

Introduction to NGS Sequencing Workshop

Botany 2015

8:30 - 8:40 am - Intro to the workshop and presenters

8:40 - 9:05 am - Intro to Sequencing Technology

9:05 - 9:30 am - Sequence Processing/Intro Bioinformatics

9:30 - 10:00 am - Group Activity

10:00 - 10:15 am - Coffee break

10:15 - 11:00 am - Genome Reduction Methods

11:00 - 11:30 am - Applications and data analysis - phylogenomics

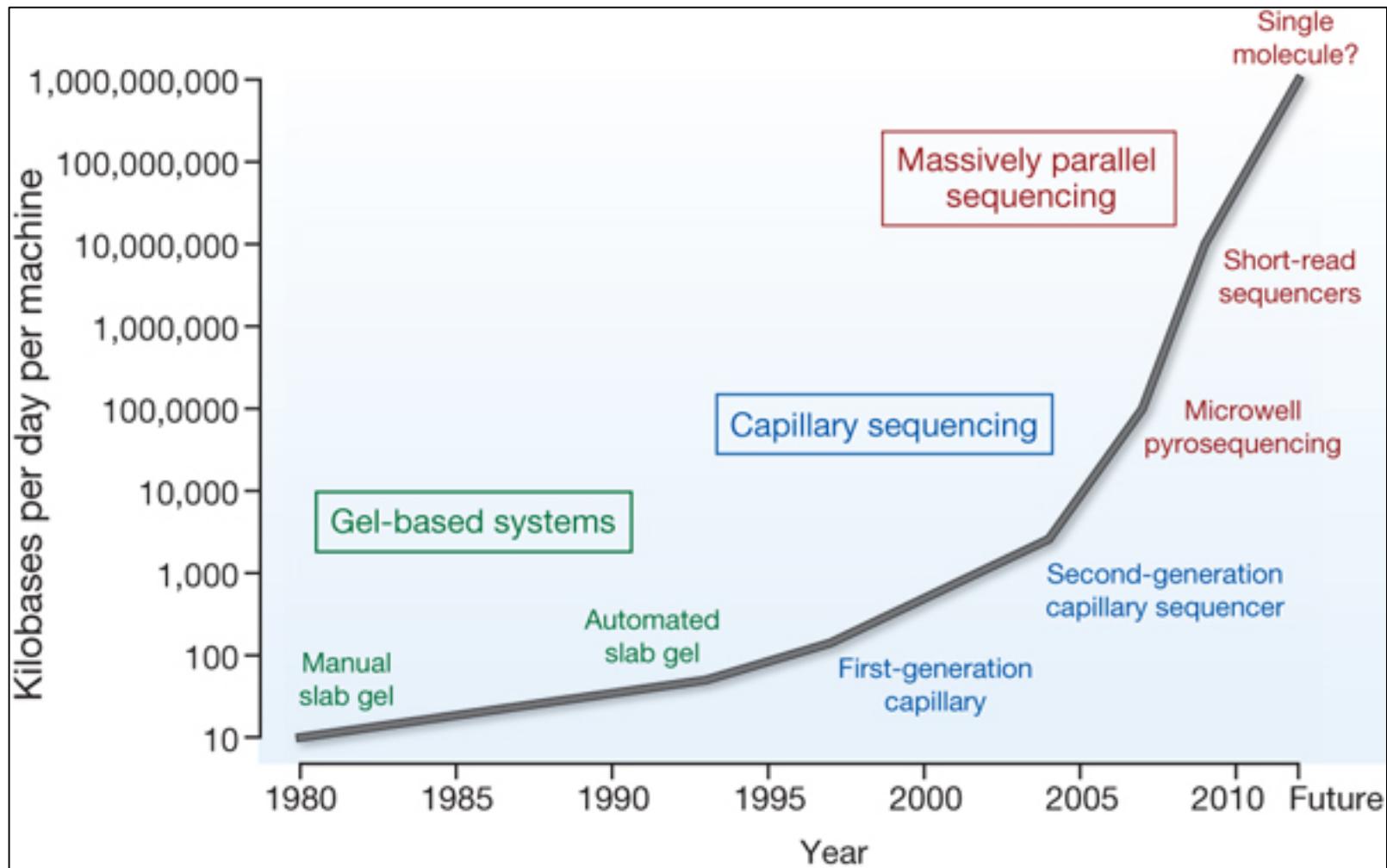
11:30 am - 12:00 pm - Applications and data analysis - population genomics

12:00 - 1:00 pm - Lunch

1:00 - 1:20 pm - Introduction to the hands-on exercises

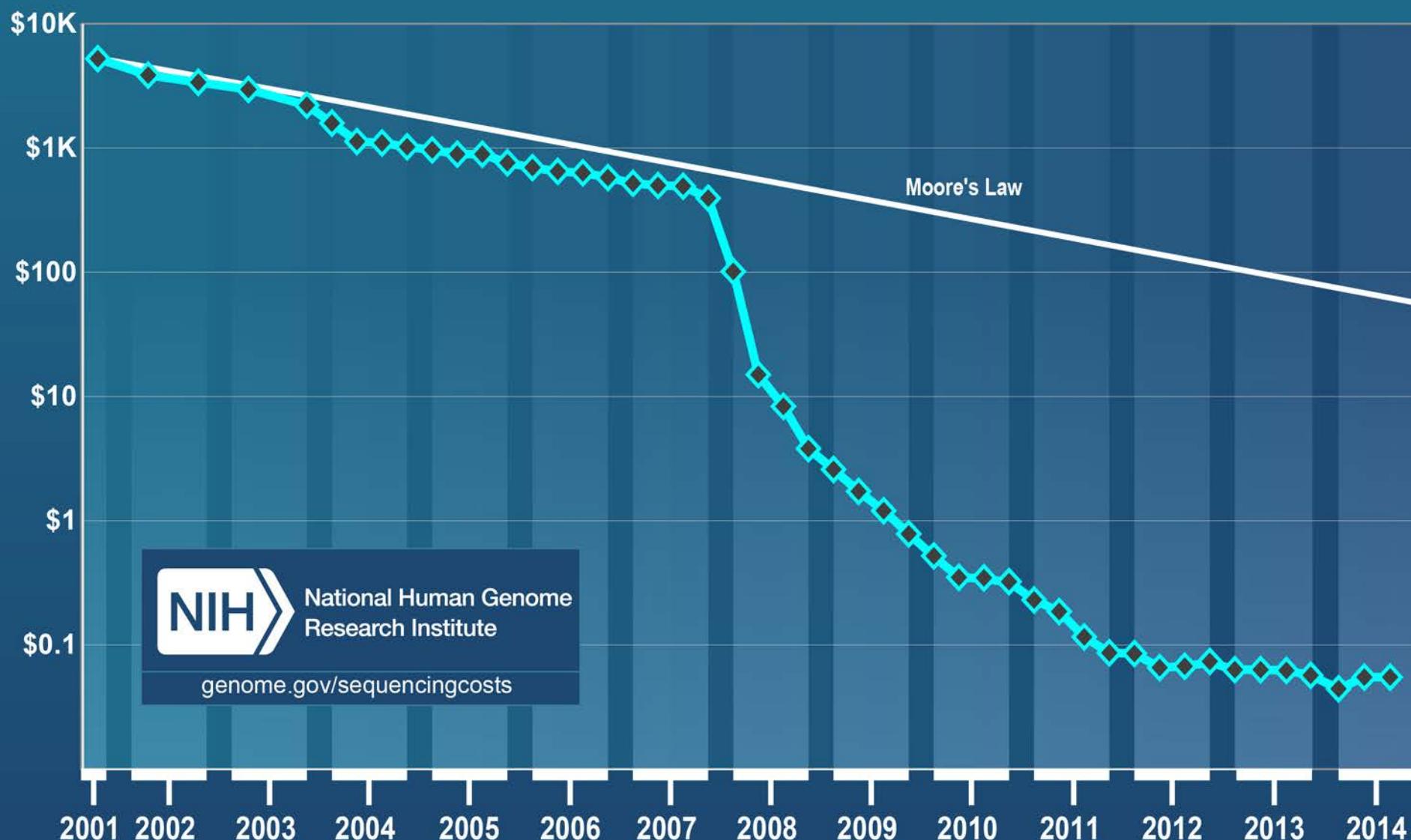
1:20 - 4:30 pm - Hands-on exercises

Improvements in the rate of DNA sequencing over the past 30 years

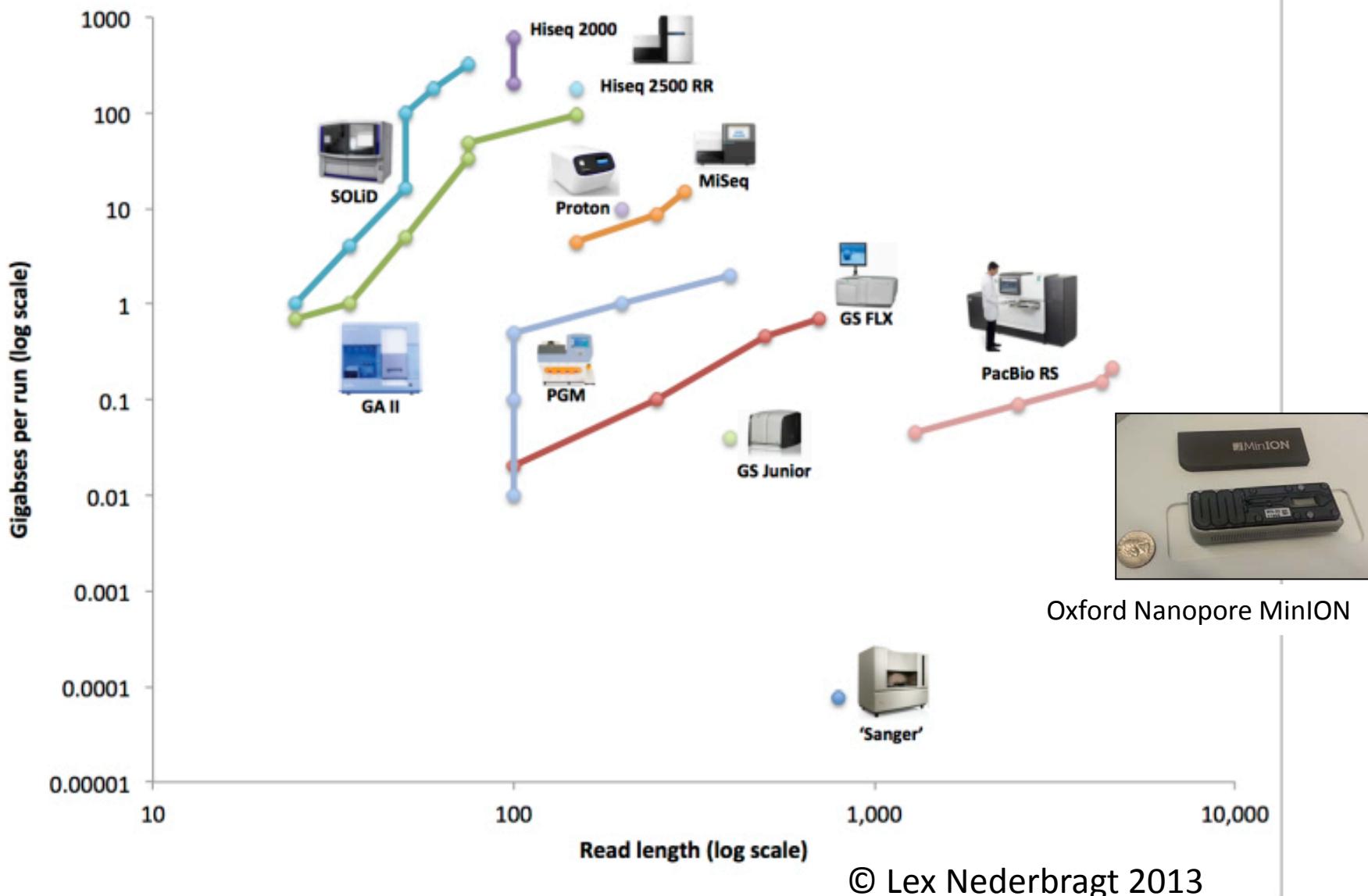


MR Stratton et al. (2009) The cancer genome. Nature **458**, 719-724

Cost per Raw Megabase of DNA Sequence



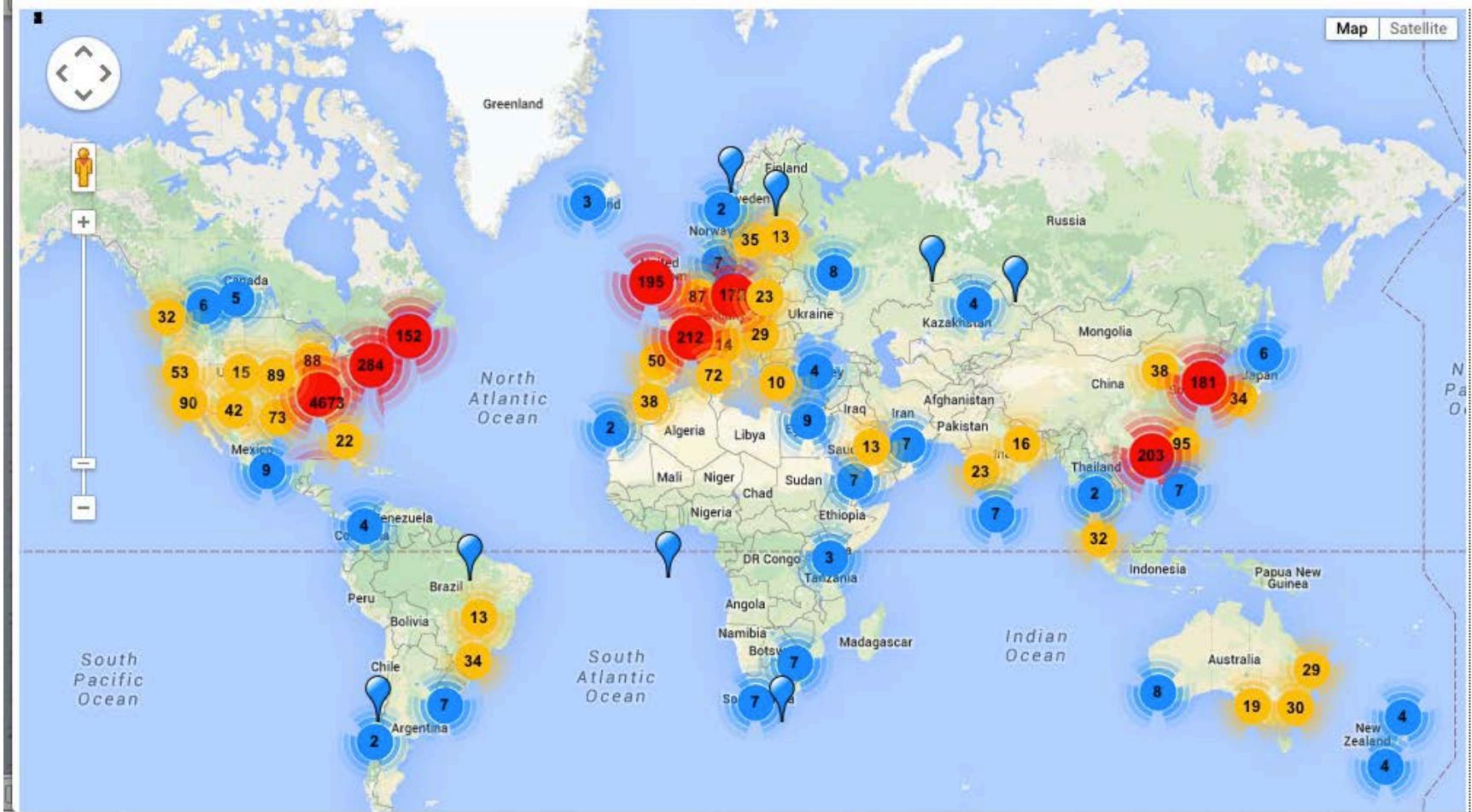
Developments in High Throughput Sequencing



<http://flxlexblog.wordpress.com/2013/10/01/developments-in-next-generation-sequencing-october-2013-edition/>

Next Generation Genomics: World Map of High-throughput Sequencers

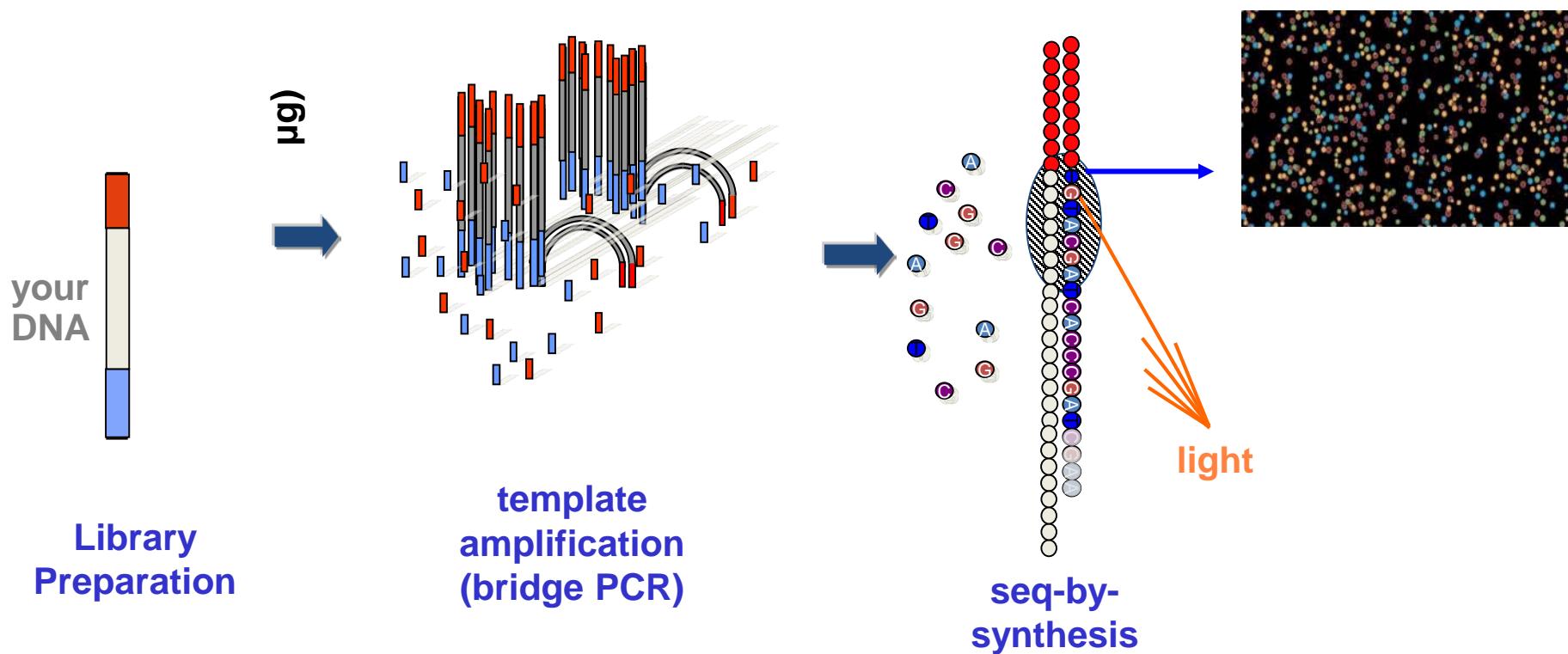
Show all platforms 454 HiSeq HiSeq X Ten Illumina GA2 Ion Torrent MiSeq MinION NextSeq PacBio Polonator Proton SOLiD Service Provider



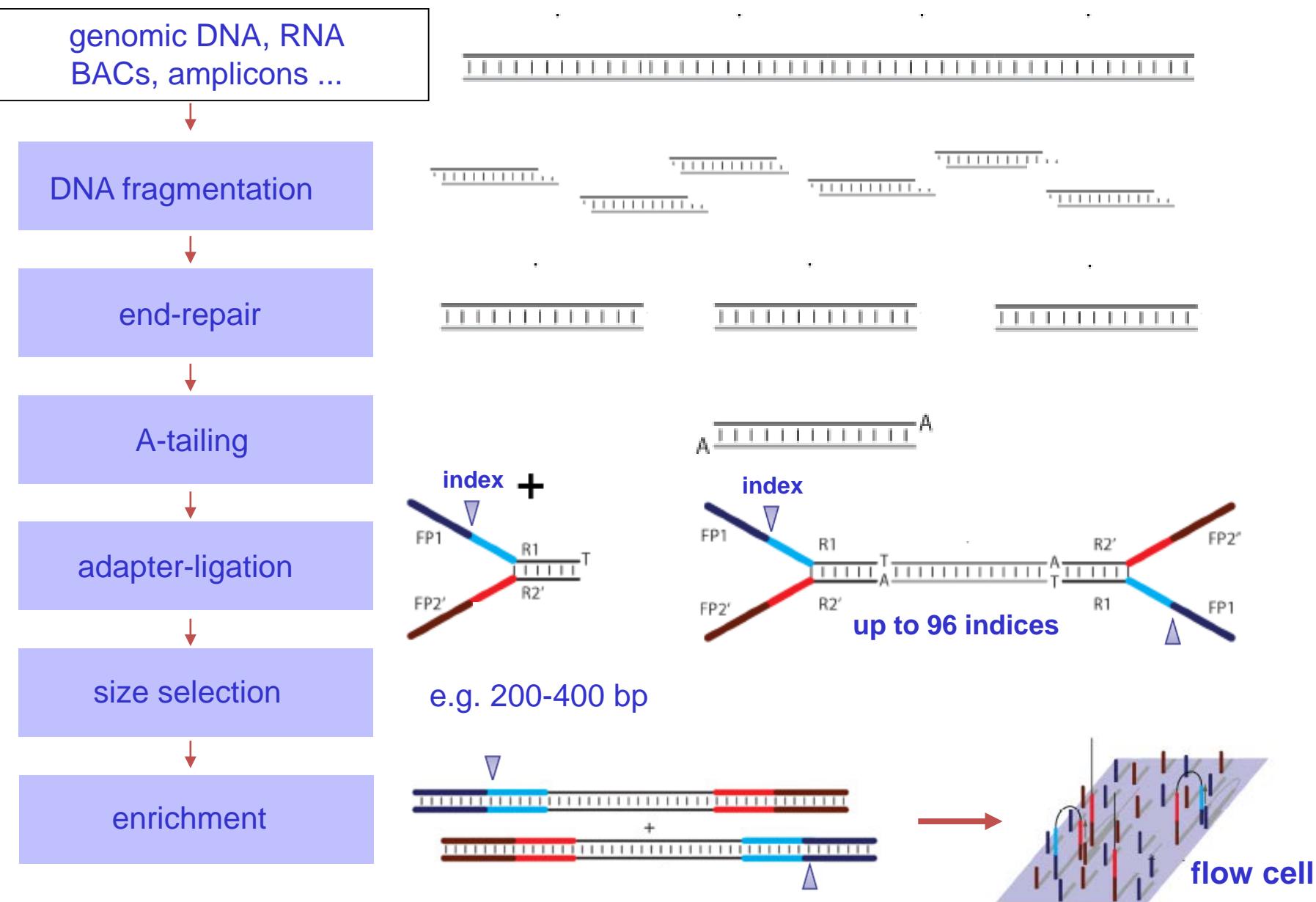
Illumina Sequencing

released in 2007 by Solexa

Template Type	Sequencing Method	Imaging Method
Clonally amplified by solid phase amplification	Sequencing by synthesis with cyclic reversible termination	Four color imaging of single events using fluorescence



“Classic” Illumina Library Prep



Illumina Library Types

Fragment Library



Paired-end Library



Separation of 200-700 bp

Mate-paired Library

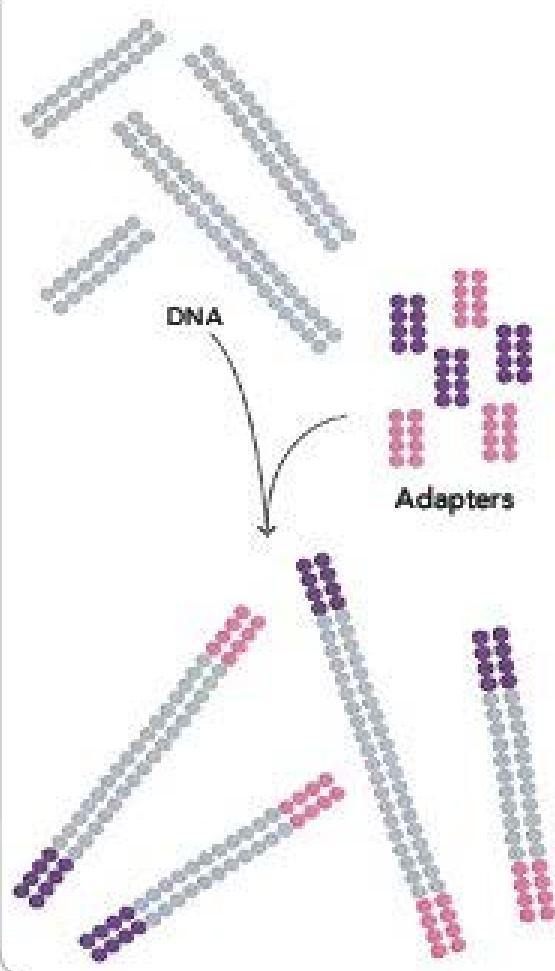


Original separation of 2-20 kb

Illumina Sequencing

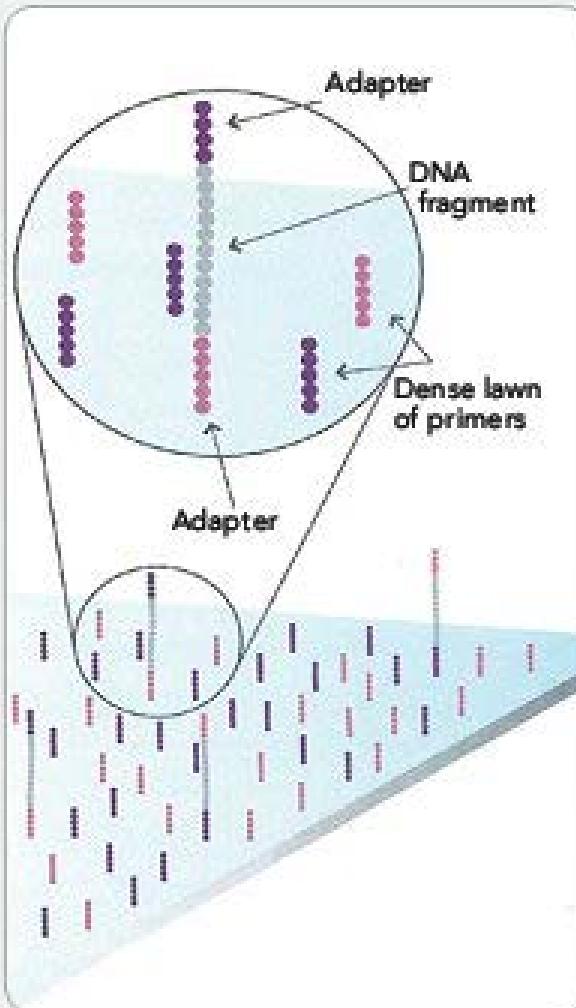
Bridge Amplification

1. PREPARE GENOMIC DNA SAMPLE



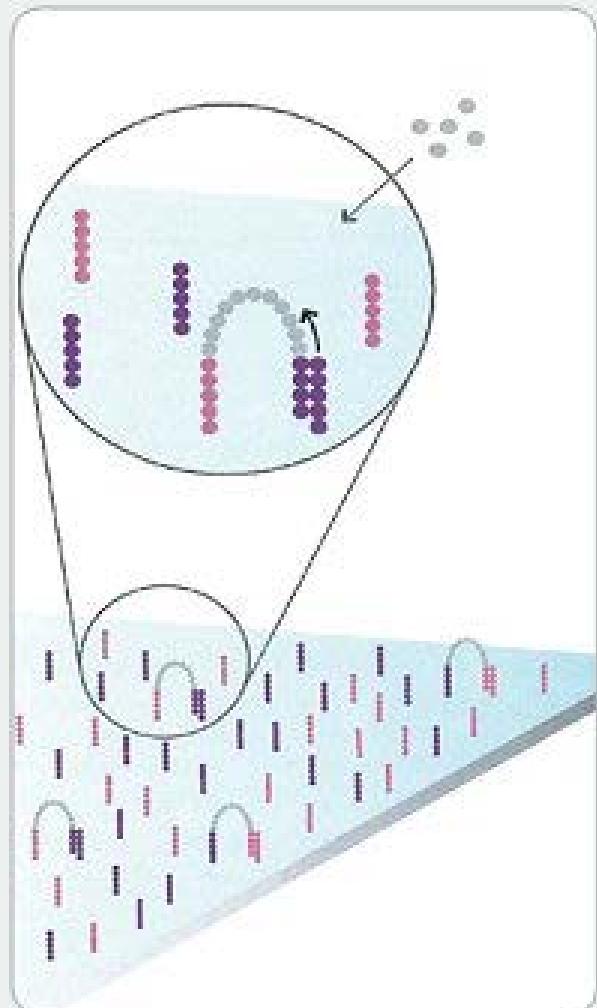
Randomly fragment genomic DNA and ligate adapters to both ends of the fragments.

2. ATTACH DNA TO SURFACE



Bind single-stranded fragments randomly to the inside surface of the flow cell channels.

3. BRIDGE AMPLIFICATION

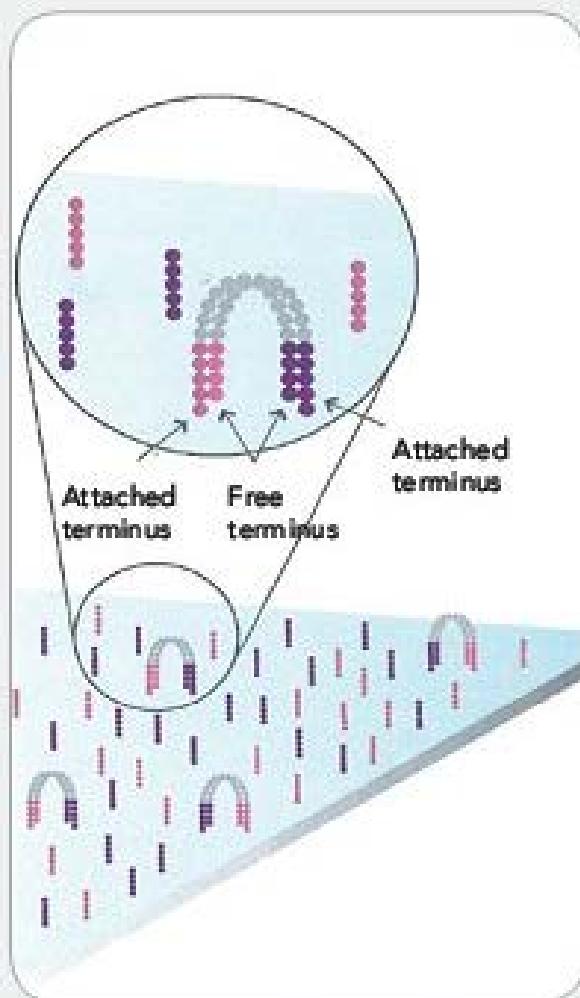


Add unlabeled nucleotides and enzyme to initiate solid-phase bridge amplification.

Illumina Sequencing

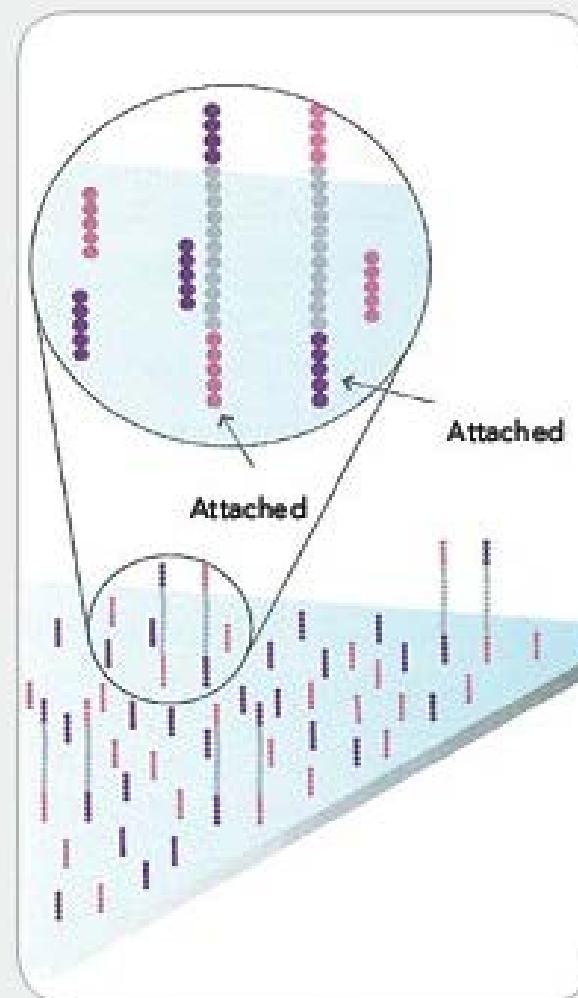
Bridge Amplification

4. FRAGMENTS BECOME DOUBLE STRANDED



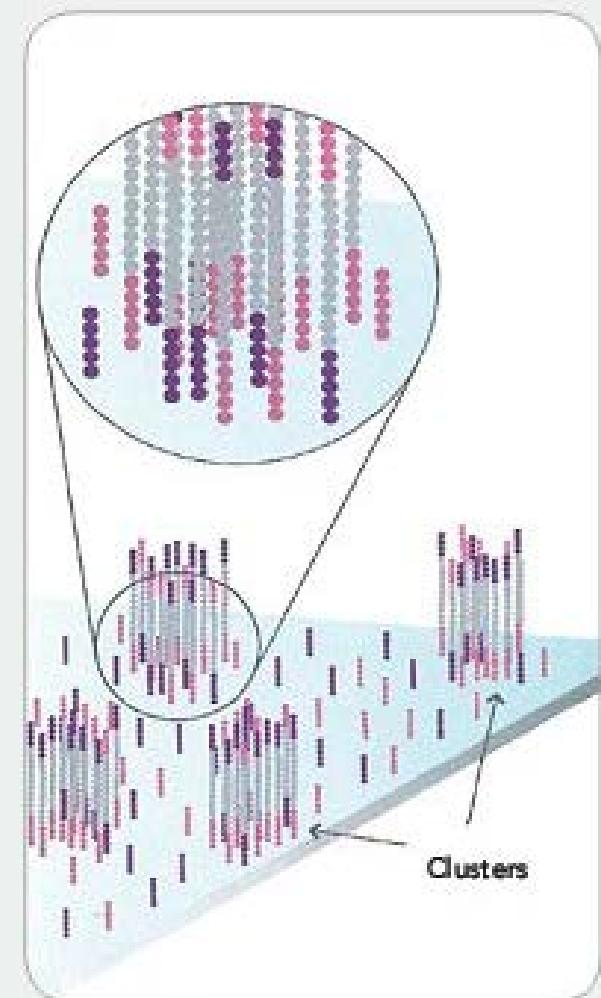
The enzyme incorporates nucleotides to build double-stranded bridges on the solid-phase substrate.

5. DENATURE THE DOUBLE-STRANDED MOLECULES



Denaturation leaves single-stranded templates anchored to the substrate.

6. COMPLETE AMPLIFICATION

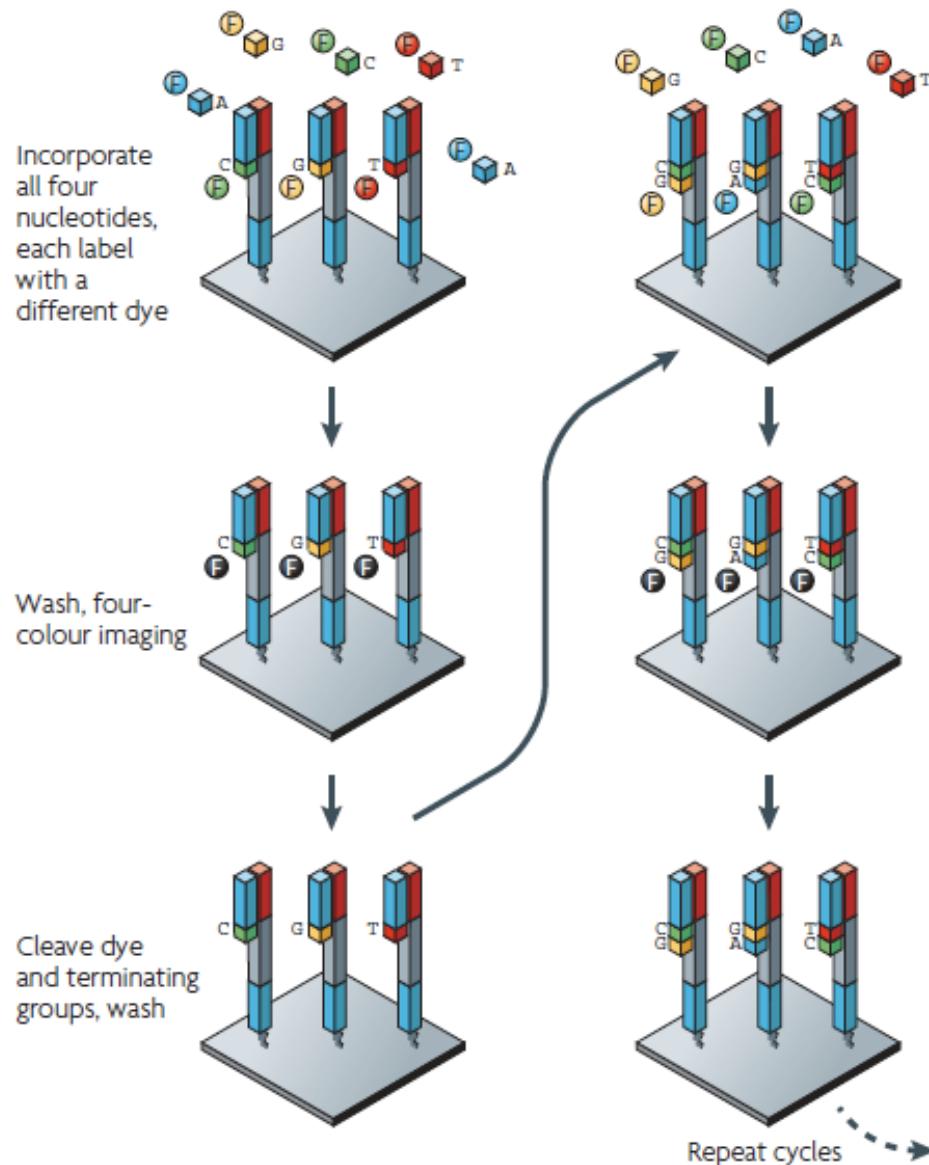


Several million dense clusters of double-stranded DNA are generated in each channel of the flow cell.

Illumina Sequencing

Cyclic reversible termination (CRT)

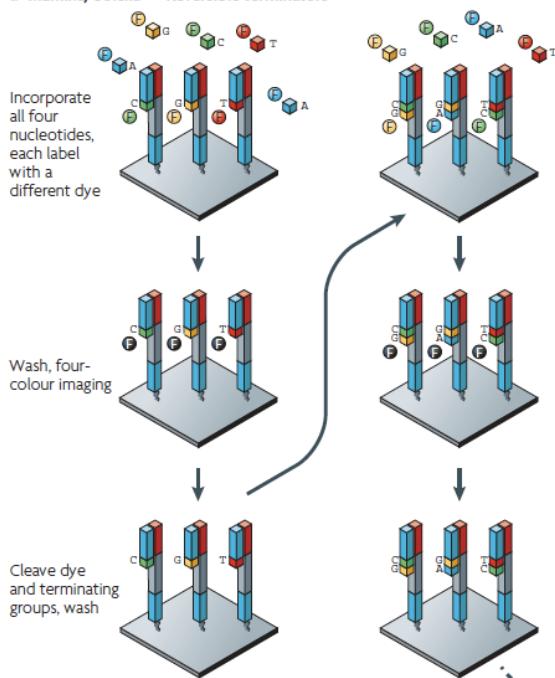
a Illumina/Solexa — Reversible terminators



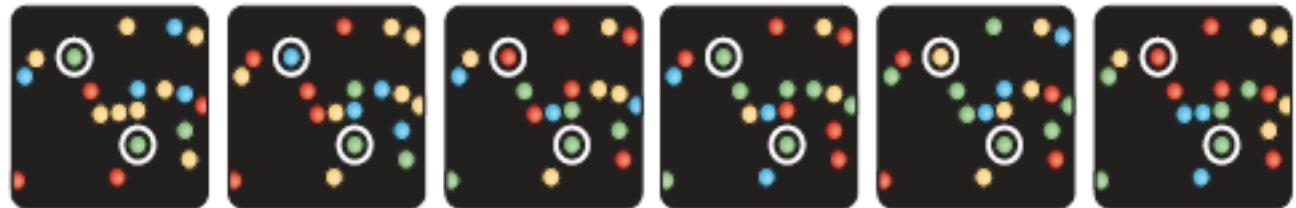
Illumina Sequencing

Cyclic reversible termination (CRT)

a Illumina/Solexa — Reversible terminators



b



Top: CATCGT
Bottom: CCCCCC

Illumina Sequencing

Sequencing power for every scale.

Find the system that's right for your lab or application.

Looking for more detail? Select a single system or two to compare.

Focused power.



[MiSeq Series ◉](#)

Flexible power.



[NextSeq Series ◉](#)

Production power.



[HiSeq Series ◉](#)

Population power.



