



# Targeted next-gen sequencing for plant phylogenomics

**Jennifer Mandel** University of Memphis

**Rebecca Dikow** Smithsonian Institution

**Vicki Funk** Smithsonian Institution

# Overview of Workshop

- Why phylogenomics and why this method?
- Getting started designing the project
- Choosing taxa and best practices
- Carrying out the lab work
- Carrying out the bioinformatics
- Costs and considerations
- Where to get help?

# *Why Phylogenomics?*

- Many, many loci can yield tremendous resolution
- Can begin to tackle questions about gene trees and species trees
- Next-gen sequencing has become affordable
- Many resources available for designing your own project

# Many, Many Loci

- How many loci?
- What type of loci?
- How many taxa?
- Depends on your questions

# Sanger vs. NGS

- Cost considerations
  - Total vs. per sample
- Amount of data possible
  - Few vs. hundreds
  - Do you need all that data?
- Data analysis/processing
  - Old hat vs. potentially challenging

# Lots of Options

## A Targeted Enrichment Strategy for Massively Parallel Sequencing of Angiosperm Plastid Genomes

Gregory W. Stull<sup>2,3,8</sup>, Michael J. Moore<sup>4</sup>, Venkata S. Mandala<sup>4</sup>, Norman A. Douglas<sup>4</sup>, Heather-Rose Kates<sup>3,5</sup>, Xinshuai Qi<sup>6</sup>, Samuel F. Brockington<sup>7</sup>, Pamela S. Soltis<sup>3,5</sup>, Douglas E. Soltis<sup>2,3,5</sup>, and Matthew A. Gitzendanner<sup>2,3,5</sup>



## A Phylogenomic Approach Based on PCR Target Enrichment and High Throughput Sequencing: Resolving the Diversity within the South American Species of *Bartsia* L. (Orobanchaceae)

Simon Uribe-Convers , Matthew L. Settles , David C. Tank

## Inferring Phylogeny and Introgression using RADseq Data: An Example from Flowering Plants (*Pedicularis*: Orobanchaceae)

Deren A. R. Eaton<sup>1,2\*</sup> and Richard H. Ree<sup>1,2</sup>

## Transcriptome Sequences Resolve Deep Relationships of the Grape Family

Jun Wen , Zhiqiang Xiong, Ze-Long Nie, Likai Mao, Yabing Zhu, Xian-Zhao Kan, Stefanie M. Ickert-Bond, Jean Gerrath, Elizabeth A. Zimmer, Xiao-Dong Fang

# Targeted Enrichment/Hyb-Seq

## Hyb-Seq: Combining target enrichment and genome skimming for plant phylogenomics<sup>1</sup>

Kevin Weitemier,<sup>2,7</sup> Shannon C. K. Straub,<sup>2,7</sup> Richard C. Cronn,<sup>3</sup> Mark Fishbein,<sup>4</sup> Roswitha Schmickl,<sup>5</sup> Angela McDonnell,<sup>4</sup> and Aaron Liston<sup>2,6</sup>

## A Target Enrichment Method for Gathering Phylogenetic Information from Hundreds of Loci: An Example from the Compositae

Jennifer R. Mandel<sup>2,9</sup>, Rebecca B. Dikow<sup>3</sup>, Vicki A. Funk<sup>4</sup>, Rishi R. Masalia<sup>5</sup>, S. Evan Staton<sup>6</sup>, Alex Kozik<sup>7</sup>, Richard W. Michelmore<sup>7</sup>, Loren H. Rieseberg<sup>8</sup>, and John M. Burke<sup>5</sup>

## Using phylogenomics to resolve mega-families: An example from Compositae

Jennifer R. Mandel<sup>1,2,†,\*</sup>, Rebecca B. Dikow  
Issue \_\_\_\_\_  
<sup>3,†</sup> and Vicki A. Funk<sup>4</sup>



Journal of Systematics

Applications in Plant Sciences 2015 3(8): 1500039

APPLICATION ARTICLE

## A PROTOCOL FOR TARGETED ENRICHMENT OF INTRON-CONTAINING SEQUENCE MARKERS FOR RECENT RADIATIONS: A PHYLOGENOMIC EXAMPLE FROM *HEUCHERA* (SAXIFRAGACEAE)<sup>1</sup>

RYAN A. FOLK<sup>2,4</sup>, JENNIFER R. MANDEL<sup>3</sup>, AND JOHN V. FREUDENSTEIN<sup>2</sup>

# Some initial considerations

- The first step for a target enrichment study is to identify all the genomic resources that exist for your group: genomes, transcriptomes, ESTs
- Consider the phylogenetic breadth you wish to evaluate (exons vs. introns/cds vs. non-cds)

# Where to start?

- What resources are available?
  - Genomes?
  - Transcriptomes?
  - EST data
  - Same or closely related species
  - Nothing?

# Genomes



The image shows the PlantGDB homepage. At the top left is the "PlantGDB" logo in large yellow letters. To its right is a magnifying glass icon over a sequence of DNA bases: ACTGA. To the right of the magnifying glass is the tagline "...resources for comparative plant genomics". Below the header is a navigation bar with seven items: Home, Sequence, Genomes, Tools, Datasets, Outreach, and Help. The "Home" item is highlighted with a light green background.

## Search - Overview

PlantGDB downloads all *Viridiplantae* plant sequence data (GenBank and UniProt) every 4 months, parses them by species and makes data available for search, download or BLAST.

► *Current PlantGDB Version: Release 187 [2011-12-15] (click for details...)*

## Search and Download Options

- Quick Search (see header box)
- [Species Search/Download](#) -- identify and retrieve sequences from any species PlantGDB's Public Plant Sequence repository.
- [ID or Keyword Search](#) -- identify specific sequences by ID or annotation keyword.
- [FTP server](#) -- batch download of sequences and PUT assemblies by species or sequence type.
- [TableMaker tool !\[\]\(67498f4429b07796de483219b69f0741\_img.jpg\)](#) -- for complex PlantGDB database queries.

# More Genomes



<https://phytozome.jgi.doe.gov/pz/portal.html>

A screenshot of the Phytozome 11 portal homepage. At the top is a navigation bar with links for Home, Help, Contact, About Us, Subscribe, Login, and Register. Below this is a secondary navigation bar with links for Search, Browse, Tools, Portals, Download, Submit, News, and ABRC Stocks. The main content area shows a breadcrumb trail: Home &gt; Portals &gt; Genome Annotation &gt; Other Genomes. A section titled "Other Genome Databases and Sequencing Efforts" is visible. The Tair logo is present in the top left corner.

[https://www.arabidopsis.org/portals/genAnnotation/other\\_genomes/](https://www.arabidopsis.org/portals/genAnnotation/other_genomes/)

A screenshot of the NCBI genome search interface. The top navigation bar includes links for NCBI, Resources, and How To. Below this is a search bar with dropdown menus for "Genome" and "Species", a search input field, and a "Search" button. There are also "Limits" and "Advanced" links below the search bar.

<http://www.ncbi.nlm.nih.gov/genome>

# Transcriptomes



# 1000 Plants

Search this site

[HOME](#)

[CONTACT INFO](#)

[GREEN PLANTS](#)

[MEDIA](#)

▼ [SUB-PROJECTS](#)

[AGRICULTURE](#)

[ANGIOSPERMS](#)

[BIOCHEMISTRY](#)

[EXTREMOPHYTE](#)

[GREEN ALGAE](#)

[MEDICINES](#)

[NON-FLOWERING](#)

[SITEMAP](#)

## Home

The 1000 plants (oneKP or 1KP) initiative is an international multi-disciplinary consortium that has generated large-scale gene sequencing data for over 1000 species of plants. Major supporters include Alberta Ministry of Innovation and Advanced Education, Musea Ventures (Somekh Family Foundation), Beijing Genomics Institute in Shenzhen (BGI-Shenzhen), China National GeneBank (CNGB), iPlant Tree-of-Life (iPToL) Grand Challenge, Compute Canada (Westgrid), Alberta Innovates Technology Futures (AITF-iCORE Strategic Chair). The sample selection was originally based on a series of overlapping sub-projects with scientific objectives that could be addressed by sequencing multiple plant species (links on left). As more collaborators joined 1KP, however, the objectives evolved and are now exemplified by the diverse collection of papers described by the links below.

# ESTs

The screenshot shows the NCBI dbEST database interface. At the top left is the NCBI logo. Below it is a navigation bar with links: PubMed, Entrez, BLAST, OMIM, Taxonomy, and Structure. A search bar contains the placeholder "Search EST for" and buttons for "Go" and "Clear". Below the search bar is a filter for "modified during the last" time period, set to "10 Years". To the left of the main content area is a sidebar with links: BI, E MAP, Human Genome Resources, and iGene. The main content area features a purple header bar with the text "► What is dbEST?". The main text describes dbEST as a division of GenBank containing sequence data and other information on single-pass cDNA sequences from various organisms. It also mentions a brief account of the history of human ESTs in GenBank and a special issue of Nature.

# Expressed Sequence Tags database

PubMed Entrez BLAST OMIM Taxonomy Structure

Search EST for  Go Clear

modified during the last 10 Years

► **What is dbEST?**

dbEST ([Nature Genetics 4:332-3;1993](#)) is a division of [GenBank](#) that contains sequence data and other information on "single-pass" cDNA sequences, or "Expressed Sequence Tags", from a number of organisms. A brief account of the history of human ESTs in GenBank is available ([Trends Biochem. Sci. 20:295-6;1995](#)). Also, consult the special "Genome Directory" issue of *Nature* (vol. 377, issue 6547S, 28 September 1995).

# Finding close relatives

Angiosperms

Flowering Plants

Pam Soltis, Doug Soltis, and Christine Edwards



<http://tolweb.org/angiosperms>

[HOME](#)

[TREES](#)

[ORDERS](#)

[FAMILIES](#)

[CHARACTERS](#)

[SEARCH](#)

[LINKS](#)

[REFERENCES](#)

**Angiosperm Phylogeny Website**

[GLOSSARY](#)

**ANGIOSPERM PHYLOGENY WEBSITE, version 13.**

<http://www.mobot.org/MOBOT/research/APweb/>

# How we approached it:



# “Those things are related?”

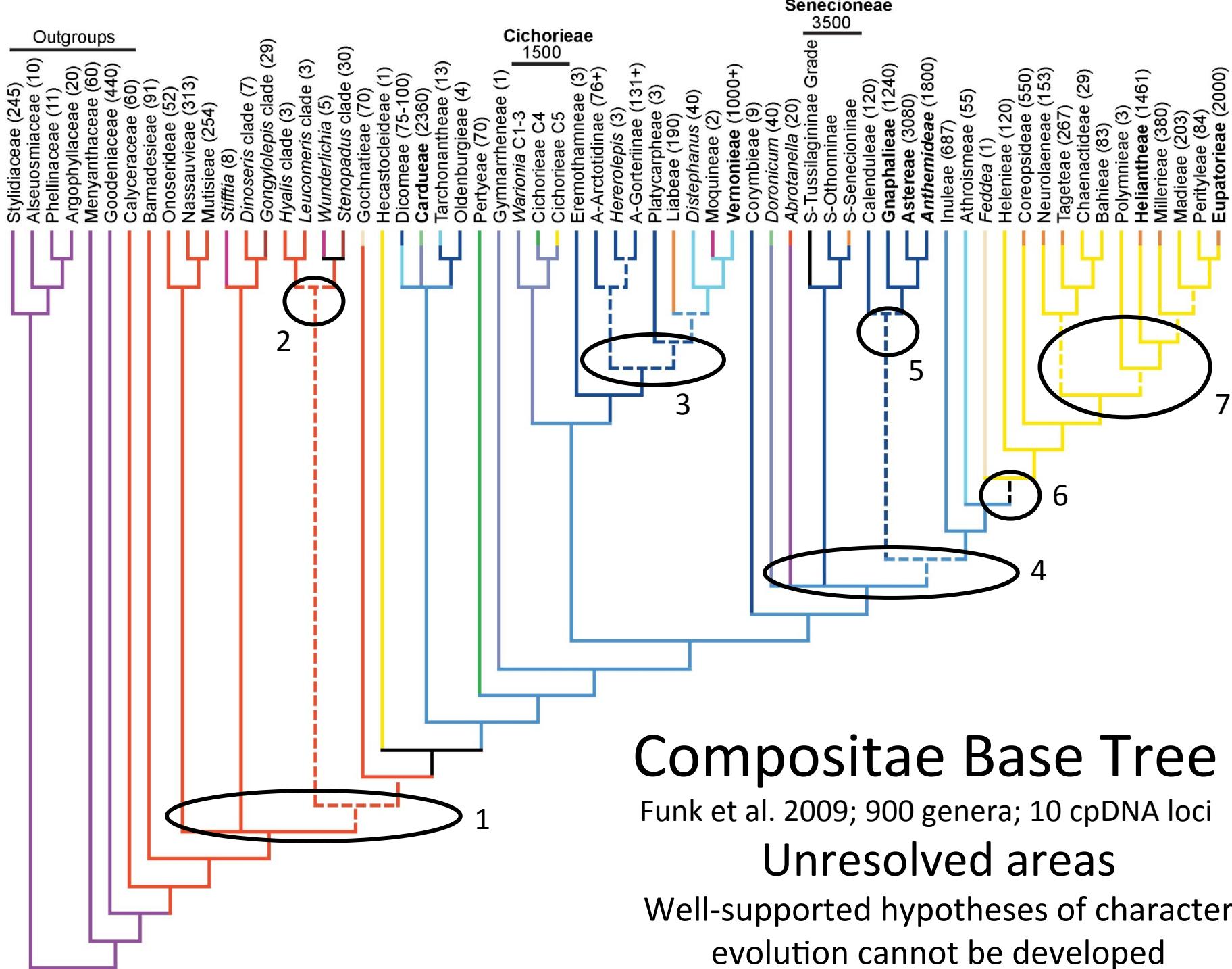
Yes, but how?



