

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

Learning Objectives

1. Review command line
2. How do you get a genome?
3. File formats
4. Discuss paper
5. How to review a paper critically

High parasite diversity accelerates host adaptation and diversification

Betts et al., 2018

Multiple parasites speed host evolution

Virtually all organisms are parasitized by multiple species, but our current understanding of host-parasite interactions is based on pairwise species interactions. Betts *et al.* address this by using the bacterium *Pseudomonas aeruginosa* and five different phage virus parasites. Increasing parasite diversity accelerated the rate of host evolution, driving both faster genomic evolution within populations and greater divergence between populations. Thus, different parasite loads prompt different evolutionary dynamics and profoundly shape host evolution by different mechanisms.

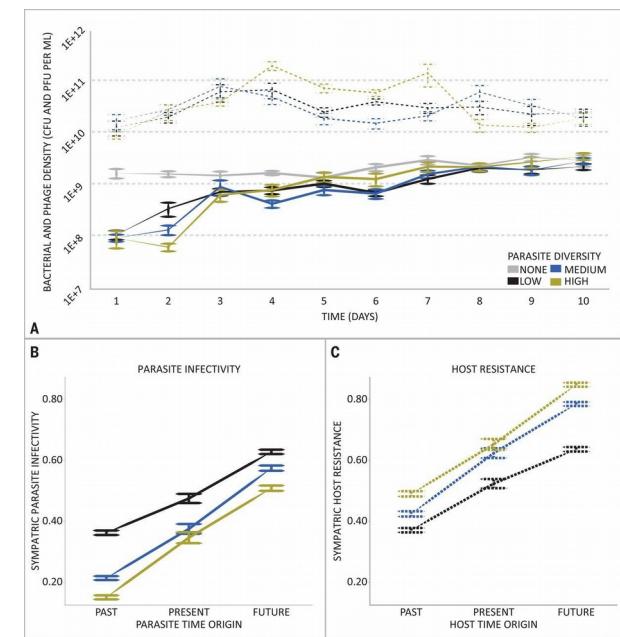


Figure 1 (from article): Ecological and coevolutionary dynamics of host-parasite interactions..

Link to research article: <http://science.sciencemag.org/content/360/6391/907>

Original article/source of figure: Betts A, Gray C, Zelek M, MacLean RC, King KC. 2018. High parasite diversity accelerates host adaptation and diversification. *Science*. 360:907-911. doi: 10.1126/science.aam9974.

Announcements

- Phytobacteriology in the News [Week 3](#) review.
- Good work on the News & Views draft titles! First draft due **Thurs Sept 13th** (midnight), please submit to folder [here](#)
- Paper for next week: Pan-genomes! Form [here](#) everybody answered, right?

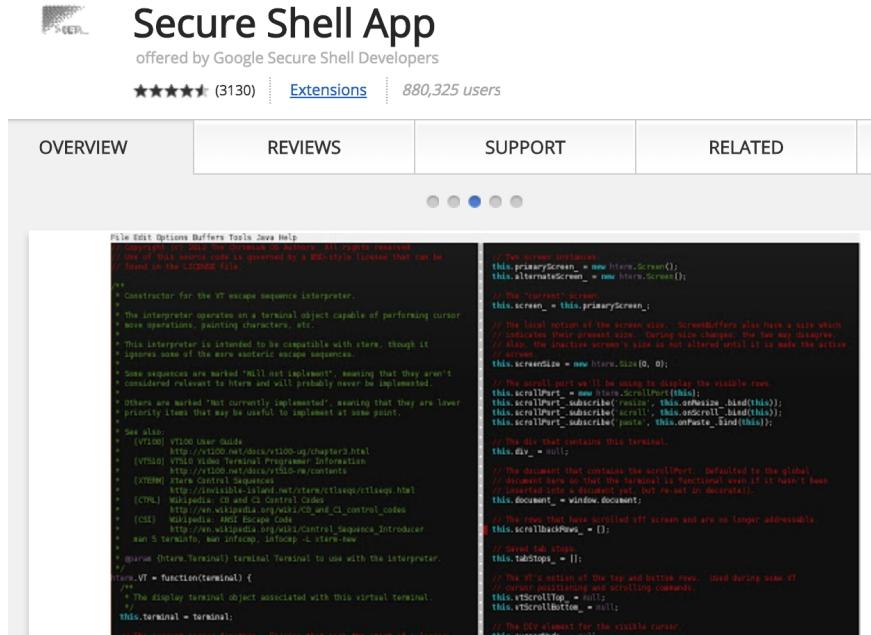
Command Line - Terminal

every have one?

