Beyond the text: Figures and Tables
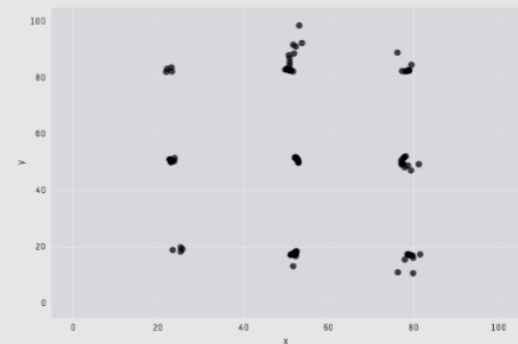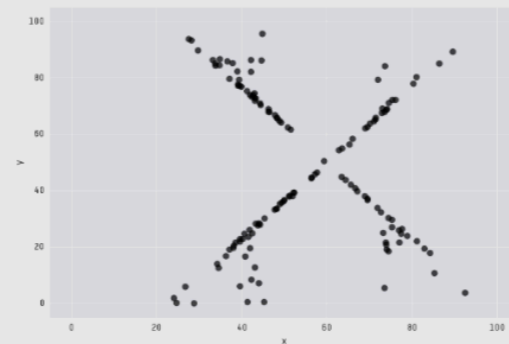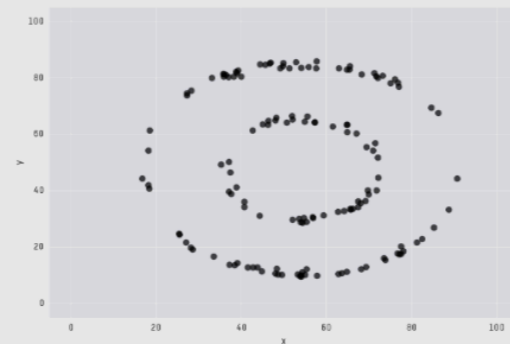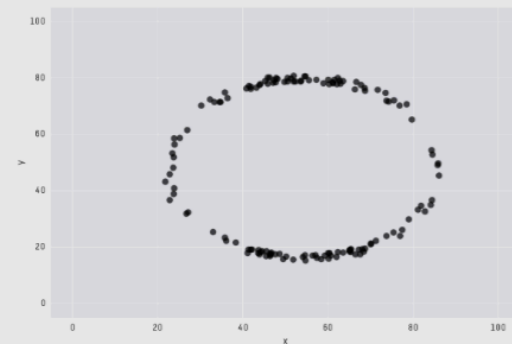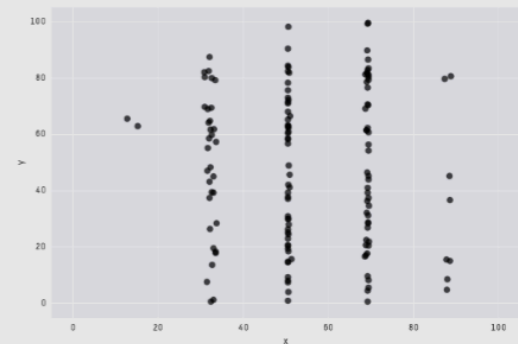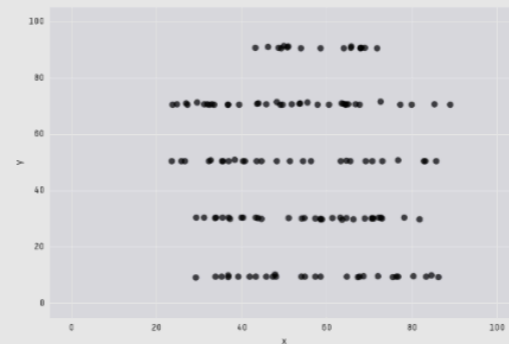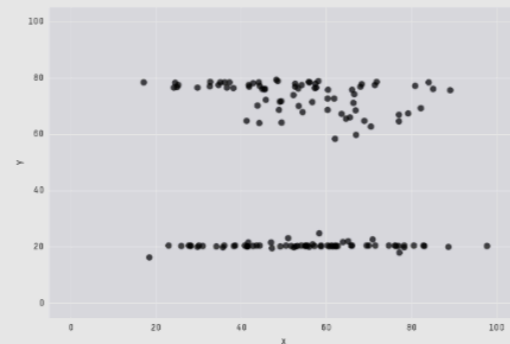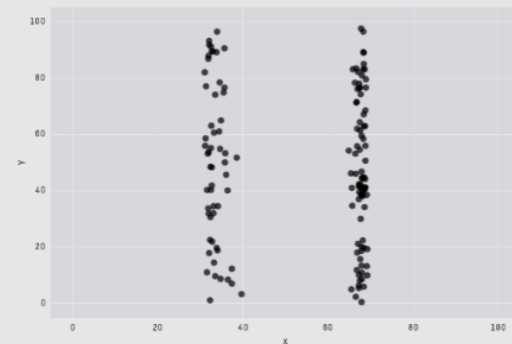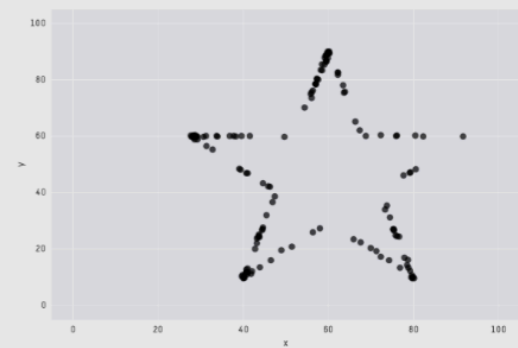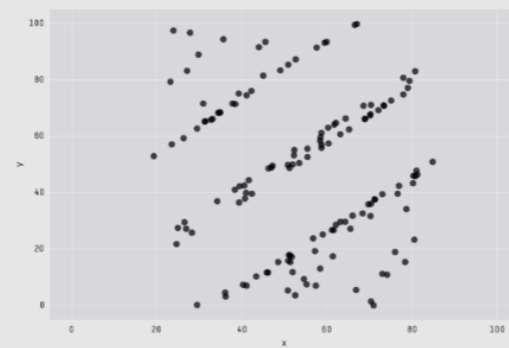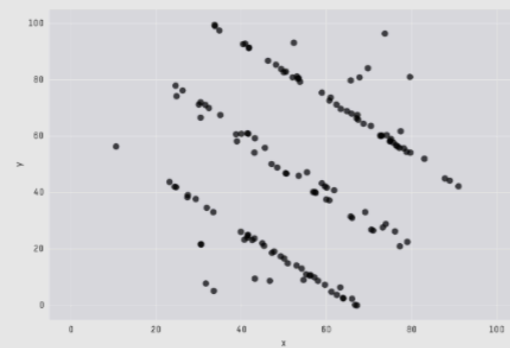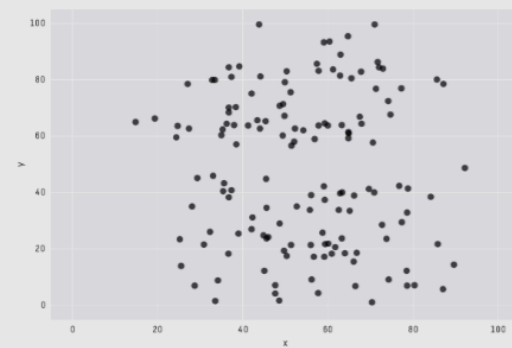
Data summaries: Where do the numbers go?