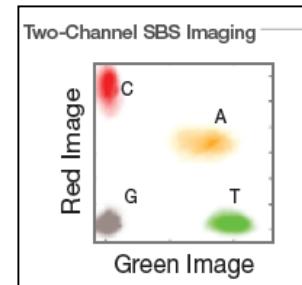
[HiSeq X Series ◉](#)

Key Methods	Small genome, amplicon, and targeted gene panel sequencing.	Everyday genome, exome, transcriptome sequencing, and more.	Production-scale genome, exome, transcriptome sequencing, and more.	Population- and production-scale human whole-genome sequencing.
-------------	---	---	---	---

Illumina Sequencing

Sequencer	Output (Gb)	Run Time	Reads/Flow Cell (M)	Max Read Length
MiSeq	0.5 - 15	4 – 55 hrs.	15 - 25	2 x 300
NextSeq 500	20 - 120	11 – 29 hrs.	130 - 400	2 x 150
HiSeq 2500	10 - 1000	<1 - 6 days	300 - 4000	2 x 150
HiSeq 4000	125 - 1500	<1 – 3.5 days	2500 - 5000	2 x 150
HiSeq X Ten	900 - 1800	< 3 days	3000 - 6000	2 x 150

<http://www.illumina.com/systems/sequencing-platform-comparison.html>



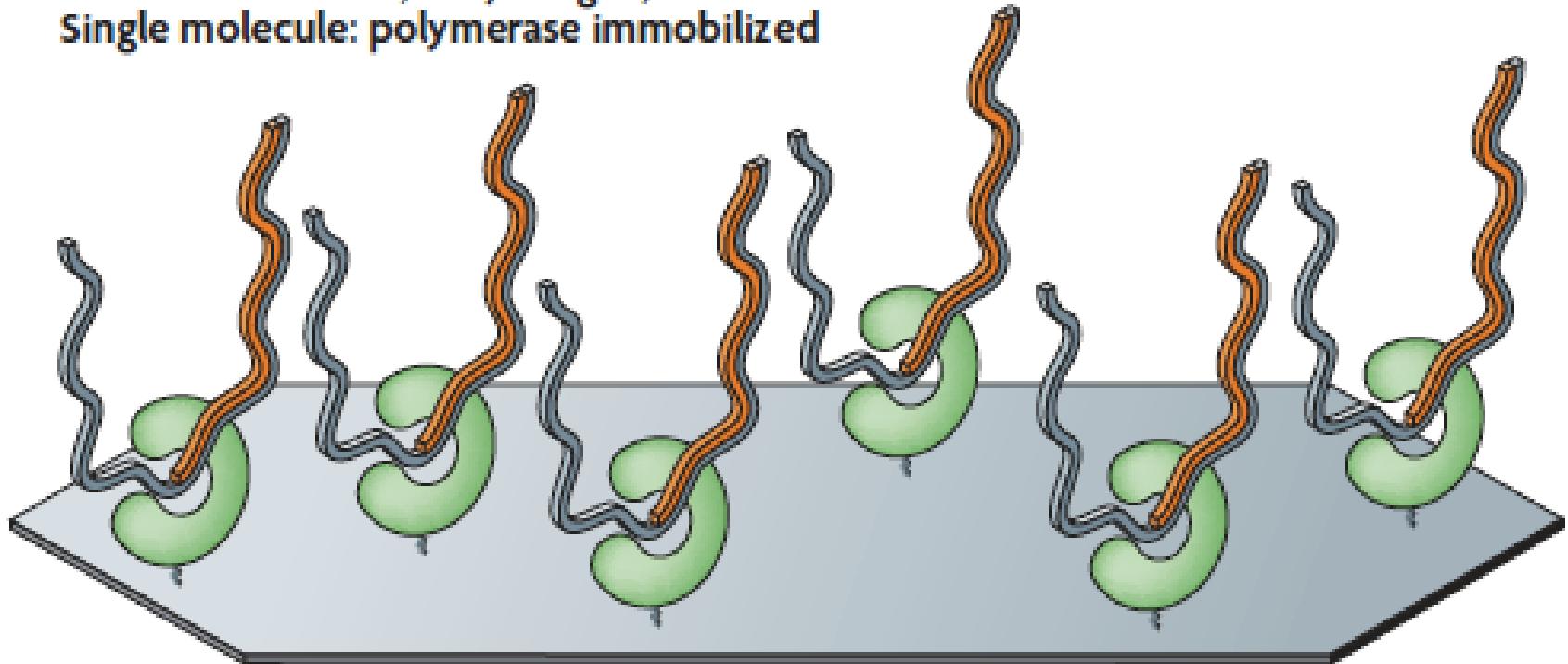
- Real Time Sequencing by Synthesis

SMRT Cells



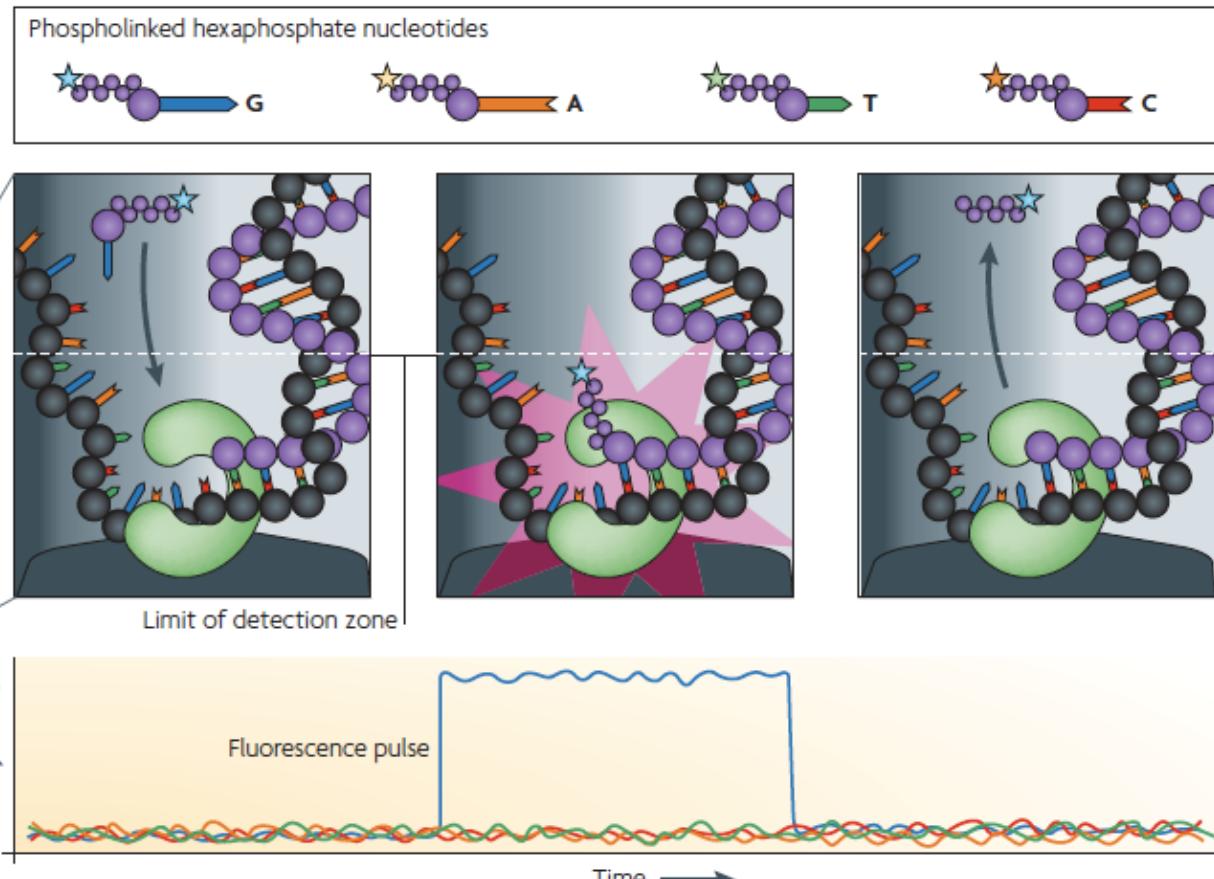
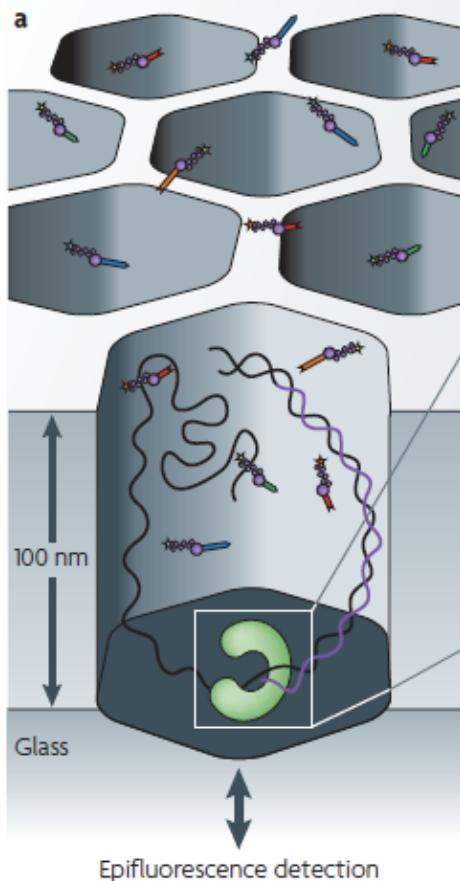
<http://smrt.med.cornell.edu>

- e Pacific Biosciences, Life/Visigen, LI-COR Biosciences
Single molecule: polymerase immobilized



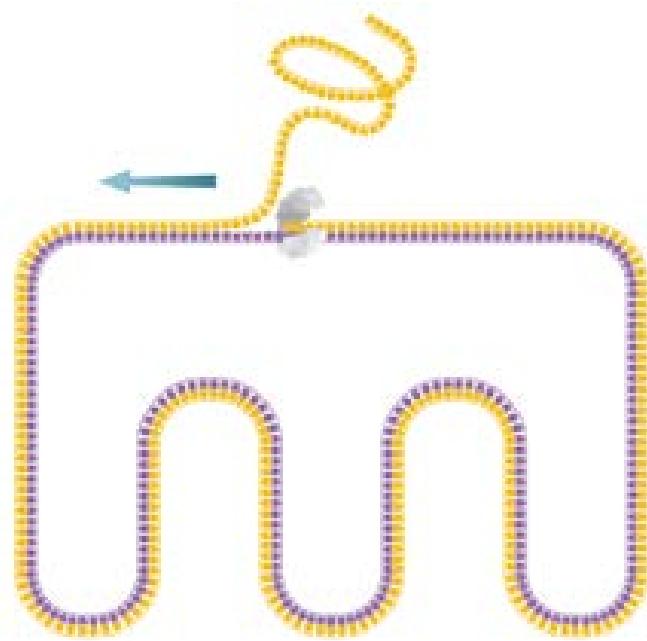
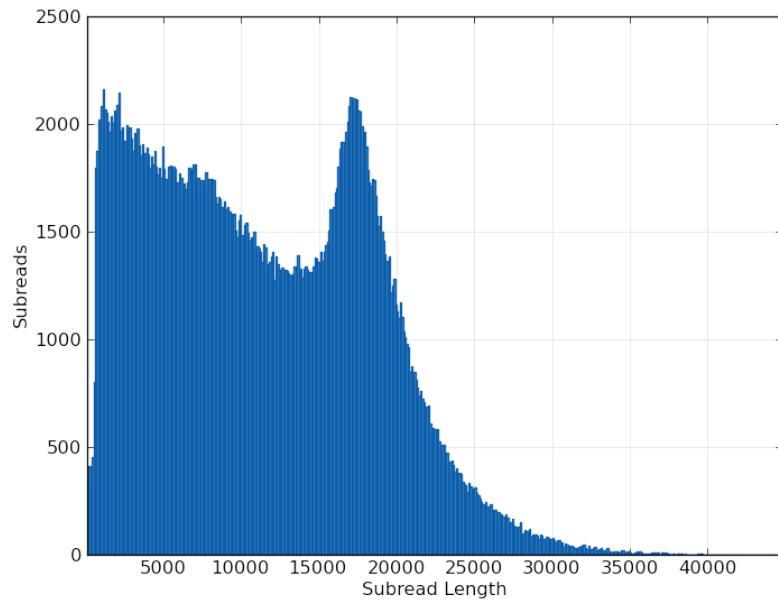
Thousands of primed, single-molecule templates

Pacific Biosciences — Real-time sequencing



Real-Time Sequencing and Re-Sequencing (circular consensus)

Mapped Subread Length Distribution for *C. elegans*



<http://www.pacificbiosciences.com/index.php?q=smrt-dna-sequencing>

<http://blog.pacificbiosciences.com/2014/10/new-chemistry-boosts-average-read.html>

Recent success:

De novo assembly of bacterial and other small genomes

Improving eukaryotic genome assemblies

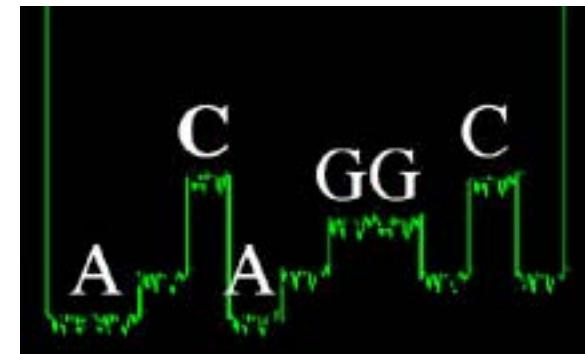
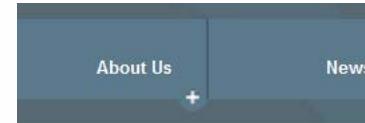
Longer reads, up to ~65,000 bp (avg. = 10-15 kb)

500 Mb – 1 Gb of sequence per run

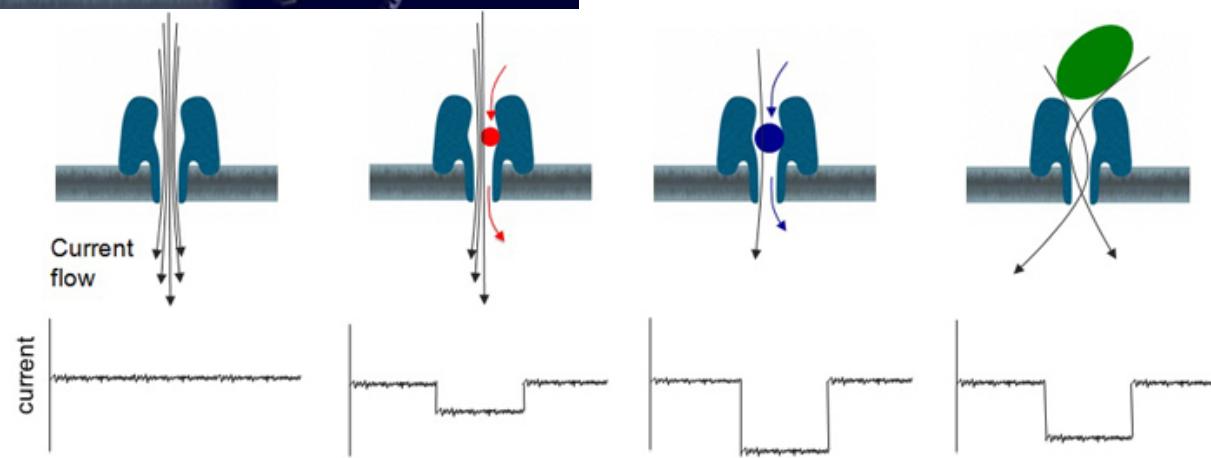
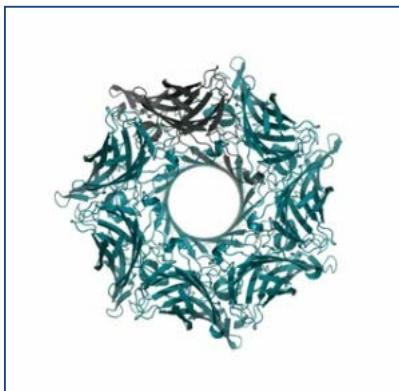
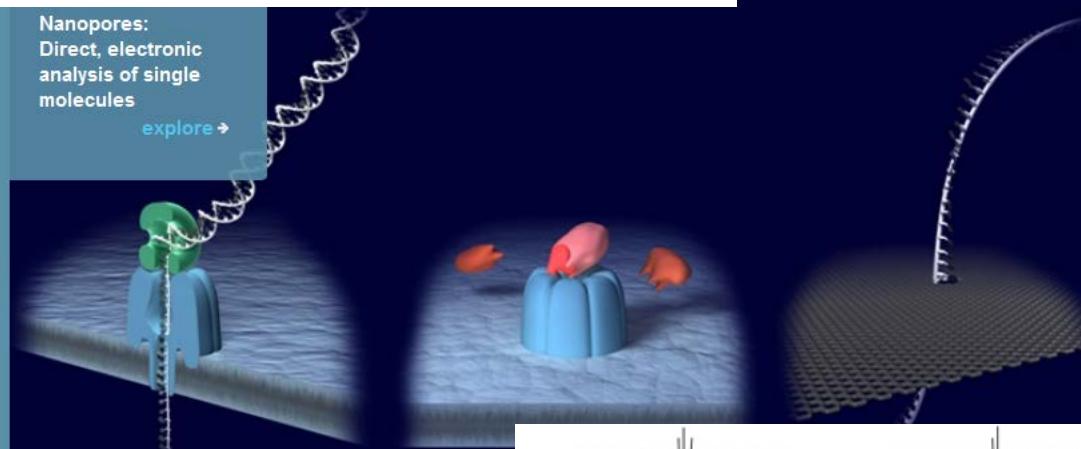
10,000 reads in 2 hours

1st Human Genome (Pendleton et al. Online 6/29/15 – Nature Methods)

Oxford Nanopore

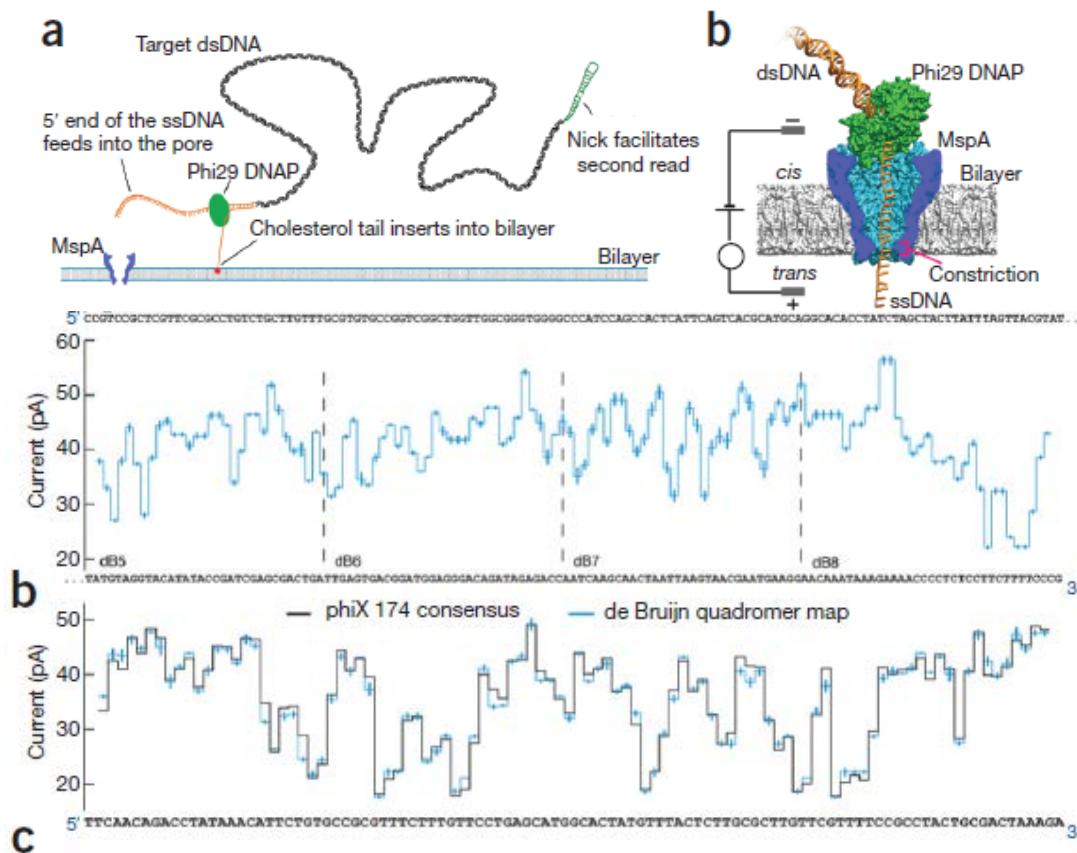


Nanopores



Decoding long nanopore sequencing reads of natural DNA

Andrew H Laszlo¹, Ian M Derrington¹, Brian C Ross¹, Henry Brinkerhoff¹, Andrew Adey², Ian C Nova¹, Jonathan M Craig¹, Kyle W Langford¹, Jenny Mae Samson¹, Riza Daza², Kenji Doering¹, Jay Shendure² & Jens H Gundlach¹



Each of the 256 possible quadromers has a unique current value

Oxford Nanopore

Nov 8th 2013 update: Daniel Turner from ONT offers several clarification. See the end of this post.



+ Follow erlichya

tumblr.

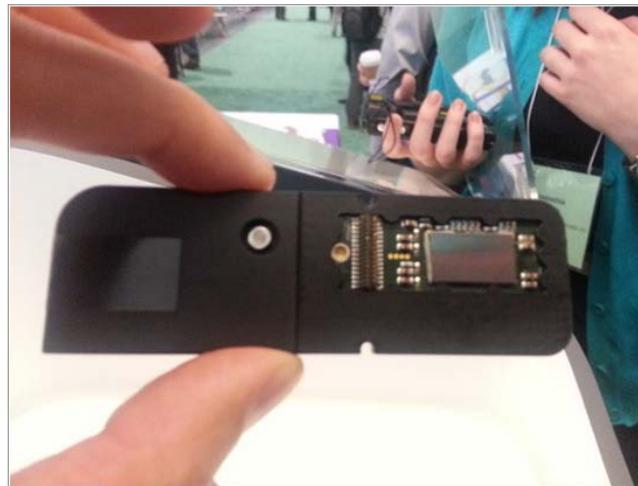
It is a nice milestone for Oxford Nanopore. Almost two years after Clive Brown first described their technology in AGBT2012, they finally set up an [early release plan](#) for the MinIon USB sequencer next month.

MinIon from front:



Avg. read length 2 kb
Read lengths >100 kb possible

MinIon from back (outside the small box):



On the way:
PromethION
GridION

The minion sequencer is going to cost about \$1000 and will last for about 6 hours of sequencing (this number might change based on the experience of early adaptors). According to OxNano, the reason for this limitation is erosion of the electrodes. The sequencing program will notify the user when the sequencer has exhausted its lifespan.

[One interesting feature they aim for is dividing the sequencer lifespan across several runs. For example, you would sequence](#)

© Yaniv Erlich

NGS Error Rates

2014 NGS Field Guide. www.molecularecologist.com

Platform	Primary Errors	Single-pass Error Rate (%)	Final Error Rate (%)	Notes
ABI 3730xl (capillary)	Substitution	0.1-1	0.1-1	
Illumina	Substitution	~0.1	~0.1	≥ 85% of reads
PacBio RS	Indel	~13	≤1	consensus of 3 reads
Oxford Nanopore	Deletion	4 (38)	4 (38)	press release vs. early pubs (e.g., Laver et al. 2015)

How to Deal with Overcapacity

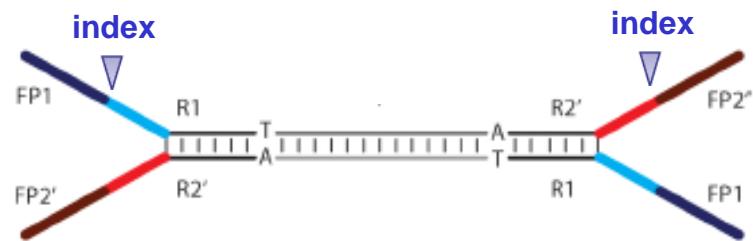
Multiplexing

Addition of a unique sequence identifier (barcode or index) allowing multiple samples to be run together on a single flow cell lane.

Internal indexes (Cronn et al 2008)

External indexes (extra round of sequencing)

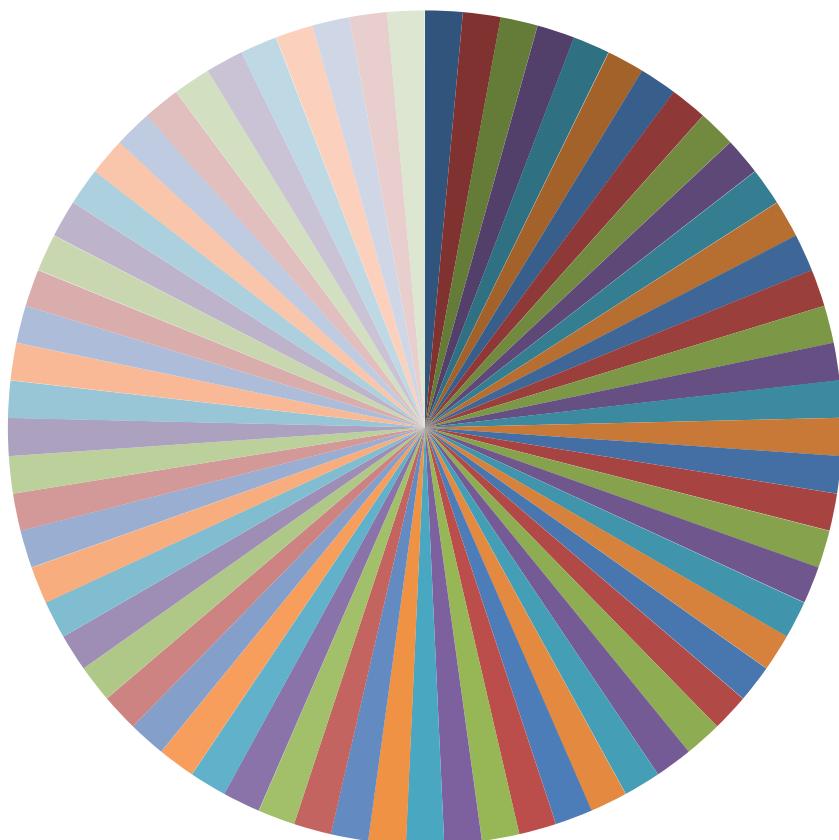
- Illumina 24 single; 96 dual
- NEB 24 single; 96 dual
- Nextflex 384 single
- NuGen 384 dual
- Swift 24 single; 96 dual



Multiplexing Example

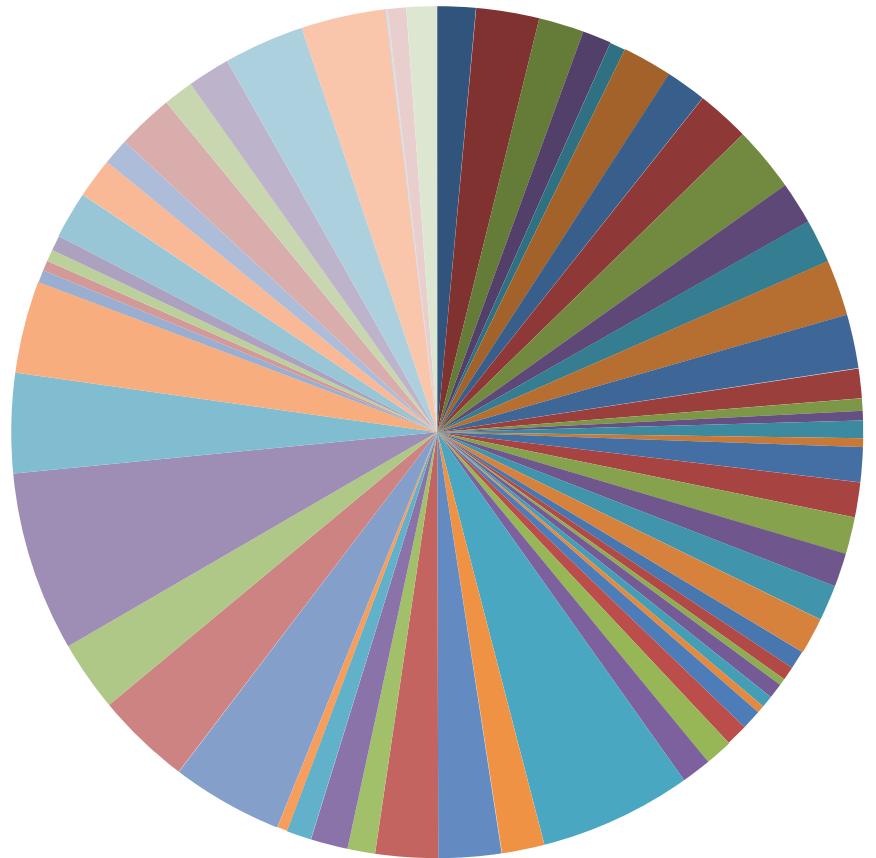
MiSeq v.3 1 Nov 2013
20.3 million paired end reads
69 *Fragaria virginiana* BACs

Ideal Equimolar Results



average = 295,000

Actual Results



1200-13.8 million reads
median = 236,000

Outsourcing Library Prep

1. Fairly standardized (although diverse library options exist)
2. Expensive equipment (sonication, quantification)
3. Automation leading to lower prices



How much will it cost?

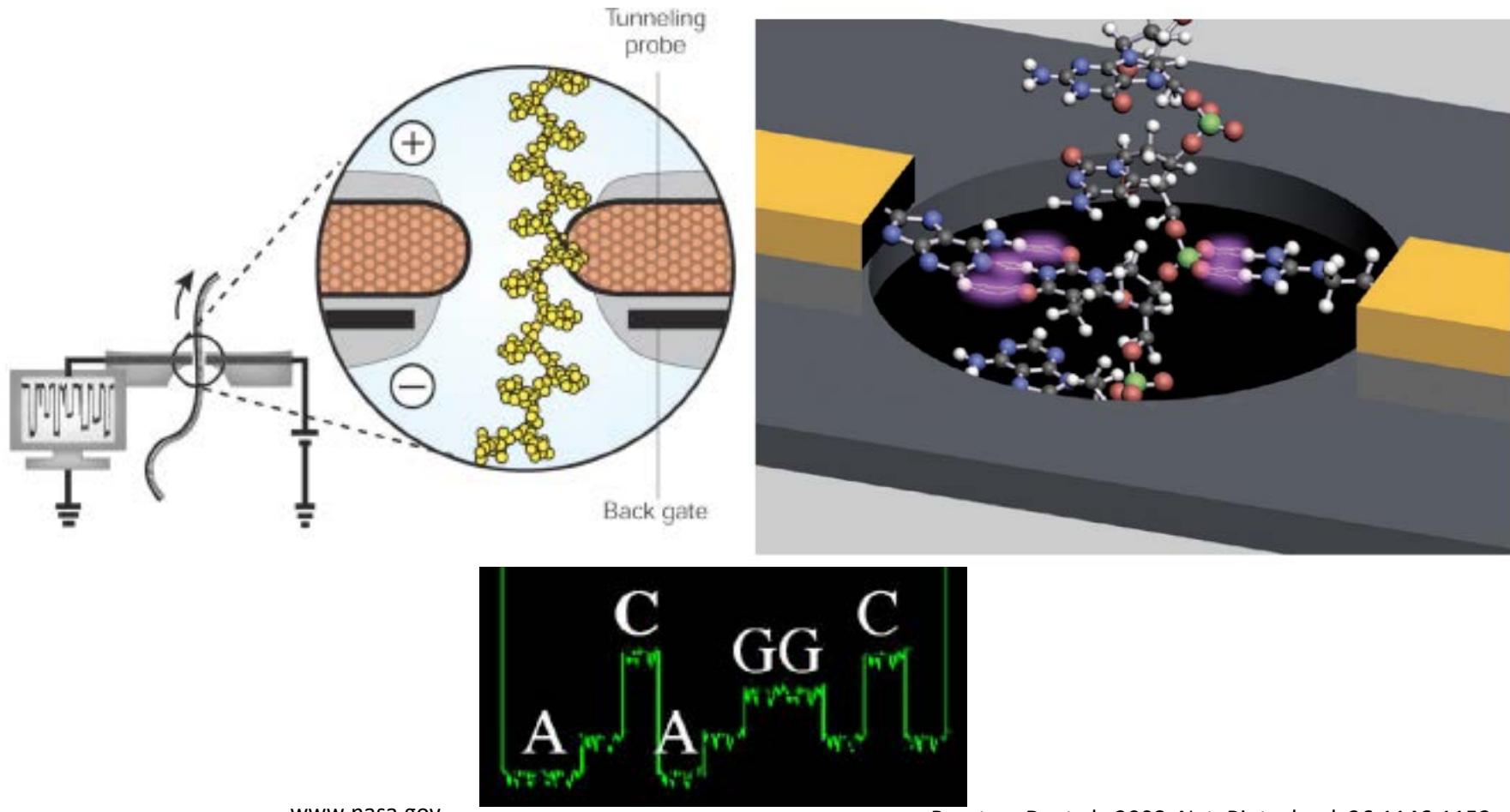
2014 NGS Field Guide. www.molecularecologist.com

Instrument	Cost Per GBp
ABI 3730xl (capillary)	\$2,307,692
PacBio RS II	\$1,111
Illumina MiSeq	\$109
Illumina HiSeq 2500	\$30
Illumina HiSeq X 10	\$7

Includes sequencing reagent costs, but excludes library preparation costs.

What's Next?

Solid State Nanopore Sequencing



Introduction to NGS Sequencing Workshop

Botany 2015

8:30 - 8:40 am - Intro to the workshop and presenters

8:40 - 9:05 am - Intro to Sequencing Technology

9:05 - 9:30 am - Sequence Processing/Intro Bioinformatics

9:30 - 10:00 am - Group Activity

10:00 - 10:15 am - Coffee break

10:15 - 11:00 am - Genome Reduction Methods

11:00 - 11:30 am - Applications and data analysis - phylogenomics

11:30 am - 12:00 pm - Applications and data analysis - population genomics

12:00 - 1:00 pm - Lunch

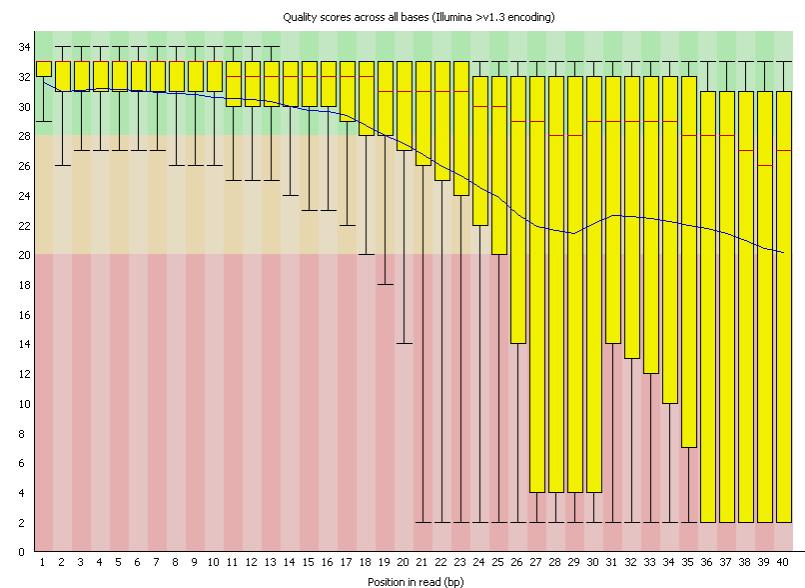
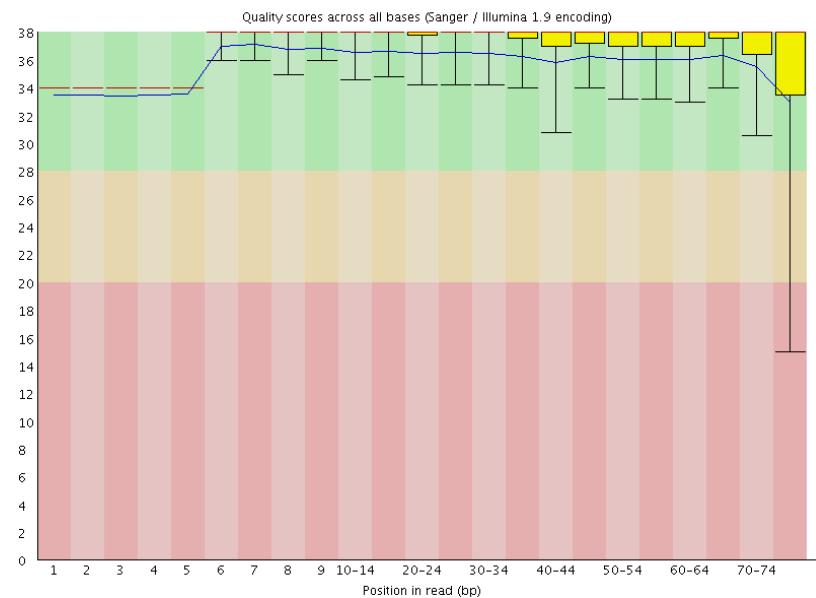
1:00 - 1:20 pm - Introduction to the hands-on exercises

1:20 - 4:30 pm - Hands-on exercises

Sequence Processing

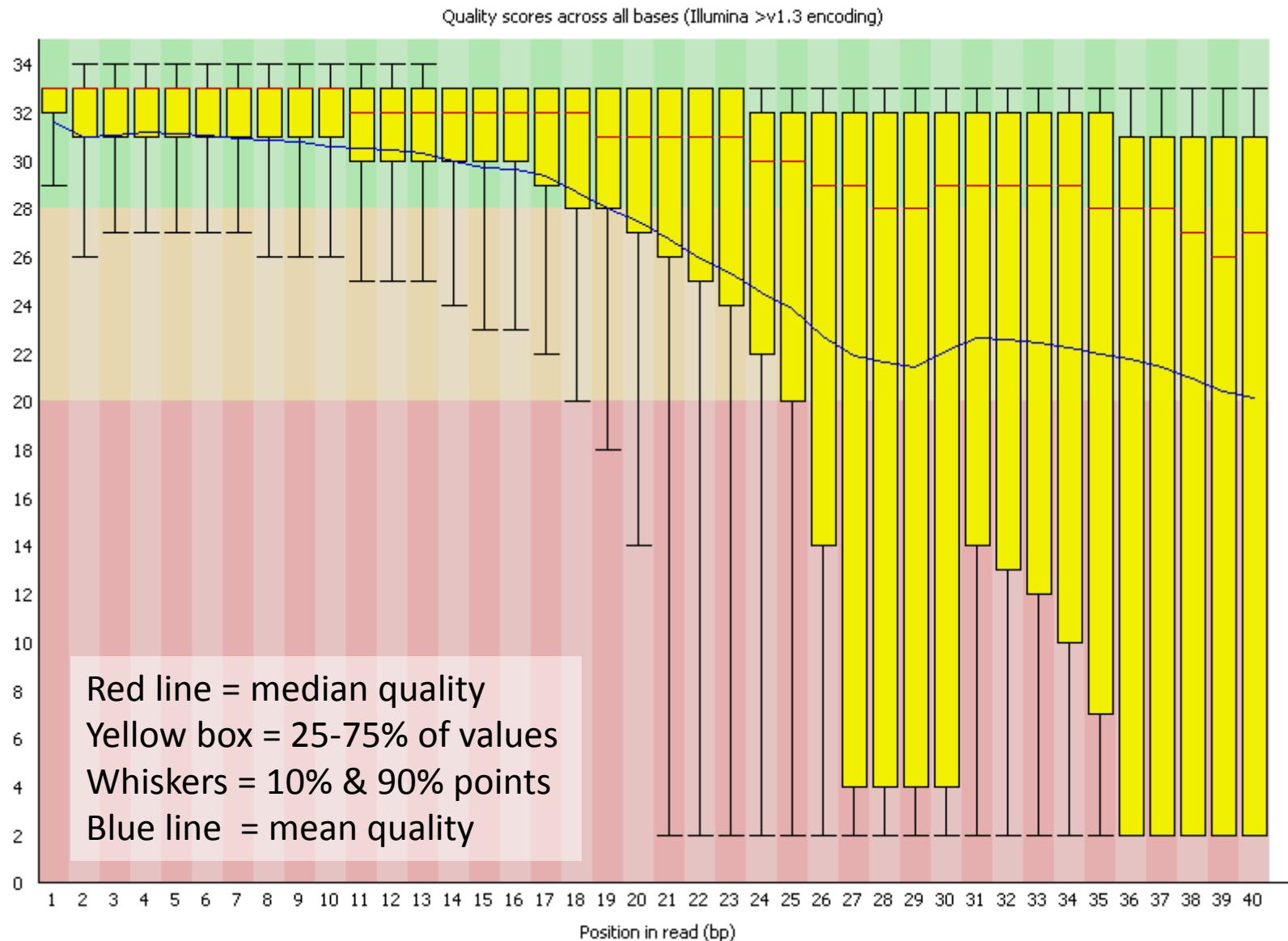
Potential Contaminants and Biases

1. Low Quality Bases
2. Adapters
3. Duplicate Reads
4. Uneven Sequencing Depth
5. Fragment Size Variation
6. Read Overlap
7. Biological Contamination



Sequence Processing

FASTQC output



Sequence Processing

Quality Assessment Tools:

Illumina BaseSpace <https://basespace.illumina.com/runs>

FastQC

<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

SGA Simpson, 2014. Bioinformatics
uses kmers to estimate:
paired end insert sizes
heterozygosity
repeat content
genome size

Sequence Processing

A Popular Read Trimming & Adapter Removal Tool:

Trimmomatic (Bolger et al 2014)

<http://www.usadellab.org/cms/?page=trimmomatic>

- trims bases at start & end of reads, using a threshold quality score
- also can use a sliding window average of base quality to trim
- filters reads below a minimum length after trimming
- keeps track of orphaned pairs
- short adapters can be trimmed in palindrome mode

Phred = 20 (1% error rate) is commonly used

But is this too stringent? (MacManes 2014)

Sequence Processing

Duplicate Removal

Optical duplicates = split clusters

PCR duplicates from final “enrichment” step of library prep

Best to limit to absolute minimum of cycles (6-12)

Deep sequencing (>500X) will result in “natural duplication”

Should these be removed? Differences of opinion exist.

See Zhou et al. Bioinformatics January 2, 2014.

Yes for SNPs; No for counting [but incorporate into model]

Tools for Duplicate Read Removal:

Picard (picard.sourceforge.net; requires mapping)

FastUniq (Xu et al 2012; requires paired sequences)

fastq_collapse.py (Weitemier, unpublished)

Sequence Processing

k -mer = substrings of length k in DNA sequence data

ACGTCA~~G~~

acg tca

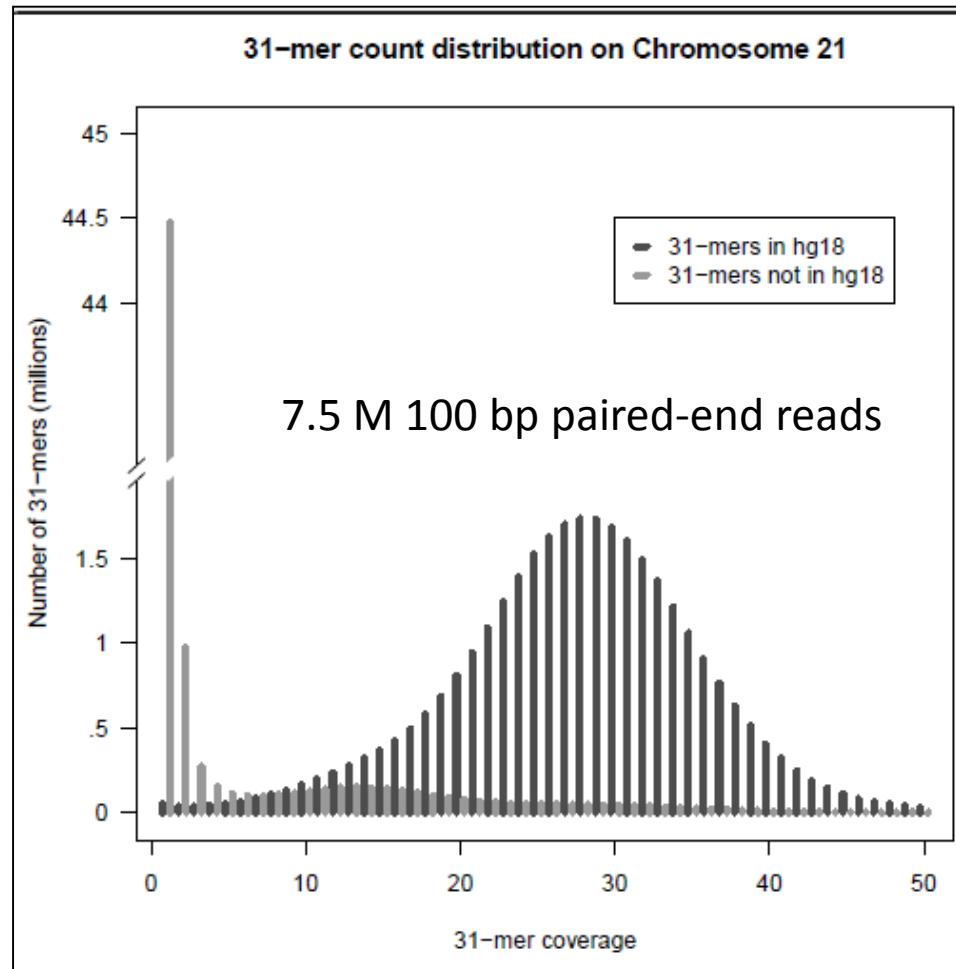
cgt cac five 3-mers

gtc acg

Counting the occurrences of all such substrings in a sequence generates a k -mer frequency distribution:

(reverse complements counted as a single k -mer)

Melsted & Pritchard (2011) Efficient counting of k -mers in DNA sequences using a bloom filter. BMC Bioinformatics 12:333.



