

NGS Applications

I. Whole Genome Sequencing & Resequencing

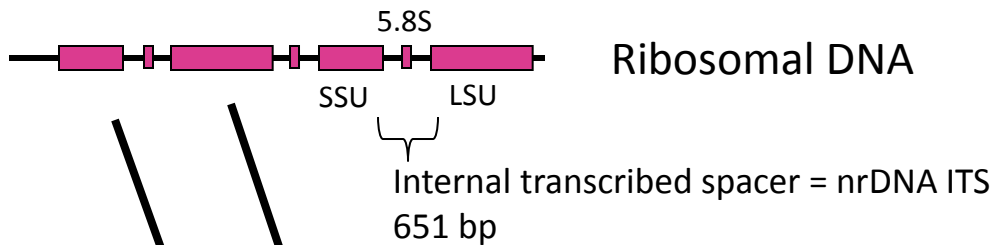
II. Targeted Sequencing (Enrichment)

Cronn et al. 2012. Targeted enrichment strategies for next-generation plant biology. *American Journal of Botany* 99:291-311.

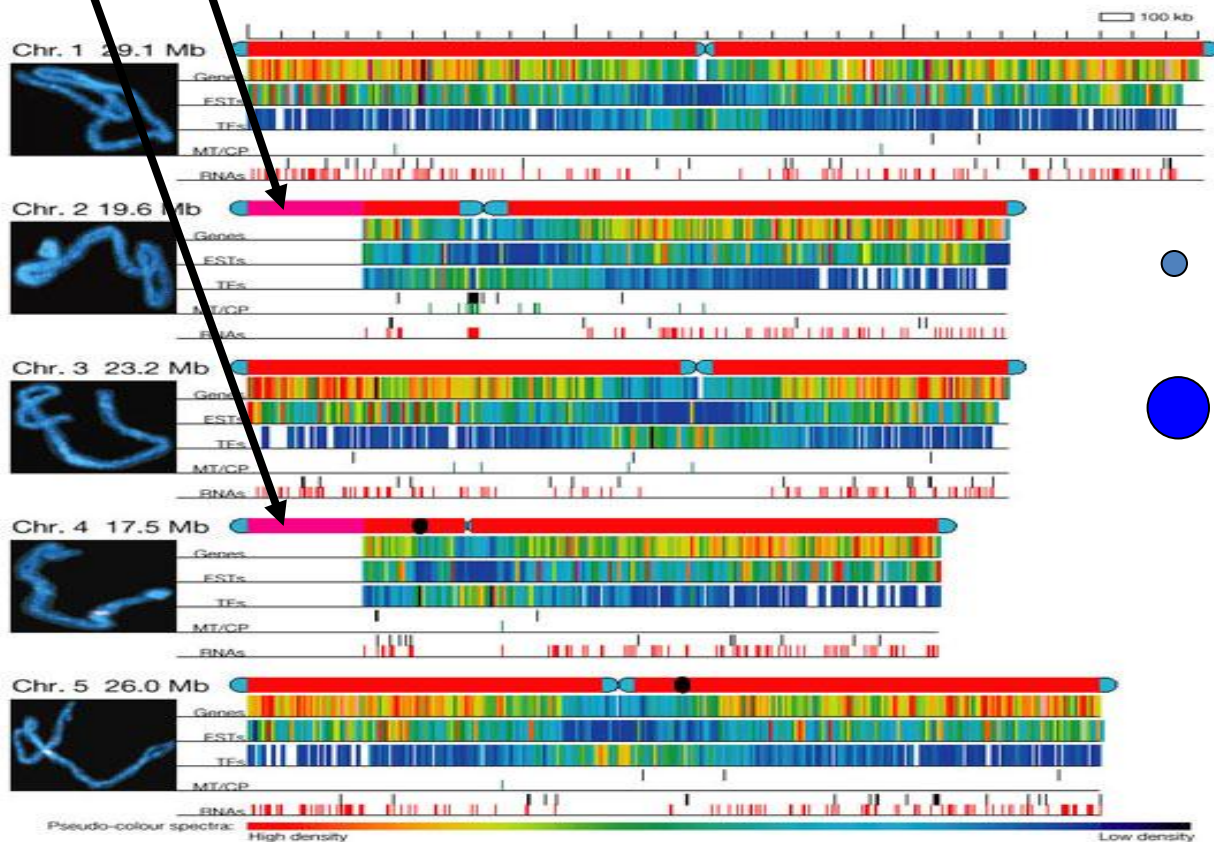
Straub et al. 2012. Navigating the tip of the genomic iceberg: next-generation sequencing for plant systematics. *American Journal of Botany* 99:349-364.

McCormack et al. 2013. Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetic & Evolution* 66:526–538.