Mac: Terminal

Windows: PUTTY

Chrome: Secure Shell (extension; runs command line in your browser! How cool is that?)



Command line - Review

cd	?	grep	?
pwd	?	sed	?
mkdir	?	rm	?
head	?	rm -r	?
tail	?	echo	?
more	?		?
less	?	>	?
mv	?	<	?
touch	?	printf	?
cp	?	>>	?
cat	?		

Command line - Review

- Make a folder with your name
- Make two files in as .txt (labeled 1 and 2)
- Write your name in file 1
- Write your name in file 2 with _a at the end
- Combine those two files to a new file
- Copy the whole folder with name, label it folder 1

Kamiak - login

ssh

friesenguest@se088039.temp.wsu.edu

XXXX

How to obtain a genome?

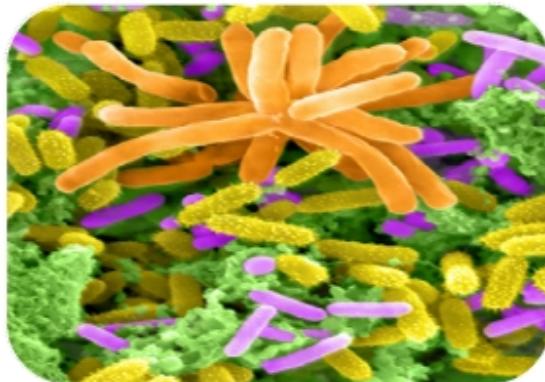


Culturing < 1% can be easily cultured



Single cell genomics

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



Metagenomic - population genome binning

De novo assembly terms

k-mer (4-mer)

ATTACCGG
TTTCGGG
GGGCCGG
TTTAATTA



Reads

De novo
assembly

TTTCGGGGGGCCGG
TTTAATTACCGG

Contigs

TTTCGGGGGGCCGG
TTTCGGG

Mapping
(Coverage)



TTTCGGGGGGCCGG
NNNNTTAATTACCGG

Scaffolds

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)

What is a draft vs complete¹¹ genome?

