Planning a successful target enrichment project

Alison M. Devault, Ph.D.

Product Manager, MYbaits & NGS

MYcroarray

Smithsonian NMNH

March 15th, 2016

We are **MYcroarray®**, a scientist-owned biotech company specializing in custom affordable microarrays, oligo libraries, and capture kits for target enrichment. We are committed to providing personalized customer service & innovative solutions for your unique applications.

MYbaits® – Target enrichment for NGS

MYreads® – NGS and target capture services

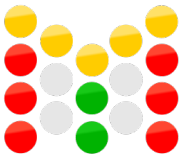
MYtxtl® – Cell-free protein expression kits

MYtags® – Labeled FISH probe libraries

MYlib® – Single- and double-stranded DNA pools

MYdna® – Error-free DNA for gene synthesis

MYcroarray® – Oligonucleotide microarrays



MYcroarray

THE OLIGO LIBRARY COMPANY

Contact information

alison@mycroarray.com

info@mycroarray.com

Skype: alison.mycroarray

1 (734)-998-0751

www.mycroarray.com/mybaits

www.mycroarray.com/myreads

5692 Plymouth Rd
Ann Arbor, MI, USA

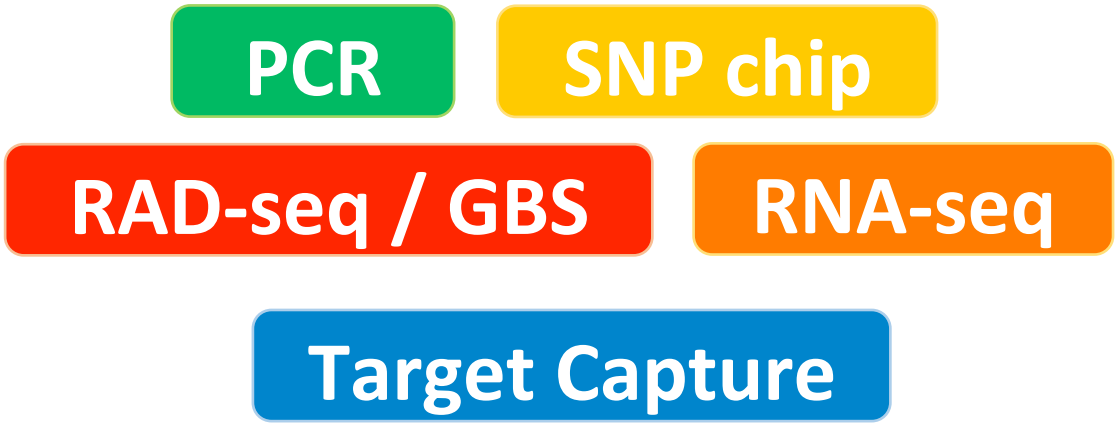
MYcroarray

Target Capture Applications and Bait Design

Ultimate research goal

→ Constraining genotyping/sequencing resources to **informative targets**

→ There are multiple avenues:



MYcroarray

What is target capture?

→ Hybridization-based enrichment of target loci prior to NGS

Your kit includes:
Custom RNA Probes
+ Capture Reagents

MYcroarray

Why target capture? Powerful

- ✓ Highly flexible + scalable
- ✓ Tolerant of bait:target divergence
- ✓ Versatile and reproducible
- ✓ Small-to-large targets feasible
- ✓ Out-of-the-box functionality
- ✓ No specialized equipment
- ✓ Compatible with any type of specimen

MYcroarray

Why target capture? <u>Affordable</u>

→ Highly efficient, highly flexible

96 samples

2Mb

Library preparation	$45	x96	$4,320
MYbaits kit (2Mb, 12 reactions)			$3,000
Enrichment (8-plex)	$55	x12	$660
Sequencing (Illumina PE100, 1 lane)			$2,650

$10,630

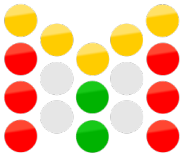
= $**111**/sample

MYcroarray

MYreads® NGS services

Planning your successful project

1. Picking & filtering loci
2. Designing your baits
3. Building your libraries
4. Executing your experiment
5. Sequencing logistics
6. Interpreting your results



MYcroarray

Denature libraries

Anneal blockers

Hybridize to baits

Bind to beads

Wash hybrids

Elute enriched library

1) DNA sequencing library is heat-denatured in the presence of adapter-specific blocking oligonucleotides

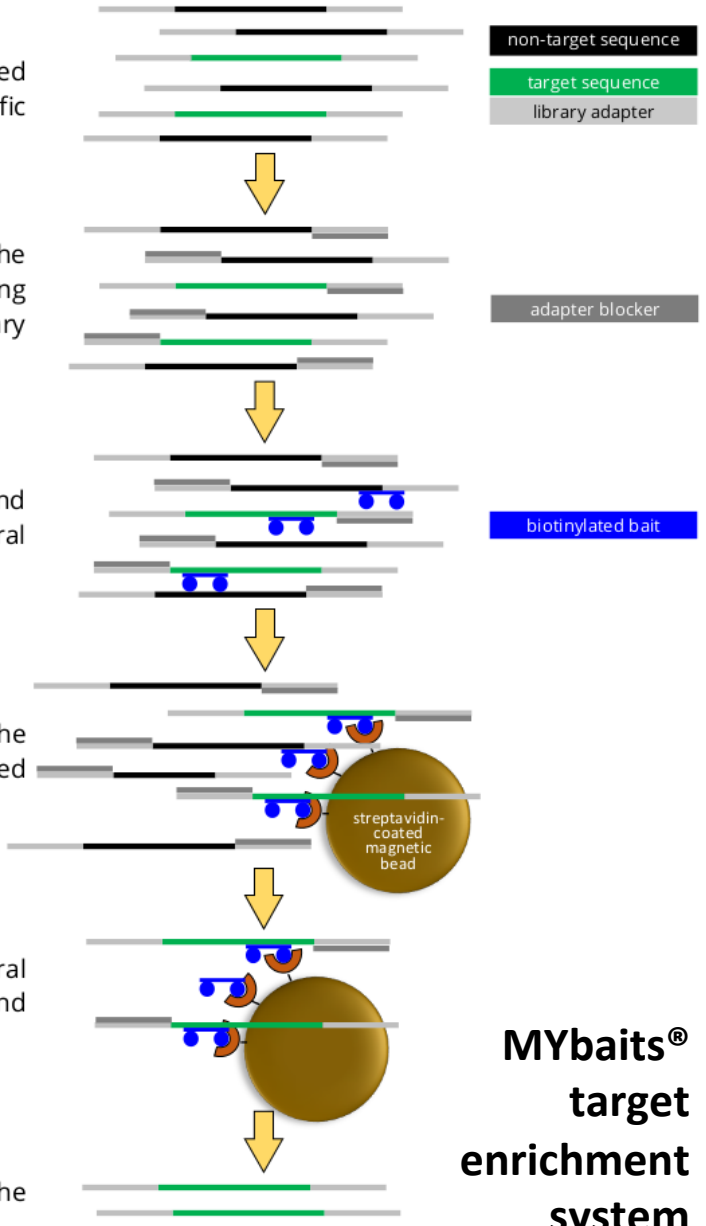
2) Library and blockers are dropped to the hybridization temperature, allowing blockers to hybridize to the library adapters

3) Biotinylated RNA baits are introduced and allowed to hybridize to targets for several hours

4) Bait-target hybrids are pulled out of the solution with streptavidin-coated magnetic beads

5) Beads are stringently washed several times to remove non-hybridized and nonspecifically-hybridized molecules