Plant Genomes



Arabidopsis thaliana

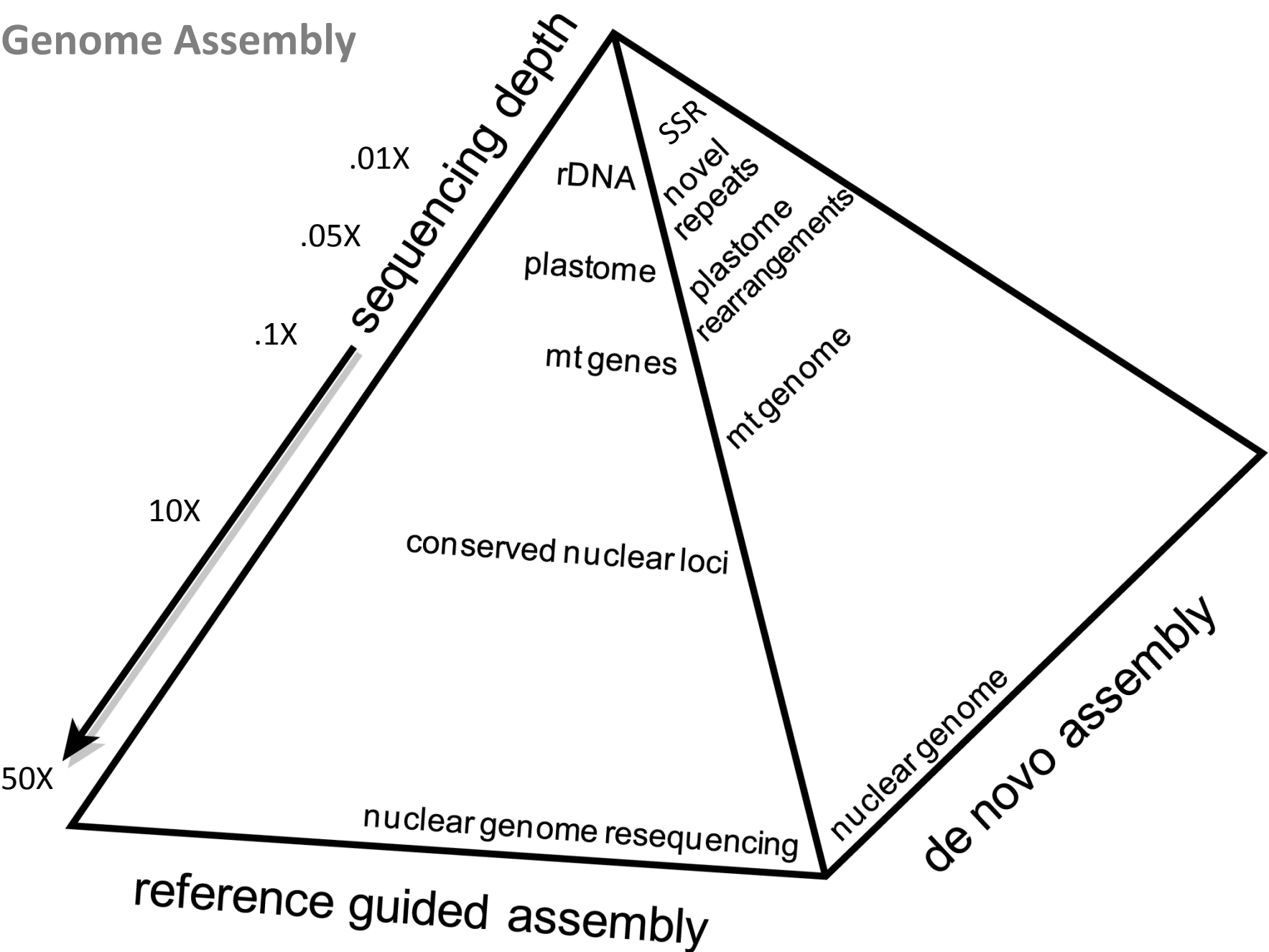


Chloroplast
genome 154
kbp

Mitochondrial
genome
367 kbp

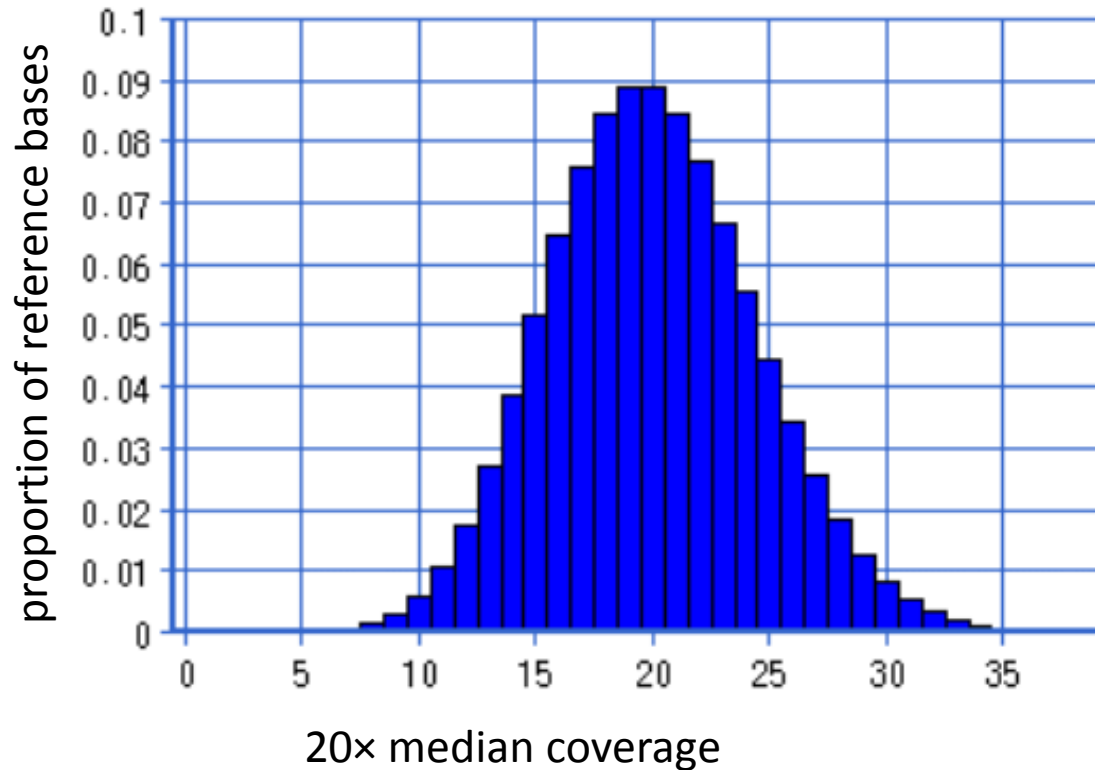
Arabidopsis Genome
Initiative. 2000. *Nature*
408:796-815.

Genome Assembly



Genome Assembly

Idealized genome coverage based on Poisson distribution
HIGH COVERAGE

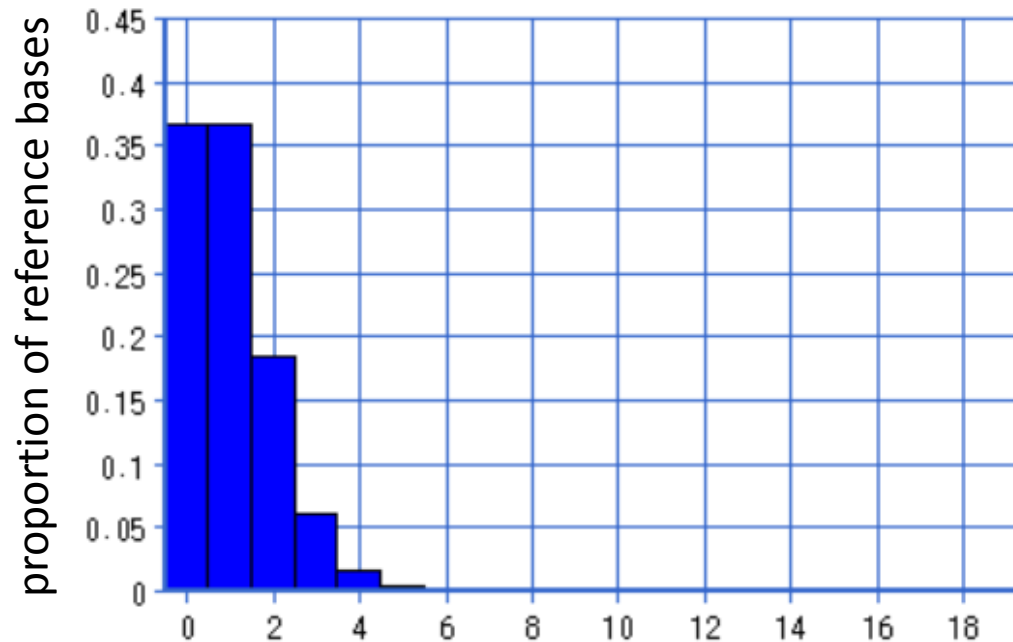


for a good overview: Wendl & Wilson 2008 BMC Bioinformatics 9:239.

