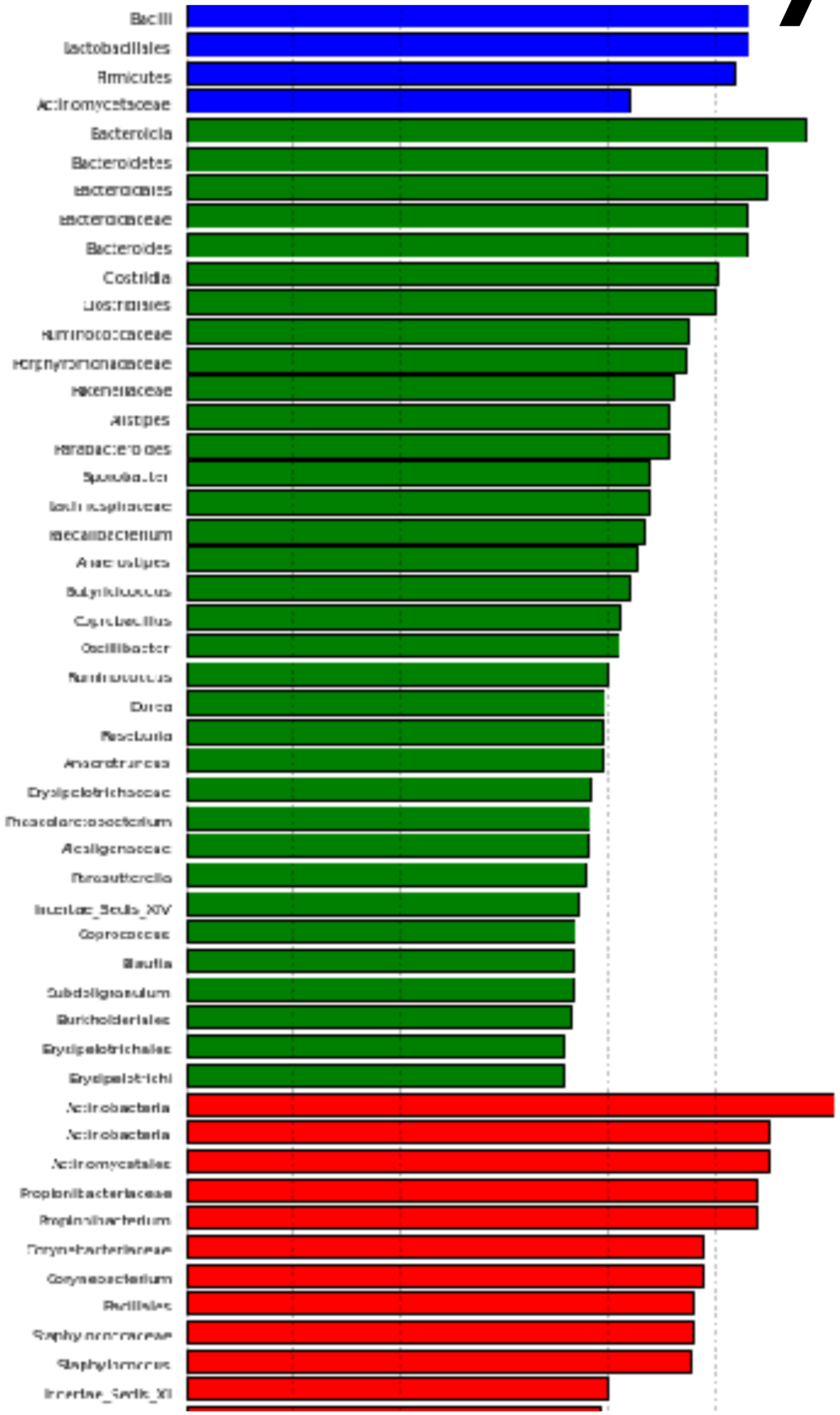
Assorted Topics and Other Stuff

March 29, 2019

Josh Granek

Biomarker Discovery

- Question: Which OTUs* have different abundance between Site A and Site B



* or higher level taxonomic groups,

LEfSe Tutorial

Biomarker Discovery

- LEfSe
- MetaBoot
- Metastats
- LIBSVM
- mRMR
- Regularized Low Rank-Sparse Decomposition (RegLRSD)

PICRUSt

Metagenomics

	What	Information	Analogy	Target Size	Cost
Amplicon	Marker Gene	Who is Present	Name	100bp - 1kb	Low
Shotgun Metagenome	Genomes	What Genes are Present	CV	100kb - 100Mb	High
Shotgun Metatranscriptome	All RNA	What Genes are Expressed	Twitter Feed	100kb - 100Mb	High

Amplicon Sequencing

PCR amplify and sequence a marker gene

	Marker Gene
Bacteria	16s rRNA
Fungi	18s or ITS rRNA
Archaea	16s rRNA
Protozoa	18s rRNA
Viruses	?????

Metagenomics

	What	Information	Analogy	Target Size	Cost
Amplicon	Marker Gene	Who is Present	Name	100bp - 1kb	Low
Shotgun Metagenome	Genomes	What Genes are Present	CV	100kb - 100Mb	High
Shotgun Metatranscriptome	All RNA	What Genes are Expressed	Twitter Feed	100kb - 100Mb	High

Metagenomics

	What	Information	Analogy	Target Size	Cost	Discovery?
Amplicon	Marker Gene	Who is Present	Name	100bp - 1kb	Low	+/-
Shotgun Metagenome	Genomes	What Genes are Present	CV	100kb - 100Mb	High	++
Shotgun Metatranscriptome	All RNA	What Genes are Expressed	Twitter Feed	100kb - 100Mb	High	++

PICRUSt

- **What I Have:**

  250bp sequence from v4 region of 16s rRNA gene

- **What I Want:**

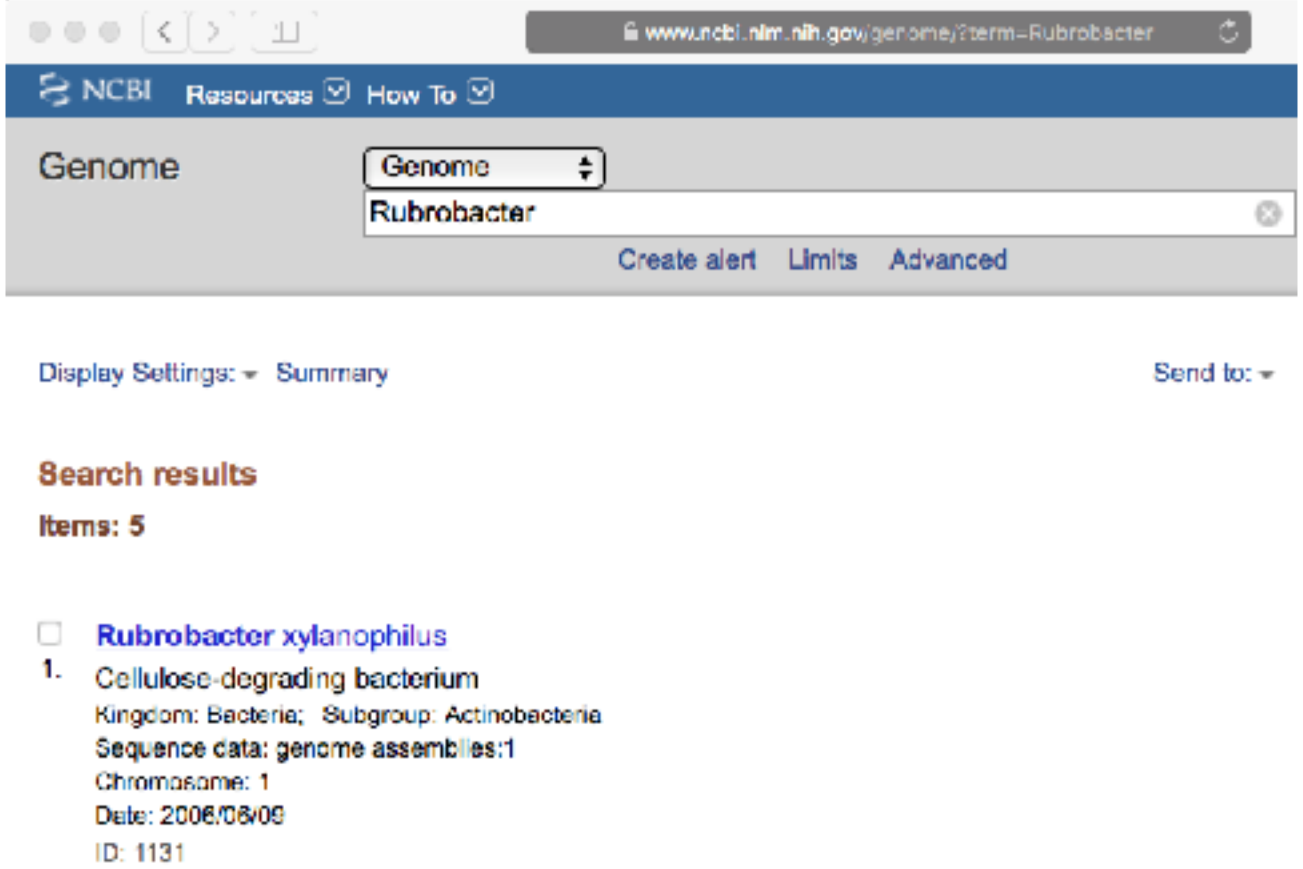
1. All the genes in the sample
2. The relative abundance of all the genes in the sample

Inferring Gene Content

16s rRNA v4

```
GCGAGCGTTGTCCGGAATTATTGGGCGTAAAGAGCGTGTAGGCGGTTCGGT
AAGTCTGCCGTGAAAACCTGGGGCTCAACCCCGGGCGTGCGGTGGATACTG
CCGGGCTAGAGGATGGTAGAGGCGAGTGGAATTCCCGGTGTAGCGGTGAAA
TGCGCAGATATCGGGAGGAACACCAGTAGCGAAGGCGGCTCGCTGGGCCAT
TCCTGACGCTGAGACGCGAAAGCTAGGGG
```