# Compositae (a.k.a. Asteraceae)

- ~ 25,000+ species worldwide
- Mediterranean climate, deserts and margins, open prairies and steppes, and mountain habitat
- Vast array of ecological, morphological, and genomic diversity
- Economically important crops, noxious weeds, endangered taxa





# A Need for Large-Scale Data

- Develop and test a workflow for generating phylogenetic information at hundreds of loci
- Utilize targeted capture & next-gen sequencing
- Data available at the time = ESTs

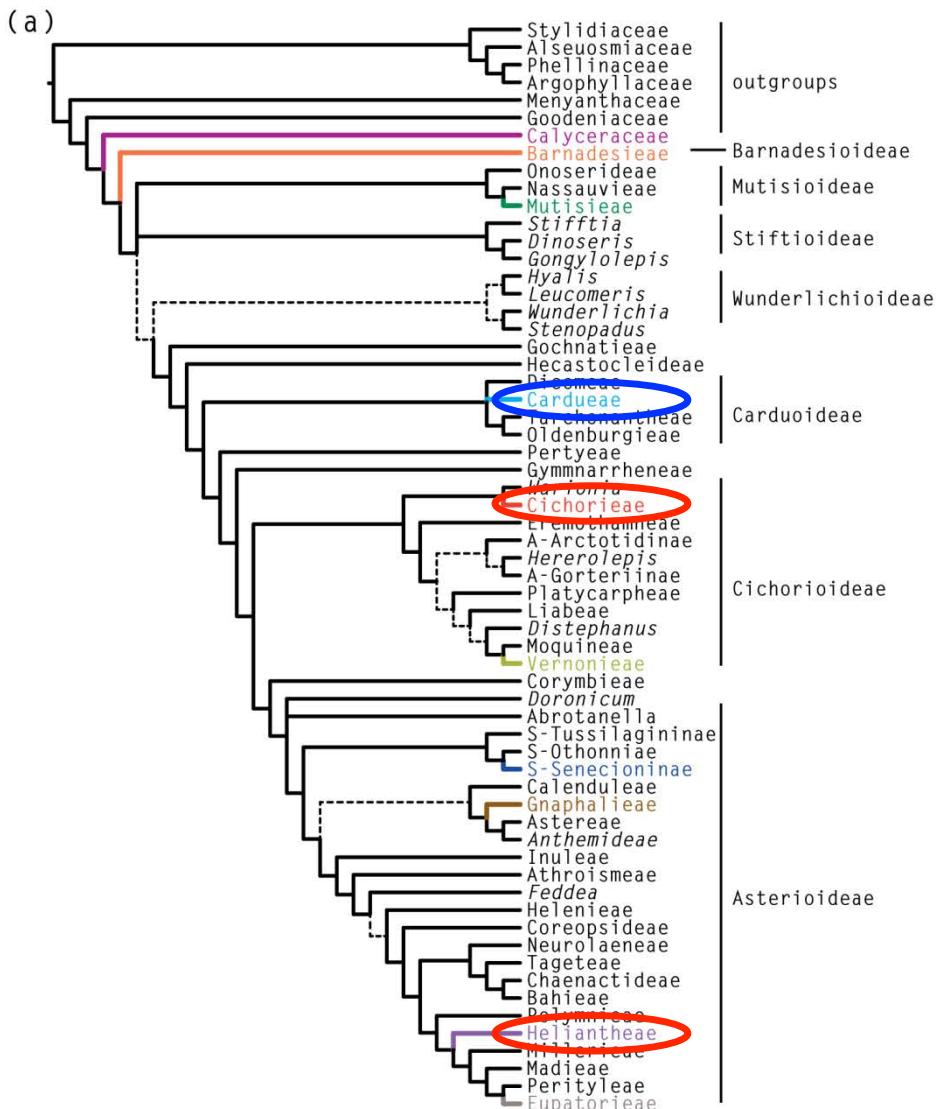


# What we did in 2011...

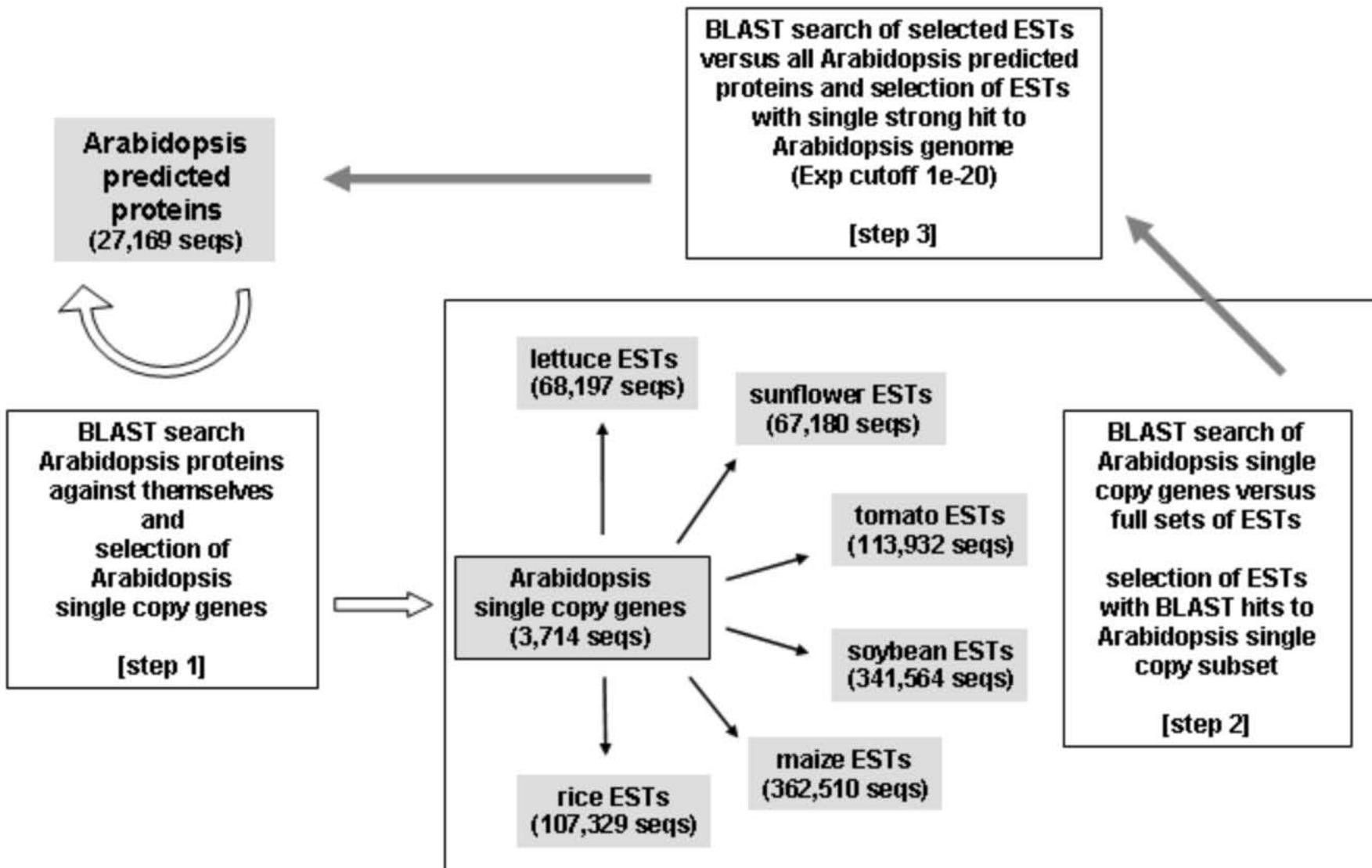
~70K ESTs for each sunflower and lettuce

From this we had  
~1,300 conserved locus alignments

~20K ESTs for each safflower



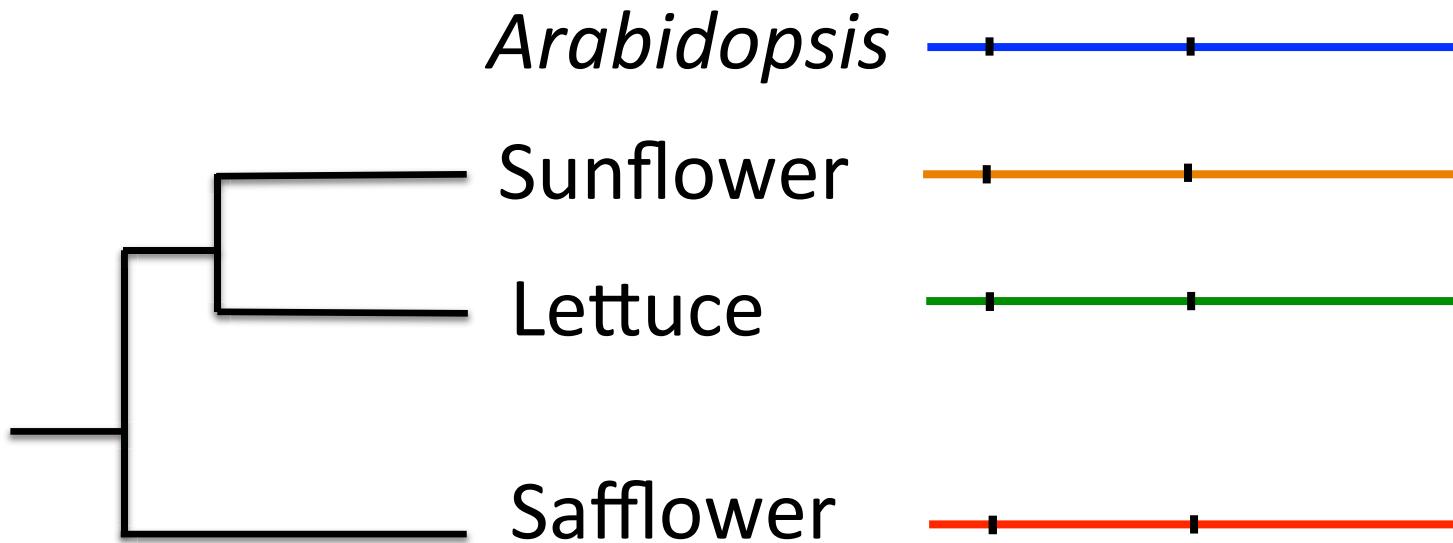
# Single Copy Ortholog Pipeline



# Conserved Ortholog Sequence Alignments

~1,300 loci

Based on ESTs



Added Safflower ESTs to generate ~700/1300 four species alignments

# Designing Probes

One conserved probe per location vs. three probes per location?

Sunflower



Lettuce



Safflower



Baits are 120-mer tiled, overlapped every 60 bases when possible

# Designing Probes

Intron Issue – used *Arabidopsis* intron position to predict putative splice sites and avoided probes spanning those regions

Sunflower



Lettuce



Safflower



Intron Avoiding Probe Design

**1,061 loci targeted**



## Agilent SureSelect Target Enrichment



*NimbleGen SeqCap Target Enrichment*



Probe Design  
What we'd do in 2016...  
Rebecca Dikow

# Much more data available

- Transcriptomes
- Genomes
- Use the data instead of ESTs
- Use orthology detection methods
- Ensure no repetitive DNA

# Resources: Orthology Detection

- **OMA**: find *de novo* orthologs from transcriptome data with or without genomes.
- **AGALMA**: a pipeline that generates orthologs and species trees from transcriptomes. Incorporates PhyloTreePruner, which distinguishes between in- and out-paralogs.
- **Orthofinder**: a program for identifying orthologous protein sequence families. python; runs as a single command taking input a directory of FASTA files, one per species. Outputs file containing orthologous groups of genes from these species.
- **OrthoMCL**: database that includes ortholog groups for a curated set of taxa.
- **HaMStR**: extends a set of orthologs (e.g. from OrthoMCL) to additional taxa using hidden markov models.

