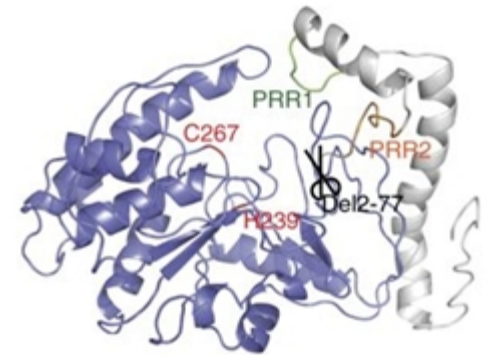


An effector protein that depletes phosphorous storage from plant cells

Lederson Gañán

Molecular studies have proven that plant pathogenic bacteria produce proteins that can be harmful for the host plants. These proteins are referred as effectors and they are delivered to the plant by secretion systems that works like a molecular syringe. Once delivered to the host cells, these effectors can either facilitate the infection process or manipulate host's metabolic processes. Blüher et al. studied the effector XopH secreted by *Xanthomonas* which is a bacteria that primarily infects tomato and pepper plants. According to the research results, XopH targets phytate which is the major phosphate storage compound in plants. It is suggested that XopH not only depletes this source of phosphorus, but also prepares the plant to receive the harmful bacteria by weakening the plant's defense. Changes in the host hormonal balance are suggested as the affected plants show problems to grow. It is also indicated that some hosts have adapted to recognize this effector and therefore a response of resistance can be observed which prevents the pathogenic bacteria from spreading further through the plant tissue, although the mechanisms are not well known yet. The study provides another answer to the question of how bacteria harm plants and how plants react to this threat.



From the News Story: "How bacteria manipulate plants."
ScienceDaily. 21 February 2018.

www.sciencedaily.com/releases/2018/02/180221122853.htm

Original scientific paper: Blüher, D., Laha, D., Thieme, S., Hofer, A., Eschen-Lippold, L., Masch, A., et al. (2017). A 1-phytase type III effector interferes with plant hormone signaling. *Nature communications*, 8(1), 2159.

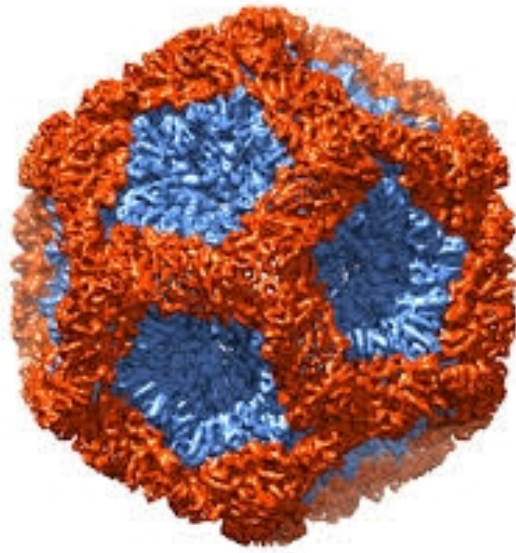
<https://www.nature.com/articles/s41467-017-02195-8>

A model of the XopH protein structure (Source: Figure 1b from original scientific paper by Blüher et al. 2017)

Announcements

- Phytobacteriology in the News [Week 4](#)
- News & Views! First draft due TODAY! **Thurs Sept 13th** (midnight), Please email to Maren
- Paper for next week:
 - Microbially-Mediated Plant Functional Traits
 - Form [here](#)

Phenomics Symposium Next Tuesday!



Guest lecturer: Dr. Richard Allen White III **Post-**
doc: Friesen lab

Learning Objectives

1. File formats
2. Assembly
3. Assembly statistics
4. Discuss paper

How to obtain a genome?

5

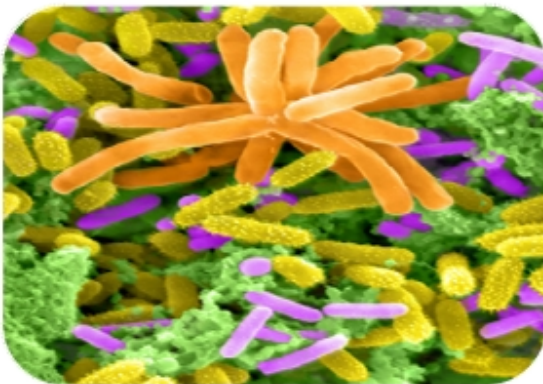


Culturing $< 1\%$ can be easily cultured



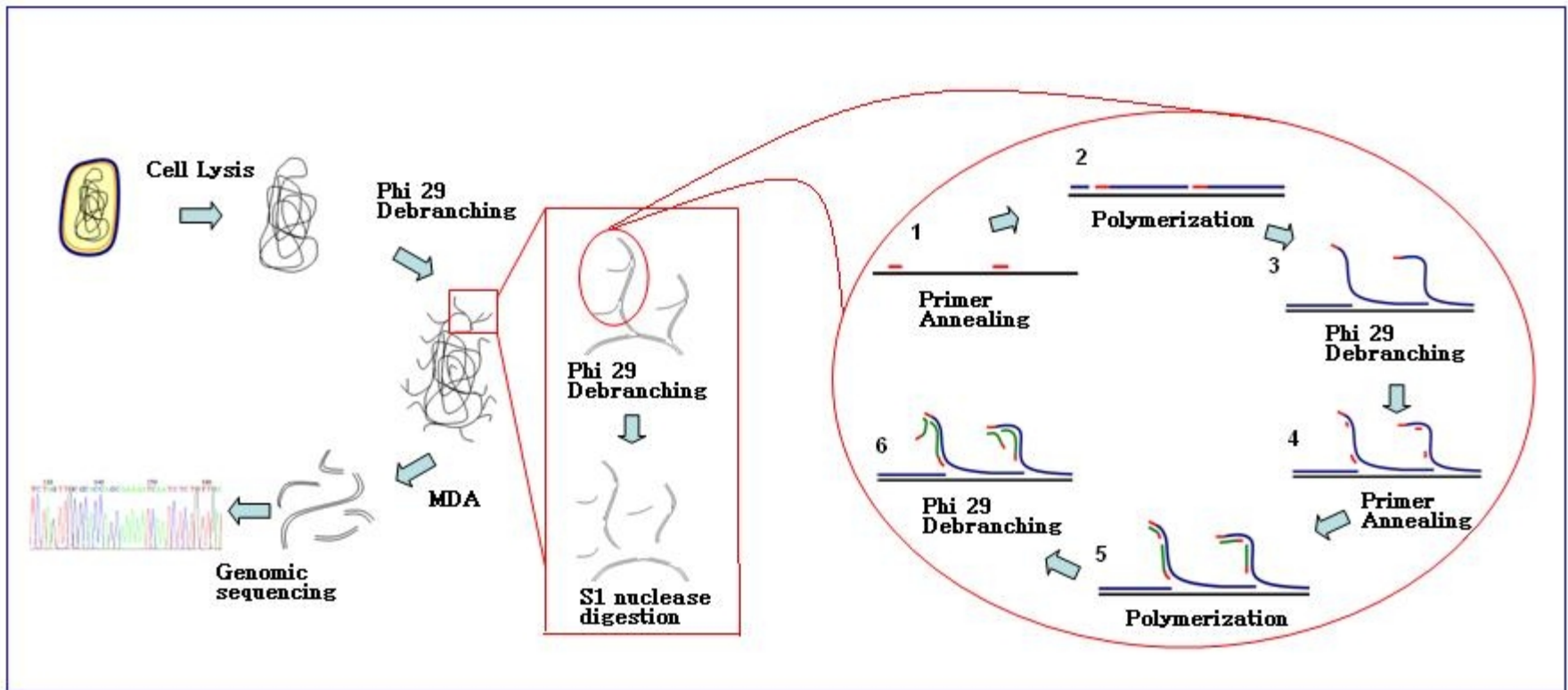
Single cell genomics

Only a few labs in the world, very incomplete genomes (30% avg, 10-90%)