6) Captured DNA library is released from the beads and amplified



MYbaits® target enrichment system

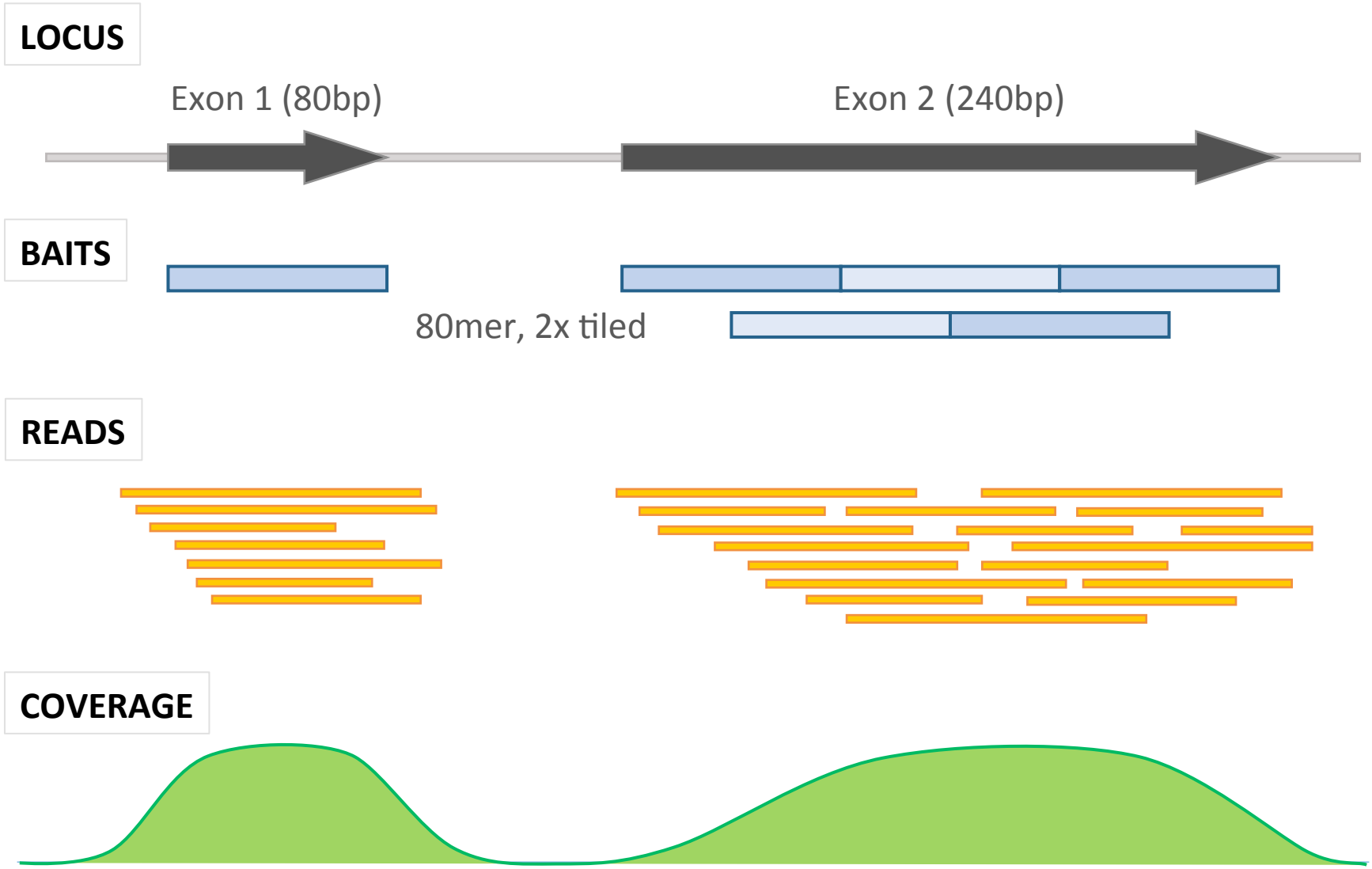
Your sequences = capture probes

→ Sequences of interest are produced as pool of short, synthetic RNA baits

→ Baits hybridize to barcoded NGS library molecules

→ Enriched molecules are identified by aligning NGS reads to the targets

MYcroarray



Target enrichment

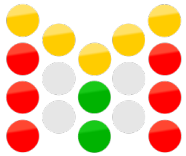
Consider molecular behavior

→ Pad loci by **~20 to 100bp** to mitigate low read coverage at edges

→ **Bait overlap** is recommended (e.g. 2×)

→ Consider higher bait tiling for:

- ❑ Short and/or rare inserts (degraded)
- ❑ Unknown exon/intron boundaries
- ❑ Unknown structural variants



MYcroarray

Choosing appropriate loci

→ Any nucleic acid sequence is eligible

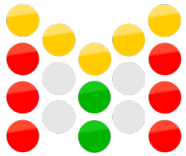
→ Most loci are good capture candidates

→ Can have assorted **lengths**

SNPs | Transcripts | Non-coding | Genes

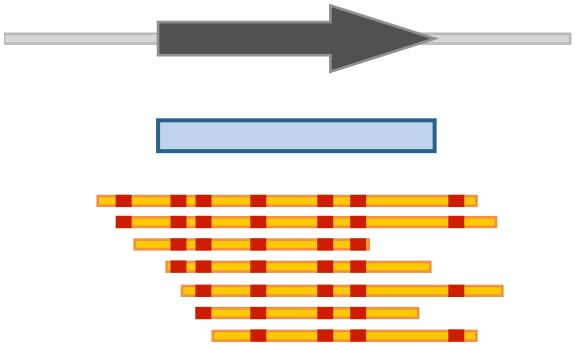
Organellar | Chromosomal

Host | Microbe



MYcroarray

Hybridization is tolerant of divergence



→ Target capture is highly flexible

→ Can tolerate bait:target divergence of **at least 10%+**

→ If sufficiently divergent, reasonable to include >1 variant for a locus or region

→ Balance with kit budget needs

MYcroarray

Providing sequences for bait design

→ FASTA format or genomic coordinates

FASTA

```
>Snake_hemoglobin_gene
AGGCTAGGCTTAGCGAGTATAGAGGCTATAGGGAGTTRRATAT
ATTCGGATCGAGTTAGCGATAGCTAGATTCGATCGACTCT...
>Snake_PCR_amplicon_1735235
TTACCTAATCTGGRTAGAGCTAGCATCGATCGATCGTAGC...
>Snake_mRNA_sequence_NCBI_77777_edited
AAGTACGTAGCTGAGAGGTCGAGGCGTAACCCTAGCTAGCGGA
TCGACTGATCGTGTAGGGATGCTGTGATCGTAGCAAAAAAAAA
```

coor

```
chr1:847584-848920:+
chr6:225046994-225086003:-
chr17:42274060-42292454:+
```

✓ Ambiguities are okay, BUT we will replace with random base before synthesis *(no mixed bases)*