Gray = errors and SNPs

Sequence Processing

Digital Normalization

Brown et al (2014)

1 error => up to k erroneous k -mers

AAAAAAGAAAAAA

5-mer example

Table 1. Digital normalization to C=20 removes many erroneous k-mers from sequencing data sets. Numbers in parentheses indicate number of true k-mers lost at each step, based on reference.

Data set	True 20-mers	20-mers in reads	20-mers at C=20	% reads kept
Simulated genome	399,981	8,162,813	3,052,007 (-2)	19%
Simulated mRNASeq	48,100	2,466,638 (-88)	1,087,916 (-9)	4.1%
<i>E. coli</i> genome	4,542,150	175,627,381 (-152)	90,844,428 (-5)	11%
Yeast mRNASeq	10,631,882	224,847,659 (-683)	10,625,416 (-6,469)	9.3%
Mouse mRNASeq	43,830,642	709,662,624 (-23,196)	43,820,319 (-13,400)	26.4%

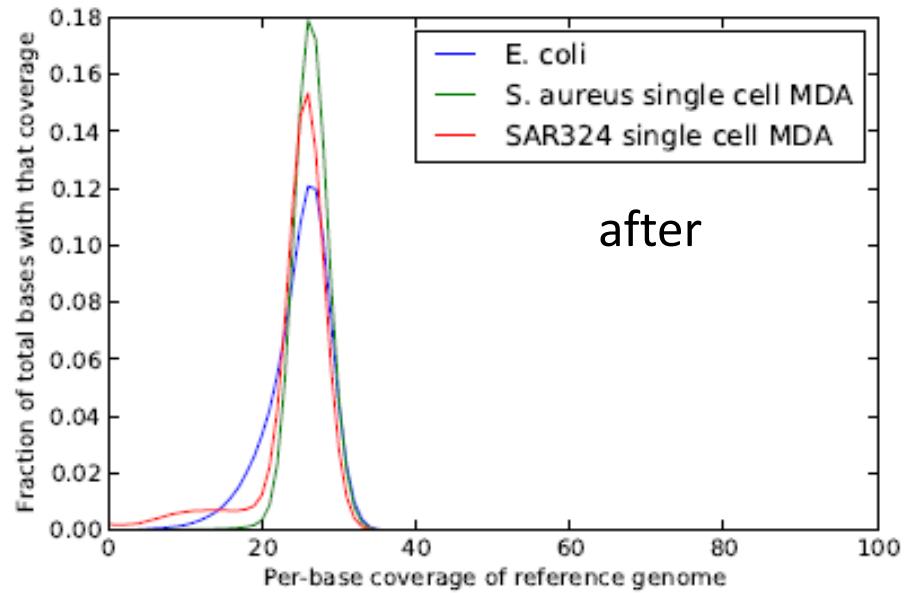
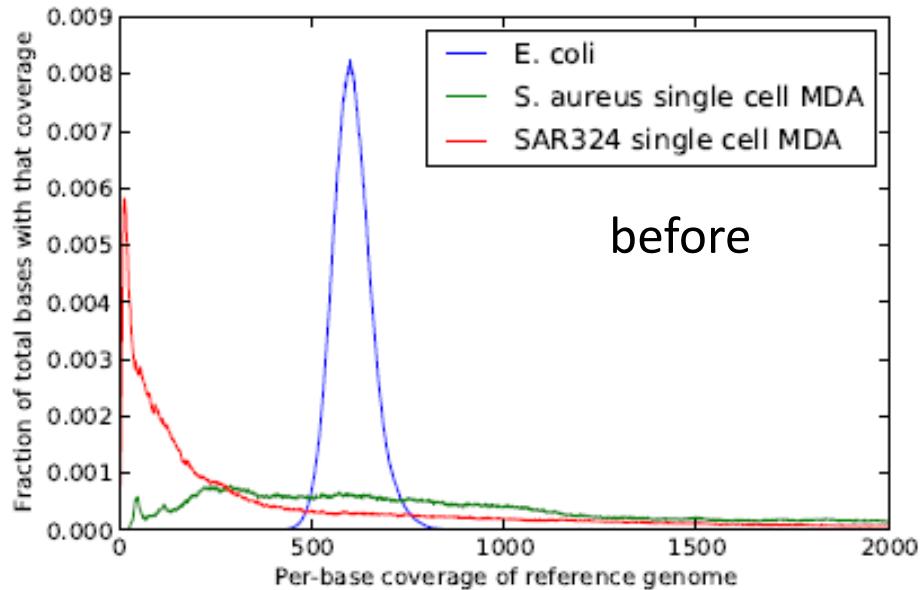
Sequence Processing

Digital Normalization

Brown et al (2014)

Retains nearly all real k -mers while discarding the majority of erroneous k -mers.

Reduces the number of over-represented k -mers (duplicates, repeats)



Sequence Processing

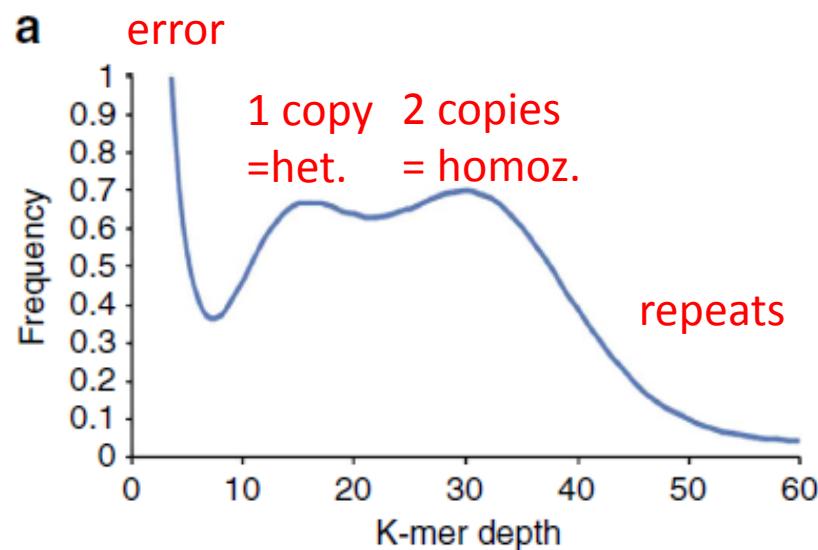
Uneven Sequencing Depth:

Heterozygosity

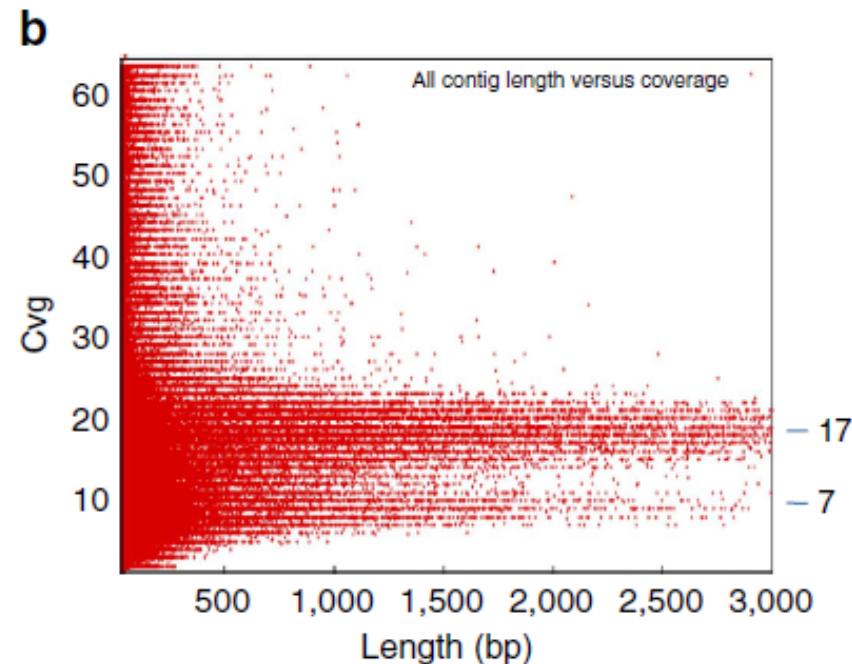
Repeated Nuclear Sequences

Organelles

Biological Contamination



Whole Genome Illumina Reads



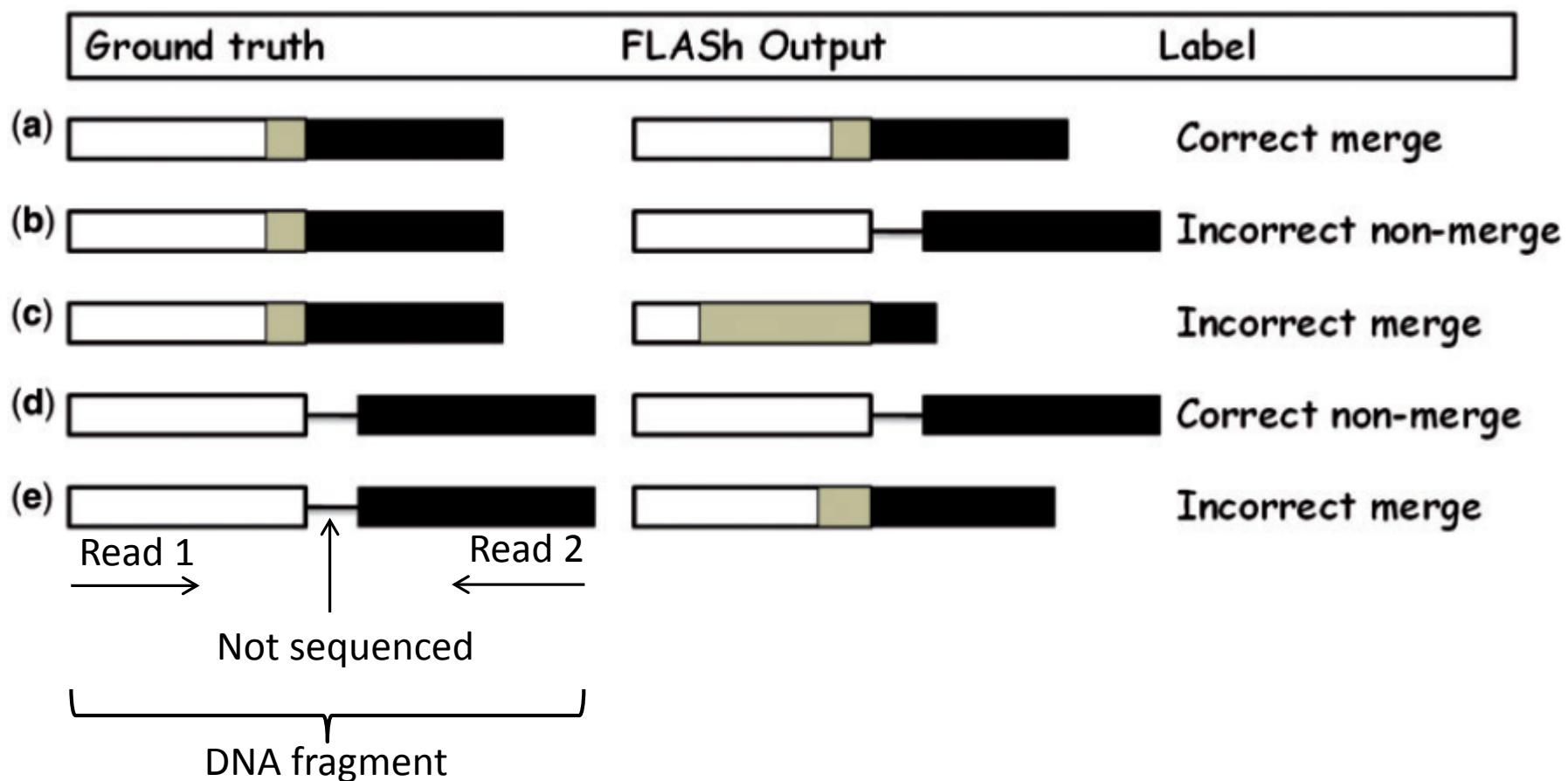
Assembled Contigs

Sequence Processing

Read Overlap

FLASH Majoc & Salzberg 2011 Bioinformatics 27:2957-2963.

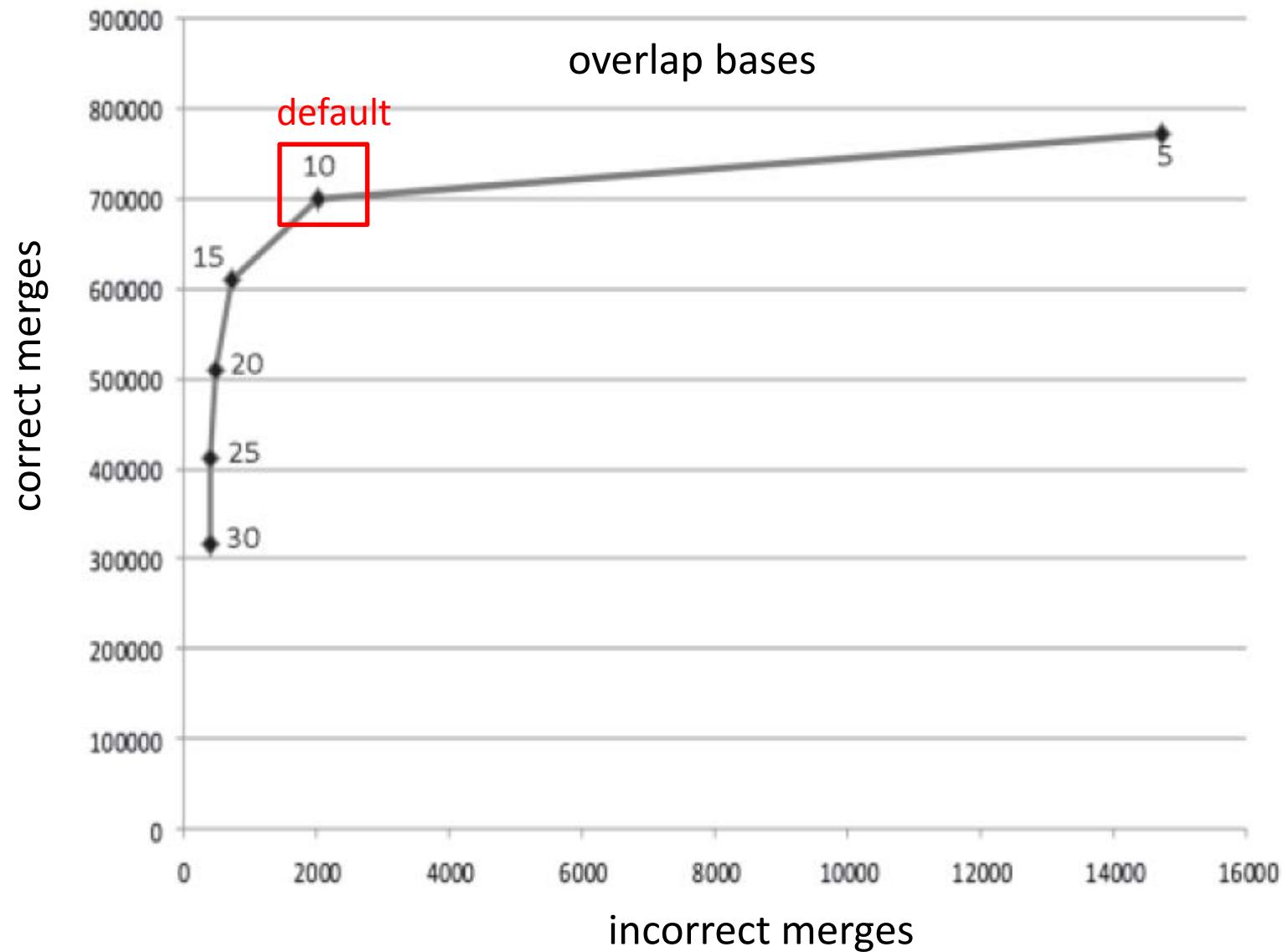
Possible Outcomes of Paired Read Merge Algorithm



Sequence Processing

Read Overlap

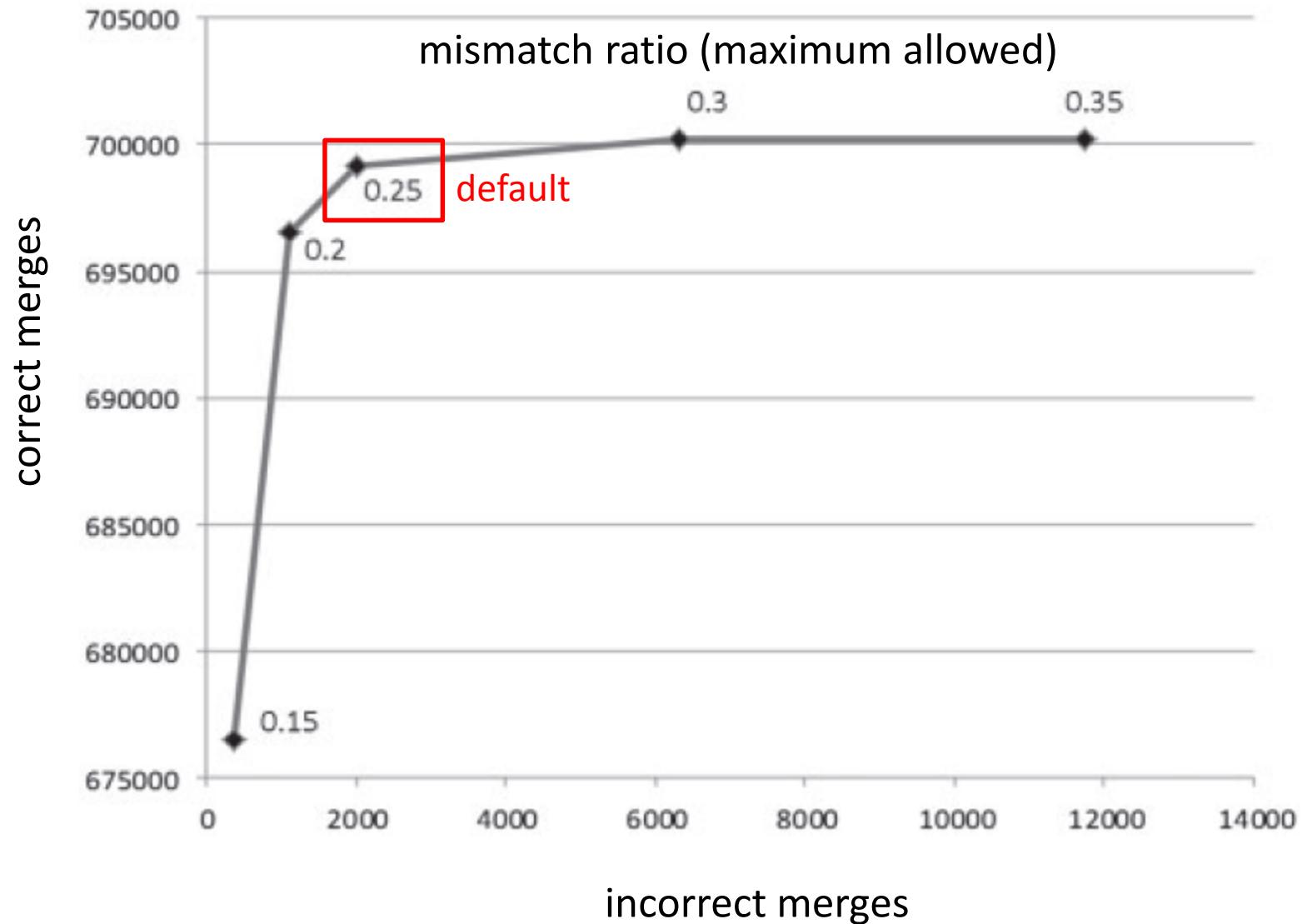
FLASH results with 1 million simulated reads, 1% error



Sequence Processing

Read Overlap

FLASH results with 1 million simulated reads, 1% error



Diverse and widespread contamination evident in the unmapped depths of high throughput sequencing data

Richard W Lusk

bioRxiv posted online January 30, 2014
Access the most recent version at doi:[10.](#)

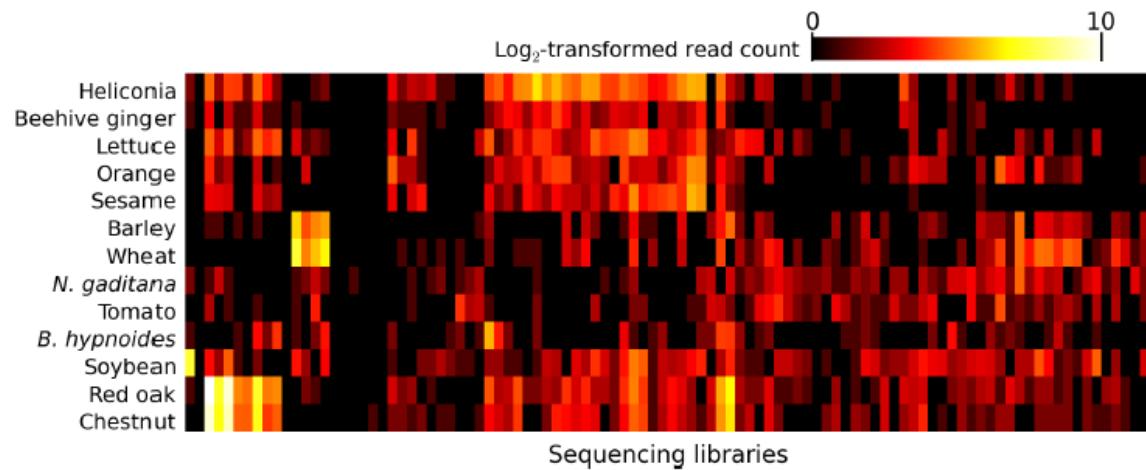


Figure 4. Heterogeneous species appear to contaminant samples from the same tissue and experiment. The “Tumor” [18] experiment dissociated 100 individual cells from a sample of a single tumor and sequenced libraries from each. Following the analysis pipeline of a study that claimed to find different plant species in different blood plasma samples from a single experiment, I used bowtie to screen each read in each library against

Bioinformatics

Illumina data from 1 lane:

200 million paired end reads

40 gigabase

800 million lines

60 gigabyte file size

Data Transfer

FTP

HTTP

portable hard drives



CORE LABS

NEWS & EVENTS

ETA

OSU Oregon State University

Main Menu

[HOME](#)

[CGRB BIOINFO](#)

[WEB ACCESS \(ETA\)](#)

[DOCUMENTATION](#)

[DONATIONS](#)

[POLICIES](#)

[ACCOUNT REQUEST](#)

[ANNOUNCEMENTS](#)

[SERVICES](#)

[WORKSHOP ACCOUNT](#)

[CLOUD STATUS](#)

[CONTACT](#)

CGRB Bioinformatics

Resources

- > 1800 AMD and Intel processors
- >900TB shared disk space
- Hardware:
 - AMD
 - Intel
 - Dell
 - Advanced HPC
 - Hewlett Packard
 - Sun Microsystem
 - APC and Eaton
- 10G main network connection
- Gigabit private network
- Secure, climate controlled

Projects

- Assembling the Fungal Tree of Life
- Marine Microbial Genomics
- CGRB Bioinformatics
- Plantontology.org
- Phytophthora-id.org



Bioinformatics

Computing Hardware

Desktop Computer with 16 GB RAM, 1 TB HD	\$1,500
48 CPU Server with 96 GB RAM	\$9,000
24 TB Storage Array	\$10,000

Bioinformatics

Data Processing Options

Commercial Packages (Geneious, CLC Bio)

\$400 and up

Bioinformaticist or Undergraduate Programmer

NGS experience or not

Web-Services

iPlant, Galaxy



Do It Yourself

Programming (R, Python, Perl, Ruby, Java ...)

Linux + Google

Bioinformatics

Open Source Software

Blat (Kent et al 2002)

Trimmomatic (Bolger et al 2014)

khmer (Brown et al 2014)

Velvet (Zerbino & Birney 2008)

YASRA (Ratan 2009)

BWA (Li & Durbin 2009)

Abyss (Simpson et al 2009)

Trinity (Grabherr et al 2011)

SAMTools (Li et al 2009)

Tablet (Milne et al 2010)

Scripting Languages:

Python, R swirlstats.com

Linux

My 6 favorite commands:

sed, grep, sort, uniq, join, awk

Sequence Similarity Searching

Short Read Quality Trimming

Digital Normalization

Short Read Assemblers

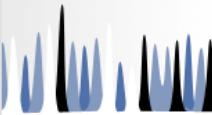
(RNA)

Analyzing Mapped Reads

Visualizing Mapped Reads

IP[y]: IPython
Interactive Computing
{swirl}
statistics with interactive R learning: an R package

For questions that we didn't answer:



SEQanswers

the next generation sequencing community

[SEQanswers > Bioinformatics](#)
[Bioinformatics](#)

User CP FAQ Community ▾ Calendar New Posts: 1,470 New Posts In ▾ Search ▾ Quick Links ▾ Log Out

New Thread

Welcome, [liston](#).
You last visited: 06-27-2014 at 02:57 PM

Threads in Forum : Bioinformatics

	Thread / Thread Starter	Last Post	Replies	Views
	Sticky: ICGC-TCGA DREAM Somatic Mutation Calling Challenge khoulahan	05-27-2014 10:47 PM by karenmartin	7	28,838
	Sticky: New Resources for 1000 Genomes (1 2) laura	12-05-2013 02:37 AM by laura	35	19,914
	Sticky: SEQwiki dan	08-09-2010 11:56 PM by scofield_qao	3	20,421
	Sticky: Software packages for next gen sequence analysis (1 2 3 ... Last Page) sci_guy	12-25-2009 06:45 PM by ECO	236	357,003
	How to blast a transposon sequence to a MiSeq transposon library helelein	Today 05:06 PM by Brian Bushnell	3	84
	KEGG Pathway menegidio	Today 04:21 PM by menegidio	0	27
	Help with Bowtie2 aakriti	Today 04:14 PM by GenoMax	9	122
	Remote blast+ skipping query sequences arundurvasula	Today 04:07 PM by GenoMax	1	69
	unmark duplicates SWP	Today 03:04 PM by mebbert	2	126
	Mitochondria genome-denovo assembly bioman1	Today 02:24 PM by francicco	12	895
	RNA-Seq Pathway and Gene-set Analysis Workflows in R/Bioconductor with GAGE/Pathview (1 2 3 ... Last Page) bigmw	Today 01:49 PM by tigerxu	82	7,528
	BioGPS, command line? sindrie	Today 01:44 PM by sindrie	4	238
	bootstrapping gene-content trees with discrete binary data using fseqboot (EBASSY) someperson	Today 01:20 PM by someperson	0	63
	HMMSearch on 6-frame translated genome vs protein sequences yields different results Loddi	Today 12:51 PM by Loddi	0	54

Introduction to NGS Sequencing Workshop

Botany 2015

8:30 - 8:40 am - Intro to the workshop and presenters

8:40 - 9:05 am - Intro to Sequencing Technology

9:05 - 9:30 am - Sequence Processing/Intro Bioinformatics

9:30 - 10:00 am - Group Activity

10:00 - 10:15 am - Coffee break

10:15 - 11:00 am - Genome Reduction Methods

11:00 - 11:30 am - Applications and data analysis - phylogenomics

11:30 am - 12:00 pm - Applications and data analysis - population genomics

12:00 - 1:00 pm - Lunch

1:00 - 1:20 pm - Introduction to the hands-on exercises

1:20 - 4:30 pm - Hands-on exercises

Whole Genome Sequencing (not covered here)

Genome Reduction

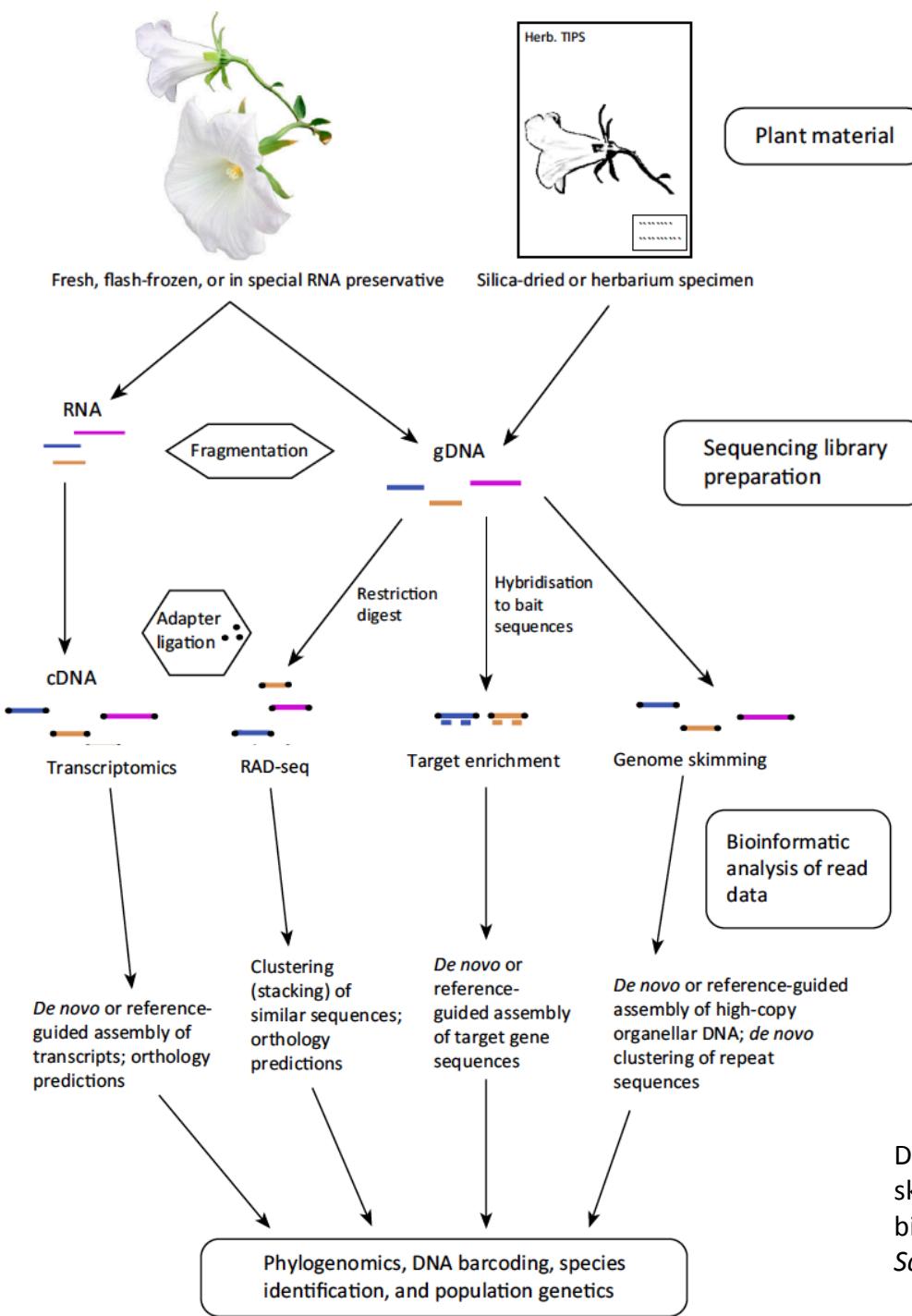
- A. PCR, fosmids, BACs
- B. Low Coverage (=Genome Skimming)
- C. Restriction Digest Methods (GBS, RAD, etc.)
- D. Target Capture (Hyb-Seq)
- E. Transcriptome (RNA-Seq)

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.

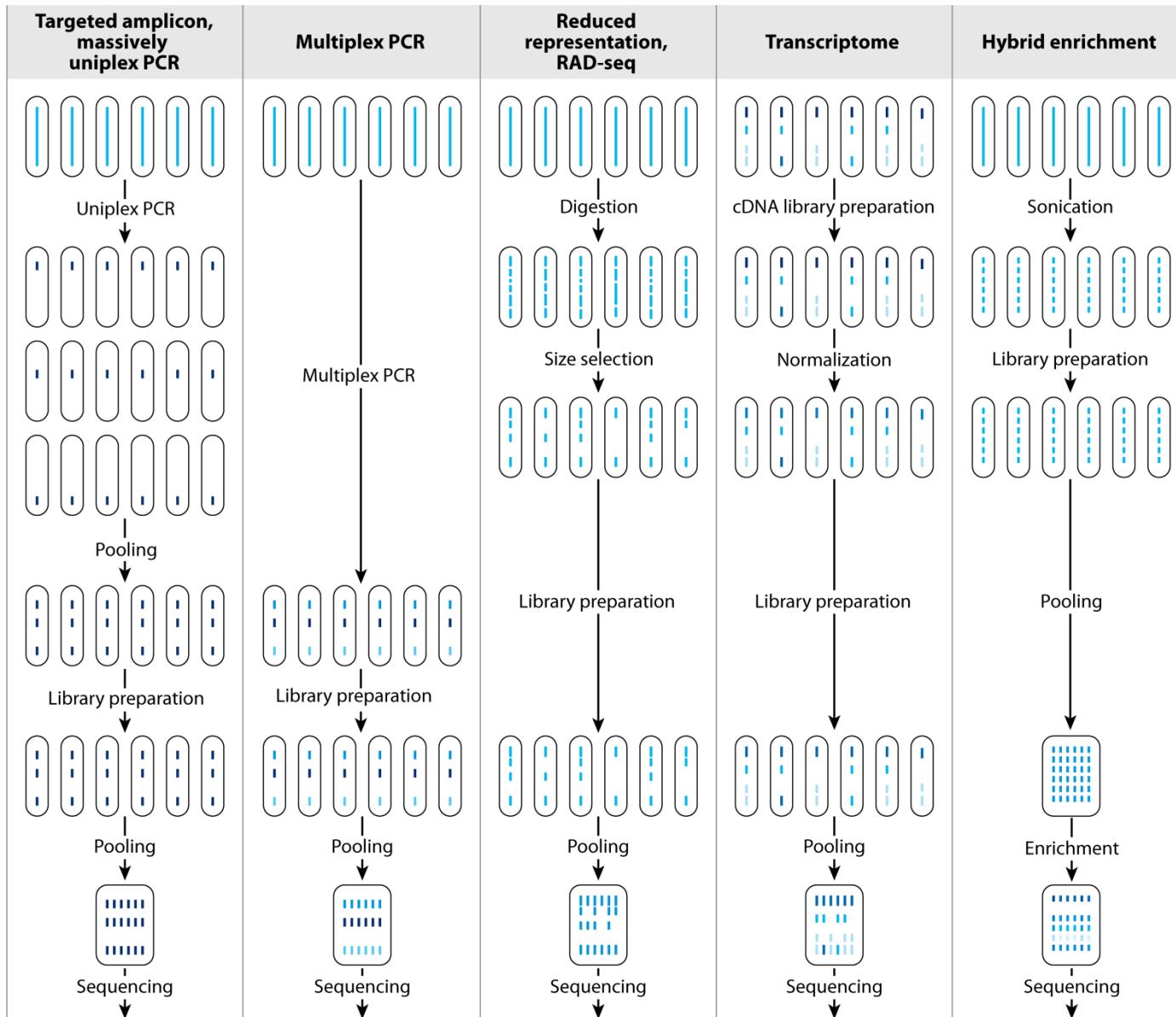
Lemmon & Lemmon. 2013. High-throughput genomic data in systematics and phylogenetics. Annual Review of Ecology, Evolution, and Systematics 44:99-121.

Ellegren H. 2014. Genome sequencing and population genomics in non-model organisms. Trends Ecol. Evol. 29:51–63.



Doddsworth, S. In Press. Genome skimming for next-generation biodiversity analysis. *Trends in Plant Science*.

Genome Reduction Approaches



A. Amplicons and BACs

= genome reduction prior to NGS library preparation

1. Amplicons Fluidigm automation

Uribe-Convers et al 2014. A Long PCR-Based Approach...
Appl. Plant Sci. 2:1300063.

48 amplicons X 48 samples = \$600

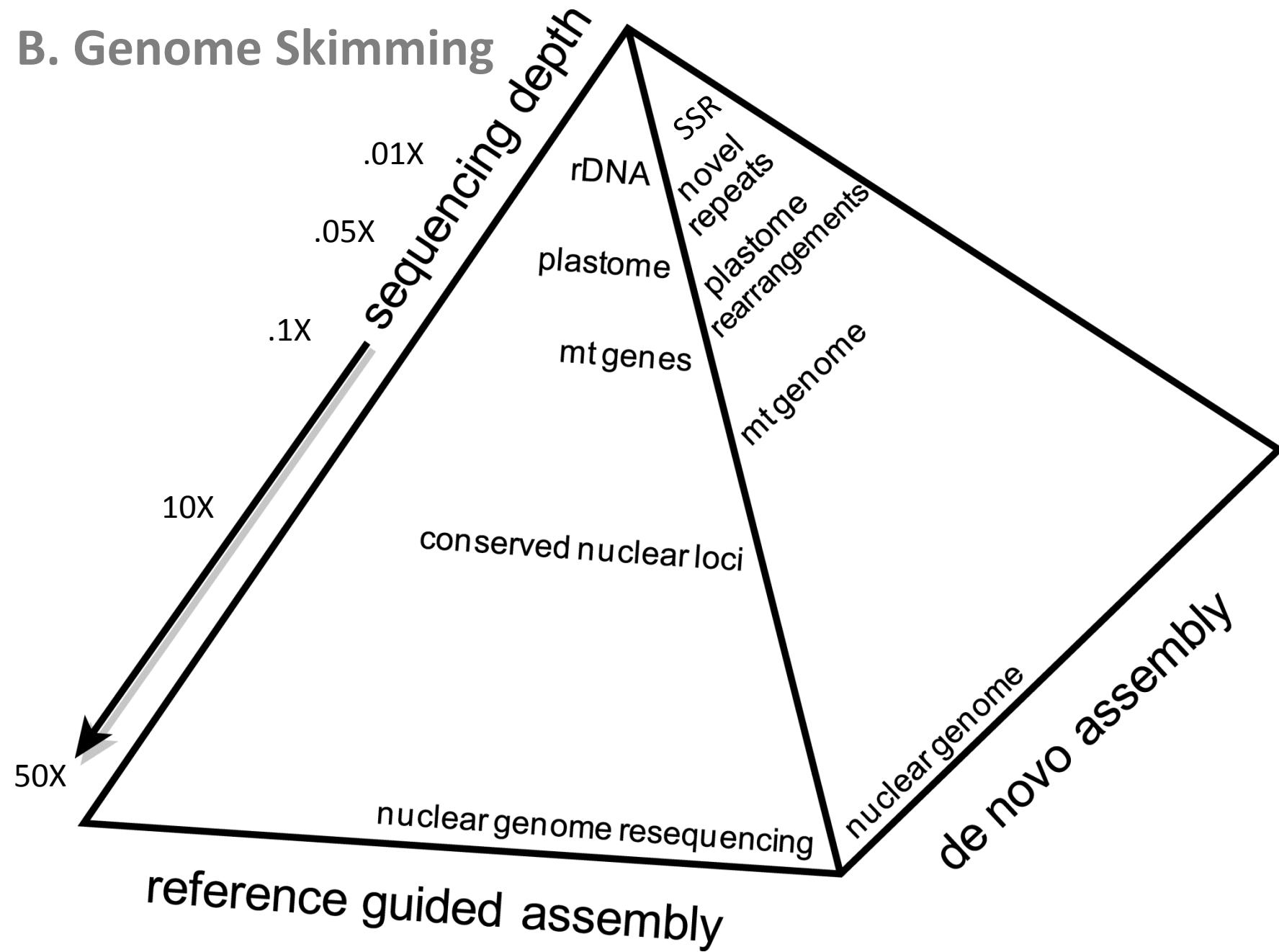
1000+ can be sequenced on MiSeq run = \$1500

2. Fosmids (10-40 kbp) and BACs (100-150 kbp)

Relatively expensive (\$10,000 for a BAC library)

Standard approach in large (well-funded) genome projects.