<http://keisan.casio.com/has10>

Genome Assembly

Idealized genome coverage based on Poisson distribution
LOW COVERAGE

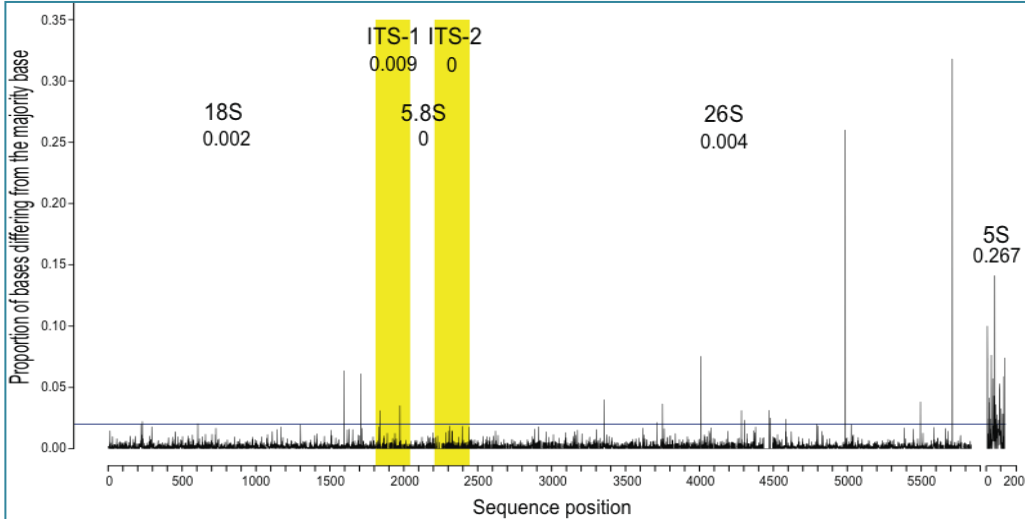


1x median coverage => 0.3% at 5x

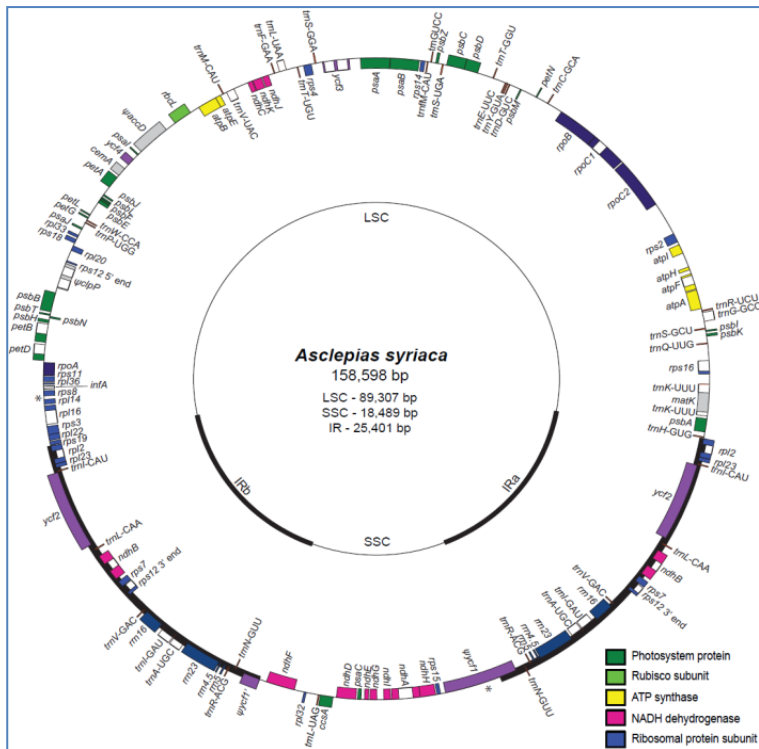
2x median coverage => 3.6% at 5x

Genome Skimming in *Asclepias syriaca*

- 20 million 40 bp reads
- 1x sequencing depth
- Reference guided assembly
- Nuclear Ribosomal DNA
- Chloroplast genome
- Mitochondrial genes
- Conserved single-copy nuclear genes