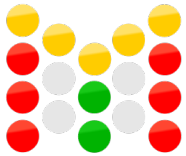
MYcroarray

Including diversity in your baitset

→ Can include multiple variants/alleles per locus

  → e.g. within population or different species

→ Send alignments or individual sequences



MYcroarray

Including diversity in your baitset

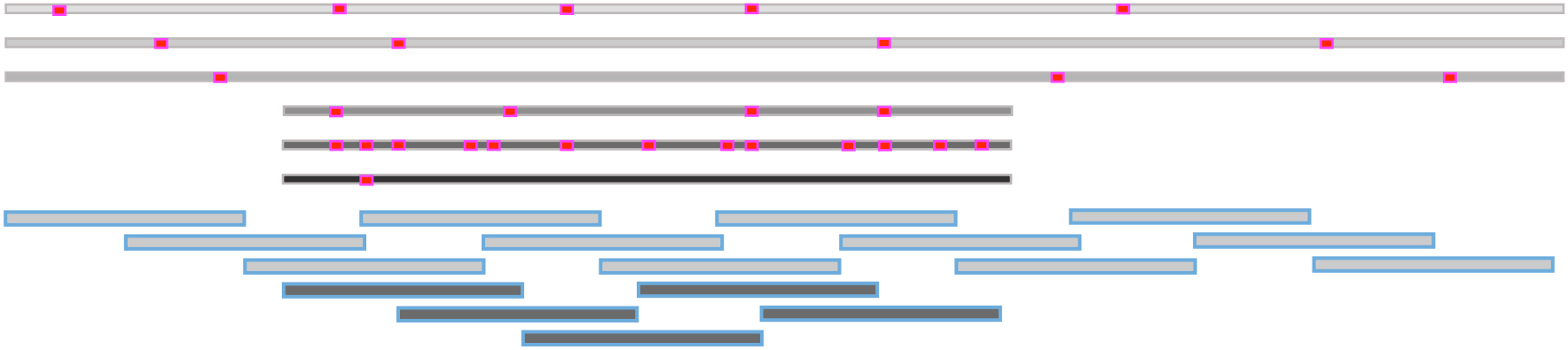
```
>Plant_genome_locus_sequence_species1
CGAGATACGACGAGTATAGCTAGACTGACTACGATATTAGGCGAGCTAGTTGATAGCGGCGCGATCGTATTTCG
>Plant_genome_locus_sequence_species2
CGACATACGACGAGTATAGCTAGACTGAGTACGATATTAGGCGAGGTAGTTGATAGCGGCGCGATCCTATTTCC
>New_transcript_species3
CGAGTTACGAGGAGTATnnnnnnnnnnnnnnCGATA--AGGCGAGCTAGTTGATA-CGGCGCnnnnnnnnnnnn
>New_transcript_species4
CGTGTTACGAGGAG---nnnnnnnnnnnnnnCGATA--AGGCGAGGTAGTTGATA-CGGCGCnnnnnnnnnnnn
>ExonB_species5
------------------------------CGTTATTAGGCCAGGTAGTTGATAGCGGGGC-----------
```

MYcroarray

Including diversity in your baitset

MYcroarray

Including diversity in your baitset



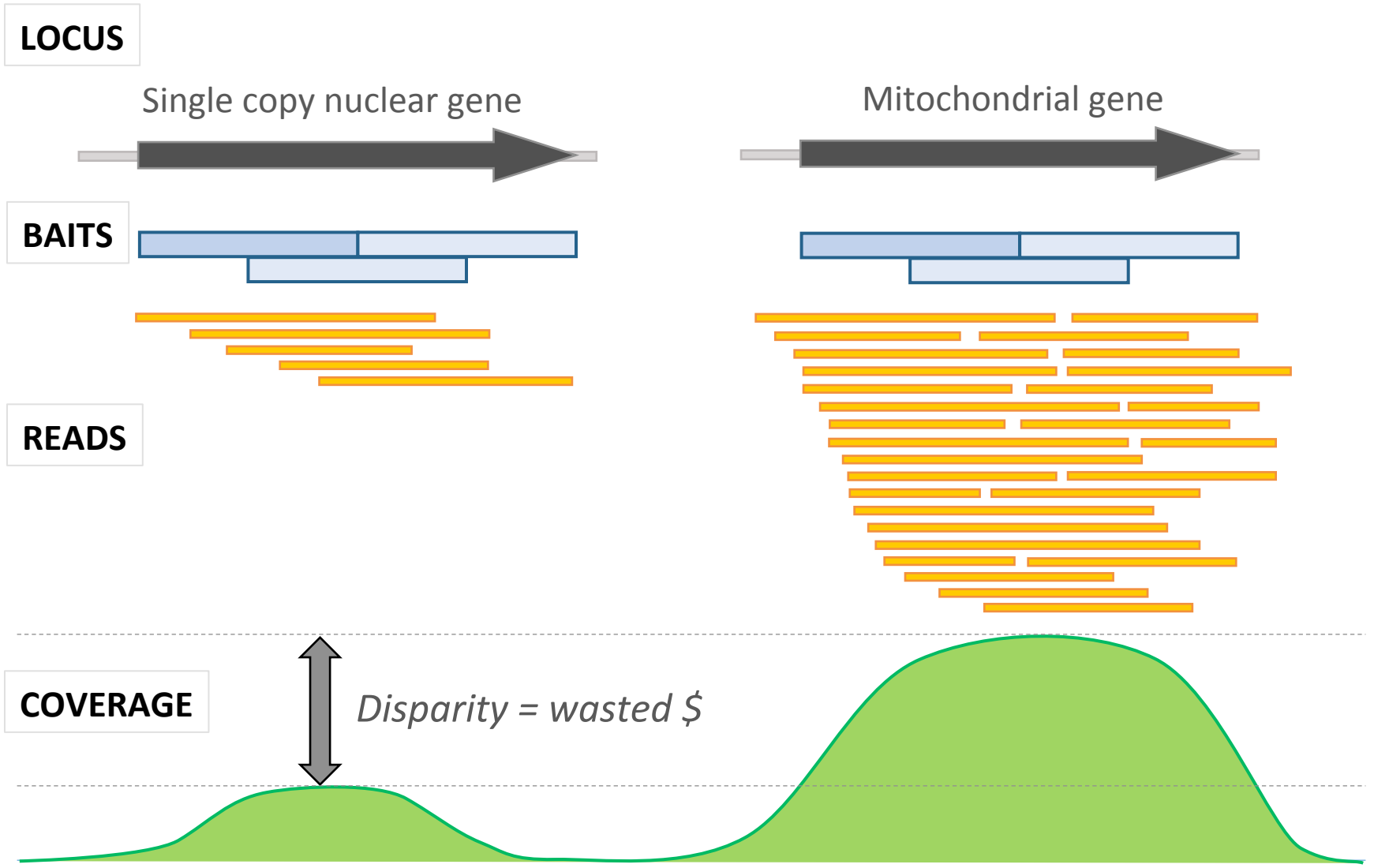
MYcroarray

Important to filter your loci & baits

→ Efficient baitsets have targets of similar copy number

- e.g. chromosomal vs. plastid
- e.g. single-copy vs. multi-copy

MYcroarray



NGS efficiency is important to your project budget

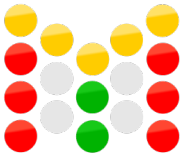
Options for efficiently targeting loci with different copy numbers

Gene families

Known repeats

Organellar

- ❑ Capture separately
  - ➢ Pool in new ratio before sequencing
- ❑ Replicate baits differently
  - ➢ *(or, "spike-in" high-copy baits)*
- ❑ Assemble high-copy loci from your off-target reads
  - ➢ e.g. "Hyb-Seq" genome skimming

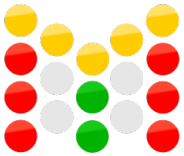


MYcroarray

Methods of filtering loci

- Repetitive elements
- Low complexity
- Common motifs
- High-copy loci

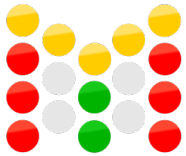
1. Genome-wide BLAST of candidate baits
   - ✓ e.g. MYbaits® complementary screen
2. RepeatMasker.org
3. Taxon-specific repeat databases
4. Low coverage shotgun data
5. Iterative experimental process



MYcroarray

Example BLAST filtering results

Probe	40.0-60.0	62.5	65.0	67.5	70.0	>70.0	Selected
Bait1	0	0	0	1	0	0	true
Bait2	0	0	0	0	0	0	true
Bait3	345	45	32	8	21	12	false
Bait4	2	0	2	1	0	0	true



MYcroarray

How much will my kit cost?