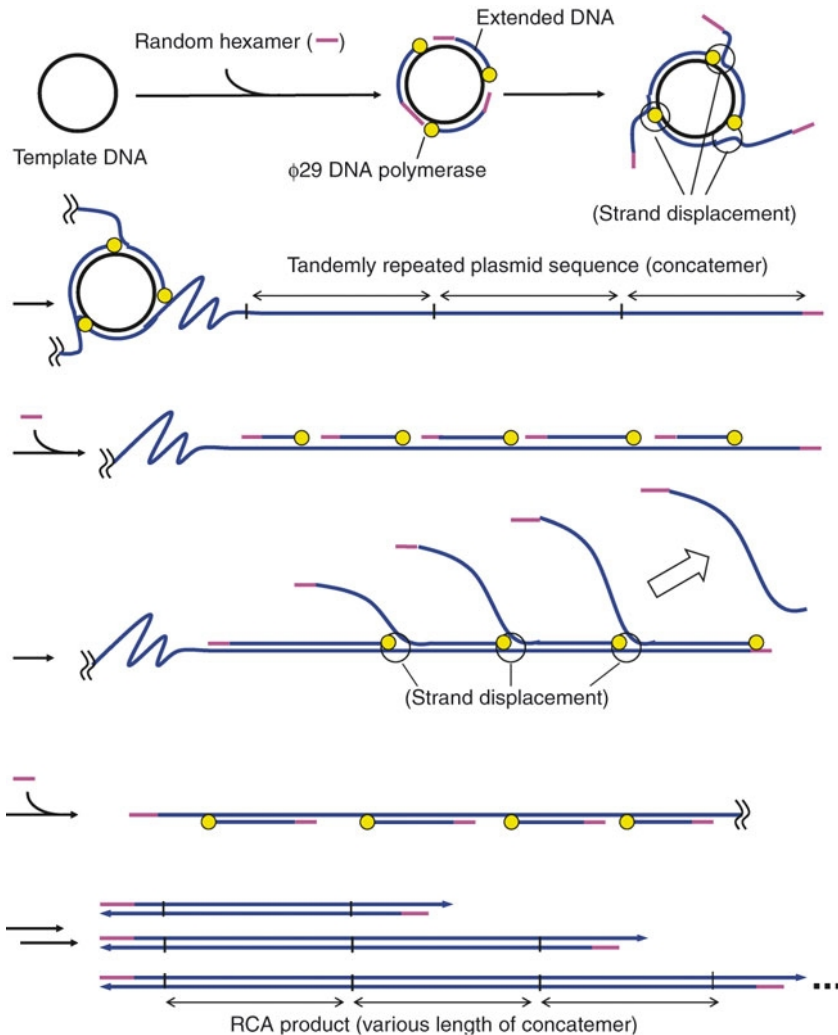
Metagenomic - population genome binning

Single cell genomics WGA/



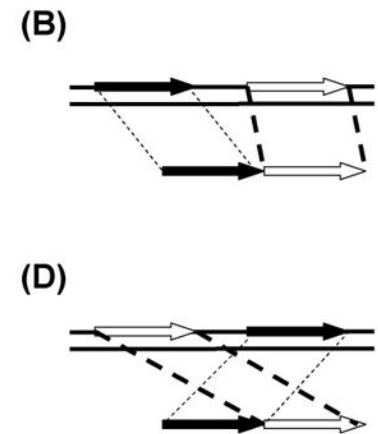
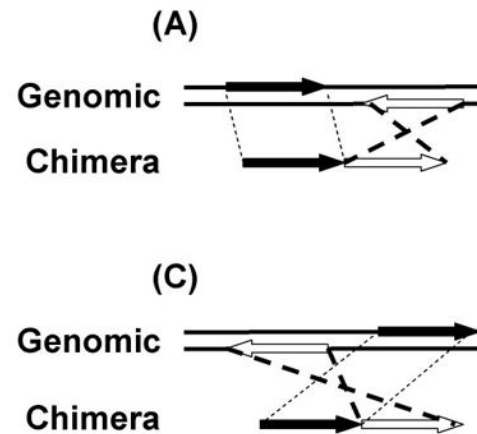
WGA via MDA for single

7



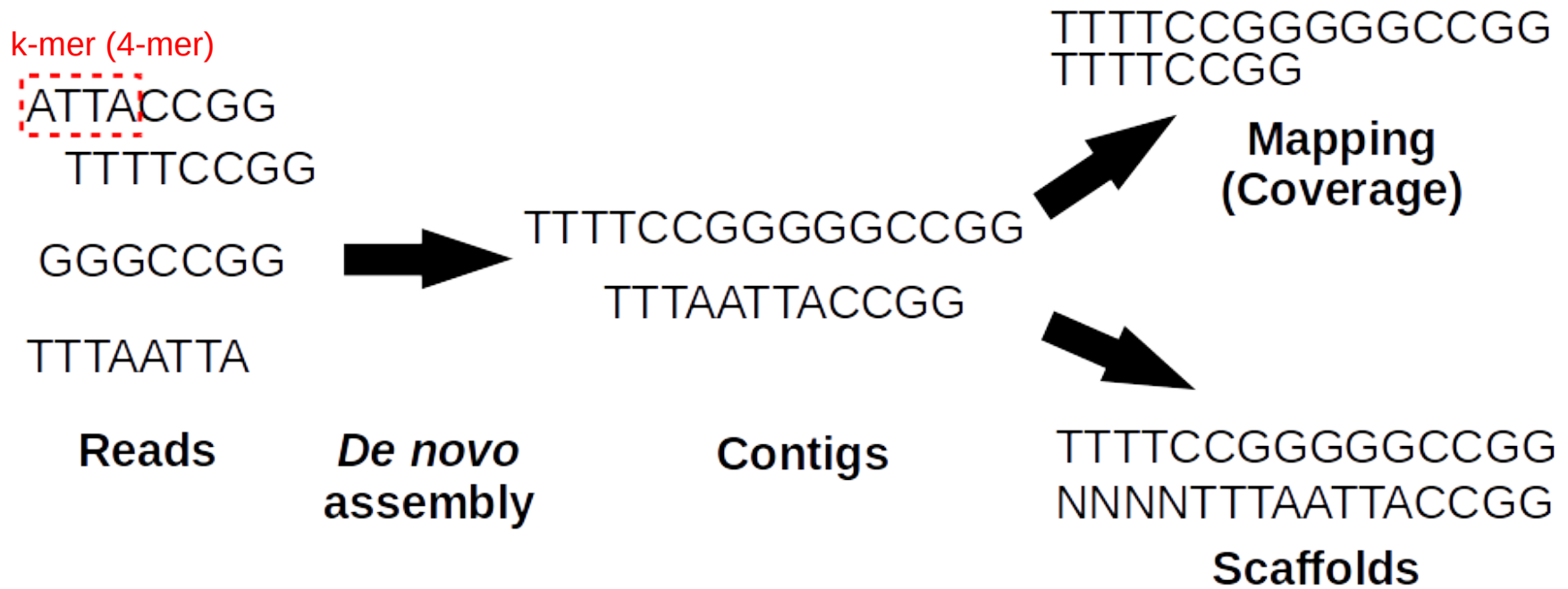
85% Inverted sequences

15% Direct sequences



<https://bmcbiotechnol.biomedcentral.com/articles/10.1186/1472-6750-7-19>

De novo assembly terms



Metagenomic (DNA) and metatranscriptomic (RNA) assembly is the process of taking a large number of short sequences from many different organisms then reconstructing them into larger sequences (i.e contigs)