"Golden-mantled tree-kangaroo home ranges were on average 1.7 km^2 (± 0.9 SD), while Lumholtz's tree-kangaroos had home ranges of 3.3 km^2 (± 1.8), Bennett's tree-kangaroos 1.6 km^2 (± 0.7), and grizzled tree-kangaroos 2.2 km^2 (± 1.1)".

Species	Home range size (mean)	SD
Golden-mantled	1.7	0.9
Lumholtz's	3.3	1.8
Bennett's	1.6	0.7
Grizzled	2.2	1.1

Tree kangaroo home ranges
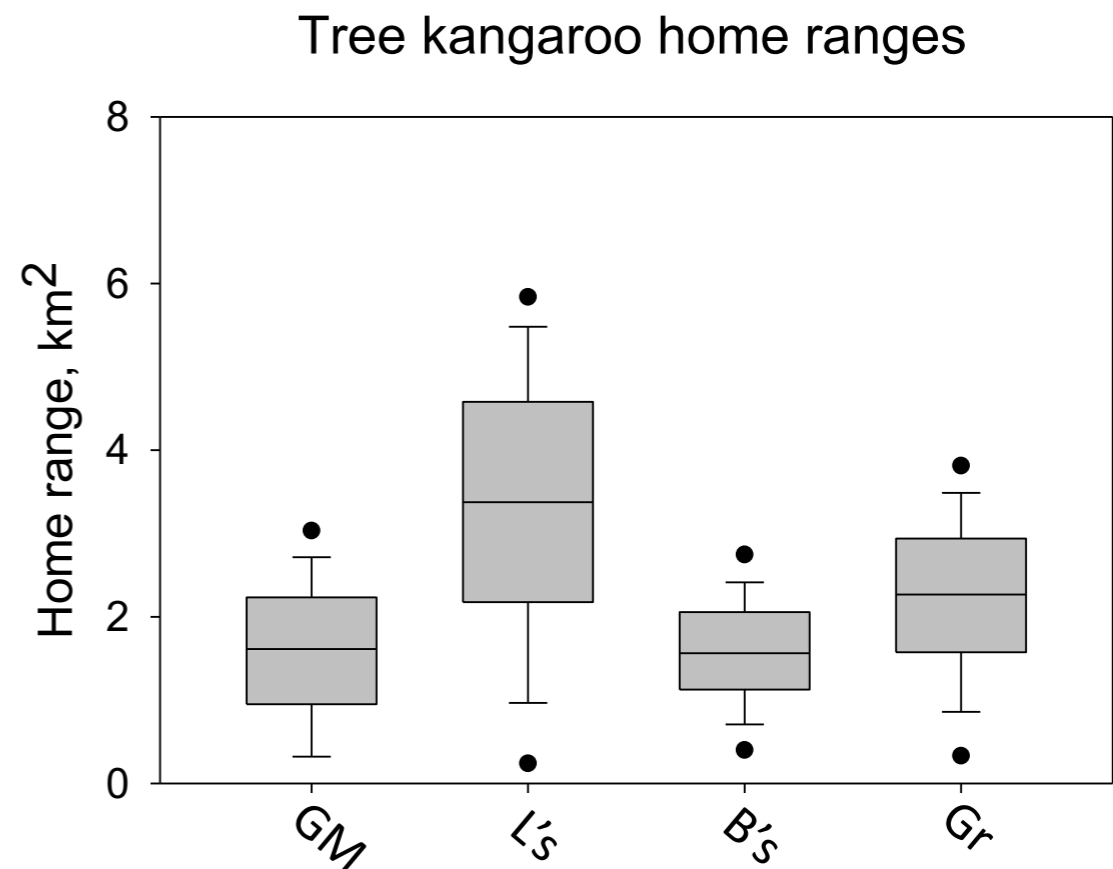
Data summaries: Where do the numbers go?

