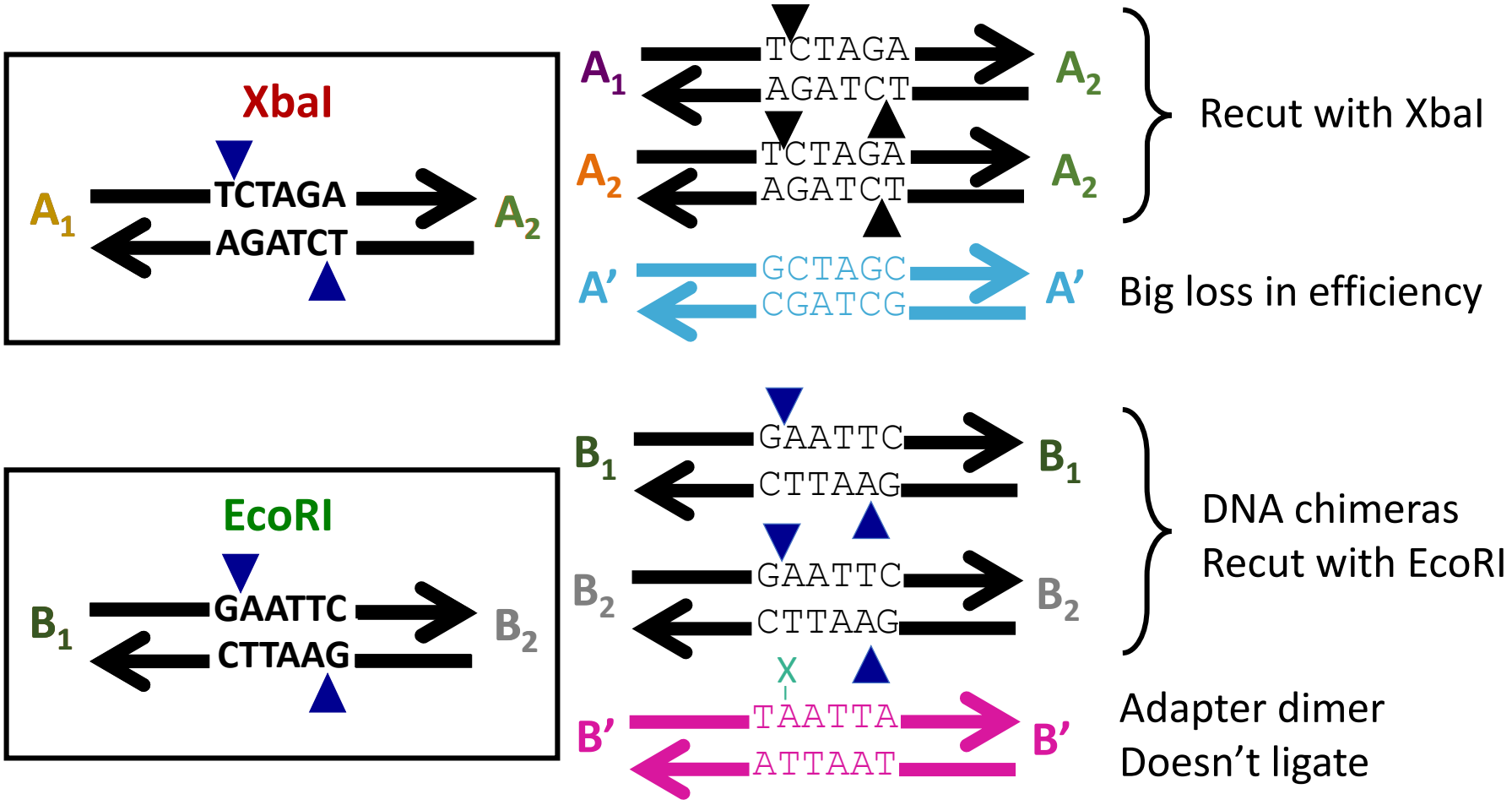
- Immediately add master mix to digested samples.
- We now have all the ingredients to do extra cycles of digestion and ligation in the tube.
- Cycle between 22 ℃ ligation for 20 mins and 37 ℃ digestion for 10 mins 2 times

Why have simultaneous digestion and ligation? The 3rd enzyme

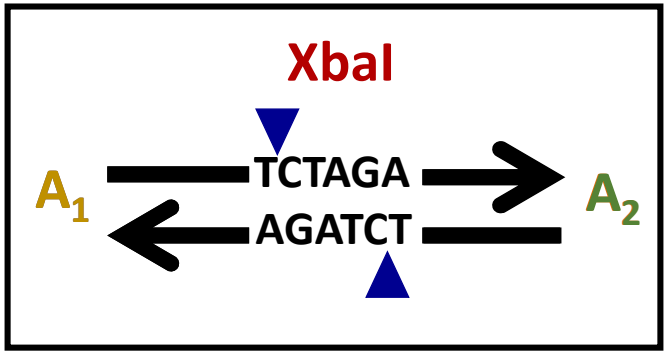


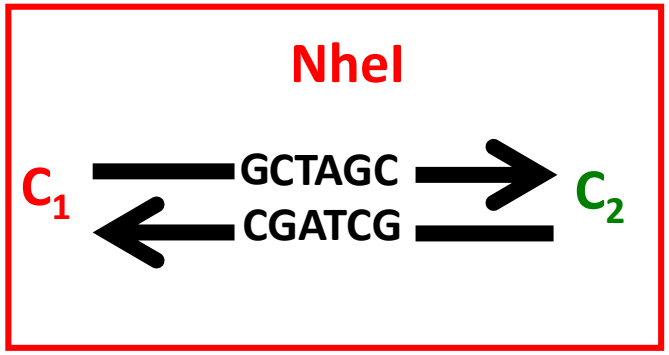
https://www.youtube.com/watch?v=ZOmwOtfP3N4&t=4s

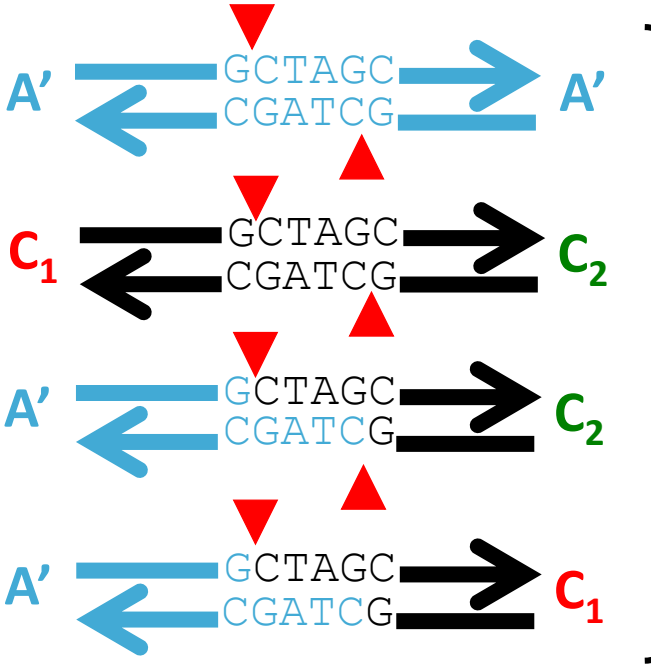
Why have simultaneous digestion and ligation? The 3rd enzyme