Finished bacterial genomes

contigs = # replicons

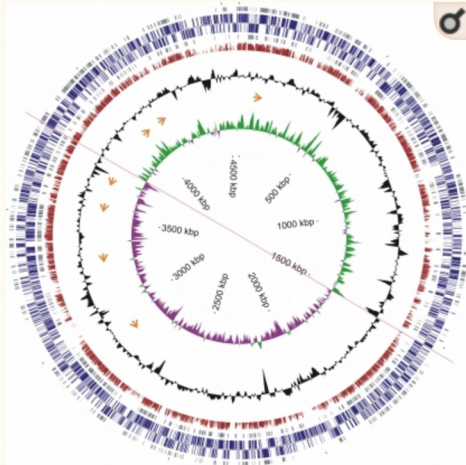
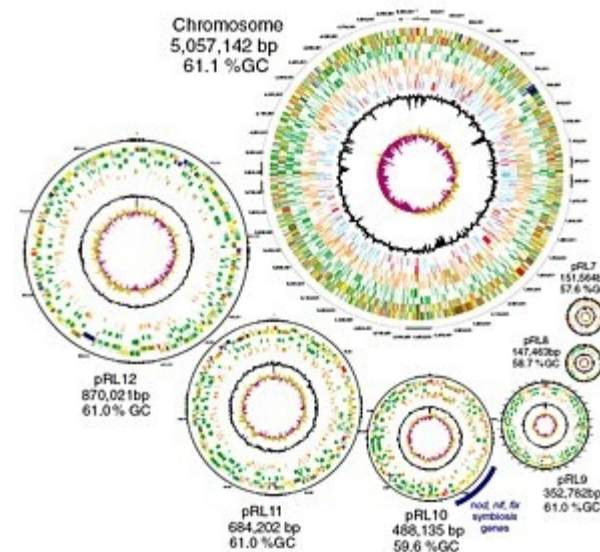


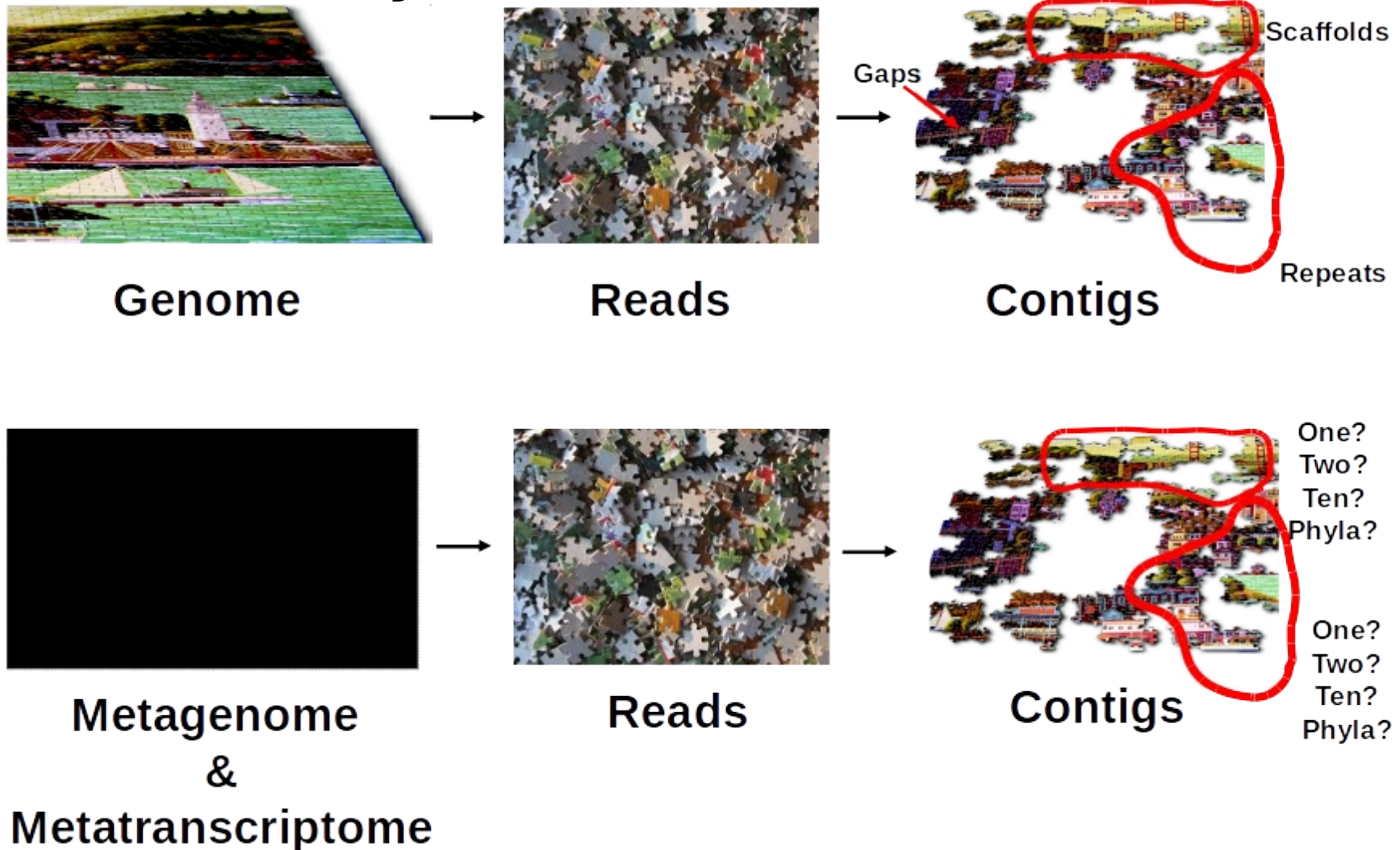
Fig. 1

Circular representation of the *P. ananatis* AJ13355 chromosome. Circles are numbered from 1 (outermost) to 8