1. Number of bait sequences
2. Number of capture reactions

# reactions	~1 Mb (20K probes)	~5 Mb (100K probes)	~10 Mb (200K probes)
12	$2,400	$4,800	$7,200
48	$5,760	$11,520	$17,280

MYcroarray

Planning for project costs

→ UPSTREAM: Library preparation

→ DOWNSTREAM: Sequencing

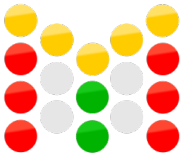
→ Mitigating costs of capture:

1. Optimizing number of targets
2. Tweaking tiling density
3. Multiplexing and/or diluting baits

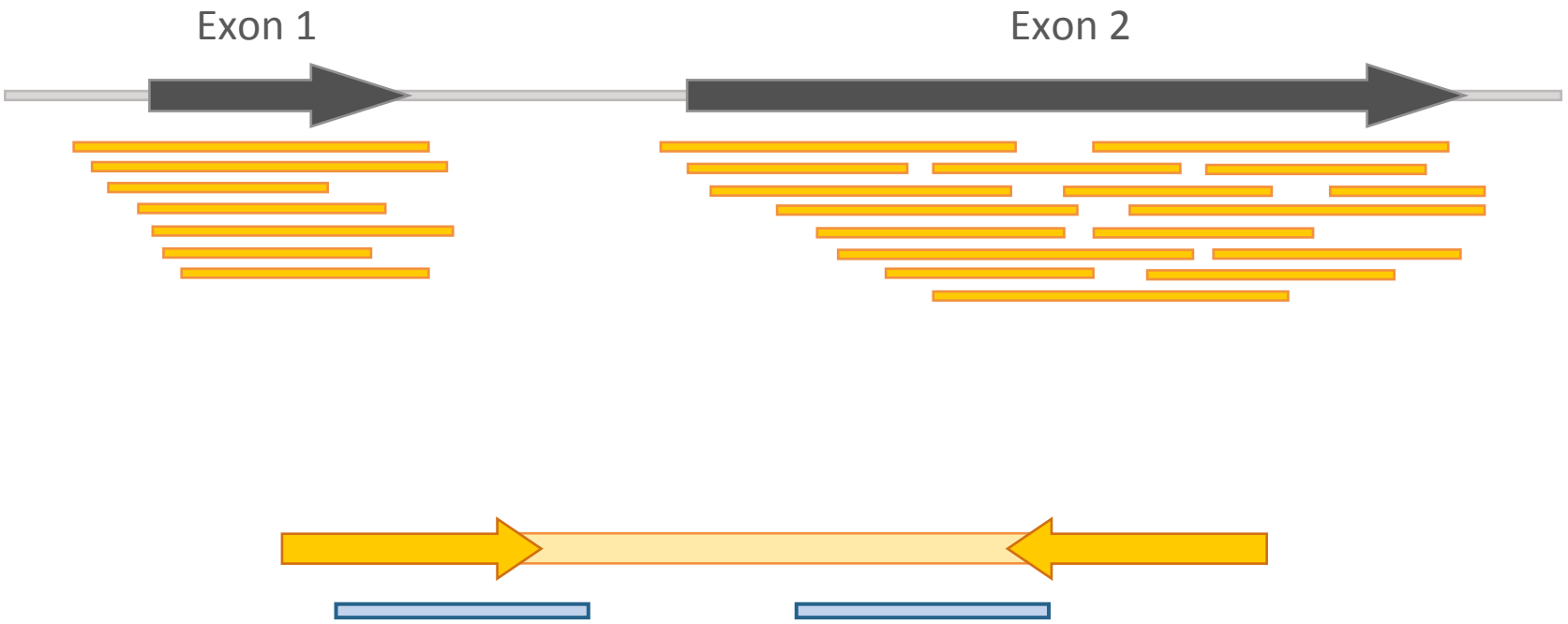
MYcroarray

Next up,
capture-specific considerations for

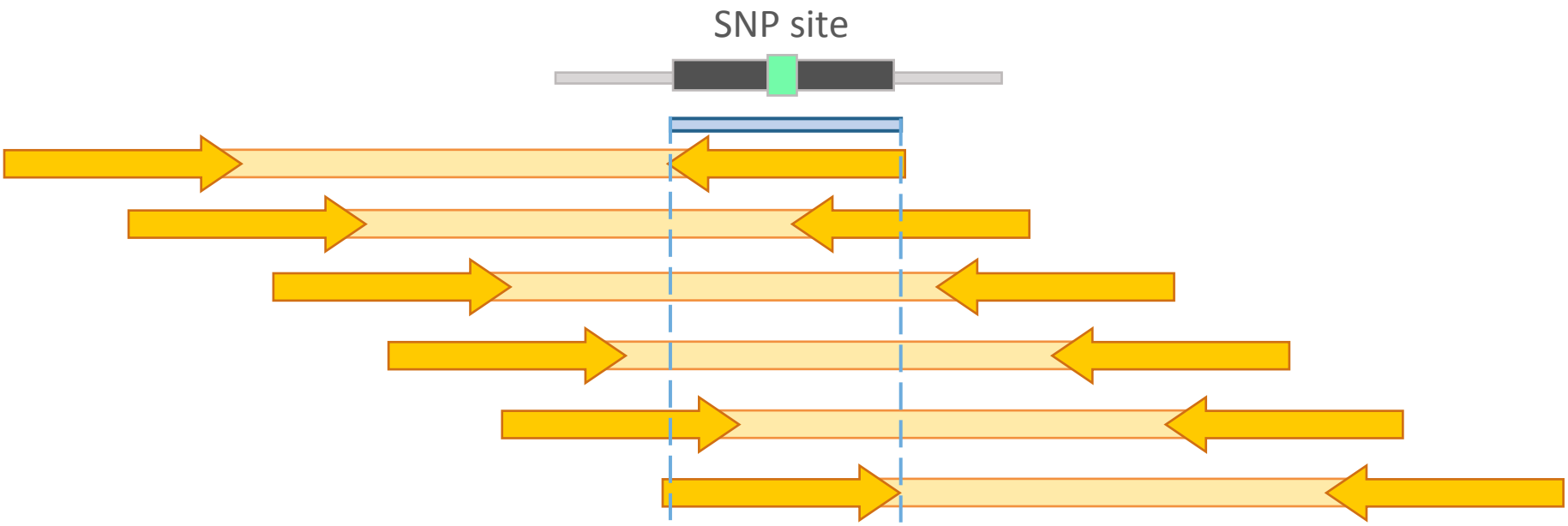
- ⇨ gDNA preparation
- ⇨ library preparation
- ⇨ capture setup & cleanup
- ⇨ postcapture amplification
- ⇨ sequencing
- ⇨ performance analysis