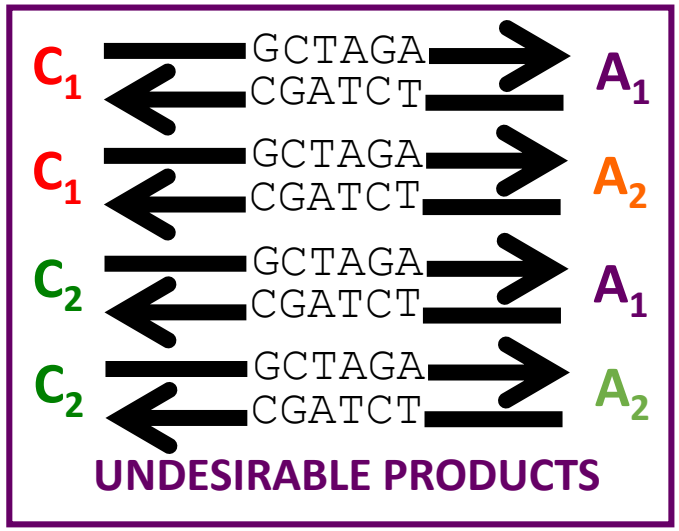
Why have simultaneous digestion and ligation? The 3rd enzyme







Recut by NheI



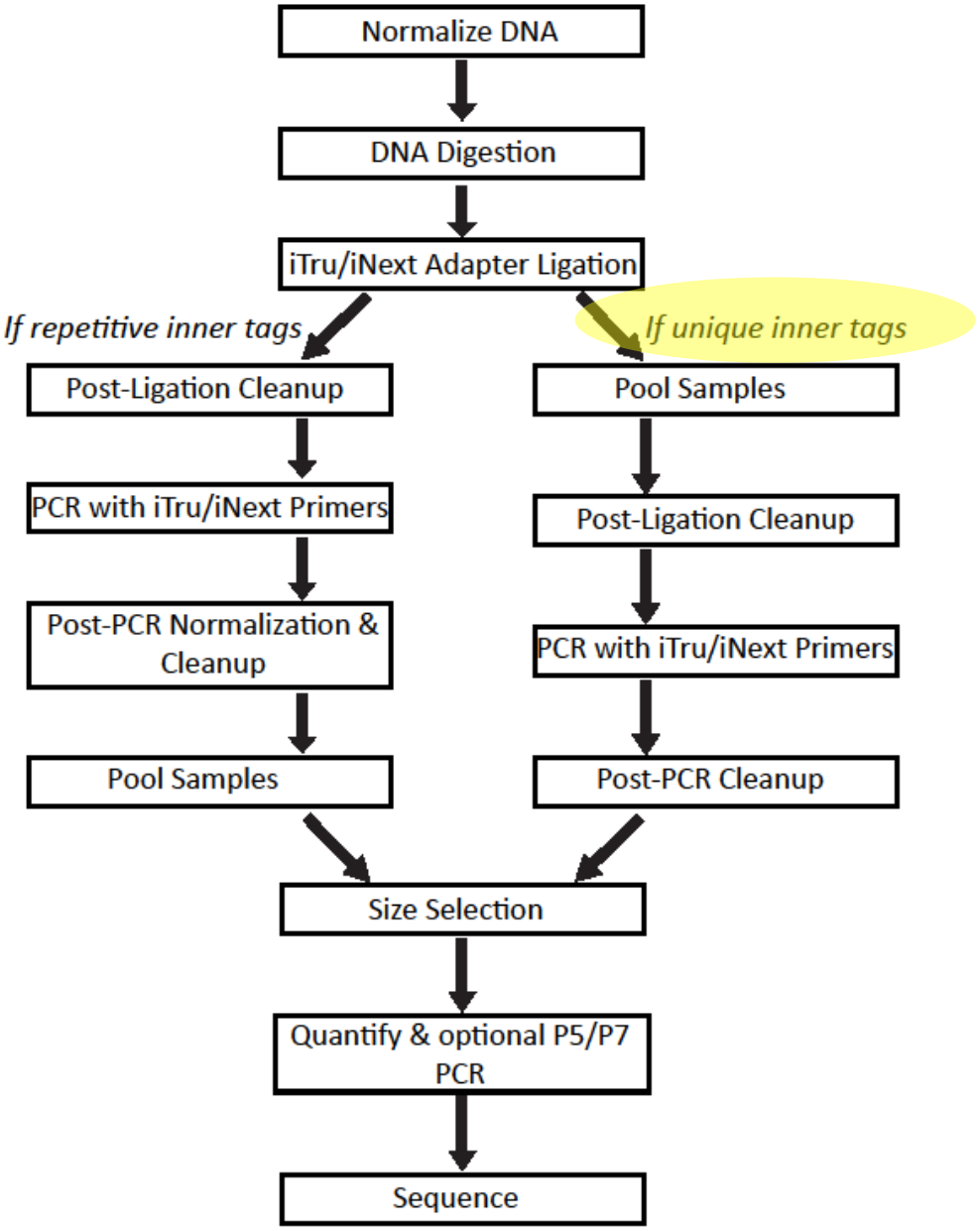
Why have simultaneous digestion and ligation? The 3rd enzyme

- 3rd enzyme increases efficiency – fewer chimeras and adapter dimers.
- Therefore can use low input DNA concentrations.

What is next? Depends.