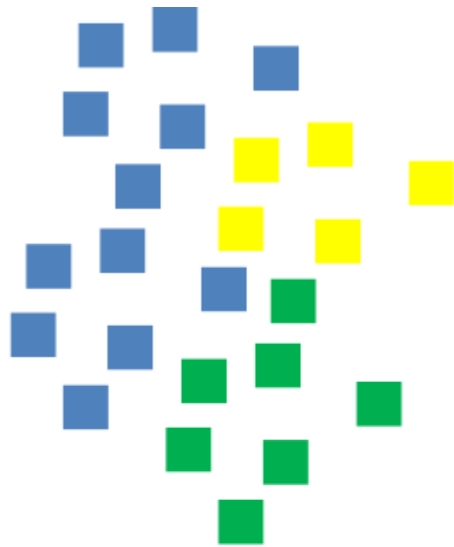
Not finished → “DRAFT” genome., i.e., when there are more contigs than replicons (you can do pulse-field gels to figure this out)

De novo assembly genome vs. community



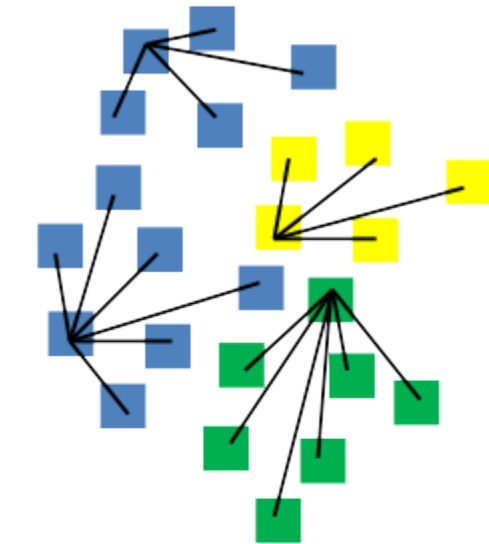
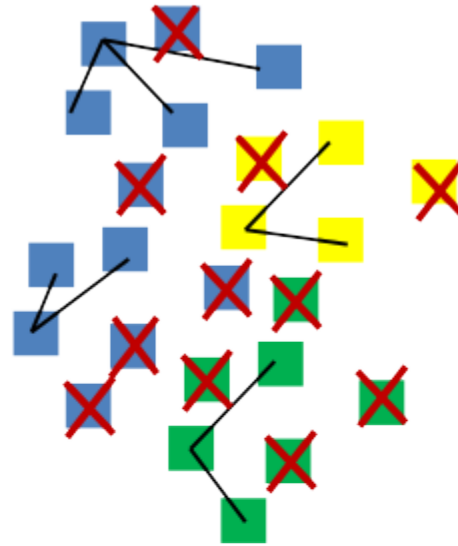


Metagenomic assembly of soil



Divide and conquer

9 contigs >10 kbp, 10% mapping¹



Store succinct and go

100 contigs >10 kbp, 30% mapping²

- No genomes binned from soil (only a few from lower diversity permafrost)
- Few long contigs from soils means full length 16S rRNAs and genomic bins cannot be reconstructed
- Software isn't enough to over-come the short comings of short reads

¹Howe *et al.*, 2014 PNAS.

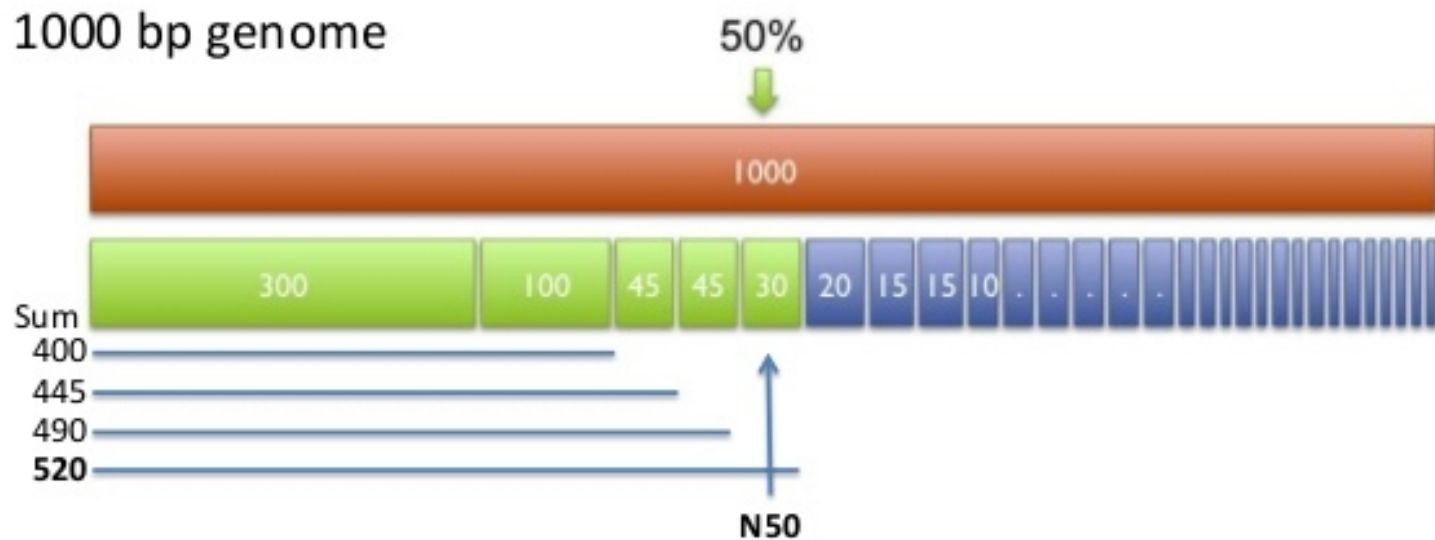
²Li *et al.*, 2015 Bioinformatics.

Assembly statistics

- No. of contigs
- Max contig length
- Total assembly length
- N50
 - That 50% of the entire assembly is contained in contigs or scaffolds equal to or larger than this value
- N90
 - Is the length for which the collection of all contigs of that length or longer contains at least 90% of the sum of the lengths of all contigs, and for which the collection of all contigs of that length or shorter contains at least 10% of the sum of the lengths of all contigs
- L50
 - Is defined as the smallest number of contigs whose length sum produces N50
- NG50 (not used that often)
 - Similar to N50 but based on assembly size rather than the genome size. Is the same as N50 except that it is 50% of the known or estimated genome size that must be of the NG50 length or longer.