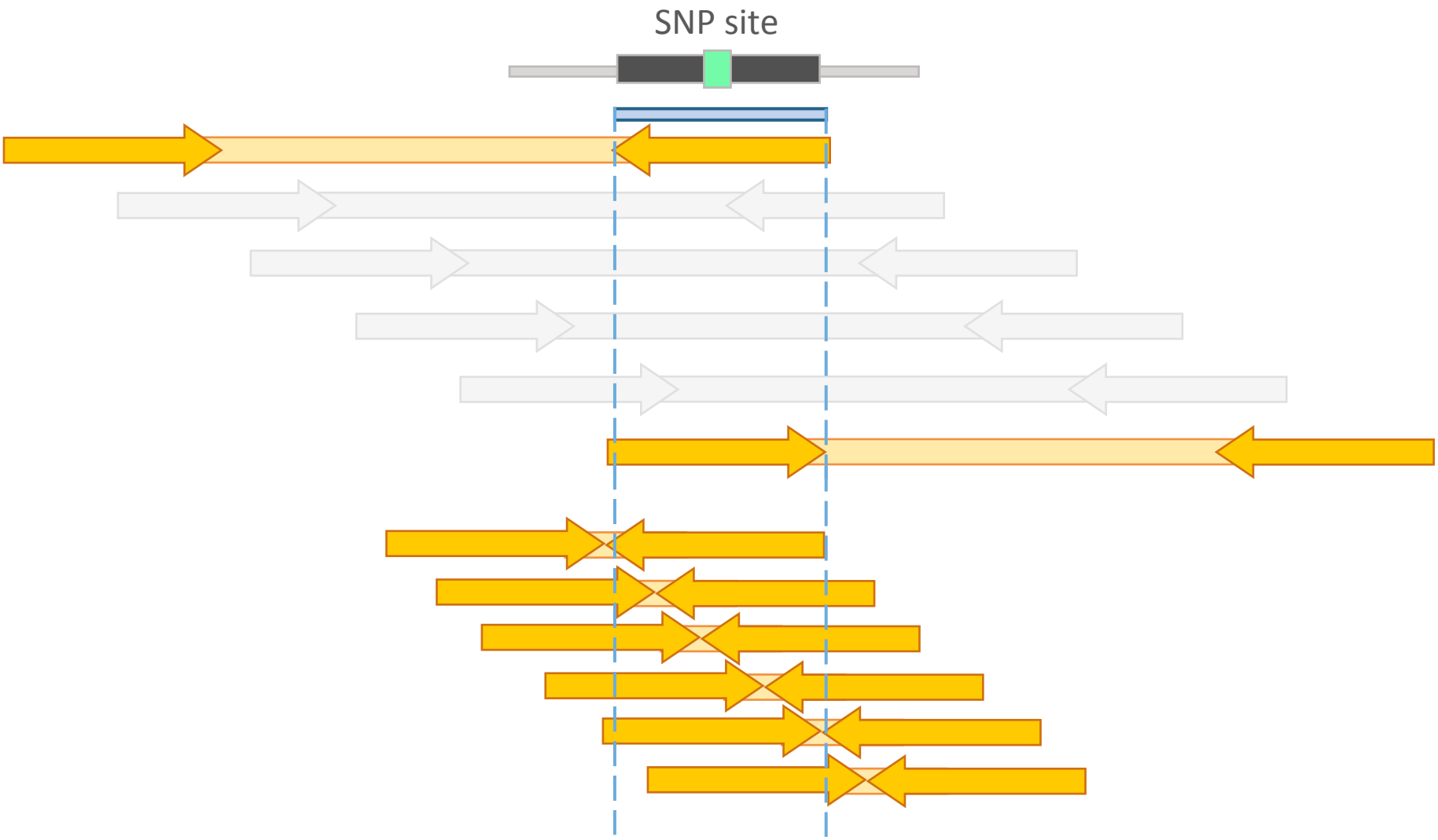


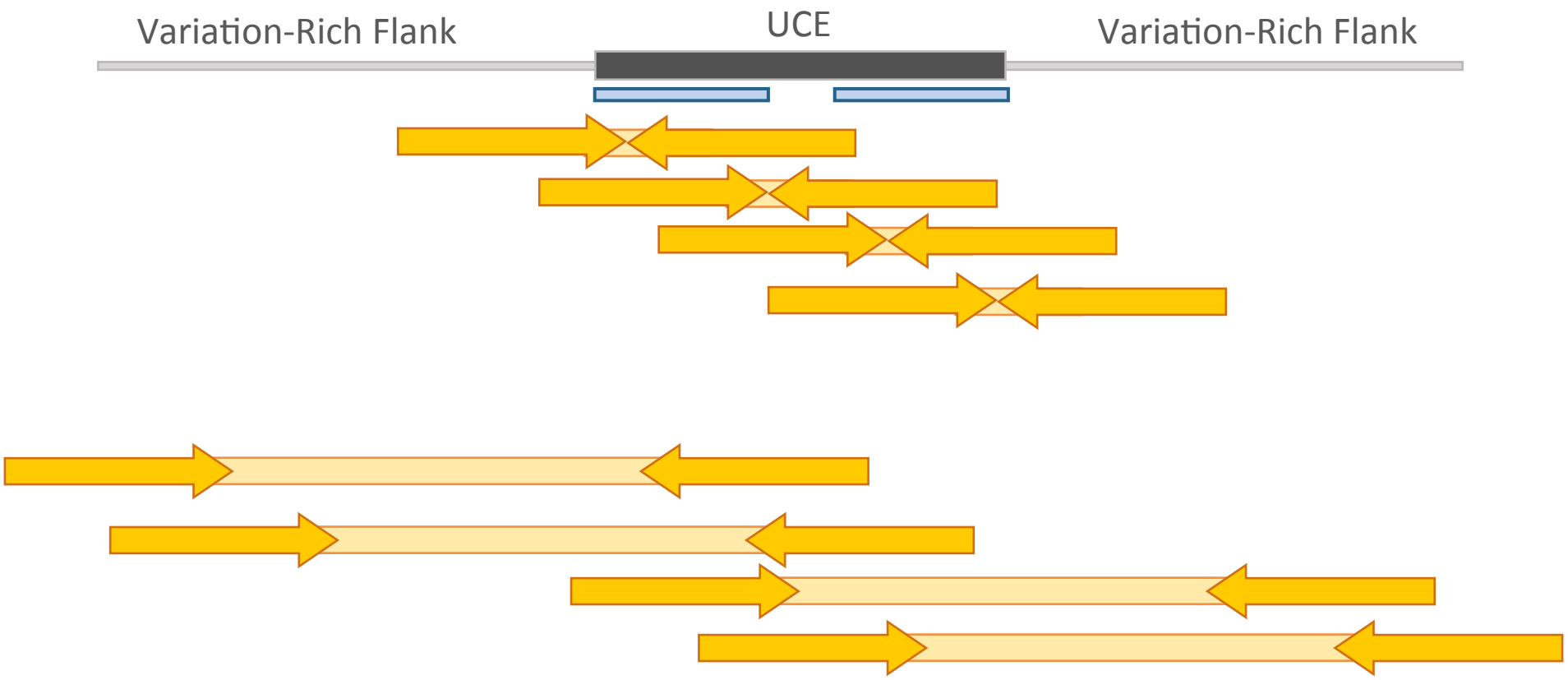
MYcroarray



Start by considering your targets


SNP site

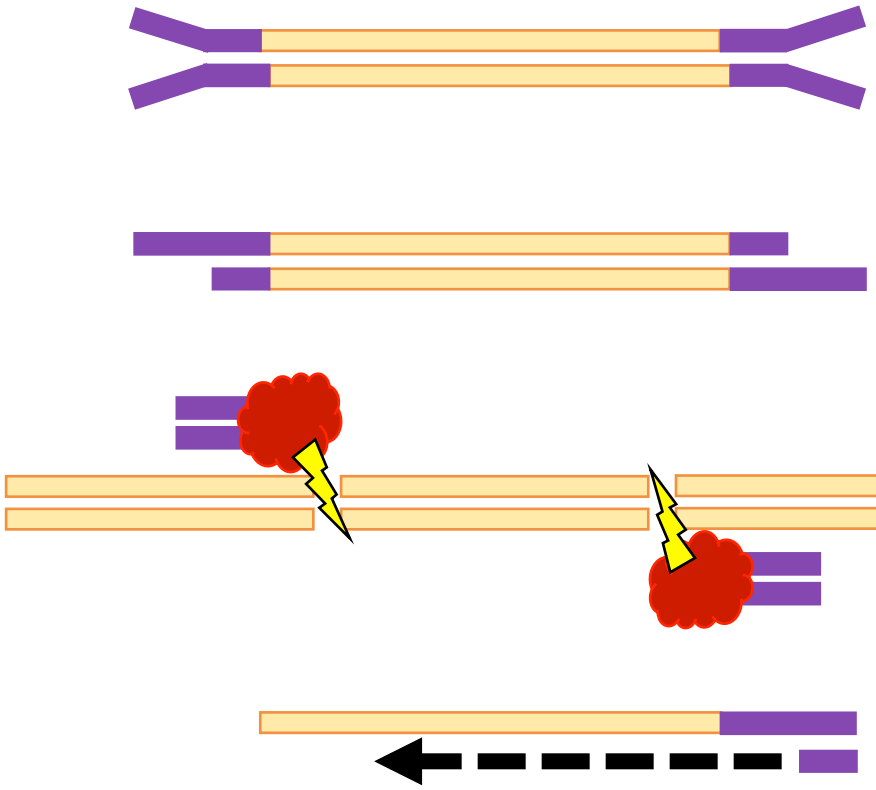

SNP site



Targets inform sonication, size-selection, and sequencing protocol

Library Preparation

Library preparation



sticky-end
(Illumina®)

blunt-end
(Torrent®, dDNA)

transposase
(Nextera®)

ssDNA
(aDNA, Swift Bio)