Easier and cheaper to be on the right side & pool earlier

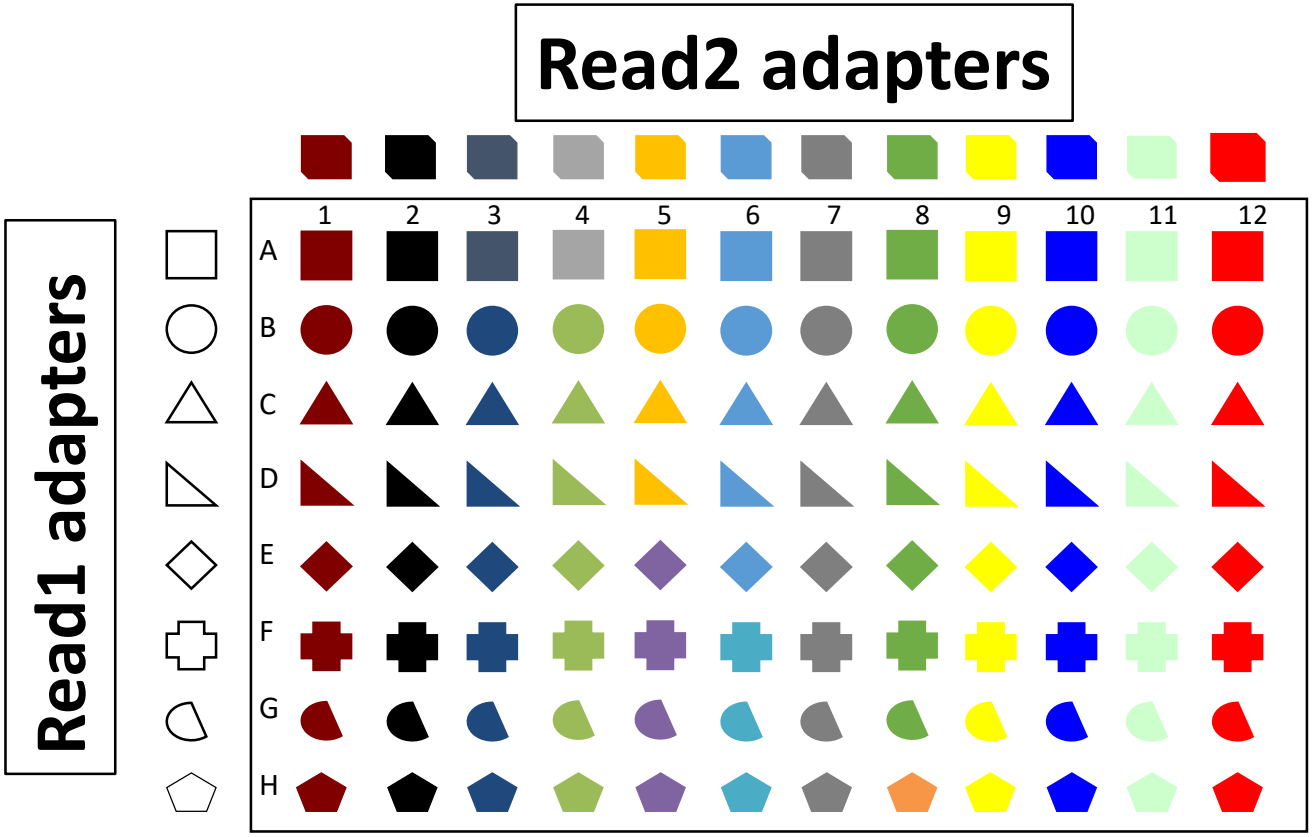
But if you want to ensure proper library prep for each sample before sequencing, go left.



Pool samples

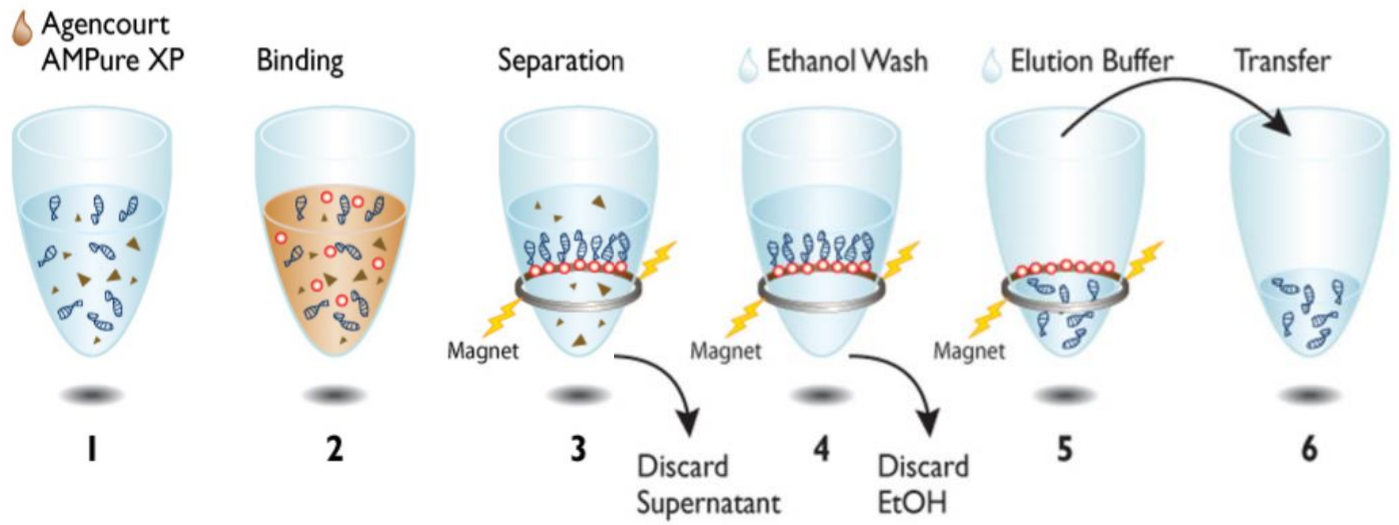
- Using a multichannel pipette, pool by equal volume (amount of DNA was already normalized).

- In 3RAD, you can pool up to 96 samples at this step.



Clean samples

- Remove enzymes, ligase, buffers, and leftover adaptors from the DNA.
- Use magnetic beads (or more expensive column-based methods).

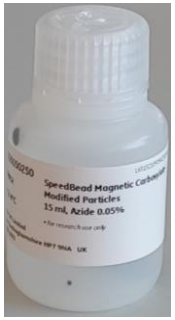


Agencourt AMPure XP PCR Purification Manual, 2016

Money saving tip: dilute SpeedBeads

GE Healthcare Sera-Mag SpeedBeads

- 15 mL bottle: $476.92
- Resuspend 1 mL into 50 mL polyethylene glycol (PEG) and salt
- Makes 15x50 = **750mL**

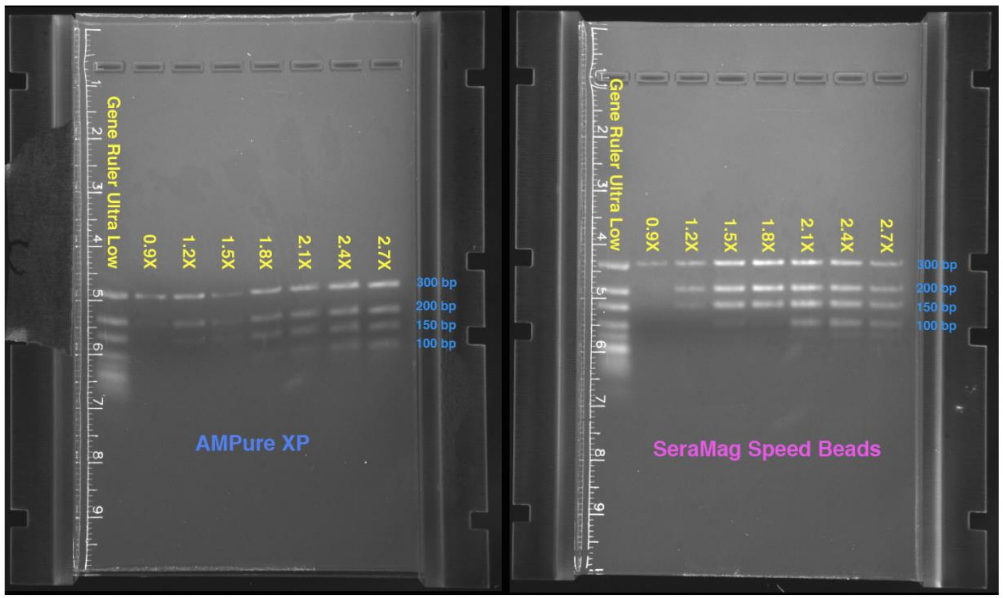


Beckman Coulter AMPURE XP

- **60. mL** bottle: $1,481.55
- Use directly

Clean samples

- Use X as much Speedbead volume as DNA volume
- Volumes of Speedbeads and DNA will affect sizes of DNA fragments you recover
  - PEG causes the negatively-charged DNA to bind with the carboxyl groups on the bead surface
  - Increasing the ratio of PEG (beads) volume to sample volume will increase the efficiency of binding smaller fragments



Clean samples

- Wash 1-2x with 70-80% ethanol.
- Hygroscopic - when opened the ethanol will both evaporate and absorb water over time.
  - Make fresh each time or make 80% ethanol
- Miscible - measuring 70 mL of ethanol and topping off to 100 mL with water will generate ~65% ethanol. Measuring 70 mL ethanol and 30 mL water separately, then combining them will generate ~95 mL of 70% ethanol.