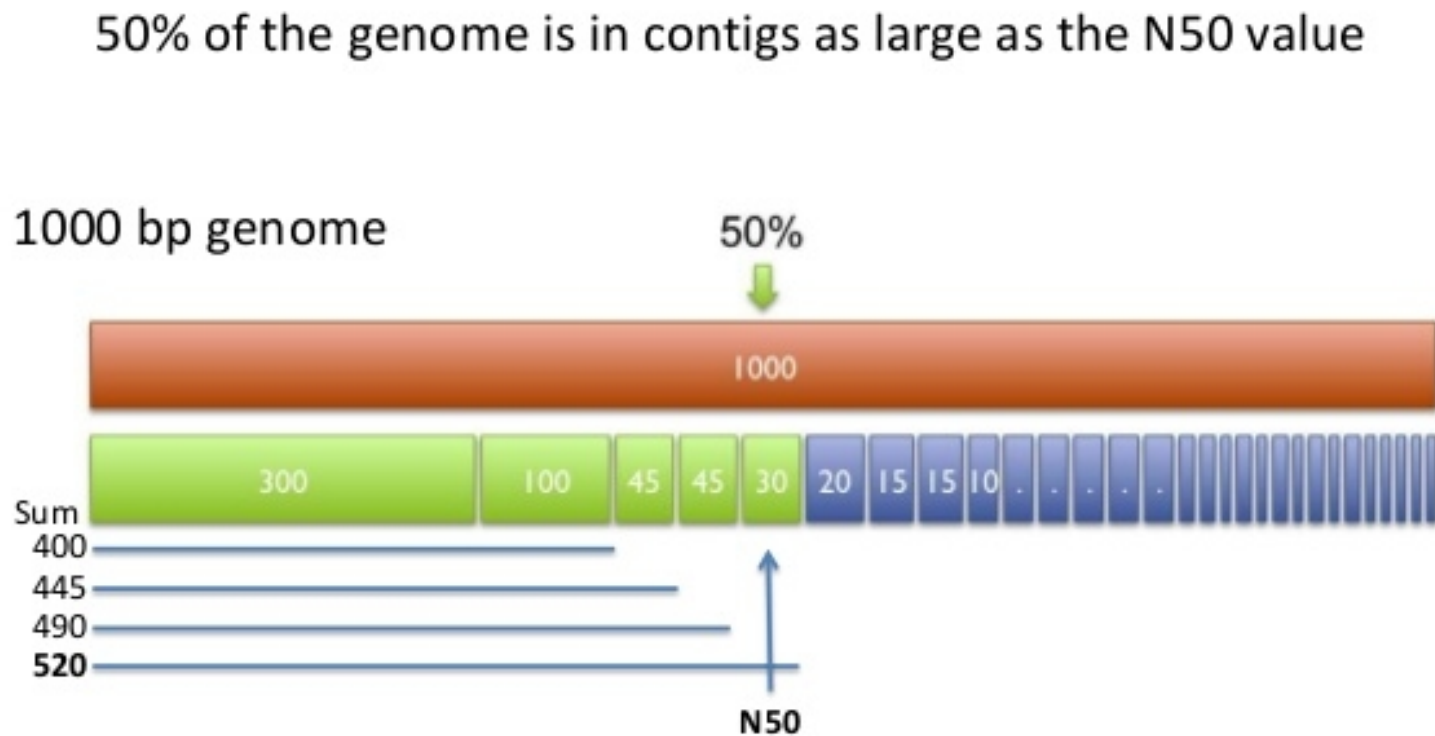
N50 explained visually

50% of the genome is in contigs as large as the N50 value



Courtesy of Michael Schatz, CSHL

N50: Test your knowledge! What is the N50 for an¹⁵
E coli strain with a finished genome 5Mbp large?



Courtesy of Michael Schatz, CSHL

Common types of files

*.fna, .fasta, .fa	sequence
*.faa	sequence (amino acid)
*.gff	gene feature format (annotation)
*.gbk	genbank format (annotation)
*.fq or fastq	sequence + quality
*.gtf	Gene transfer format (annotation)

File formats in genomics

(fasta)

>sequence1

ATACTCTACTCGTCTCATATCAT

>sequence2

GCGCGCGN CAGCGATCTCTCA

>sequence3

TTTCGCGNN CAGCGATCTCTC

File formats in genomics (gff/gtf)

GFF

```
X Ensembl Repeat 2419108241912842 . .  
hid=trf; hstart=1; hend=21
```

GTF

```
1 transcribed_unprocessed_pseudogene gene  
11869 14409 . + . gene_id "ENSG00000223972";  
gene_name "DDX11L1"; gene_source "havana";  
gene_biotype  
"transcribed_unprocessed_pseudogene";
```

Let's look at the Pantoea

Genome > **Genome Information by Organism**

<https://www.ncbi.nlm.nih.gov/genome/browse#!/overview/pantoea>

pantoea



Search

[Download Reports from FTP site](#)

Overview (25); [Eukaryotes \(0\);](#) [Prokaryotes \(227\);](#) [Viruses \(0\);](#) [Plasmids \(0\);](#) [Organelles \(0\)](#)

Filters

Download

#	Organism Name	Organism Groups	Size(Mb)	Chromosome	Organelles	Plasmids	Assemblies
1	Candidatus Pantoea carbekii	Bacteria;Proteobacteria;Gammaproteobacteria	1.19695	1	-	-	2
2	Pantoea	Bacteria;Proteobacteria;Gammaproteobacteria	6.31278	1	-	5	68
3	Pantoea agglomerans	Bacteria;Proteobacteria;Gammaproteobacteria	5.11524	1	-	3	29
4	Pantoea alhagi	Bacteria;Proteobacteria;Gammaproteobacteria	4.3163	1	-	1	1
5	Pantoea allii	Bacteria;Proteobacteria;Gammaproteobacteria	5.24114	-	-	-	4
6	Pantoea ananatis	Bacteria;Proteobacteria;Gammaproteobacteria	4.70337	1	-	-	55
7	Pantoea anthophila	Bacteria;Proteobacteria;Gammaproteobacteria	4.60068	-	-	-	1
8	Pantoea brenneri	Bacteria;Proteobacteria;Gammaproteobacteria	4.93318	-	-	-	1
9	Pantoea calida	Bacteria;Proteobacteria;Gammaproteobacteria	4.30845	1	-	-	2
10	Pantoea coffeiphila	Bacteria;Proteobacteria;Gammaproteobacteria	5.87688	-	-	-	1
11	Pantoea conspicua	Bacteria;Proteobacteria;Gammaproteobacteria	5.12279	-	-	-	2
12	Pantoea cypripedii	Bacteria;Proteobacteria;Gammaproteobacteria	6.55382	-	-	-	1
13	Pantoea deleyi	Bacteria;Proteobacteria;Gammaproteobacteria	4.61385	-	-	-	1
14	Pantoea dispersa	Bacteria;Proteobacteria;Gammaproteobacteria	4.76458	-	-	-	12
15	Pantoea eucrina	Bacteria;Proteobacteria;Gammaproteobacteria	3.93988	-	-	-	2
16	Pantoea gaviniae	Bacteria;Proteobacteria;Gammaproteobacteria	4.52761	1	-	-	2
17	Pantoea latae	Bacteria;Proteobacteria;Gammaproteobacteria	4.96042	-	-	-	1
18	Pantoea rodasii	Bacteria;Proteobacteria;Gammaproteobacteria	5.75146	-	-	-	3
19	Pantoea rwandensis	Bacteria;Proteobacteria;Gammaproteobacteria	4.32761	1	-	-	2
20	Pantoea septica	Bacteria;Proteobacteria;Gammaproteobacteria	4.30872	-	-	-	8

Let's look at the Pantoea

BioSample	BioProject	Assembly	Level	Size (Mb)	GC%	Replicons	WGS	Scaffolds	Gene	Protein	Release Date
SAMN06329872	PRJNA374633	GCA_002101395.1	●	4.3163	53.54	chromosome:NZ_CP019706.1/CP019706.1 plasmid pPALTyr11Z:NZ_CP019707.1/CP019707.1	-	-	4103	3880	2017/04/20

[Organism Overview](#) ; [Genome Assembly and Annotation report](#)