MYcroarray

Library preparation

KAPABIOSYSTEMS evolving better science — ***Hyper/HyperPlus*** — S

NEW ENGLAND BioLabs GmbH — ***NEBNext® Ultra™*** — S

NEBNext® Fast — B

www.cshprotocols.org
Vol. 2010, Issue 6, June

Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing

Matthias Meyer[1] and Martin Kircher

B

Ultimately:

MYcroarray

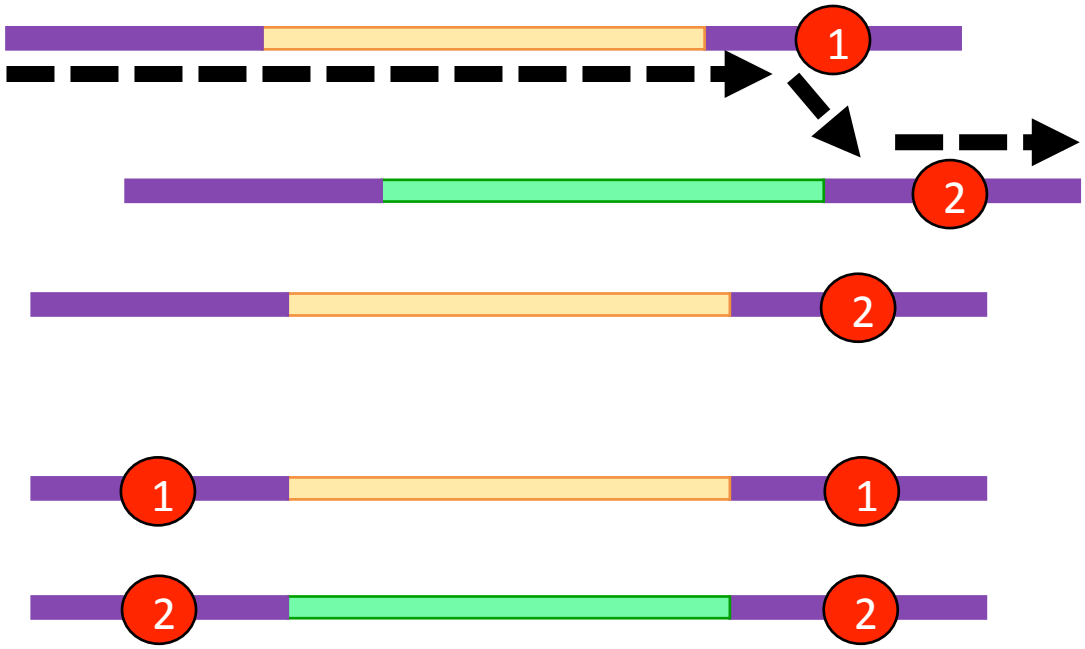
Library preparation – multiplex captures

Nucleic Acids Research, 2011, 1–8
doi:10.1093/nar/gkr771

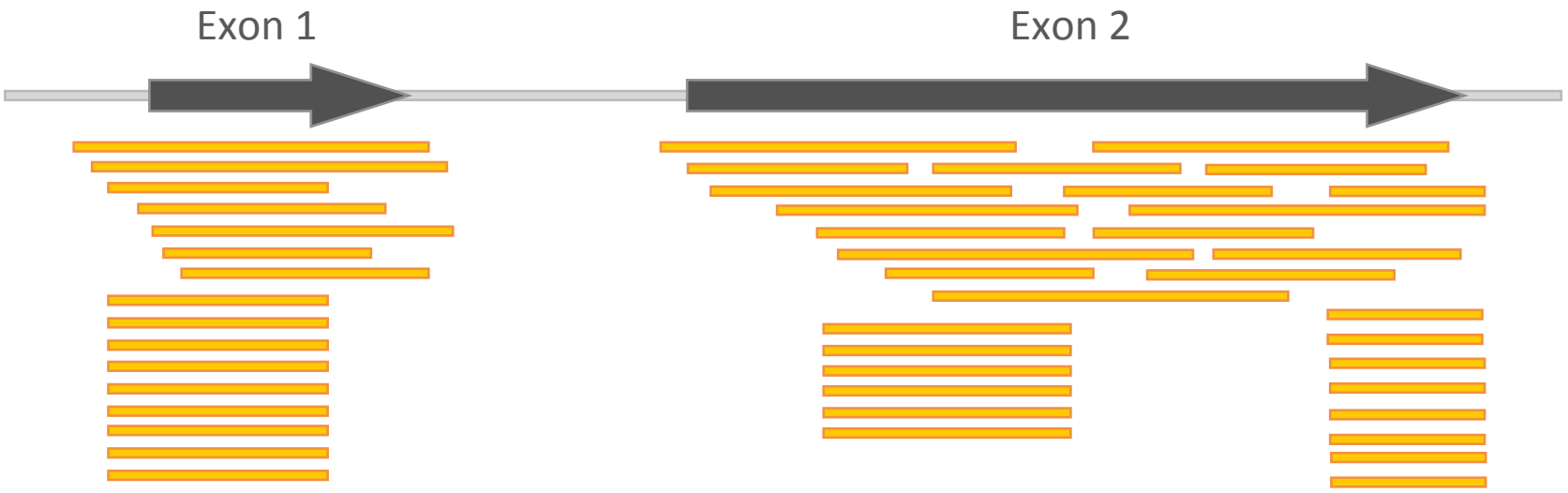
Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform

Martin Kircher*, Susanna Sawyer and Matthias Meyer*

Max Planck Institute for Evolutionary Anthropology, Department of Evolutionary Genetics, 04103 Leipzig, Germany



MYcroarray



DNA input to library: **maximize**

Library amplification: **moderate**

Library amplification

⇨ Use low-bias enzymes

- Degraded? Use uracil non-stalling
  - e.g., KAPA® HiFi Uracil+

⇨ Moderate cycles

- 500 ng input + 8 cycles → 1-2 µg lib

MYcroarray

Target Capture

Capture materials

⇨ Libraries

⇨ Baits (+ hybridization & wash kit) MYbaits®

⇨ Streptavidin-coated magnetic beads

Dynabeads®
MyOne™ C1

⇨ Magnetic rack

Diagenode® DiaMag02

ALPAQUA® 96S

MYcroarray

Capture materials

⇨ Water bath or hybridization oven

⇨ Thermal cycler (×2)

⇨ Low-bind PCR tubes (Axygen™)

⇨ Recommended:

- Heat block, multi-channel pipettor

MYcroarray

Capture materials considerations

⇨ COT-1 DNA (MYbaits Block #1 is *H. sapiens*)