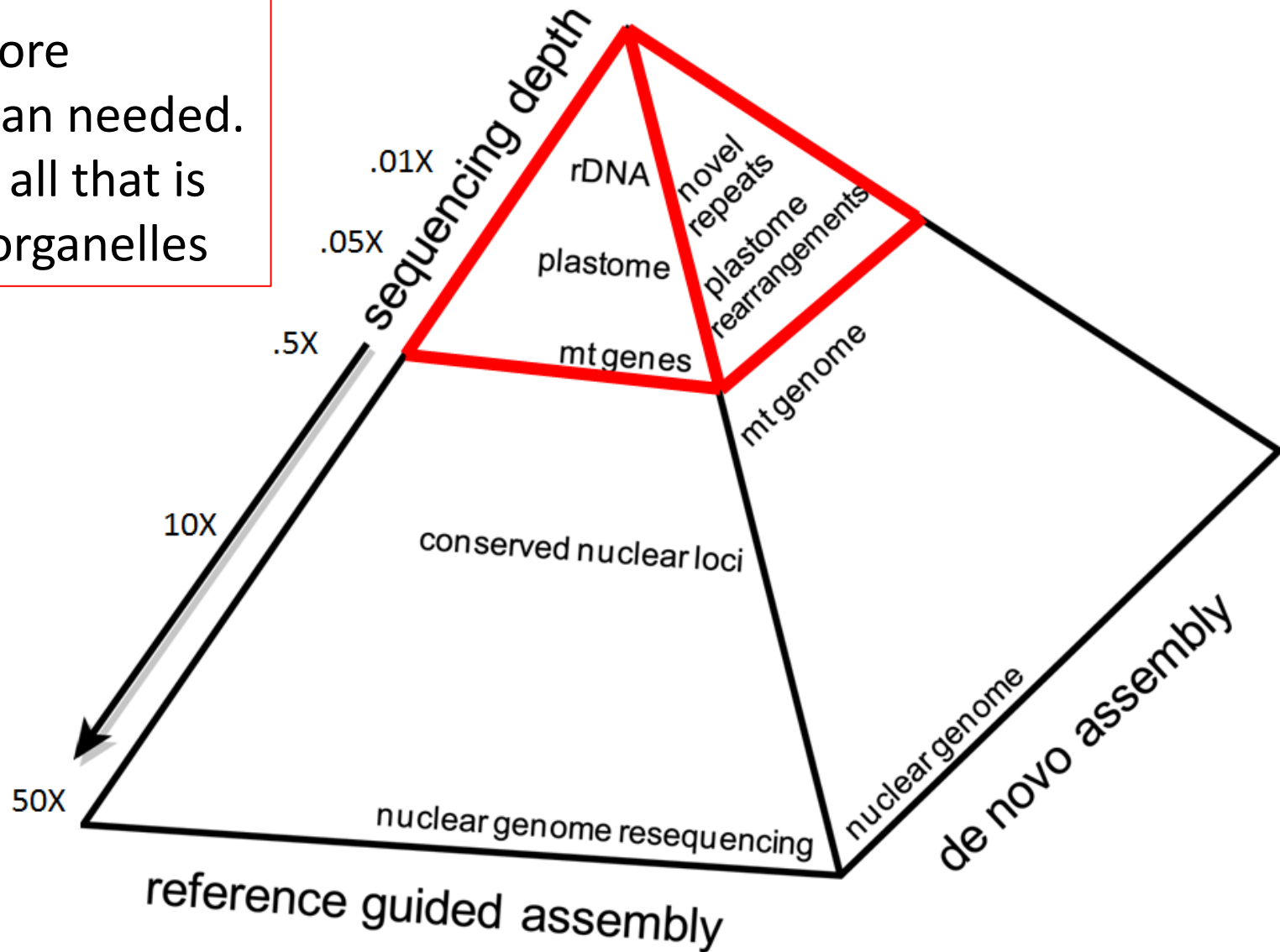


Straub et al. 2011. Building a model: Developing genomic resources for common milkweed (*Asclepias syriaca*) with low coverage genome sequencing. BMC Genomics 12: 211.



Genome Skimming

Important:
Don't use more
sequence than needed.
100-125× is all that is
needed for organelles



Genome Skimming

Microsatellite Development

(Jennings et al. 2011. Molecular Ecology Resources 11: 1060 – 1067)

Requires paired end (80 bp minimum) or long single end (>200 bp) reads

Low sequencing depth is sufficient for assembly-free methods



Targeted Sequencing

Restriction Digest Approaches (RAD-Seq, GBS)

<http://www.maizegenetics.net/>



Buckler Lab for Maize Genetics and Diversity

A USDA-ARS Lab with Cornell's Institute for Genomic Diversity

Main Menu

- [Home](#)
- [People](#)
- [Research](#)
- [Bioinformatics](#)
- [Publications](#)
- [Germplasm](#)
- [Jobs](#)
- [About Us](#)
- [Links](#)
- [Contact Us](#)
- [Site Map](#)
- [Repository](#)

Search Site

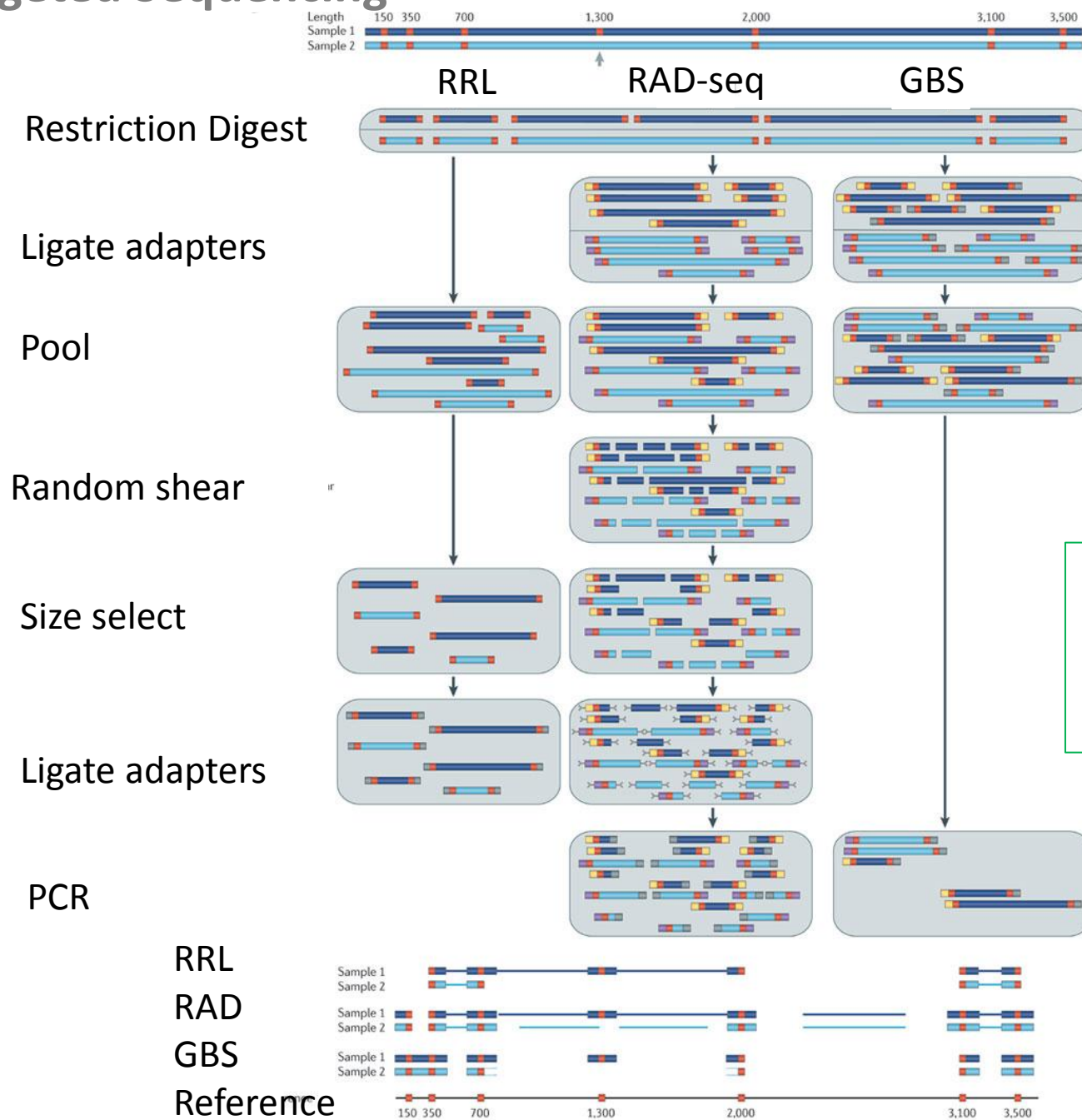
GBS Overview

Genotyping by sequencing (GBS) is a simple highly-multiplexed system for constructing reduced representation libraries for the Illumina next-generation sequencing platform developed in the Buckler lab by [Rob Elshire](#). Key components of this system are: reduced sample handling, fewer PCR and purification steps, no size fractionation and inexpensive barcoding. We use restriction enzymes to reduce genome complexity and avoid the repetitive fraction of the genome.