Use a figure to
- Show trends, patterns, and relationships across and between data sets, when the trend is more important than the precise data values
- Summarize the results of your research
- Visually explain a sequence of events, phenomena, characteristics, or geographical features

Use a table to
- Show many and precise numerical values in a small space
- Compare/contrast values or characteristics of items that are related or that share several characteristics or variables
- Show the presence or absence of specific characteristics

Use text when
- The data you intend to present is not relevant to the main findings of your study
- The data is not too complex or large, and can be easily incorporated in your manuscript
- Your data, if presented as a table, would warrant only 1 or 2 columns

Figures
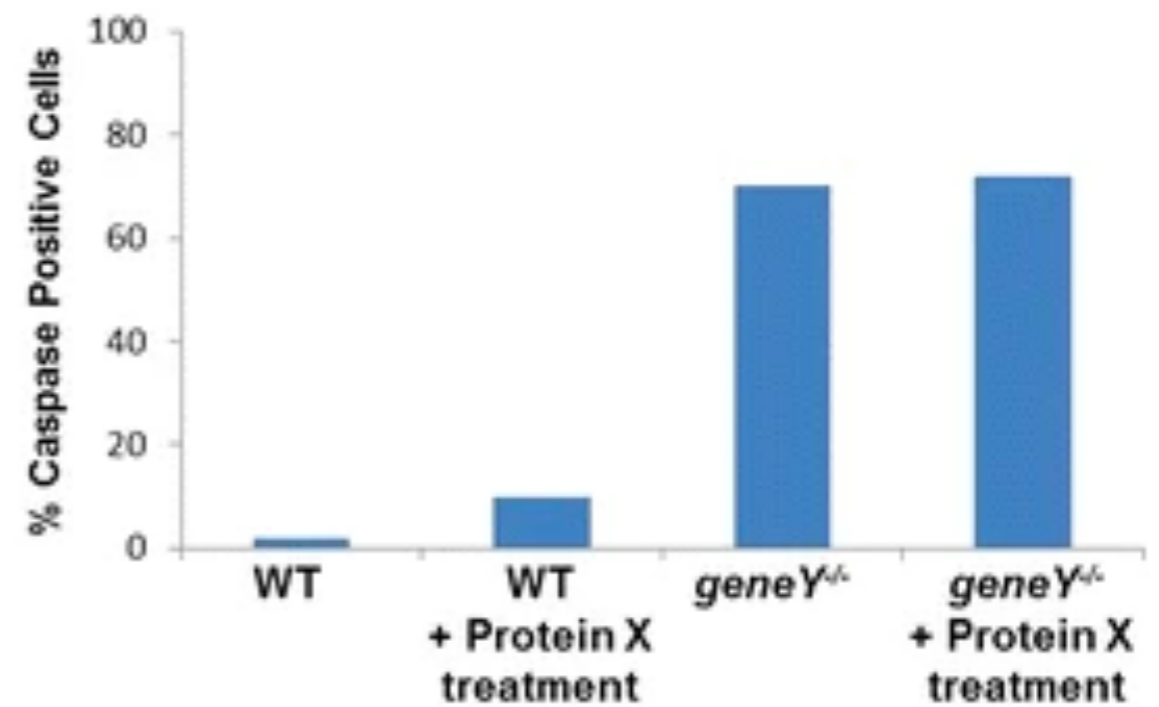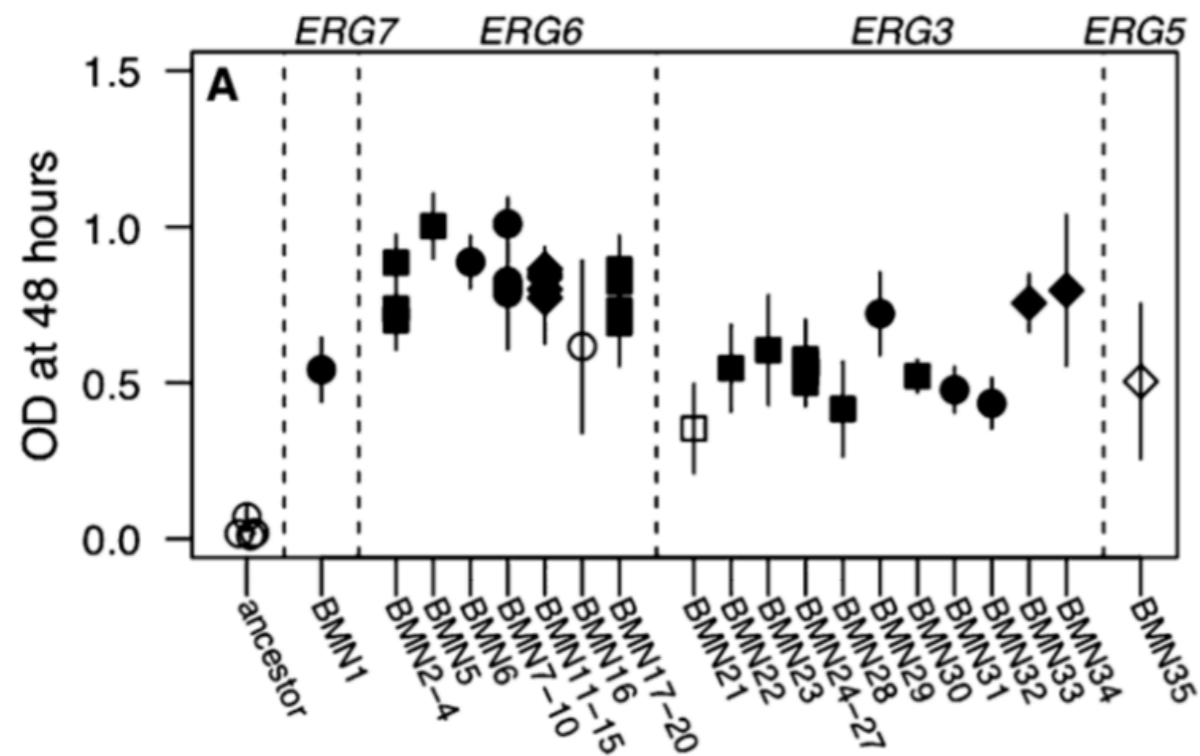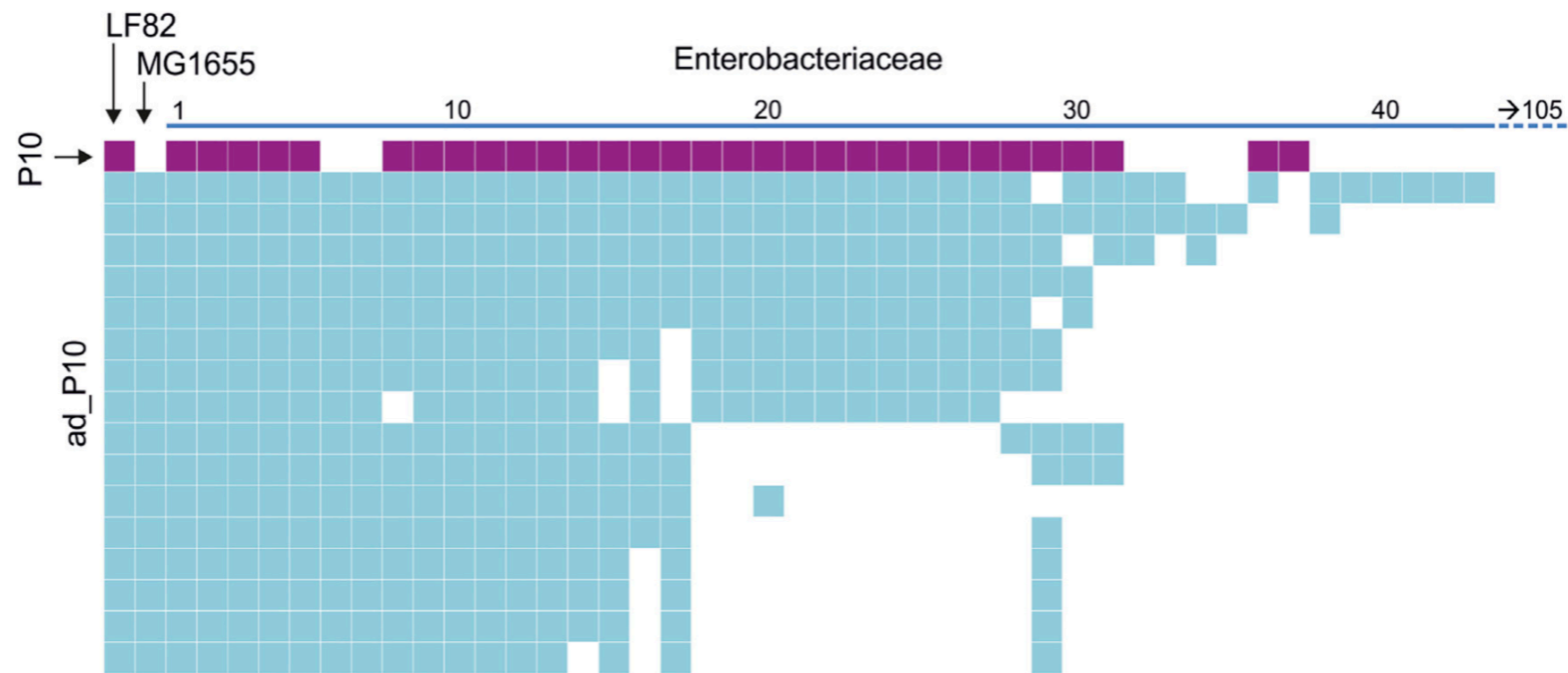
Many elements to figure design
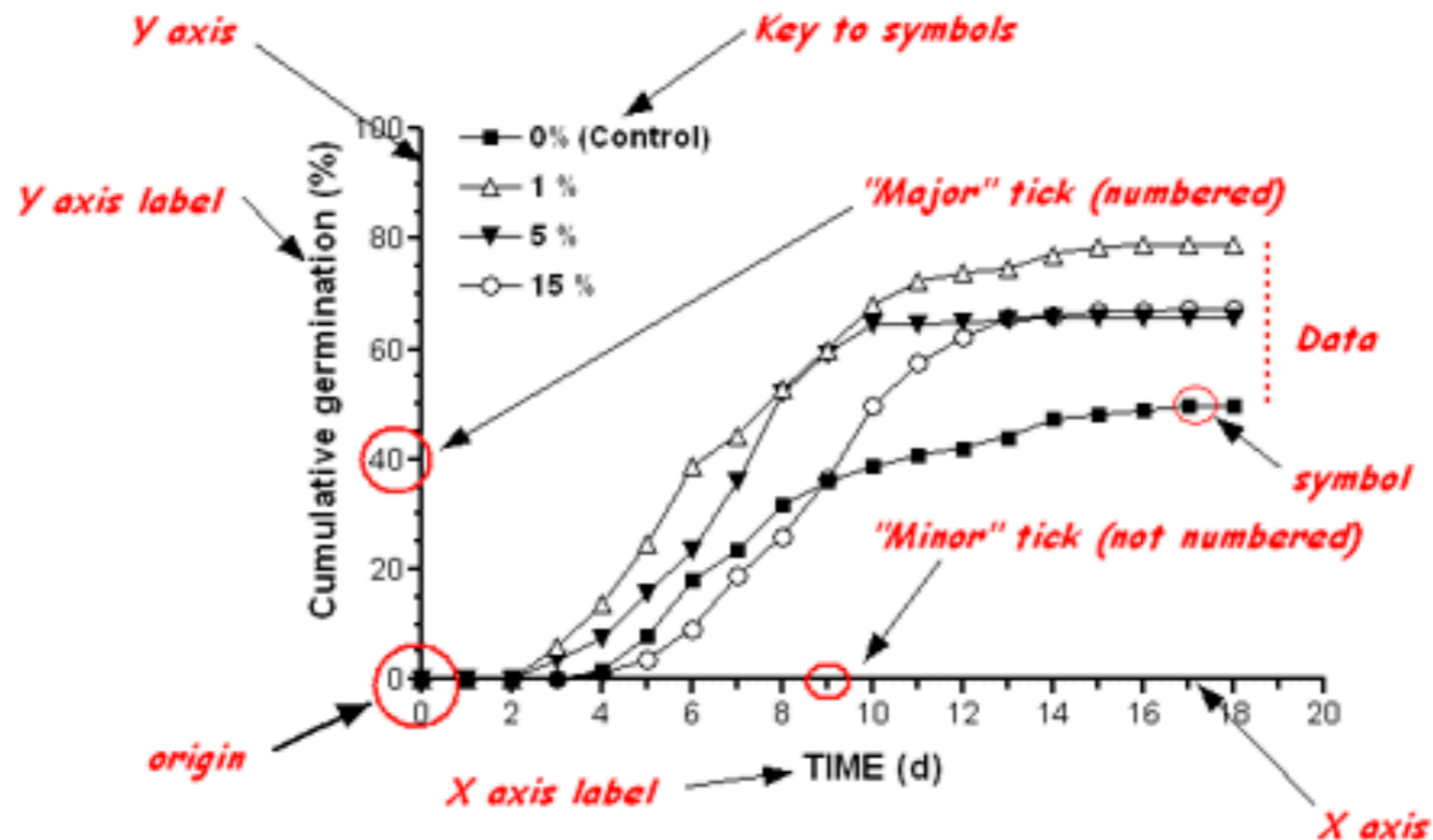
Figure 1. Cumulative germination of *Chenopodium* seeds after pregermination treatment of 2 day soak in NaCl solutions. n = 1 trial per treatment group (100 seeds/trial.)