# Post Orthology Detection

- Check for approximate copy number and intron boundaries by mapping orthologs to a reference genome. Tools: BWA, TopHat, GSNAP.
- Choose single or low copy-number orthologs.
- Align orthologs across the taxonomic divergence you want to capture.
- [MYcroarray suggests around 10% divergence to guarantee capture.](#)

# Resources: Finding Repetitive DNA

- RepeatMasker
- RepeatFinder
- RepSeek
- DAWG-PAWS
- Many, many more

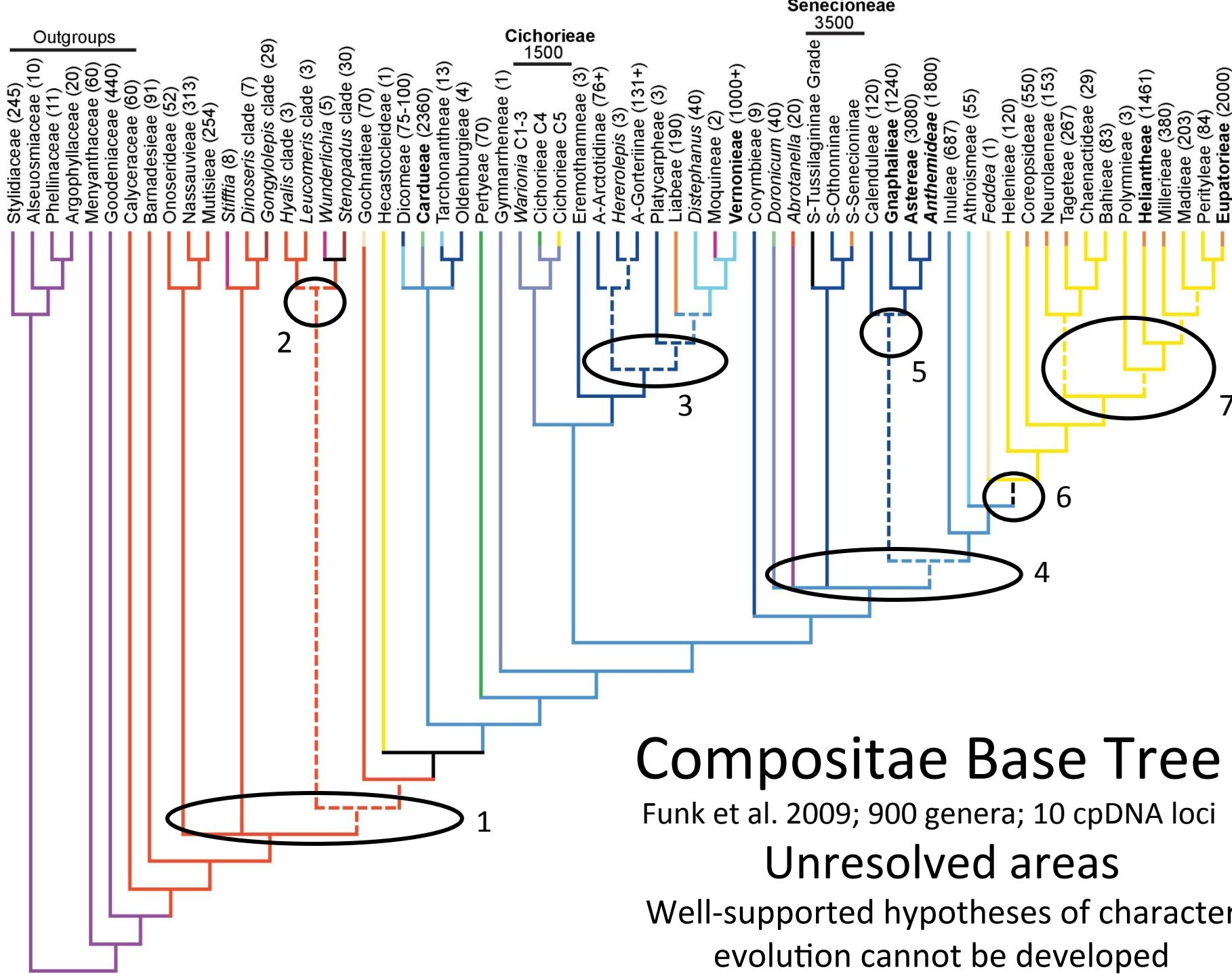
Good review: Lerat, E. 2009. Identifying repeats and transposable elements in sequenced genomes: how to find your way through the dense forest of programs. *Heredity* (2010) 104, 520–533.

[http://www.nature.com/hdy/journal/v104/n6/fig\\_tab/hdy2009165t1.html#figure-title](http://www.nature.com/hdy/journal/v104/n6/fig_tab/hdy2009165t1.html#figure-title)

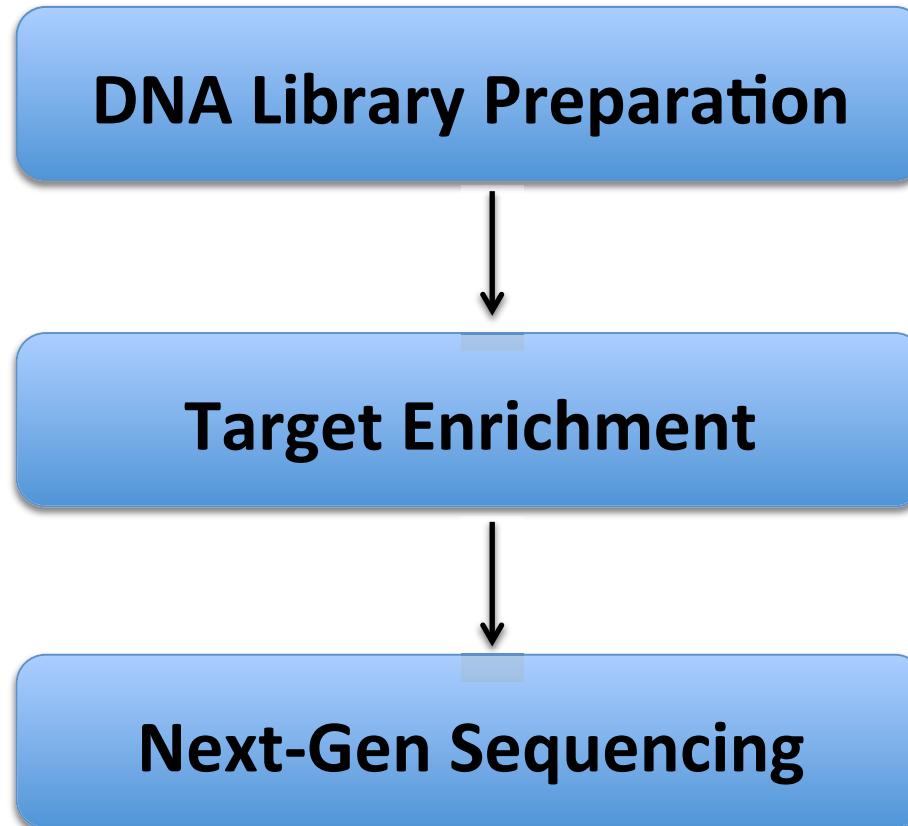
# Choosing Taxa and Best Practices

## Vicki Funk





# What is the wet lab workflow?



# DNA Input for Library Preparation

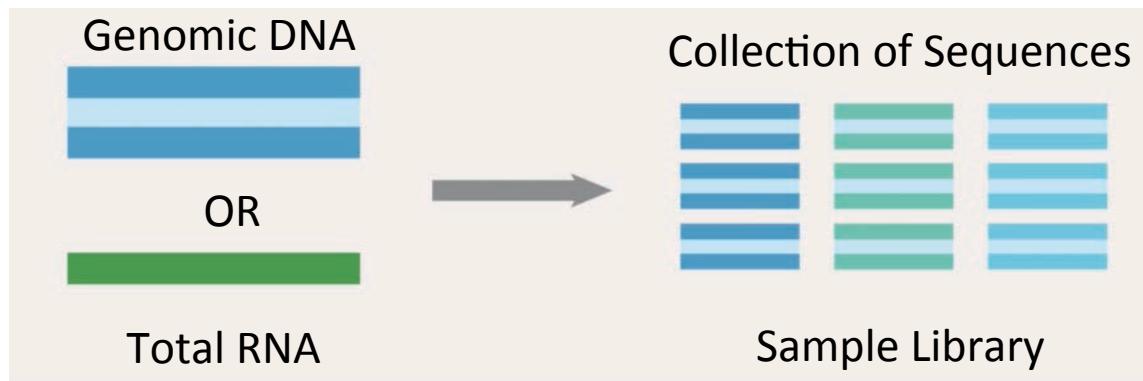
- Quality?
  - Needs to be relatively good
- Quantity?
  - 1 $\mu$ g is safe
- Nano for kicks (and quality), BUT MUST QUBIT!
- Herbarium specimens (1940s and beyond)
  - Degraded DNA, DNA Damage in Plant Herbarium Tissue, *Staats et al. 2011 PLoS One*
- Review on library prep: DNA Damage in Plant Herbarium Tissue, *Head et al. 2014, BioTechniques*

# Bias During Library Preparation

- Bias in sample composition
- GC-neutral fragments may be amplified more efficiently
- Some protocols are PCR-free (need lots of DNA)
- Library preparation methods for next-generation sequencing: Tone down the bias, *Dijk et al. 2014 Exp Cell Res*
  - Most current NGS library preparation protocols introduce significant biases
  - Simple modifications of existing protocols are often sufficient to reduce bias
  - PCR is a major source of bias but **Kapa HiFi polymerase** performs remarkably well (better than **Phusion**)

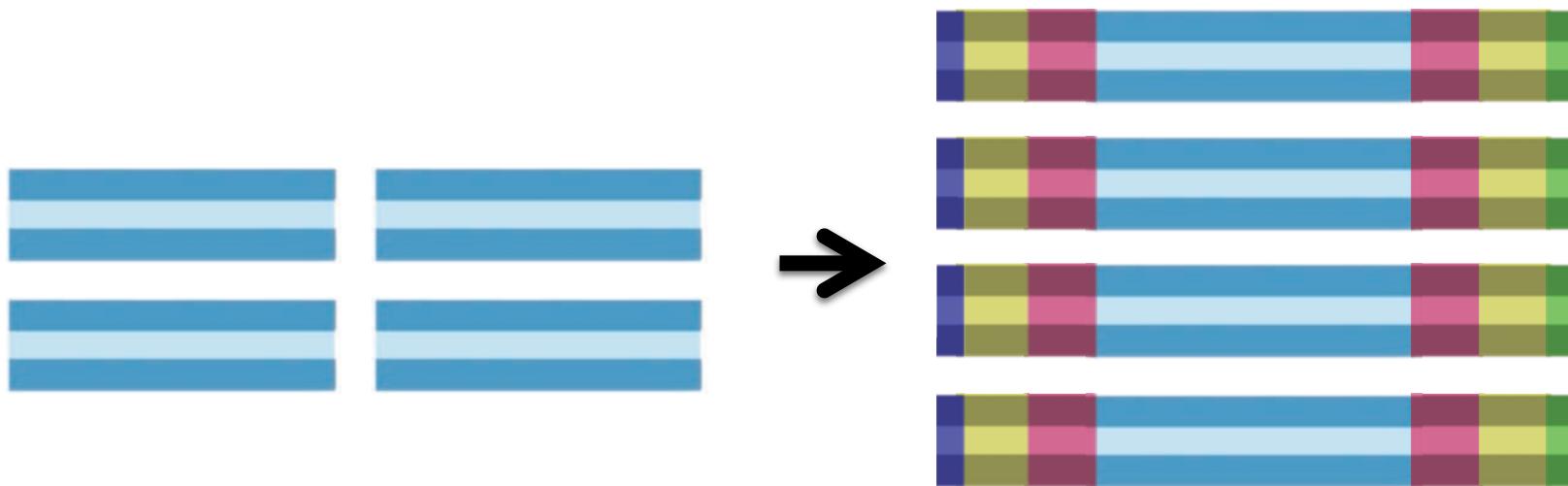
# Preparation of Sequencing Libraries

DNA is prepared into a “Sample Library”

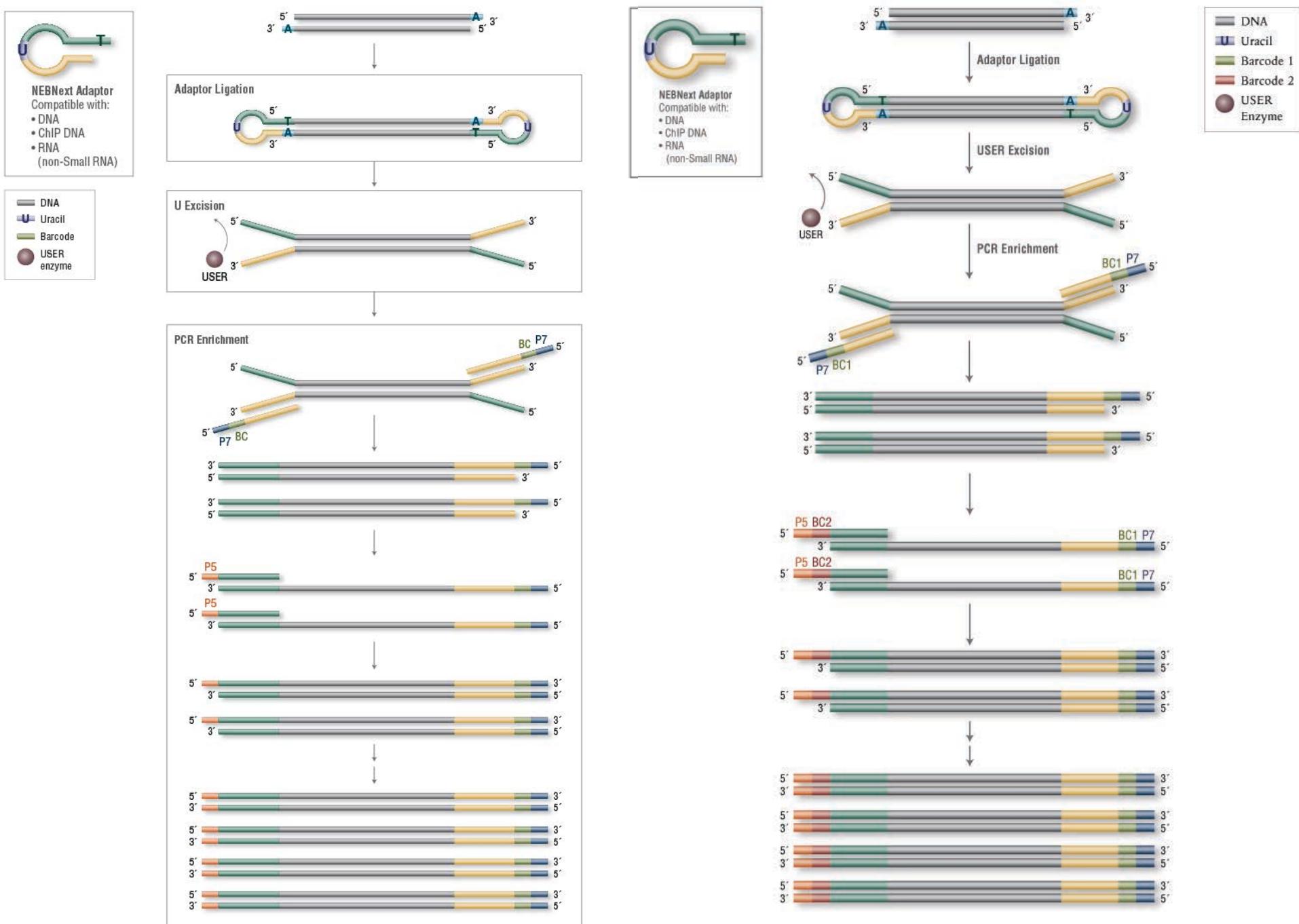


Genomic DNA is randomly sheared  
Covaris, sonicator (see costs later)