Quick Links:

- [GBS Bioinformatics](#)
- [Workshop Videos](#)
- [FAQ](#)
- [GBS Method Paper](#)
- [Presentation on GBS](#)
- [96 Plex GBS Protocol](#)
- [Dilution Calculator](#)
- [Bar Coded Adapter Generator](#) (outside link)
- [384 Plex ApeKI Adapters](#) (Updated May 11, 2012 to correct two bad bar codes.)

Targeted Sequencing



Modified from
Davey et al. 2011
Nature Reviews.
Genetics 12: 499 – 510.

Why We Don't Like Restriction Digest approaches

1. Generally requires at least 1 μg of good quality DNA.
2. Restriction digestion adds another variable to the library prep.
3. Does not target specific genes or SNPs.
4. References are easily obtained, and soon to be widely available.
5. Loci are generally not transferrable among species.

Hyb-Seq in *Asclepias*

Species	# unique reads	Targets hit	Targets assembled	Total length of assembly	% plastome assembled	% nrDNA cistron assembled
<i>Asclepias connivens</i>	830044	921	802	919877	98.3	100
<i>Asclepias engelmanniana</i>	1804956	924	901	2141384	97.8	98.3
<i>Asclepias eriocarpa</i>	384595	919	433	192346	81.9	94.4
<i>Asclepias flava</i>	1301608	922	875	1600197	98.4	100
<i>Asclepias humistrata</i>	843463	920	822	1129215	93.1	97
<i>Asclepias involucrata</i>	645580	924	752	748878	90.5	99.4
<i>Asclepias masonii</i>	971606	918	822	1003142	99.1	100
<i>Asclepias nyctaginifolia</i>	2295691	925	914	2203378	96	100
<i>Asclepias scheryi</i>	1295739	924	895	1962775	98.7	100
<i>Asclepias subaphylla</i>	21064	641	1	176	2.2	0
<i>Asclepias tomentosa</i>	1111909	918	834	1162912	95.2	99.7
<i>Calotropis procera</i>	1135014	917	857	1953095	96	100

Hyb-Seq in *Asclepias*

Species	# unique reads	Targets hit	Targets assembled	Total length of assembly	% plastome assembled	% nrDNA cistron assembled
<i>Asclepias connivens</i>	830044	921	802	919877	98.3	100
<i>Asclepias engelmanniana</i>	1804956	924	901	2141384	97.8	98.3
<i>Asclepias eriocarpa</i>	384595	919	433	192346	81.9	94.4
<i>Asclepias flava</i>	1301608	922	875	1600197	98.4	100
<i>Asclepias humistrata</i>	843463	920	822	1129215	93.1	97.0
<i>Asclepias involucrata</i>	645580	924	752	748878	90.5	99.4
<i>Asclepias masonii</i>	971606	918	822	1003142	99.1	100
<i>Asclepias nyctaginifolia</i>	2295691	925	914	2203378	96.0	100
<i>Asclepias scheryi</i>	1295739	924	895	1962775	98.7	100
<i>Asclepias subaphylla</i>	21064	641	1	176	2.2	0
<i>Asclepias tomentosa</i>	1111909	918	834	1162912	95.2	99.7
<i>Calotropis procera</i>	1135014	917	857	1953095	96.0	100
Average:		99.6%	87.5%	1.4 Mbp	95.0 %	99.0 %

Why we like the Hyb-Seq Approach

1. Like SNPs, sequence-based and easily combined/extended.
2. A single laboratory procedure and bioinformatics pipeline can be used for phylogenetics (deep and shallow), population genetics and genetic linkage mapping.
3. A relatively distant (e.g. plant family) reference can be used.
4. Candidate genes can be targeted.
5. Can be scaled from hundreds of genes to entire exomes (25,000 - 30,000 genes).
6. Comparable costs to other genotyping by sequencing approaches.

Targeted Sequencing: Hyb-Seq

Solution Phase Hybridization

Examples: Agilent SureSelect, Mycroarray MyBait

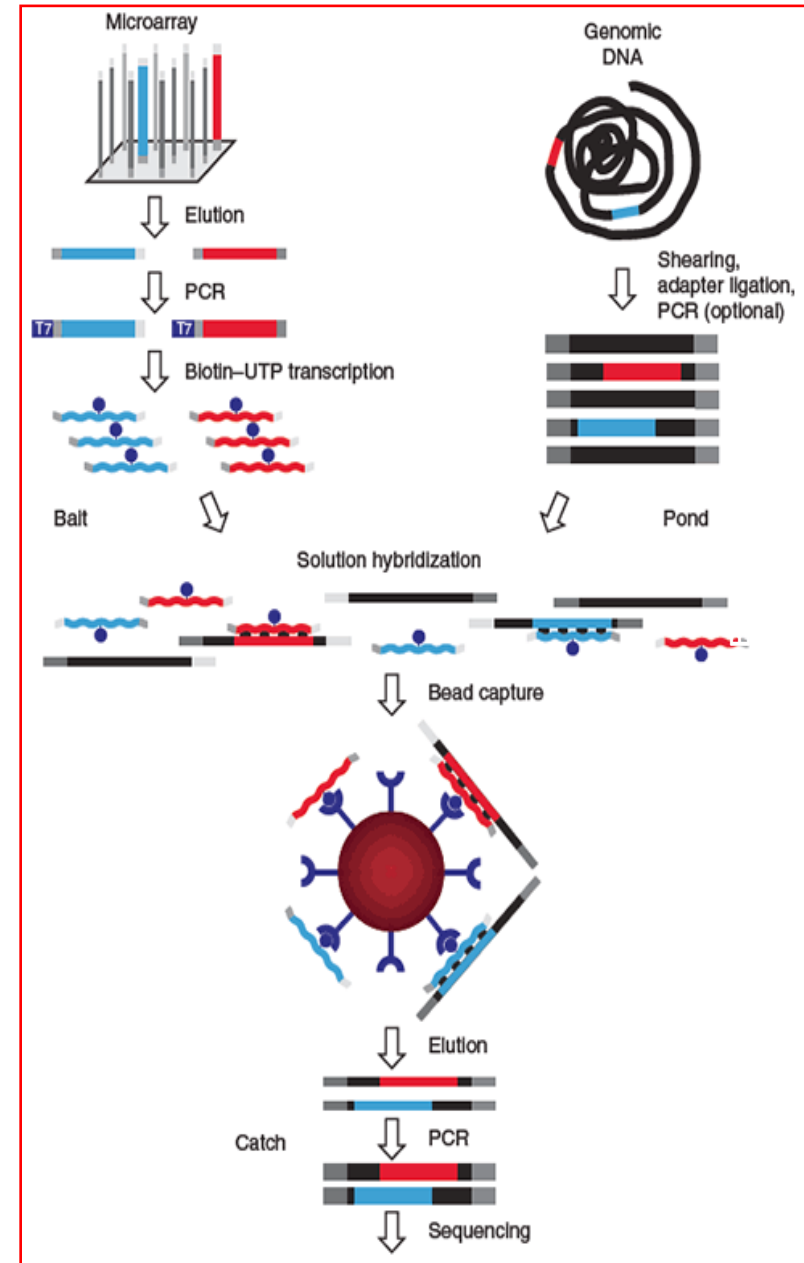
- ‘Baits’ synthesized on arrays
- 80-120 bp RNA probes
- Hybridization in solution
- Immobilization via biotin-streptavidin capture
- 100 – 500 ng DNA of input library

Cronn et al. 2012 Amer J Bot 99: 291-311

Lemmon et al. 2012 Syst. Biol.