B. Genome Skimming



Genome Coverage



Low Coverage



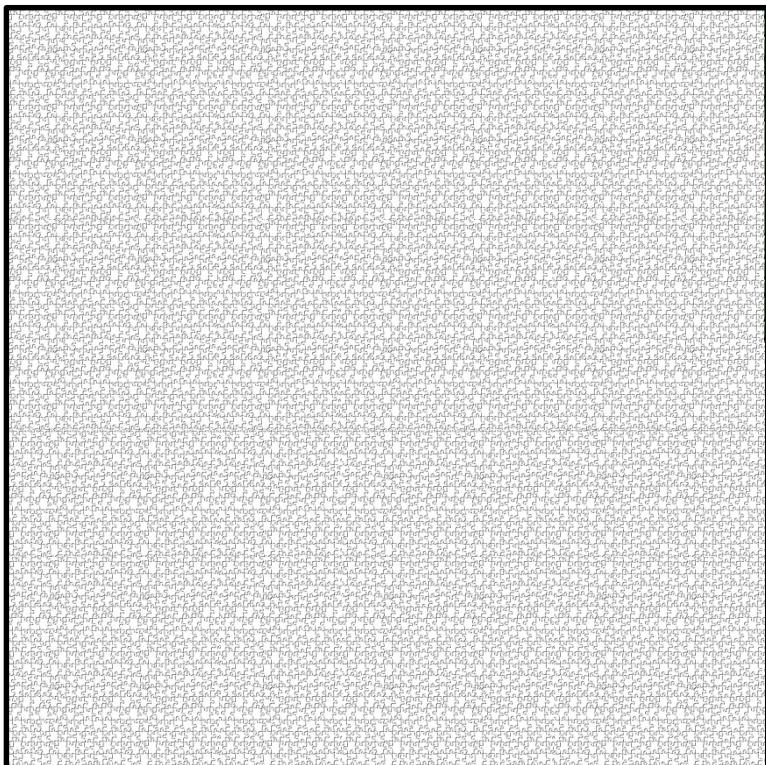
High Coverage

Genome Assembly

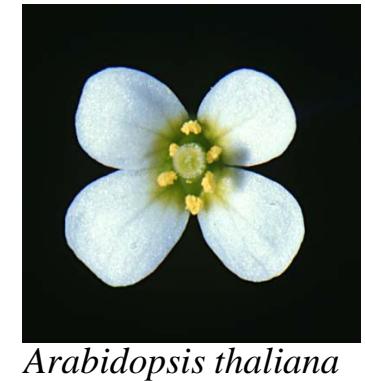
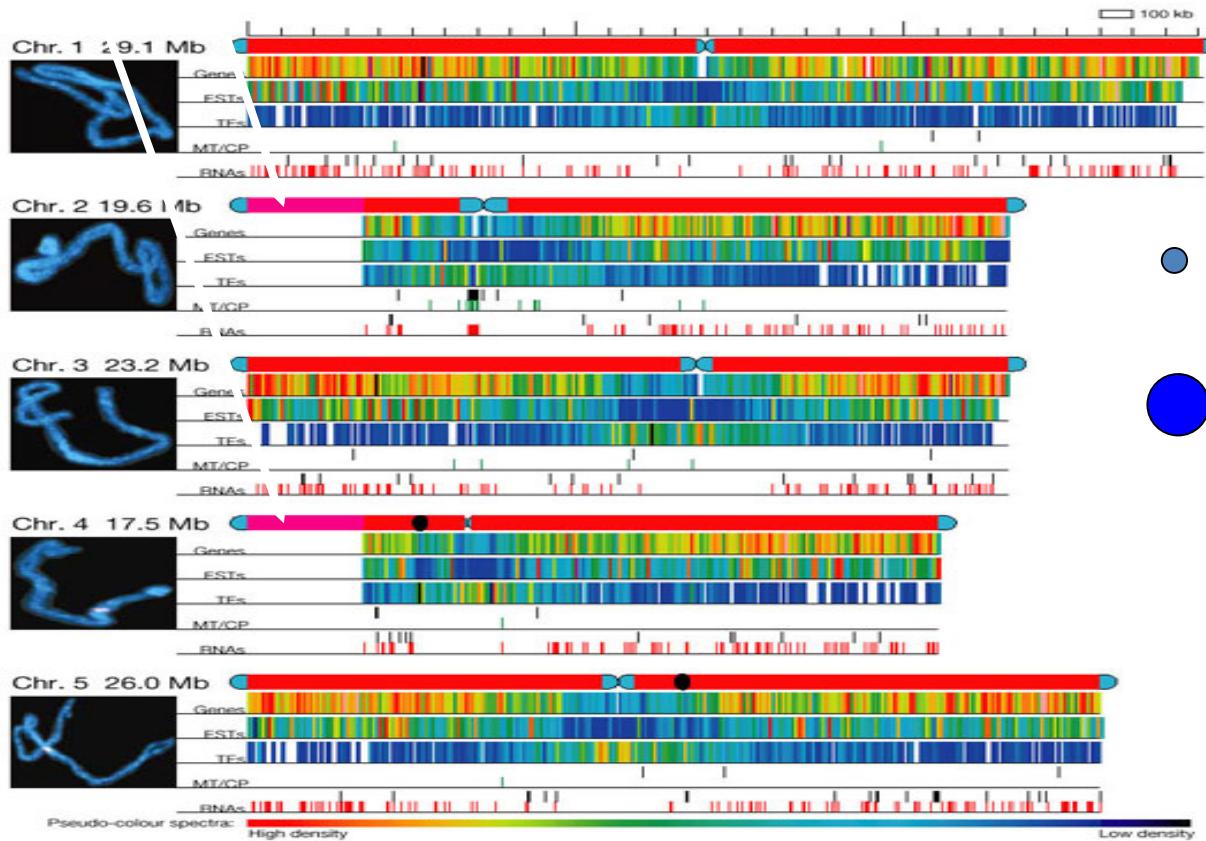
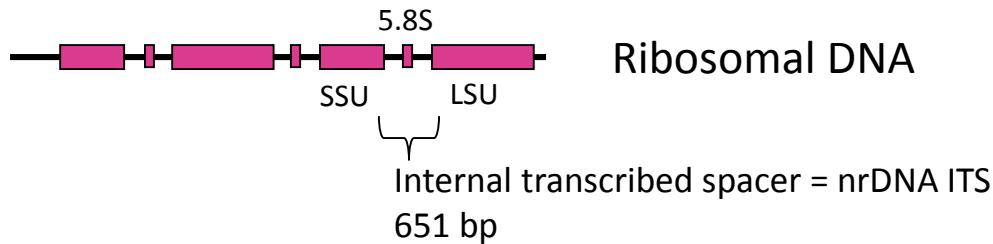
Reference guided assembly



De novo assembly



Plant Genomes

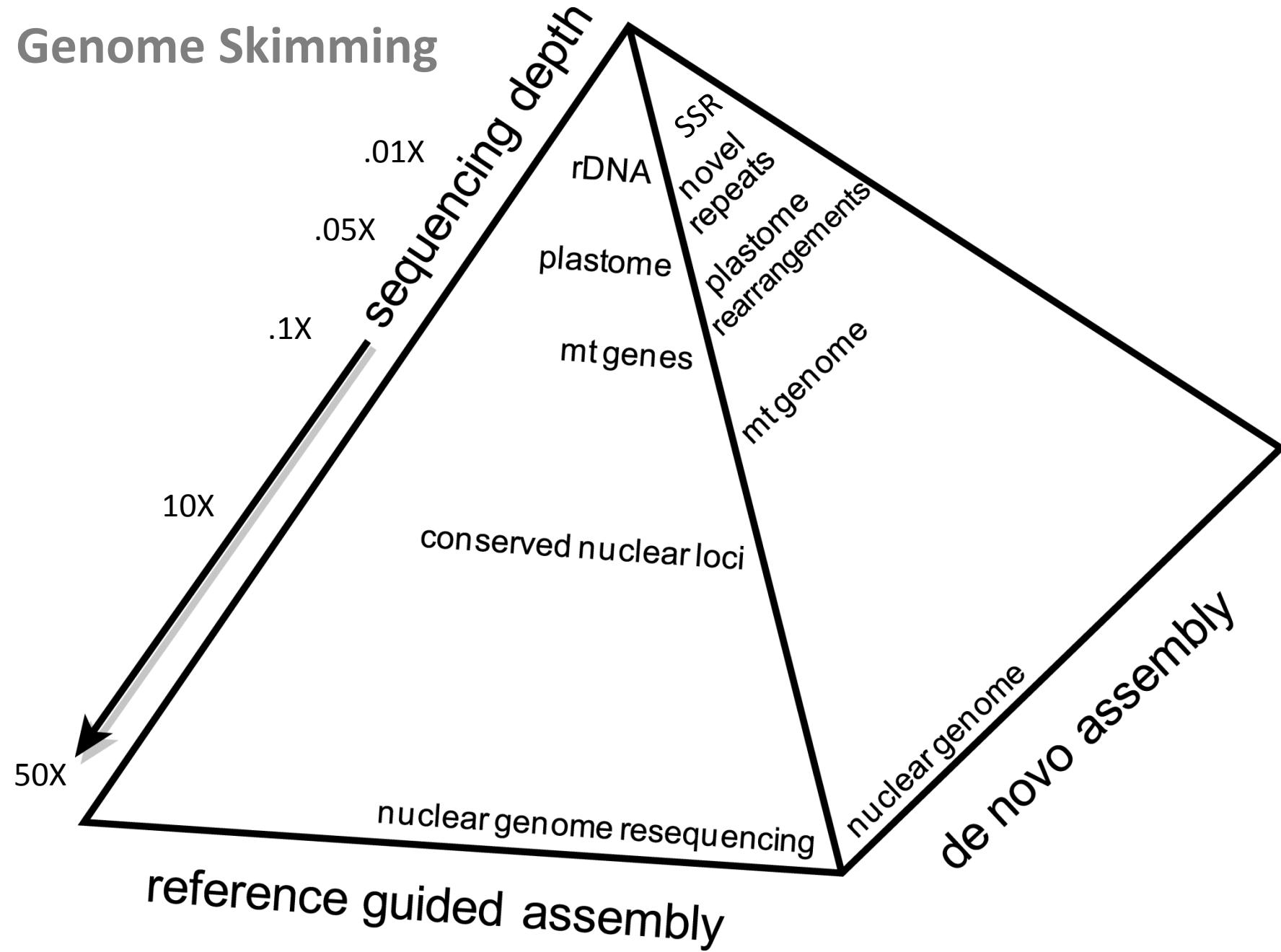


Chloroplast genome 154 kbp

Mitochondrial genome 367 kbp

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

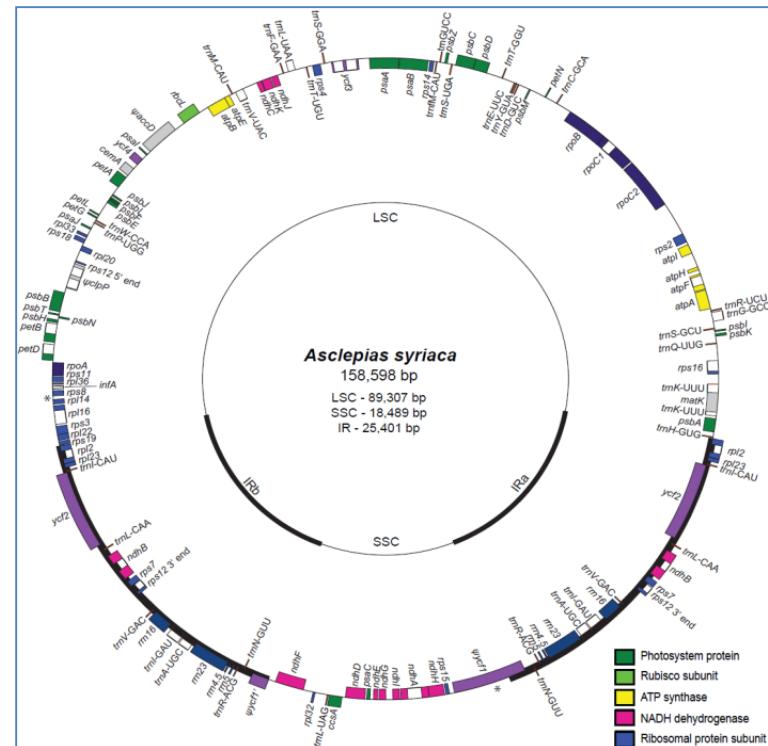
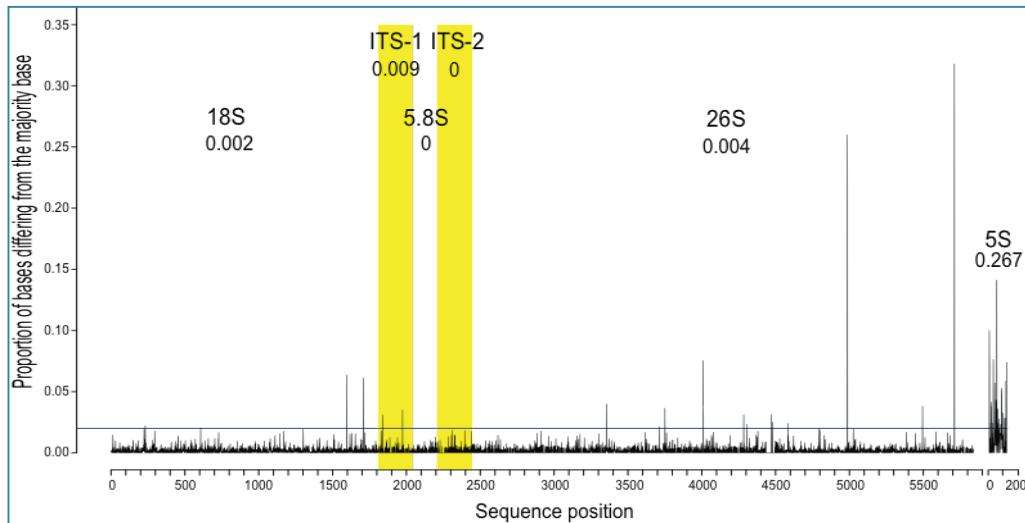
Genome Skimming



Genome Skimming in *Asclepias syriaca*

20 million 40 bp reads
1× sequencing depth
Reference guided assembly

Nuclear Ribosomal DNA
Chloroplast genome
Mitochondrial genes
Conserved single-copy nuclear genes

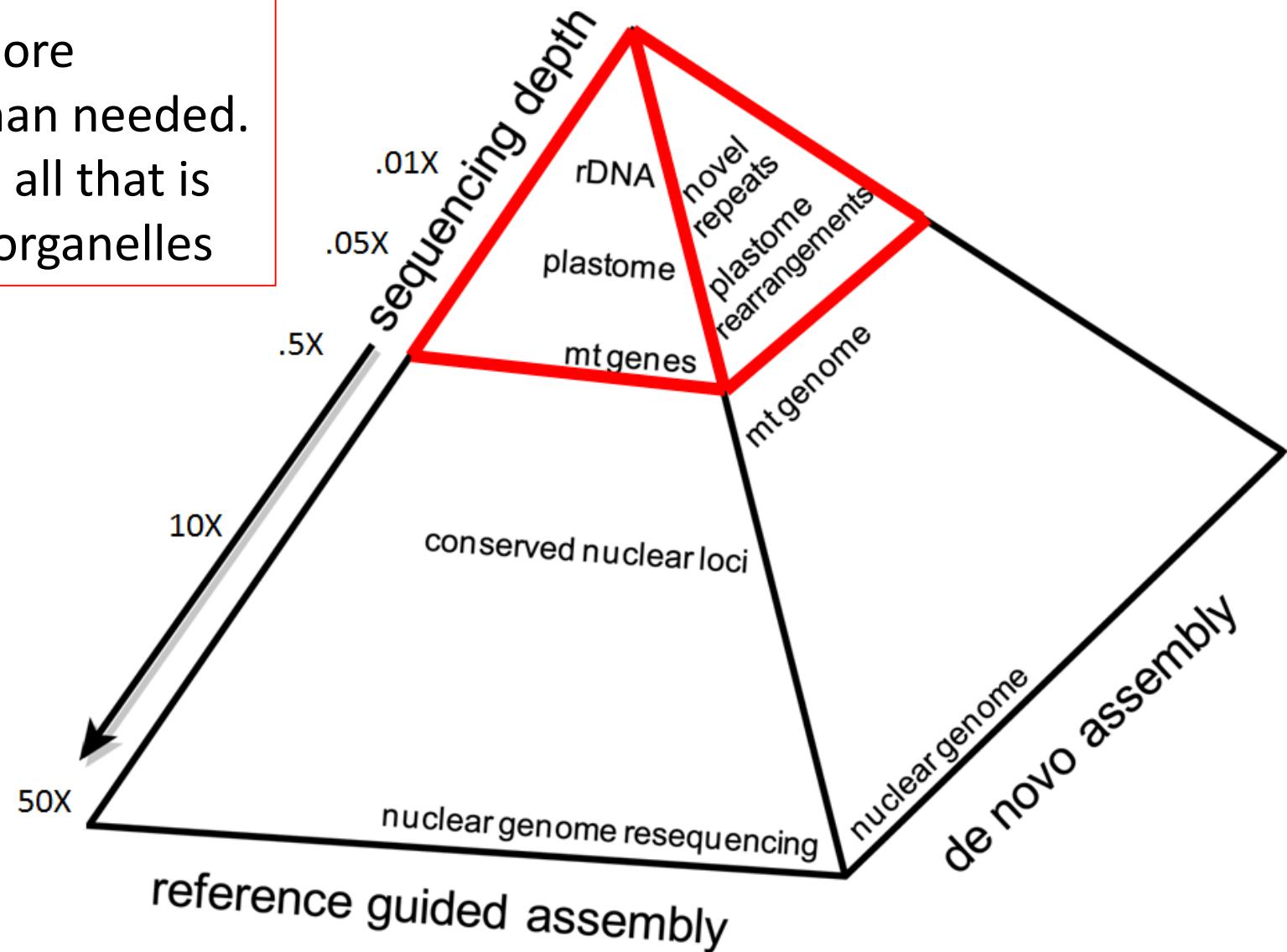


Straub et al. 2011. Building a model: Developing genomic resources for common milkweed (*Asclepias syriaca*) with low coverage genome sequencing. BMC Genomics 12: 211.

Genome Skimming

Important:

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



Genome Skimming

...NNNNNNNNNNNATATATATATATATATATATATATNNNNNNNNN...

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



C. Restriction Digest Approaches (RAD-Seq, GBS, 2b-RAD)

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



Main Menu

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

Search Site

GBS Overview



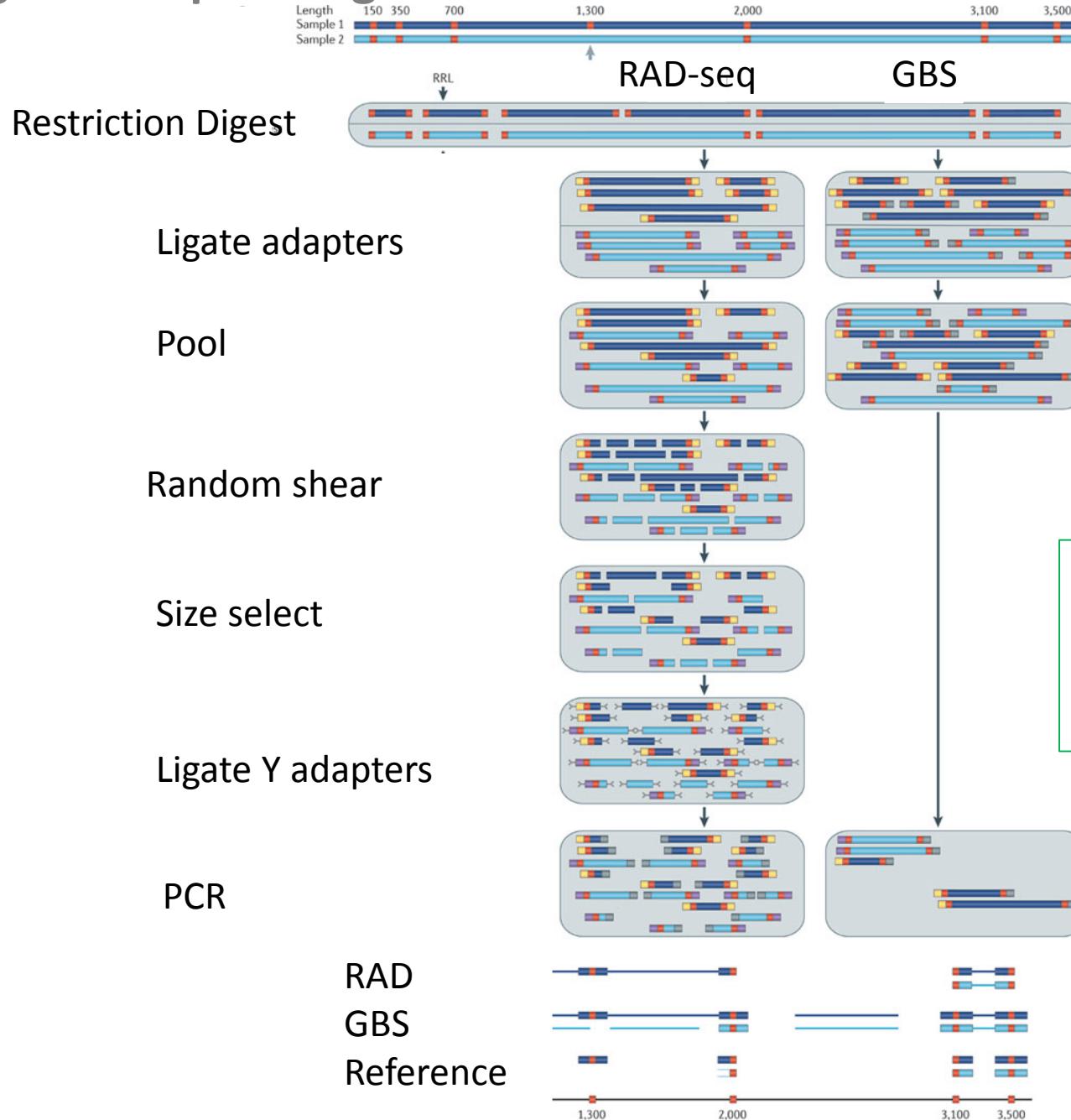
Genotyping by sequencing (GBS) is a simple highly-multiplexed system for constructing reduced representation libraries for the Illumina next-generation sequencing platform developed in the Buckler lab by [Rob Elshire](#). Key components of this system are: reduced sample handling, fewer PCR and purification 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.

Advantages of Restriction Digest approaches

1. Can obtain thousands of SNPs without a reference genome
2. Inexpensive library prep = hundreds to thousands of individuals
3. Discovery of candidate loci associated with phenotypic traits (quantitative trait loci = QTL) in natural populations (genome wide association studies = GWAS)
4. Analytical pipelines available
5. Hundreds of studies published to date.

Drawbacks of Restriction Digest approaches

1. Generally requires at least 1 µg of good quality DNA.
2. Restriction digestion adds another variable to the library prep.
 - a. Restriction site polymorphism results in missing data.
 - b. Short fragments are cut less efficiently.
3. Does not target specific genes or SNPs.
4. References are easily obtained, and soon to be widely available.
5. Loci are not transferrable among species.
6. High potential for ascertainment bias.

Restriction Site Bias

RADSEQ INDUCES ASCERTAINMENT BIAS 3187

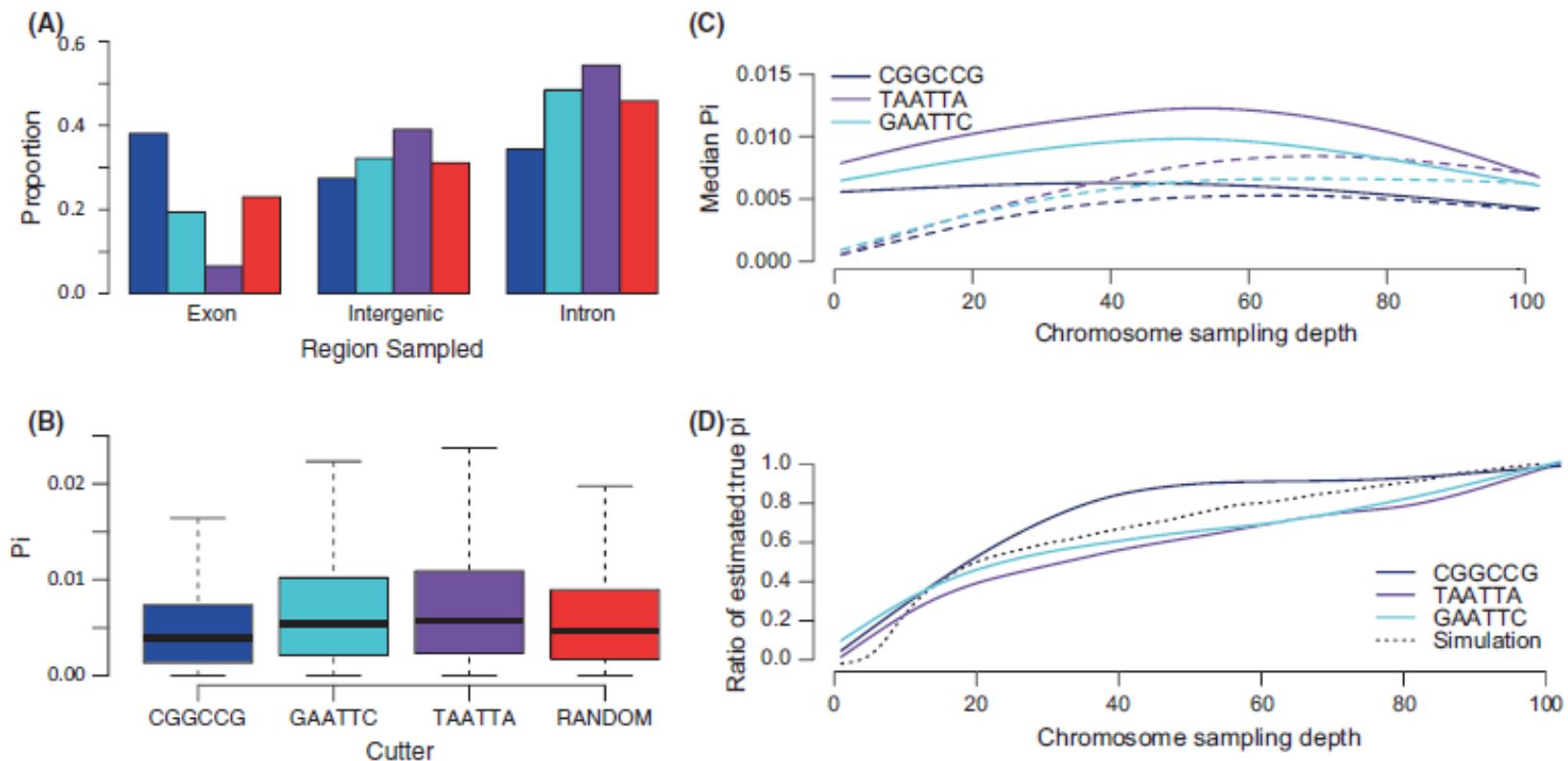
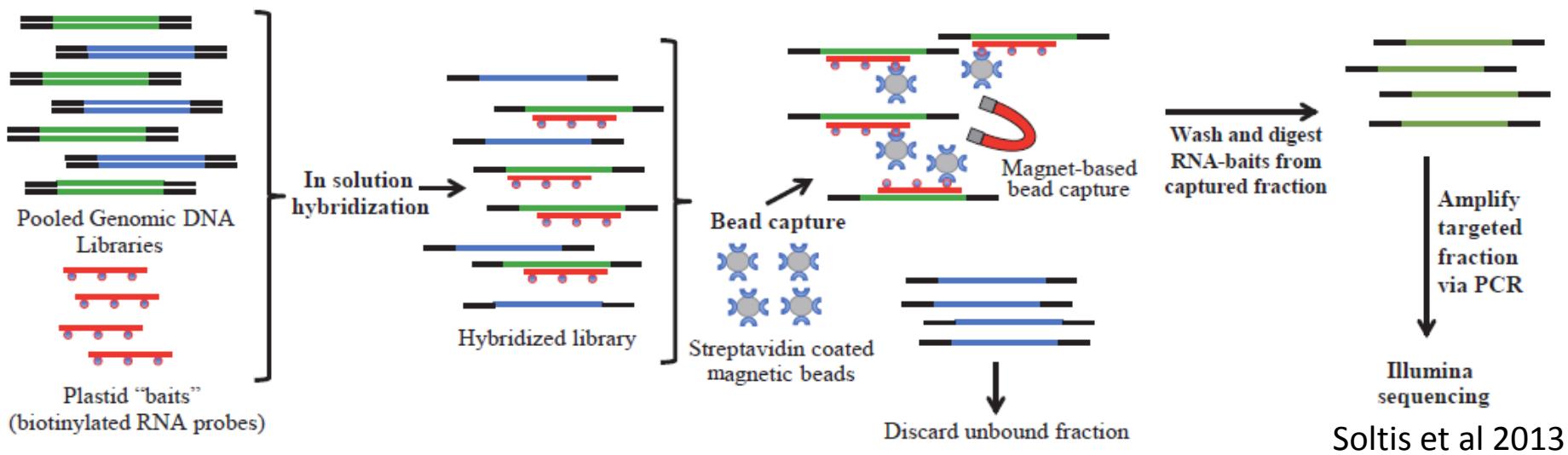


Fig. 8 Results for the *in silico* digests 102 *Drosophila melanogaster* genomes. (A) Proportion of sites located in distinct regions of the genome when *in silico* digests are performed with different enzyme recognition sequences. GC-rich recognition sequences sample more exons, whereas AT-rich recognition sequences sample comparatively more introns and intergenic regions. ‘Random’ values are calculated from fragments selected at random throughout the genome. (B) Box plots of true π for regions sampled by enzymes with different recognition sequences. (C) The median true π (solid line) and estimated π (dashed line) as a function of chromosome sampling depth for three different recognition sequences. (D) Median of the ratio of estimated π to true π as a function of the number of sampled chromosomes. Dark blue, purple and cyan lines represent the three different restriction enzymes used in the *in silico* digest of the *D. melanogaster* genomes, and the dotted black line is from simulations with $p = 0.1$ per bp, $\theta = 0.01$ per bp

D. Targeted Sequence Capture = Hyb-Seq



12-24 individuals per hybridization

0.45-2 Mbp / individual

48-96 individuals per MiSeq lane

=250-1000 nuclear genes
or 6500-20,000 SNPs

384 individuals per HiSeq lane

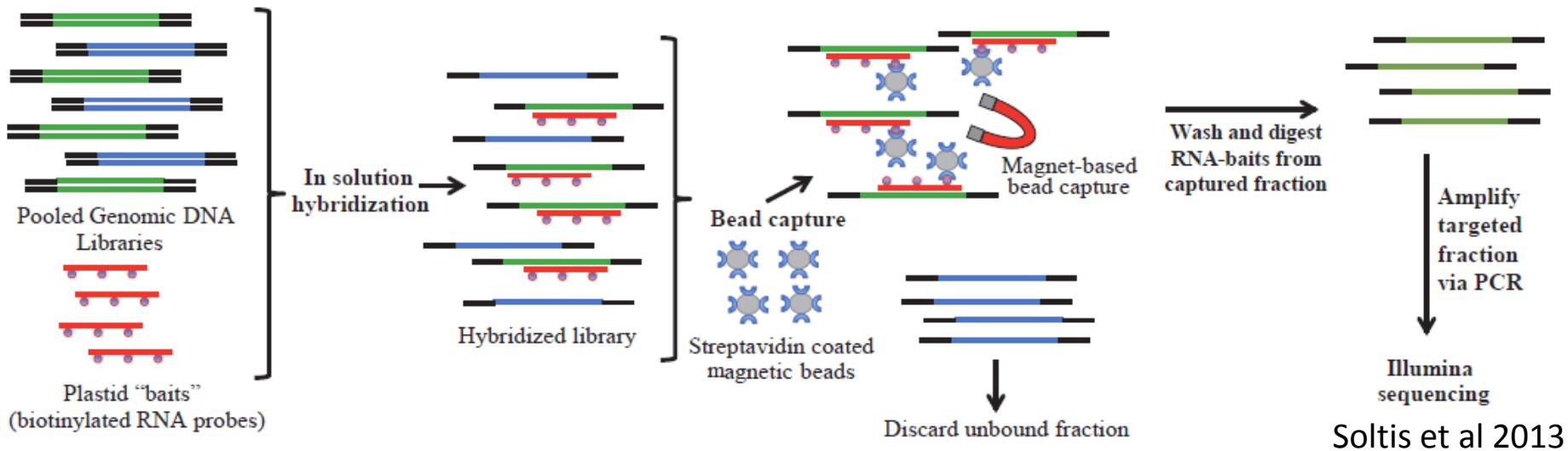
~50% efficiency
= 50% off-target genome
skimming

\$50-\$75 per individual

Weitemier et al. 2014 APPS 2(9): 1400042.

Soltis et al 2013

D. Targeted Sequence Capture = Hyb-Seq



12-24 individuals per hybridization

0.45-2 Mbp / individual

48-96 individuals per MiSeq lane

=250-1000 nuclear genes
or 6500-20,000 SNPs

384 individuals per HiSeq lane

~50% efficiency

= 50% off-target genome
skimming*

\$50-\$75 per individual

Weitemier et al. 2014 APPS 2(9): 1400042.

*new protocols >90% efficient, so spike with unenriched gDNA library (1:1 on HiSeq3000 and above) prior to sequencing if plastomes and repeats are desired

Advantages of Hyb-Seq

1. A single laboratory procedure and bioinformatics pipeline can be used for phylogenetics (deep and shallow), population genetics and genetic linkage mapping.
2. A relatively distant (e.g. plant family) reference can be used.
3. Candidate genes can be targeted.
4. Can be scaled from 250 genes to an entire exome (25,000 - 30,000 genes).
5. Data sets can be easily combined and extended (in contrast to GBS, SNP chips).
6. Minor potential for ascertainment bias.

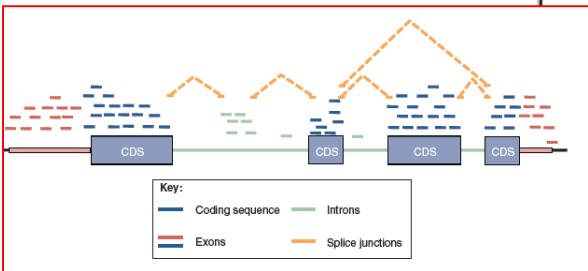
Genome Reduction

- A. PCR, fosmids, BACs
- B. Low Coverage (=Genome Skimming)
- C. Restriction Digest Methods (GBS, RAD, etc.)
- D. Target Capture (Hyb-Seq)
- E. Transcriptome (RNA-Seq)

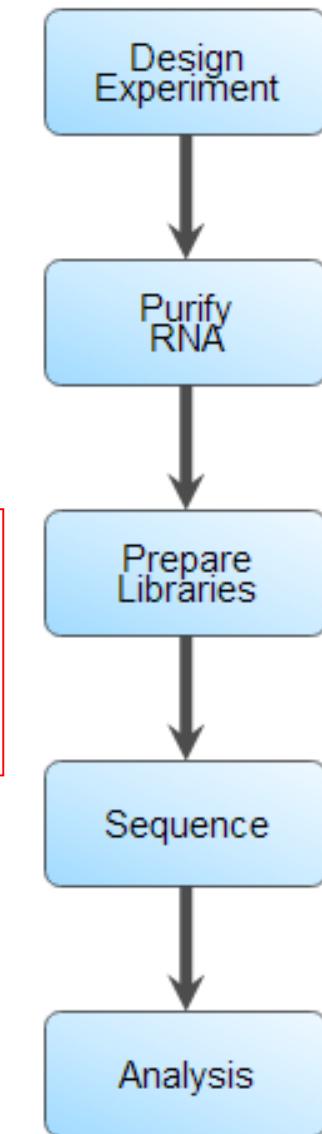
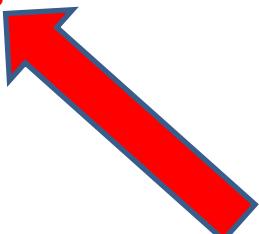
RNA-SEQ For Everyone

<http://rnaseq.uoregon.edu/index.html>

quantify gene expression



read mapping



1. Carefully design the experiment.

2. Isolate and purify input RNA.

3. Convert the RNA to cDNA and add sequencing adapters.

4. Sequence cDNAs using one of the available NGS platforms.

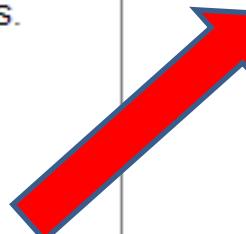
5. Analyze the resulting short-read sequences.

<http://rnaseq.uoregon.edu/index.html>

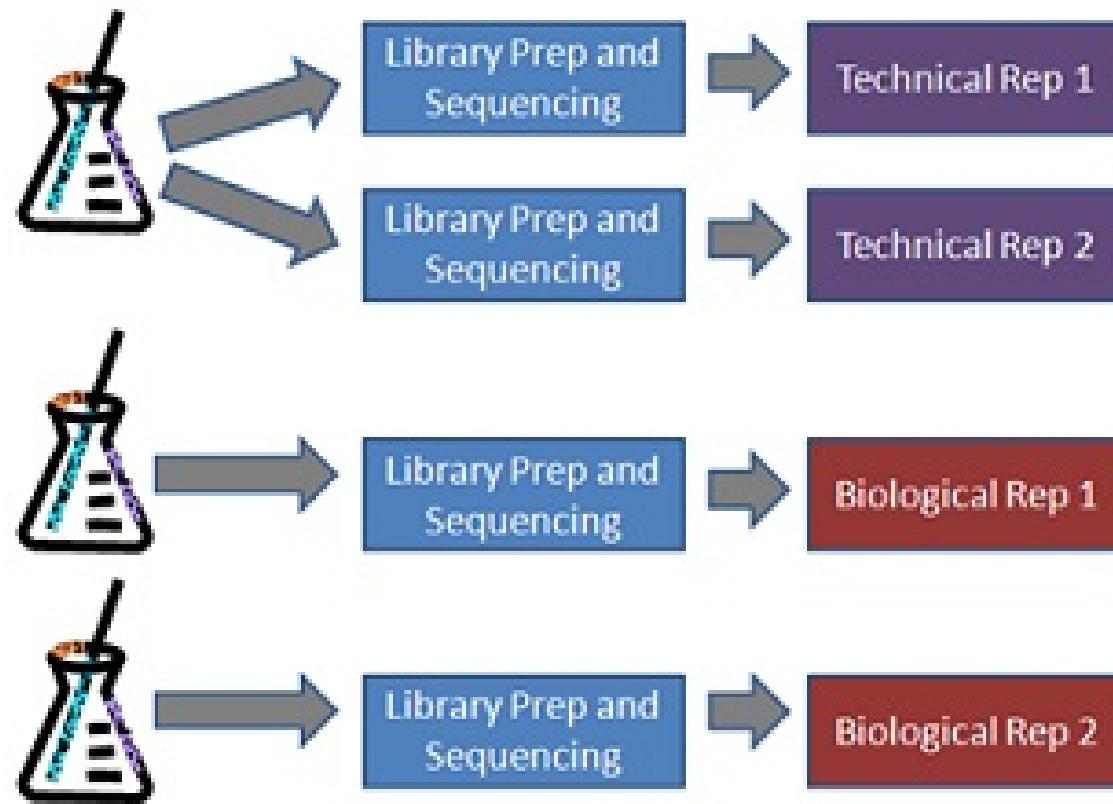
transcriptome reference genome reduction



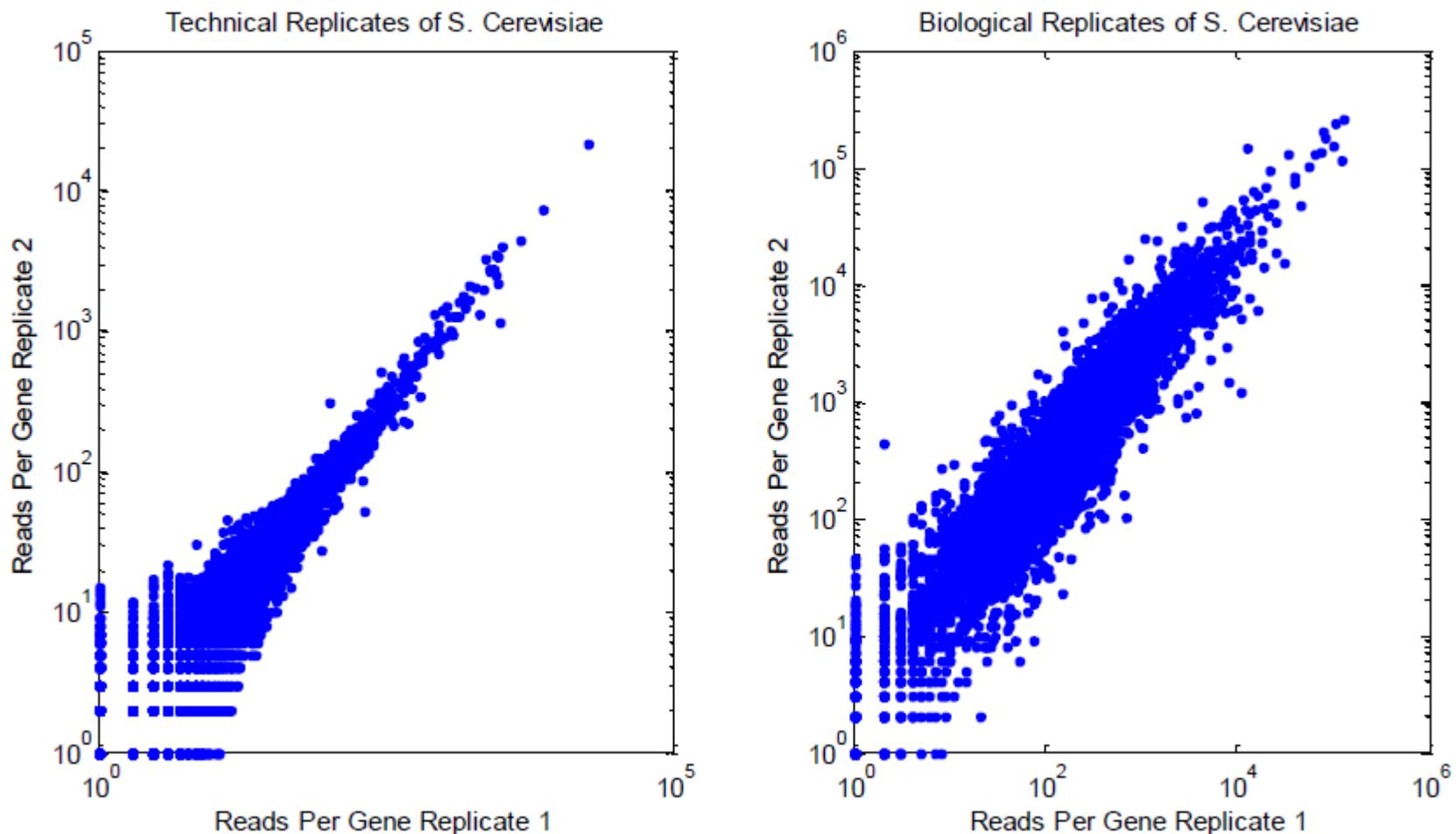
de novo assembly



Experimental Design



RNA-Seq Technical vs. Biological Replicates

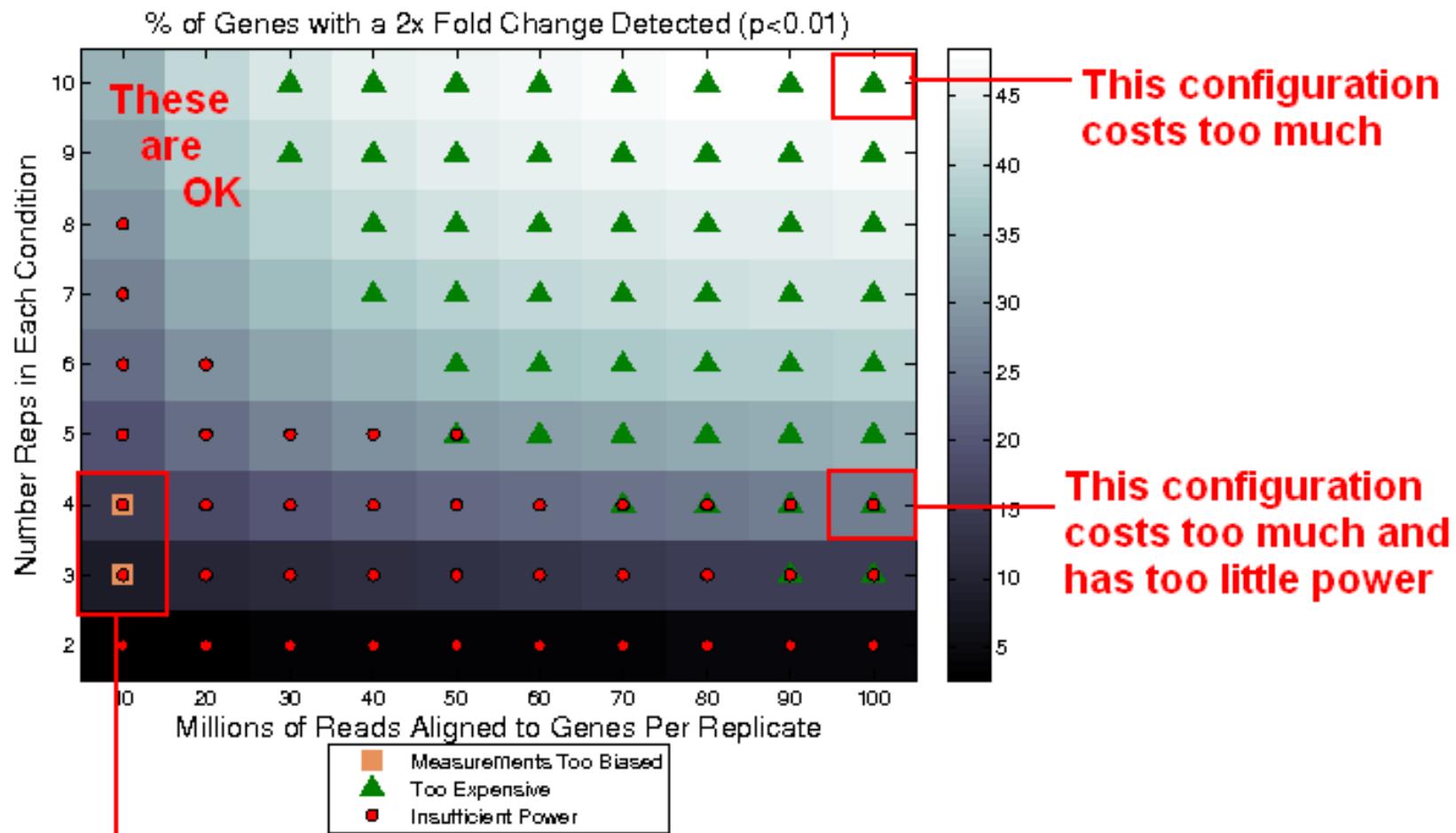


Busby et al. 2011. Expression divergence measured by transcriptome sequencing of four yeast species. BMC Genomics 12: 635.