# Preparation of Sequencing Libraries



Size Selection  
Using hybridization and PCR:  
Adapters, barcodes, sequencing primers ligated to DNA

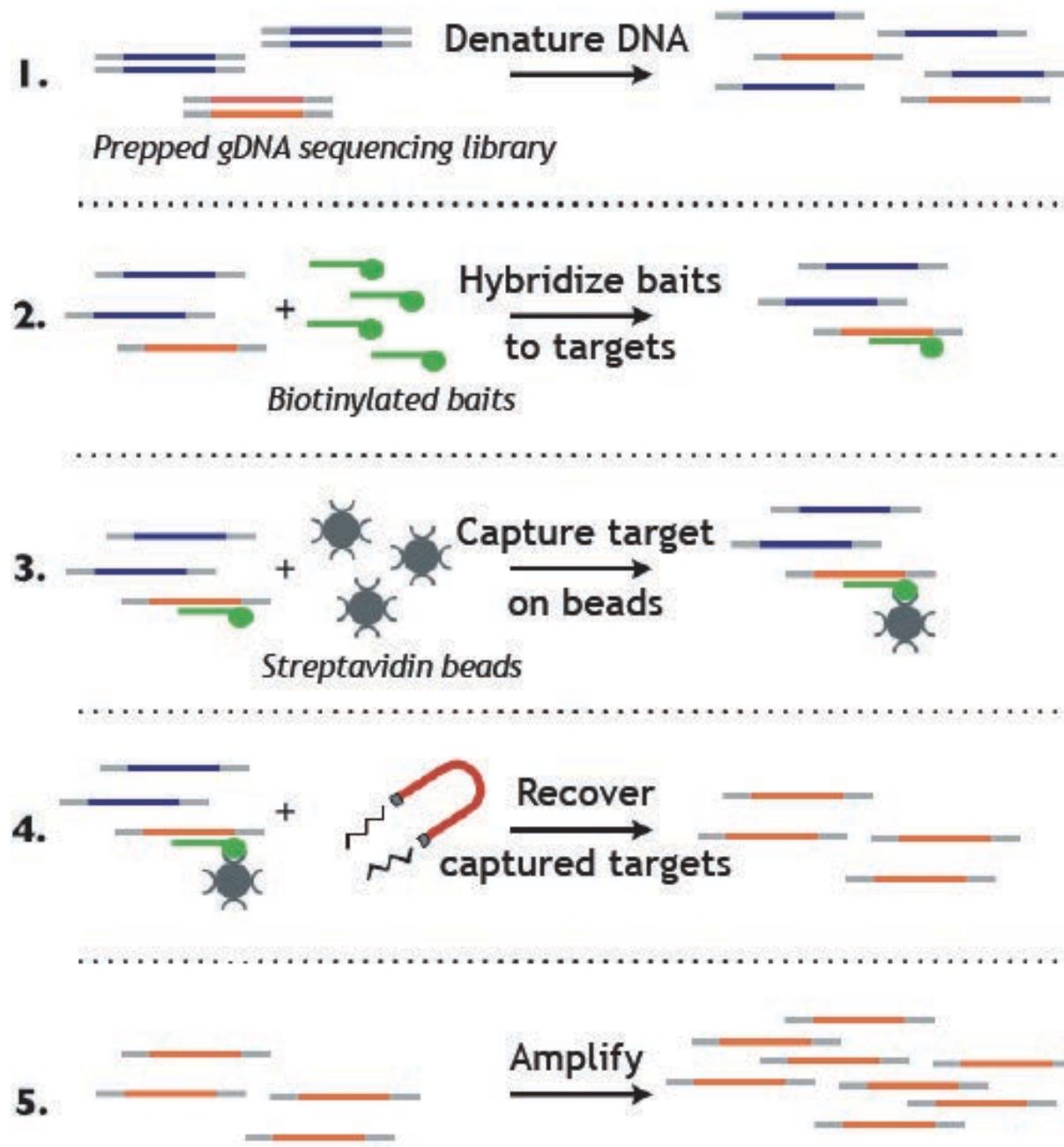


## NEB Library Prep for Illumina

# MYbaits

Sequence Enrichment for Targeted Sequencing

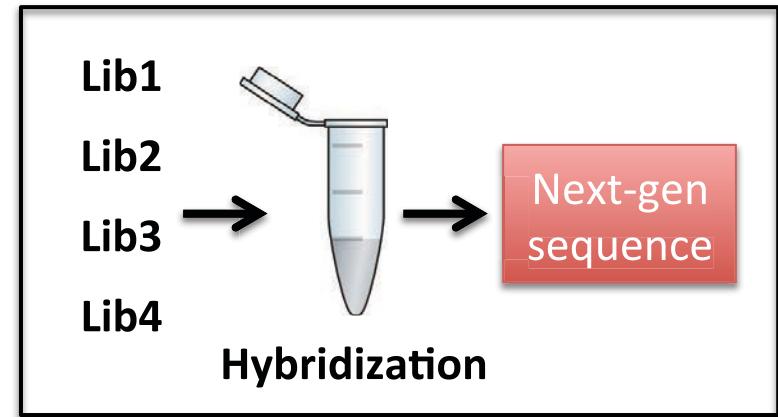




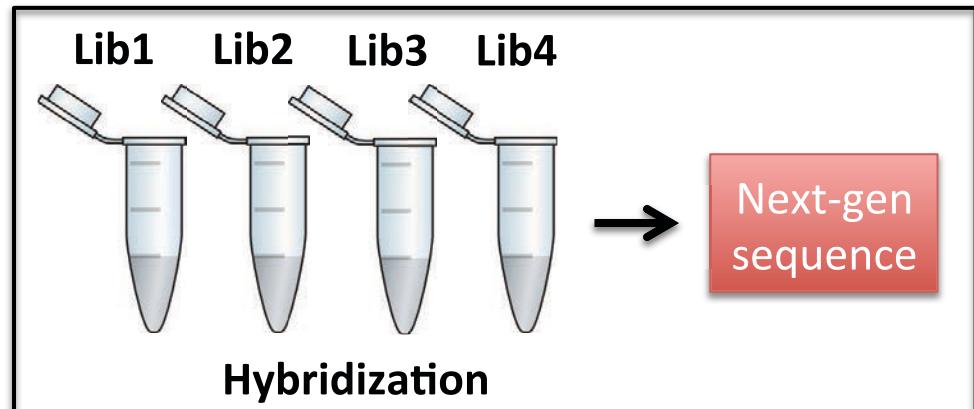
Target Enrichment/Sequence Capture

# Pooling Samples

- Pool before you MyBaits
  - Require smaller amount of each sample library
  - Cost reduction (pool 4 = 1/4 MyBaits cost)
  - Less flexibility



- Pool after you MyBaits
  - Uses more library
  - Cost is more
  - Flexibility



# Other Pooling Considerations

- Relationship of capture taxa to bait design taxa:

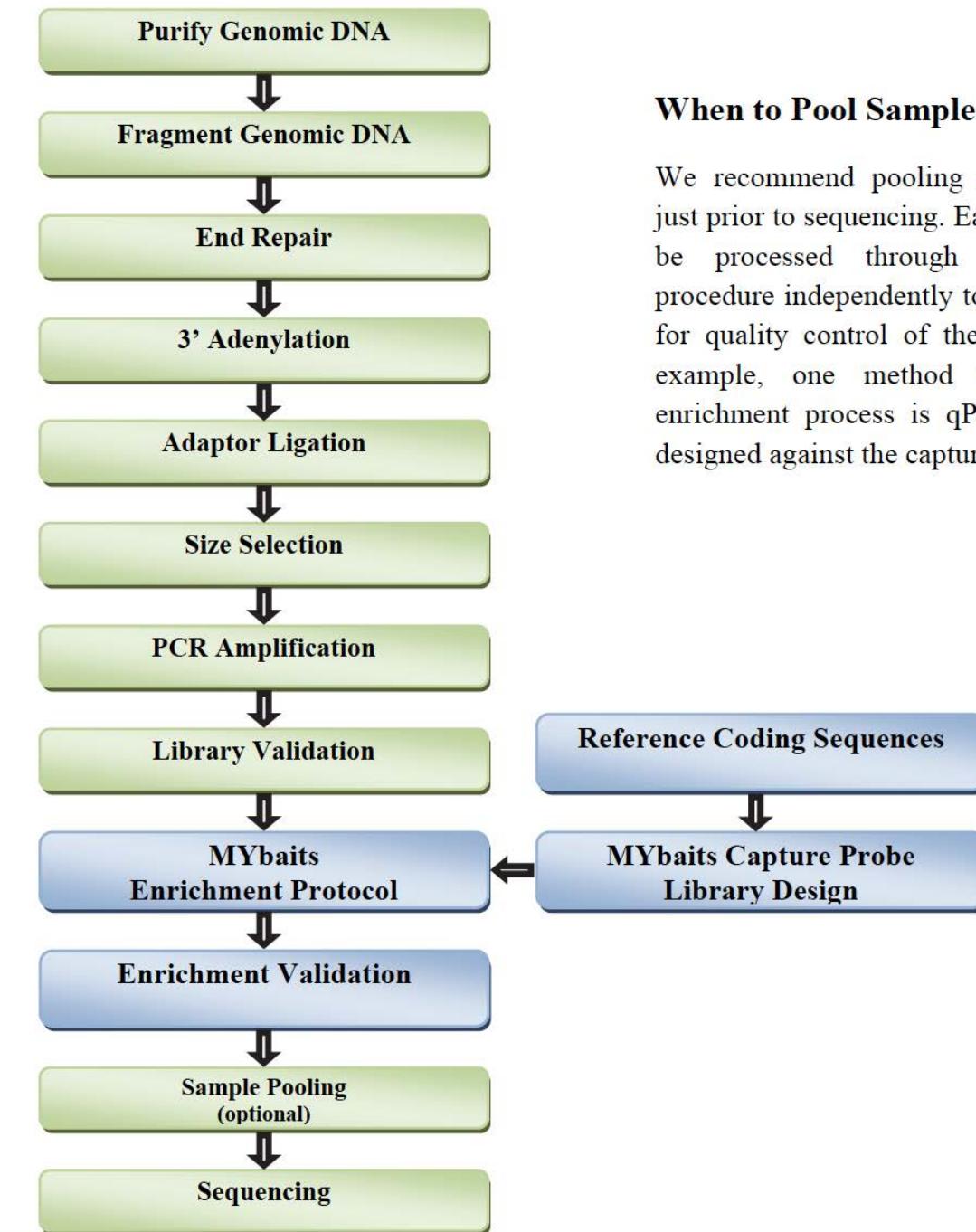
“...the two outgroup samples that were pooled in a hybridization with *Artocarpus* in the first sequencing run had much lower enrichment efficiency than ingroup samples. This suggests that multiplexing at the hybridization stage should be nonrandom, and only libraries of taxa that are relatively equidistant from the taxa used to design the bait sequences should be pooled...Therefore, pooling outgroup samples together in separate hybridizations may be advisable in future Hyb-Seq analyses.”

