CAIRIBU
Collaborating for the Advancement of Interdisciplinary Research in Benign Urology

How to write a scientific paper

Dr. Nicole J. De Nisco
Assistant Professor
The University of Texas at Dallas

August 9th, 2023
CAIRIBU ARCTICS Community Forums

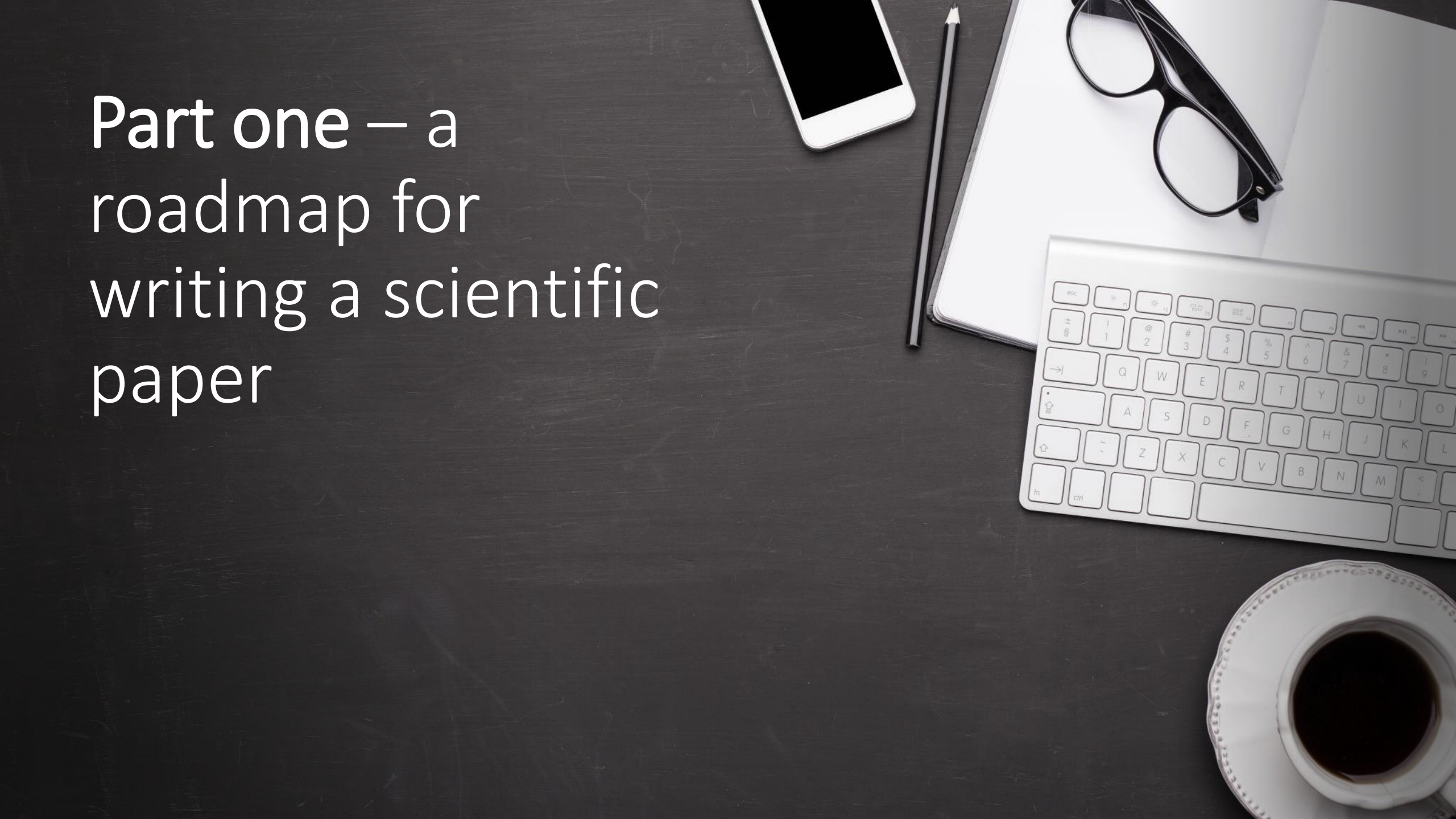
CAIRIBU ARCTICS Community
Advancing the Research Capacity of Trainees and Investigators at early Career Stages
Virtual Forums

Audience survey

- Raise your hand (virtually) if you struggle with scientific writing!

- Please share what aspects of scientific writing are most difficult for you over the chat….

Part one – a roadmap for writing a scientific paper



When should you start writing?

- A common struggle with manuscript writing is knowing when to start
- Many scientists feel they need to have every experiment complete before they should start writing
- This is not the case!
- The act of organizing your work into a manuscript will reveal missing experiments or logic gaps in your work
- ***Start writing early*** so that you can identify these gaps before you are up against a tough deadline!

What are the common components of Scientific Papers

- Title
- Authors and affiliations
- Keywords

Title page

- Abstract
- Introduction
- Methods
- Results
- Discussion
- Figure and Table Legends
- References
- Acknowledgements
- Author contributions

- Depending on the journal, the figures and legends will either be inserted within the results section or will be put towards the end of the manuscript

- Most journals offer copyediting and will format your original image files and figure legends into the final PDF version

How do you start writing?

- **Please share in the chat your ideas on the best way to start writing a manuscript?**
  - For example, what section should you start with first?

Steps in scientific manuscript writing

1. **Prepare the Main Text Figures and Tables**
   - Don’t forget to write figure and table legends!

2. **Organize the Figures and Tables so that they follow a logical narrative – Make a figure outline**
   - The best order for your figures/tables may not follow the chronological order in which the experiments or analyses were performed!
   - *At this point you may notice gaps in the logic of your work that need to be filled in with more experiments or analysis*

Steps in scientific manuscript writing

3. Write the methods

- *This should be the easiest part* – you should have your methods well-documented in your lab notebook.
  - You can write these along the way while doing experiments and you always come back and revise
  - Having typed up protocols will help you when it comes to writing the methods!

- Methods should be sufficiently clear for someone to **replicate** your work

- Some journals require very detailed methods (e.g. STAR methods) or will allow supplemental methods so you can provide more detail

Steps in scientific manuscript writing contd.

4. Write the Results

- Describe the data presented in each figure and table in order
  - It is good practice to report both the effect size and the statistical significance
  - The word ”significant” should only be used in the case where hypothesis testing is performed, and the p value is less than alpha
- Separate into sections either by figure or by major result
  - If a figure is large and contains many panels, you may have more than one section per figure
- Section titles should state the major findings of that section
  - If no major findings (e.g. describing patient demographic data) then should be descriptive of the contents of the section
- *While writing this section you may realize that you are missing experiments or analyses – if so, complete them and then go back to Step 1.*

Steps in scientific manuscript writing contd.