Identify Your Message

#1 tip to anyone trying to hone their SciComm skills: the most effective images **say one thing and say it well**. So, before you start, come up with a single message or story that your image will convey and show it in the **simplest** way possible. If you need to show *multiple* concepts, break it down into multiple figures.

Savanna Jackson. Improving your Science Figures: What I learned from conversations with 100+ scientists
https://blog.quartzy.com/improving-your-science-figures-what-i-learned-from-conversations-with-100-scientists

Graphs should "represent all the data that is needed to see and understand what's meaningful."

Few S (2011) The chartjunk debate, a close examination of recent findings. Visual Business Intelligence Newsletter.
http://www.perceptualedge.com/articles/visual_business_intelligence/the_chartjunk_debate.pdf.

GOAL 1: Clarity. Present the data in a way that the audience can easily see what is meaningful.

What type of data do you have?

Table 2.1: Types of variables encountered in typical data visualization scenarios.

Type of variable	Examples	Appropriate scale
quantitative/numerical continuous	1.3, 5.7, 83, 1.5×10^{-2}	continuous
quantitative/numerical discrete	1, 2, 3, 4	discrete
qualitative/categorical unordered	dog, cat, fish	discrete
qualitative/categorical ordered	good, fair, poor	discrete
date or time	Jan. 5 2018, 8:03am	continuous or discrete
text	The quick brown fox jumps over the lazy dog.	none, or discrete

Fundamentals of Data Visualization, Claus O. Wilke, https://serialmentor.com/dataviz/

Amounts
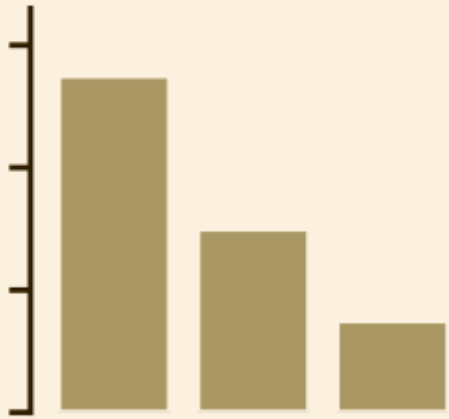
Bars

Dots
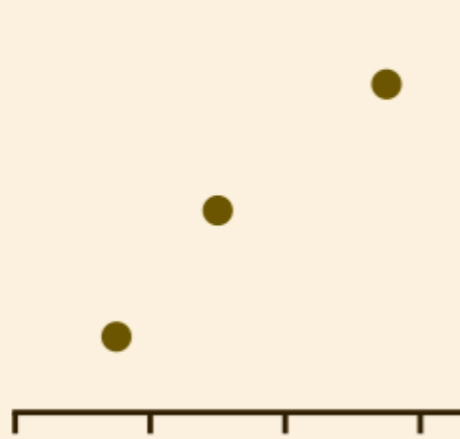
Grouped Bars
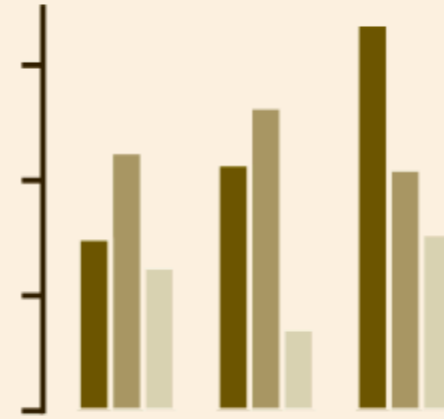
Stacked Bars
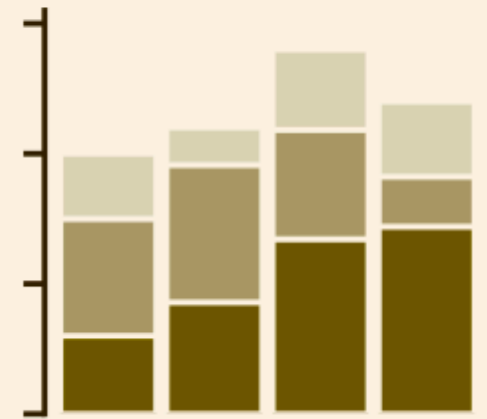
Distributions