--Johnson *et al.* 2016 APPS

- Any variation in genome sizes

# MycroArray Recommends:

Figure 1. Library Preparation Workflow

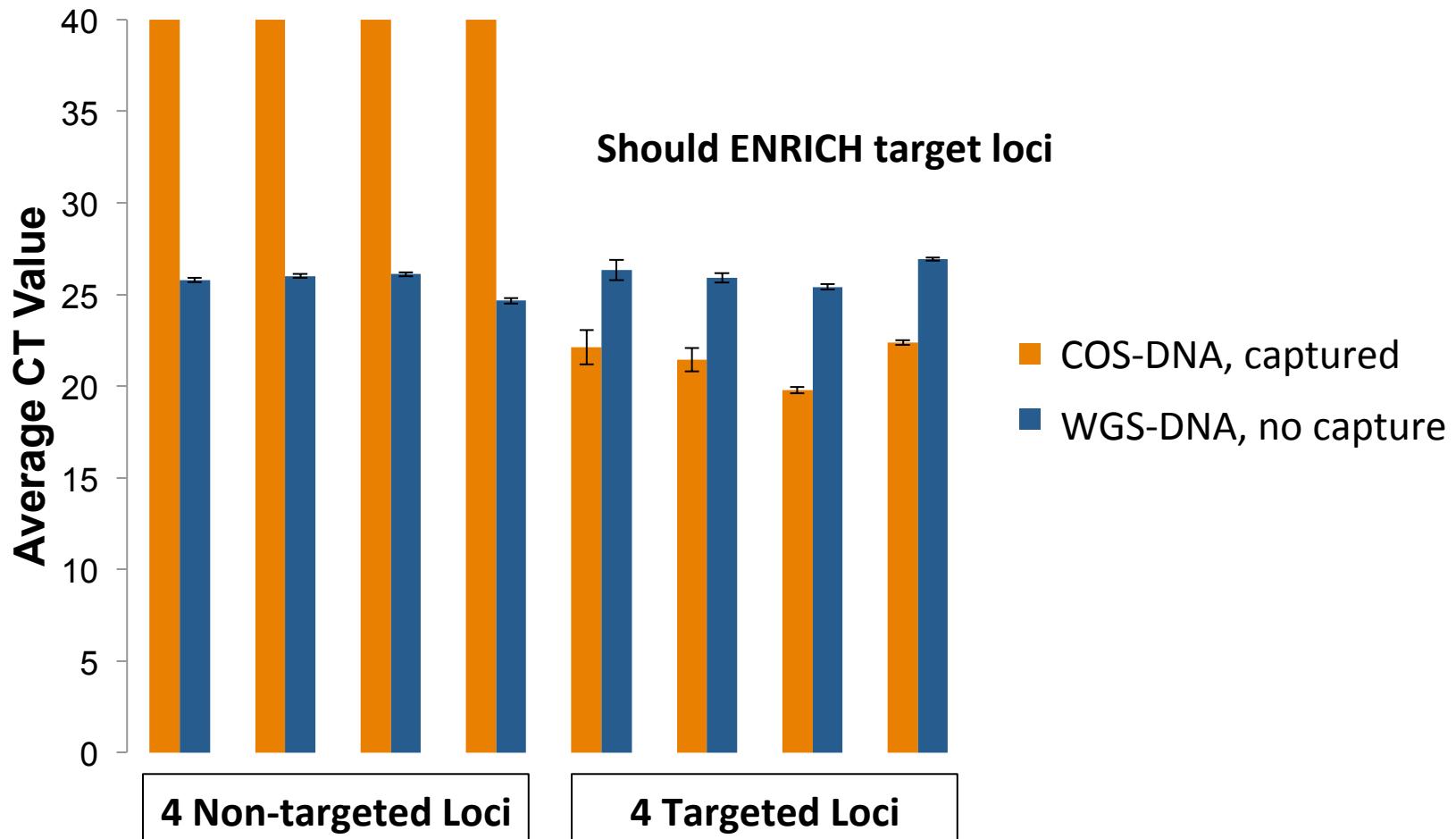


## When to Pool Samples?

We recommend pooling multiple samples just prior to sequencing. Each sample should be processed through the enrichment procedure independently to allow the ability for quality control of the enrichment. For example, one method to validate the enrichment process is qPCR with primers designed against the captured material.

# Enrichment of Targeted Loci - QPCR

Should NOT recover non-target loci



# Sequencing Platforms

- Illumina
- Ion Torrent
- Roche 454
- SOLiD
- PacBio

# Sequencing Platforms

Company (former companies)	Platforms	Carrier of library molecules or beads during sequencing	Sequencing principle	Nucleotide modifications	Signal detection method	Dominant type of sequencing error
Roche (454 until 2006)	454 FLX Titanium 454 FLX+ 454 GS Junior Titanium	Picotiterplate	Pyrosequencing	None (except for dATP, which is added as thiol derivative dATP $\alpha$ S)	Optical detection of light, emitted in secondary reactions initiated by release of PP <sub>i</sub> upon nucleotide incorporation	Indels in homopolymeric regions
Illumina (Solexa until 2007)	Illumina GAIx Illumina HiSeq1000 Illumina HiSeq1500 Illumina HiSeq2000 Illumina HiSeq2500 Illumina MiSeq Illumina NextSeq 500 Illumina HiSeq X ten	Flow cell	Reversible terminator sequencing by synthesis	End-blocked fluorescent nucleotides	Optical detection of fluorescent emission from incorporated dye-labeled nucleotides	Substitutions, in particular at the end of the read
Life Technologies (Agencourt until 2006, Applied Biosystems until 2008)	SOLiD 4 SOLiD 5500 SOLiD 5500xl SOLiD 5500 W SOLiD 5500xl W	FlowChip	Sequencing by ligation	2-base encoded fluorescent oligonucleotides	Optical detection of fluorescent emission from ligated dye-labeled oligonucleotides	Substitutions, in particular at the end of the read
Life Technologies (Ion Torrent until 2010)	Ion PGM Ion Proton	Ion Chip, a semiconductor chip	Semiconductor-based sequencing by synthesis	None	Transistor-based detection of H <sup>+</sup> shift upon nucleotide incorporation	Indels
Pacific biosciences	PacBio RS	SMRT cell	Single-molecule, real-time DNA sequencing by synthesis	Phosphor-linked fluorescent nucleotides	Real-time optical detection of fluorescent dye in polymerase active site during incorporation	Indels

<https://www.youtube.com/watch?v=HMyCqWhwB8E>

# Illumina Sequencing Overview

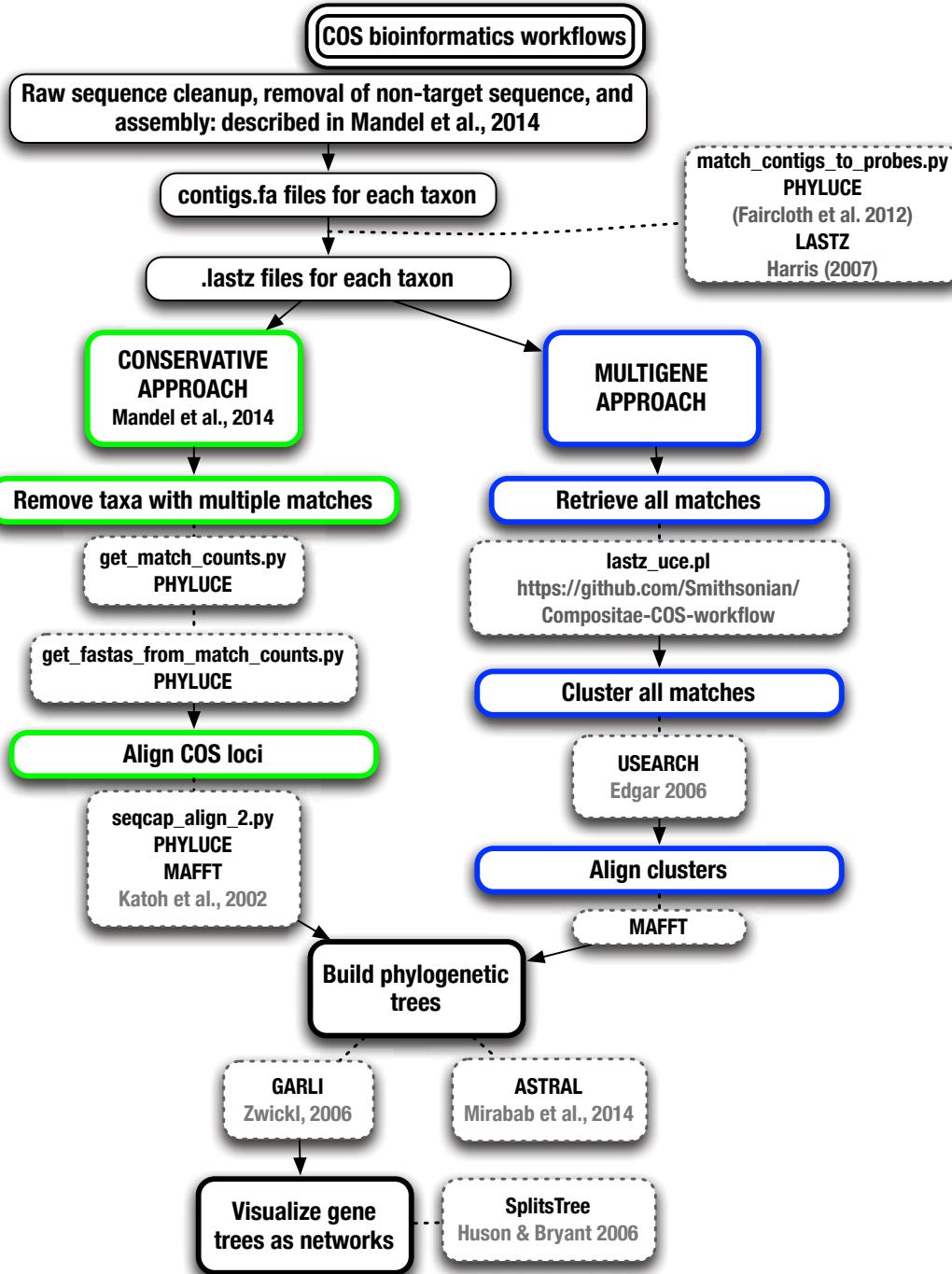
# Expected amount of data

- Illumina Sequencing
- MiSeq (500 cycles v2 kit) PE250 39 h 250 + 250  
30-40 million reads, 7.5-8.5Gb
- HiSeq 2500 Rapid Run PE100 27 h 100 + 100 600  
million reads, 60-70 Gb
- Coverage varies: recent MiSeq, 10-100+X/locus
- Will need to consider how many targets, how  
long, how many samples pooling, seq. platform

Questions before we go on?

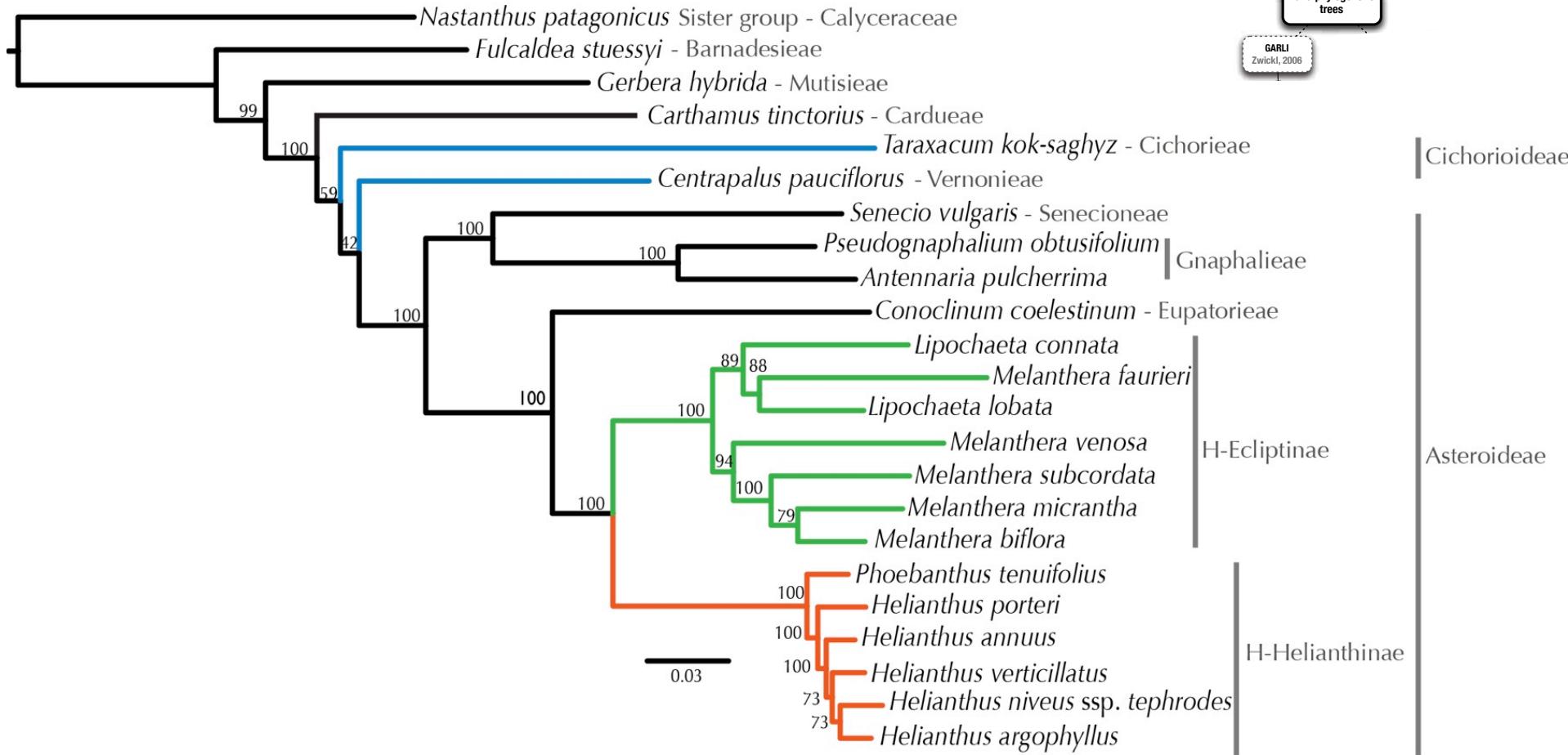
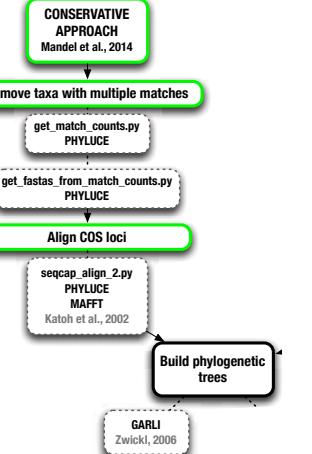
# Overview of Our Analysis/Results