5. Organize the supplemental material

- While writing the results, you will identify where supplemental material (figures or tables) is needed.
- You may also identify potential supplemental figures while organizing your figure panels in **Step 2**
- Supplemental material includes data that is necessary to further support the conclusions drawn in the results section, to demonstrate rigor, data files needed to reproduce analyses, and that will enable the use of your dataset by others
- Large supplemental data files like raw sequencing data should be deposited in the appropriate public repository. For most journals this must be done ***before submission***

Steps in scientific manuscript writing contd.

6. **Write Discussion**
    - Steps 6 and 7 are interchangeable in my opinion
    - Discuss your findings in the context of the existing literature
    - This is a good place to describe a model figure if you have one.
    - Identify study limitations and lines of research that need to be continued
7. **Write the Introduction**
    - The introduction should provide sufficient background so that the reader understands why your study is important/new/interesting
    - Define your hypotheses and objectives in the context of what is known
    - Summarize your most important findings in the final paragraph to entice the reader to keep reading

Pointers for writing introductions and discussions

- Make sure your introduction and discussion are well-cited.
  - For science journals there is usually not a citation limit, so make sure ***each statement of fact is supported by a citation***
  - For medical journals there are often citation limits - these can be tricky to navigate but word limits (i.e. introduction scope) are also usually restricted
  - Citations should be for the article where the **primary data** to support the statement was first published, not a review article. DO NOT JUST CITE REVIEWS!

- Present limitations of your study in the discussion
  - Many journals require this as a separate section withing the discussion
  - It is very important to clearly convey the limitations of your study to prevent overinterpretation

Steps in writing a scientific manuscript contd.

8. **Write the Abstract, compose the title, and select keywords for indexing**
    - Title should accurately convey major findings, cute/funny titles should be reserved for reviews or commentary
    - Abstract format and length depends on the journal

9. **Make sure the references are in the correct journal format**
    - you can download journal-specific reference styles in Endnote, Mendeley etc.

10. **Write acknowledgements and author contributions**
    - Journals often have guidelines for how to format author contributions
    - Acknowledgements need to contain funding information including grant numbers

What is the most important part of a scientific paper?

- **Please share your thoughts in the chat**
- **The figures!**
  - Your data is the centerpiece of your manuscript
  - You need to present your data in a way that is accurate, easy to interpret, and accessible
  - It also can help to present data in a way that is visually appealing
  - ***Every conclusion you make needs to be supported by data provided in a figure or table***

How to turn your data into a figure

- **What software do you use to make figures?**
  - Adobe Illustrator (composing multi-panel figures)
  - ImageJ/FIJI (microscopy data)
  - GraphPad Prism (plots of various types and performing basic biostatistics)
  - R packages – ggplot (various types of plots if familiar with R)
  - Biorender, ChemDraw (for diagrams, workflows and models)
- **What software should you NOT use**
  - Photoshop – data can be manipulated
  - Powerpoint – produces raster low dpi graphics

WHAT is a Raster graphic?

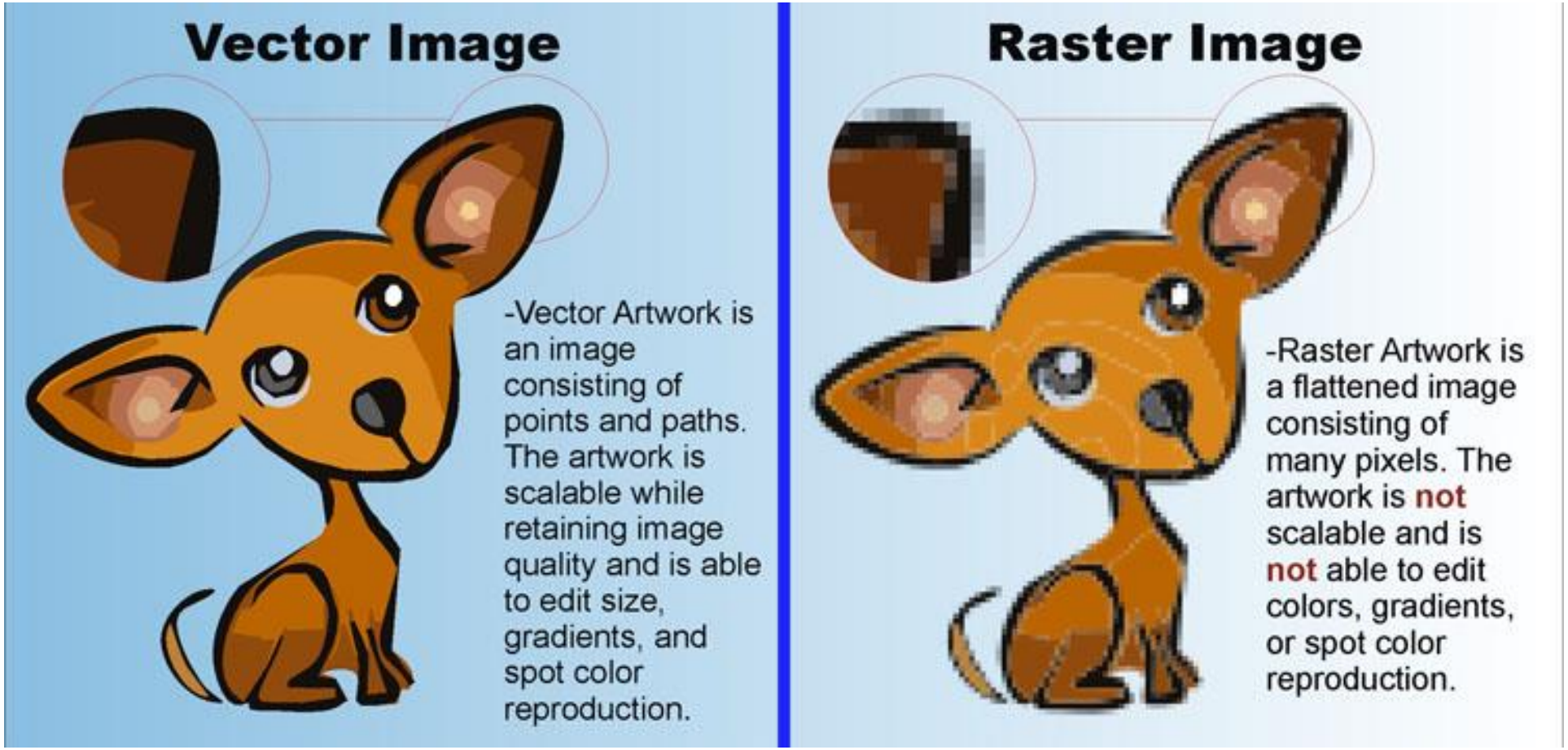
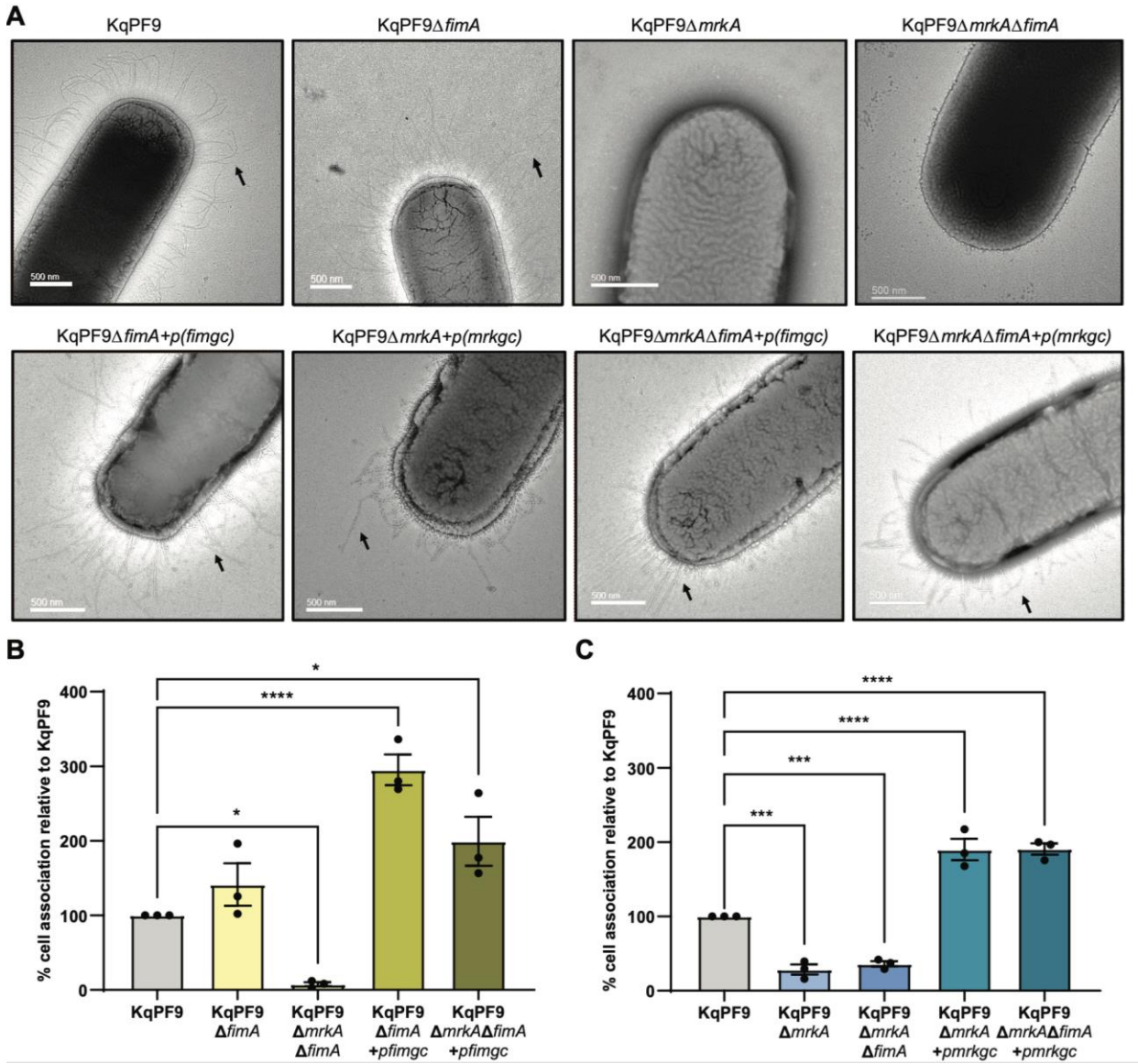