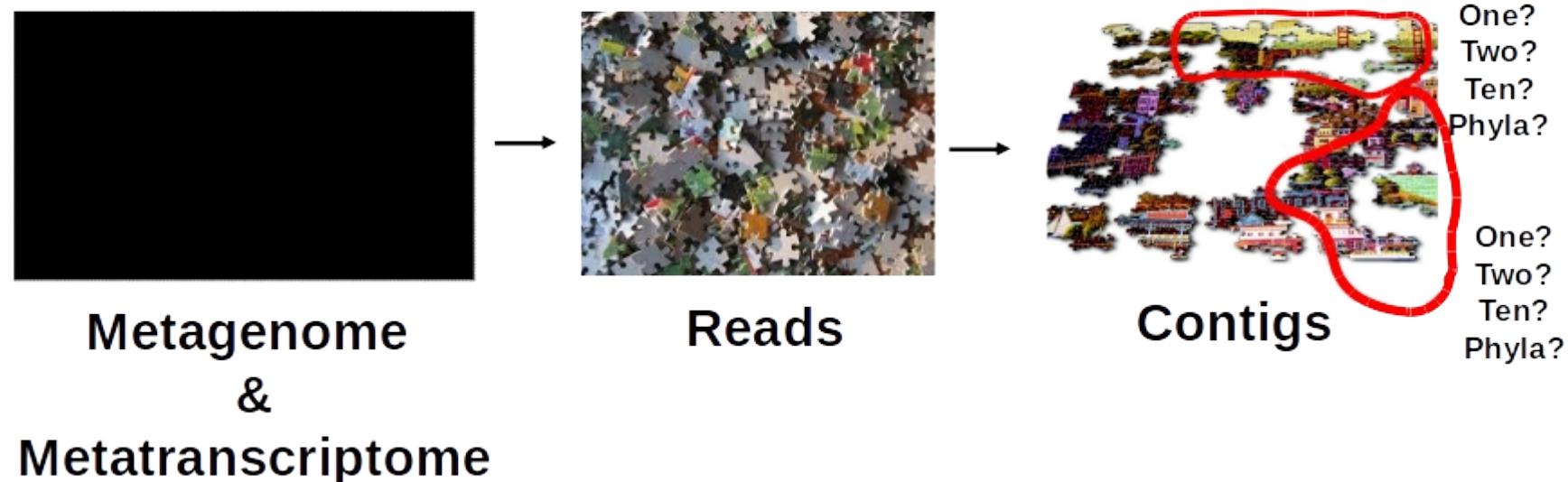
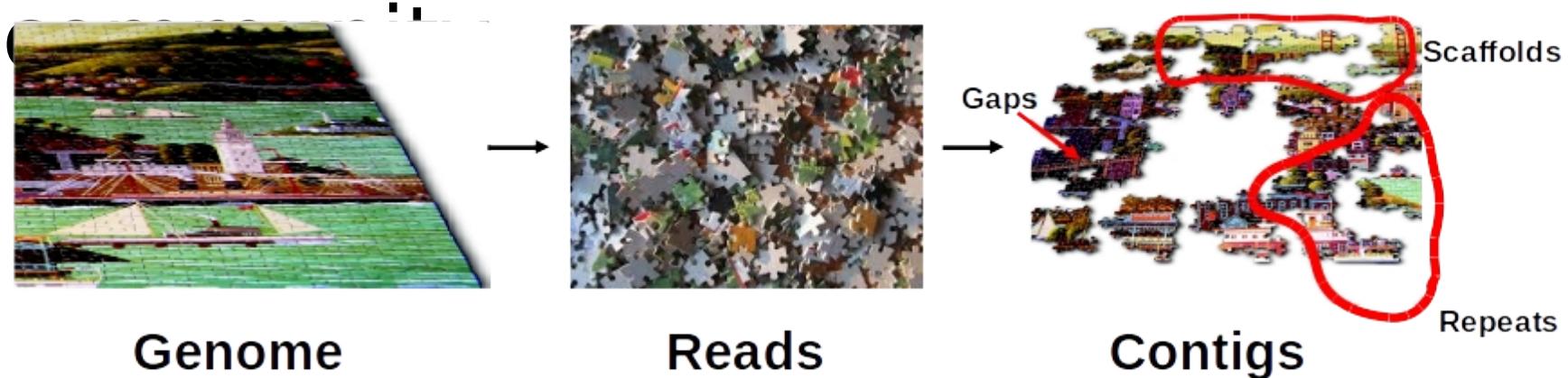
- Draft genome contains many contigs (>10)
- Complete genome in bacteria, is usually (<5 contigs) often containing a circular chromosome but can contain a linear chromosome (e.g., *Rhodobacter*), with plasmids.

Why do we need complete genomes?

12

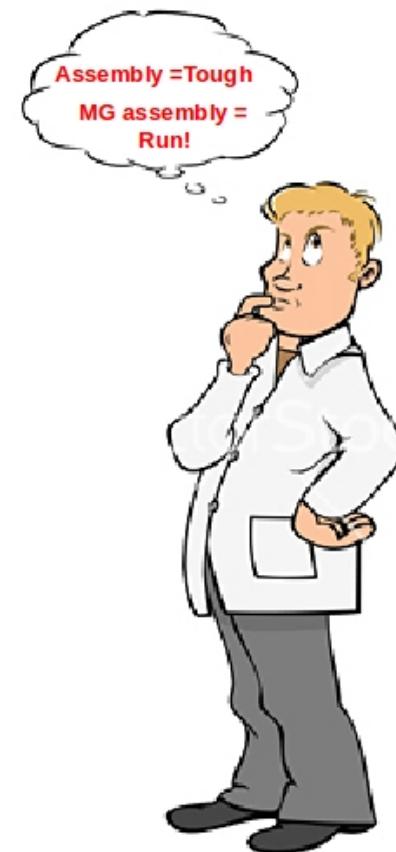
- Functional studies demand an error-free genome sequence as a starting point¹
- Availability of data on genome organization provides biological insights¹
- A complete genome sequence is a permanent, valuable scientific resource¹

De novo assembly genome vs.