## Jennifer Mandel



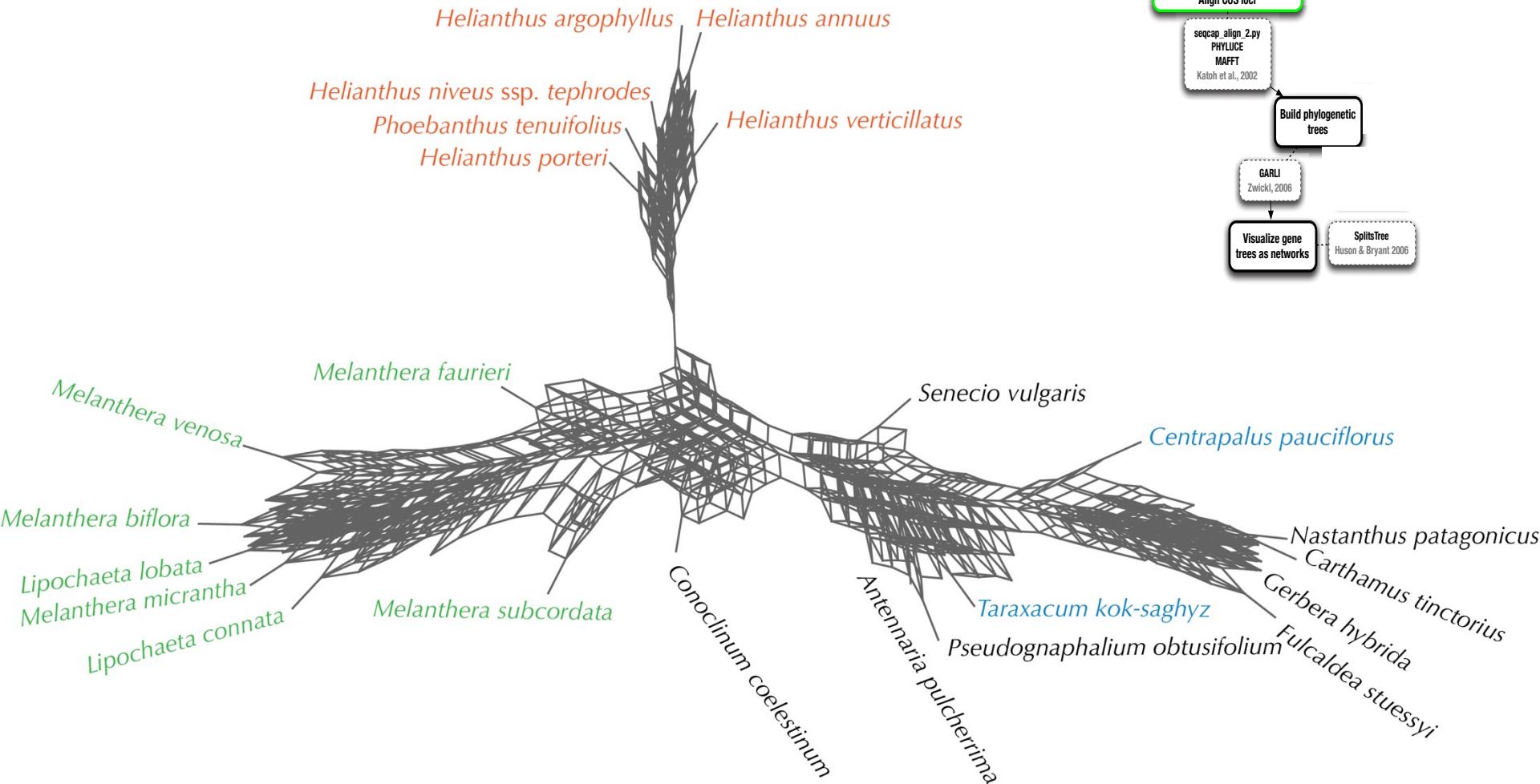
# Phylogeny based on conserved loci

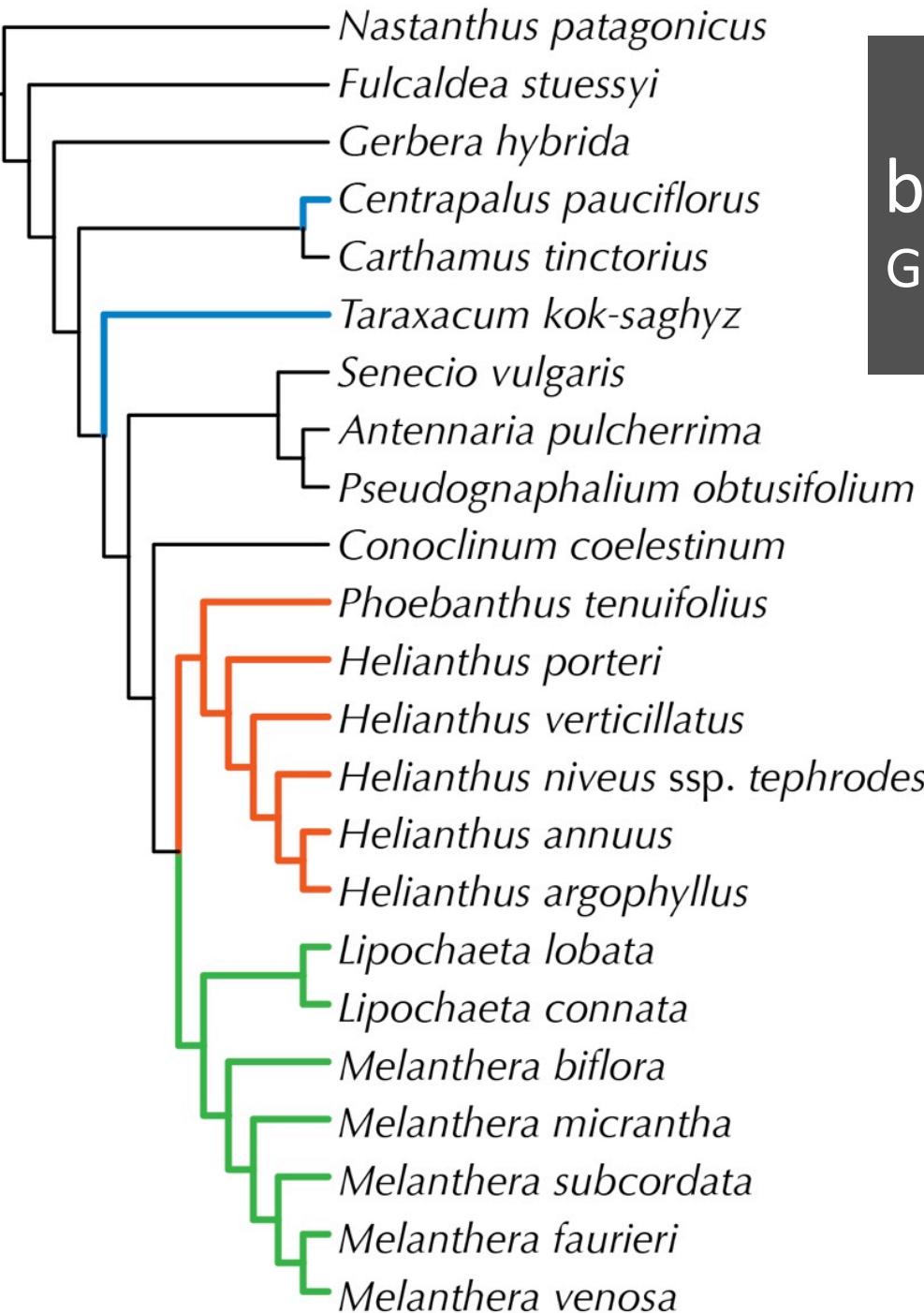
## Concatenated data set



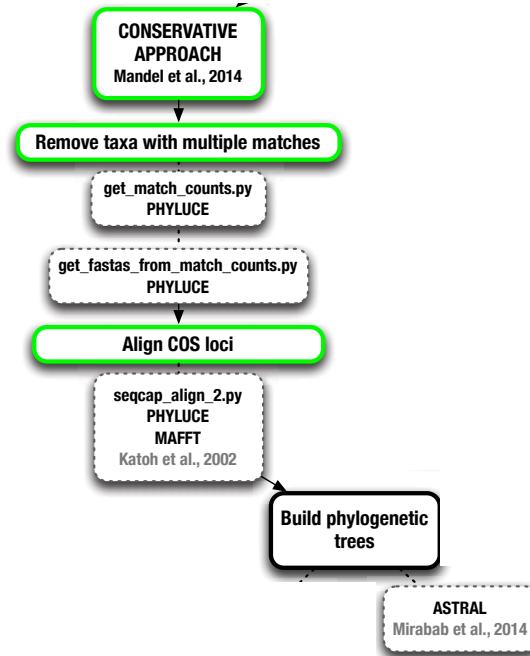
# Phylogeny based on conserved loci

## SuperNetwork of gene trees

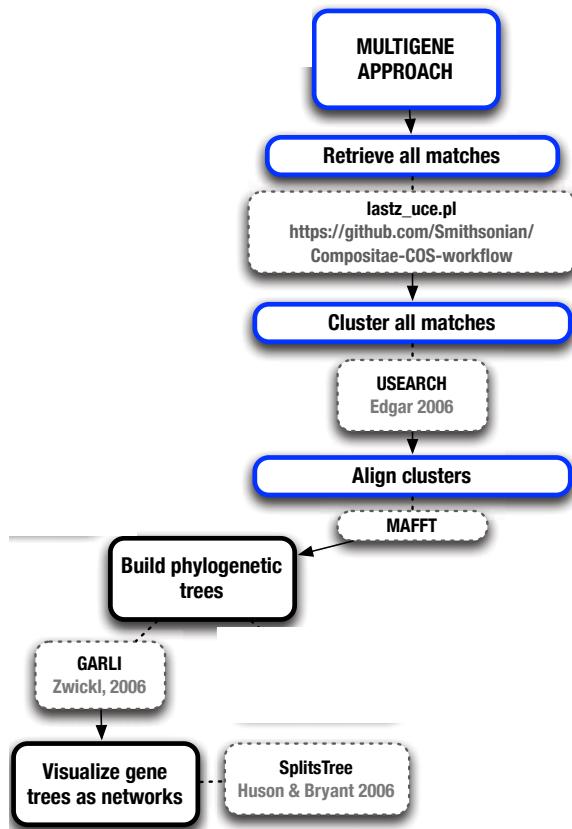




**ASTRAL phylogeny**  
 based on conserved loci  
 Gene trees analyzed to produce  
 a “species” tree



# What about a SuperNetwork of all Clusters?

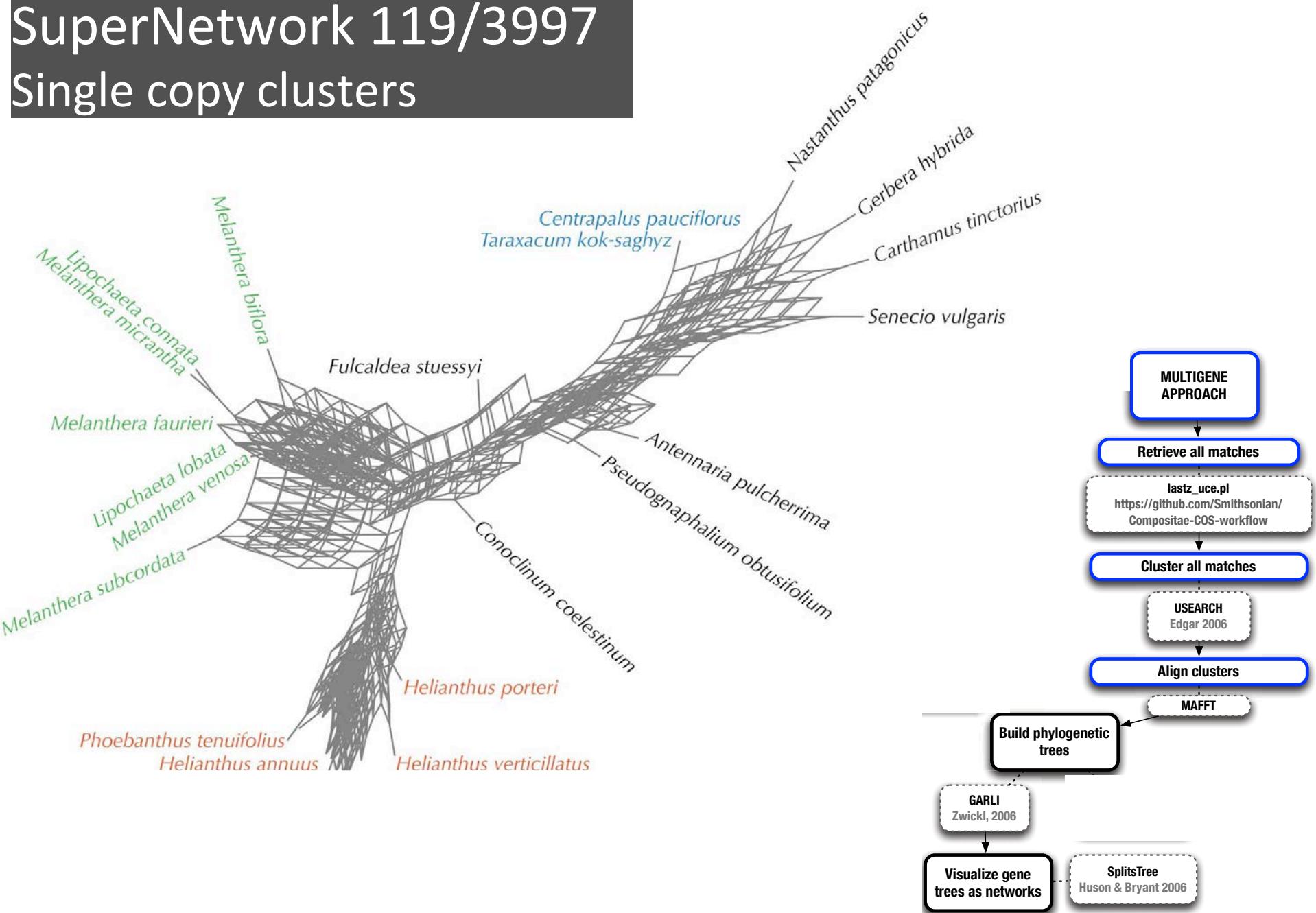


Species A:  
4 copies of Cluster 1  
1 copy of Cluster 2

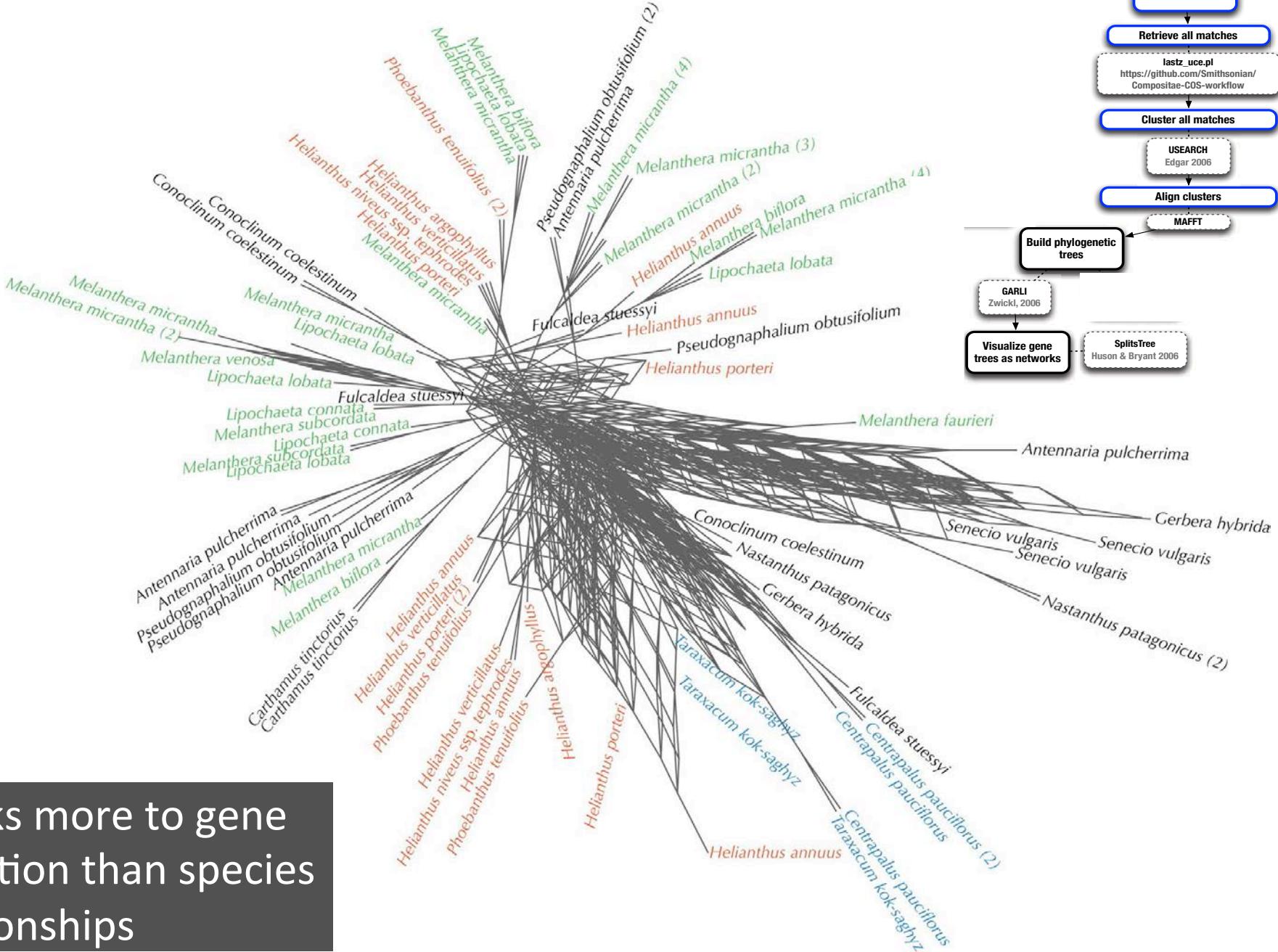
Species B:  
1 copy of Cluster 1  
6 copies of Cluster 2  
THEN multiplied by  
3997 clusters and  
23 species