Figure or Table?

- How do you decide whether to present your data in tabular or graphical format?
  - For data that can be plotted, graphs are usually preferable because they allow you to visualize effect sizes, data distribution and summary stats like median
  - Tables are useful for reporting summary values, binary, or categorical outcomes that cannot easily be plotted.
  - Some journal may require raw data to be supplied as supplemental material – this is always good practice!



Panel A images were processed in ImageJ, panel B plots were made in Prism, and the figure was composed together in Adobe Illustrator

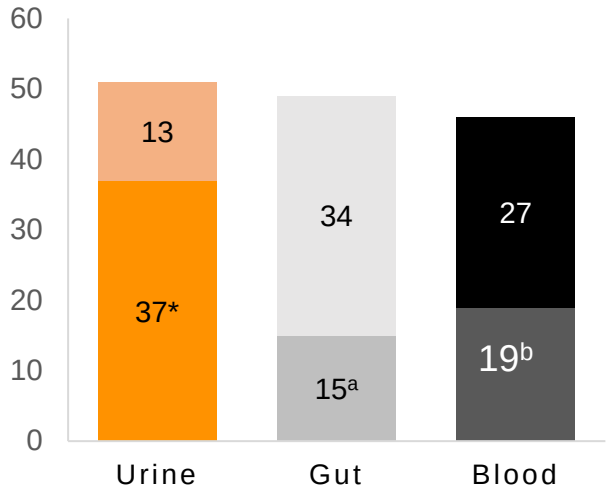
What are attributes of a good figure?

- **Figures must be easy to interpret without reading the text**
  - Clear labeling and legends (e.g. for heatmaps) are key
  - Use legible font sizes!
- **A good figure directs the eye of the reader to the important data**
  - For example, use arrowheads to denote important features in images
  - Present results of statistical tests
- **A good figure does not mislead the reader**
  - Keep scales consistent between graphs
  - Select representative images, not just ones that support your hypothesis
    - ***It's a good idea to quantify imaging data when possible to avoid bias***
  - Apply the correct statistical tests and show data distribution (e.g. individual data points or box and whisker plots)

Let’s look at an example!

Figure 1 – First Draft

A

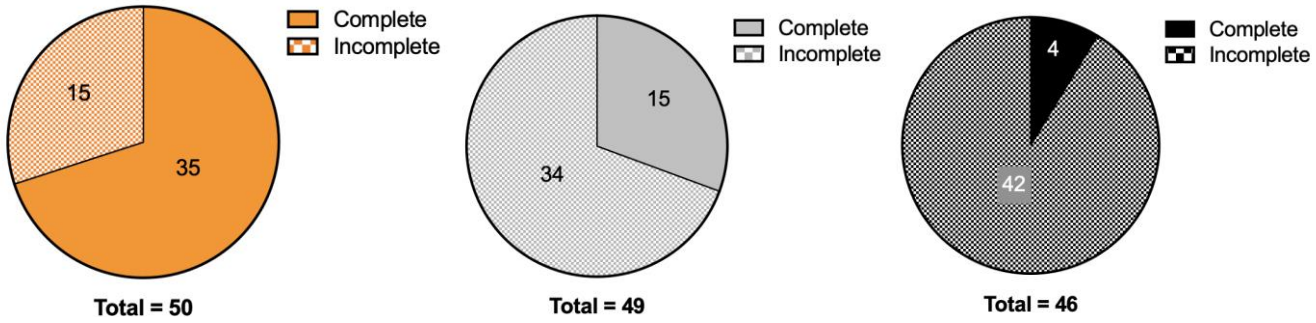


* Isolates sequenced as part of this study
[a] BioRxiv
[b] Van Tyne et al.