We will stop at this point in the protocol on Saturday

Add sequencing primers

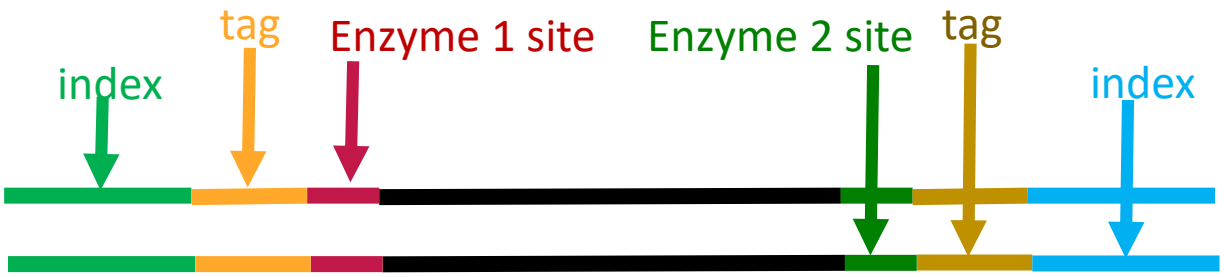
- Do PCR to add oligonucleotide sequences (P5, P7) to the template strands that allow hybridization to the Illumina flow-cell surface -> generate “full-length libraries”
  - Includes index position, called i5 and i7
- Enrich for properly ligated template strands—those that have an adapter at both ends.
- Increase the amount of library available for sequencing.
- Generate enough DNA for accurate quantification.

Adapterama I: Universal stubs and primers for 384 unique dual-indexed or 147,456 combinatorially-indexed Illumina libraries (iTru & iNext)

Travis C. Glenn, Roger A. Nilsen, Troy J. Kieran, Jon G. Sanders, Natalia J. Bayona-Vásquez, John W. Finger Jr., Todd W. Pierson, Kerin E. Bentley, Sandra L. Hoffberg, Swarnali Louha, Francisco J. García-De León, Miguel Angel Del Río-Portilla, Kurt D. Reed, Jennifer L. Anderson, Jennifer K. Meece, Samuel E. Aggrey, Romdhane Rekaya, Magdy Alabady, Myriam Bélanger, Kevin Winker, Brant C. Faircloth