# SuperNetwork 119/3997

## Single copy clusters

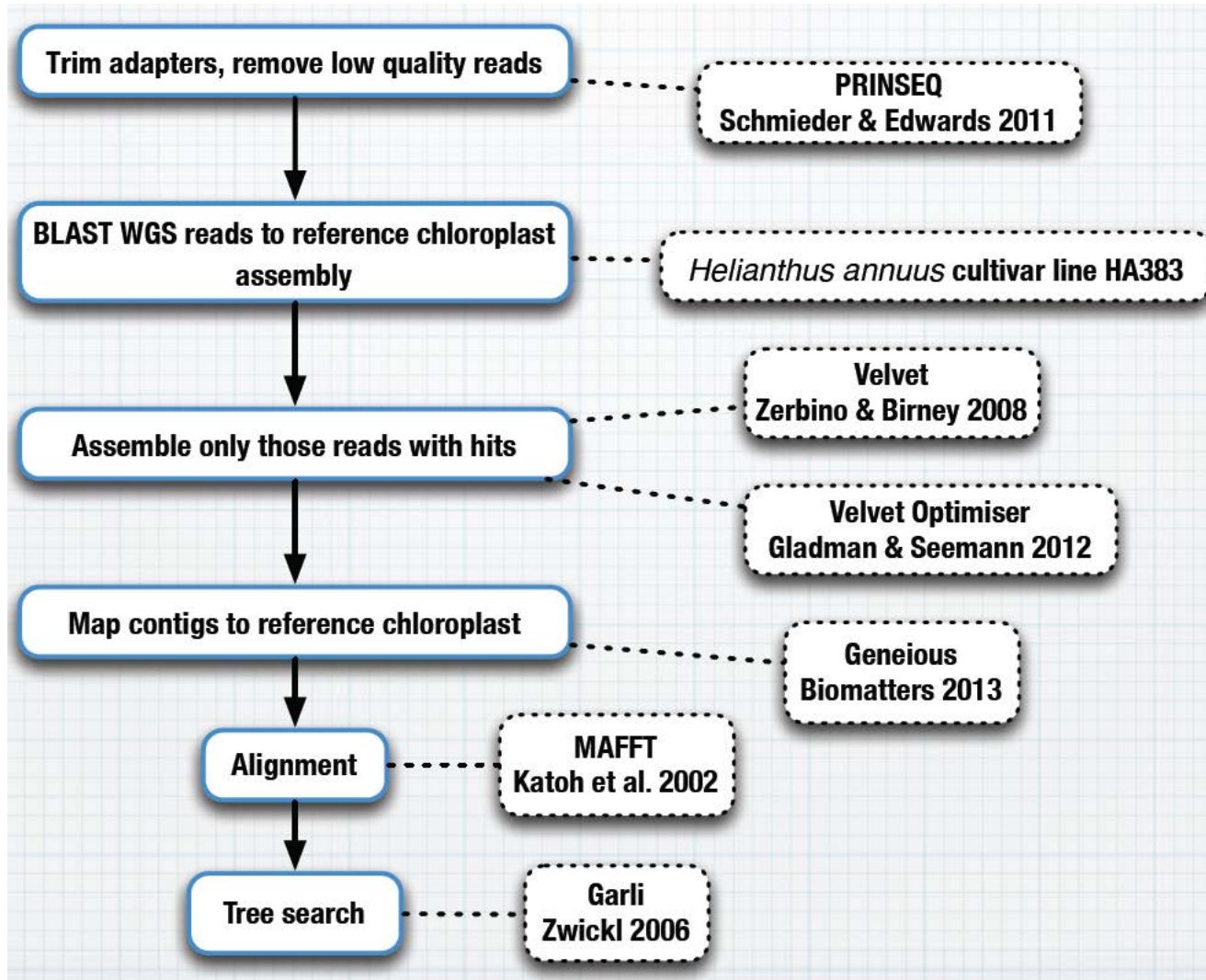


# SuperNetwork for 100 random clusters

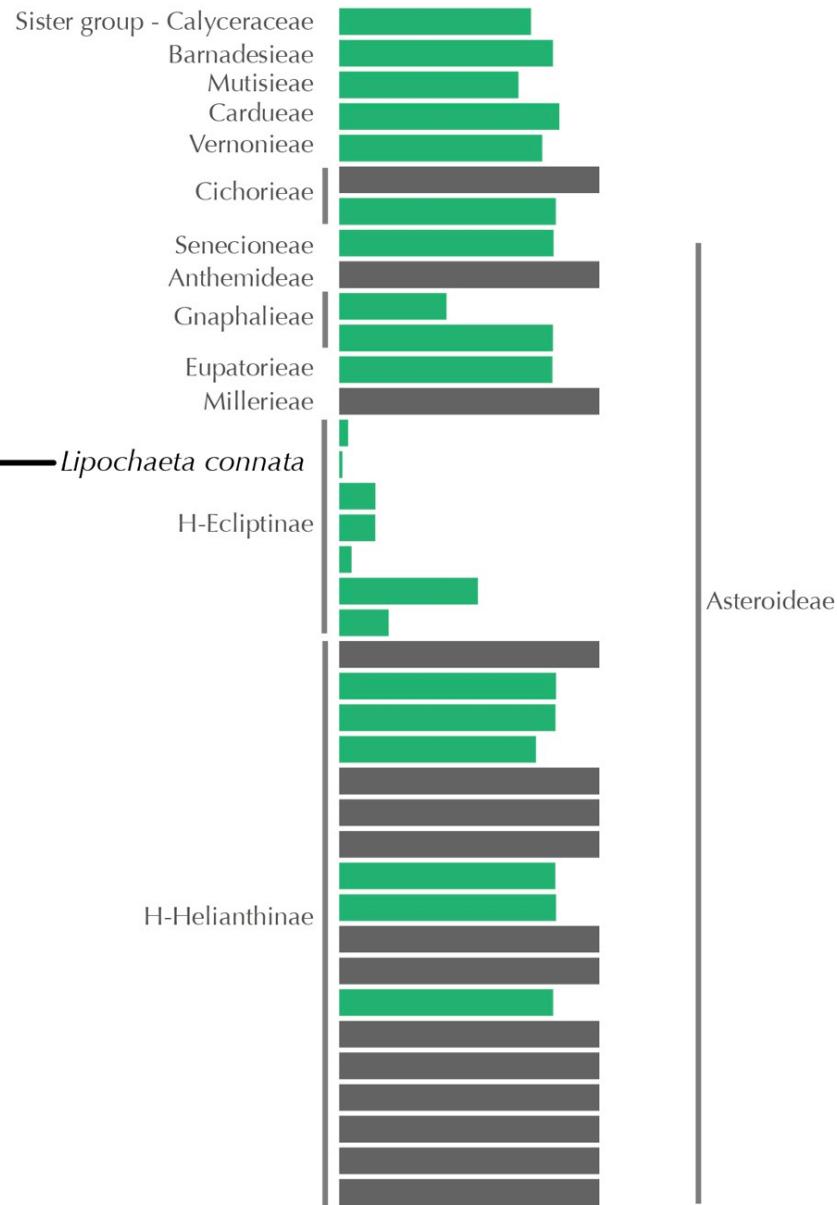


Speaks more to gene evolution than species relationships

# Chloroplast Assembly from WGS Added Bonus



# Chloroplast genome phylogeny with extra taxa



0.0090



Latest phylogeny based  
on conserved loci  
Concatenated data set

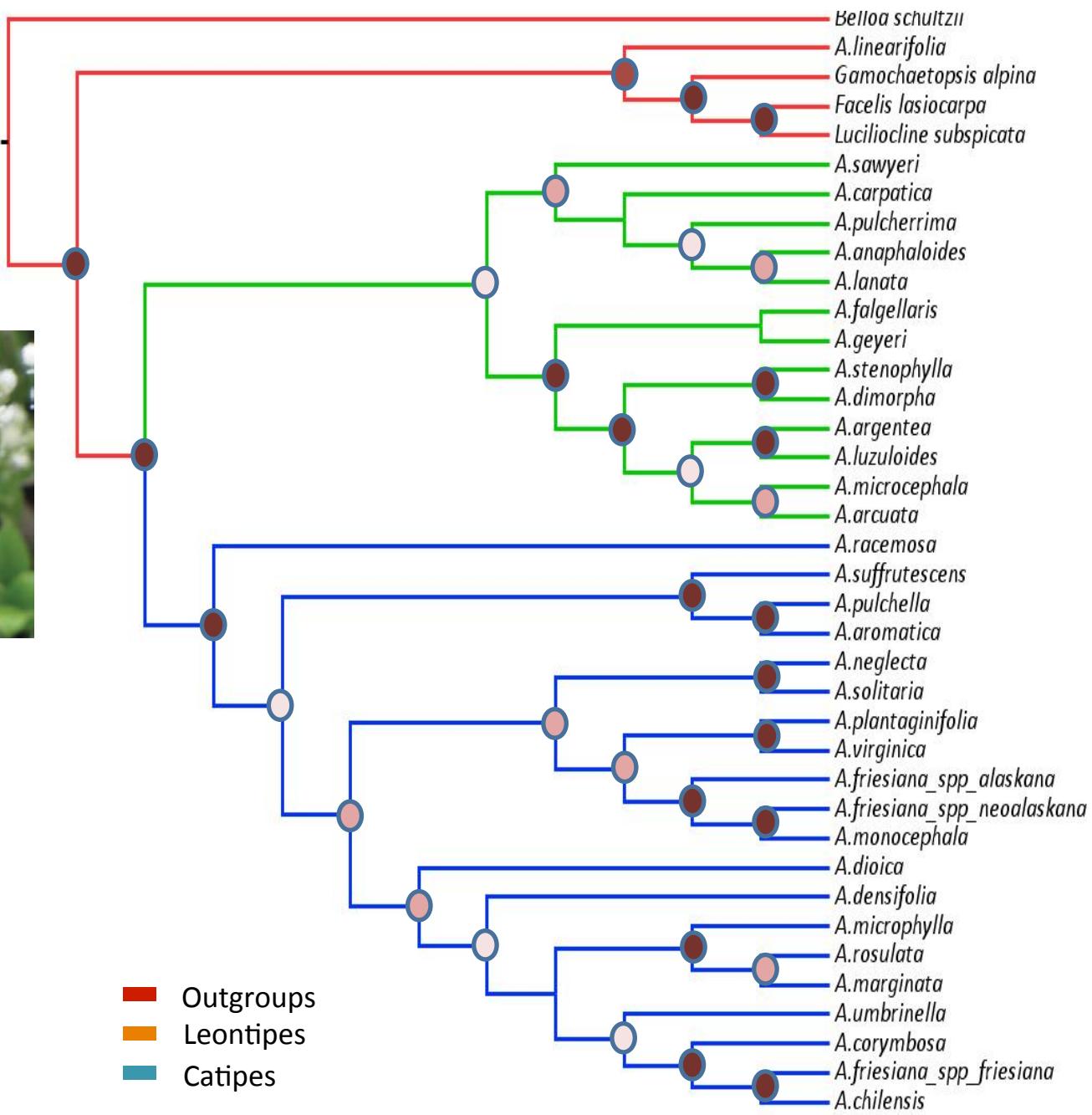
# *Antennaria* Intrageneric phylogeny



Support Values

- > 60%
- > 70%
- > 80%
- ■ > 90%

Maximum likelihood  
nuclear consensus tree  
825 COS loci



# Reconstruction of Phylogenetic Relationship in Antennaria (Asteraceae) using Data from Hundreds of Loci

Ram Thapa, Randall Bayer, Jennifer Mandel

Date: Tuesday, August 2nd, 2016

Time: 10:30 AM

Location: 201/Savannah International Trade and  
Convention Center

Abstract ID:429

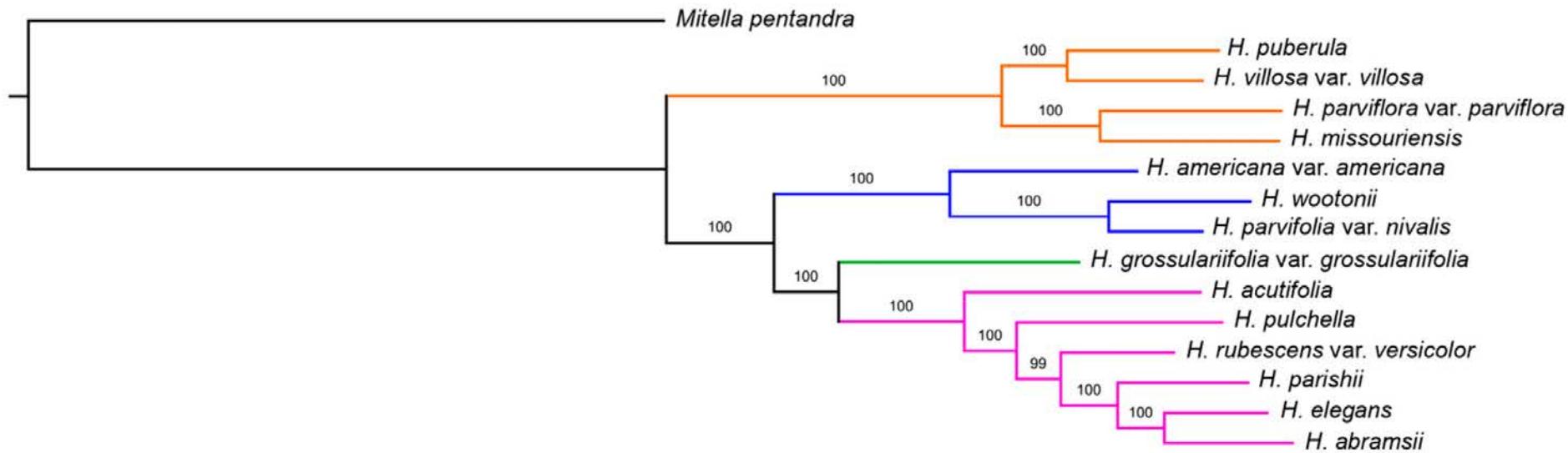


# *Heuchera* Study

Folk et al. 2015 APPS



- Targeted intron containing exons
- Recovered nuclear, cp-, and mt-DNA phylogenies



# Quick Break?

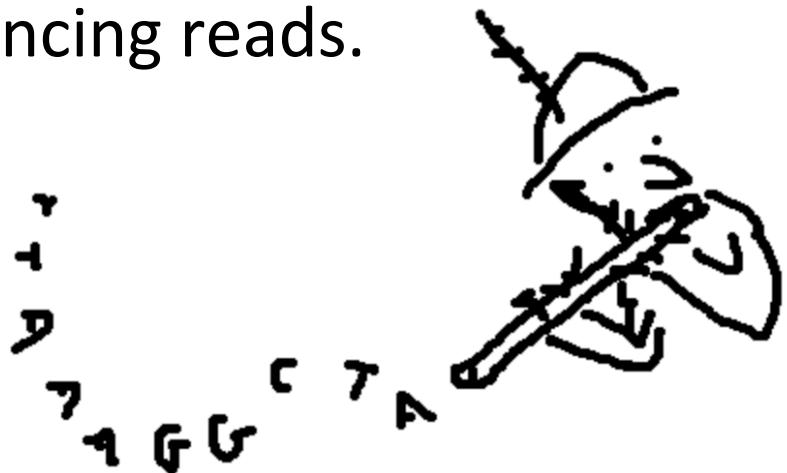
# Bioinformatics Nitty Gritty

## Rebecca Dikow

# HybPiper

Johnson et al. 2016 APPS

- HybPiper was designed for targeted sequence capture, in which DNA sequencing libraries are enriched for gene regions of interest, especially for phylogenetics. HybPiper is a suite of Python scripts that wrap and connect bioinformatics tools in order to extract target sequences from high-throughput DNA sequencing reads.



# Costs and Considerations

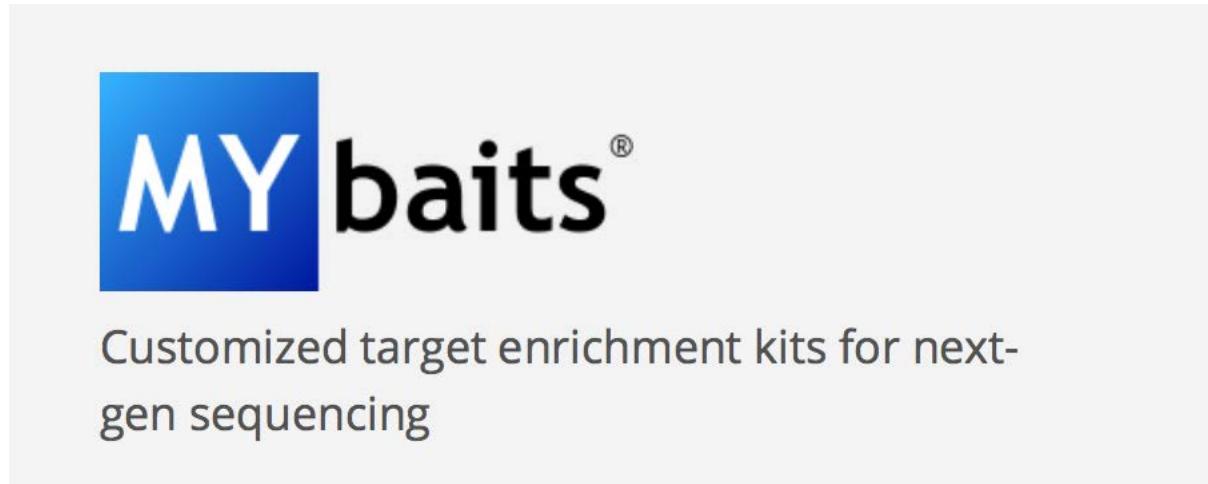
- How will you construct libraries?
- How many will you pool?
- Where will you sequence?

# Library Preparation

- 24 per lane with traditional TruSeq Kits
  - ~\$800/24 preps, \$33/each
- 96 indexes with newer kits (e.g., NEB)
  - \$28/each
- Sky is (mostly) the limit with indexes for custom libraries
  - <http://protocols.readthedocs.io/projects/protocols-libprep/en/latest/>
  - ~\$15/each, big upfront investment

# MyBaits Capture Probes: MycroArray

- Your custom design
- Few to thousands of genes can be targeted
- Target transcriptome, genomic, introns/exons



# MyBaits Capture Probes: MycroArray

Number of reactions	Baitset Tier ( <i>Maximum # of bait sequences</i> )				
	MYbaits-1 (20,000)	Per unpooled capture	Pooling 4 per capture	MYbaits-5 (50,000)	MYbaits-10 (200,000)
12	\$2,400	\$200	\$3,600	\$50	\$4,800
24	\$3,600	\$150	\$5,400	\$38	\$7,200
36	\$4,740	\$132	\$7,109	\$33	\$9,478
48	\$5,760	\$120	\$8,640	\$30	\$11,520
96	\$8,640	\$90	\$12,960	\$23	\$17,280
192	\$13,440	\$70	\$20,160	\$18	\$26,880
384	\$23,040	\$60	\$34,560	\$15	\$46,080
768	\$38,400	\$50	\$57,600	\$13	\$76,800

Kit item number: MYbaits-[# of modules]-[# of reactions]

<http://www.mycroarray.com/mybaits/mybaits-custom.html>

# MyBaits Capture Probes: MycroArray

- Pooling SAVES!
- Catalog CompCOS (1,061 loci, 20K probes)