B

NCBI Genome Assembly Inclusion Criteria	
Isolation Source	Human - gut, fecal, urine, urogenital tract
Collection	Non-longitudinal, isolated from unique individuals
Assembly Coverage	>30x
Contig N50	>80,000
No. of Contigs	<100
Publication	Published or listed in literature

C



Feedback –

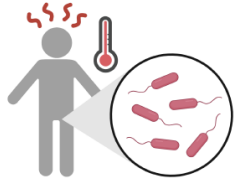
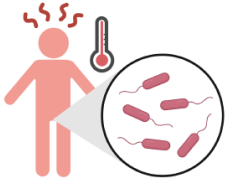
- Provenance of urine *E. faecalis* strains unclear
- Missing y-axis legend
- Color scheme could be improved

Figure 1 – Final Draft

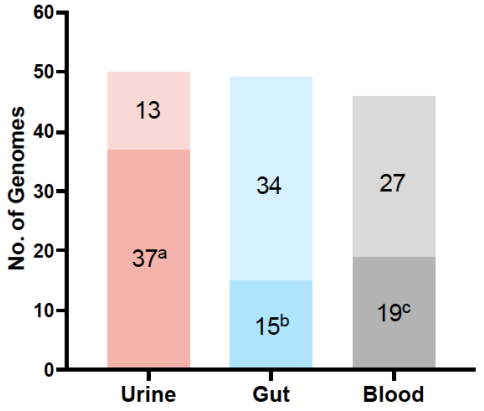
How feedback was addressed:

- Added panel A so source and host status of clinical *E. faecalis* isolates clearly identified, number per group is defined
- The y-axis for panel B is now defined
- Color scheme is unified and visually appealing

A

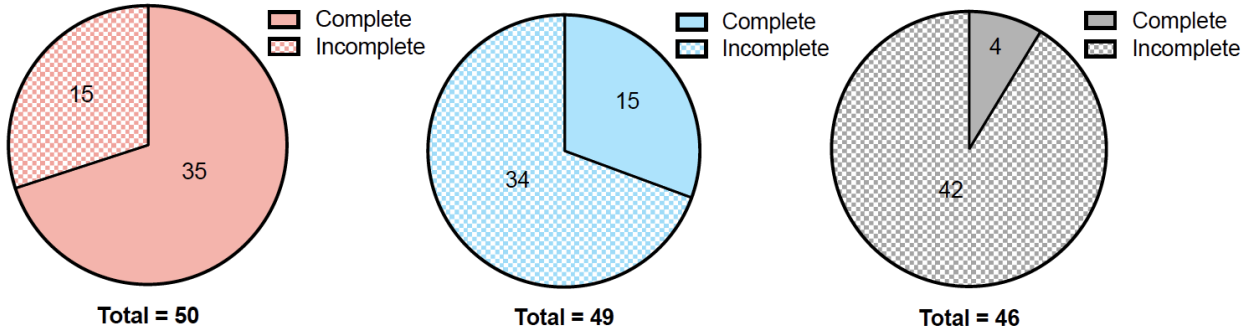
	Never	Sporadic	Remission	Relapse
Cohort				
UTI History	None	UTI	RUTI	RUTI
Symptoms	None	Symptomatic	None	Symptomatic
Urinalysis	Negative	Positive	Negative	Positive
n	4	3	17	13

B



NCBI Genome Assembly Inclusion Criteria	
Isolation Source	Human - gut, fecal, urine, urogenital tract
Collection	Non-longitudinal, isolated from unique individuals
Assembly Coverage	>30x
Contig N50	>80,000
No. of Contigs	<100
Publication	Published or listed in literature

C



bioRxiv - doi: https://doi.org/10.1101/2023.05.18.541374

Figure Legend pointers:

- Figure legend title should summarize the major finding or the overall point of the figure
  - Here this figure depicts the overall isolate collection used for genomic analysis in this paper
- The legend should describe any abbreviations, acronyms or notations
  - Superscript a, b, c described in legend
- If applicable legend should include the names of statistical tests used to generate p-values
  - No statistical tests performed in this figures