doi: https://doi.org/10.1101/049114

Add sequencing primers



The problem with PCR

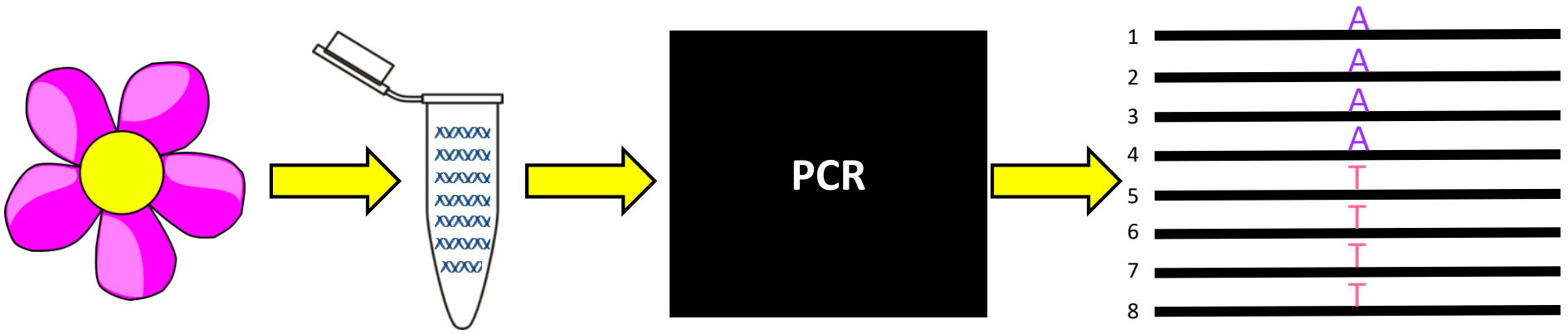
A/T heterozygote

Biological Sample

Whole genomic DNA

Library Preparation

Sequence Reads



The problem with PCR

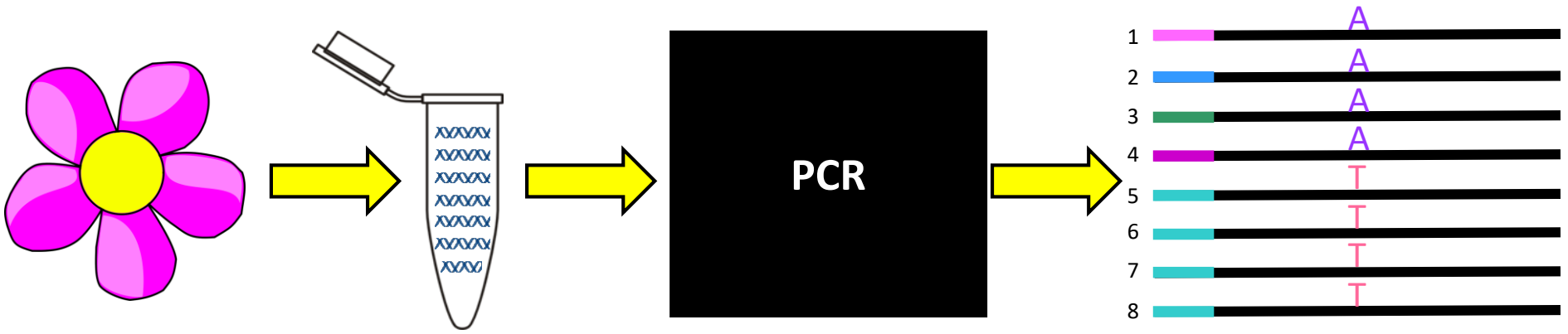
A homozygote

Biological Sample

Whole genomic DNA

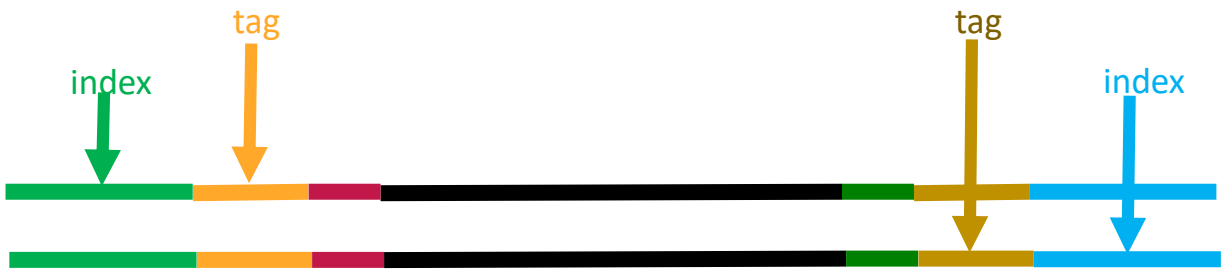
Library Preparation

Sequence Reads



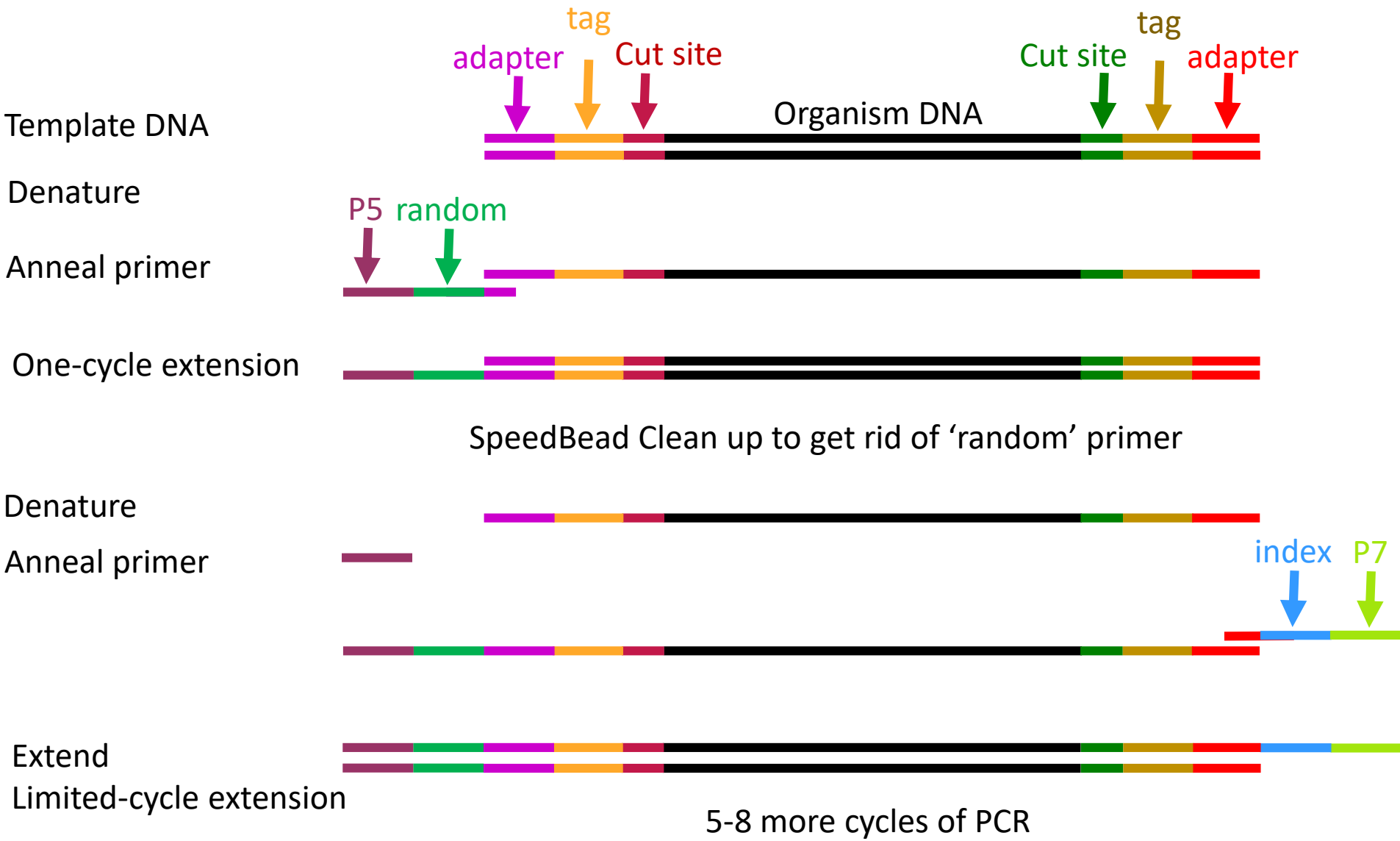
Therefore, add a PCR duplicate identifier pre-PCR