- Species-proximate option if available

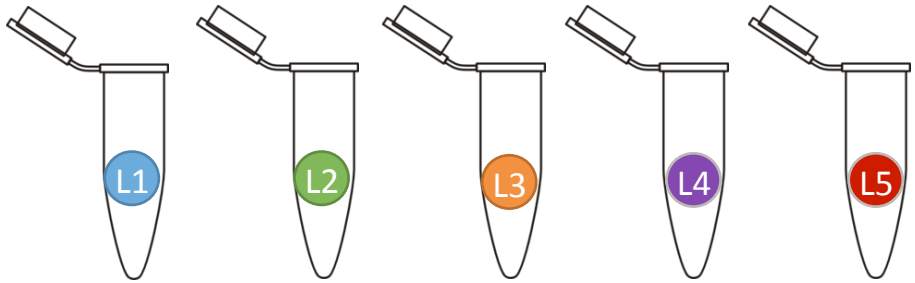
⇨ Adapter-specific blockers (MYbaits Block #3)

- Index-specific high-bind (*BiooScientific, IDT*)

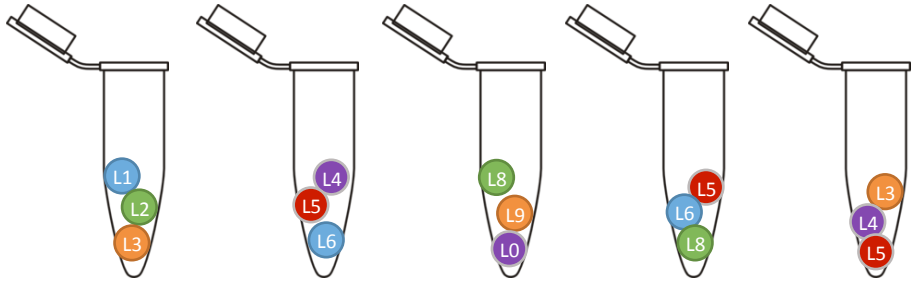
⇨ KAPA® HiFi or AccuPrime™ Pfx polymerases

- Allows for on-bead amplification postcapture

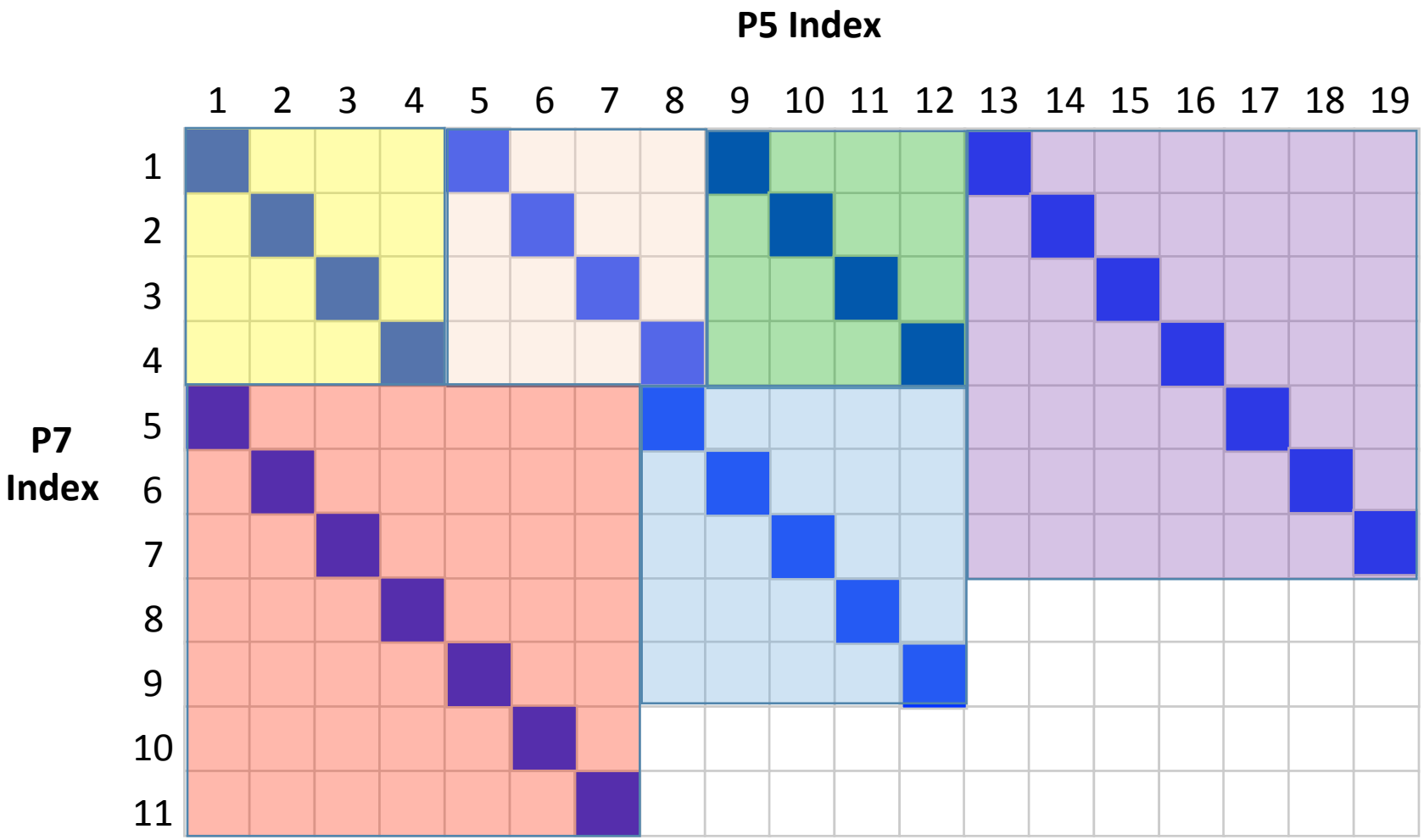
MYcroarray



Library input to capture: 100-500 ng
(14 - 72 ng/µL in 7 µL)



Degraded/rare DNA or Multiplexing:
up to 2 µg total per rxn

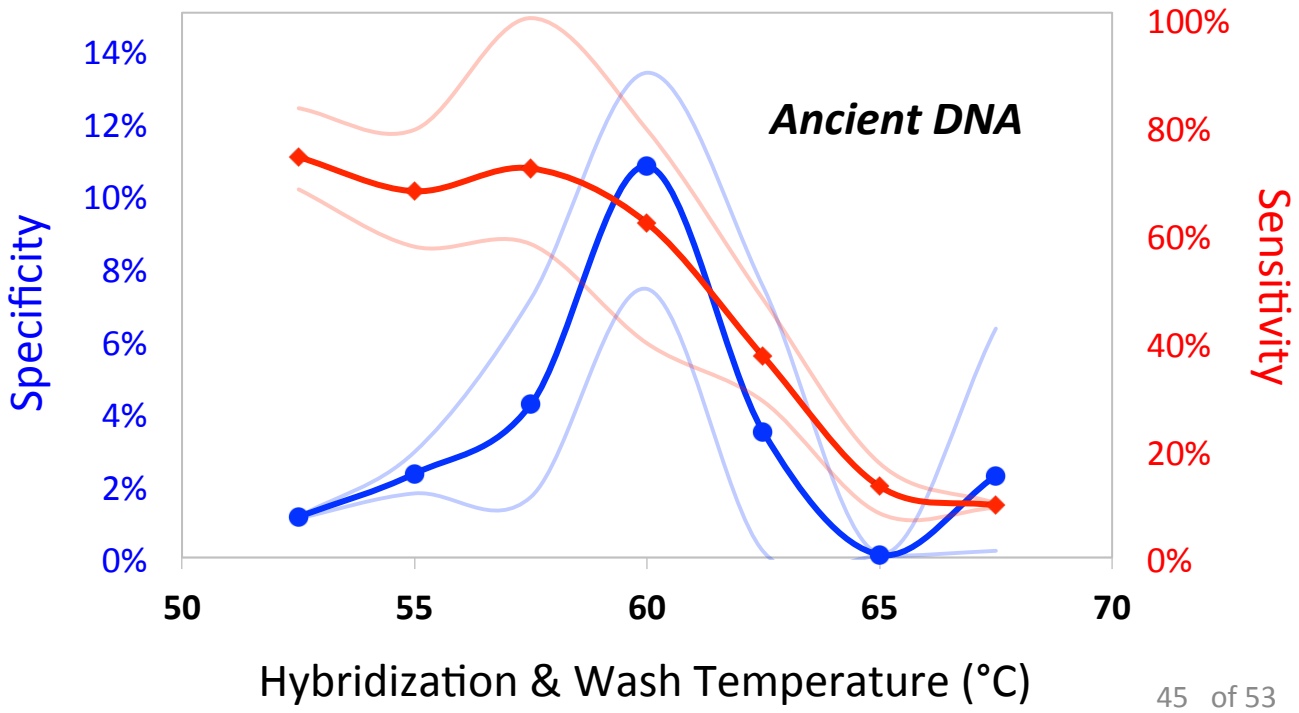


Multiplexing – careful with your index combos!