Scotty - Power Analysis for RNA Seq Experiments

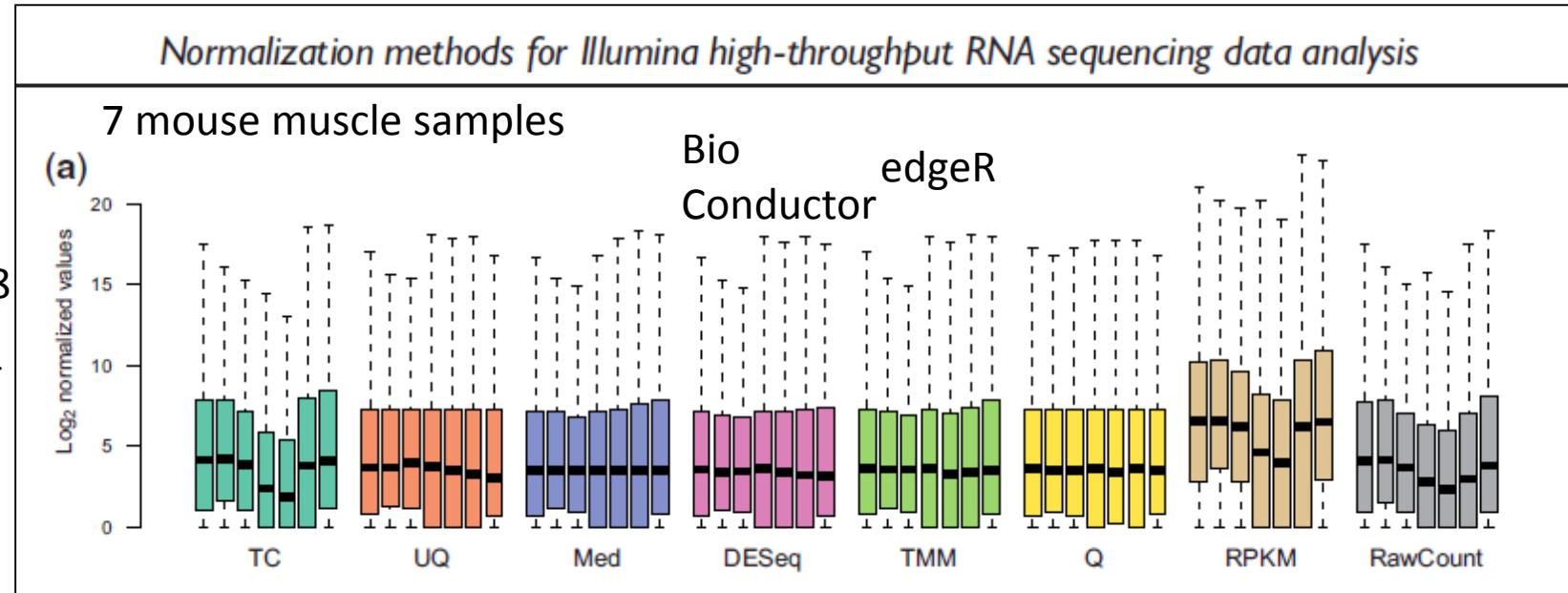
Scotty is a tool to assist in the designing of RNA Seq experiments that have adequate power to detect differential expression at the level required to achieve experimental aims.

<http://bioinformatics.bc.edu/marthlab/scotty/scotty.php>



These have too little power and too much measurement bias

Count Normalization



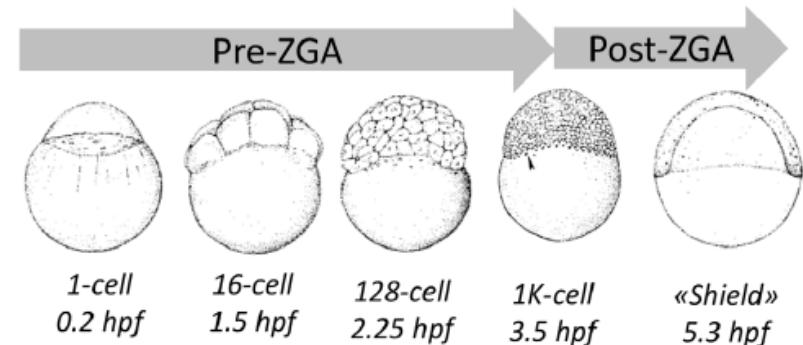
RPKM = reads per kilobase per million reads (most popular method, performs the worst)

All of these methods assume a constant denominator (total RNA per cell)

Dillies et al. 2013. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Briefings in Bioinformatics* 14: 671-683.

Total RNA Amount Varies Among Samples

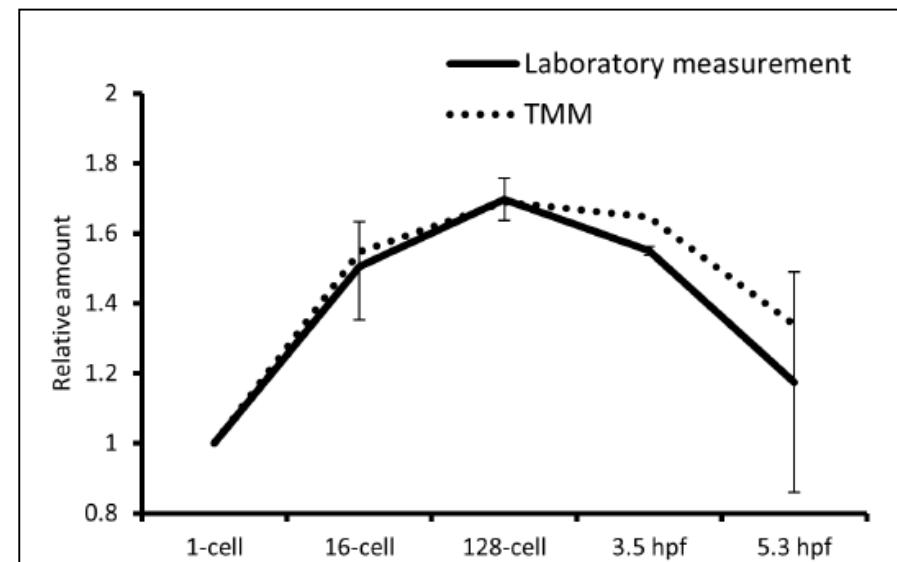
different developmental stages
different ploidy levels
(Ilut et al, 2012)



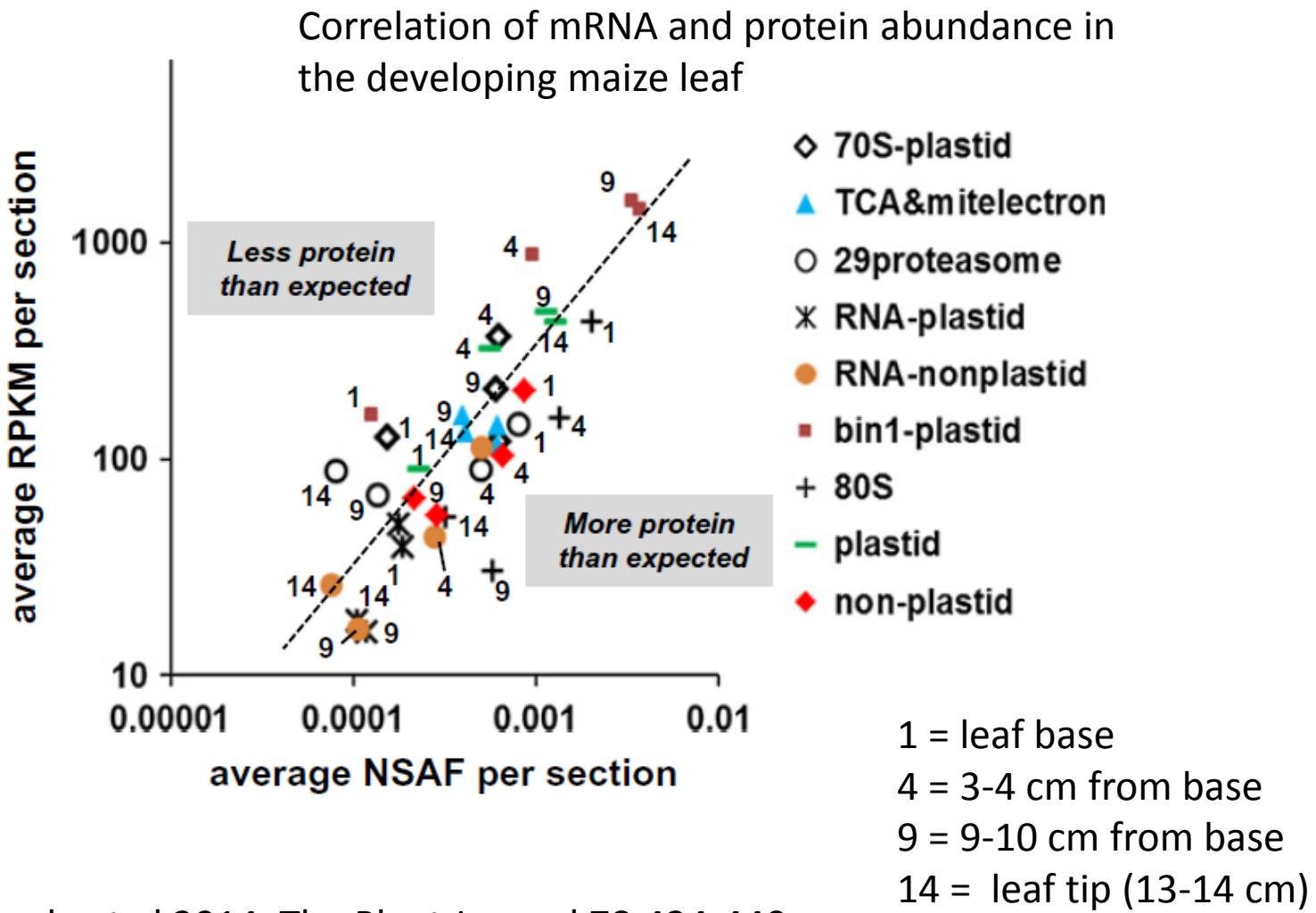
Potential Solutions:
Scale by Total RNA / Cell

Include Internal RNA Standards

Aanes et al. 2014. Normalization of RNA-Sequencing Data from Samples with Varying mRNA Levels. PLOS ONE 9: e89158.



RNA and proteins not perfectly correlated



Ponnala et al 2014. The Plant Journal 78:424-440.

RNA-Seq

Pros:

Powerful method with multiple applications:

SNPs, differential expression, genome annotation,
phylogenomics, alternative splicing, RNA editing

Cons:

Requires living tissue

RNA extraction is more involved than DNA

RNA expression biases (biological and analytical) are far from understood

Analysis of isoforms and paralogs is challenging

Introduction to NGS Sequencing Workshop

Botany 2015

8:30 - 8:40 am - Intro to the workshop and presenters

8:40 - 9:05 am - Intro to Sequencing Technology

9:05 - 9:30 am - Sequence Processing/Intro Bioinformatics

9:30 - 10:00 am - Group Activity

10:00 - 10:15 am - Coffee break

10:15 - 11:00 am - Genome Reduction Methods

11:00 - 11:30 am - Applications and data analysis - phylogenomics

11:30 am - 12:00 pm - Applications and data analysis - population genomics

12:00 - 1:00 pm - Lunch

1:00 - 1:20 pm - Introduction to the hands-on exercises

1:20 - 4:30 pm - Hands-on exercises

Phylogenomics

Mark Fishbein
Oklahoma State University
26 July 2015

I've got 100 samples and 1000 genes. Now what do I do?

Concatenation is out
From gene trees to species trees
Limitations of commonly used approaches
New approaches

Against concatenation

Genomes are mosaics
Introgressive hybridization
Horizontal gene transfer
Transposable elements
Incomplete lineage sorting

Gene Trees to Species Trees

What is a gene?

Non-recombining sequence

For most purposes, vertically inherited

Is there heterozygosity?

Polyplody?

Examine set of gene trees for congruence

BUCKy can be helpful

Coalescent-based analyses

Deep coalescence

Polymorphisms persist through speciation events

Random fixation

Age of allelic ancestor is older than species ancestor and allele tree may conflict with species tree

Multilocus methods

Parsimony minimize deep coalescence (MDC)

Bayesian modeling of gene trees and allele coalescence (*BEAST)

Maximum pseudolikelihood (MP-EST)

Also STEM, STAR

Limitations to pure coalescent methods

Do not account for non-vertical processes, e.g., introgression, HGT

Computational limitations

*BEAST, MP-EST limited to ~100 species, loci

Coping with Introgression

Detection—pop gen

Estimate migration time among populations: IMa2

Detection—phylogenetic

Eaton & Ree's quartet method; Syst Biol, 2013, 62:689-706

JML; Joly. Mol Ecol Resources, 2012, 12:179-184 (evaluates introgression vs. ILS)

HybTree; Gerard et al. BMC Evolutionary Biology 2011, 11:291 (evaluates introgression vs. ILS)

New Methods

Overcoming computational limits

Random subsampling loci with *BEAST: BBCA (Boosted Bin Coalescent Analysis); Zimmermann et al. BMC Genomics 2014, 15(Suppl 6):S11

Random subsampling species with MP-EST: DACTAL-boosting, SSG-boosting; Bayzid et al. BMC Genomics 2014, 15(Suppl 6):S7

Random resampling possible

Constrained searches: ASTRAL-II; Mirabab & Warnow. Bioinformatics 2015, 31:144-152

New Methods

Species trees from SNPs
RADseq, GBS, etc.

SNAP-MCMC, Bryant et al., Molec Biol Evol, forthcoming

More to ponder

Gene families, paralogy, loss, expansion

Genome structure, synteny

TE content, organization, evolution

Introduction to NGS Sequencing Workshop

Botany 2015

8:30 - 8:40 am - Intro to the workshop and presenters

8:40 - 9:05 am - Intro to Sequencing Technology

9:05 - 9:30 am - Sequence Processing/Intro Bioinformatics

9:30 - 10:00 am - Group Activity

10:00 - 10:15 am - Coffee break

10:15 - 11:00 am - Genome Reduction Methods

11:00 - 11:30 am - Applications and data analysis - phylogenomics

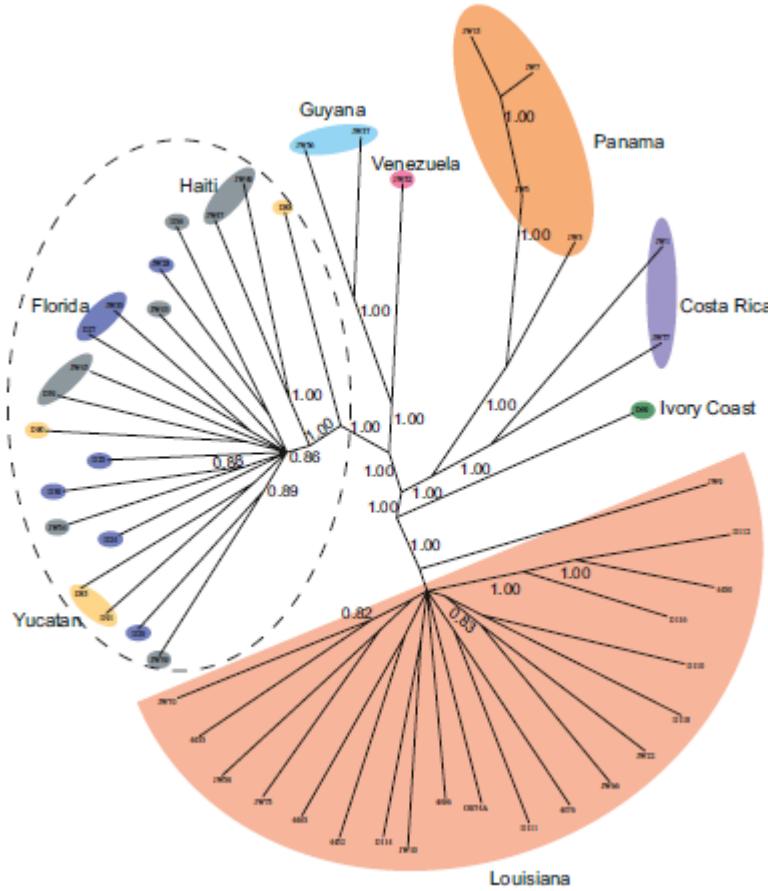
11:30 am - 12:00 pm - Applications and data analysis - population genomics

12:00 - 1:00 pm - Lunch

1:00 - 1:20 pm - Introduction to the hands-on exercises

1:20 - 4:30 pm - Hands-on exercises

Population Genomics with High-Throughput Sequencing



Botany 2015
Edmonton, Alberta
Introduction to Next-generation Sequencing

Kevin Weitemier

FROM THE COVER

Rapid genetic adaptation precedes the spread of an exotic plant species

KATRIEN VANDEPITTE,* TIM DE MEYER,† KENNY HELSEN,* KASPER VAN ACKER,*
ISABEL ROLDÁN-RUIZ,‡ JOACHIM MERGEAY§ and OLIVIER HONNAY*

Adaptive differences between
native and invasive populations

390 contemporary individuals
plus 42 herbarium records
(1829–1955).

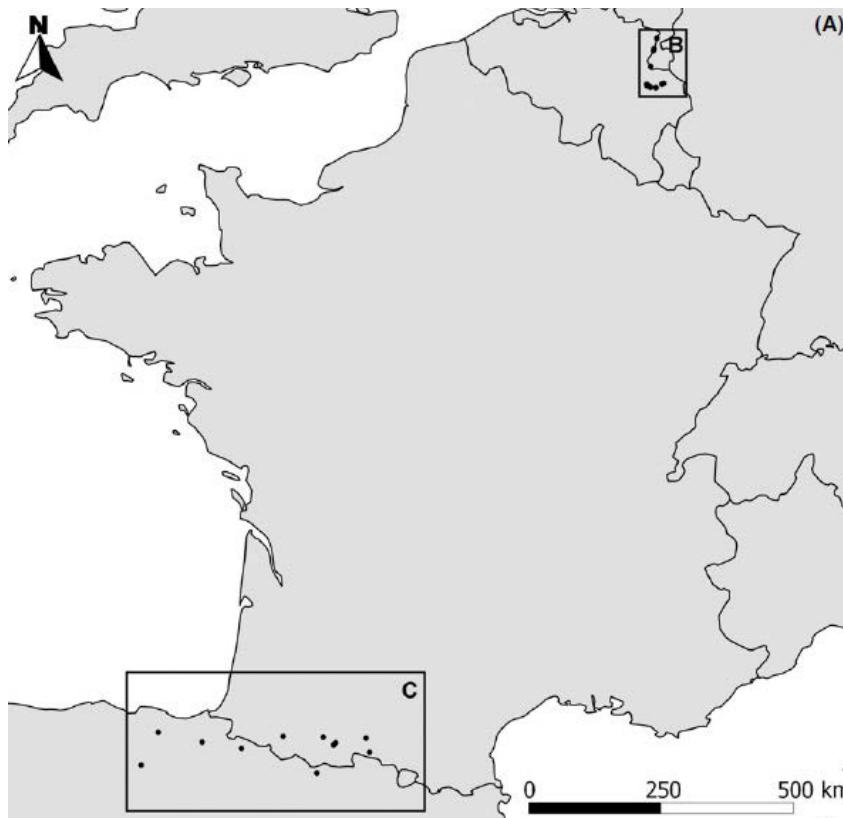


"*Sisymbrium austriacum*" CC BY-SA flickr Joan Simon

FROM THE COVER

Rapid genetic adaptation precedes the spread of an exotic plant species

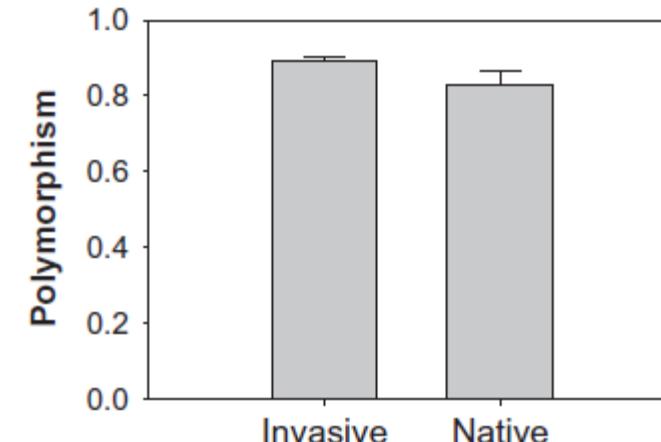
KATRIEN VANDEPITTE,* TIM DE MEYER,† KENNY HELSEN,* KASPER VAN ACKER,*
ISABEL ROLDÁN-RUIZ,‡ JOACHIM MERGEAY§ and OLIVIER HONNAY*



Previous RAD-Seq study developed ~14,000 SNPs.

Here focus on 300 SNPs related to genes with certain GO terms.

Assayed using KASPar (allele-specific PCR)



FROM THE COVER

Rapid genetic adaptation precedes the spread of an exotic plant species

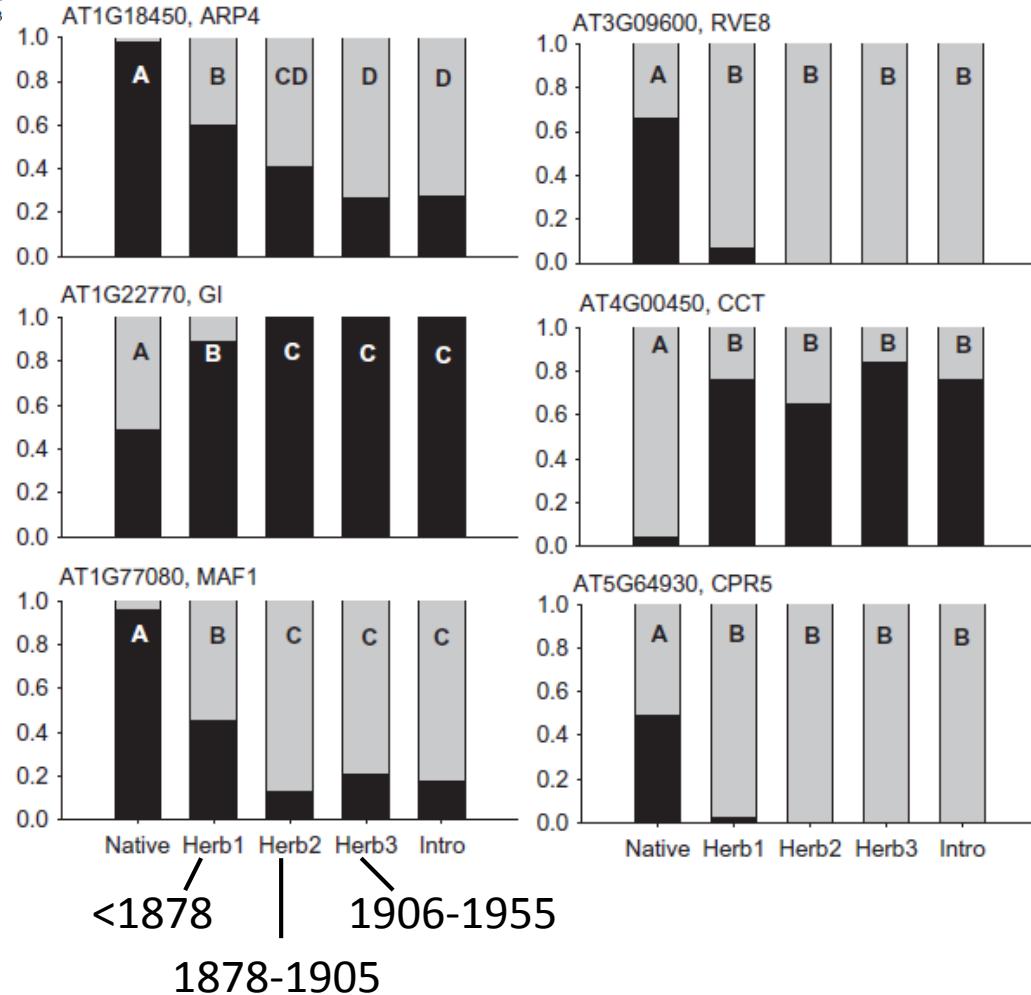
KATRIEN VANDEPITTE,* TIM DE MEYER,† KENNY HELSEN,* KASPER VAN ACKER,*
ISABEL ROLDÁN-RUIZ,‡ JOACHIM MERGEAY§ and OLIVIER HONNAY*

Significant change in frequency
of 6 alleles over time.

Genes related to
photoperiodism and flowering

Good example of:

- Using existing genomic resources (previous SNP study)
- Use of resources from related organisms (*Arabidopsis*)
- Use of degraded/historical DNA



Population genomics and local adaptation in wild isolates of a model microbial eukaryote

Christopher E. Ellison^a, Charles Hall^a, David Kowbel^a, Juliet Welch^a, Rachel B. Brem^b, N. L. Glass^a, and John W. Taylor^{a,1}

PNAS | February 15, 2011 | vol. 108 | no. 7 | 2831–2836

Ascomycete *Neurospora crassa*

48 isolates from Louisiana,
Caribbean, South America, Africa

Whole transcriptome sequencing
(RNA-Seq)

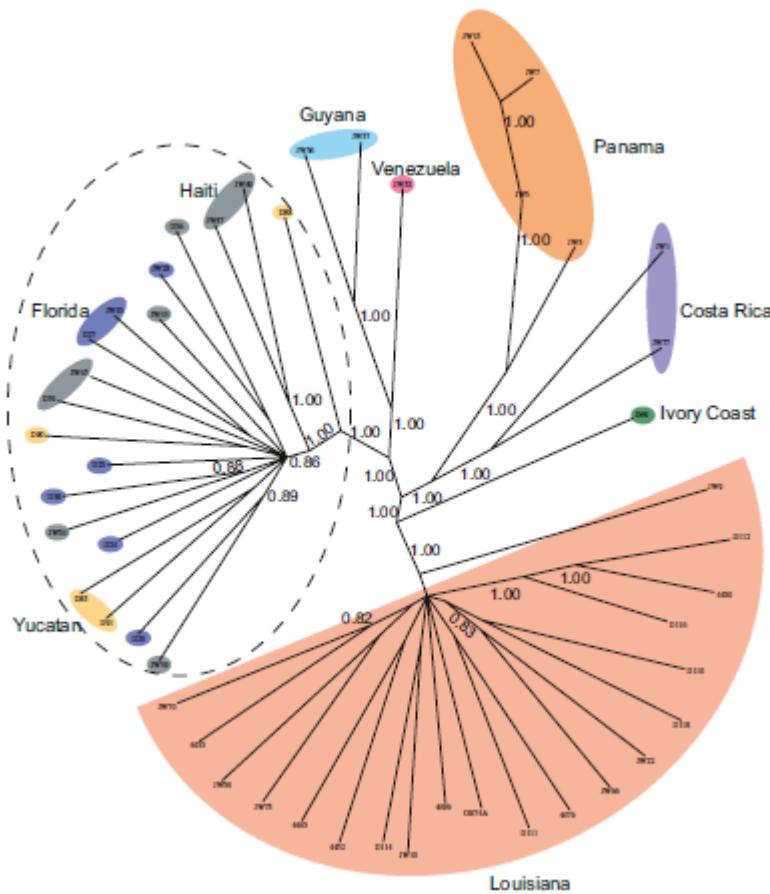
Identified ~135,000 SNPs

Genome available



Population genomics and local adaptation in wild isolates of a model microbial eukaryote

Christopher E. Ellison^a, Charles Hall^a, David Kowbel^a, Juliet Welch^a, Rachel B. Brem^b, N. L. Glass^a, and John W. Taylor^{a,1}



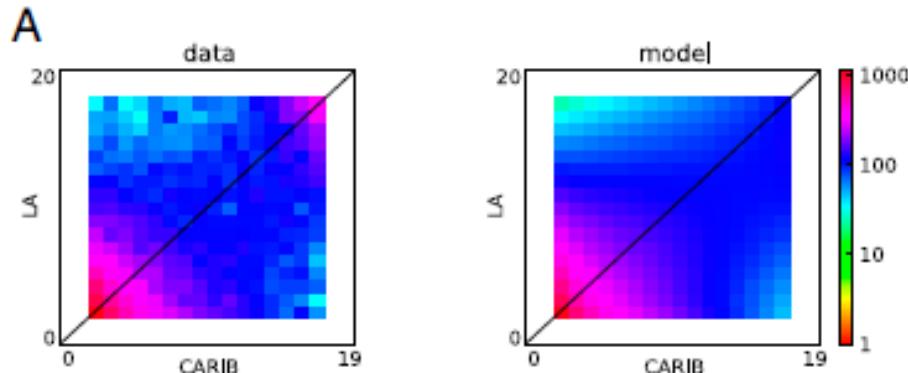
Identification of two cryptic populations:

Louisiana and Caribbean
Phylogenetic and allele frequency methods

High F_{ST} (0.19)

Population genomics and local adaptation in wild isolates of a model microbial eukaryote

Christopher E. Ellison^a, Charles Hall^a, David Kowbel^a, Juliet Welch^a, Rachel B. Brem^b, N. L. Glass^a, and John W. Taylor^{a,1}



Analysis of allele frequency spectrum between populations.

Compared to model of isolation with asymmetric migration.

Multiple comparisons with *Saccharomyces* demonstrating increased levels of outcrossing.

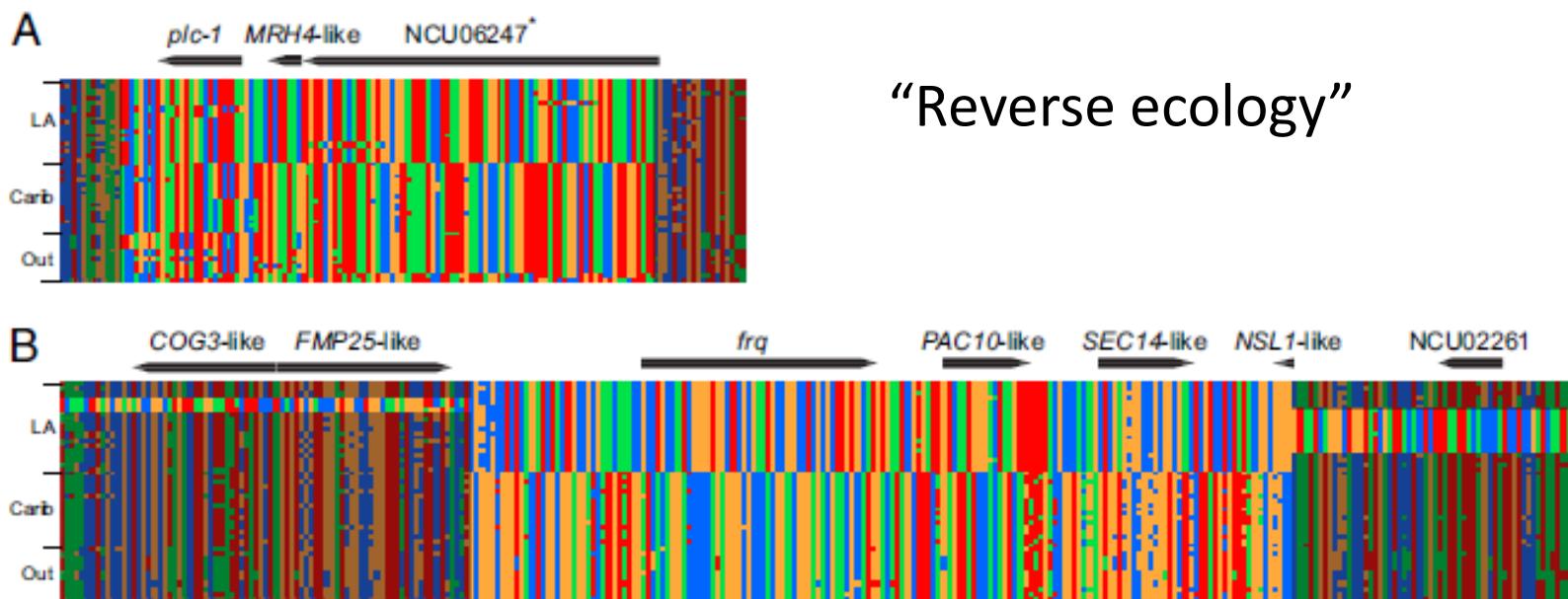
Population genomics and local adaptation in wild isolates of a model microbial eukaryote

Christopher E. Ellison^a, Charles Hall^a, David Kowbel^a, Juliet Welch^a, Rachel B. Brem^b, N. L. Glass^a, and John W. Taylor^{a,1}

Since reads could be mapped to genome, differences across the genome could be analyzed using a sliding window (e.g., F_{ST} , Tajima's D, Dxy).

Two regions were outliers in all three measures.

Contain genes related to temperature and circadian rhythm.



Speciation by genome duplication: Repeated origins and genomic composition of the recently formed allopolyploid species *Mimulus peregrinus*

Mario Vallejo-Marín,^{1,2} Richard J. A. Buggs,³ Arielle M. Cooley,⁴ and Joshua R. Puzey⁵

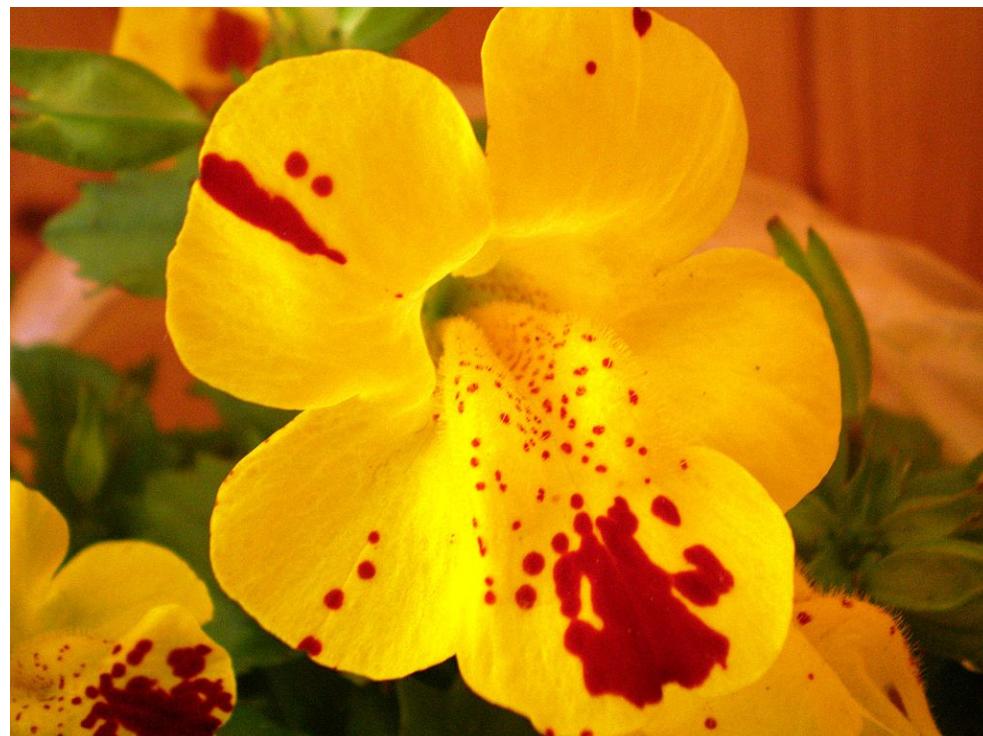
Evolution 69-6: 1487-1500

Examined both parental species,
sterile hybrid, and fertile
allopolyploid

Targeted capture of 1200 loci

4 individuals each

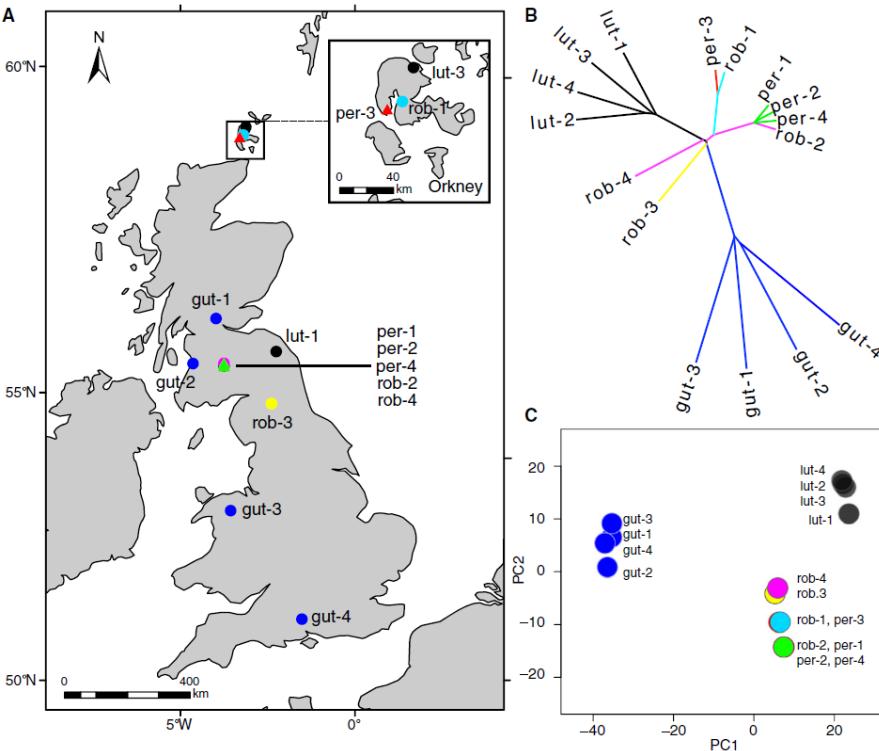
Genome available



"Mimulus001" CC BY-SA 2.5 Hugo.arg

Speciation by genome duplication: Repeated origins and genomic composition of the recently formed allopolyploid species *Mimulus peregrinus*

Mario Vallejo-Marín,^{1,2} Richard J. A. Buggs,³ Arielle M. Cooley,⁴ and Joshua R. Puzey⁵



M. guttatus
→
M. x robertsii → *M. peregrinus*
M. luteus

Targeted capture with Agilent
SureSelect

20,749 SNPs genotyped across 772
loci

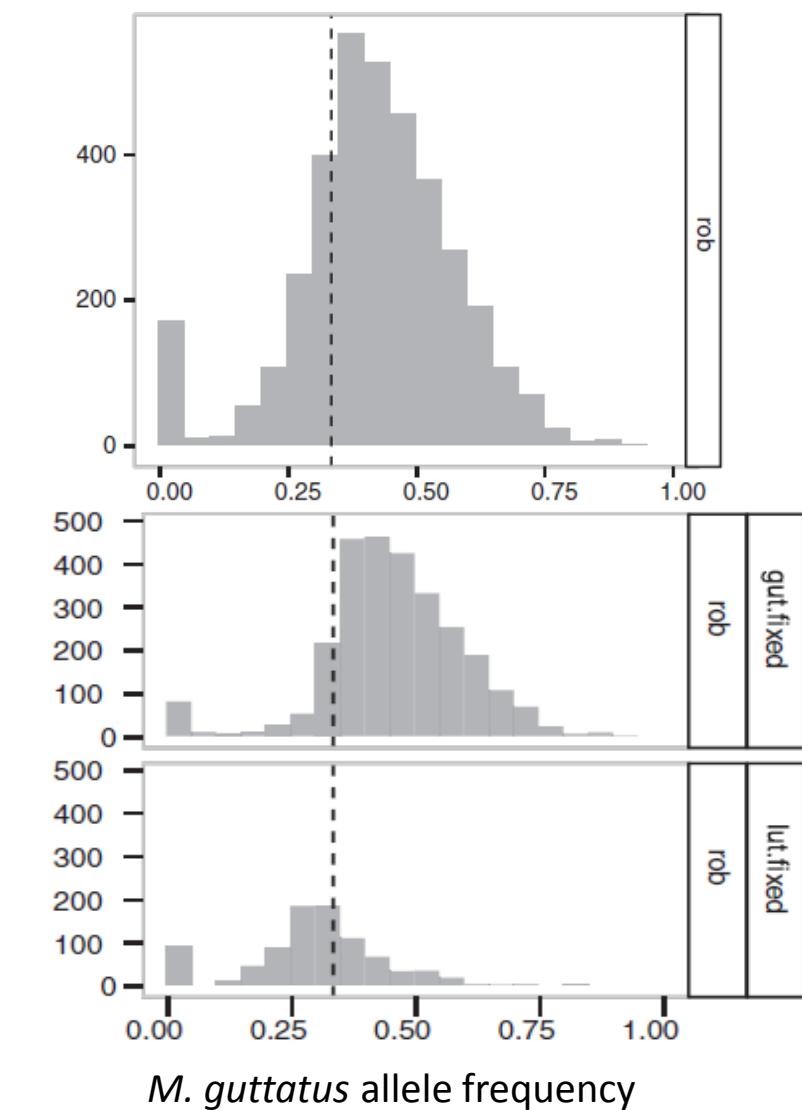
Note:
Genomic resources aren't
perfect:

Only 64% of loci developed from
Mimulus genome v.1 mapped
uniquely in v.2.