Rubrobacter



Rubrobacter

Search results

Items: 5

□ Rubrobacter xylanophilus

1. Cellulose-degrading bacterium

Kingdom: Bacteria; Subgroup: Actinobacteria

Sequence data: genome assemblies:1

Chromosome: 1

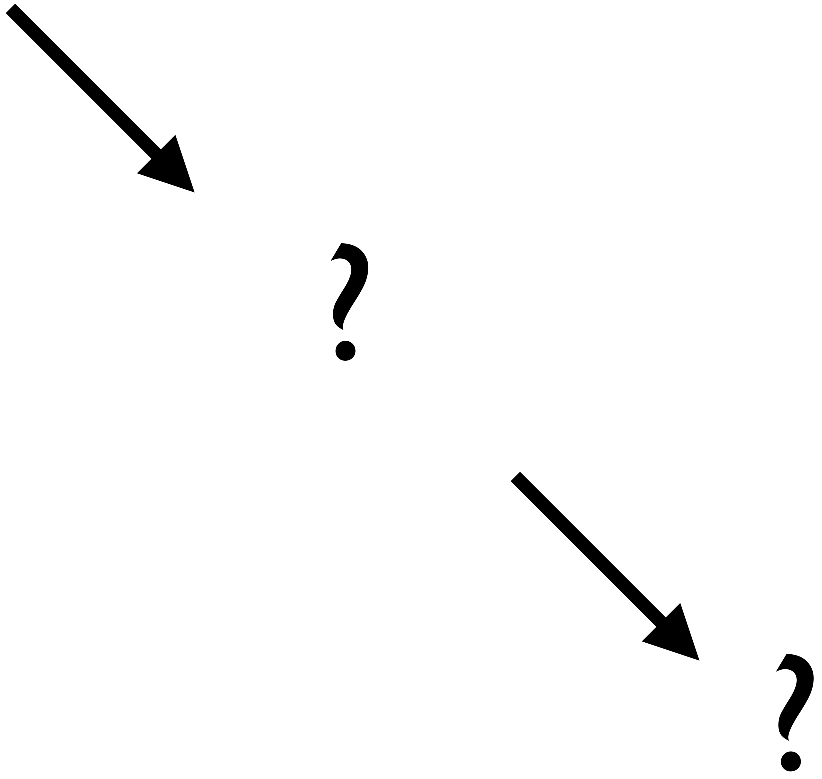
Date: 2006/06/09

ID: 1131

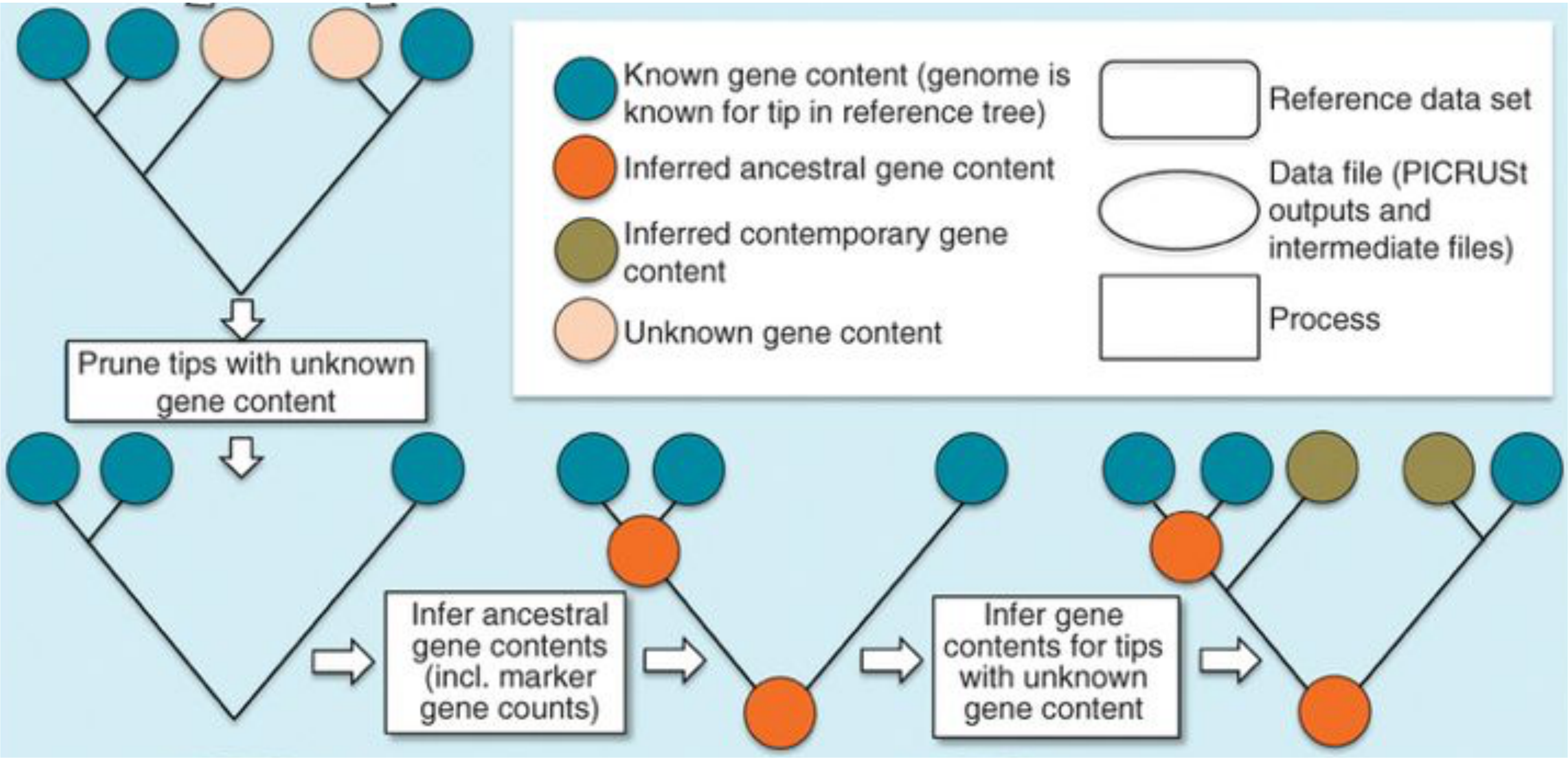
Inferring Gene Content

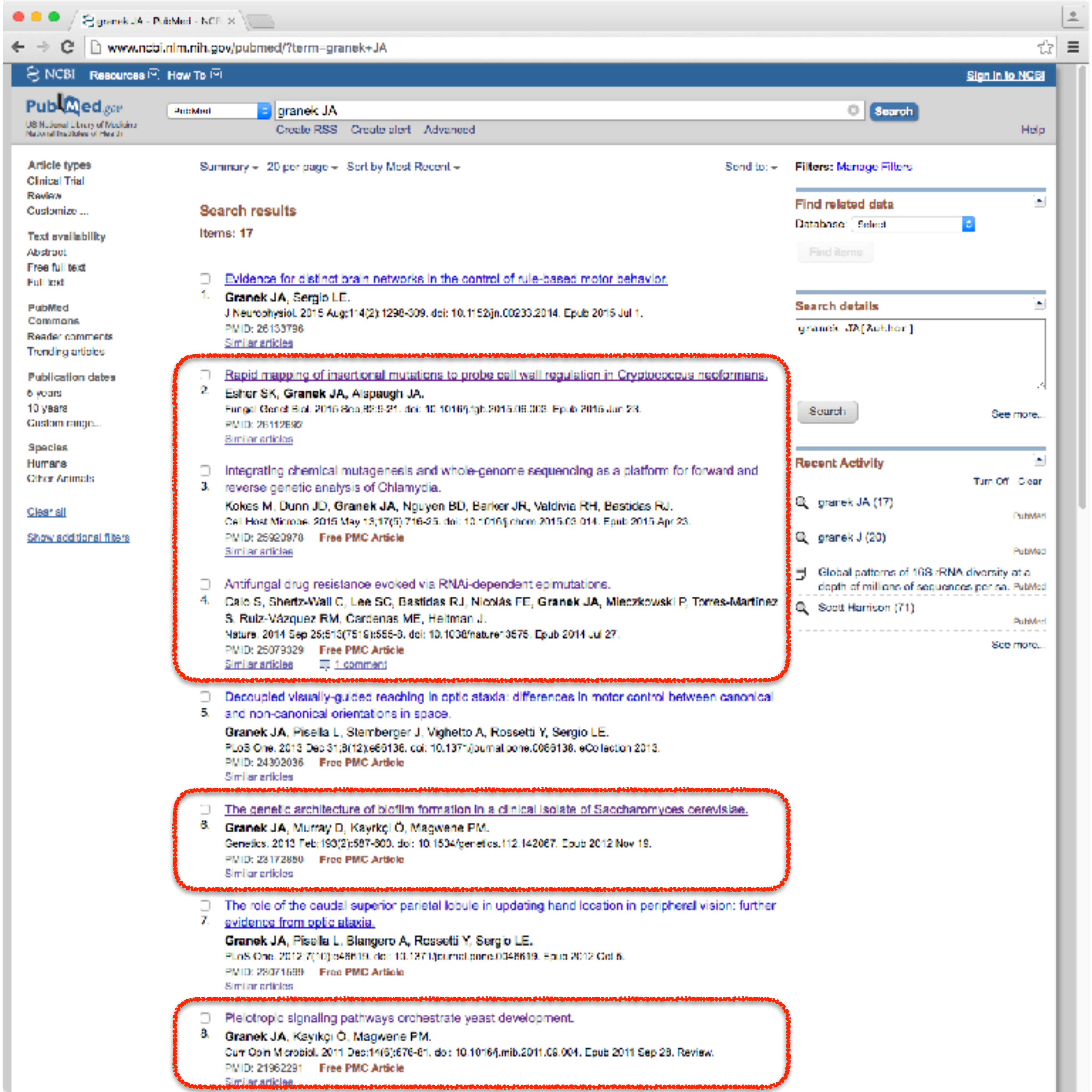
16s rRNA v4

```
GCGAGCGTTAATCGGAATTACTGGGCGTAAAGGGCGCGTAGGCGGTGAAGT
AAGTCGGGTGTGAAAGCCCCGGGCTCAACCTGGGAACTGCATCCGATACTG
CTTCGCTAGAGTATGGTAGAGGGAAGCGGAATTCCGGGTGTAGCGGTGAAA
TGCGTAGATATCCGGAGGAACACCAGTGGCGAAGGCGGCTTCCTGGACCAA
TACTGACGCTGAGGCGCGAAAGCGTGGGG
```



Inferring Gene Content


Known gene content (genome is known for tip in reference tree)
Inferred ancestral gene content
Inferred contemporary gene content
Unknown gene content
Reference data set
Data file (PICRUSt outputs and intermediate files)
Process
Prune tips with unknown gene content
Infer ancestral gene contents (incl. marker gene counts)
Infer gene contents for tips with unknown gene content


granek JA - PubMed - NCBI
www.ncbi.nlm.nih.gov/pubmed/?term=granek+JA
NCBI
Resources
How To
Sign in to NCBI
PubMed.gov
US National Library of Medicine
National Institutes of Health
PubMed
Search
Create RSS
Create alert
Advanced
Help

granek JA

Article types
Clinical Trial
Review
Customize ...