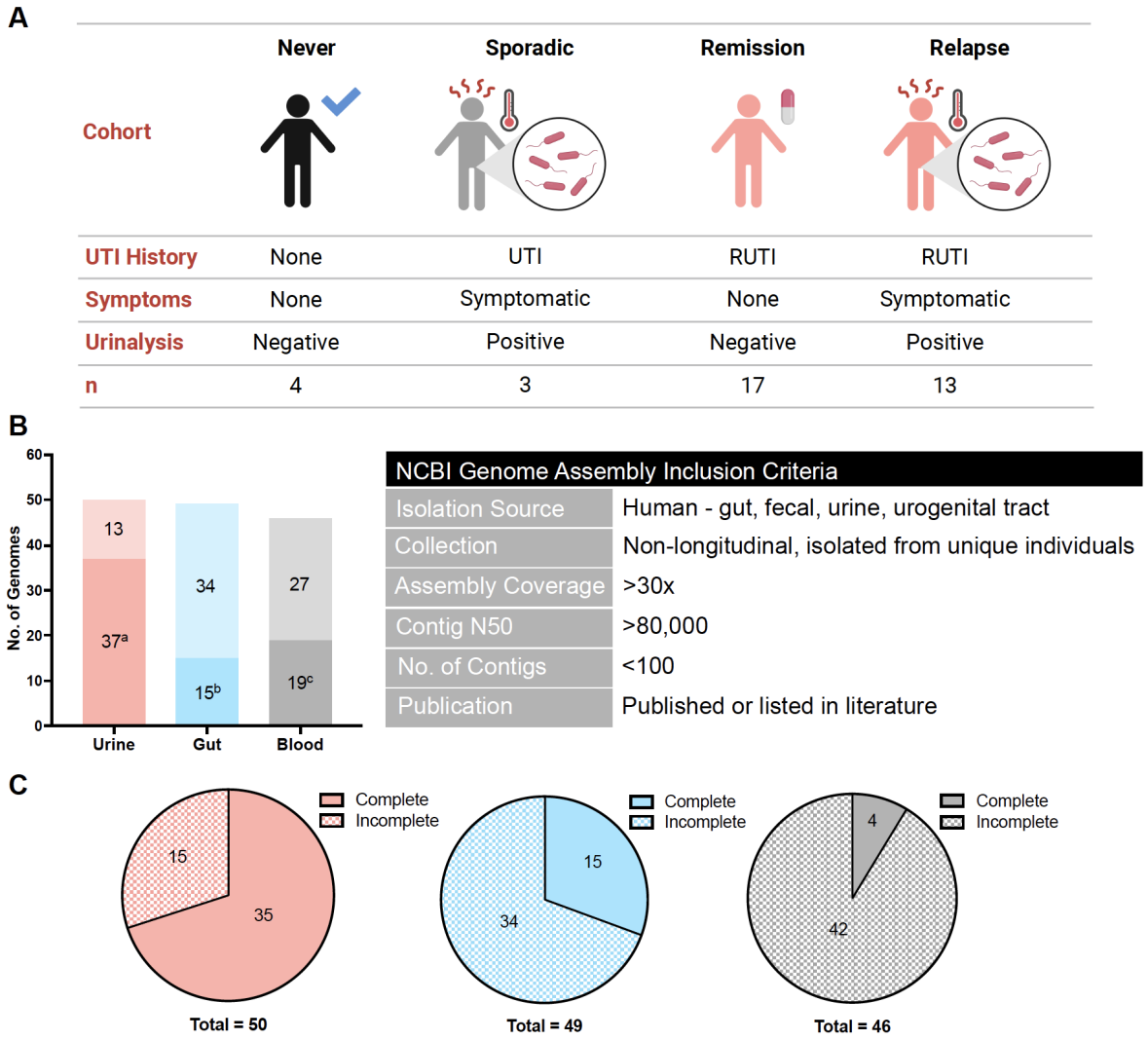
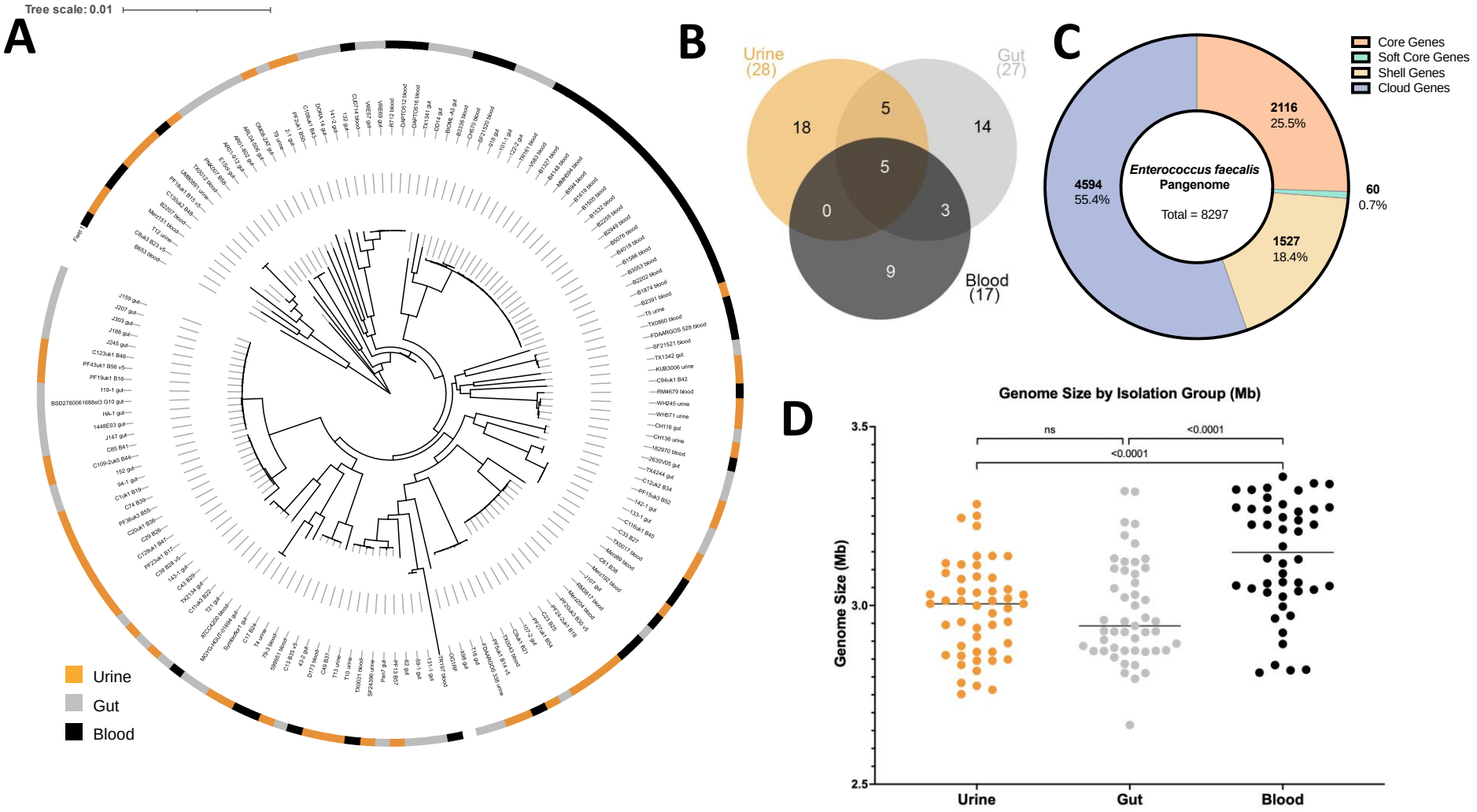


Figure 1. Clinical cohorts and isolated collection of *Enterococcus faecalis*. (**A**) Patient cohorts were stratified based on patient UTI history, symptoms, and urinalysis results at time of specimen collection. n denotes number of isolates. (**B**) Number of isolate genomes from each isolation group. [a]Isolates sequenced as part of this study. [b]Isolates from Palacious Araya *et al.* (73). [c]Isolates from Van Tyne *et al.* (21). Genome assembly criteria used for selection of comparator genomes. (**C**) Counts of complete and incomplete genome assemblies in the urine (pink), gut (blue), and blood (grey) isolation groups.

MAIN-TEXT

Figure 2 – First Draft



Feedback –

- Tree text is impossibly small to read
- Color scheme can be improved

When should a figure be supplemental?

- When it is not necessary for understanding the main narrative/major conclusions of the manuscript
- But provides important *additional* supporting evidence for the major conclusions or is needed to support minor findings
- Negative data is also often included in the supplement
  - E.g. data that does not show a certain phenotype or difference
- Data needed to reproduce major analyses is usually provided as a supplemental table (or deposited on a public repository)

Figure 2

- Figure 2C provides overall pangenome breakdown for *E. faecalis*
- Does not show pangenomes by isolation source