McCormack et al. 2012 Syst. Biol.

Bi et al. 2012 BMC Genomics (*Tamias*)



**Size-selected library
Adapters Ligated
100 – 500 ng**

Targeted Sequencing: Hyb-Seq Enrichment

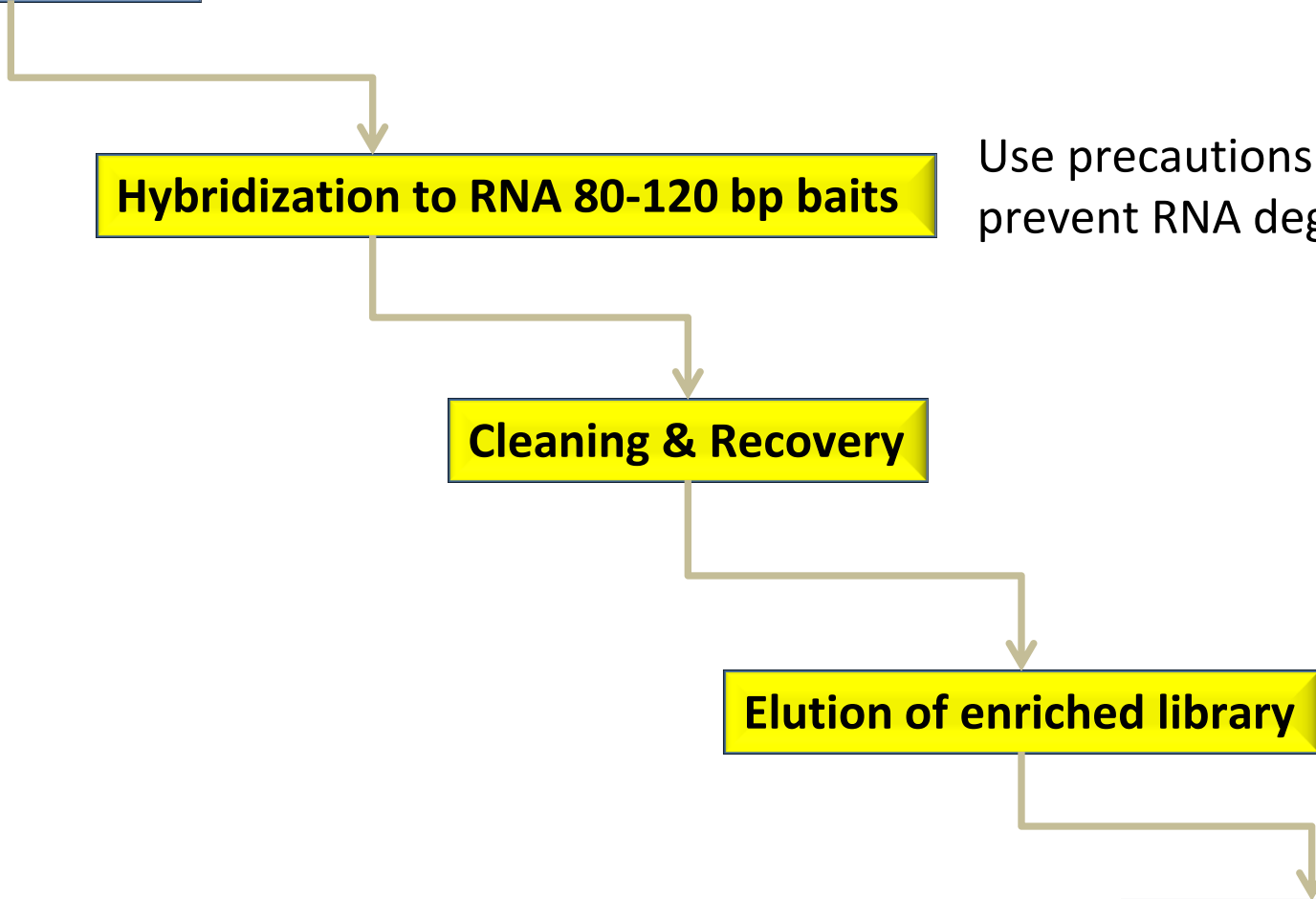
Hybridization to RNA 80-120 bp baits

Use precautions to
prevent RNA degradation!

Cleaning & Recovery

Elution of enriched library

PCR Enrichment



Slides from Kevin Go Here

Hyb-Seq Bait Design for 257 Genes in Rosaceae

<u>Genome</u>	<u>Exons</u>	<u>Baits</u>	<u>Target (bp)</u>
strawberry	1419	6307	448,163
peach	1425	6234	444,042
apple	1254	5857	422,886

Putatively orthologous exons in strawberry, peach and apple identified by sequence similarity