A single genome isn't a community metagenome?

- High number of types and sizes of genomes present with very sparse sampling leading to low coverage
- Lots of data required usually short reads are used which requires high RAM (working memory) required to assemble
- De novo assembly of soil metagenomes (most complex) yield few contigs and few of the contigs have representative coverage (few mapped reads)
- Few or any computational tools that scale to the large amounts of data required for analysis of data



File formats in genomics

*.fna, .fasta, .fa ?

*.faa ?

*.gff ?

*.gbk ?

*.fq or fastq ?

*.gtf Gene transfer format (GTF)

Read vs. sequence?

Contig?

kmer?

File formats in genomics (fasta)

>sequence1

ATACTCTACTCGTCTCATATCAT

>sequence2

GCGCGCG**N**CAGCGATCTCTCA

>sequence3

TTTCGCG**NN**CAGCGATCTCTC

File formats in genomics (fastq)

Example 1

```
@SIM:1:FCX:1:15:6329:1045 1:N:0:2
TCGCACCTAACGCCCTGCATATGACAAGACAGAAC
+
<>;##=><9=AAAAAAAAAA9#:<#<;<<<????#=
```

Example 2

```
@SRR001666.1 071112_SLXA-EAS1_s_7:5:1:817:345
length=36
GGGTGATGGCCGCTGCCGATGGCGTCAAATCCCACC
+SRR001666.1 071112_SLXA-EAS1_s_7:5:1:817:345
length=36
```