  - 8 captures: \$80/each, or \$20/each
  - 48 captures: \$44/each, or \$11/each
  - 96 captures; \$32/each, or \$8/each

# NGS Cost Considerations

- What platform? <http://www.molecularecologist.com/next-gen-fieldguide-2014/>
  - Illumina
  - 454
  - SOLid
  - Ion Torrent
  - PacBio
- Where to do the sequencing?
  - Do some shopping around
  - E.g., HiSeq lane ~\$2K-3,500
- How many samples to run in one lane? PE, SE?
  - Can load up to 24 per lane with traditional TruSeq Kits
  - Can load up to 96 with newer kits
  - Can load any number with custom libraries:  
<http://protocols.readthedocs.io/projects/protocols-libprep/en/latest/>
  - Depending on platform: think about coverage?
- Who pools, you or them?

# Some Specifics on Illumina Costs

- Shop around
- Get recommendations
- In house/off-campus
- \$1,600 to \$4000, MiSeq to HiSeq
- <http://www.biotech.ufl.edu/services/service-fees/#nextgen>
- <http://dna.uga.edu/services/illumina-sequencing/prices/>
- <http://www.biotech.uiuc.edu/htdna/pricing>

# Extra, Hidden Costs

- Shearing DNA, tubes for Covaris ~\$5/each
- Quibit (for DNA quantity)
- Ampure beads (need for library prep)
- Dyna beads (need for capture)
- KAPA HiFi Taq (for prep/capture)
- Microtubes, use non-stick (capture)
- QPCR quantifying for pooling (unless seq. facility doing)

# Costs and Considerations

<b>Supplier</b>	<b>Product</b>	<b>Product #</b>	<b>Price</b>	<b># units</b>	<b>Price/unit</b>	<b>Needed for 1</b>
Invitrogen	Qubit dsDNA HS Assay	Q32851	79	100	0.79	3.16
Invitrogen	Qubit Assay Tubes	Q32856	70.25	500	0.14	0.56
Covaris	microTUBE 6x16	520045	125	25	5.00	5.00
NEB	Library Prep Kit	E7370L	2045	96	21.30	21.30
NEB	Index Primers 96	E6609S	650	96	6.77	6.77
Beckman Coulter	Agencourt AMPure XP 5 ml	A63881	300	5000	0.06	9.00
MycroArray	Capture Kit Custom	N/A	3130	96	32.60	8.15
			161.2			Pooling 4
KAPA	HiFi HotStart ReadyMix PCR Kit	KK2601	8	100	1.61	2.42
Fisher	MICRO-TUBES 2.0ML nonstick	50-212-313	41.4	500	0.08	0.83
Life Technologies	Dynabeads® MyOne™ Streptavidin C1	65002	1624	10,000	0.16	12.18
					Pre-Seq Total:	69.37
Sequencing (HiSeq)	Varies depending on facility		2800	48	58.33	58.33
					<b>Total per ind.</b>	<b>\$127.71</b>

# Where to get help?

- Companies
- Other researchers who have done this
- Workshop: Introduction to Next Generation Sequencing Straub et al.
- Us!
  - [jmandel@memphis.edu](mailto:jmandel@memphis.edu)
  - [DikowR@si.edu](mailto:DikowR@si.edu)
  - [FunkV@si.edu](mailto:FunkV@si.edu)

Other helpful links:

<http://core-genomics.blogspot.com/2014/04/choosing-ngs-library-prep-kit-provider.html>

[http://epigenome.usc.edu/services/nextgen/making\\_libraries.html](http://epigenome.usc.edu/services/nextgen/making_libraries.html)

[http://www.bioinformatics.nl/courses/bit2/docs\\_2013/NGS-18022013.pdf](http://www.bioinformatics.nl/courses/bit2/docs_2013/NGS-18022013.pdf)

# Questions?