Choosing hyb/wash temperature

⇨ Majority of projects: 65°C

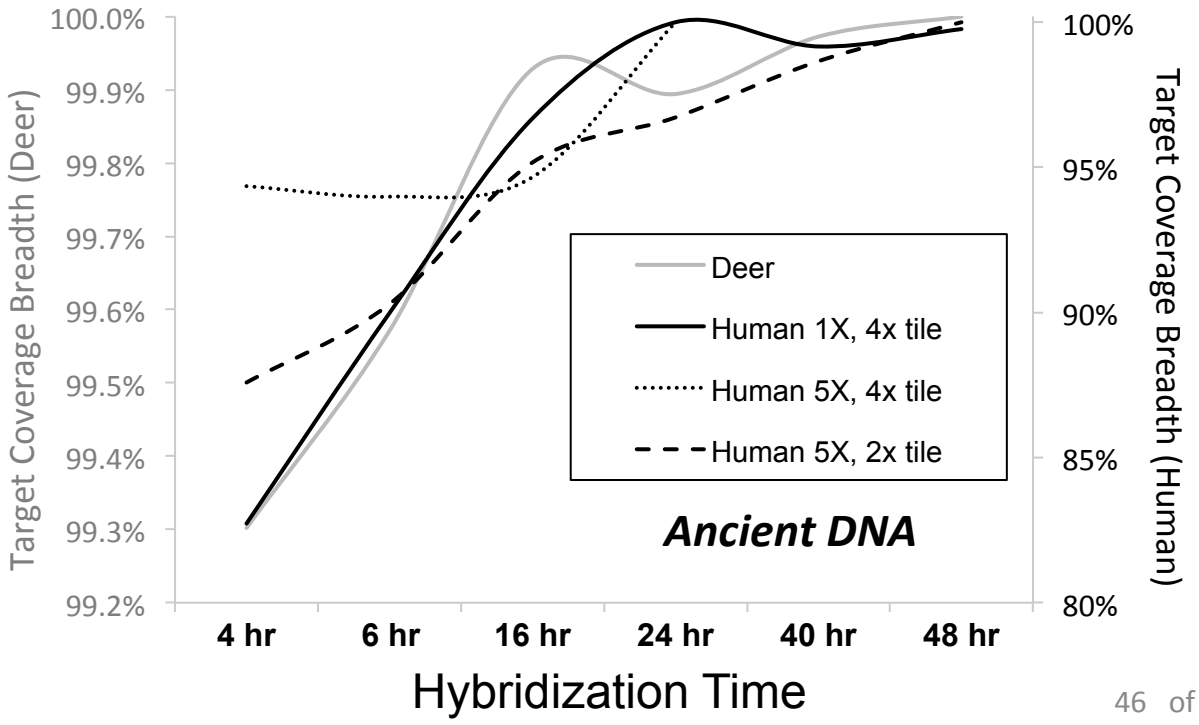
⇨ Very rare and/or degraded templates: 55-65°C



MYcroarray

Choosing hyb time

- ⇨ Majority of projects: Overnight hybridization
- ⇨ Very rare and/or degraded templates: 24-72 hr



MYcroarray

$$$ Budgeting – DNA to enriched library

Step	Materials	Time
gDNA extraction	$ 10	0.5 days
sonication / size-selection	$ 10	1 day
library preparation	$ 50	1 day
capture – bait kit	$ 150	2 days
capture – magbeads	$ 10	-
capture – reamplification	$ 5	-
library quantification	$ 5	0.5 days
TOTAL	**$240/sample**	**5 days** (~48 samples)

Sequencing

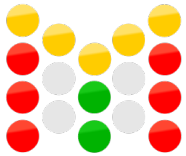
How should I sequence?

⇨ Protocol: how long of reads?

- Consider library insert size – point to long PE?
- Consider target type – SNPs, flanks, or contigs?

⇨ Depth - three important numbers:

- Size of target
- Required coverage depth
- Expected % on-target*



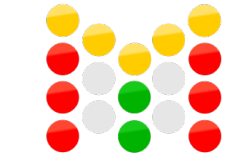
How deep to sequence?

required reads =	example
target length	*2 Mbp*
÷	÷
read length	*200 (PE100)*
×	×
X coverage desired	*50 X*
÷	÷
expected % on-target	*40%*
÷	÷
expected % unique	*20%*
	=6.25 M reads/sample

Measuring success

⇨ **Specificity** (measure: % reads on-target)

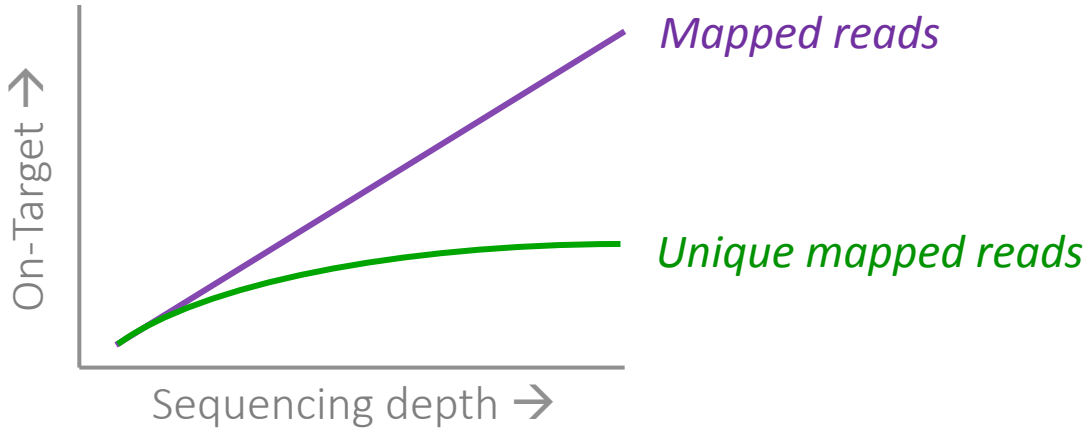
- Frequently ranges 15-95% on-target
- If low: use higher temps (up to 68°C)
- If low: do a second round of capture
- If low: check your bait design



MYcroarray

Measuring success

⇨ **Sensitivity** (measure: library complexity)

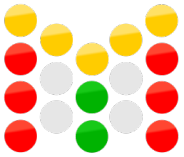


- If low: more input into library prep/capture
- If low: use lower temps (down to 55°C)
- If low: increase hyb time (up to 72 hr)

MYcroarray

Project Design & Troubleshooting

Contact us!



MYcroarray

THE OLIGO LIBRARY COMPANY

http://crossroads.uni-koeln.de/images/neb.jpg

https://www.kapabiosystems.com/assets/logo.png

https://www.diagenode.com/img/product/reagents/diamag02.png

http://tools.thermofisher.com/content/sfs/prodImages/high/65001_650x600.jpg

http://www.alpaqua.com/Portals/0/Images/Magnet%20Plates/MP021-260x140.jpg

http://www.clker.com/clipart-eppendorf-tube.html