|||||||9IG9IC

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";
```

Pull files from NCBI

- For *Rhodobacter sphaeroides* 2.4.1 (pubmed.com)
 - Get contig fasta
 - Gbk file
 - Gff or gtf file
 - Protein fasta

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

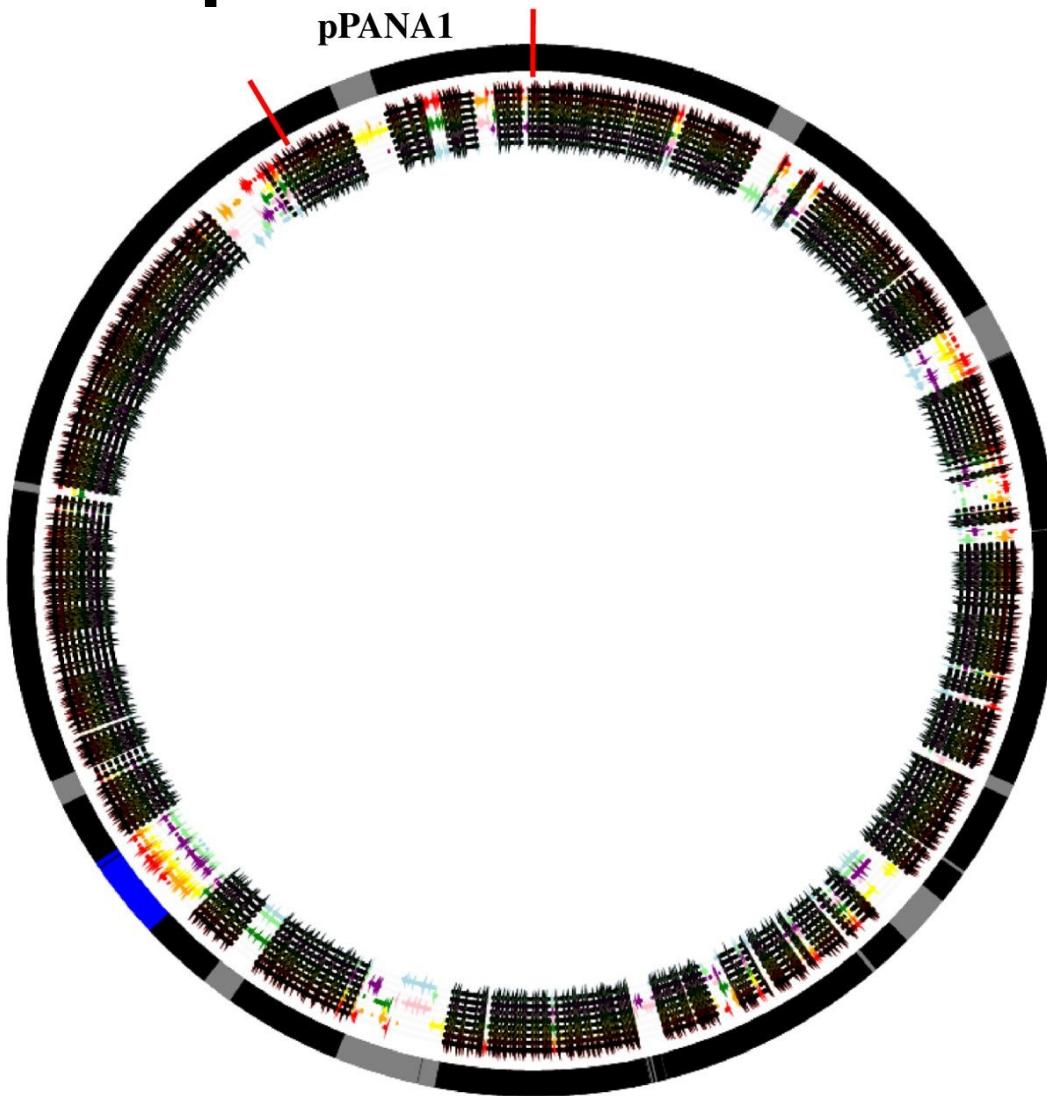
Paper review - Terms

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

Paper review