Text availability
Abstract
Free full text
Full text

PubMed
Commons
Reader comments
Trending articles

Publication dates
5 years
10 years
Custom range...

Species
Humans
Other Animals

Clear all

Show additional filters

Summary ▾ 20 per page ▾ Sort by Most Recent ▾ Send to: ▾

Search results

Items: 17

□ 1. Evidence for distinct brain networks in the control of rule-based motor behavior.
Granek JA, Sergio LE.
J Neurophysiol. 2015 Aug;114(2):1298-309. doi: 10.1152/jn.00233.2014. Epub 2015 Jul 1.
PMID: 26133796
Similar articles

□ 2. Rapid mapping of insertional mutations to probe cell wall regulation in Cryptococcus neoformans.
Esher SK, **Granek JA**, Alspaugh JA.
Fungal Genet Biol. 2015 Sep;82:9-21. doi: 10.1016/j.fgb.2015.06.003. Epub 2015 Jun 23.
PMID: 26112692
Similar articles

□ 3. Integrating chemical mutagenesis and whole-genome sequencing as a platform for forward and reverse genetic analysis of Chlamydia.
Kokes M, Dunn JD, **Granek JA**, Nguyen BD, Barker JR, Valdivia RH, Bastidas RJ.
Cell Host Microbe. 2015 May 13;17(5):716-25. doi: 10.1016/j.chom.2015.03.014. Epub 2015 Apr 23.
PMID: 25920978 **Free PMC Article**
Similar articles

□ 4. Antifungal drug resistance evoked via RNAi-dependent epimutations.
Calo S, Shertz-Wall C, Lee SC, Bastidas RJ, Nicolás FE, **Granek JA**, Mieczkowski P, Torres-Martinez S, Ruiz-Vázquez RM, Cardenas ME, Heitman J.
Nature. 2014 Sep 25;513(7519):555-8. doi: 10.1038/nature13575. Epub 2014 Jul 27.
PMID: 25079329 **Free PMC Article**
Similar articles 1 comment

□ 5. Decoupled visually-guided reaching in optic ataxia: differences in motor control between canonical and non-canonical orientations in space.
Granek JA, Pisella L, Stemberger J, Vighetto A, Rossetti Y, Sergio LE.
PLoS One. 2013 Dec 31;8(12):e86138. doi: 10.1371/journal.pone.0086138. eCollection 2013.
PMID: 24392036 **Free PMC Article**
Similar articles

□ 6. The genetic architecture of biofilm formation in a clinical isolate of Saccharomyces cerevisiae.
Granek JA, Murray D, Kayrkçi Ö, Magwene PM.
Genetics. 2013 Feb;193(2):587-600. doi: 10.1534/genetics.112.142067. Epub 2012 Nov 19.
PMID: 23172850 **Free PMC Article**
Similar articles

□ 7. The role of the caudal superior parietal lobule in updating hand location in peripheral vision: further evidence from optic ataxia.
Granek JA, Pisella L, Blangero A, Rossetti Y, Sergio LE.
PLoS One. 2012;7(10):e46619. doi: 10.1371/journal.pone.0046619. Epub 2012 Oct 5.
PMID: 23071599 **Free PMC Article**
Similar articles

□ 8. Pleiotropic signaling pathways orchestrate yeast development.
Granek JA, Kayıkçı Ö, Magwene PM.
Curr Opin Microbiol. 2011 Dec;14(6):676-81. doi: 10.1016/j.mib.2011.08.004. Epub 2011 Sep 23. Review.
PMID: 21962291 **Free PMC Article**
Similar articles

Filters: Manage Filters

Find related data
Database: Select
Find items

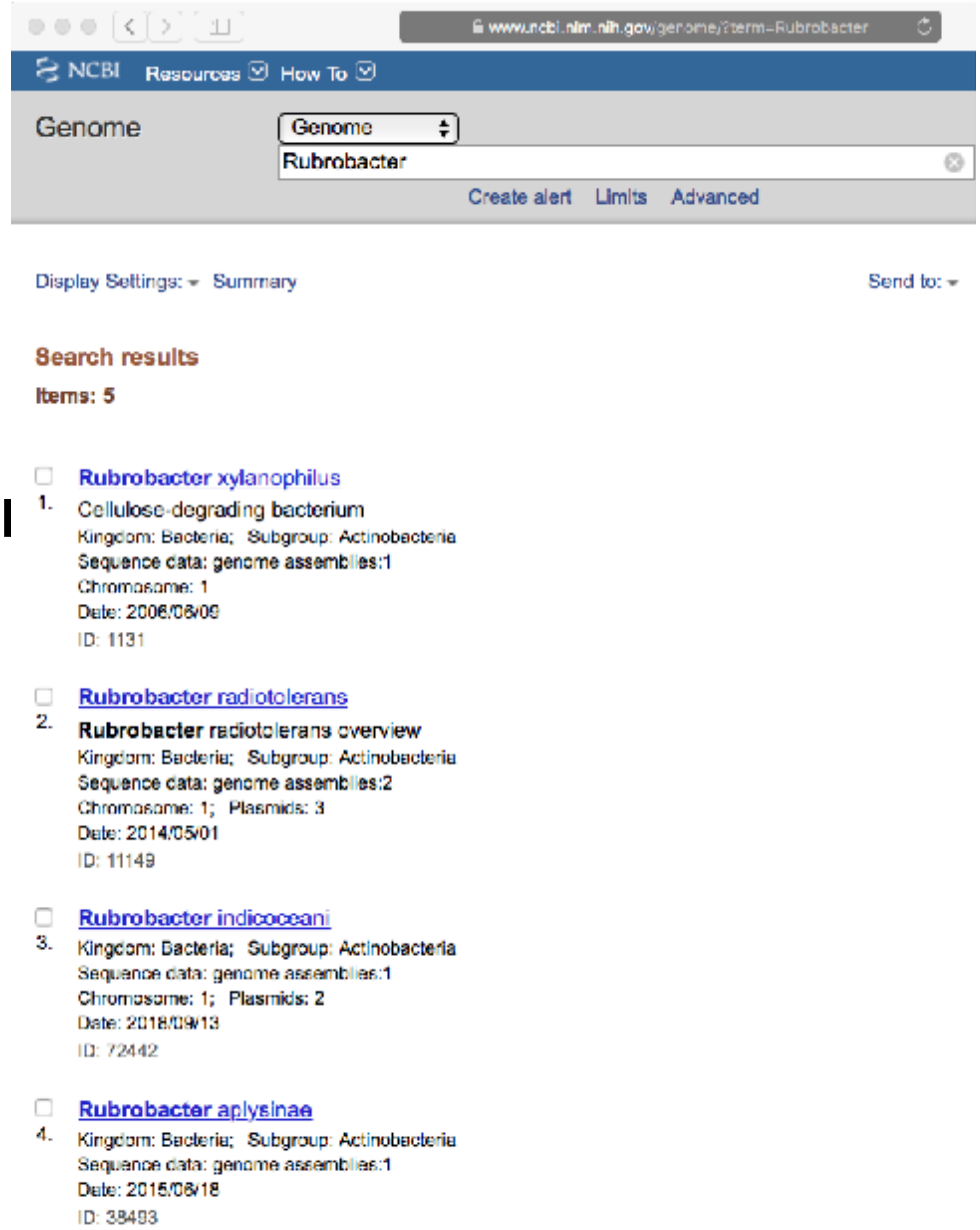
Search details
granek JA[Author]
Search
See more...

Recent Activity
Turn Off Clear
granek JA (17) PubMed
granek J (20) PubMed
Global patterns of 16S rRNA diversity at a depth of millions of sequences per sa... PubMed
Scott Harrison (71) PubMed
See more...