Pantoea alhagi

Pantoea alhagi Genome sequencing and assembly

Lineage: [Bacteria\[22698\]](#); [Proteobacteria\[7158\]](#); [Gammaproteobacteria\[2685\]](#); [Enterobacterales\[467\]](#); [Erwiniaceae\[52\]](#); [Pantoea\[25\]](#); [Pantoea alhagi\[1\]](#)

Summary

Submitter: Northwest A&F University
Assembly level: Complete Genome
Assembly: [GCA_002101395.1](#) ASM210139v1 **scaffolds: 2 contigs: 2 N50: 4,239,970 L50: 1**
BioProjects: [PRJNA374633](#)
Statistics: total length (Mb): 4.3163
protein count: 3880
GC%: 53.5399

Replicon Info

Type	Name	RefSeq	INSDC	Size (Mb)	GC%	Protein	rRNA	tRNA	Other RNA	Gene	Pseudogene
Chr	-	NZ_CP019706.1	CP019706.1	4.24	53.6	3,820	22	84	10	4,037	101
Plasm	pPALTyr11Z	NZ_CP019707.1	CP019707.1	0.08	50.2	60	-	-	-	66	6

Let's look at the Pantoea

Pantoea alhagi strain LTYR-11Z, complete genome

NCBI Reference Sequence: NZ_CP019706.1

[FASTA](#) [Graphics](#)

Go to: 

LOCUS NZ_CP019706 4239970 bp DNA circular CON 22-APR-2017
 DEFINITION Pantoea alhagi strain LTYR-11Z, complete genome.
 ACCESSION NZ_CP019706
 VERSION NZ_CP019706.1
 DBLINK BioProject: [PRJNA224116](#)
 BioSample: [SAMN06329872](#)
 Assembly: [GCF_002101395.1](#)
 KEYWORDS RefSeq.
 SOURCE Pantoea alhagi
 ORGANISM [Pantoea alhagi](#)
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacterales;
 Erwiniaceae; Pantoea.
 REFERENCE 1 (bases 1 to 4239970)
 AUTHORS Zhang,L.
 TITLE Complete genome sequence of the drought resistance-promoting
 endophyte Pantoea alhagi LTYR-11Z
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 4239970)
 AUTHORS Zhang,L.
 TITLE Direct Submission

Customize view 

Analyze this sequence

Run BLAST

Pick Primers

Highlight Sequence Features

Related information

Assembly

BioProject

BioSample

Protein

Taxonomy

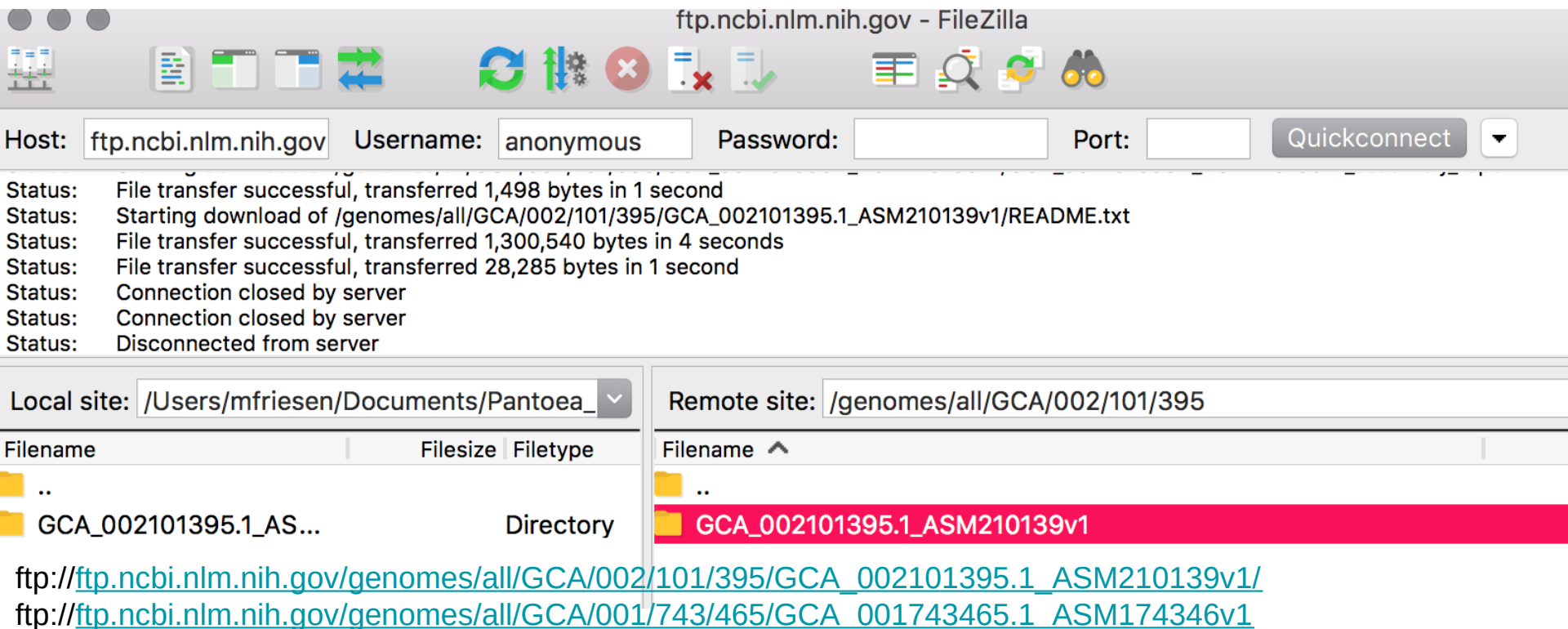
Components (Core)

Genome

Identical GenBank Sequence

Moving files via GUI

Maren likes Filezilla



Let's look at the Pantoea genomes!

- Get your command line out and ssh into the friesen lab machine
- Check out the folder of genomes!
- Copy (cp) one to a new subfolder in your personal folder in PLP514_2018
- Investigate the files!
- How many contigs are there per genome? How many genes? How many bases?
 - Hint: use the “man” command to figure out what “wc” does... it could be very helpful!

Paper review

De Maayer et al. *BMC Genomics* 2014, **15**:404
<http://www.biomedcentral.com/1471-2164/15/404>



RESEARCH ARTICLE

Open Access

Analysis of the *Pantoea ananatis* pan-genome reveals factors underlying its ability to colonize and interact with plant, insect and vertebrate hosts

Pieter De Maayer^{1,2*}, Wai Yin Chan², Enrico Rubagotti³, Stephanus N Venter², Ian K Toth^{2,4}, Paul R J Birch^{2,4,5} and Teresa A Coutinho²