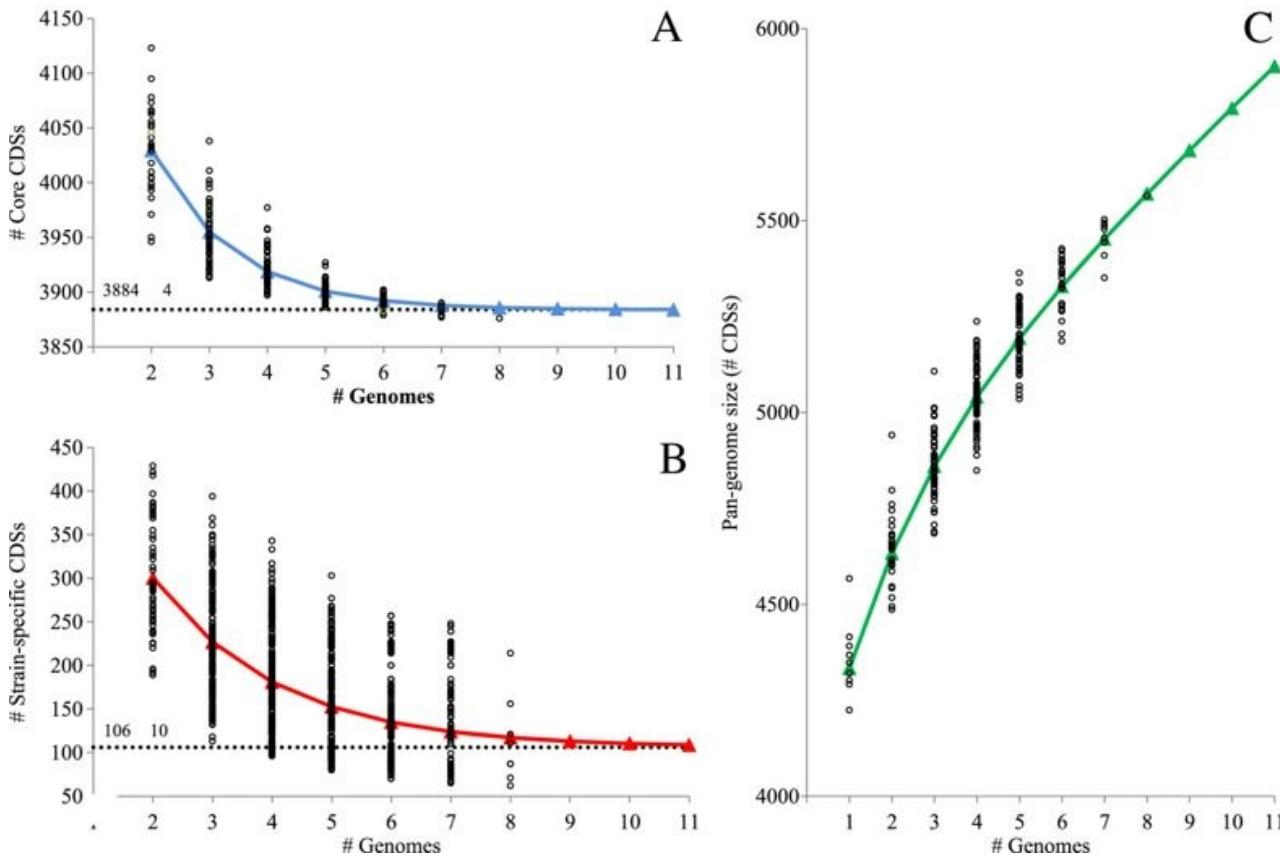
- What type of paper is it?
- What is the central hypothesis of this paper? *
- What is the key scientific advance of this paper? *
- What surprised you about this paper? *
- What new vocabulary or concepts did you learn while reading this paper? List & define. *
- If you were the author of this paper, what would you do next? *

Paper review - figure 1