Speciation by genome duplication: Repeated origins and genomic composition of the recently formed allopolyploid species

Mimulus peregrinus

Mario Vallejo-Marín,^{1,2} Richard J. A. Buggs,³ Arielle M. Cooley,⁴ and Joshua R. Puzey⁵



Alternative alleles fixed in the parents should be heterozygous in hybrids

Use read depth of each allele as a proxy for dosage in genome

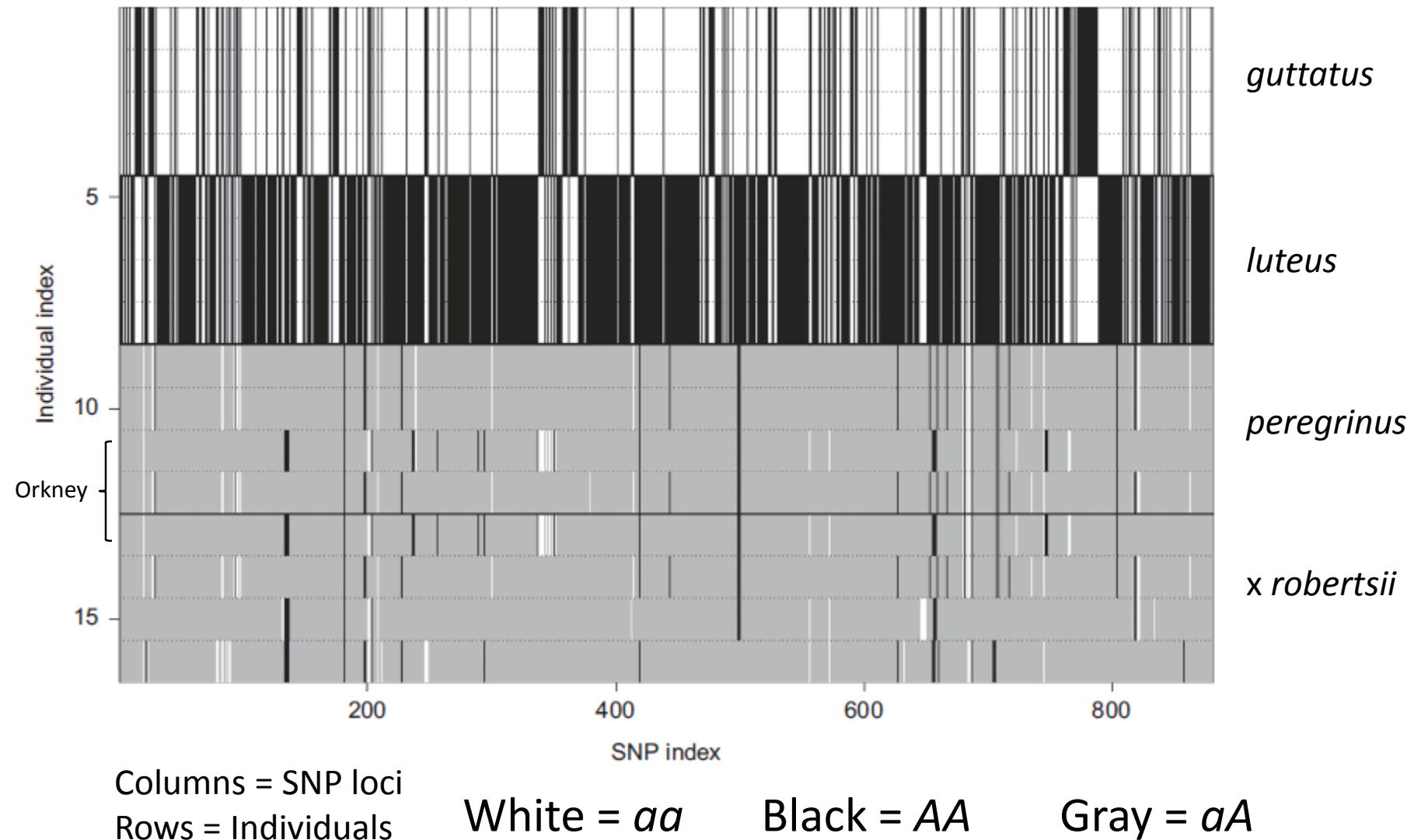
Expect *M. guttatus* to have ~33%

Possible bias of target enrichment...

...or effect of diploidization in some *M. luteus* loci.

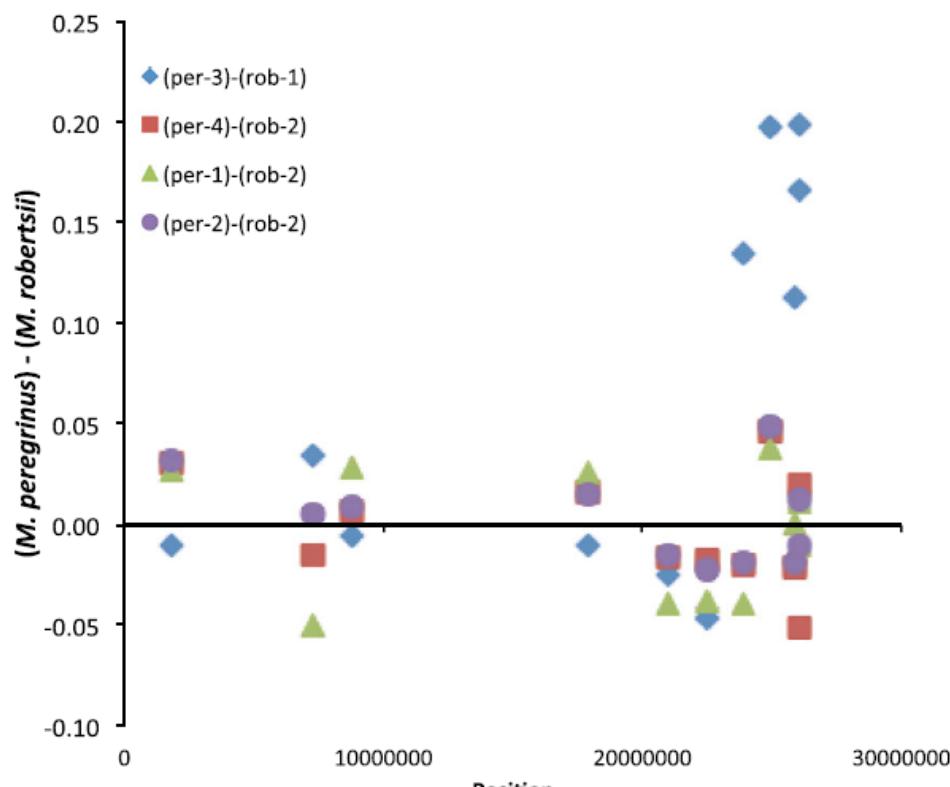
Speciation by genome duplication: Repeated origins and genomic composition of the recently formed allopolyploid species *Mimulus peregrinus*

Mario Vallejo-Marín,^{1,2} Richard J. A. Buggs,³ Arielle M. Cooley,⁴ and Joshua R. Puzey⁵



Speciation by genome duplication: Repeated origins and genomic composition of the recently formed allopolyploid species *Mimulus peregrinus*

Mario Vallejo-Marín,^{1,2} Richard J. A. Buggs,³ Arielle M. Cooley,⁴ and Joshua R. Puzey⁵



Mimulus linkage group 14

Difference in allele frequency between *xrobertsii* and *peregrinus* could indicate change in genome composition due to polyploidization.

DIVERGENCE IS FOCUSED ON FEW GENOMIC REGIONS EARLY IN SPECIATION: INCIPIENT SPECIATION OF SUNFLOWER ECOTYPES

Rose L. Andrew^{1,2} and Loren H. Rieseberg^{1,3}

Evolution 67-9: 2468–2482

Identify genomic differences between dune and non-dune ecotypes of *Helianthus petiolaris*.

120+ individuals

RAD-Seq; 19,000-28,000 SNPs

Genome available, able to map ~60% of SNPs



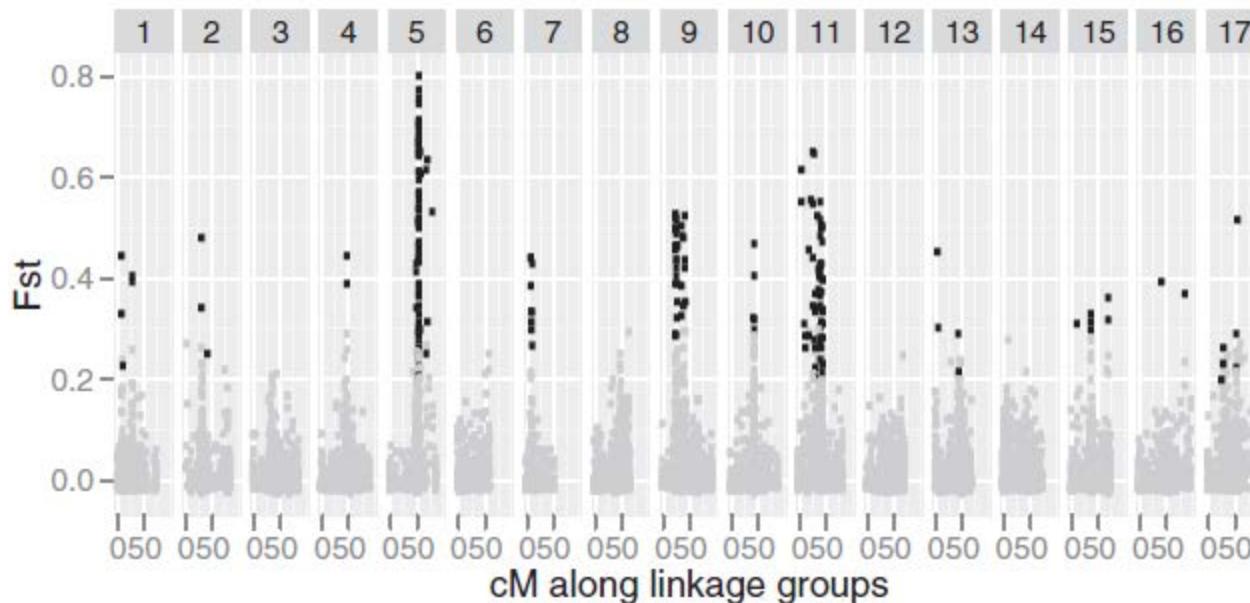
"Sunflower" CC BY-NC flickr Bryant Olsen

DIVERGENCE IS FOCUSED ON FEW GENOMIC REGIONS EARLY IN SPECIATION: INCIPIENT SPECIATION OF SUNFLOWER ECOTYPES

Rose L. Andrew^{1,2} and Loren H. Rieseberg^{1,3}

Evolution 67-9: 2468–2482

Excellent example of applying pop gen statistics across the genome rather than a whole-genome summary.



DIVERGENCE IS FOCUSED ON FEW GENOMIC REGIONS EARLY IN SPECIATION: INCIPIENT SPECIATION OF SUNFLOWER ECOTYPES

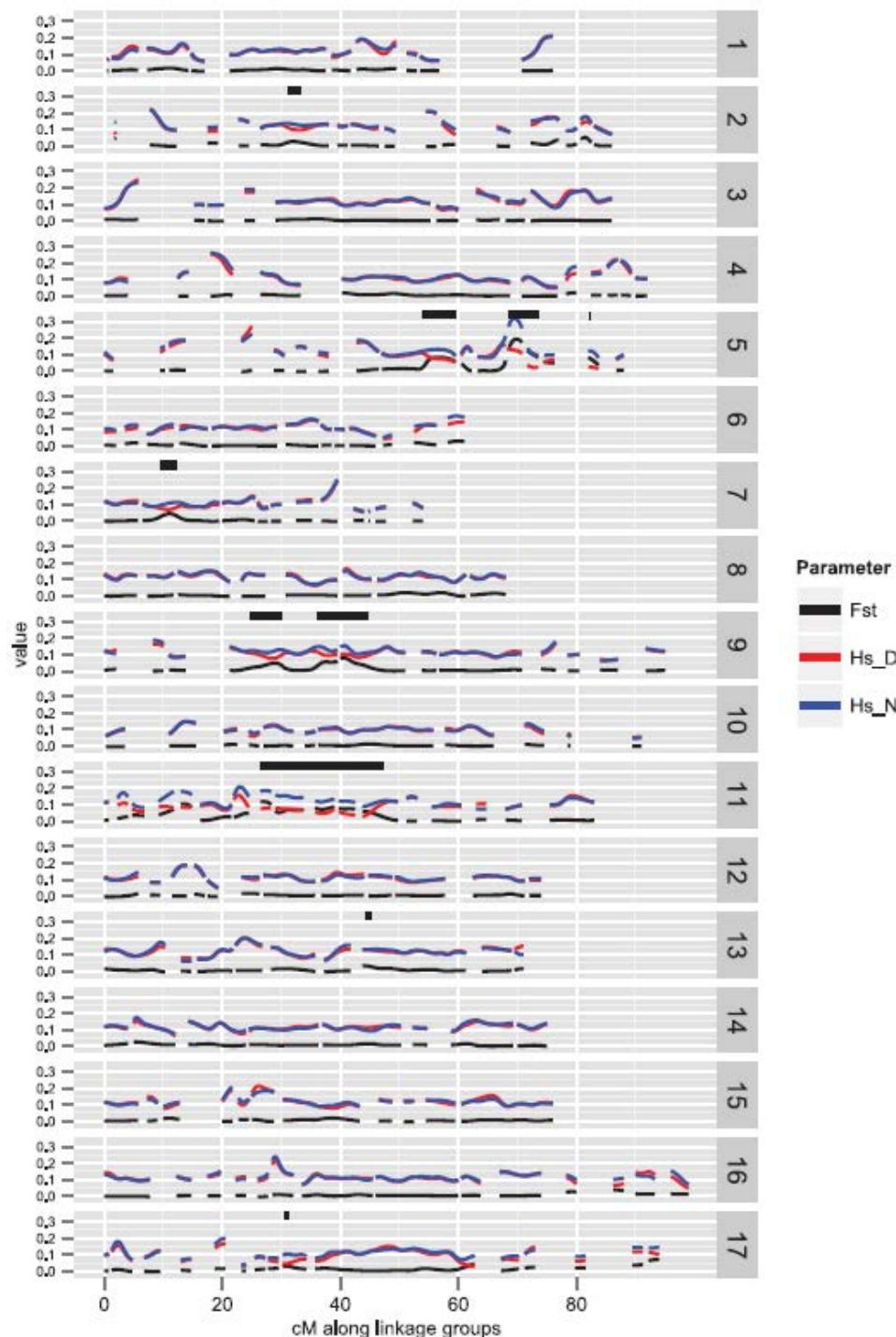
Rose L. Andrew^{1,2} and Loren H. Rieseberg^{1,3}

Evolution 67(9): 2468–2482

Sliding window across linkage groups

Black bars are regions of high F_{ST} and possible “islands of divergence.”

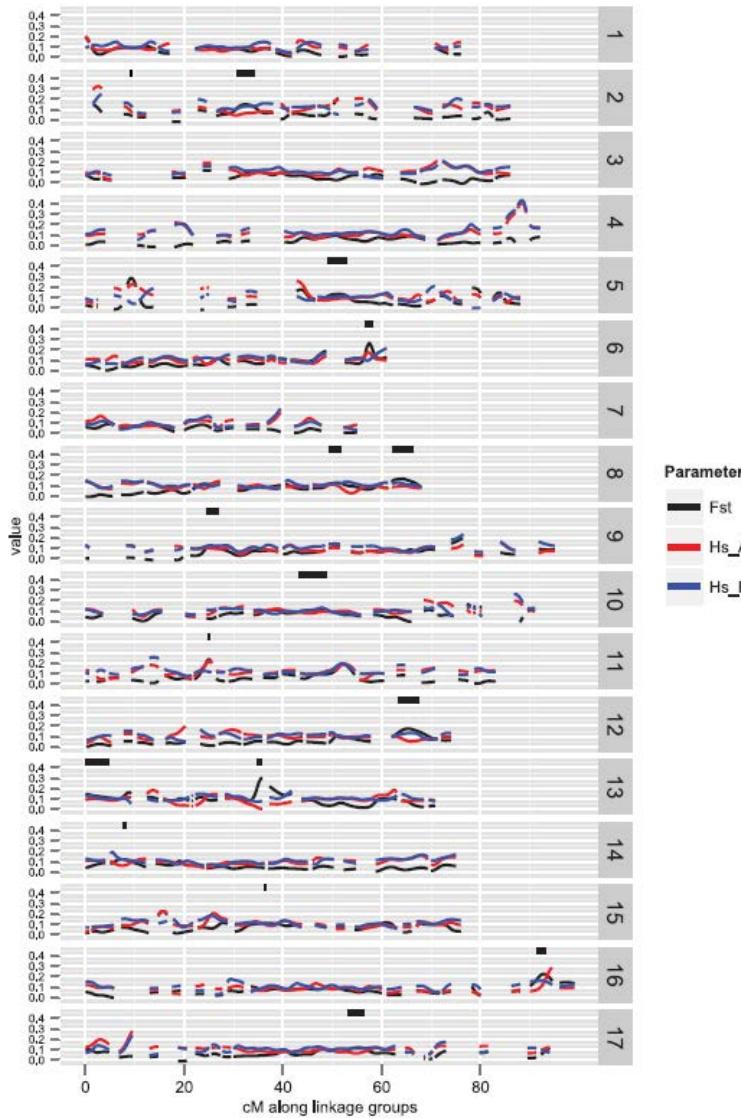
Red and blue lines denote nucleotide diversity.



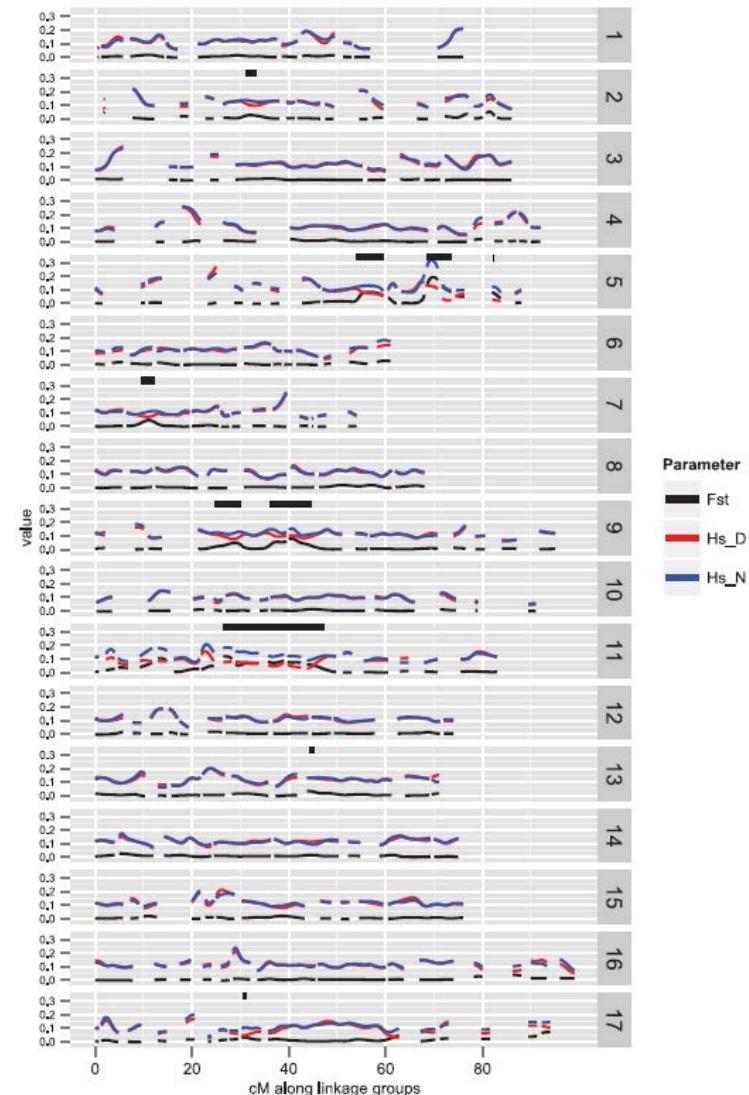
DIVERGENCE IS FOCUSED ON FEW GENOMIC REGIONS EARLY IN SPECIATION: INCIPIENT SPECIATION OF SUNFLOWER ECOTYPES

Rose L. Andrew^{1,2} and Loren H. Rieseberg^{1,3}

Evolution 67(9): 2468–2482



Between species

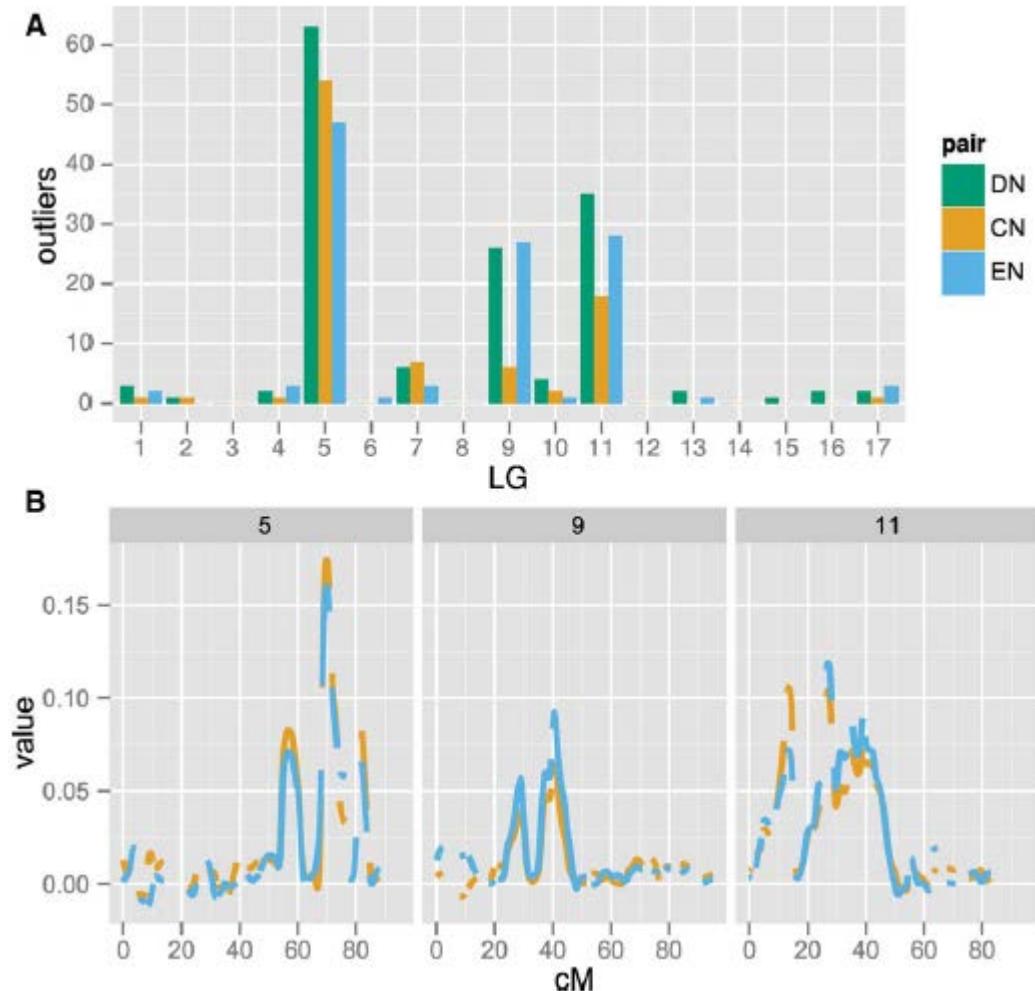


Between ecotypes

DIVERGENCE IS FOCUSED ON FEW GENOMIC REGIONS EARLY IN SPECIATION: INCIPIENT SPECIATION OF SUNFLOWER ECOTYPES

Rose L. Andrew^{1,2} and Loren H. Rieseberg^{1,3}

Evolution 67-9: 2468–2482



Similar measures were made for subsets of populations.

Here, divergence between dune and non-dune ecotypes are split between populations:

- at the edge of the dunes (blue)
- far from the edge (gold)

Introduction to NGS Sequencing Workshop

Botany 2015

8:30 - 8:40 am - Intro to the workshop and presenters

8:40 - 9:05 am - Intro to Sequencing Technology

9:05 - 9:30 am - Sequence Processing/Intro Bioinformatics

9:30 - 10:00 am - Group Activity

10:00 - 10:15 am - Coffee break

10:15 - 11:00 am - Genome Reduction Methods

11:00 - 11:30 am - Applications and data analysis - phylogenomics

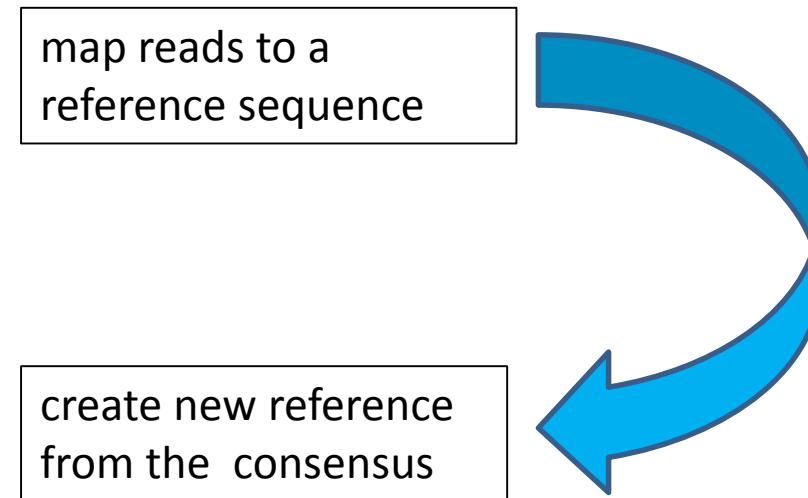
11:30 am - 12:00 pm - Applications and data analysis - population genomics

12:00 - 1:00 pm - Lunch

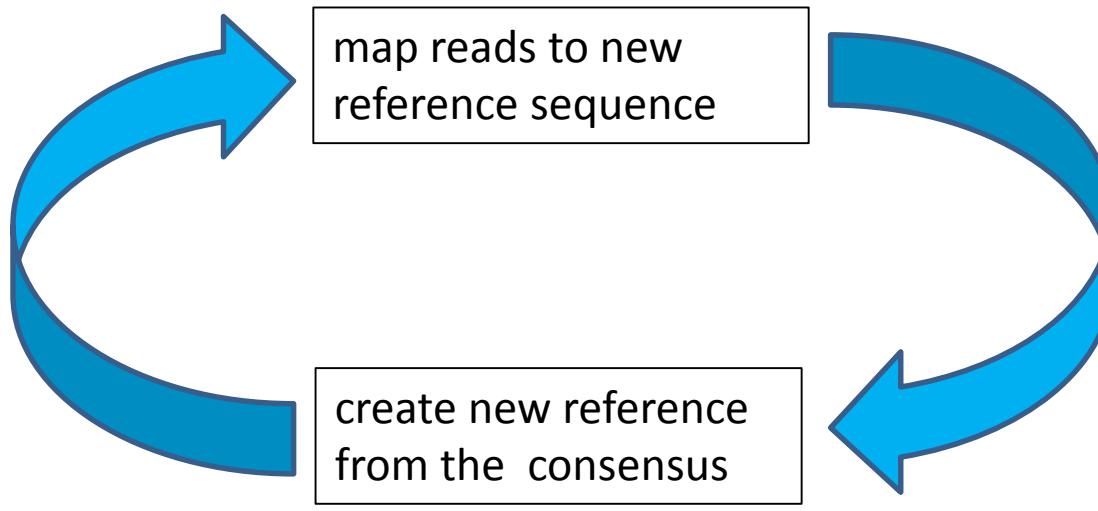
1:00 - 1:20 pm - Introduction to the hands-on exercises

1:20 - 4:30 pm - Hands-on exercises

Reference Guided Assembly



Reference Guided Assembly

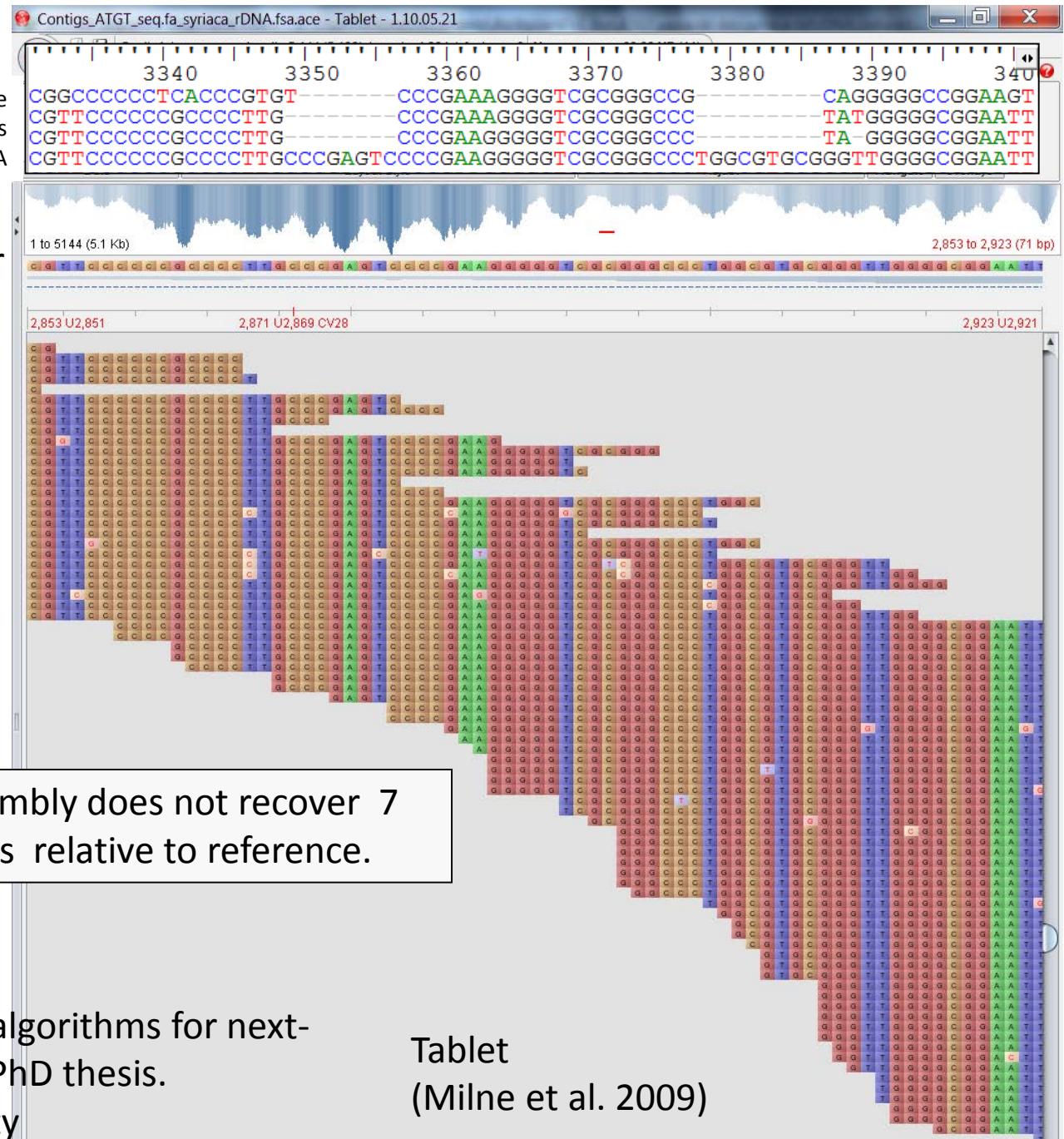


Repeat until no further improvement

Machine learning = A branch of artificial intelligence, is a scientific discipline concerned with the design and development of algorithms that allow computers to evolve behaviors based on empirical data. Wikipedia

yet another short read aligner

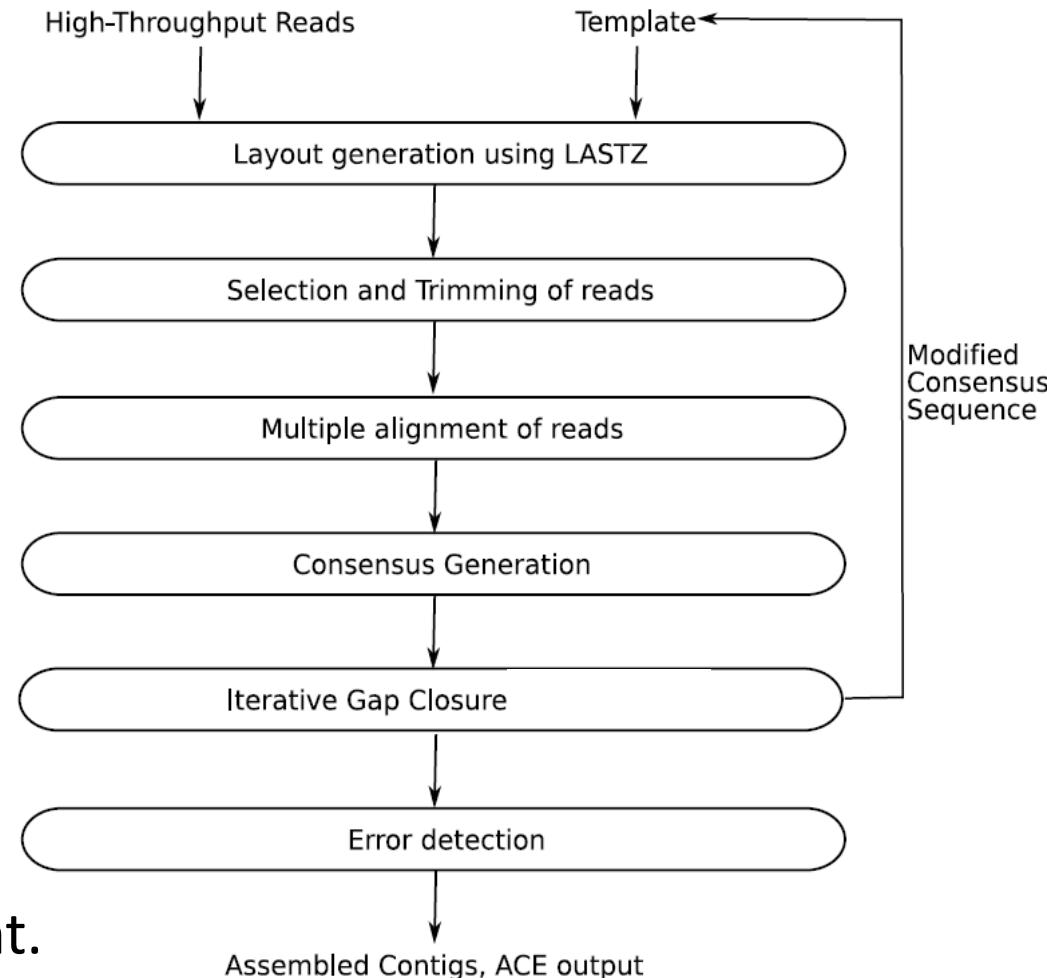
reference
other aligners
YASRA



Reference Guided Assembly

YASRA (Ratan, 2009)

1. Reference can be 80-90% divergent.
2. Map reads to reference followed by de novo assembly of unmapped reads.
3. Closes gaps with overlap-layout consensus.
4. Creates a new reference.
5. Repeats the process until no additional improvement.



Reference Guided Assembly

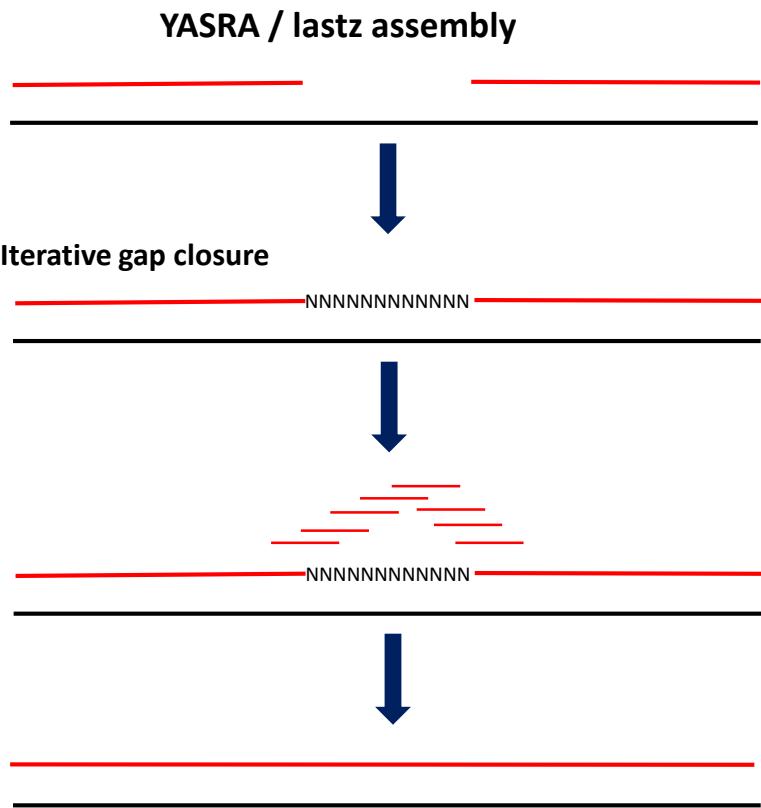
Used for Chloroplast Genome Assembly

AlignReads

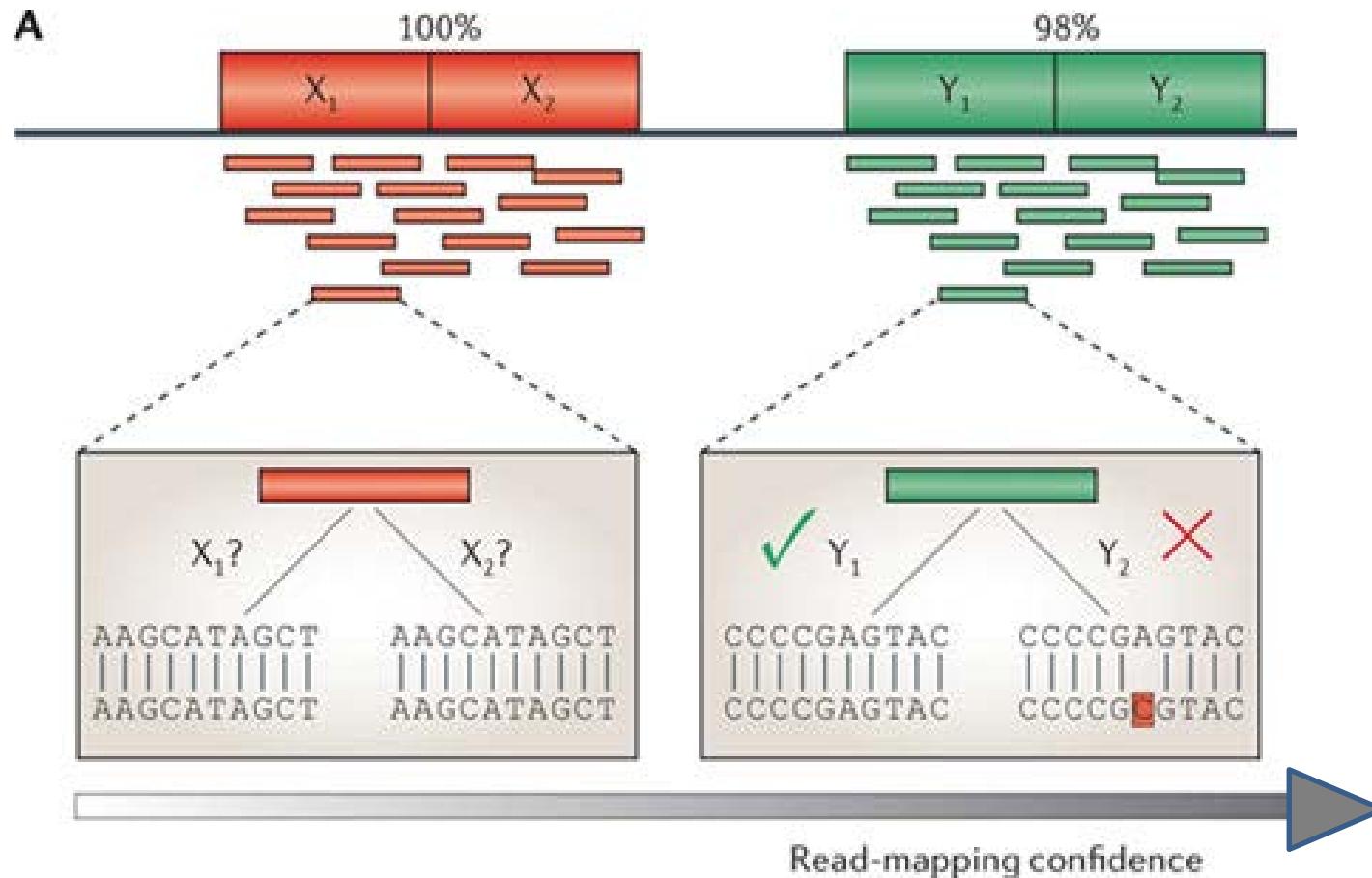
A three-stage automated pipeline for assembly of short reads:

- 1) Iterative Assembly (YASRA, Ratan, 2009)
- 2) Contig Mapping (Mummer, Kurtz et al. 2004)
- 3) Post-processing (scripts written by Zach Foster)

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

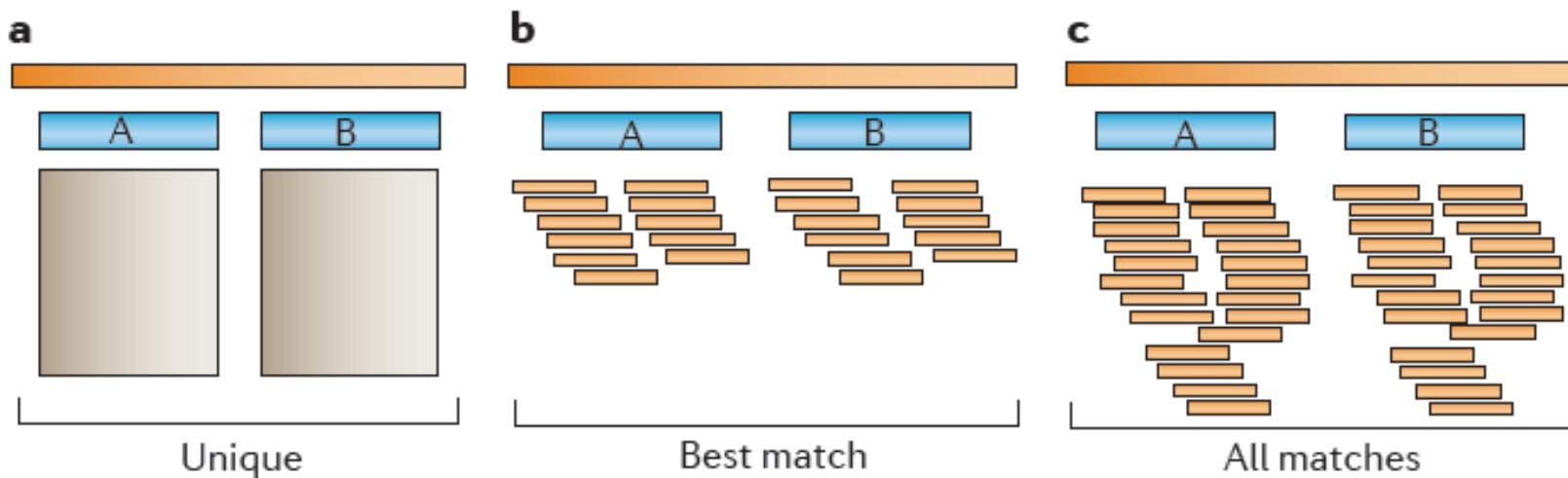


Multi-Reads Map to Identical or Similar Repeats



Treangen & Salzberg 2012. Repetitive DNA and next-generation sequencing: computational challenges and solutions Nature Review Genetics 13:36-46

Multi-Reads Map to Identical or Similar Repeats



- a. only report unique matches (read ignored)
- b. randomly distribute repeat matches (1 per read)
- c. report all repeat matches (many per read)

Treangen & Salzberg 2012. Repetitive DNA and next-generation sequencing: computational challenges and solutions Nature Review Genetics 13:36-46

Introduction to NGS Sequencing Workshop

Botany 2015

8:30 - 8:40 am - Intro to the workshop and presenters

8:40 - 9:05 am - Intro to Sequencing Technology

9:05 - 9:30 am - Sequence Processing/Intro Bioinformatics

9:30 - 10:00 am - Group Activity

10:00 - 10:15 am - Coffee break

10:15 - 11:00 am - Genome Reduction Methods

11:00 - 11:30 am - Applications and data analysis - phylogenomics

11:30 am - 12:00 pm - Applications and data analysis - population genomics

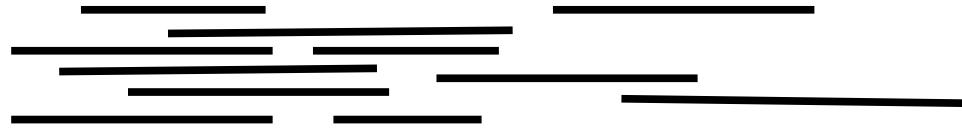
12:00 - 1:00 pm - Lunch

1:00 - 1:20 pm - Introduction to the hands-on exercises

1:20 - 4:30 pm - Hands-on exercises

Introduction to Hyb-Seq

- 1) Probe development
- 2) Target assembly

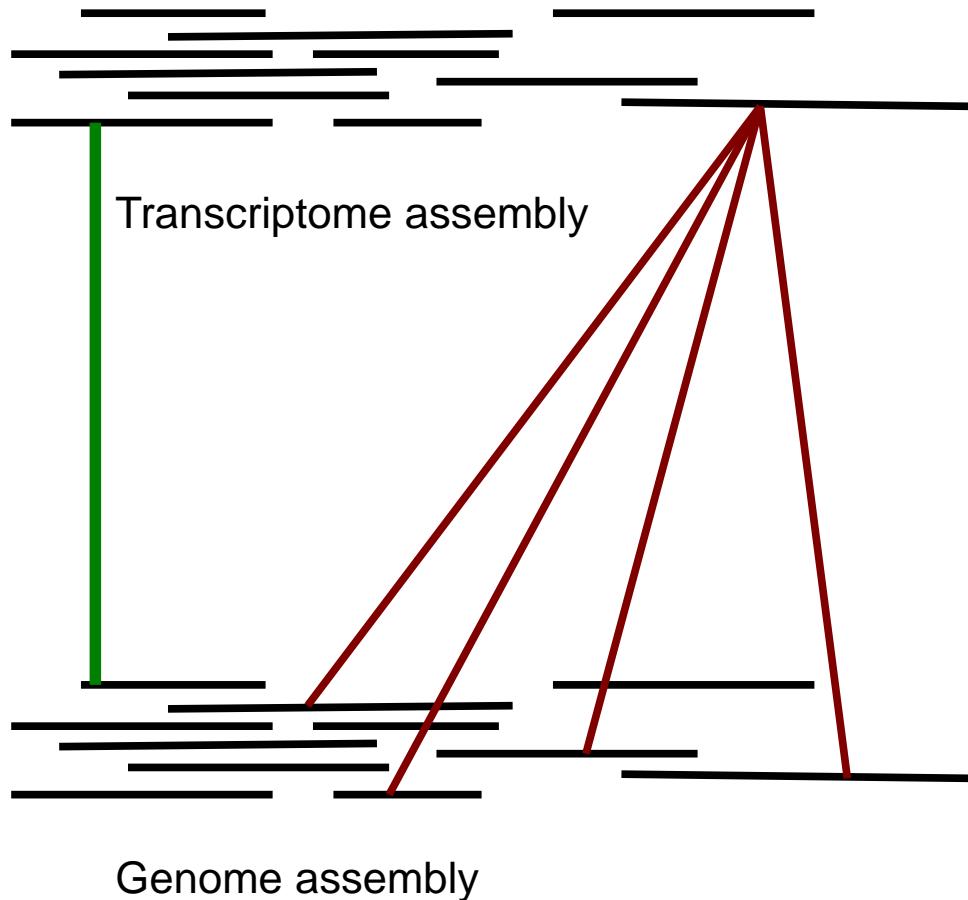


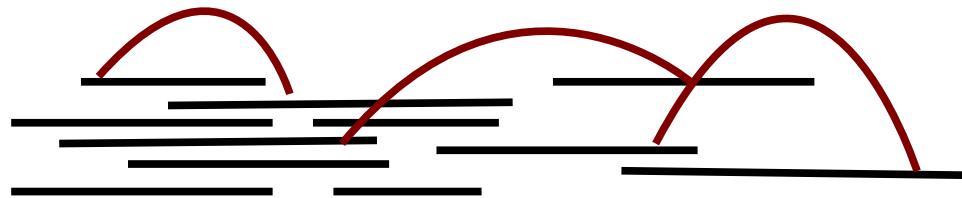
Transcriptome assembly



Genome assembly

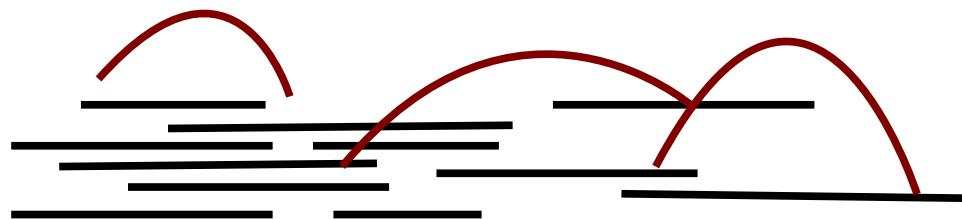
Compare (BLAT, BLAST) genome and transcriptome to each other.
Retain only matches that are unique.