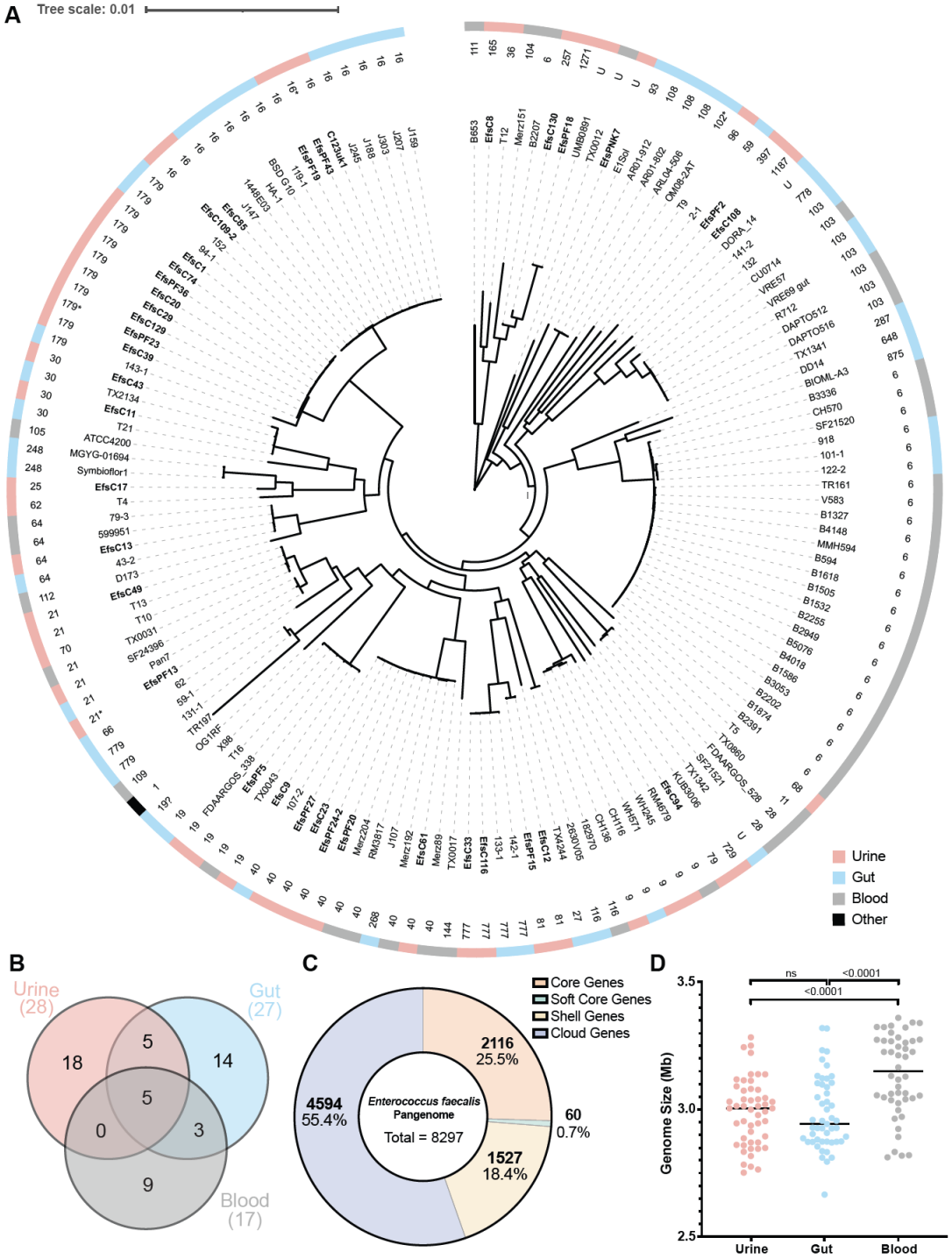


Figure 2 Phylogenetics, MLST, pangenome and genome size distributions of urinary, gut, and blood *E. faecalis* isolates. (**A**) MinVAR-rooted maximum likelihood phylogenetic tree constructed from the core gene alignment of all 146 E. faecalis strains in this study. Isolate names are listed at leaves. Dots represent complete genome assembly. MLST is depicted next to isolate names. U = unknown, * = ST has a novel allele, ? = ST uncertain. Outermost ring is color-coded by isolation source: urine (pink), gut (blue), blood (grey), reference (black). (**B**) MLST Venn diagram depicting the total number of distinct sequence types in each of the isolation groups. Totals are listed below group names. (**C**) Pangenome analysis summary. Core genes present in >99% of isolates, Soft Core genes present in 95-99% of isolates, Shell genes present in 15-95% of isolates, and Cloud genes present in <15% of isolates. (**D**) Distributions of genome size in megabases of all isolates in each isolation group. Each dot on the graph is an isolate. Statistical significance was determined using ordinary one-way ANOVA with multiple comparisons $p<0.05$ is significant.
[a]Isolate name shortened for simplicity. MGYG-01694 = MGYG-HGUT-01694, BSD G10 = BSD2780061688st3_G10.

bioRxiv - doi: https://doi.org/10.1101/2023.05.18.541374

SUPPLEMENT

Supplemental Figure S2

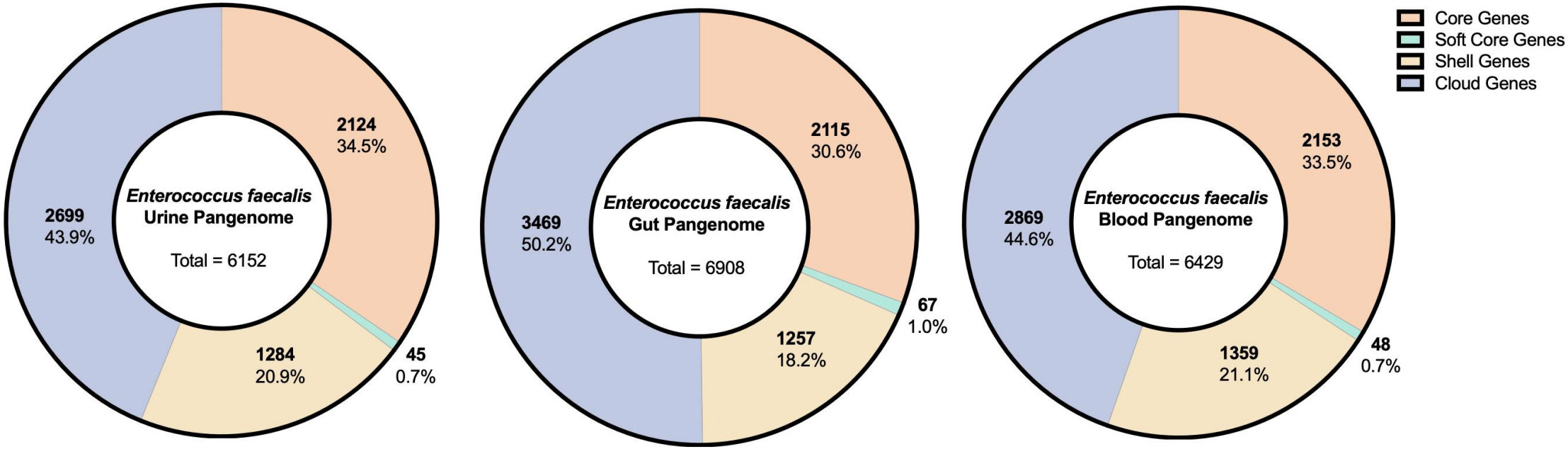


Figure S2. Pangenomes of urinary, gut, and blood isolation groups. Pangenome was determined for isolates in each isolation group. Core genes present in >99% of isolates, Soft Core genes present in 95-99% of isolates, Shell genes present in 15-95% of isolates, and Cloud genes present in <15% of isolates.