Histogram
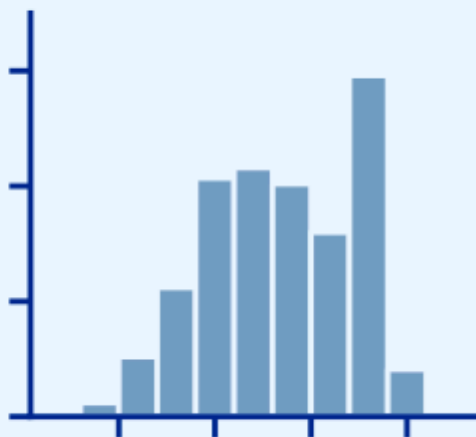
Sina Plots
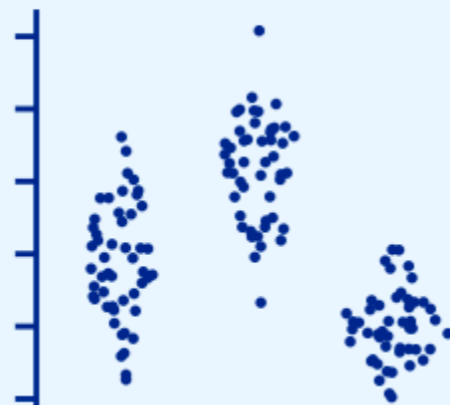
Overlapping Densities
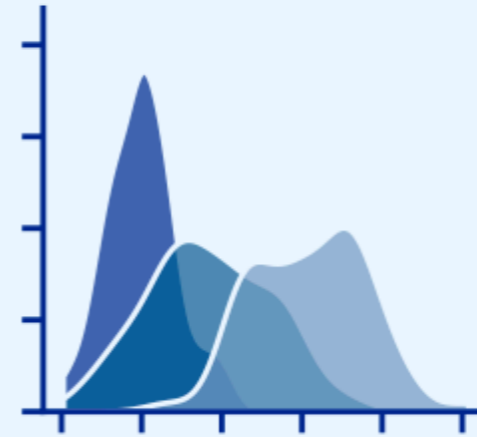
Ridgeline Plot