- 8 cycle PCR, with a clean up step in the middle

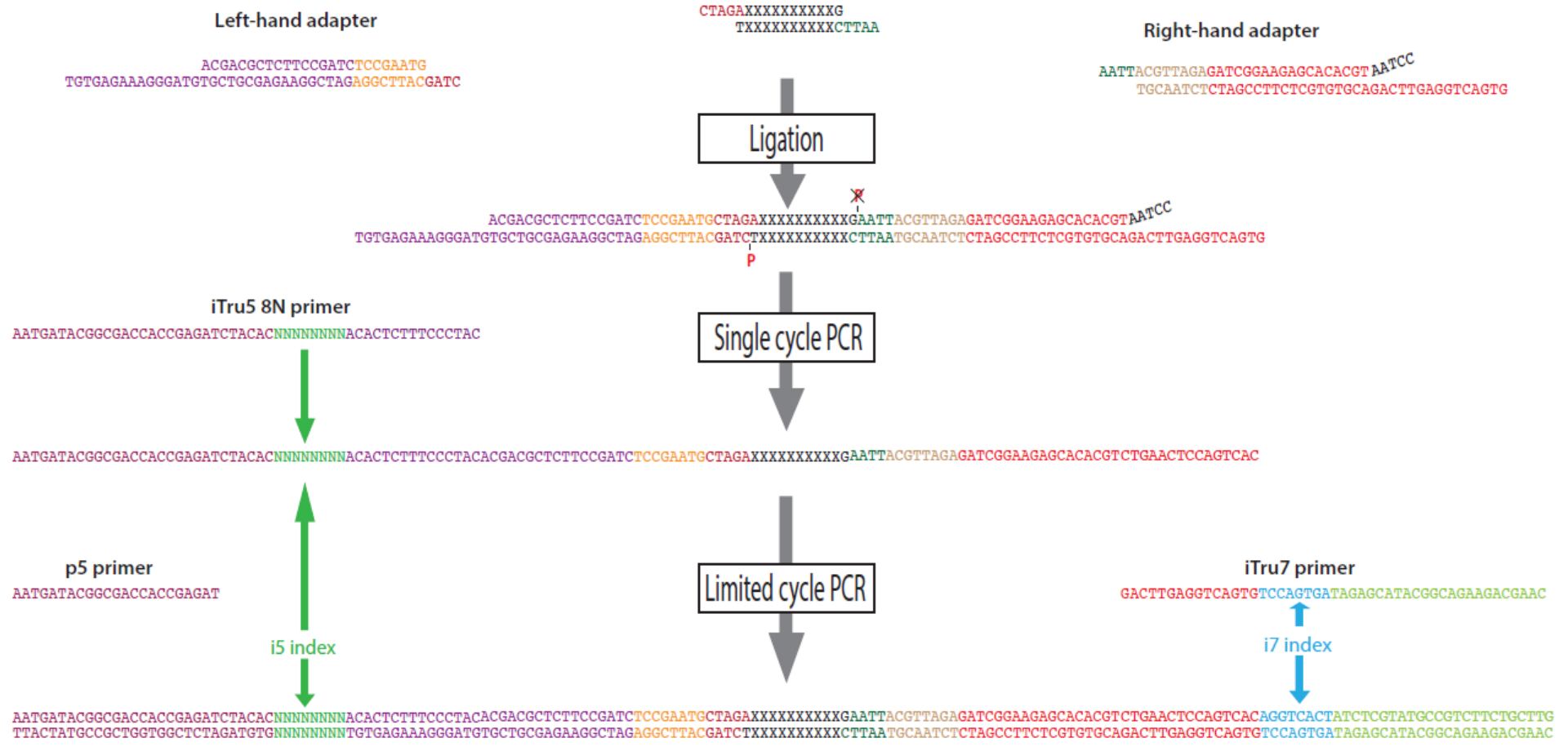


4^8=65,536

How to add PCR duplicate identifier



How to add PCR duplicate identifier



Size selection

- Further reduce the part of the genome you sequence.
- Get sequences of appropriate size on Illumina machine – fragments over 1kb cluster less efficiently.
- Important to be consistent to assess the same loci across samples.

Size selection

- PippinPrep (Sage Science).
- Dual size selection with beads.
- Extract fragments from agarose gel.

PippinPrep

- Sets of 96 samples
- 500 “tight” size selection: +/- 10%
- 550 “broad” size selection: +/- 12.5%
- Running a gel with standards so the machine detects fragments sizes, and once your target is detected it directs them into a separate well (elution well).
- Then, you can recover the fragments of the desired size.

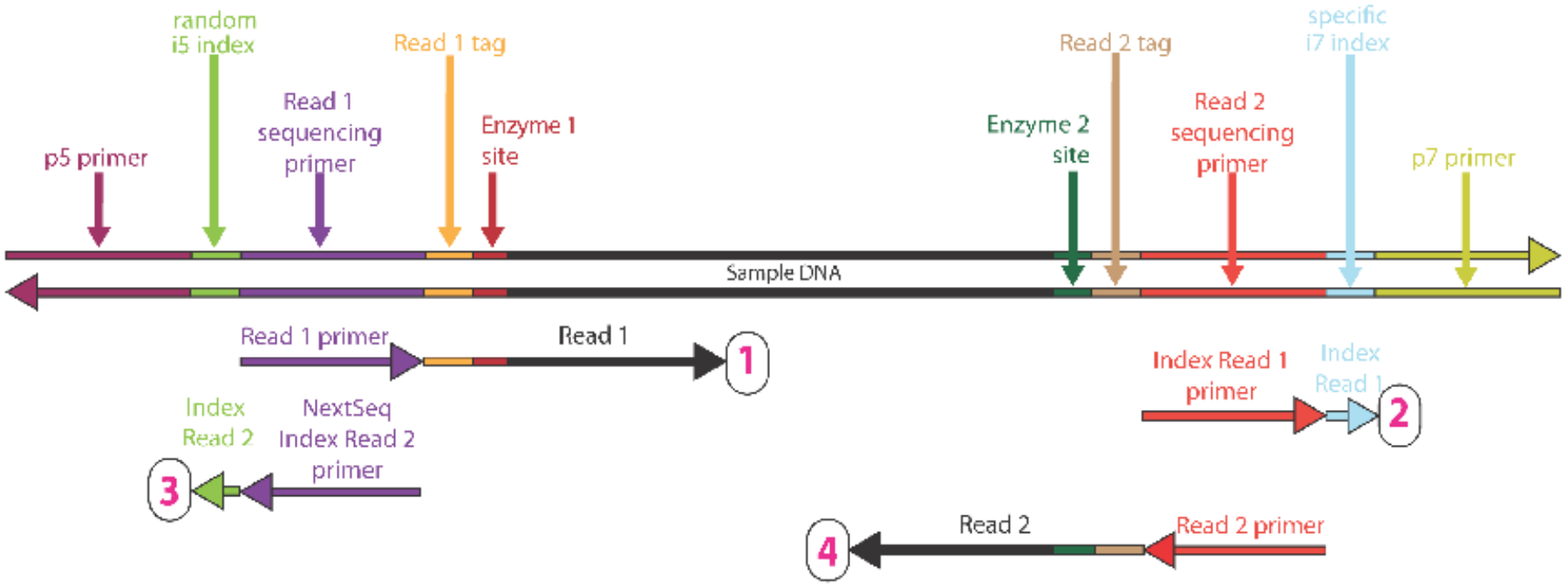
Pippin Prep video

https://www.youtube.com/watch?v=4lg_6qInvbE

After PippinPrep

- You can perform a P5 / P7 PCR after size selection to increase the amount of library to pool for sequencing.
- If you do so, purify with Speed Beads afterwards.

Complete library molecule



Quantify DNA

- Accurate quantification is vital to pool equal amounts of libraries
  - Too low cluster density reduces yield & increases per-base cost of sequencing
  - Too high cluster density reduces yield due to cluster overlap

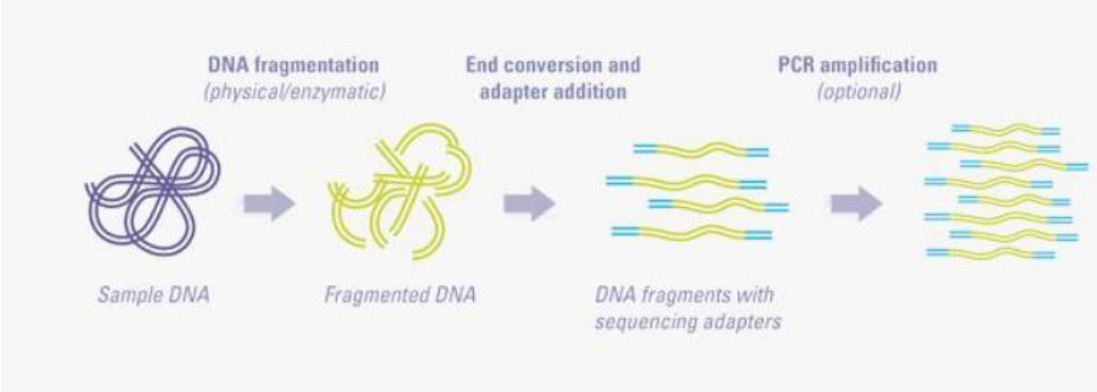
Agilent Bioanalyzer

Invitrogen Qubit

Money saving tips

- Choose cheap enzymes (check our enzyme designs).
- Multiplex samples early.
- Dilute SPRI beads (or homebrew Speedbeads).
- Share adaptors.
- Share sequencing runs by the use of different iTrus (check our primers and diversity calculator).
- Research prices at sequencing facilities.