Evaluate the data decisions

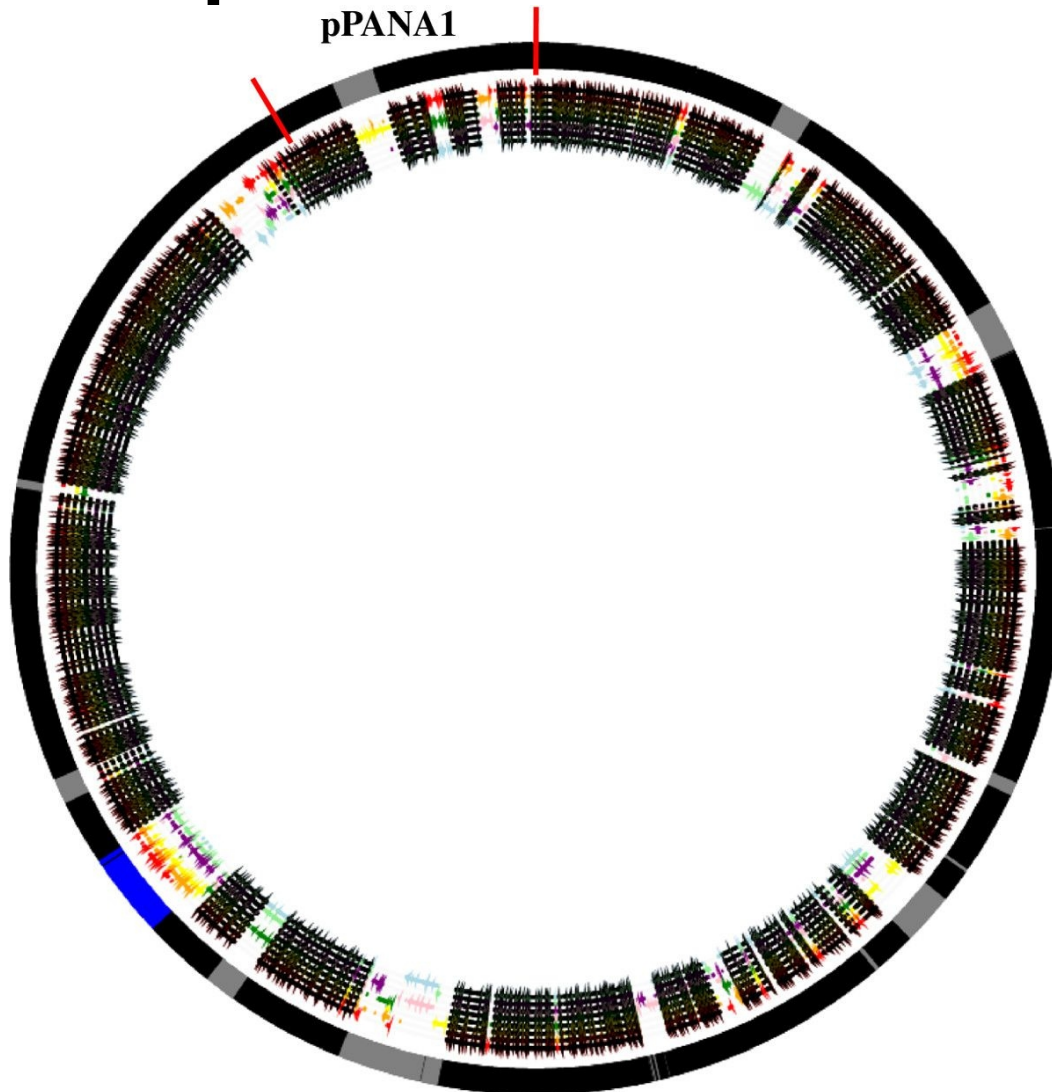
- What kind of sampling biases might be present in this dataset?
- Evaluate the sequencing strategy and data analysis
- In your opinion, is this a rigorous study?

Paper review - Terms

- Core genome?
- Pan genome?
- Accessory genome?
- Core vs. Pan
- Accessory vs. Pan
- Mobilome?

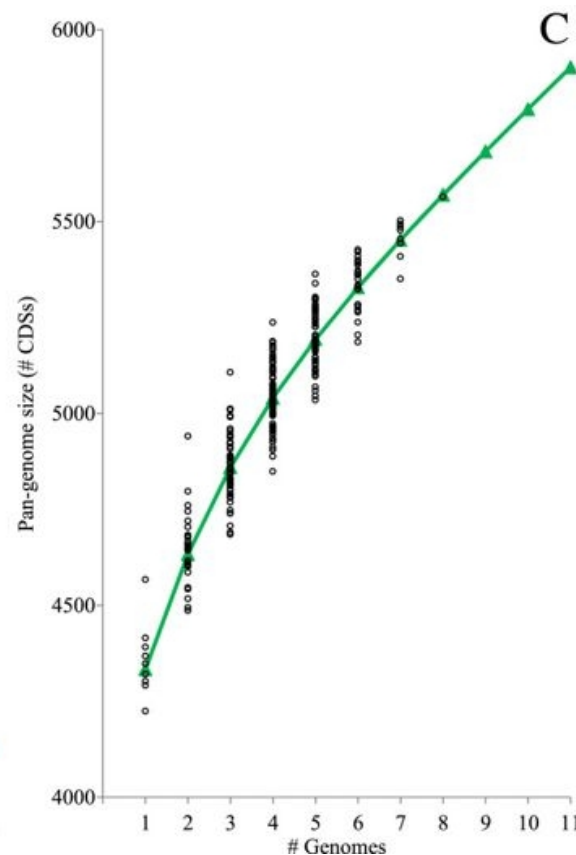
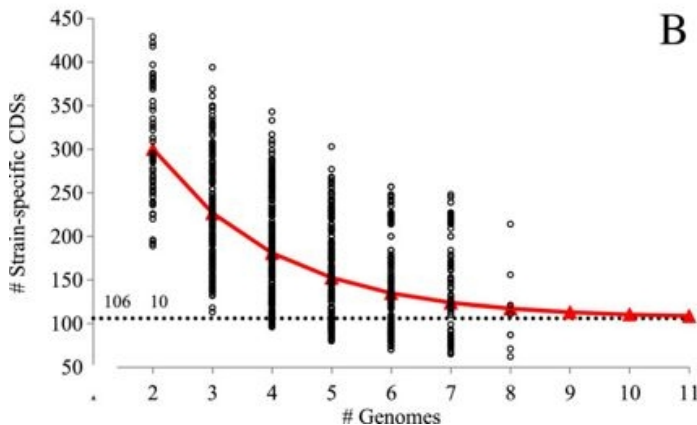
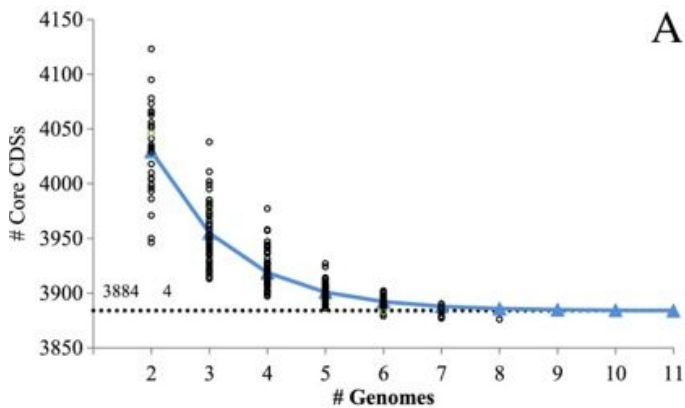
What is the key hypothesis of this paper?