Proportions
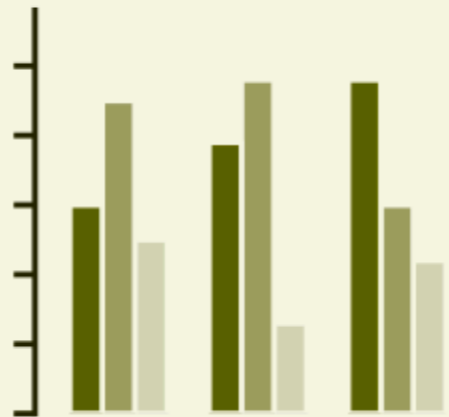
x-y relationships
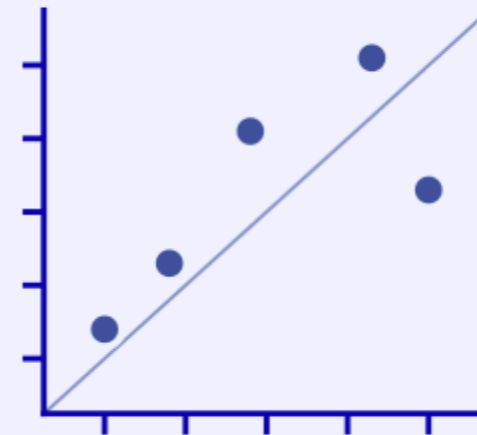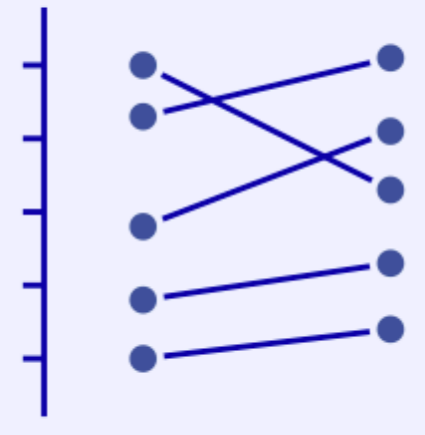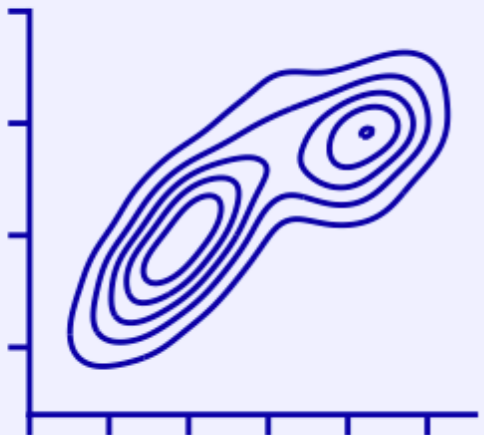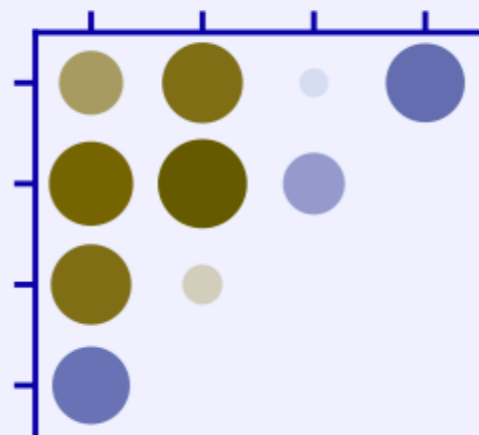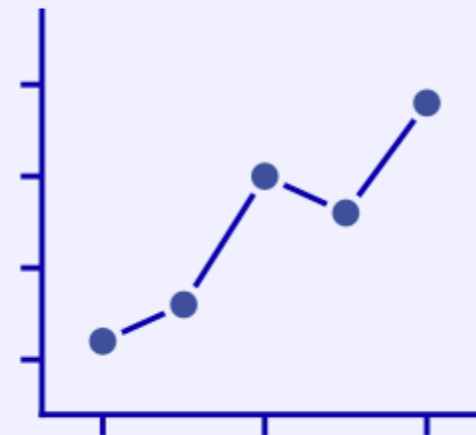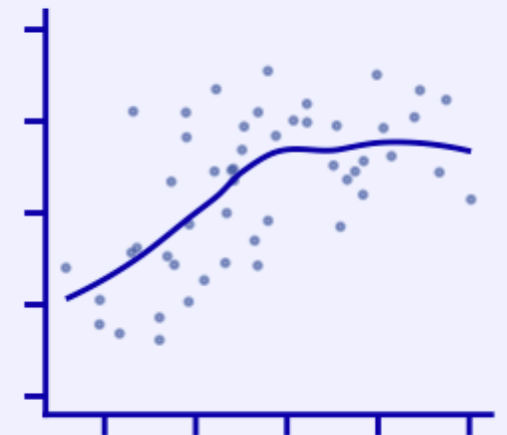
Uncertainty
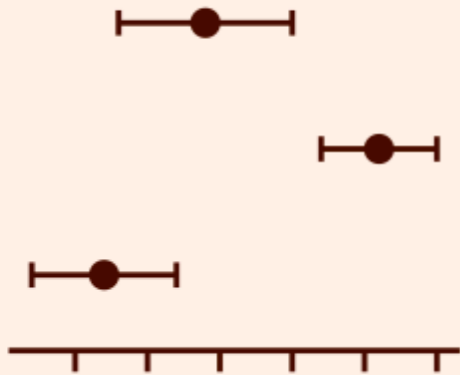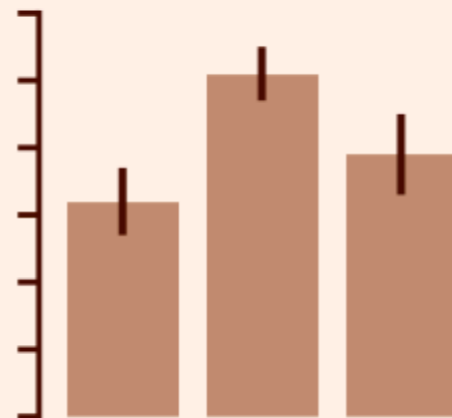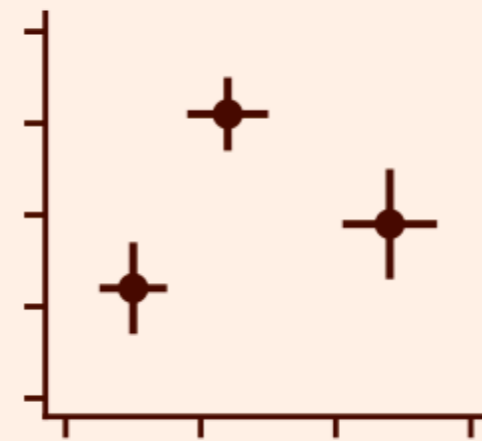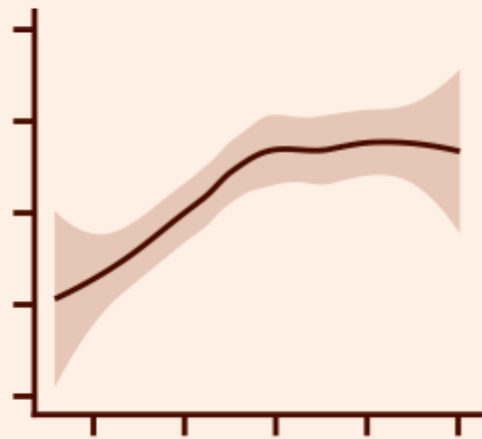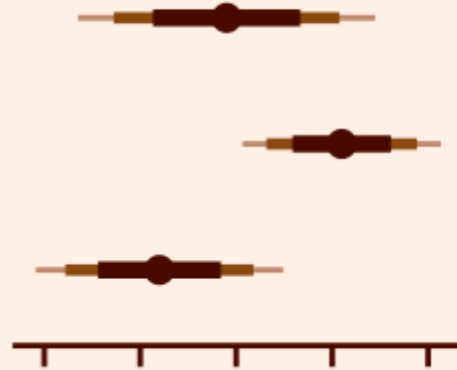
GOAL 2: Precision. Represent data accurately.