16s rRNA v4

```
GCGAGCGTTGTCCGGAATTATTGGGCGTAAAGAGCGTGTAGGCGGTTCGGT
AAGTCTGCCGTGAAAACCTGGGGCTCAACCCCGGGCGTGCGGTGGATACTG
CCGGGCTAGAGGATGGTAGAGGCGAGTGGAATTCCCGGTGTAGCGGTGAAA
TGCGCAGATATCGGGAGGAACACCAGTAGCGAAGGCGGCTCGCTGGGCCAT
TCCTGACGCTGAGACGCGAAAGCTAGGGG
```

Rubrobacter
(genus)



Search results

Items: 5

□ 1. Rubrobacter xylanophilus

Cellulose-degrading bacterium

Kingdom: Bacteria; Subgroup: Actinobacteria

Sequence data: genome assemblies:1

Chromosome: 1

Date: 2006/08/09

ID: 1131

□ 2. Rubrobacter radiotolerans

Rubrobacter radiotolerans overview

Kingdom: Bacteria; Subgroup: Actinobacteria

Sequence data: genome assemblies:2

Chromosome: 1; Plasmids: 3

Date: 2014/05/01

ID: 11149

□ 3. Rubrobacter indicoceani

Kingdom: Bacteria; Subgroup: Actinobacteria

Sequence data: genome assemblies:1

Chromosome: 1; Plasmids: 2

Date: 2018/09/13

ID: 72442

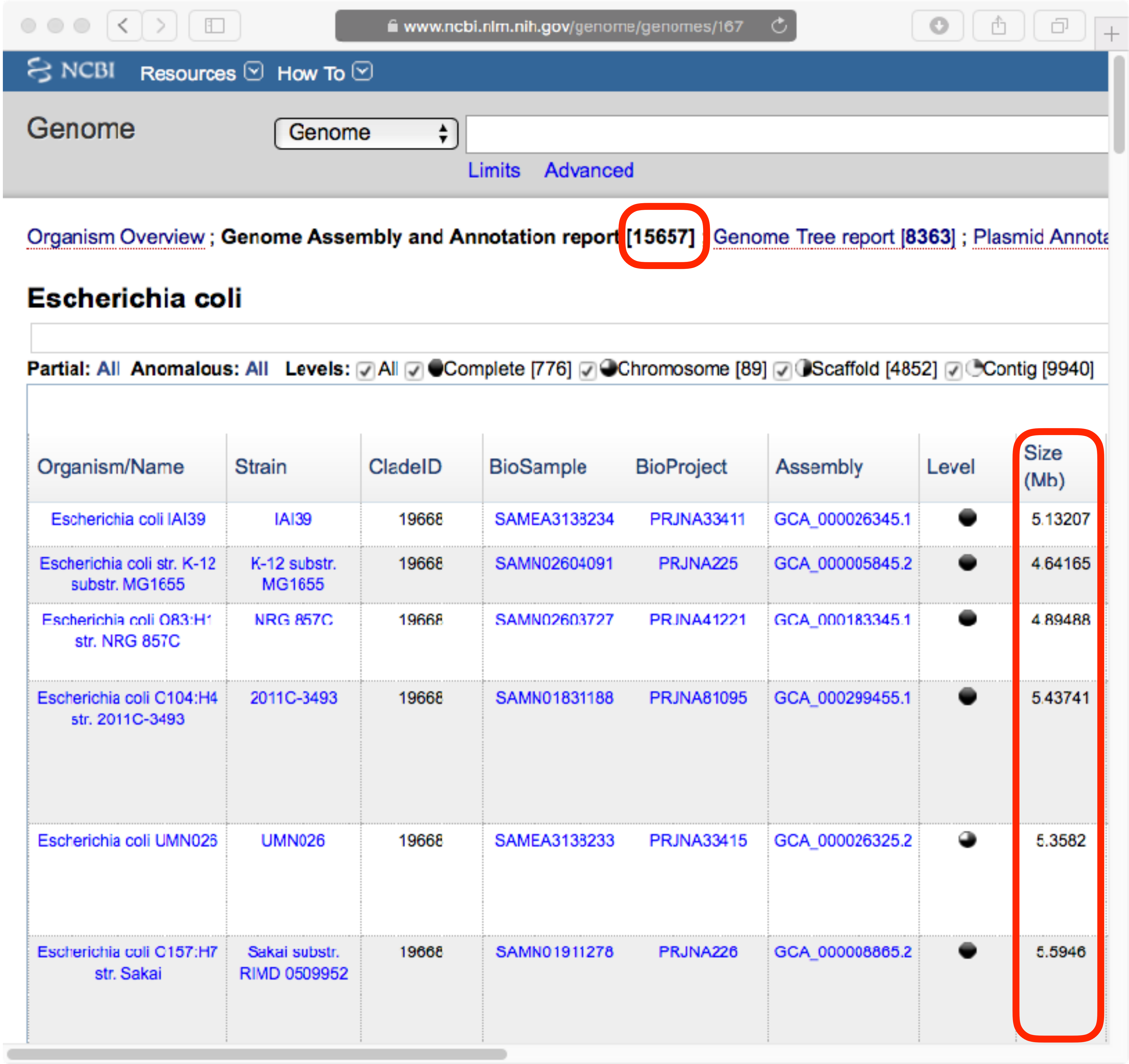
□ 4. Rubrobacter aplysinae

Kingdom: Bacteria; Subgroup: Actinobacteria

Sequence data: genome assemblies:1

Date: 2015/06/18

ID: 38493



Organism Overview ; **Genome Assembly and Annotation report [15657]** ; Genome Tree report [8363] ; Plasmid Annota

Escherichia coli

Partial: All **Anomalous:** All **Levels:** ☑ All ☑ ●Complete [776] ☑ ◕Chromosome [89] ☑ ◑Scaffold [4852] ☑ ◔Contig [9940]

Organism/Name	Strain	CladeID	BioSample	BioProject	Assembly	Level	Size (Mb)
Escherichia coli IAI39	IAI39	19668	SAMEA3138234	PRJNA33411	GCA_000026345.1	●	5.13207
Escherichia coli str. K-12 substr. MG1655	K-12 substr. MG1655	19668	SAMN02604091	PRJNA225	GCA_000005845.2	●	4.64165
Escherichia coli O83:H1 str. NRG 857C	NRG 857C	19668	SAMN02603727	PRJNA41221	GCA_000183345.1	●	4.89488
Escherichia coli O104:H4 str. 2011C-3493	2011C-3493	19668	SAMN01831188	PRJNA81095	GCA_000299455.1	●	5.43741
Escherichia coli UMN026	UMN026	19668	SAMEA3138233	PRJNA33415	GCA_000026325.2	◕	5.3582
Escherichia coli O157:H7 str. Sakai	Sakai substr. RIMD 0509952	19668	SAMN01911278	PRJNA226	GCA_000008865.2	●	5.5946

PICRUSt is . . .

- PICRUSt is to Bacteria
- Googling is to People

HTS Applications

- DNA-Seq
- RNA-Seq
- Amplicon Sequencing
- Many More
  - ChIP-Seq
  - Ribo-Seq
  - Hi-C
  - MethylC-Seq

https://www.illumina.com/content/dam/illumina-marketing/documents/applications/ngs-library-prep/for-all-you-seq-dna.pdf

DNA-Seq

- De *Novo* Genome Sequencing
- Genotyping
  - GWAS
  - Genetic risk factors
- Mutation identification

https://www.illumina.com/content/dam/illumina-marketing/documents/applications/ngs-library-prep/for-all-you-seq-rna.pdf

RNA-Seq

- Transcriptome: “Which genes are expressed in this sample?”
  - Differential Expression
  - Genome Annotation
- SNPs
- Gene Fusions

RNA-Seq

- Bulk RNA-Seq
- Single-Cell RNA-Seq (scRNA-Seq)

Amplicon Sequencing

- CRISPER Barcode Seq
- 16s rRNA

*-Seq Comparison

Method	Molecule	Target	Target Size (in humans)
DNA-Seq	DNA	Whole Genome	2 x 10^9 bp
RNA-Seq	RNA	Transcriptome	<3 x 10^7 bp
Amplicon	DNA?	Target Region	10 - 10,000bp

HTS Applications

- DNA-Seq
- RNA-Seq
- Amplicon Sequencing
- Many More
    - ChIP-Seq
    - Ribo-Seq
    - Hi-C
    - MethylC-Seq

Comparing Technologies

Method	Read length	Accuracy	Reads per run	Max Output	Cost ($/Mb)	Pros	Cons
Sanger	400-900 bp	99.9%	1	900 bp	$2400	Longer reads.	Expensive. Low Output
Illumina	600 bp (300bp PE)	99.9%	$20x10^9$	6000 Gb	$0.01	High yield per base cost	Equipment expense. Short reads
PacBio	>10kb ave. >40kb max	99%	$5x10^5$	10 Gb	$0.08	Very long reads	Homopolymer errors. Moderate Output. Equipment expense.
Nanopore	>100 kb N50 >1Mb Max	92%	$1x10^6$	5 Gb	$0.10	Very long reads Portable Cheap Equipment	Homopolymer errors. Moderate Output.

https://en.wikipedia.org/wiki/DNA_sequencing
https://blog.genohub.com/2017/06/16/pacbio-vs-oxford-nanopore-sequencing/