Transcriptome assembly

Compare candidate regions to the rest of the genome/transcriptome. Remove any matching regions.



Genome assembly

Target Assembly

Gene 1

Exon 1



Exon 2



Exon 3



Gene 2

Exon 1



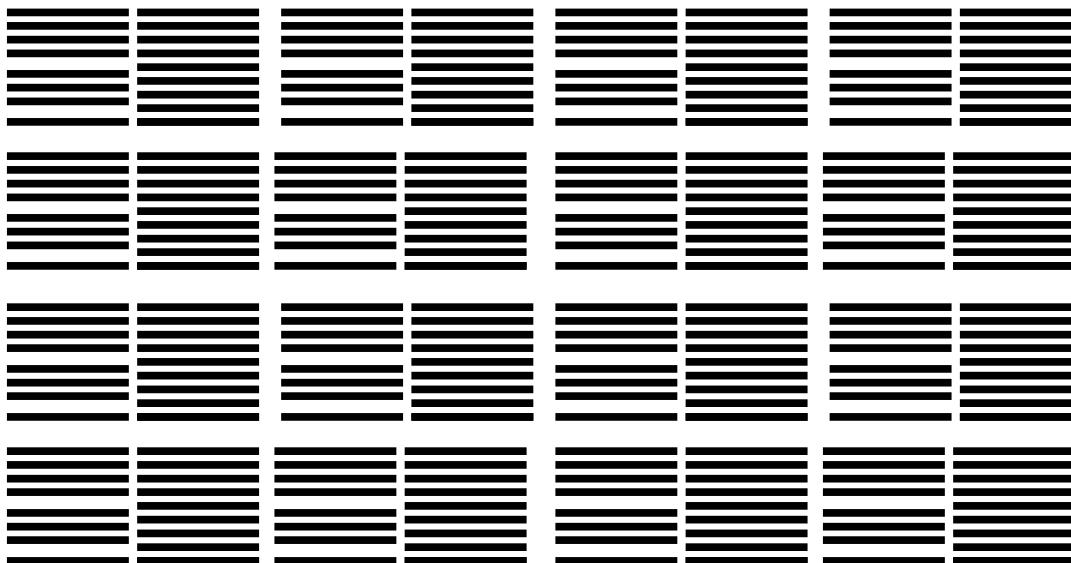
Exon 2



Exon 3



Sequence reads



Weitemier et al, 2014

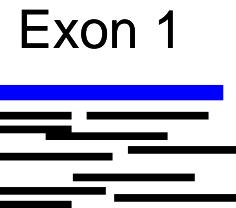
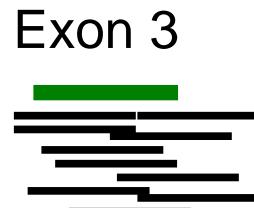
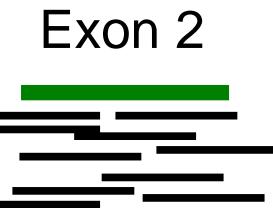
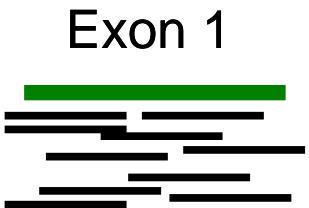
Sandovac:

Schmickl et al, in press

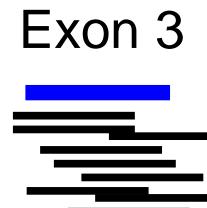
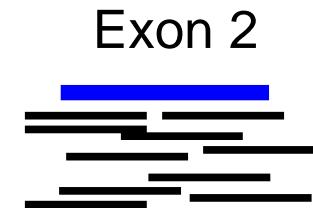
<https://github.com/V-Z/sondovac/wiki>

Map to Reference

Gene 1



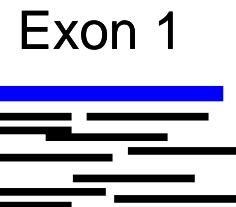
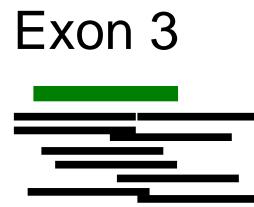
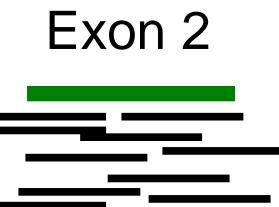
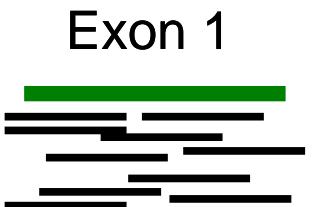
Gene 2



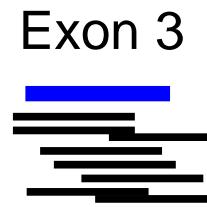
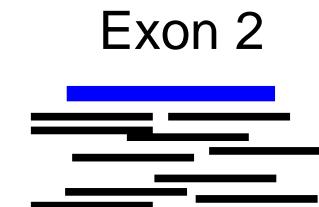
Develop consensus exon for each sample

Map to Reference

Gene 1



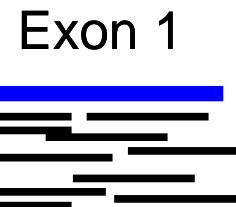
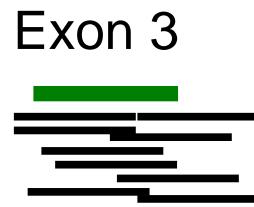
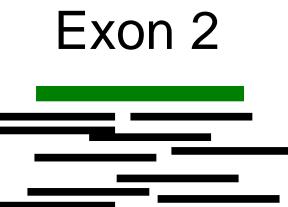
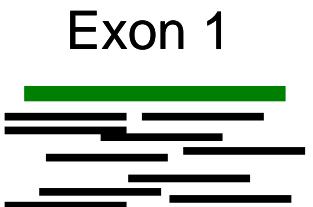
Gene 2



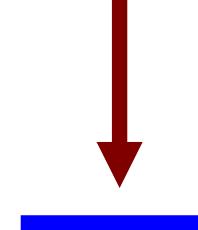
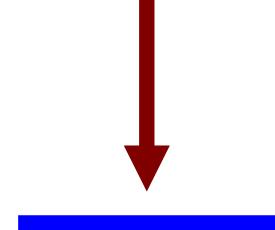
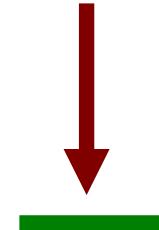
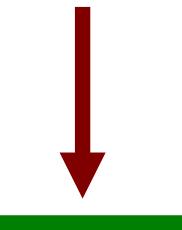
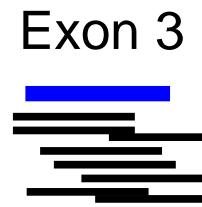
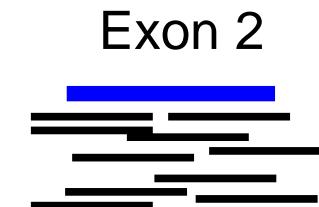
Bin assembled exons together by gene

Map to Reference

Gene 1



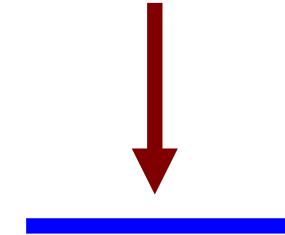
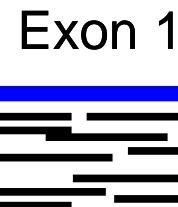
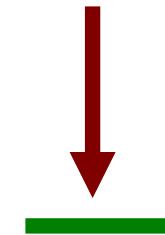
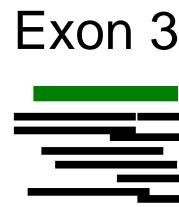
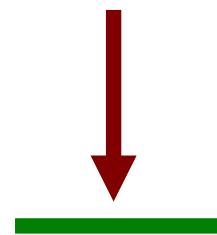
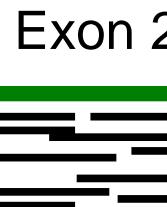
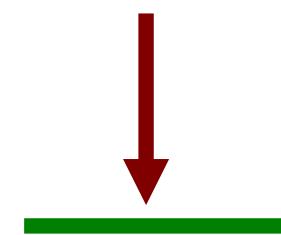
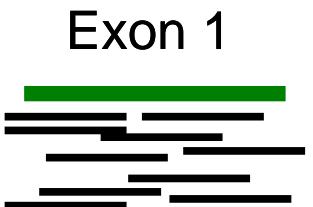
Gene 2



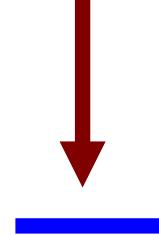
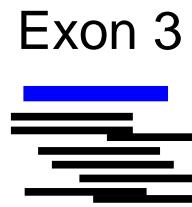
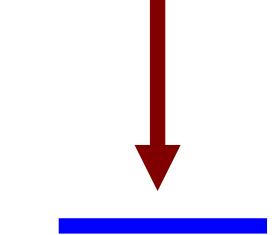
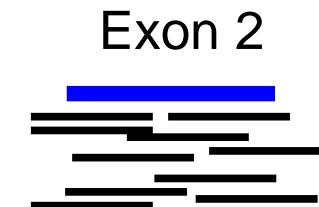
Concatenate exons into a gene.

Map to Reference

Gene 1



Gene 2



Align multiple samples



Introduction to NGS Sequencing Workshop

Botany 2015

8:30 - 8:40 am - Intro to the workshop and presenters

8:40 - 9:05 am - Intro to Sequencing Technology

9:05 - 9:30 am - Sequence Processing/Intro Bioinformatics

9:30 - 10:00 am - Group Activity

10:00 - 10:15 am - Coffee break

10:15 - 11:00 am - Genome Reduction Methods

11:00 - 11:30 am - Applications and data analysis - phylogenomics

11:30 am - 12:00 pm - Applications and data analysis - population genomics

12:00 - 1:00 pm - Lunch

1:00 - 1:20 pm - Introduction to the hands-on exercises

1:20 - 4:30 pm - Hands-on exercises



Next Generation Sequencing Workshop BOTANY 2015

Makenzie Mabry



Applications

- Agriculture, Plant ,and Crop Science
- Ecology and Conservation Genetics
- Genomics
- Microbiology, Virology, and Infectious Diseases
- Population Genetics and Evolutionary Biology
- Synthetic Biology



Why Use Geneious?

Data. Your way.



Drag and drop to convert and import sequence data.



NGS? Sanger? Hybrid? Paired-end?
Barcode? Fragments? Annotations?
Whatever your data, you're covered.



Cross-platform: Mac, Windows, Linux.



Arrange and browse your data library
however you like.

Industry-leading tools, visualized.



Beautiful genome browser, reference
mapping and sequence assembly.



Tree building and viewing without juggling
files.



Streamlined microsatellite genotyping.



Awesome alignment visualization and
editing.



Powerful SNP detection and variant calling.



Why Use Geneious?

Customizable, extendible, awesome.

-  Download and install free plugins written by developers and the community.
-  Add your favourite algorithm, database or visualization to Geneious.
-  Share your program with thousands of Geneious users through our plugin community.
-  Create a top secret plugin that gives your researchers an edge over the competition.

- Phylogenetics
 - RAxML, MrBayes
- Nucleotide Analysis
 - Phobos
- Protein Analysis
 - Coiled Coils
- Services and Searching
 - GenBank Submission
- Alignment
 - MAFFT
- Assembly
 - Velvet



Features

- Sequence Analysis, Annotation, and Prediction
- Alignment
- Assembly and Reference Mapping
- Tree Building and Viewing
- Primer Design
- Workflows
- Molecular Cloning
- Microsatellite Analysis



Geneious

Geneious 8.0.5 – For non-commercial use only

Back Forward Sequence Search Agents Align/Assemble Tree Primers Cloning Back Up Support Help

Search

1 of 104 selected

Sources

- Local (1)
 - Anna Simpson (3)
 - Backbone (0)
 - cpDNA (399)
 - OLD (235)
 - Raw Reads (132)
 - rDNA (399)
 - Everett (162)
 - Kristen (0)
 - Grasses (147, 54 unread)
 - Kristen, Amsinckiae (6)
 - Makenzie Mabry (0)
 - FinalQCReadPool (81)
 - Full Dataset (0)
 - Concat (31)
 - cpDNA (82)
 - cpDNA1_29 (82)
 - cpDNA_consensus_sequences (97)
 - mtDNA (373)
 - rDNA (62)
 - cistron_annotation (1)
 - nrDNA_1_29 (82)
 - nrDNA_Consensus Sequences (104)
 - Genbank (104)
 - ITSData (3)
 - Previous Analysis (294)
 - Reduced Dataset (16)
 - Extracted cpDNA genes (43)
 - extracted mtDNA genes (34)
 - FullTaxa_reducedbp (48)
 - reducedTaxa_fullbp (12)
 - references (2)
 - Ripg data (226)
 - Sample Documents (623)
 - Siekkinen (6907, 59 unread)
 - Deleted Items (3794, 59 unread)
 - Shared Databases
 - Operations
 - NCBI
 - Gene
 - Genome
 - Nucleotide
 - PopSet
 - Protein
 - PubMed
 - SNP
 - Structure
 - Taxonomy
 - UniProt

Name HQ% Description Sequence Len... Ref Seq Name Post-Trim # Sequences & Min Seq... Max Seq... % pairwise id... [Edit](#)

Backup of Nucleotide alignment Copy-EditedByEye (stripped) (modified) ... - Alignment of 81 sequences 6,683 - - 81 5,619 5,628 98.7% Nucleotide alignment (modified) ... - Alignment of 81 sequences 6,683 - - 81 6,384 6,506 97.1% Nucleotide alignment (modified) (stripped) 2 RAXML Bootstrapping Trees ... - Alignment of 81 sequences 5,608 - - 81 5,608 5,608 98.9% Nucleotide alignment (modified) (stripped) 2 RAXML Tree ... - Alignment of 81 sequences 5,608 - - 81 5,608 5,608 98.9% Nucleotide alignment (modified) RaxML Bootstrapping Trees ... - Alignment of 81 sequences 5,608 - - 81 5,608 5,608 98.9% Nucleotide alignment (modified) RaxML Tree ... - Alignment of 81 sequences 6,683 - - 81 6,384 6,506 97.1% Nucleotide alignment Copy-EditedByEye ... - Alignment of 81 sequences 6,440 - - 81 6,323 6,370 97.5% Nucleotide alignment Copy-EditedByEye (stripped) (modified) ... - Alignment of 81 sequences 5,645 - - 81 5,621 5,635 98.7% Nucleotide alignment Copy-EditedByEye (stripped) (modified) Copy-Regio... - Alignment of 81 sequences 5,638 - - 81 5,619 5,628 98.7% Nucleotide alignment Copy-EditedByEye (stripped) (modified) Copy-Regio... - 5,638 - - 81 5,619 5,628 98.7% Nucleotide alignment Copy-EditedByEye (stripped) (modified) Copy-Regio... - 5,638 - - 81 5,619 5,628 98.7% Nucleotide alignment Copy-EditedByEye (stripped) (modified) Copy-Regio... - 5,638 - - 81 5,619 5,628 98.7%

Alignment View Annotations Virtual Gel Distances Text View Info

Consensus Identity

1. *A_intermedia*_SDSU20756 assembled ... ETS 18S rRNA 26S rRNA

2. *C_aaffinis*_SD199070 assembled to ... ETS 18S rRNA 26S rRNA

3. *C_clevelandii*_SDSU20782 assembled ... ETS 18S rRNA 26S rRNA

4. *Pec_penicillata* assembled to Cons... ETS 18S rRNA 26S rRNA

5. *C_alfalfa*_CONC163659 assembled ... ETS 18S rRNA 26S rRNA

6. *C_capitelliflora*_CONC166914 assem... ETS 18S rRNA 26S rRNA

7. *C_gloemerulifera*_CONC166867 as... ETS 18S rRNA 26S rRNA

8. *C_gloemerata*_SGO146941 assem... ETS 18S rRNA 26S rRNA

9. *C_alyssoides*_CONC156553 assem... ETS 18S rRNA 26S rRNA

10. *C_aspera*_MO4317599 assembl... ETS 18S rRNA 26S rRNA

11. *C_kingii*_SGO123832 assembled ... ETS 18S rRNA 26S rRNA

12. *C_gnaphalooides*_SGO146002 as... ETS 18S rRNA 26S rRNA

13. *C_crassispala*_SDSU20623 assem... ETS 18S rRNA 26S rRNA

14. *C_keiseyana*_SDSU20630 assem... ETS 18S rRNA 26S rRNA

15. *C_minima*_SDSU20629 assembled ... ETS 18S rRNA 26S rRNA

16. *C_fendleri*_SDSU20114 assembled ... ETS 18S rRNA 26S rRNA

17. *C_recurvata*_UCR225245 assem... ETS 18S rRNA 26S rRNA

18. *C_clevelandii*_RSA710334 assem... ETS 18S rRNA 26S rRNA

19. *C_diffusa*_MERL56799 assembled ... ETS 18S rRNA 26S rRNA

General

Colors: ACGT - Edit

Graphs Options >

Annotations Options >

Consensus Options >

Highlighting Options >

Complement Options >

Translation Options >

Linear View

Wrap

Show Names

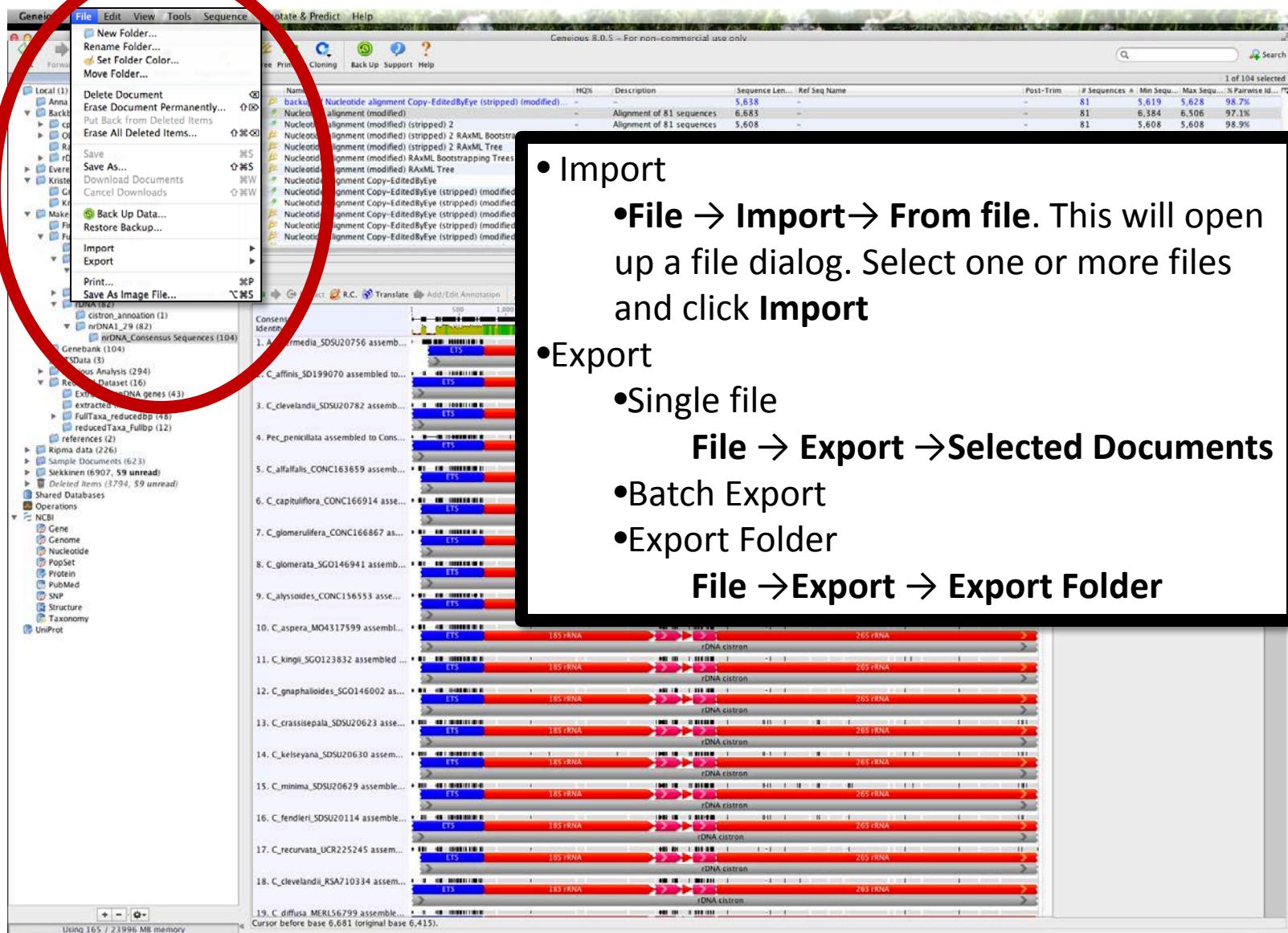
Show Description

Show sequence numbers

100% [Zoom In](#) [Zoom Out](#)

Using 165 / 23996 MB memory

Importing/Exporting Sequences





Document Types

Document type	Geneious Icon
Nucleotide sequence	
Oligo sequences	
Enzyme Sets	
Chromatogram	
Contig	
Protein sequence	
Phylogenetic tree	
3D structure	
Sequence alignment	
Journal articles	
PDF	
Other documents	

Viewing Sequences

Genieous 8.0.5 – For non-commercial use only

Sources

- Local (1)
 - Anna Simpson (3)
 - Backbone (0)
 - cpDNA (399)
 - OLD (235)
 - Raw Reads (132)
 - rRNA (399)
 - Everett (162)
 - Kristen (0)
 - Grasses (147, 54 unread)
 - Kristen, Amsinckiae (6)
 - Makenzie Mabry (0)
 - FinalQCReadPool (81)
 - Full Dataset (0)
 - Concat (31)
 - cpDNA (82)
 - cpDNA1_29 (82)
 - cpDNA_consensus_sequences (97)
 - mtDNA (373)
 - rRNA (62)
 - cistron_annotation (1)
 - nrDNA_1_29 (82)
 - nrDNA_Consensus Sequences (104)
 - Genebank (104)
 - ITSData (3)
 - Previous Analysis (294)
 - Reduced Dataset (16)
 - Extracted cpDNA genes (43)
 - extracted mtDNA genes (34)
 - FullTaxa_reducedbp (48)
 - reducedTaxa_Fullbp (12)
 - references (2)
 - Ripg data (226)
 - Sample Documents (623)
 - Siekkinen (6907, 59 unread)
 - Deleted Items (3794, 59 unread)
 - Shared Databases
 - Operations
 - NCBI
 - Gene
 - Genome
 - Nucleotide
 - PopSet
 - Protein
 - PubMed
 - SNP
 - Structure
 - Taxonomy
 - UniProt

Table View

Name	HQ%	Description	Sequence Len...	Ref Seq Name	Post-Trim	# Sequences	Min Sequ...	Max Sequ...	% Pairwise id...
backup of Nucleotide alignment Copy-EditedByEye (stripped) (modified) ...	-	Alignment of 81 sequences	6,683	-	-	81	6,384	6,506	97.1%
Nucleotide alignment (modified)	-	Alignment of 81 sequences	5,608	-	-	81	5,608	5,608	98.9%
Nucleotide alignment (modified) (stripped) 2 RAXML Bootstrapping Trees	-	Alignment of 81 sequences	5,608	-	-	81	5,608	5,608	98.9%
Nucleotide alignment (modified) (stripped) 2 RAXML Tree	-	Alignment of 81 sequences	5,608	-	-	81	5,608	5,608	98.9%
Nucleotide alignment (modified) (stripped) RaxML Bootstrapping Trees	-	Alignment of 81 sequences	6,683	-	-	81	6,384	6,506	97.1%
Nucleotide alignment (modified) RaxML Tree	-	Alignment of 81 sequences	6,683	-	-	81	6,384	6,506	97.1%
Nucleotide alignment Copy-EditedByEye	-	Alignment of 81 sequences	6,440	-	-	81	6,323	6,370	97.5%
Nucleotide alignment Copy-EditedByEye (stripped) (modified)	-	Alignment of 81 sequences	5,645	-	-	81	5,621	5,635	98.7%
Nucleotide alignment Copy-EditedByEye (stripped) (modified) Copy-Regio...	-	Alignment of 81 sequences	5,638	-	-	81	5,619	5,628	98.7%
Nucleotide alignment Copy-EditedByEye (stripped) (modified) Copy-Regio...	-	Alignment of 81 sequences	5,638	-	-	81	5,619	5,628	98.7%
Nucleotide alignment Copy-EditedByEye (stripped) (modified) Copy-Regio...	-	Alignment of 81 sequences	5,638	-	-	81	5,619	5,628	98.7%
Nucleotide alignment Copy-EditedByEye (stripped) (modified) Copy-Regio...	-	Alignment of 81 sequences	5,638	-	-	81	5,619	5,628	98.7%
1 of 104 selected									

Sequence View

Consensus Identity

Annotations

Virtual Gel

Distances

Text View

Info

Colors: ACGT - Edit

General

Graphs Options >

Annotations Options >

Consensus Options >

Highlighting Options >

Complement Options >

Translation Options >

Linear View

Wrap

Show Names

Show Description

Show sequence numbers

Cursor before base 6,681 (original base 6,415).

Viewing Sequences

Genieous 8.0.5 – For non-commercial use only

1 of 104 selected

Sequence Search Agents Align/Assemble Tree Primers Cloning Back Up Support Help

Source

Alignment View Annotations Virtual Gel Distances Text View Info

General

Colors: ACGT - Edit

Graphs Options >

Annotations Options >

Consensus Options >

Highlighting Options >

Complement Options >

Translation Options >

Linear View

Wrap

Show Names

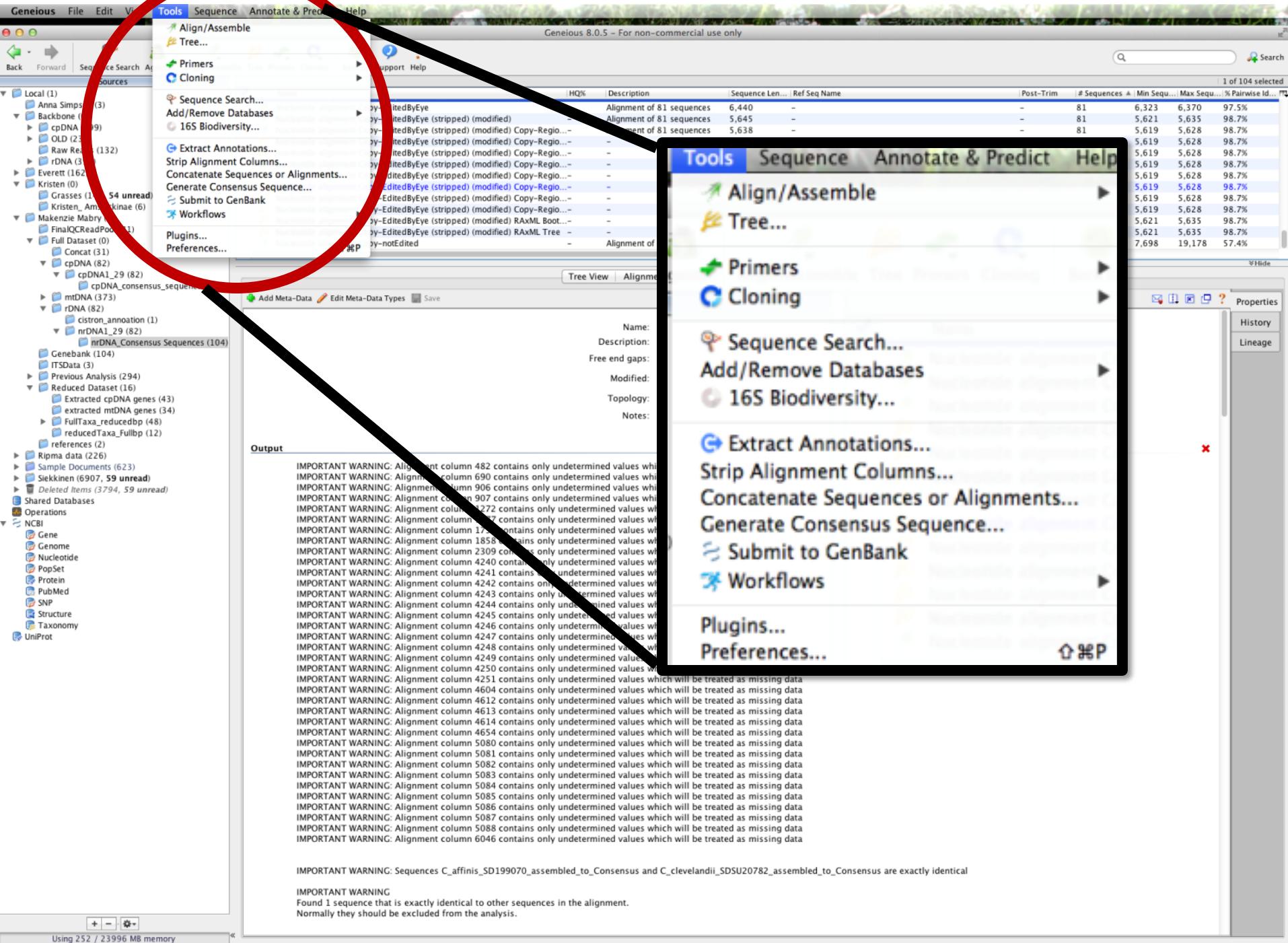
Show Description

Show sequence numbers

Cursor before base 6,681 (original base 6,415).

Using 165 / 23996 MB memory

The screenshot shows the Genieous 8.0.5 software interface for viewing and analyzing DNA sequences. The main window displays a list of projects on the left and a detailed sequence viewer on the right. The sequence viewer shows 19 rRNA genes from various species, each with its 5' and 3' ends labeled (ETS and 26S rRNA), and internal rRNA cistrons. The viewer includes a color-coded consensus bar at the top and various analysis tools like Extract, R.C., Translate, and Add/Edit Annotation along the top toolbar.



Viewing Sequences

The figure shows the Geneious 8.0.5 software interface. The left sidebar displays a hierarchical tree of projects and databases. The main workspace shows a sequence alignment viewer with 19 entries, each representing a different species assembly. Each entry consists of a blue ETS region, a red 18S rRNA region, and a pink 26S rRNA region. The right side features a detailed phylogenetic tree with various nodes highlighted in different colors (blue, green, yellow, orange) and labeled with taxon names like "C. diffusa MERL56799 assembled...". A status bar at the bottom indicates "Using 165 / 23996 MB memory".

Viewing Sequences

Genieous 8.0.5 – For non-commercial use only

Sources

- Local (1)
 - Anna Simpson (3)
 - Backbone (0)
 - cpDNA (399)
 - OLD (235)
 - Raw Reads (132)
 - rRNA (399)
 - Everett (162)
 - Kristen (0)
 - Grasses (147, 54 unread)
 - Kristen, Amsinckiae (6)
 - Makenzie Mabry (0)
 - FinalQCReadPool (81)
 - Full Dataset (0)
 - Concat (31)
 - cpDNA (82)
 - cpDNA1_29 (82)
 - cpDNA_consensus_sequences (97)
 - mtDNA (373)
 - rRNA (62)
 - cistron_annotation (1)
 - nrDNA_1_29 (82)
 - nrDNA_consensus Sequences (104)
 - Genebank (104)
 - ITSData (3)
 - Previous Analysis (294)
 - Reduced Dataset (16)
 - Extracted cpDNA genes (43)
 - extracted mtDNA genes (34)
 - FullTaxa_reducedbp (48)
 - reducedTaxa_fullbp (12)
 - references (2)
 - Ripg data (226)
 - Sample Documents (623)
 - Siekkinen (6907, 59 unread)
 - Deleted Items (3794, 59 unread)
 - Shared Databases
 - Operations
 - NCBI
 - Gene
 - Genome
 - Nucleotide
 - PopSet
 - Protein
 - PubMed
 - SNP
 - Structure
 - Taxonomy
 - UniProt

Annotations

Name	HQ%	Description	Sequence Len...	Ref Seq Name	Post-Trim	# Sequences	Avg Min Seq...	Max Seq...	% Pairwise Id...
backup of Nucleotide alignment Copy-EditedByEye (stripped) (modified)	-	Alignment of 81 sequences	6,683	-	-	81	6,384	6,506	97.1%
Nucleotide alignment (modified) (stripped) 2	-	Alignment of 81 sequences	5,608	-	-	81	5,608	5,608	98.9%
Nucleotide alignment (modified) (stripped) 2 RAxML Bootstrapping Trees	-	-	5,608	-	-	81	5,608	5,608	98.9%
Nucleotide alignment (modified) (stripped) 2 RAxML Tree	-	-	5,608	-	-	81	5,608	5,608	98.9%
Nucleotide alignment (modified) RAxML Tree	-	-	6,683	-	-	81	6,384	6,506	97.1%
Nucleotide alignment Copy-EditedByEye	-	Alignment of 81 sequences	6,440	-	-	81	6,323	6,370	97.5%
Nucleotide alignment Copy-EditedByEye (modified)	-	Alignment of 81 sequences	5,645	-	-	81	5,621	5,635	98.7%
Nucleotide alignment Copy-EditedByEye (stripped) (modified) Copy-Regio...	-	Alignment of 81 sequences	5,638	-	-	81	5,619	5,628	98.7%
Nucleotide alignment Copy-EditedByEye (stripped) (modified) Copy-Regio...	-	-	5,638	-	-	81	5,619	5,628	98.7%
Nucleotide alignment Copy-EditedByEye (stripped) (modified) Copy-Regio...	-	-	5,638	-	-	81	5,619	5,628	98.7%
Nucleotide alignment Copy-EditedByEye (stripped) (modified) Copy-Regio...	-	-	5,638	-	-	81	5,619	5,628	98.7%

Virtual Gel

Distances

Text View

Info

Consensus Identity

1. *A_intermedia_SD5U20756 assembled* rRNA cistron

2. *C_aaffinis_SD199070 assembled* rRNA cistron

3. *C_clevelandii_SD5U20782 assembled* rRNA cistron

4. *Pec_pencillata assembled to Consensus* rRNA cistron

5. *C_alfalfa CONC163659 assembled* rRNA cistron

6. *C_capitelliflora CONC166914 assembled* rRNA cistron

7. *C_gloemerulifera CONC166867 assembled* rRNA cistron

8. *C_gloemerata_SGO146941 assembled* rRNA cistron

9. *C_alyssoides CONC156553 assembled* rRNA cistron

10. *C_aspera MO4317599 assembled* rRNA cistron

11. *C_kingii SGO123832 assembled* rRNA cistron

12. *C_gnaphaloides SGO146002 assembled* rRNA cistron

13. *C_crassispala SD5U20623 assembled* rRNA cistron

14. *C_keiseyana SD5U20630 assembled* rRNA cistron

15. *C_minima SD5U20629 assembled* rRNA cistron

16. *C_fendleri SD5U20114 assembled* rRNA cistron

17. *C_recurvata UCR225245 assembled* rRNA cistron

18. *C_clevelandii RSA710334 assembled* rRNA cistron

19. *C_diffusa MERL56799 assembled* rRNA cistron

General

Colors: ACGT - Edit

Graphs Options >

Annotations Options >

Consensus Options >

Highlighting Options >

Complement Options >

Translation Options >

Linear View

Wrap

Show Names

Show Description

Show sequence numbers

Viewing Sequences

Genieous 8.0.5 – For non-commercial use only

Sources

- Local (1)
 - Anna Simpson (3)
 - Backbone (0)
 - cpDNA (399)
 - OLD (235)
 - Raw Reads (132)
 - rDNA (399)
 - Everett (162)
 - Kristen (0)
 - Grasses (147, 54 unread)
 - Kristen, Amsinckiae (6)
 - Makenzie Mabry (0)
 - FinalQCReadPool (81)
 - Full Dataset (0)
 - Concat (31)
 - cpDNA (82)
 - cpDNA1_29 (82)
 - cpDNA_consensus_sequences
 - mtDNA (373)
 - rDNA (62)
 - cistron_annotation (1)
 - nrDNA_1_29 (82)
 - nrDNA_consensus Sequences (1)
 - Genebank (104)
 - ITSData (3)
 - Previous Analysis (294)
 - Reduced Dataset (16)
 - Extracted cpDNA genes (43)
 - extracted mtDNA genes (34)
 - FullTaxa_reducedbp (48)
 - reducedTaxa_Fullbp (12)
 - references (2)
 - Ripon data (226)
 - Sample Documents (623)
 - Siekkinen (6907, 59 unread)
 - Deleted Items (3794, 59 unread)
 - Shared Databases
 - Operations
 - NCBI
 - Gene
 - Genome
 - Nucleotide
 - PopSet
 - Protein
 - PubMed
 - SNP
 - Structure
 - Taxonomy
 - UniProt

Table View

Name	HQ%	Description	Sequence Len...	Ref Seq Name	Post-Trim	# Sequences	Min Seq...	Max Seq...	% Pairwise id...
backup of Nucleotide alignment Copy-EditedByEye (stripped) (modified)...	-	-	5,638	-	-	81	5,619	5,628	98.7%
Nucleotide alignment (modified)	-	Alignment of 81 sequences	6,683	-	-	81	6,384	6,506	97.1%
Nucleotide alignment (modified) (stripped) 2	-	Alignment of 81 sequences	5,608	-	-	81	5,608	5,608	98.9%
Nucleotide alignment (modified) (stripped) 2 RAXML Bootstrapping Trees	-	-	5,608	-	-	81	5,608	5,608	98.9%
Nucleotide alignment (modified) (stripped) 2 RAXML Tree	-	-	5,608	-	-	81	5,608	5,608	98.9%
Nucleotide alignment (modified) (stripped) RaxML Bootstrapping Trees	-	-	6,683	-	-	81	6,384	6,506	97.1%
Nucleotide alignment (modified) RaxML Tree	-	-	6,683	-	-	81	6,384	6,506	97.1%
Nucleotide alignment Copy-EditedByEye	-	Alignment of 81 sequences	6,440	-	-	81	6,323	6,370	97.5%
Nucleotide alignment Copy-EditedByEye (stripped) (modified)	-	Alignment of 81 sequences	5,645	-	-	81	5,621	5,635	98.7%
Nucleotide alignment Copy-EditedByEye (stripped) (modified) Copy-Regio...	-	Alignment of 81 sequences	5,638	-	-	81	5,619	5,628	98.7%
Nucleotide alignment Copy-EditedByEye (stripped) (modified) Copy-Regio...	-	-	5,638	-	-	81	5,619	5,628	98.7%
Nucleotide alignment Copy-EditedByEye (stripped) (modified) Copy-Regio...	-	-	5,638	-	-	81	5,619	5,628	98.7%
Nucleotide alignment Copy-EditedByEye (stripped) (modified) Copy-Regio...	-	-	5,638	-	-	81	5,619	5,628	98.7%
Nucleotide alignment Copy-EditedByEye (stripped) (modified) Copy-Regio...	-	-	5,638	-	-	81	5,619	5,628	98.7%

Sequence View

Alignment View Annotations Virtual Gel Distances Text View Info

Consensus Identity

1. *A_intermedia*_SDSU20756 assembled ...
2. *C_aaffinis*_SD199070 assembled to ...
3. *C_clevelandii*_SDSU20782 assembled ...
4. *Pec_penicillata* assembled to Cons...
5. *C_alfalfa*_CONC163659 assembled ...
6. *C_capitelliflora*_CONC166914 assem...
7. *C_gloemerulifera*_CONC166867 as...
8. *C_gloemerata*_SGO146941 assem...
9. *C_alyssoides*_CONC156553 assem...
10. *C_aspera*_MO4317599 assembl...
11. *C_kingii*_SGO123832 assembled ...
12. *C_gnaphalooides*_SGO146002 as...
13. *C_crassispela*_SDSU20623 assem...
14. *C_keiskeyanana*_SDSU20630 assem...
15. *C_minima*_SDSU20629 assembled ...
16. *C_fendleri*_SDSU20114 assembled ...
17. *C_recurvata*_UCR225245 assem...
18. *C_clevelandii*_RSA710334 assem...
19. *C_diffusa*_MERL56799 assembled ...