Paper review - figure 1



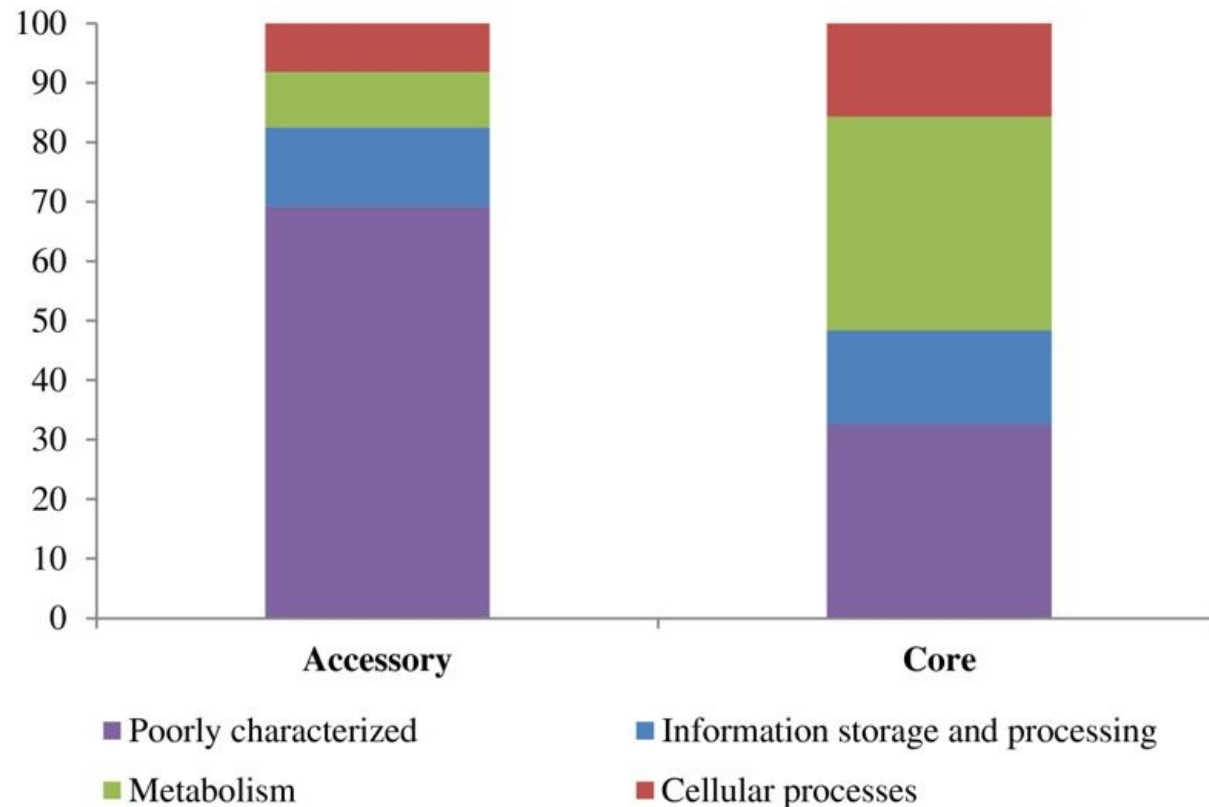
- What is this figure showing?
- Is the program correct for visualizing the data?
- Strengths and weaknesses of this figure

Paper review - figure 2



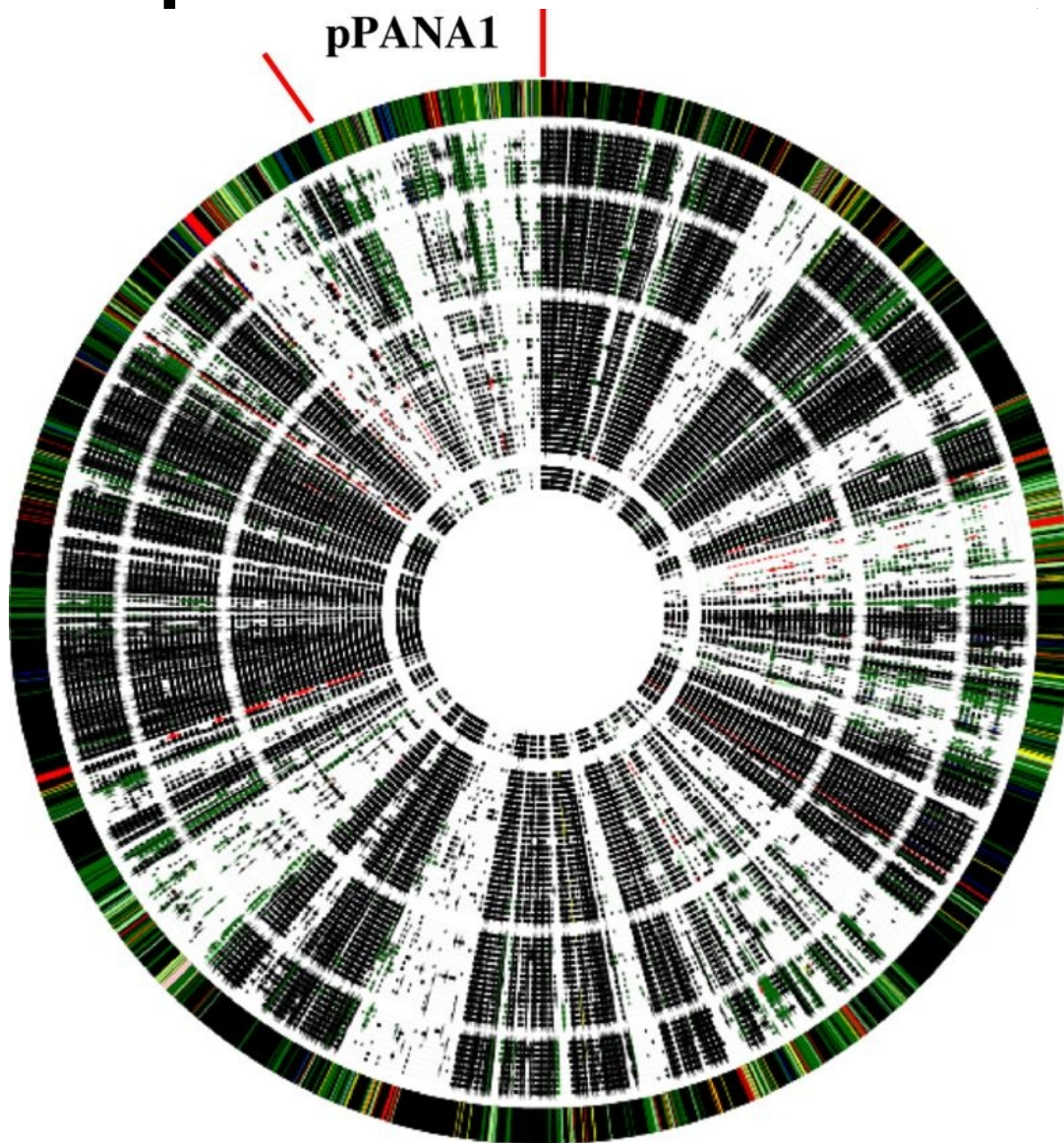
- What is this figure showing?
- Is the program correct for visualizing the data?
- Strengths and weaknesses of this figure

Paper review - figure 3



- What is this figure showing?
- Is the program correct for visualizing the data?
- Strengths and weaknesses of this figure

Paper review - figure 4



- What is this figure showing?
- Is the program correct for visualizing the data?
- Strengths and weaknesses of this figure

Paper review - figure 5

- What is this figure showing?
- Is the program correct for visualizing the data?
- Strengths and weaknesses of this figure