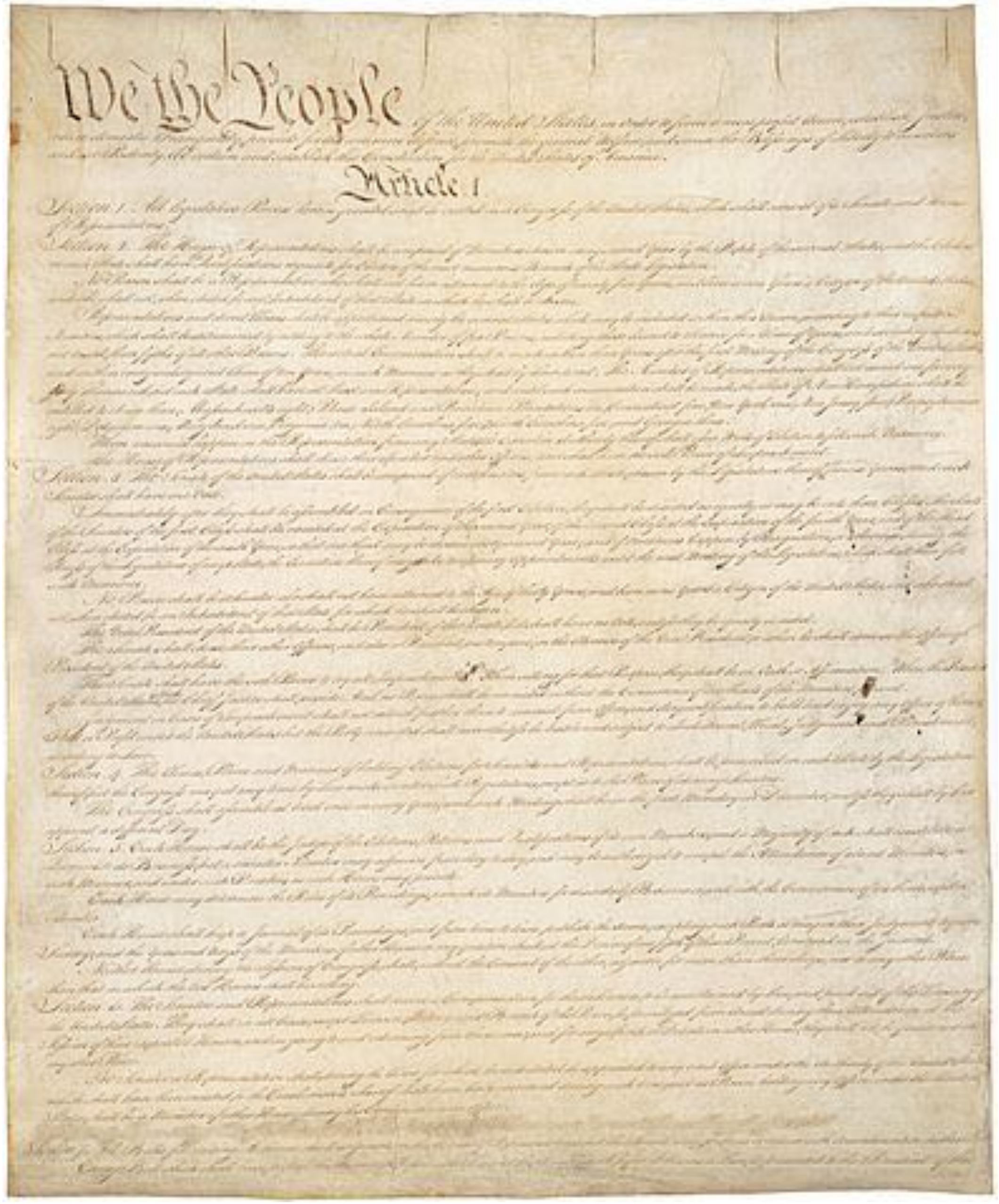
Why Long Reads?

- Structural Variation
    - Large Insertions or Deletions
    - Duplications
    - Translocations
- De Novo Genome Assembly
- Phasing

Short Reads

```
e of the U
stablish J
Union, est
nited Stat
 to form a
rder to fo
e perfect 
ion, estab
eople of t
 the Peopl
```

“Genome” Reference



We the People

Article 1

Reference Based Mapping

```
We the People of the United States, in Order to form a more perfect Union, establish Justice, insur
```

```
e of the U
stablish J
Union, est
nited Stat
 to form a
rder to fo
e perfect 
ion, estab
eople of t
 the Peopl
```

Reference Based Mapping

```
We the People of the United States, in Order to form a more perfect Union, establish Justice, insur



   the Peopl
        eople of t
            e of the U
                      nited Stat
                                        rder to fo
                                             to form a
                                                          e perfect
                                                                    Union, est
                                                                      ion, estab
                                                                            stablish J
```

De Novo Assembly

Overlapping Random Fragments

```
rious disg
Age.  "You
rinking Ag
uises of A
the portra
ugh the po
of every D
nking Age.
r various
, under va
```

Assemble Contigs

```
        Age.  “You
rinking Ag
  nking Age.
```

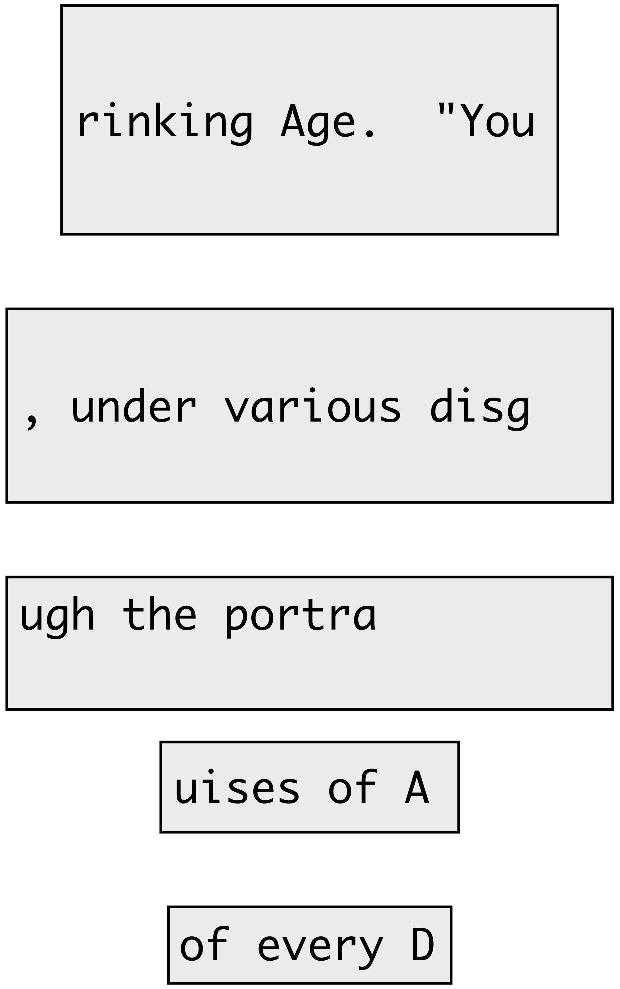
```
          rious disg
      r various
, under va
```

```
    the portra
ugh the po
```

```
uises of A
```

```
of every D
```

Assemble Contigs



rinking Age. "You

, under various disg

ugh the portra

uises of A

of every D

Assemble Contigs

rinking Age. "You

, under various disg

ugh the portra

uises of A

of every D

ed, under various disguises of Art, through the portraits of every Drinking Age. "You are a little

More Reads

```
rious disg
Age."You
rinking Ag
uises of A
the portra
ugh the po
of every D
nking Age.
r various
, under va
Age."Yo
rough the
rinking Ag
ed, under
ugh the po
ry Drinkin
sguises of
 u are a li
"You are
, under va
```

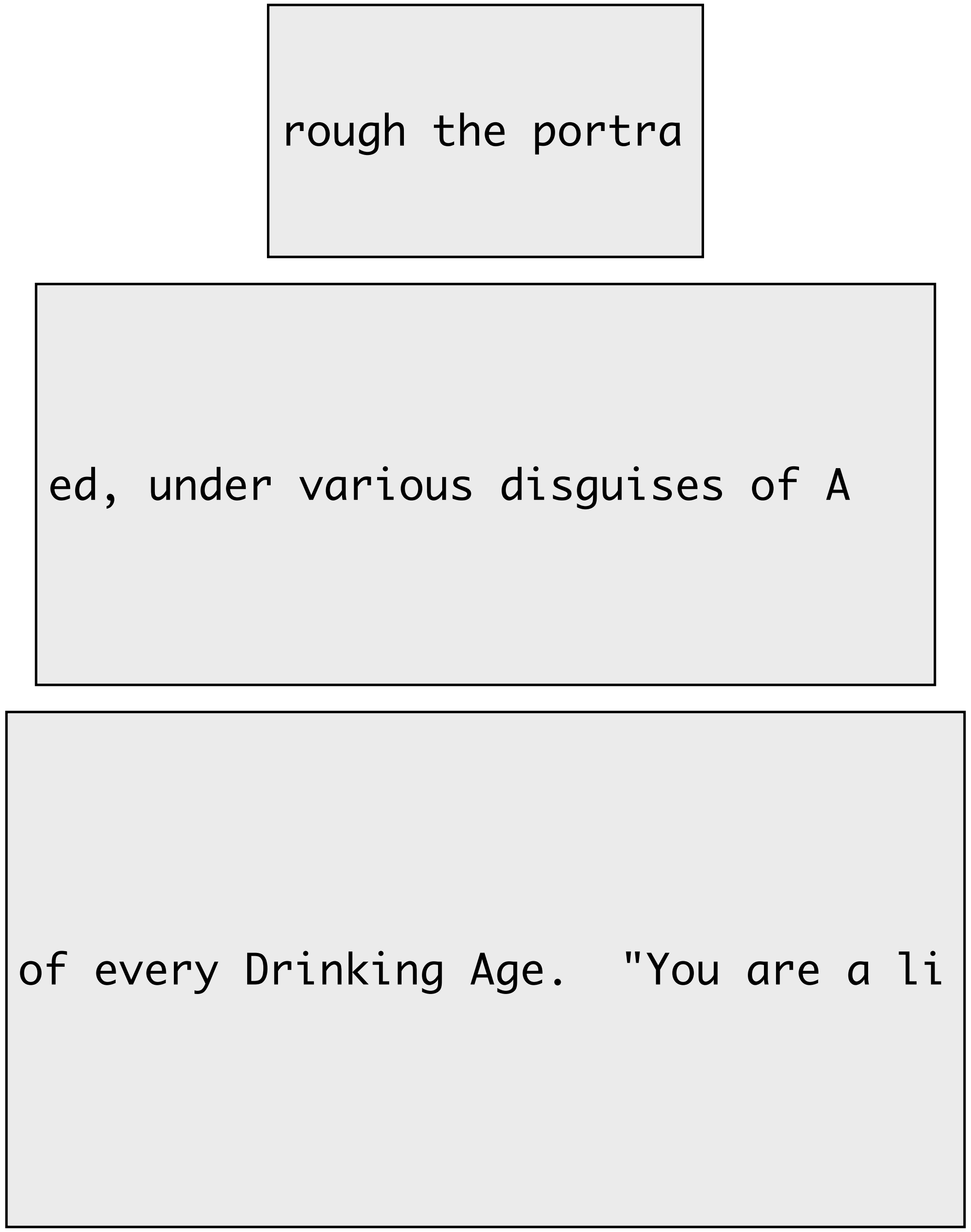
More Reads

```
rough the
  ugh the po
  ugh the po
      the portra
```

```
ed, under
  , under va
  , under va
        r various
            rious disg
                    sguises of
                      uises of A
```

```
of every D
     ry Drinkin
         rinking Ag
         rinking Ag
           nking Age.
                 Age.  "You
                 Age.  "Yo
                       "You are
                          u are a li
```