<15% sequence pairwise divergence across the genomes

>10% divergence among loci in a genome

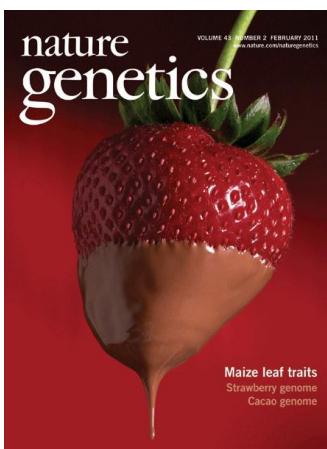
Single copy in strawberry and peach

Most have 2 or more copies in apple; a single paralog was chosen arbitrarily

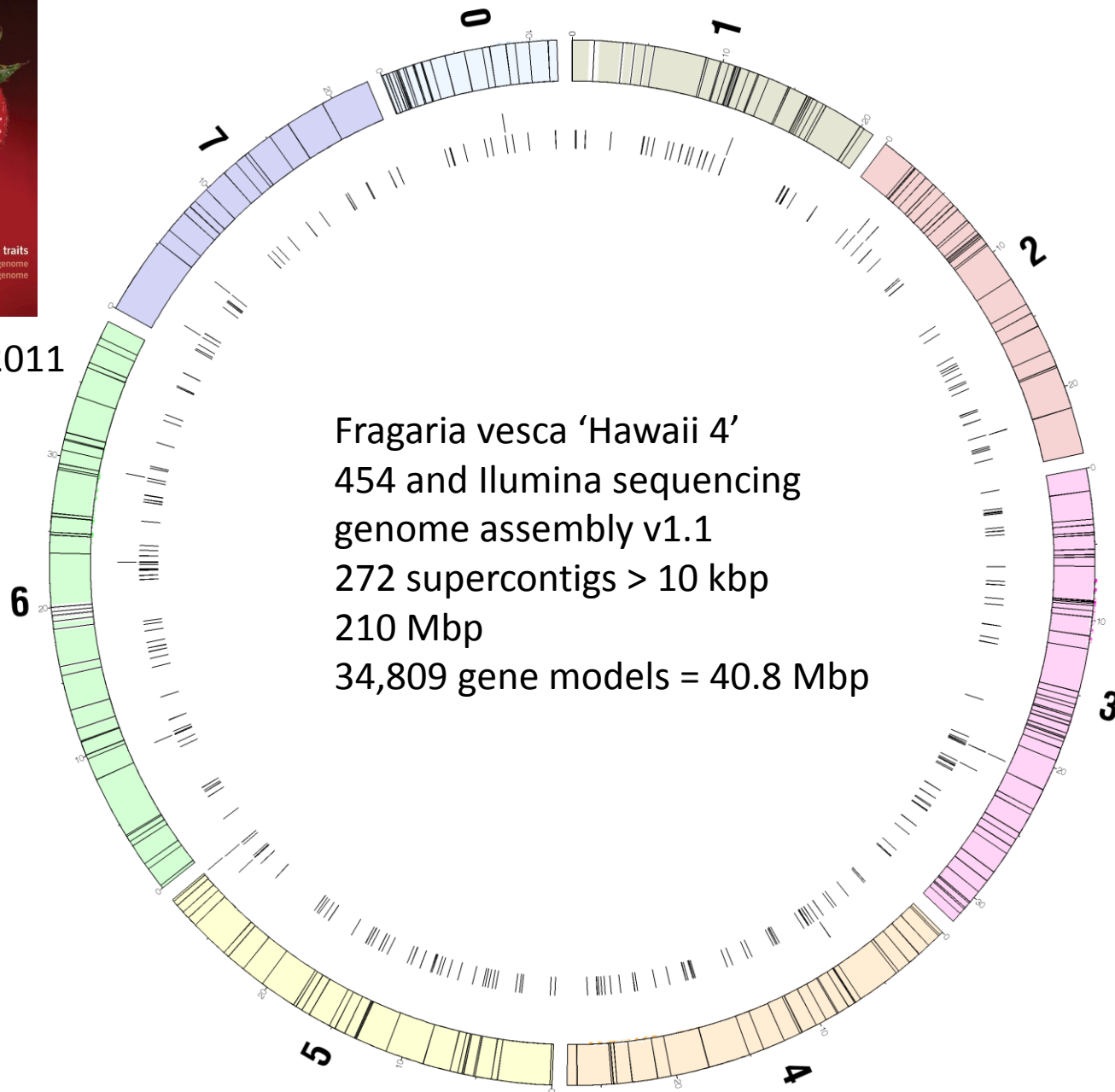
The average locus is 1704 bp in 1-20 exons (mean = 5.3)