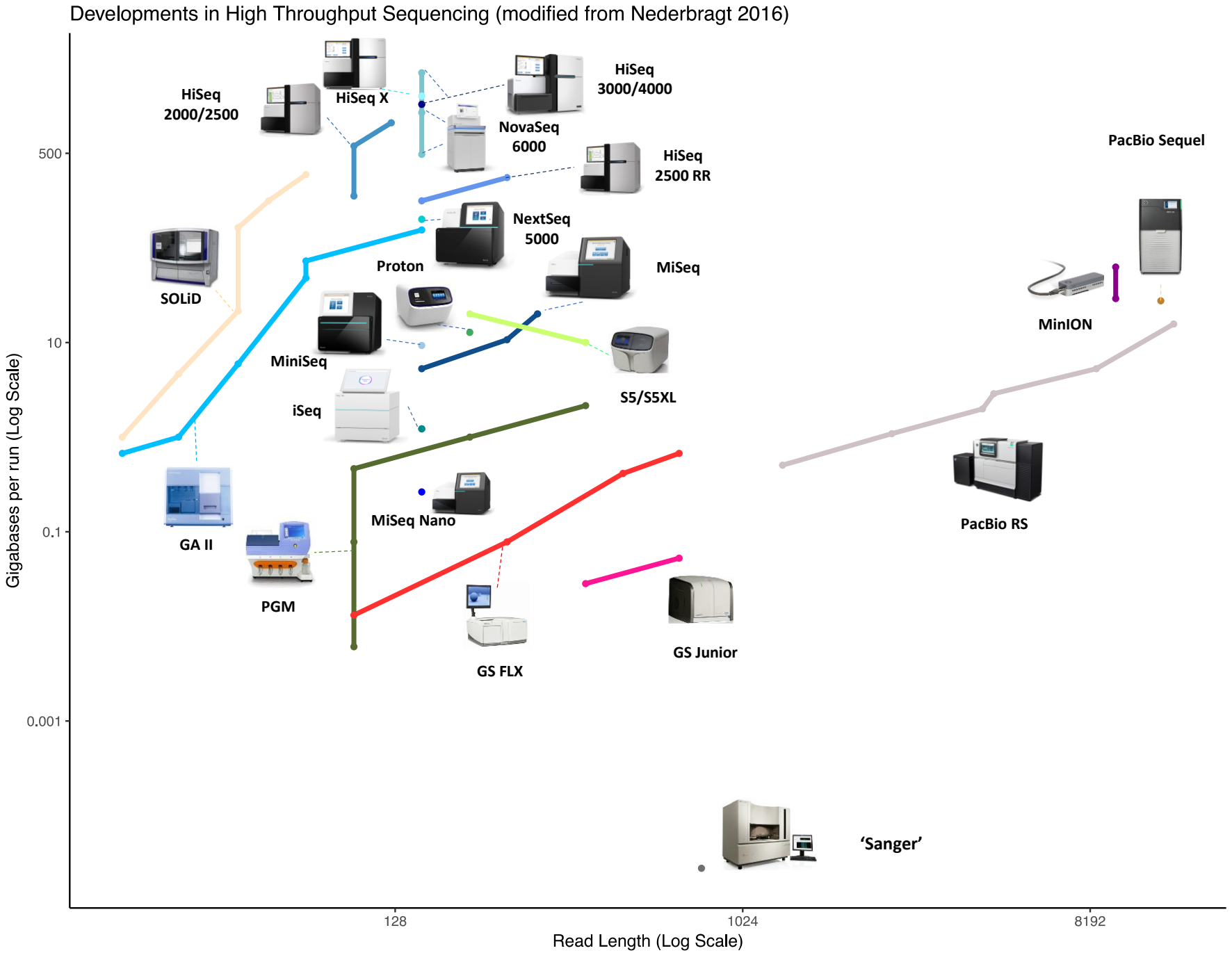
Library prep is designed for Illumina sequencing



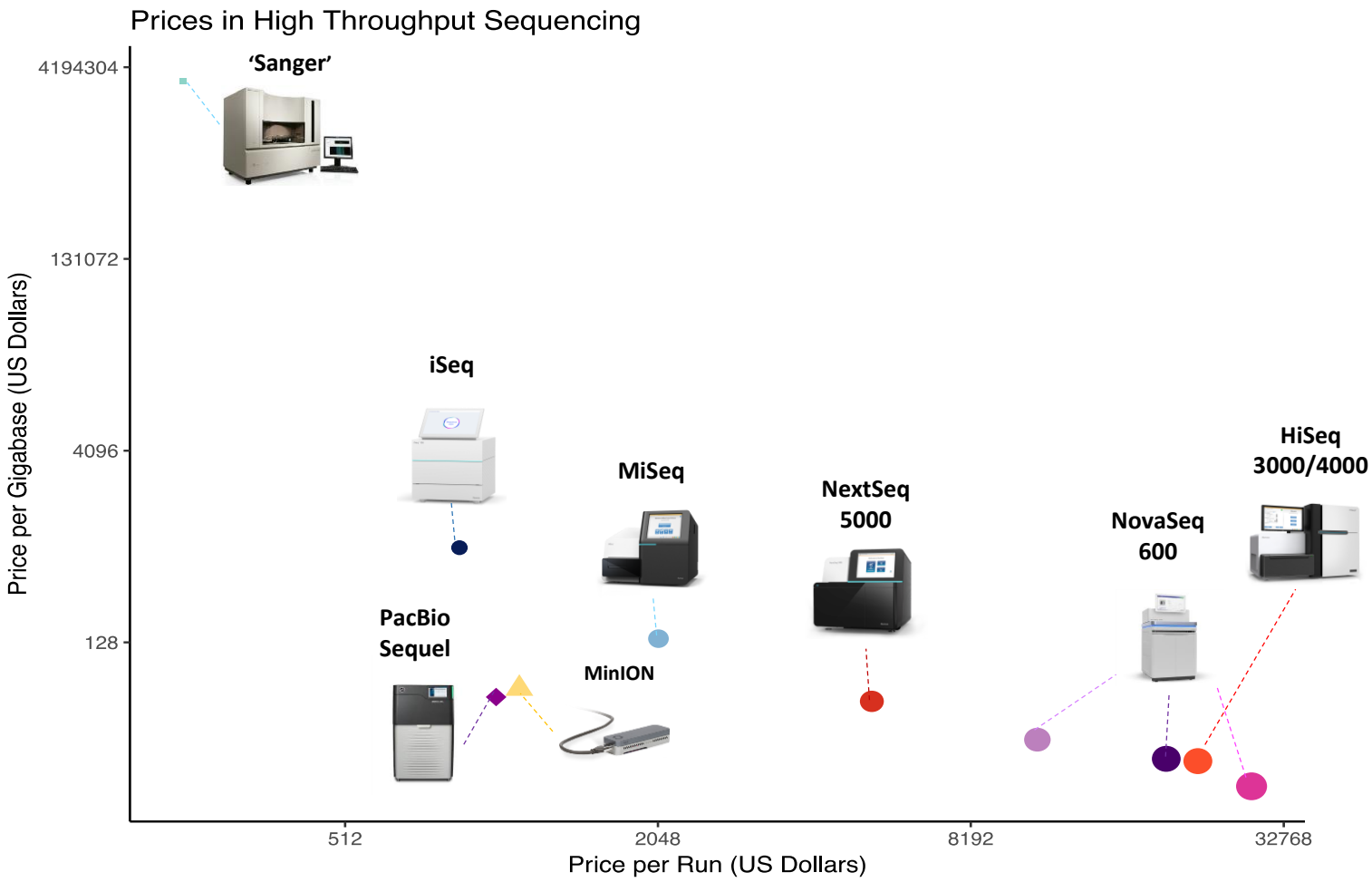
Preparing for sequencing

- If you are sending your samples to a sequencing center, you generally submit >20 µL of pooled DNA at ≥10 nM (check our pooling guide).
- Sequencing facility will do QC on a Bioanalyzer to quantify DNA and see size distribution.
- They will mix DNA with PhiX control (1-5%) and denature DNA before loading onto machine.