More Reads

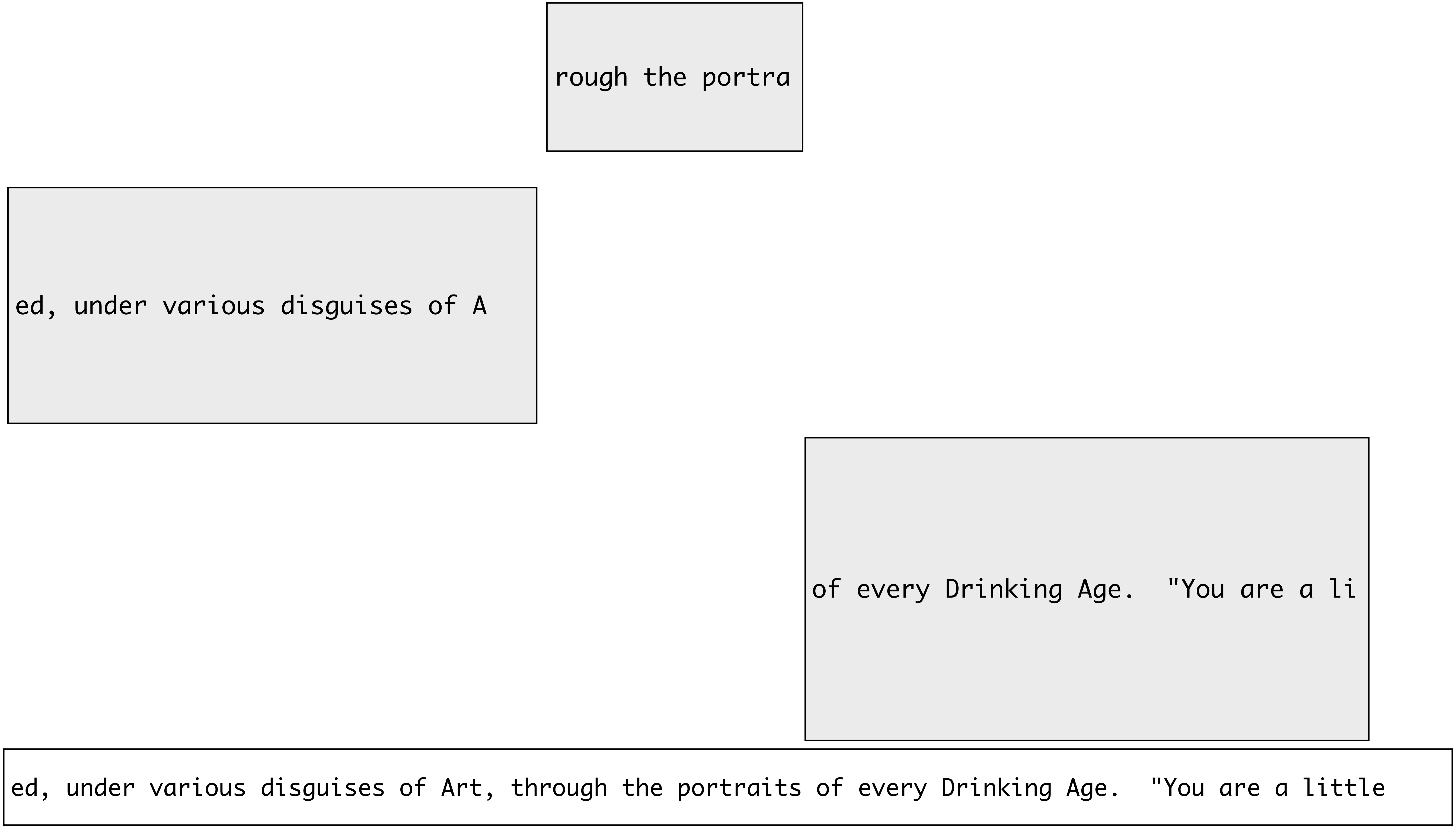


rough the portra

ed, under various disguises of A

of every Drinking Age. "You are a li

More Reads



rough the portra

ed, under various disguises of A

of every Drinking Age. "You are a li

ed, under various disguises of Art, through the portraits of every Drinking Age. "You are a little

Longer Reads

```
various disguises of
 Drinking Age."You
 every Drinking Age.
sguises of Art, thro
 ough the portraits o
through the portrai
raits of every Drink
 ery Drinking Age."
 er various disguises
, under various disg
```

Longer Reads

```
, under various disg
     er various disguises
        various disguises of
                  sguises of Art, thro
                                  through the portrai
                                     ough the portraits o
                                                  raits of every Drink
                                                           every Drinking Age.
                                                             ery Drinking Age.  "
                                                                 Drinking Age.  "You
```

Longer Reads

, under various disguises of Art, through the portraits of every Drinking Age. "You

ed, under various disguises of Art, through the portraits of every Drinking Age. "You are a little

Fragmentation

```
"You
Age.
Art,
Drinking
a
are
disguises
ed,
every
little
of
of
portraits
the
through
under
various
```

Problem Sequences

- Repeats
  - Transposons
  - Centromeres
- Homologs
- Duplications

De novo “Reference”

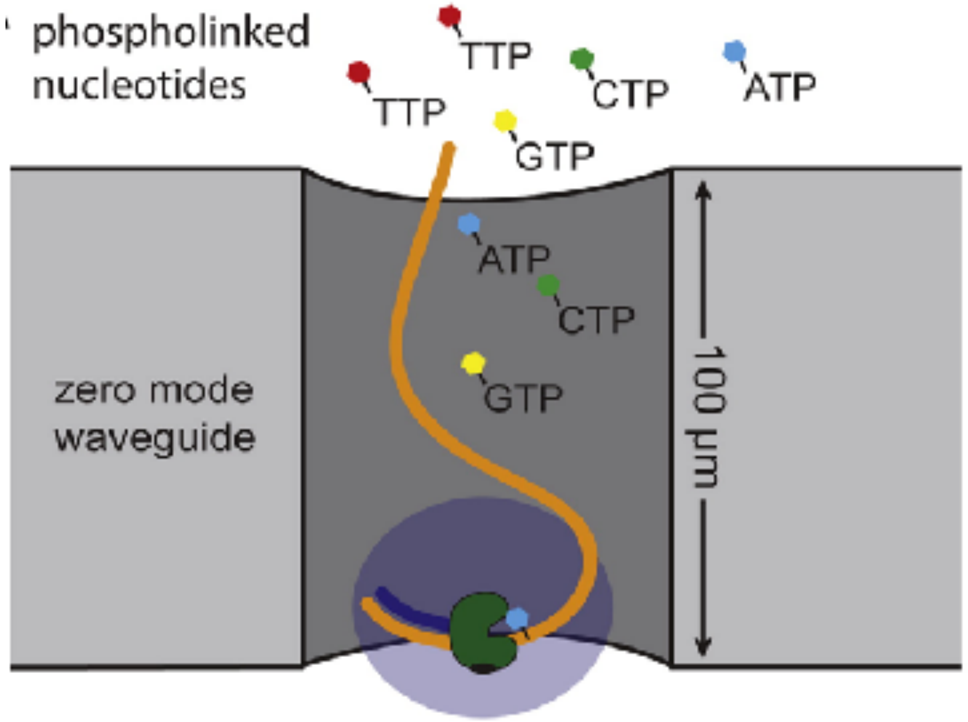
ed, under various disguises of Art, through the portraits of every Drinking Age. "You are a little

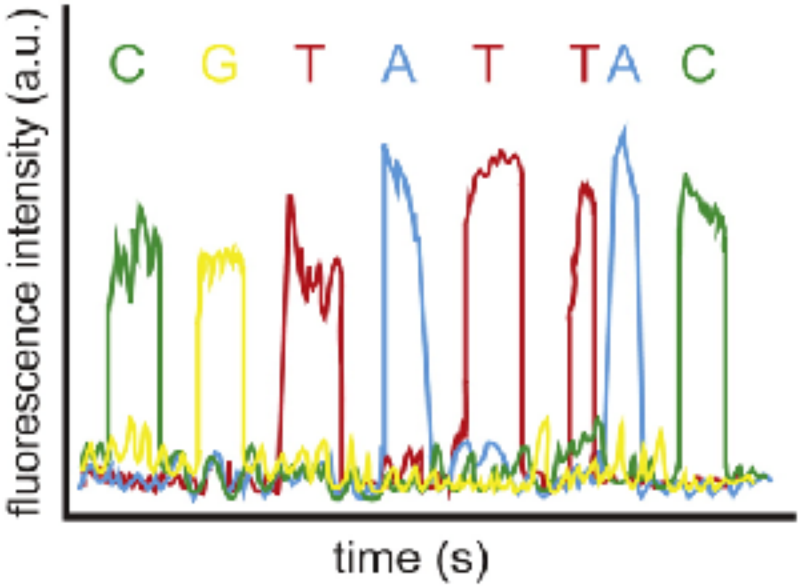
— *A Tale of Two Cities*

Single Molecule Technologies

1st Generation	2nd Generation	3rd Generation
Chemical (Maxim-Gilbert)	Pyrosequencing (454)	Single molecule real time (PacBio)
Chain Termination (Sanger)	Chain Termination (Illumina)	Nanopore sequencing (Oxford Nanopore)
Pyrosequencing	Sequencing by ligation (SOLiD sequencing)	
	Ion semiconductor (Ion Torrent)	