80-120 bp baits with 1.5X tiling

Position of the 257 Genes in the Strawberry Genome

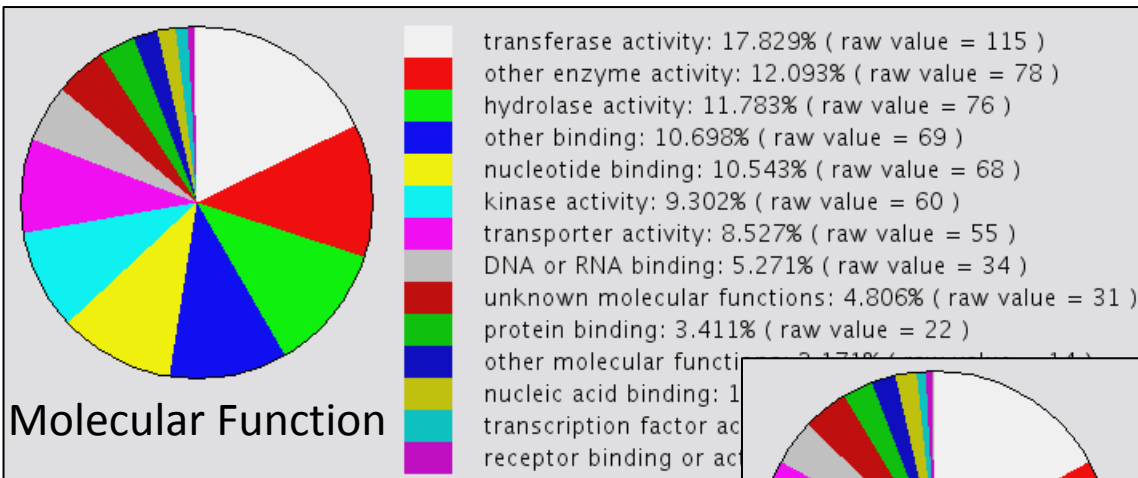


Shulaev et al. 2011

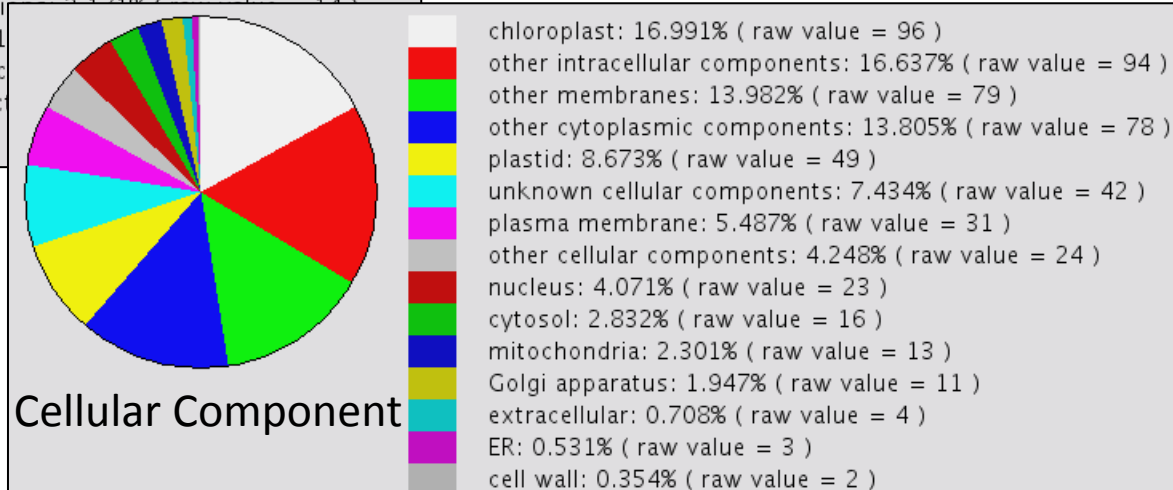


Gene Ontology Annotation For 245 of the 257 genes

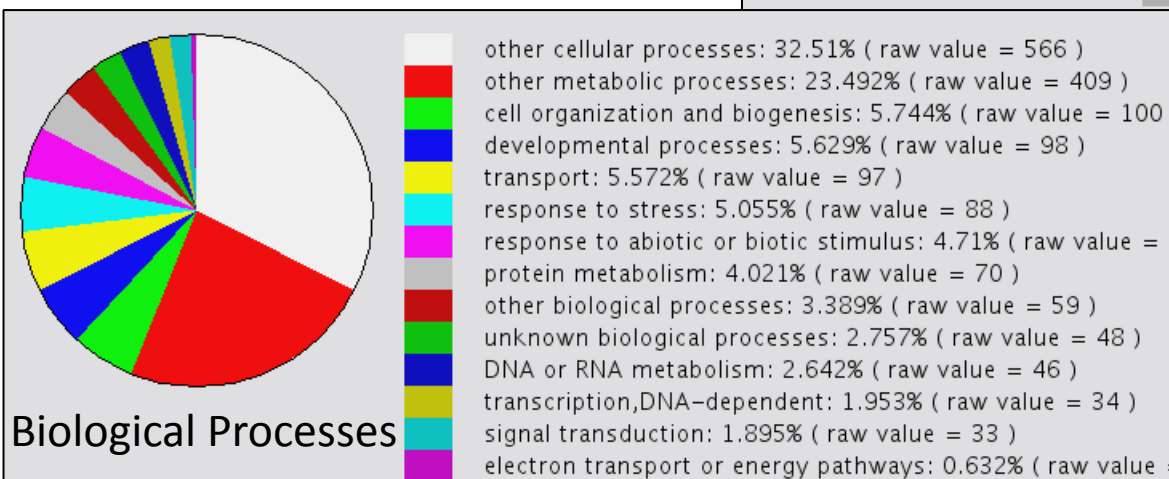
Molecular Function



Cellular Component

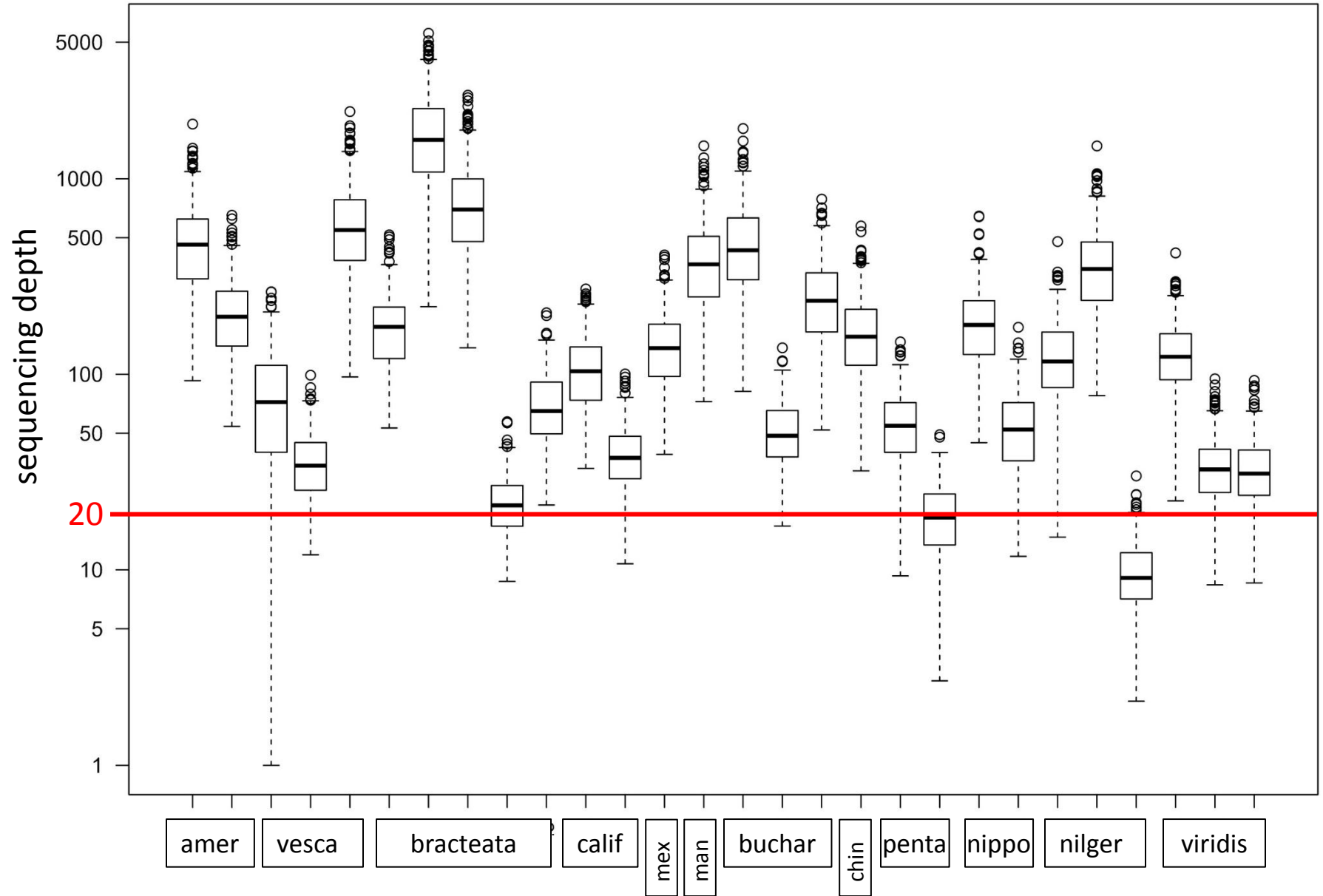


Biological Processes



www.arabidopsis.org/tools/bulk/go/

Sequencing depth at 257 nuclear loci (1419 exons)
in 28 individuals of 12 *Fragaria* species/ subspecies



Fragaria diploids
257 nuclear loci
– concatenated

Mean assembly =
365,048 bp / sample

18.9% missing data

RAxML 7.2.8

GTRCAT rapid bootstrap

GTR- γ ML search

lnL = -1083000.77195

18,023 parsimony
informative sites

one tree

length = 70,659

CI_x = 0.627

Number Loci Supporting Each Node (70% MP bootstrap)