- The pangenome composition did not vary much when genomes binned by isolation source
- So individual isolation source *E. faecalis* pangenomes graphs presented in a supplemental figure

Supplemental tables

- Supplemental tables can either be provided within the supplemental material file as a “Word” table or if sufficiently large as separate excel file
- For example, additional summarized clinical metadata or strain lists may be provided as a word format table
- Intermediate data files or large metadata that are in tabular format are often supplied as excel files
  - E.g. Results of “big data” analysis that are summarized into main text figures
  - Metadata files that are too large to be readable in word table format
- Supplemental tables must have descriptive titles – legends may also be necessary to describe any acronyms or annotations.

Supplemental table example – excel format

Descriptive title

Legend

Supplemental Table S3 - All *Enterococcus faecalis* genomes metadata

Urine, gut, and blood *E. faecalis* genomes used in this study are listed along with NCBI accessions, genome size (in basepairs), average assembly coverage, number of contigs, contig N50 (where applicable), sequencing method, isolation date, geographical isolation location, and DOIs of published work citing these genomes.

Clearly labeled column headers

Group	Strain	BioSample	BioProject	Genome Size (bp)	Average Coverage	No. Contigs	Contig N50 (Draft genomes only)	Method	Isolation Date	Geographical Location	Publication DOI
Urine	KUB3006	SAMD00113788	PRJDB6823	3,138,474	300x	4	-	Sequel, MiSeq	2017	Japan	10.3389/fmicb.2018.02576
Urine	FDAARGOS_338	SAMN06173351	PRJNA231221	2,861,022	26.32x	1	-	Illumina, PacBio	3/6/15	USA (DC)	10.1038/s41467-019-11306-6
Urine	WH571	SAMN00809137	PRJNA88889	3,283,528	140x	47	324397	Illumina	1986	USA (West Haven, CT)	10.1371/journal.pone.0000582
Urine	T9	SAMN00847610	PRJNA157759	2,991,936	141x	10	467027	Illumina	≤ 1992	Japan	10.1371/journal.pone.0000582
Urine	UMB0891	SAMN08193660	PRJNA316969	3,073,698	340x	43	343482	HiSeq	2015	USA (IL)	N/A
Urine	CH136	SAMN00809220	PRJNA89049	3,222,352	256x	41	233060	Illumina	1988	USA (MA)	10.1371/journal.pone.0000582
Urine	WH245	SAMN00809218	PRJNA89045	3,250,986	144x	39	207232	Illumina	1987	USA (West Haven, CT)	10.1371/journal.pone.0000582
Urine	T10	SAMN00847611	PRJNA157761	2,783,388	146x	12	396278	Illumina	≤ 1992	Japan	10.1371/journal.pone.0000582
Urine	T12	SAMN00847603	PRJNA157743	3,004,508	295x	25	420825	Illumina	≤ 1992	Japan	10.1371/journal.pone.0000582
Urine	T13	SAMN00809223	PRJNA89053	2,775,355	283x	15	321792	Illumina	≤ 1992	Japan	10.1371/journal.pone.0000582
Urine	T4	SAMN00839768	PRJNA158749	2,912,684	140x	18	426876	Illumina	≤ 1992	Japan	10.1371/journal.pone.0000582
Urine	T5	SAMN00847609	PRJNA157757	2,870,381	289x	16	346197	Illumina	≤ 1992	Japan	10.1371/journal.pone.0000582
Urine	SF24396	SAMN00809226	PRJNA89059	2,816,993	146x	12	532137	Illumina	2001	USA (MI)	10.1371/journal.pone.0000582
Urine	EfsPF23	SAMN33741218	PRJNA944190	3,138,108	102x, 514x	4	-	Illumina, MinION	9/10/18	USA (Dallas, TX)	This Study
Urine	EfsPF24-2	SAMN33741219	PRJNA944190	3,114,654	136x, 704x	4	-	Illumina, MinION	4/1/19	USA (Dallas, TX)	This Study
Urine	EfsC1	SAMN33741190	PRJNA944190	3,077,295	75x, 51x	3	-	Illumina, MinION	5/7/18	USA (Dallas, TX)	This Study
Urine	EfsC9	SAMN33741192	PRJNA944190	3,045,092	121x, 342x	3	-	Illumina, MinION	6/4/18	USA (Dallas, TX)	This Study
Urine	EfsC11	SAMN33741193	PRJNA944190	3,004,778	121x, 342x	2	-	Illumina, MinION	6/11/18	USA (Dallas, TX)	This Study
Urine	EfsC17	SAMN33741196	PRJNA944190	2,954,172	97x, 215x	2	-	Illumina, MinION	7/2/18	USA (Dallas, TX)	This Study
Urine	EfsC23	SAMN33741198	PRJNA944190	3,019,668	576x, 356x	3	-	Illumina, MinION	7/23/18	USA (Dallas, TX)	This Study

Needed to reproduce the genome collection described in the study and to demonstrate rigor of genome selection