Pacific Biosciences

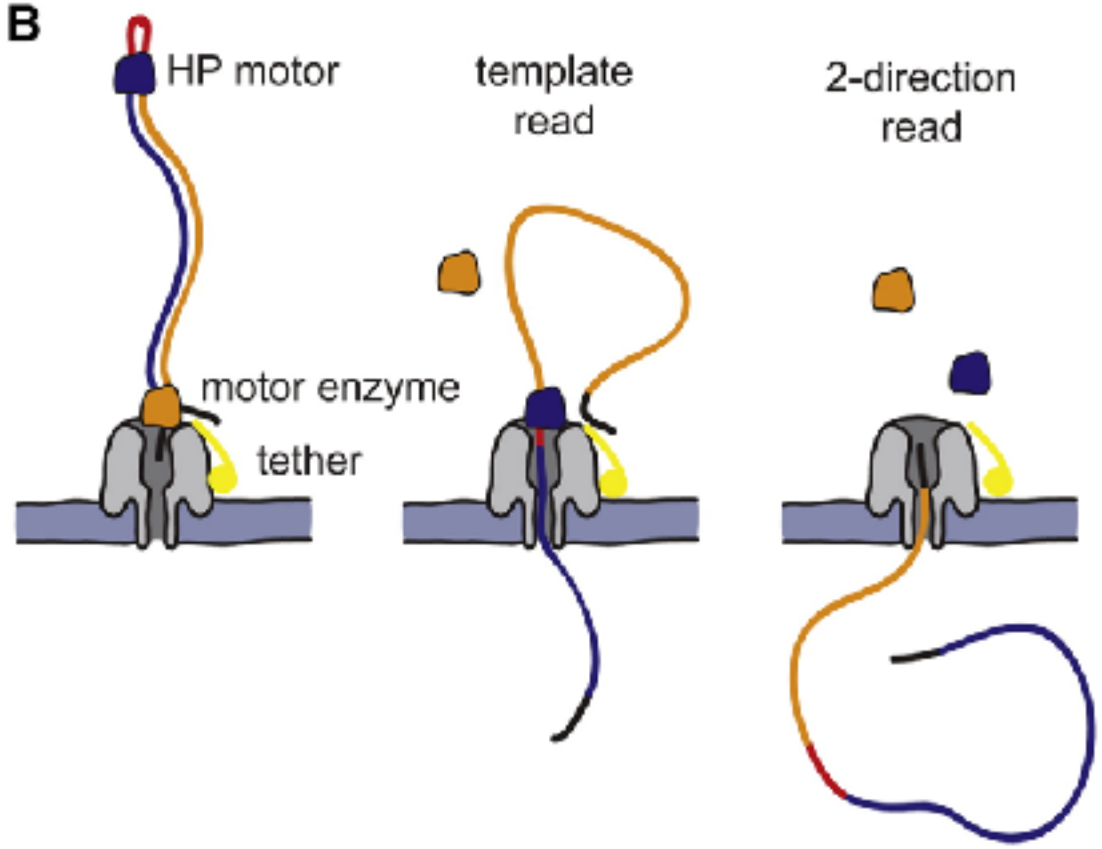


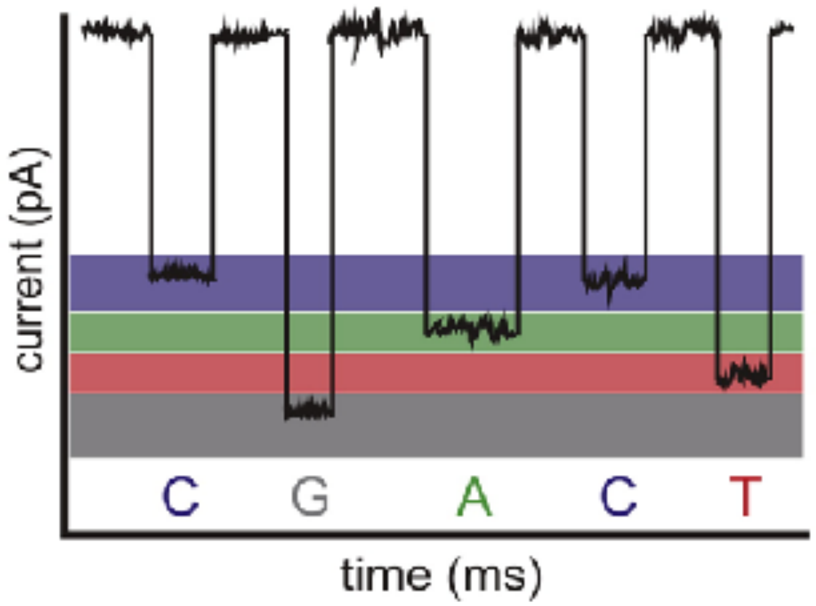


Reuter, et al 2015

1st Generation	2nd Generation	3rd Generation
Chemical (Maxim-Gilbert)	Pyrosequencing (454)	Single molecule real time (PacBio)
Chain Termination (Sanger)	Chain Termination (Illumina)	Nanopore sequencing (Oxford Nanopore)
Pyrosequencing	Sequencing by ligation (SOLiD sequencing)	
	Ion semiconductor (Ion Torrent)	

Oxford Nanopore





Reuter, et al 2015

Sequencers

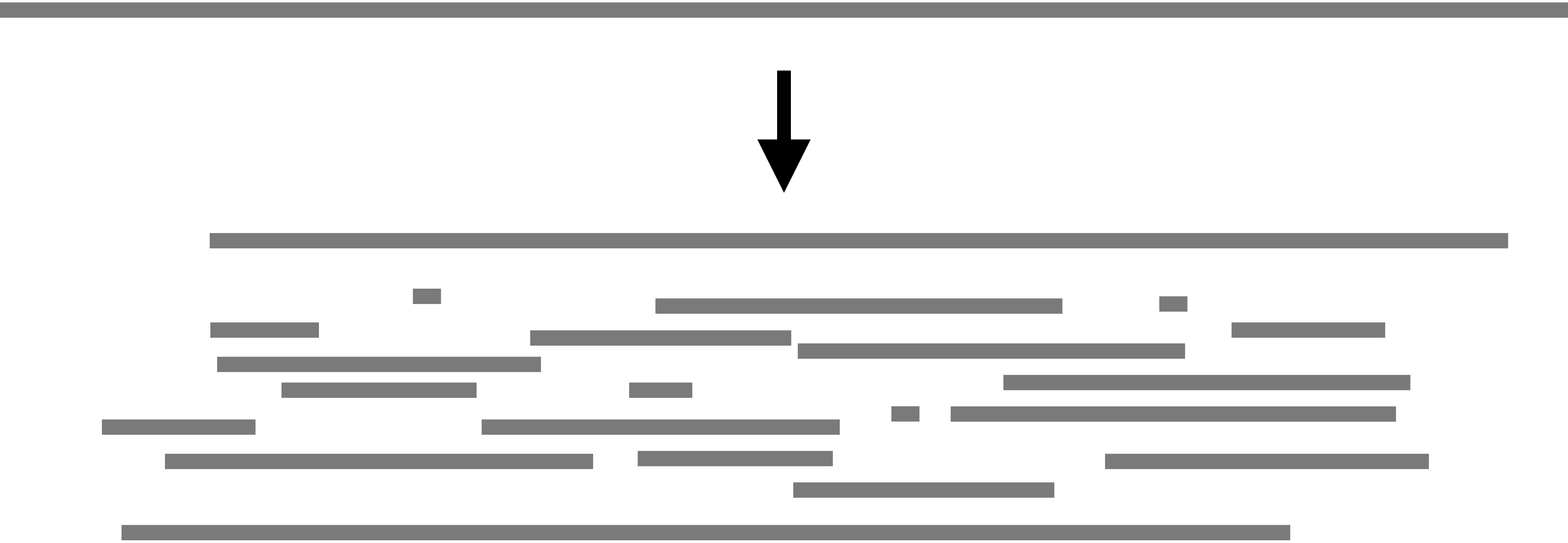
https://www2.nanoporetech.com/images/product-page/MinIOn-Banner.jpg
http://www.gatc-biotech.com/en/gatc/sequencing-technologies/pacbio-rs-ii.html
http://www.dnavision.com/illumina.php