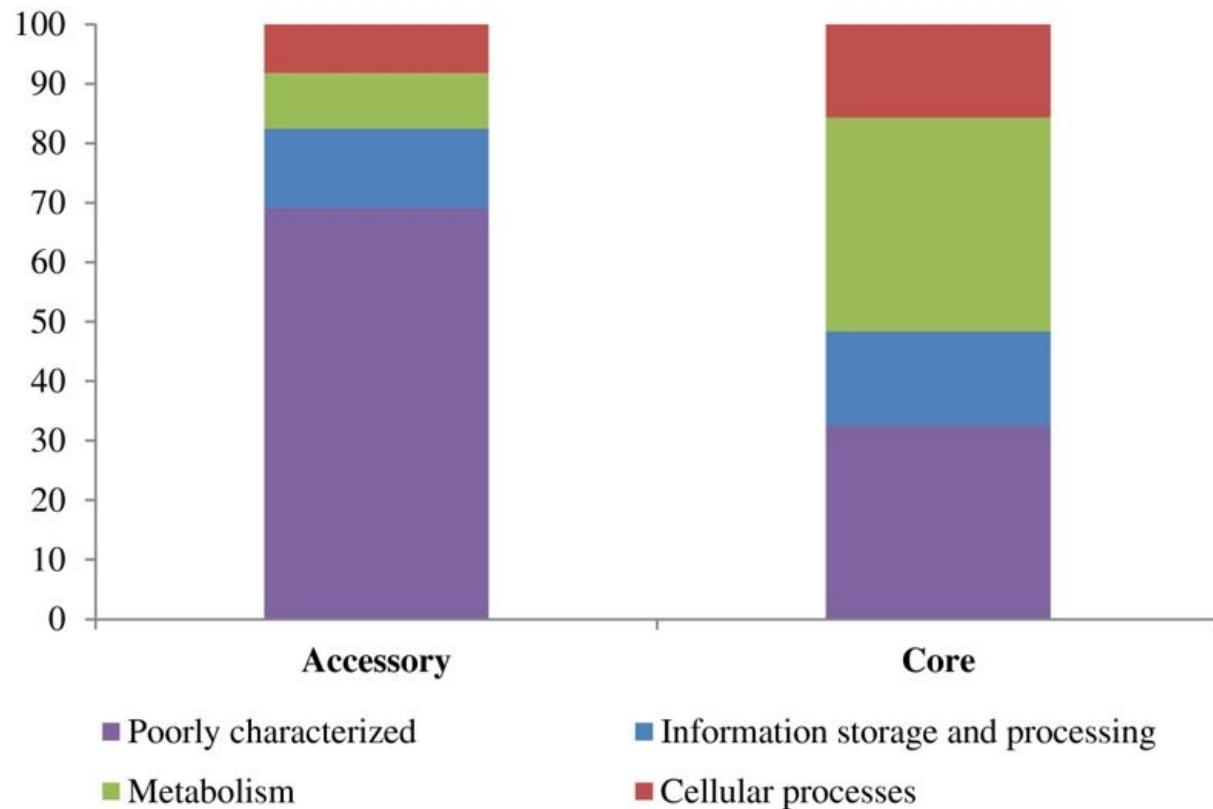
- What is this figure showing?
- Is the program correct for visualizing the data?
- Strengths and weaknesses of this figure

Paper review - figure 2



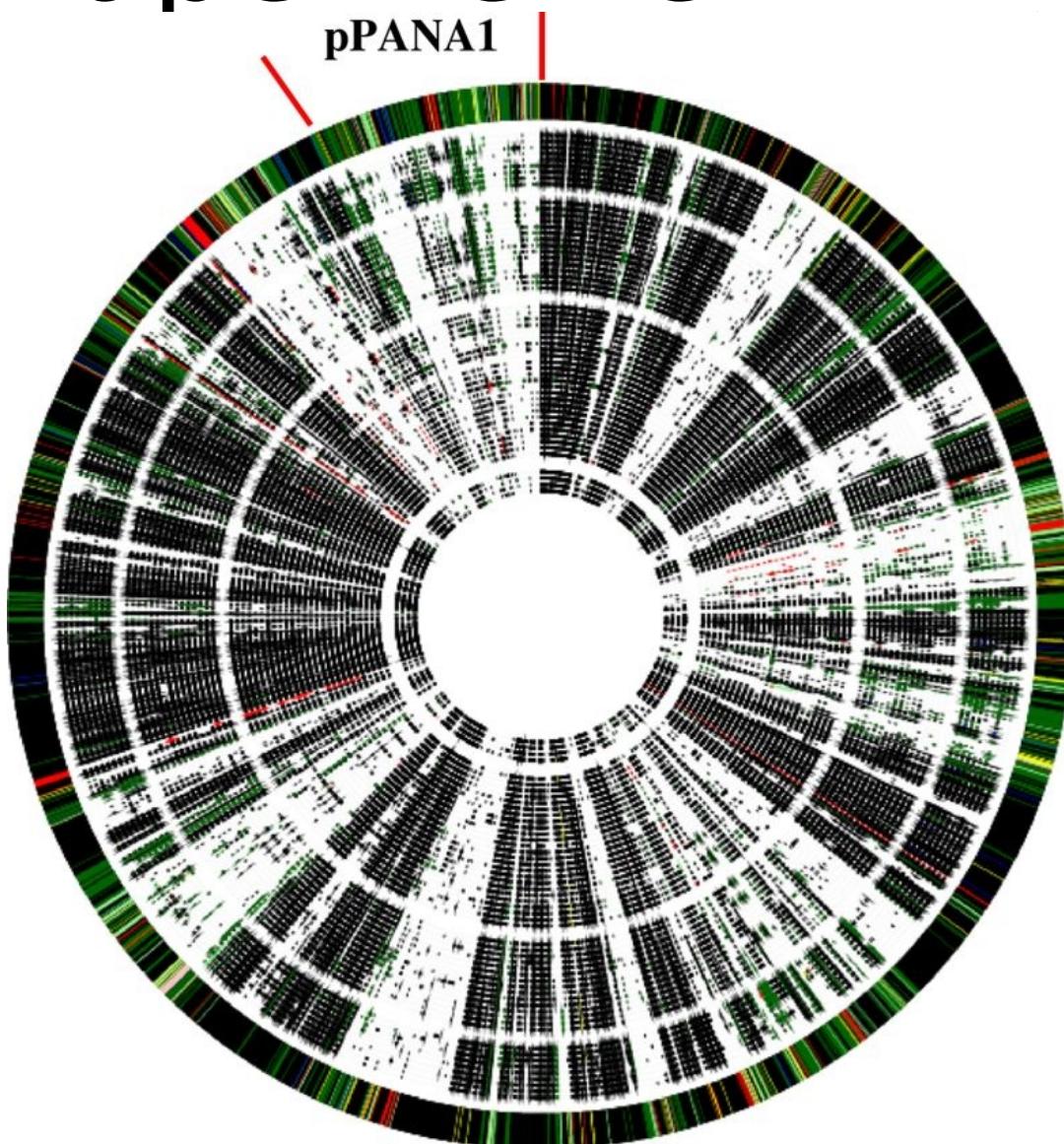
- What is this figure showing?
- Is the program correct for visualizing the data?
- Strengths and weaknesses of this figure

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