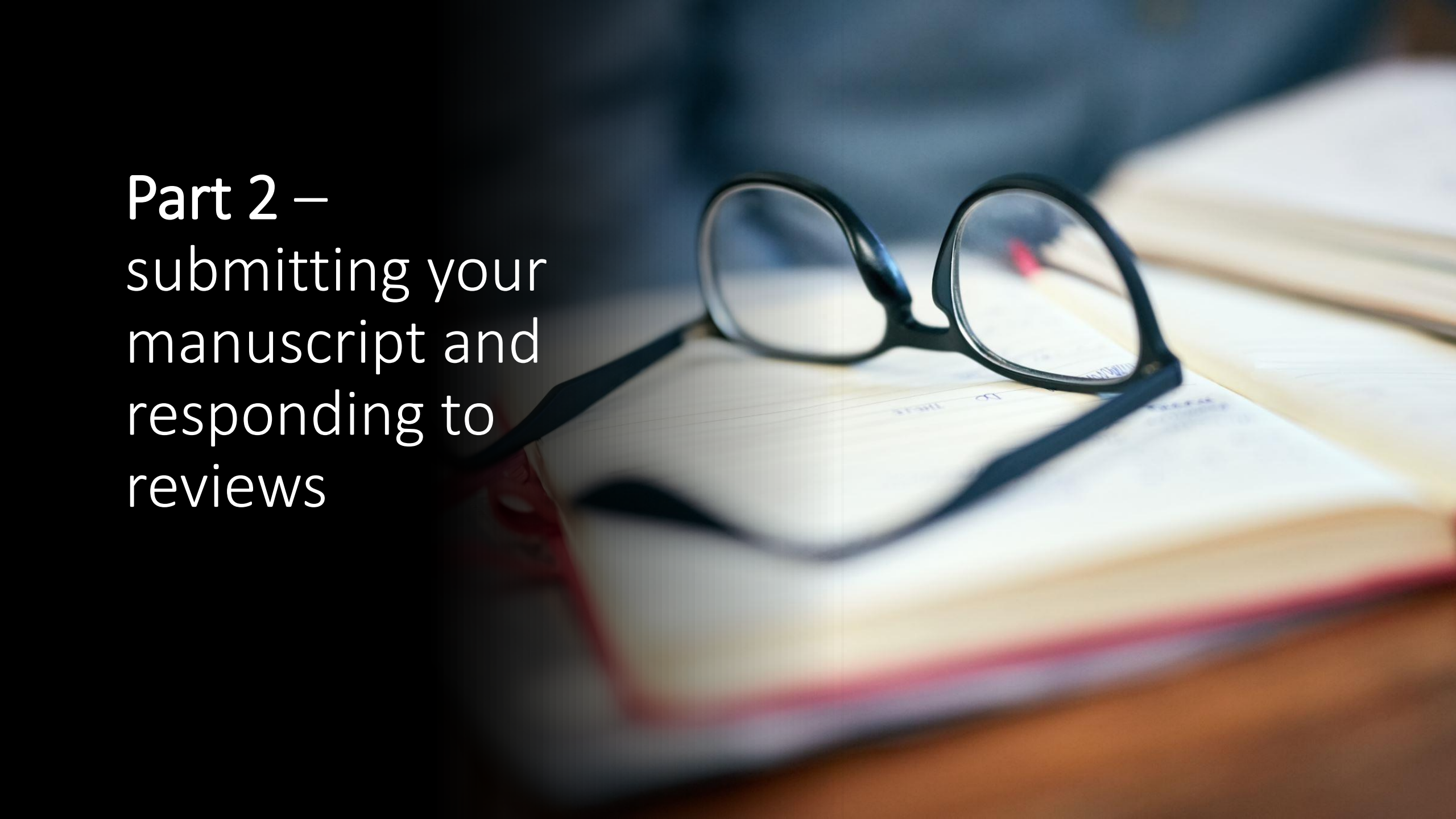
Concluding remarks – manuscript writing

- Start manuscript writing early
  - Don’t be afraid to make figure panel outlines before all experiments are complete – can use placeholders and will go through many revisions before its final!
- Clearly communicate with your research mentor about expectations
  - The scope of your project may change – discuss expectations/thresholds for when your project may be publishable
  - What word processor do they want you to use? Should you always track changes? What programs to use for figure preparation?
  - Do they want to see a single complete (not final) draft, or do they want you to send individual sections as they are complete?
- Try not to be a perfectionist
  - Your draft does not have to be perfect before sharing
  - Your mentor will likely edit your first draft heavily - this is normal!! Do not be discouraged!

Comments, Questions?

Please raise your hand or type in the chat!

Part 2 – submitting your manuscript and responding to reviews



Manuscript submission

- Usually, your advisor will submit the manuscript since they are ***corresponding author***
- Your advisor will have to compose a cover letter to include with the submission – I usually ask my lead author to review my cover letter
- Each journal will have different submission requirements, your advisor may ask you to update parts of the manuscript to meet them
- A helpful thing to do is to make a separate document with author information – full name, email address, and affiliation that your advisor can use to fill in author fields
- During submission everything is compiled into a PDF and tables and figures can sometimes get altered – check the preview to make sure conversion is OK

The Peer Review Process

- After submission, your manuscript will either go out for peer review or will be editorially rejected
  - Editorial rejection means that the editor does not think your paper fits the scope of their journal
- Once sent for peer review four decisions are usually possible: accept, minor revision, major revision, reject
  - Some journals have an option called “reject but allow resubmission”
  - If flat out rejected, you have to move on to another journal
  - Very unlikely that your paper will be accepted after the first round of review
  - Peer review and revisions are a normal and healthy part of the scientific process

Tips for responding to reviews

- Read through the reviews with your advisor and come up with a strategy on how to address them
  - Take a couple of days to digest the comments
- You do not have to do everything the reviewers ask!
  - A good editor will clearly state the experiments they expect to be done or points that will need to be addressed in a successful revision
  - If a reviewer asks for an experiment that is clearly out of the scope of the paper, you do not have to do it – your advisor may ask the editor
  - Try to see the reviewer's perspective when considering which revisions to make and which to refute
  - Make sure your arguments refuting requested experiments or revisions are logical, well-supported, and communicated in a ***neutral tone***

Tips for responding to reviews contd.

- Paste the reviewer comments into a word document
- Respond individually to each comment in a different color to make it easy to distinguish your responses (make it easy!)
  - Some journals require a tracked changes version of the manuscript
  - Indicate all textual changes made in the response to reviewers manuscript and provide line numbers
  - Provide supporting citations if refuting textual changes requested by the reviewer
  - If a reviewer's request is unclear, paraphrase what you think they are asking and politely respond – your advisor can also contact the editor to clarify
  - Thank the reviewers for their suggestions but do not be overly saccharin in your responses – aim to be be clear, succinct and logical

Example response to reviewers document

- You can find a public response to reviewers document to one for one of our recent publications here: https://www.microbiologyresearch.org/content/journal/acmi/10.1099/acmi.0.000508.v3

5) On line 175, the authors mention using ResFinder. For an underrepresented species, I recommend using ABRicate with the CARD database (using a 50% pident cutoff (--mincov 50)). I also recommend using the VFDB database with ABRicate.

ABRicate was used to query the CARD database at identity and coverage cutoffs set to 50%. Resistance gene results are reflected in table 1 and further revisions are in lines 433-437. ABRicate was also used to query the VFDB database at default cutoffs. Results are specified in lines 423-429. The methods have been revised in lines 197-199.

Example response to reviewers document - rebuttal

7) I am also concerned about the author's choice of a critical $\alpha < 0.05$, and strongly advise them to use $\alpha < 0.005$ (see https://www.nature.com/articles/s41562-017-0189-z). Could the authors also explain how they calculated their power statistic? I am aware of how lowering α to 0.005 will affect the COG results. As such, I recommend extracting the GO terms from the Prokka (or Bakta) tsv files and performing a hypergeometric using GOATools to better investigate potential functional differences. This can be performed without a multiple tests correction if the authors use a cut-off of 0.005. I am sorry that this creates a relatively unpleasant amount of work, but I am not convinced of the statistical results presented. This is not a fault of the authors, COG categories are too broad to be meaningful. To help with this, if the authors load their significant GO terms into REVIGO, it can help find common motifs amongst similar GO sets.

Thank you for your thoughtful recommendations. Our COG analysis is exploratory and is a survey providing a starting point for investigating differential gene content in the megaplasmid of *E. raffinosus*. With only 4 complete genomes, there is insufficient data for a more robust analysis. Therefore, we include in the supplemental material table nominal p-values, leaving significance interpretation to the reader. We revised the discussion about significant functional enrichments to better reflect the limitations of this statistical analysis and we would like to note that the word “significant” is absent from the revised manuscript. Please refer to the section starting in line 446 discussing these results. We appreciate the reviewer’s suggestion to extract GO terms, but we think this additional in-depth analysis would be better employed once there are more complete *E. raffinosus* genomes available to analyze. We will keep these suggestions in mind for future work in this area.

Thank you for listening!!

Comments, Questions?

Please raise your hand or type in the chat!