IBIEM2018 Docker Image

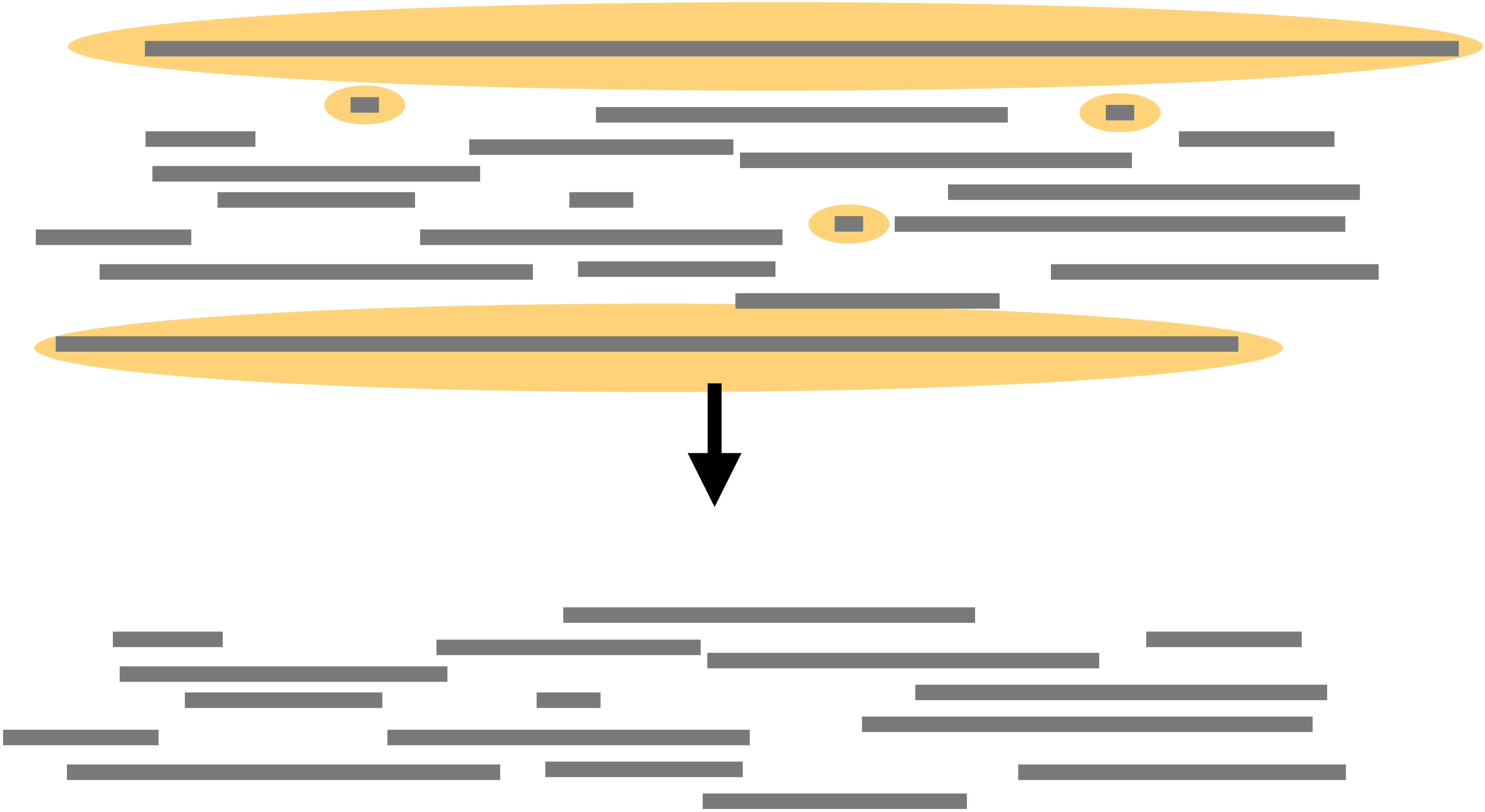
DNA-Seq Library Prep

Purified DNA

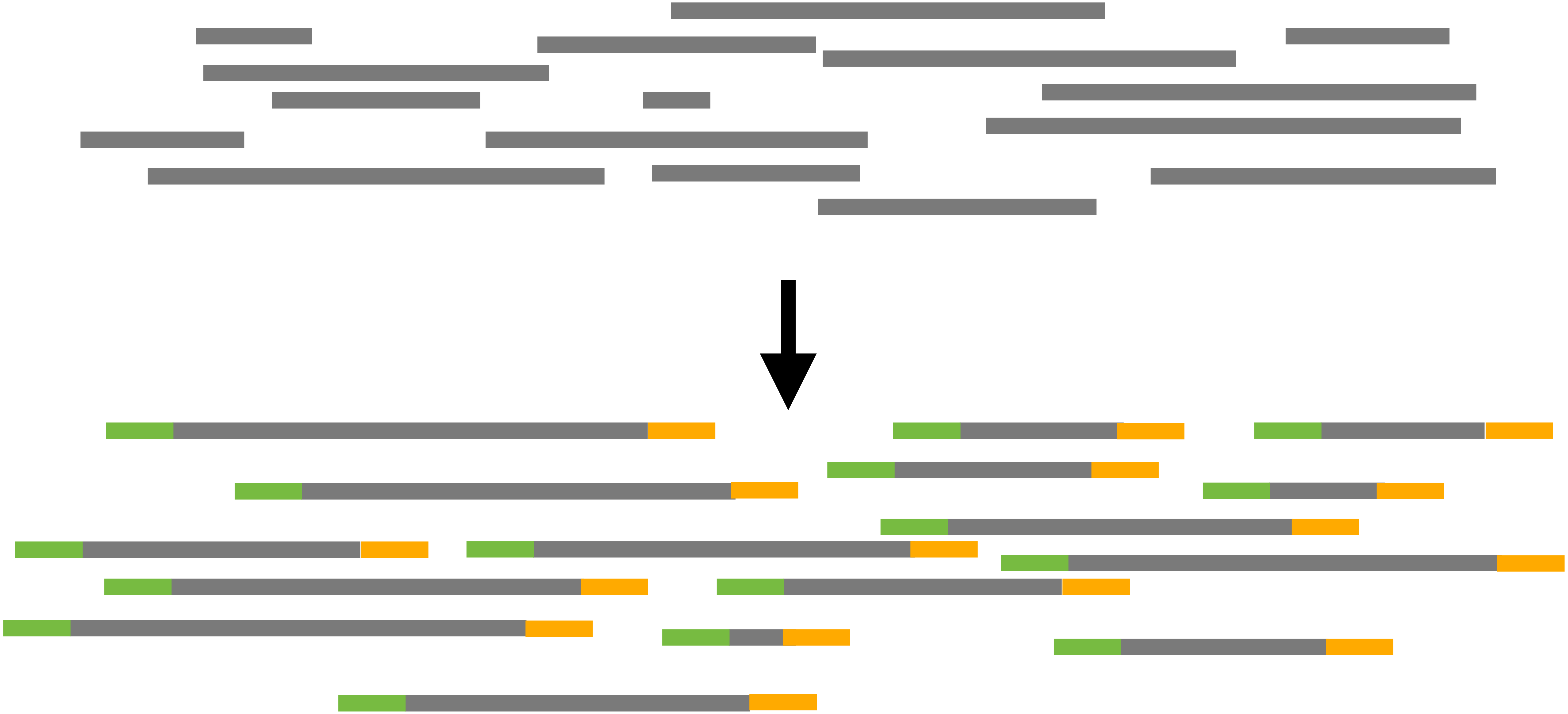
Fragmentation



Size Selection



Adapter Ligation



RNA-Seq Library Prep

DNA-Seq

1. Purify DNA
2. Fragment
3. Size Select
4. Adapter Ligation

RNA-Seq

1. Purify RNA
2. Fragment
3. Size Select
4. Make DNA From RNA
5. Adapter Ligation

Amplicon Library Prep

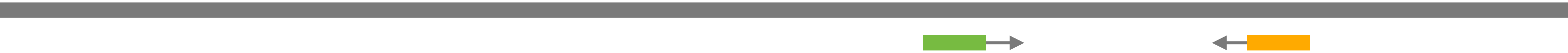
DNA-Seq

1. Purify DNA
2. Fragment
3. Size Select
4. Adapter Ligation

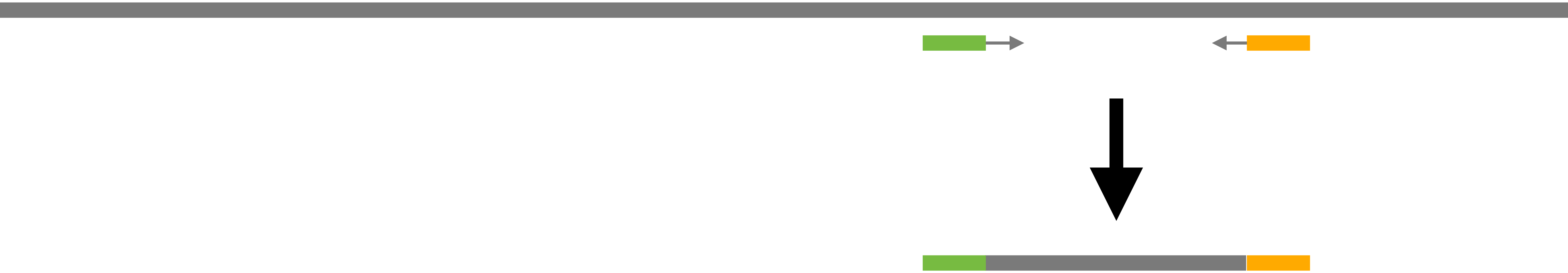
Amplicon-Seq

1. Purify DNA
2. PCR Amplify with Adapters

Purified DNA

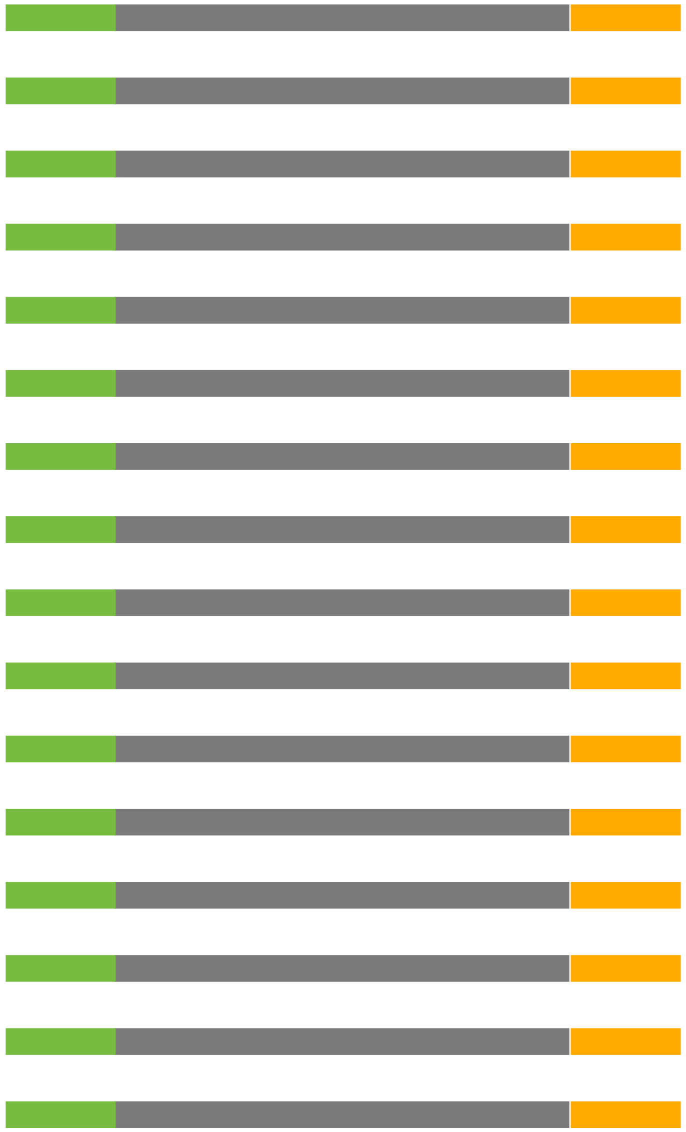


PCR Amplification



Sequencing Library

Amplicon Library



Shotgun Library