General

Colors: ACGT - Edit

Graphs Options >

Annotations Options >

Consensus Options >

Highlighting Options >

Complement Options >

Translation Options >

Linear View

Wrap

Show Names

Show Description

Show sequence numbers

Viewing Sequences

Geneious 8.0.5 – For non-commercial use only

1 of 104 selected

Sources

Name HQ% Description Sequence Len... Ref Seq Name Post-Trim # Sequences Min Sequ... Max Sequ... % Pairwise Id...

Local (1)

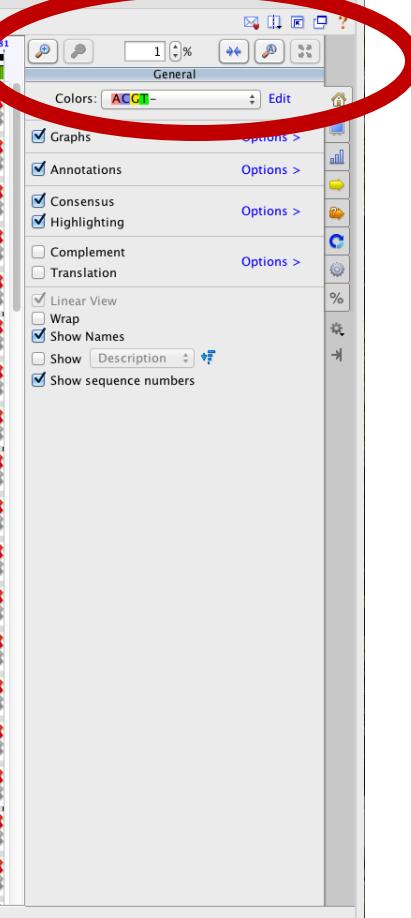
- Anna Simpson (3)
- Backbone (0)
- cpDNA (399)
- OLD (235)
- Raw Reads (132)
- rDNA (399)
- Grasses (147, 54 unread)
- Kristen (0)
- Kristen_Amsincinae (6)
- Mackenzie Mabry (0)
- FinalQCReadPool (81)
- Full Dataset (0)
 - Concat (31)
 - cpDNA (82)
 - cpDNA_1_29 (82)
 - mtDNA
 - rDNA (cist...
 - cist...
 - nrD...
- Genbank
- ITSDATA (3)
- Previous A...
- Reduced D...
- Extract...
- extract...
- FullTax...
- reduce...
- references...
- Ripma data (2)
- Sample Docu...
- Siekkinen (690)
- Deleted Items
- Shared Databases
- Operations
- NCBI
 - Gene
 - Genome
 - Nucleotide
 - PopSet
 - Protein
 - PubMed
 - SNP
 - Structure
 - Taxonomy
 - UniProt

Zooms in to fit the selected region in the available viewing area.

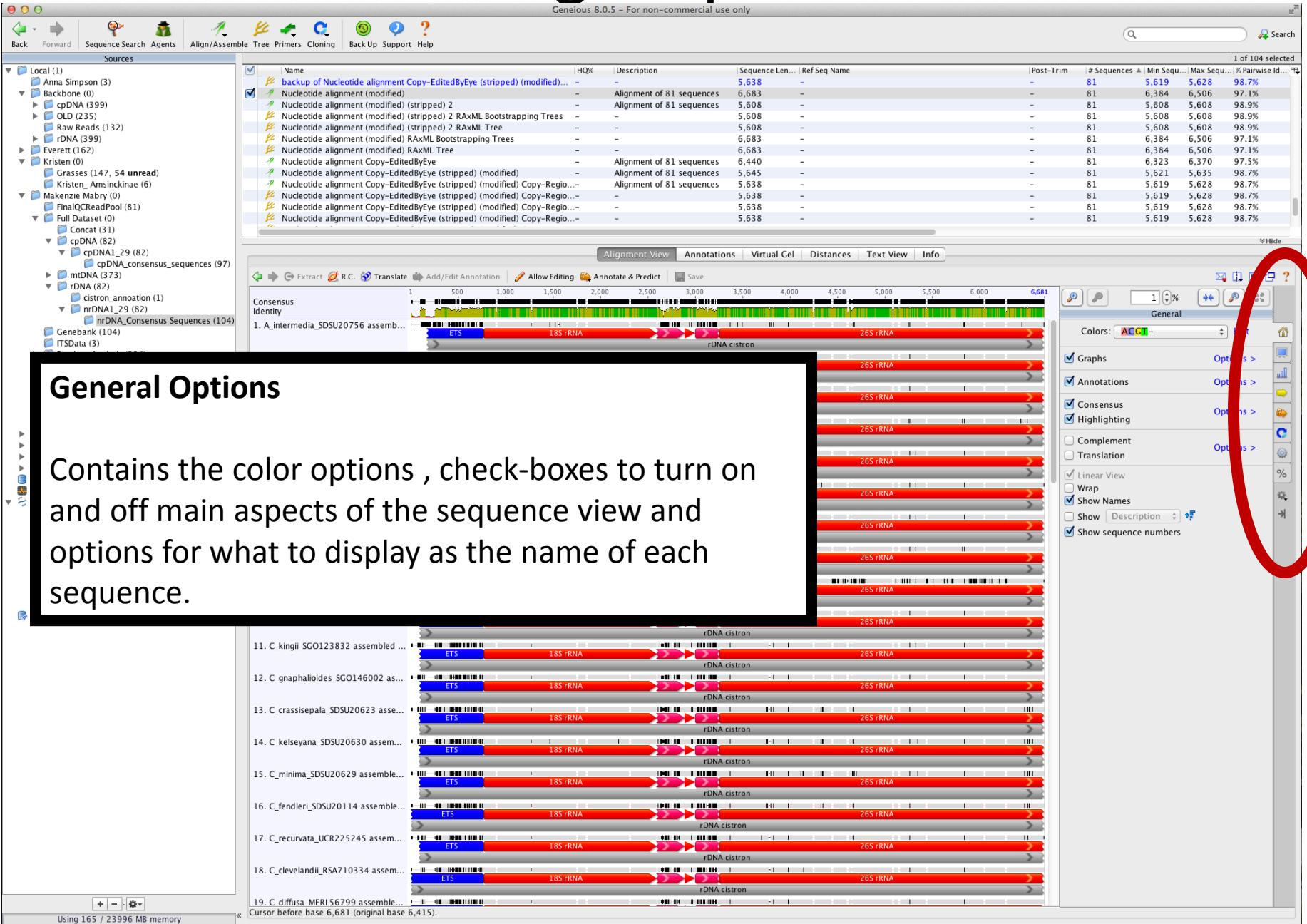
Zooms to 100%. The 100% zoom level allows for comfortable reading of the sequence.

Zooms out so as to fit the entire sequence in the available viewing area.

Alignment View Annotations Virtual Gel Distances Text View Info



Viewing Sequences



Viewing Sequences

Geneious 8.0.5 – For non-commercial use only

Sources

Local (1) Anna Simpson (3) Backbone (0) cpDNA (399) OLD (235) Raw Reads (132) rDNA (399) Everett (162) Kristen (0) Grasses (147, 54 unread) Kristen_Amsinckiae (6) McKenzie Mabry (0) FinalCReadPool (81) Full Dataset (0) Concat (31) cpDNA (82) cpDNA_1..29 (82) cpDNA_consensus_sequences (97) mtDNA (373) rDNA (269) Genbank (104) ITS Data (3) Previous Assemblies Reduced Data Extracted extracted Full Taxa reduced references Riparia data (2) Sample Document Siekkinen (690) Deleted Items Shared Databases Operations NCBI Gene Genome Nucleotide PopSet Protein PubMed SNP Structure Taxonomy UniProt

Display

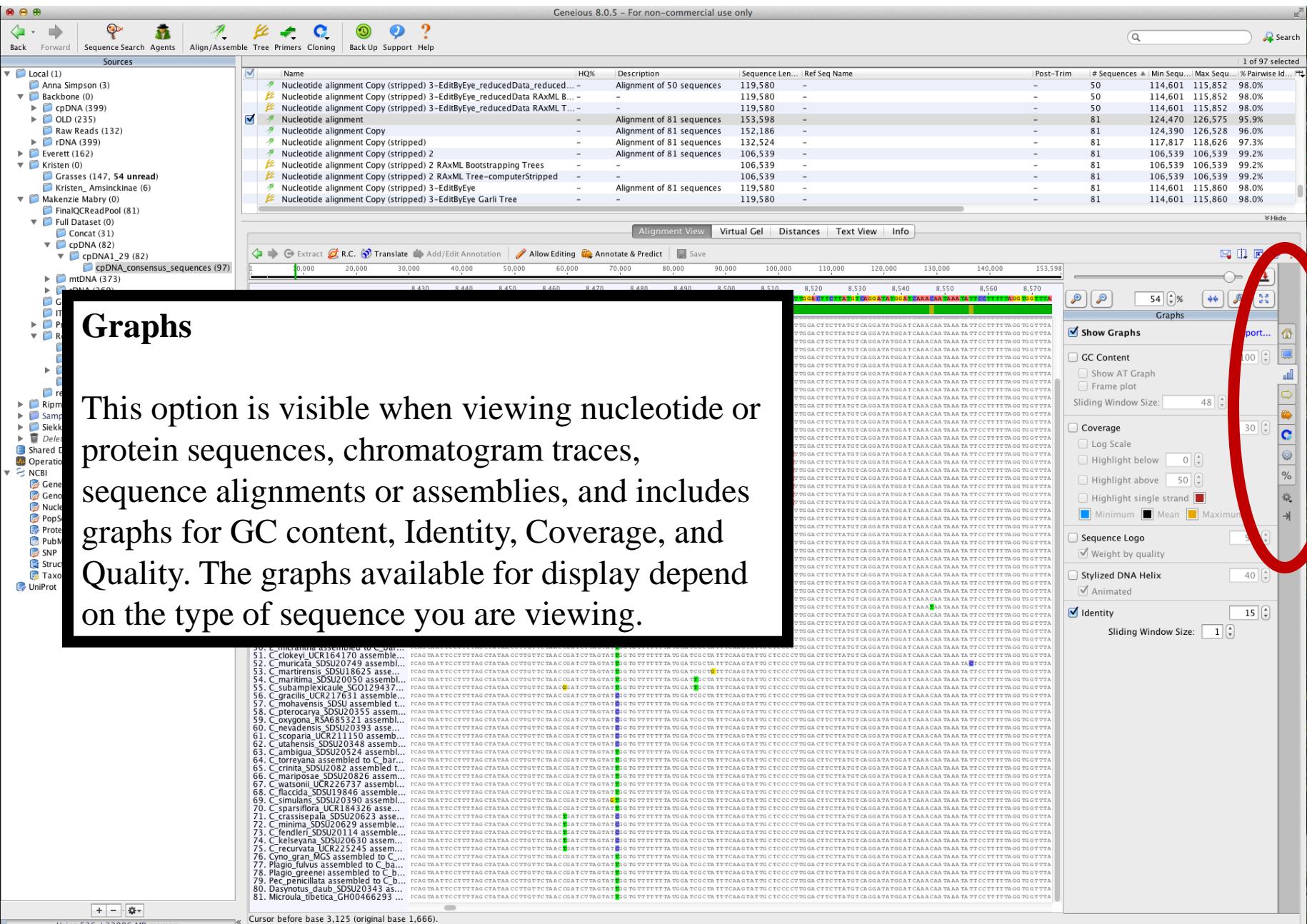
Contains options for displaying the translation and/or complement of a sequence, and turning off the original nucleotide sequence.

Display Options

- Consensus
- Ignore Gaps
- Assign Quality
- If no coverage call
- Disagreement... to Consensus
- Go to next disagreement (#D)
- Use dots
- Nucleotides
- Complement
- Translation
- Translation Options
 - Frame: By selection or annotation
 - Genetic Code: Standard
 - Relative to: Sequence
 - Colors: ARND- Edit

Cursor before base 3,125 (original base 1,666).

Viewing Sequences



Viewing Sequences

Geneious 8.0.5 – For non-commercial use only.

Back Forward Sequence Search Agents Align/Assemble Tree Primers Cloning Back Up Support Help

Sources

Local (1)
Anna Simpson (3)
Backbone (0)

Name HQ% Description Sequence Len... Ref Seq Name Post-Trim % Pairwise Id... % Identical Si... Alignment m... Tree built

CrytorL68 assembled to O humilis reference ITS 99.3% 38,289 reads assembled to... 6,386 O humilis reference ITS 6,386 92.5% 9.5%

Annotations

On sequences containing annotations this tab will show a yellow arrow. It contains controls for turning on and off annotations of each type, customizing the way each type is displayed, and filtering based on annotation name or type.

Ripma data (226)
Shared Documents (623)
Siekkinen (6907, 59 unread)
Deleted Items (3794, 59 unread)
Shared Databases
Operations
NCBI
Gene
Genome
Nucleotide
PopSet
Protein
PubMed
SNP
Structure
Taxonomy
UniProt

View Info

Annotations and Tracks

Show Annotations (9)
 Annotated (1)
 Deletion (2)
 Misc Feature (1)
 Misc_rRNA (2)
 rRNA (3)

Columns Track

Name	Type	Minimum %
CTCTCCCGAACCGA...	Editing Hi...	87
26S rRNA	rRNA	05
internal transcribed s...	misc_rna	65
5.8S rRNA	rRNA	262
internal transcribed s...	misc_rna	73
18S rRNA	rRNA	5
ETS	annotated	<1
rDNA cistron	misc_feat...	<1
NNNNNNNNNNNNNN...	Editing Hi...	1

Alt click on a sequence position or annotation, or select a region to zoom in. Alt-shift click to zoom out.

Using 554 / 23996 MB memory

Viewing Sequences

Geneious 8.0.5 – For non-commercial use only

Back Forward Sequence Search Agents Align/Assemble Tree Primers Cloning Back Up Support Help

Sources

Local (1)
Anna Simpson (3)
Backbone (0)
cpDNA (399)
OLD (235)

Name Description Sequence Length Ref Seq Name Post-Trim % Pairwise Id... % Identical Si... Alignment m... Tree built

CrytorL68 assembled to O humilis reference ITS 99.3% 38,289 reads assembled to... 6,386 O humilis reference ITS 6,386 92.5% 9.5% - -

Live Annotate and Predict

Contains real-time annotation generators such as Annotate from Database, Find ORFs and Transfer Annotations. If you want to save the annotations permanently on the sequence click Apply.

Siekkinen (6907, 59 unread)
Deleted Items (3794, 59 unread)
Shared Databases
Operations
NCBI
Gene
Genome
Nucleotide
PopSet
Protein
PubMed
SNP
Structure
Taxonomy
UniProt

Annotations Dotplot (Self) Text View Info

Live Annotate & Predict

Annotate from...
Similarity: 2 %
Source: Typha Chloro
Some features ignored
Apply Advanced

Find ORFs
Minimum size: 100
Genetic code: Standard
Start codons: CTG, TTG, ATG
Include interior ORFs
Continue outside sequences
Apply Restore Default

The screenshot shows the Geneious 8.0.5 software interface. The main window displays a sequence viewer with a red bar representing a rRNA gene structure. Key features visible include: a navigation toolbar at the top; a 'Sources' panel on the left listing local projects and databases; a central sequence table; a 'Live Annotate and Predict' callout box containing text about real-time annotation tools; a detailed annotation panel on the right with tabs for 'Live Annotate & Predict', 'Annotate from...', and 'Find ORFs'; and a status bar at the bottom providing memory usage information. A red circle highlights the 'Find ORFs' section of the annotation panel.

Viewing Sequences

Geneious 8.0.5 – For non-commercial use only

Back Forward Sequence Search Agents Align/Assemble Tree Primers Cloning Back Up Support Help

Sources

Local (1) Anna Simpson (3)
Backbone (0) cpDNA (399)
OLD (235)

Name HQ% Description Sequence Len... Ref Seq Name Post-Trim % Pairwise Id... % Identical Si... Alignment m... Tree built!
CrytortL68 assembled to O. humilis reference ITS 99.3% 38,289 reads assembled to... 6,386 O. humilis reference ITS 6,386 92.5% 9.5%

Restriction Analysis

This behaves similarly to the Live Annotate & Predict section above.

Genebank (104)
ITSData (3)
Previous Analysis (294)
Reduced Dataset (16)
Extracted cpDNA genes (43)
extracted mtDNA genes (34)
FullTaxa_reducedbp (48)
reducedTaxa_Fullbp (12)
references (2)
Ripma data (226)
Sample Documents (623)
Siekkinen (6907, 59 unread)
Deleted items (3794, 59 unread)
Shared Databases
Operations
NCBI
Gene
Genome
Nucleotide
PopSet
Protein
PubMed
SNP
Structure
Taxonomy
UniProt

Annotations | Dotplot (Self) | Text View | Info

Restriction Analysis

Find Restriction Sites

Candidate Enzymes: commonly used
Enzymes must match 1 to 2 times
Cut Anywhere
1 and 2
Apply Advanced...

Sequence View

500 1,000 1,500 2,000 2,500 3,000 3,500 4,000 4,500 5,000 5,500 6,000 6,386

ETS 18S rRNA 5.8S rRNA 26S rRNA

internal transcribed spacer 1 internal transcribed spacer 2 rDNA cistron

Alt click on a sequence position or annotation, or select a region to zoom in. Alt-shift click to zoom out.

Using 565 / 23996 MB memory

Viewing Sequences

Geneious 8.0.5 – For non-commercial use only

Back Forward Sequence Search Agents Align/Assemble Tree Primers Cloning Back Up Support Help

Sources

Name HQ% Description Sequence Length Ref Seq Name Post-Trim % Pairwise Id... % Identical Si... Alignment m... Tree built

Local (1) Anna Simpson (3) CrytorL68 assembled to O humilis reference ITS 99.3% 38,289 reads assembled to 6,386 O humilis reference ITS 6,386 92.5% 9.5%

Backbone (0)

cpDNA (399)

1 of 1 selected

Advanced

Contains advanced options for controlling the look of sequences and alignments, including wrapping, numbering, annotation placement and font sizes.

Annotations Dotplot (Self) Text View Info

Advanced

Layout

Linear view on circular sequences (checked)
Wrap sequence
Spaces every 10 bases

Properties

Numbering: Standard (checked)
Mini-map
Outline bases when zoomed out

Annotations

Labels: Inside or Outside (dropdown)
Overlay when zoomed out (checked)
Lock annotation layout (checked)
Compress annotations
Show arrow tips
Hide excessive labels

Sizes

NCBI

FullTaxa_reducedbp (48)
reducedTaxa_fullbp (12)
references (2)

Ripma data (226)
Sample Documents (623)
Siekkinen (6907, 59 unread)
Deleted items (3/94, 59 unread)

Shared Databases
Operations

FullTaxa (2)
reducedTaxa (12)
references (2)

Operations

NCBI

Gene
Genome
Nucleotide
PopSet
Protein
PubMed
SNP
Structure
Taxonomy
UniProt

1 500 1,000 1,500 2,000 2,500 3,000 3,500 4,000 4,500 5,000 5,500 6,000 6,386

ETS 18S rRNA 5.8S rRNA 26S rRNA

internal transcribed spacer 1 internal transcribed spacer 2 rDNA cistron

Alt click on a sequence position or annotation, or select a region to zoom in. Alt-shift click to zoom out.

Using 585 / 23996 MB memory

The screenshot shows the Geneious 8.0.5 software interface. The main window displays a sequence alignment of rRNA genes, specifically 18S, 5.8S, and 26S rRNA, along with their internal transcribed spacers (ITS1 and ITS2). The sequence is shown as a horizontal bar with labels indicating positions from 1 to 6,386. Annotations like 'ETS' and 'rDNA cistron' are also present. On the left, there's a sidebar with a file tree showing local sources, reduced taxonomies, and various databases like NCBI and UniProt. The top menu bar includes standard options like Back, Forward, Sequence Search, Align/Assemble, and Help. The right side of the interface features the 'Advanced' settings panel, which is highlighted with a black box. This panel contains sections for Layout (checkboxes for linear view on circular sequences, wrap sequence, and spaces every 10 bases), Properties (checkboxes for Numbering, Mini-map, and outline bases when zoomed out), Annotations (checkboxes for labels, overlay when zoomed out, lock annotation layout, compress annotations, show arrow tips, and hide excessive labels), and Sizes. A red circle specifically highlights the 'Annotations' section within the 'Advanced' panel. The bottom status bar indicates memory usage (Using 585 / 23996 MB memory) and provides zooming instructions.

Viewing Sequences

Geneious 8.0.5 – For non-commercial use only

Back (X) Forward Sequence Search Agents Align/Assemble Tree Primers Cloning Back Up Support Help

Sources

1 of 1 selected

Name %HQ% Description Sequence Length Ref Seq Name Post-Trim %Pairwise Id... %Identical St... Alignment m... Tree build

Anna Simpson (3)
Backbone (0)
cpDNA (399)
OLD (23)

Statistics

Displays statistics about the sequence or alignment currently being viewed, such as length, molecular weight and nucleotide, codon and amino acid frequencies.

reducedTaxa_Fullbp (12)
references (226)
Sample Documents (623)
Siekkinen (6907, 59 unread)
Deleted Items (3794, 59 unread)
Shared Databases
Operations
NCBI
Gene
Genome
Nucleotide
PopSet
Protein
PubMed
SNP
Structure
Taxonomy
UniProt

1 500 1,000 1,500 2,000 2,500 3,000 3,500 4,000 4,500 5,000 5,500 6,000 6,386 5.85 rRNA
18S rRNA
ETS
internal transcribed spacer 1
internal transcribed spacer 2
26S rRNA
rDNA cistron

Length: 6,386
Confidence Mean: 41.0
Expected Errors: 0.52
At least Q20: 100.0%
At least Q30: 100.0%
At least Q40: 99.3%
Freq % non-ambig
A: 1,656 23.0% 22.0%
C: 1,558 24.4% 24.4%
G: 1,902 29.8% 29.8%
T: 1,450 22.7% 22.7%
X: 1 0.0% 0.0%
R: 1 0.0% 0.0%
M: 2 0.0% 0.0%
GCT: 3,460 54.3% 54.3%
Rough Tm: 87.0°C
Molecular weight:
ssDNA: 1979.956 kDa
dsDNA: 3946.176 kDa
Amino Acids & Codons Options
Freq %
A: 17 6.3%
C: 64 3.0%
D: 80 3.9%
E: 11 0.5%
F: 55 2.6%
G: 180 8.5%
H: 44 2.1%
I: 120 5.3%
K: 79 3.7%
L: 201 9.5%
M: 10 0.5%
N: 68 3.2%
P: 132 6.2%
Q: 58 2.7%
R: 166 7.8%
S: 166 7.8%
T: 105 5.0%
V: 195 8.8%
W: 44 2.1%
Y: 24 1.1%
*: 91 4.3%
T: 7
Codon AA % of AA Freq
GCA A 20.6% 28
GCG A 26.5% 36
GCG C 32.1% 44
GCT A 20.6% 28
TGC C 50.8% 32
TGC G 49.2% 31
GAC D 47.5% 38
GAT D 52.5% 42
GAG E 57.1% 48
GAG G 42.9% 36
TTT F 50.9% 28
TTT F 49.1% 27
CTT G 29.1% 26
GGT G 20.6% 37
GGT G 32.2% 58
GGT G 21.7% 39
CAC H 41.5% 30
CAT H 54.5% 24
ATA I 28.3% 26
ATA I 32.7% 31
ATT I 38.0% 35
AAA K 43.0% 34
AAA K 57.0% 45
CTA L 30.3% 37
CTG L 23.4% 43
CTG L 21.4% 43
CTG L 16.7% 39
TTA L 10.9% 22
TTA L 23.4% 47
TTA L 23.4% 47
AAC N 55.9% 38
AAC N 44.1% 30
CAA P 18.9% 25
CCA P 23.0% 33
CCG P 35.6% 47

Alt click on a sequence position or annotation, or select a region to zoom in. Alt-shift click to zoom out.

Using 589 / 23996 MB memory

Sequence Alignments

Geneious 8.0.5 – For non-commercial use only

Using 669 / 23996 MB memory

Geneious is a powerful bioinformatics software for sequence analysis, visualization, and annotation. It features a user-friendly interface with a central workspace for viewing and manipulating biological sequences. The software includes tools for multiple sequence alignment, de novo assembly, genome alignment, primer mapping, and cloning.

The main menu bar includes:

- Back, Forward
- Sequence Search, Agents
- Align/Assemble, Tree, Primers, Cloning
- Multiple Align..., Map to Reference..., De Novo Assemble..., Align Whole Genomes..., Map Primers...
- Back Up, Support, Help

The left sidebar displays a hierarchical tree view of local datasets, including:

- Local (1)
 - Anna Simpson (3)
 - Backbone (0)
 - cpDNA (399)
 - OLD (235)
 - Raw Reads (132)
 - rDNA (399)
 - Everett (162)
 - Kristen (0)
 - Grasses (147, 54 unread)
 - Kristen_Amsinckiae (6)
 - Makenzie Mabry (0)
 - FinalQCReadPool (81)
 - Full Dataset (0)
 - Concat (31)
 - cpDNA (82)
 - cpDNA_129 (82)
 - cpDNA_consensus_sequences (97)
 - mtDNA (373)
 - rDNA (82)
 - cistrion_annotation (1)
 - nrDNA1_29 (82)
 - nrDNA_Consensus Sequences (104)
 - Genebank (104)
 - ITSData (3)
 - Previous Analysis (294)
 - Reduced Dataset (16)
 - Extracted cpDNA genes (43)
 - extracted mtDNA genes (34)
 - FullTaxa_reducedbp (48)
 - reducedTaxa_Fullbp (12)
 - references (2)
 - Rippl data (226)
 - Sample Documents (623)
 - Siekinen (6907, 59 unread)
 - Deleted Items (3794, 59 unread)
 - Shared Databases
 - Operations
 - NCBI
 - Gene
 - Genome
 - Nucleotide
 - PopSet
 - Protein
 - PubMed
 - SNP
 - Structure
 - Taxonomy
 - UniProt

The central workspace shows a list of selected sequences with details like HQ%, Description, Sequence Length, Ref Seq Name, Post-Trim, # Sequences, Min Sequence, Max Sequence, and % Pairwise Id. A large red circle highlights the "Align/Assemble" menu item.

The bottom right panel shows a zoomed-in view of a sequence alignment with various regions labeled: ETS, 18S rRNA, internal transcribed spacer 1, 5.8S rRNA, internal transcribed spacer 2, 26S rRNA, and rDNA cistron.

The bottom status bar indicates "Using 669 / 23996 MB memory".

Assembly and Mapping

Geneious 8.0.5 – For non-commercial use only

Back Forward Sequence Search Agents Align/Assemble Tree Primers Cloning Back Up Support Help

Multiple Align... ⌘A Map to Reference... De Novo Assemble... Align Whole Genomes... Map Primers...

Local (1) Anna Simpson (3) Backbone (0) cpDNA (399) OLD (23) Raw Reads (132) rDNA (399) Everett (162) Kristen (0) Grasses (147, 54 unread) Kristen_Amsinckiae (6) Makenzie Mabry (0) FinalQCReadPool (81) Full Dataset (0) Concat (31) cpDNA (82) cpDNA_29 (82) cpDNA_consensus_sequences (97) mtDNA (373) rDNA (82) cistrion_annotation (1) nrDNA1_29 (82) nrDNA_Consensus Sequences (104) Genebank (104) ITSData (3) Previous Analysis (294) Reduced Dataset (16) Extracted cpDNA genes (43) extracted mtDNA genes (34) FullTaxa_reducedbp (48) reducedTaxa_Fullbp (12) references (2) Ripma data (226) Sample Documents (623) Siekkinen (6907, 59 unread) Deleted Items (3794, 59 unread) Shared Databases Operations NCBI Gene Genome Nucleotide PopSet Protein PubMed SNP Structure Taxonomy UniProt

81 of 104 selected

HQ% Description Sequence Len... Ref Seq Name Post-Trim # Sequences Min Sequ... Max Sequ... % Pairwise Id...

assembled to Consensus 100.0% 23,414 reads assembled to... 6,333 Consensus 6,333 - - - 98.2% assembled to Consensus 100.0% 60,626 reads assembled to... 6,330 Consensus 6,330 - - - 98.0% assembled to Consensus 100.0% 56,397 reads assembled to... 6,324 Consensus 6,324 - - - 95.7% JEP8750 assembled to Consensus 100.0% 27,823 reads assembled to... 6,336 Consensus 6,336 - - - 95.5% Plagio_jonesii_UCR2154 assembled to Consensus 100.0% 31,619 reads assembled to... 6,331 Consensus 6,331 - - - 96.9% Plagio_kingii_UC1874 assembled to Consensus 100.0% 5,619 reads assembled to... 5,619 Consensus 5,619 - - - 88.6% Nucleotide alignment Copy-Edited 5,619 reads assembled to... 5,619 Consensus 5,619 - - - 98.6% Nucleotide alignment (modified) 5,619 reads assembled to... 5,619 Consensus 5,619 - - - 98.6% Nucleotide alignment (modified) 5,619 reads assembled to... 5,619 Consensus 5,619 - - - 98.6% Nucleotide alignment (modified) 5,619 reads assembled to... 5,619 Consensus 5,619 - - - 98.7% Nucleotide alignment (modified) 6,384 reads assembled to... 6,384 Consensus 6,384 - - - 97.1% Nucleotide alignment (modified) 5,608 reads assembled to... 5,608 Consensus 5,608 - - - 98.9% Nucleotide alignment (modified) 5,608 reads assembled to... 5,608 Consensus 5,608 - - - 98.9%

Align/Assemble Tree Primers Cloning

Multiple Align... ⌘A

Map to Reference...

De Novo Assemble...

Align Whole Genomes...

Map Primers...

Alt click on a sequence position or annotation, or select a region to zoom in. Alt-shift click to zoom out.

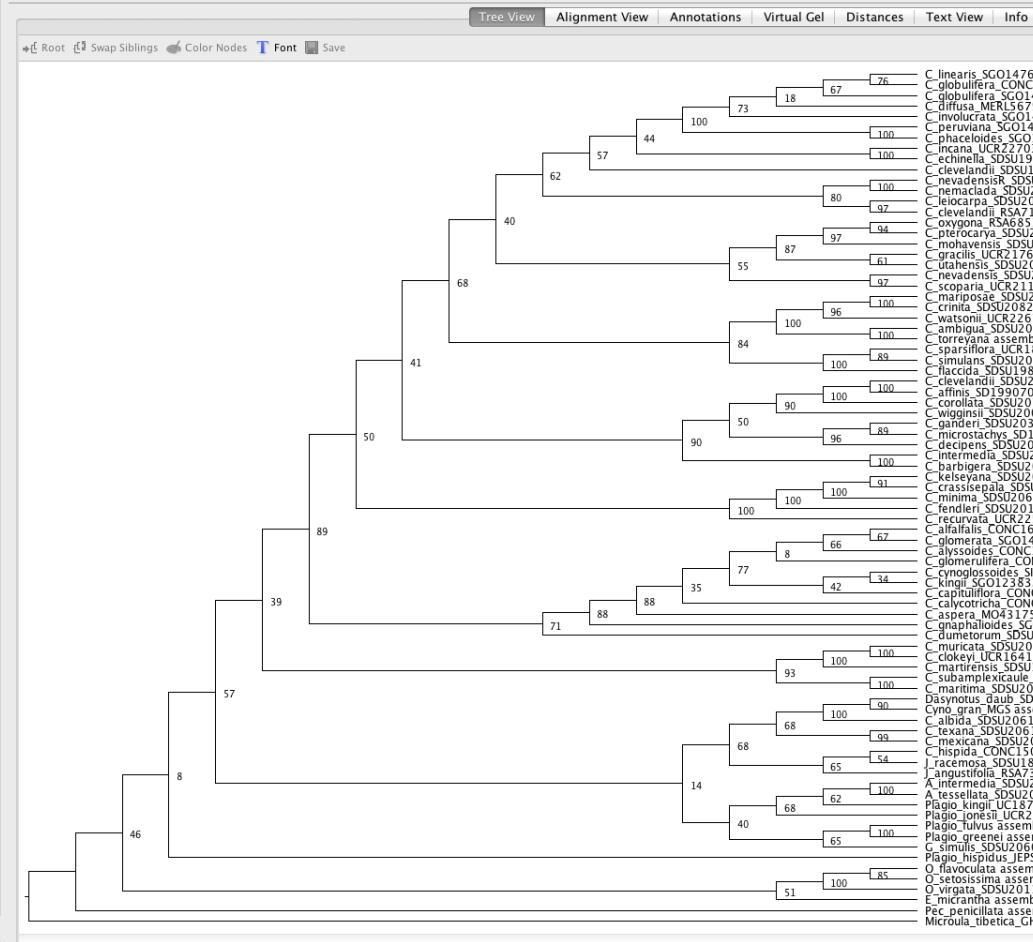
Using 669 / 23996 MB memory

Phylogenetic Trees

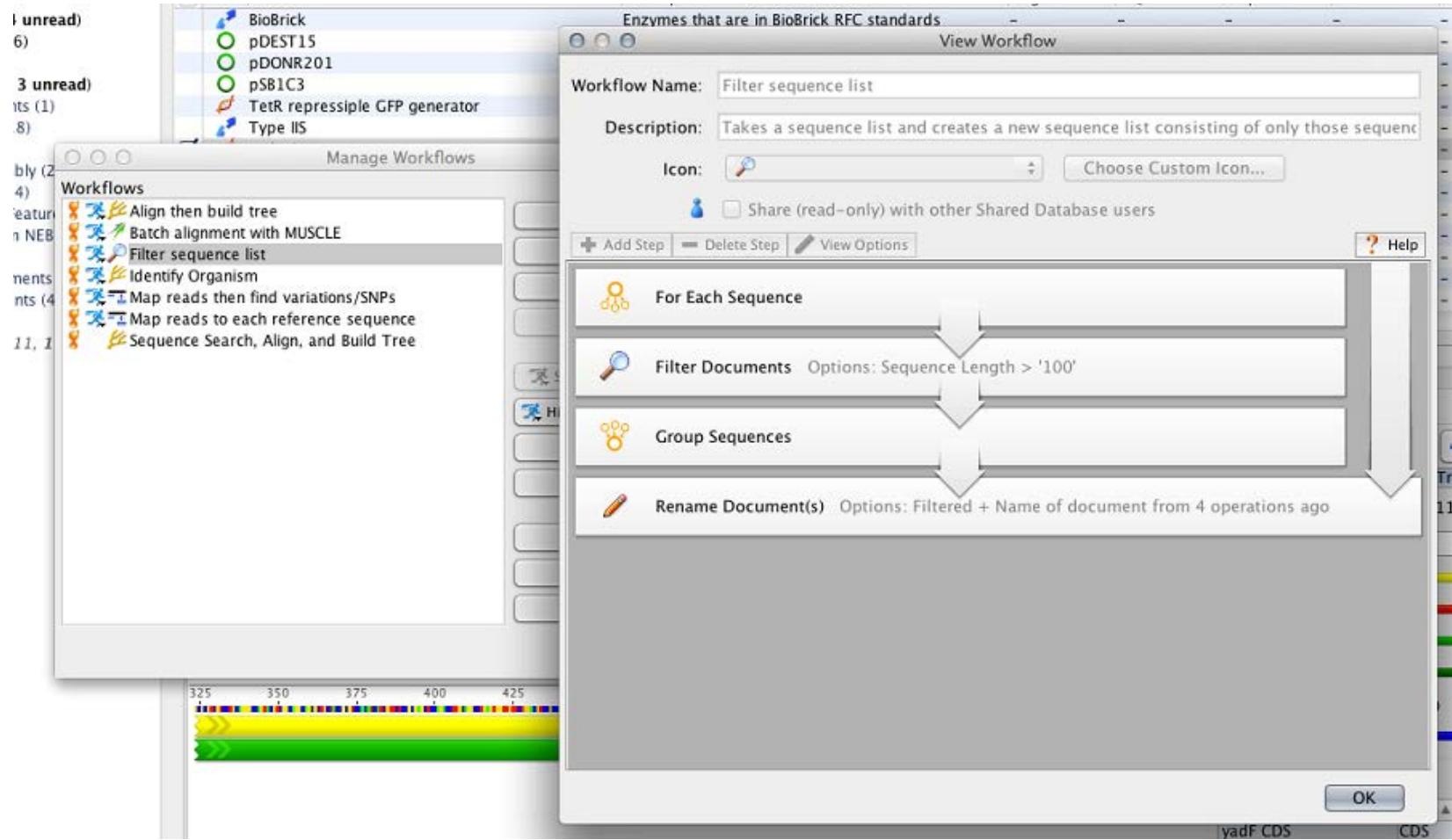
Geneious 8.0.5 – For non-commercial use only

Sources

- Local (1)
 - Anna Simpson (3)
 - Backbone (0)
 - cpDNA (399)
 - OLD (235)
 - Raw Reads (132)
 - rDNA (399)
 - Everett (162)
 - Kristen (0)
 - Grasses (147, 54 unread)
 - Kristen_Amsinckiae (6)
 - Makenzie Mabry (0)
 - FinalQCReadPool (8) Kristen_Amsinckiae folder for storing your documents.
 - Copy-EditedByEye (stripped) (modified)
 - Copy-EditedByEye (stripped) 2 RAxML Bootstrapping Trees
 - Copy-EditedByEye (stripped) 2 RAxML Tree
 - Copy-EditedByEye (stripped) RAxML Bootstrapping Trees
 - Nucleotide alignment (modified) RAxML Tree
 - Nucleotide alignment Copy-EditedByEye (stripped) (modified)
 - Nucleotide alignment Copy-EditedByEye (stripped) (modified) Copy-Regio...
 - Nucleotide alignment Copy-EditedByEye (stripped) (modified) Copy-Regio...
 - Nucleotide alignment Copy-EditedByEye (stripped) (modified) Copy-Regio...
 - Nucleotide alignment Copy-EditedByEye (stripped) (modified) Copy-Regio...
 - Full Dataset (0)
 - Concat (31)
 - cpDNA (82)
 - cpDNA1_29 (82)
 - cpDNA_consensus_sequences (97)
 - mtDNA (373)
 - rDNA (82)
 - cistrion_annotation (1)
 - nrDNA1_29 (82)
 - nrDNA_Consensus Sequences (104)
 - Genebank (104)
 - ITSData (3)
 - Previous Analysis (294)
 - Reduced Dataset (16)
 - Extracted cpDNA genes (43)
 - extracted mtDNA genes (34)
 - FullTaxa_reducedbp (48)
 - reducedTaxa_Fullbp (12)
 - references (2)
 - Ripma data (226)
 - Sample Documents (623)
 - Siekinen (6907, 59 unread)
 - Deleted Items (3794, 59 unread)
 - Shared Databases
 - Operations
 - NCBI
 - Gene
 - Genome
 - Nucleotide
 - PopSet
 - Protein
 - PubMed
 - SNP
 - Structure
 - Taxonomy
 - UniProt



Workflows



In Summary...

	CLC (Main)	CLC (Genomics)	Galaxy / UCSC	Lasergene	Vector NTI	Sequencher	Geneious
Sequence searching	✓	✓	✓	✓	✓	✓	✓
Text & literature searching	✓	✓	✗	✗	✗	✗	✓
Database creation & management	✓	✓	✓	✗	✗	✗	✓
Chromatogram editing & assembly	✓	✓	✗	✓	✓	✓	✓
Pairwise & multiple alignment	✓	✓	✓	✓	✓	✓	✓
Phylogenetics	✓	✓	✓	✓	✗	✗	✓
Molecular cloning & restriction analyses	✓	✓	✗	✓	✓	✗	✓
Predict & annotate	✓	✓	✓	✓	✓	✗	✓
Primer design & analysis	✓	✓	✗	✓	✓	✗	✓
Genome Browsing	✗	✓	✓	✓	✓	✓	✓
Reference mapping	✗	✓	✓	✓	✓	✓	✓
Variant calling	✗	✓	✓	✓	✗	✓	✓
De novo assembly	✗	✓	✓	✓	✗	✓	✓
Comparative genomics	✗	✓	✓	✓	✗	✗	✓
Microsatellite analysis	✗	✗	✗	✗	✗	✗	✓
Workflow tracking	✓	✓	✓	✗	✓	✗	✓
API (plugin development kit)	✓	✓	✓	✗	✓	✗	✓
Sequence Submission	✗	✗	✗	✗	✗	✗	✓

Introduction to NGS Sequencing Workshop

Botany 2015

8:30 - 8:40 am - Intro to the workshop and presenters

8:40 - 9:05 am - Intro to Sequencing Technology

9:05 - 9:30 am - Sequence Processing/Intro Bioinformatics

9:30 - 10:00 am - Group Activity

10:00 - 10:15 am - Coffee break

10:15 - 11:00 am - Genome Reduction Methods

11:00 - 11:30 am - Applications and data analysis - phylogenomics

11:30 am - 12:00 pm - Applications and data analysis - population genomics

12:00 - 1:00 pm - Lunch

1:00 - 1:20 pm - Introduction to the hands-on exercises

1:20 - 4:30 pm - Hands-on exercises