Meet the Illumina sequencers

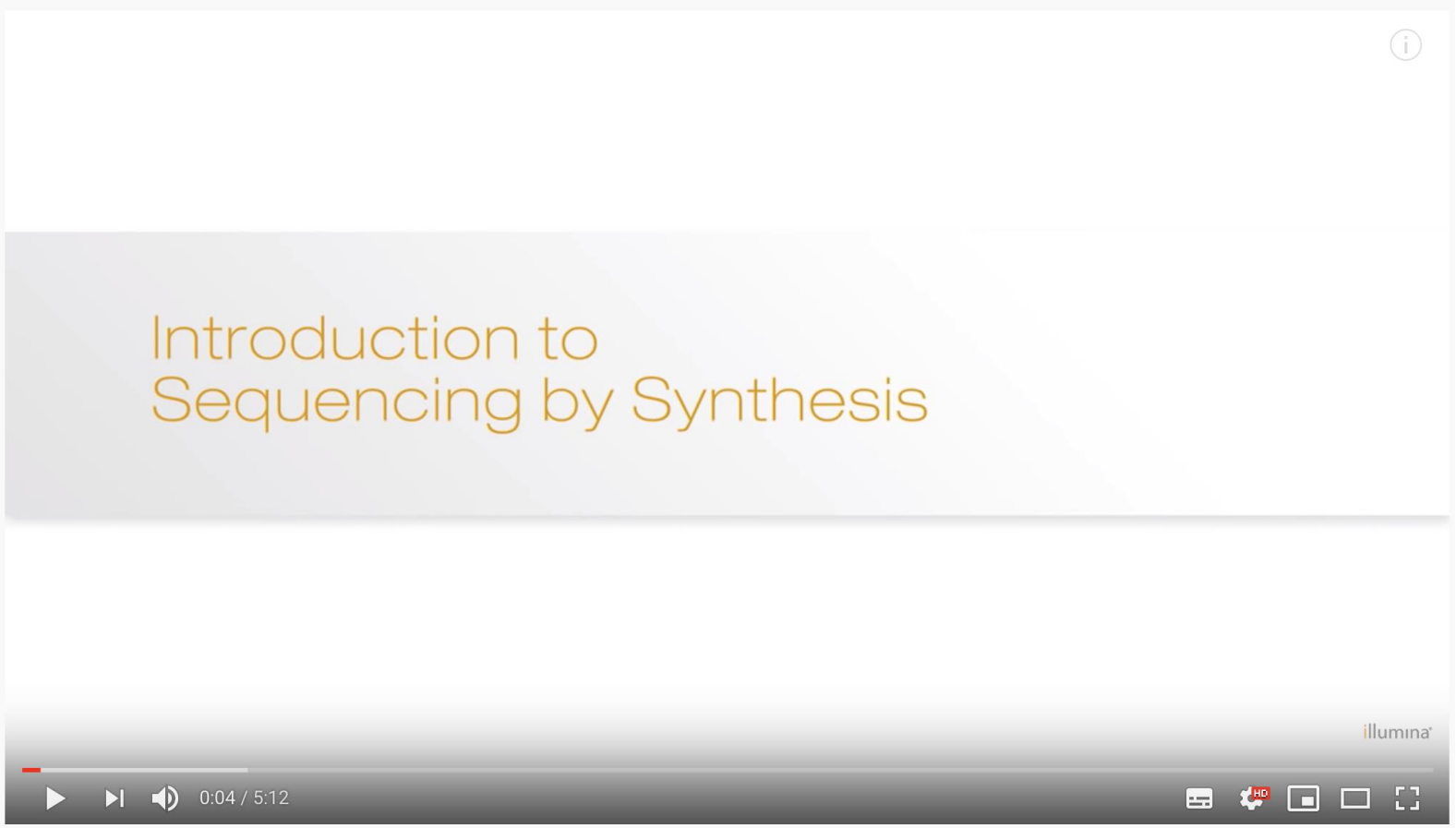

Developments in High Throughput Sequencing (modified from Nederbragt 2016)
HiSeq 2000/2500
HiSeq X
HiSeq 3000/4000
NovaSeq 6000
HiSeq 2500 RR
PacBio Sequel
NextSeq 5000
Proton
MiSeq
SOLiD
MinION
MiniSeq
S5/S5XL
iSeq
PacBio RS
MiSeq Nano
GA II
PGM
GS FLX
GS Junior
'Sanger'
500
10
0.1
0.001
Gigabases per run (Log Scale)
128
1024
8192
Read Length (Log Scale)

Next-Generation Sequencing: Capabilities vs. Cost



Shape size reflects output (Gigabases per run)

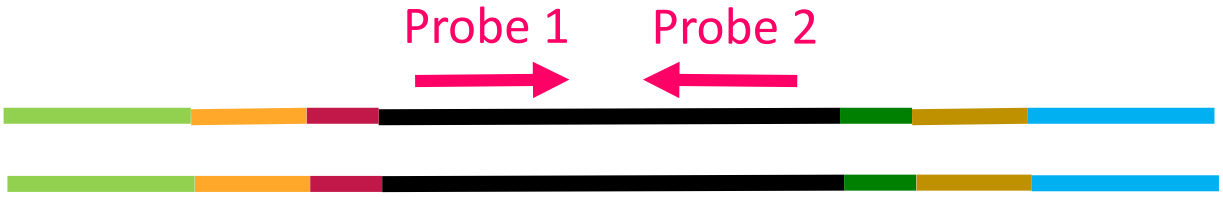
Illumina Sequencing



https://www.youtube.com/watch?v=fCd6B5HRaZ8

RADcap

- Uses baits/probes to capture RADseq loci.
- Select loci to be polymorphic and present in most samples.
- Decreases the number of loci obtained.
- Decreases amount of sequencing necessary → decreases cost.
- Multiplexing and PCR duplicate identification possible.



Hoffberg et al. 2016

RADcap

https://arborbiosci.com/

On 24-96 samples

- Do 3RAD, sequence libraries.
- Analyze reads to select loci.
- Arbor Biosciences will design and synthesize baits.

On the rest of your samples:

- Do 3RAD.
- Instead of size selection, do sequence capture.
- Sequence libraries.

Hoffberg et al. 2016