The principle of proportional ink: The sizes of shaded areas in a visualization need to be proportional to the data values they represent.
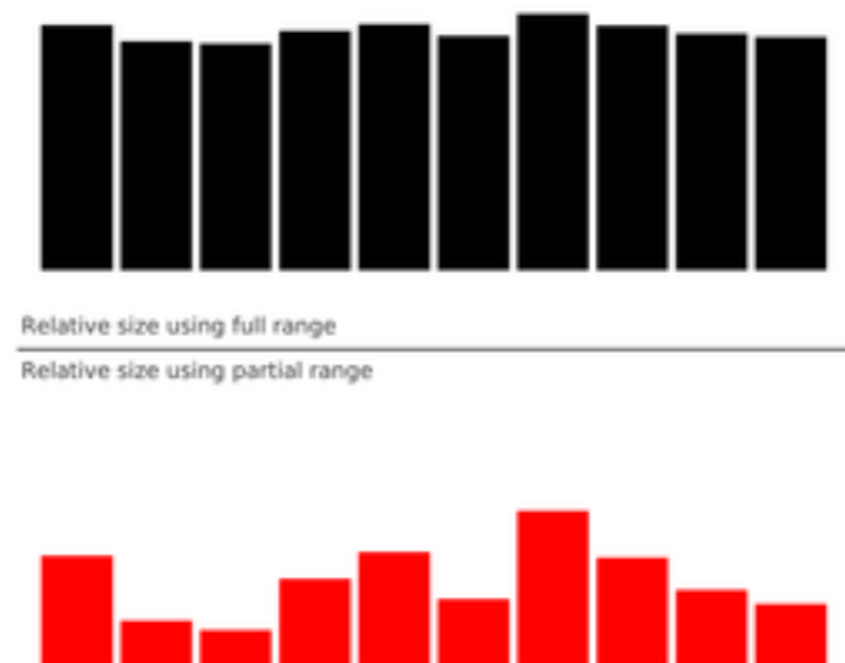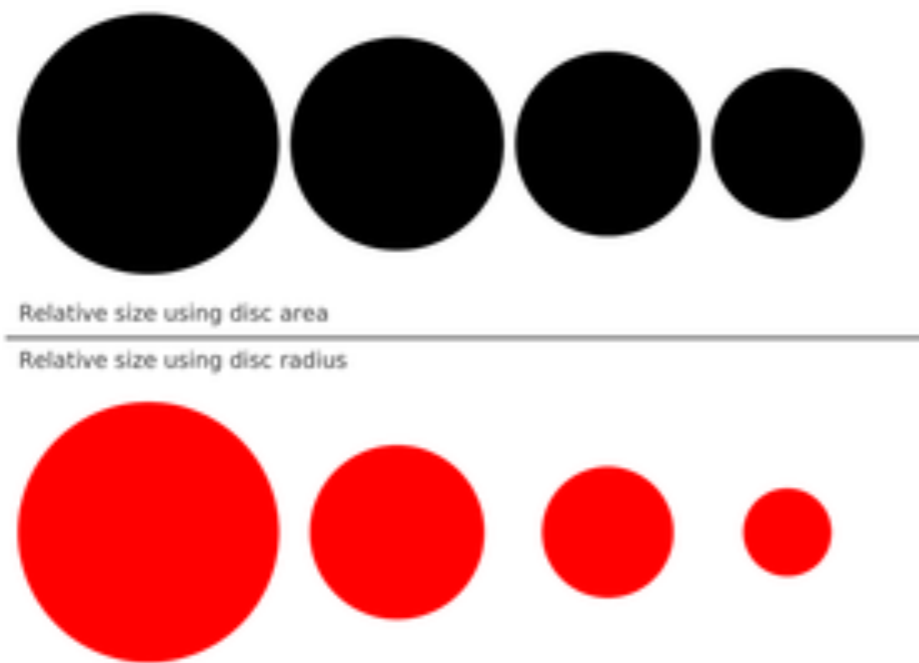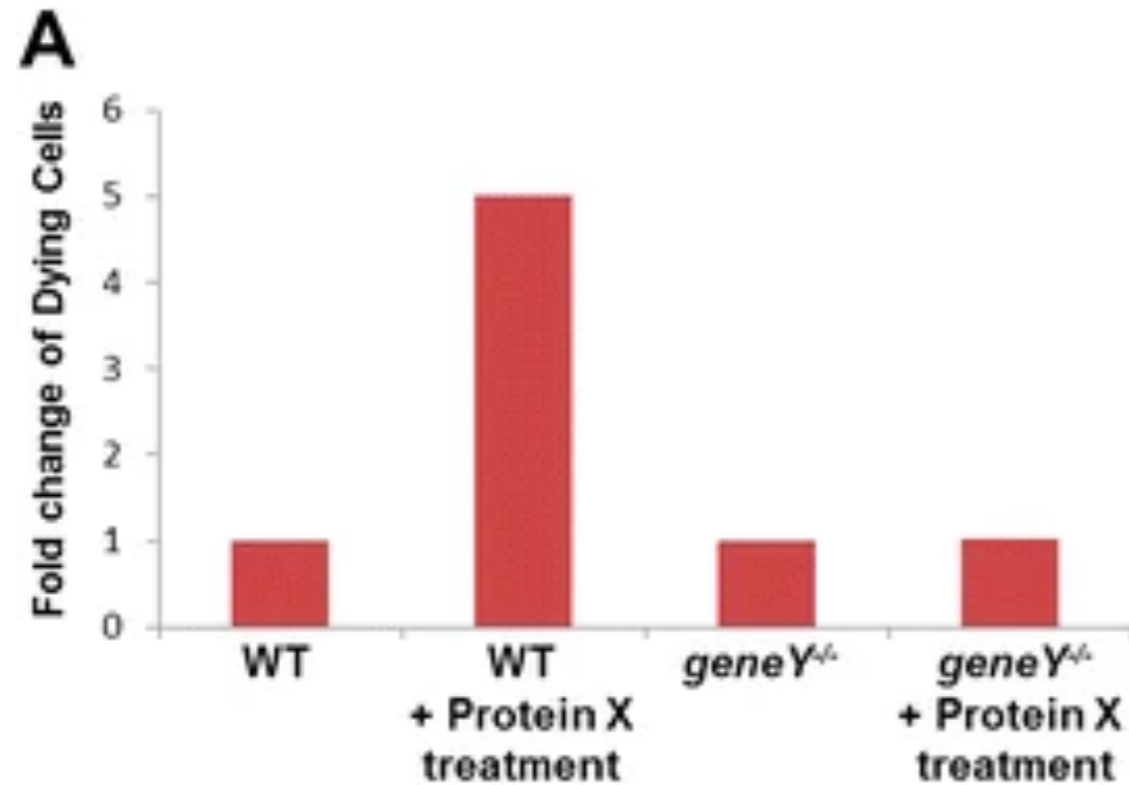
Wilke CO, Fundamentals of Data Visualization,, https://serialmentor.com/dataviz/
Rougier NP, Droettboom M, Bourne PE (2014) Ten Simple Rules for Better Figures. PLOS Computational Biology 10(9): e1003833. https://doi.org/10.1371/journal.pcbi.1003833

GOAL 2: Precision. Represent data accurately.

a Fold change of percentage of dying cells in wild-type (*WT*) and mutant samples, either treated with pro-apoptotic factor Protein X or untreated. For each genotype, cells treated with Protein X were compared with the respective untreated control.

GOAL 2: Precision. Represent data accurately.
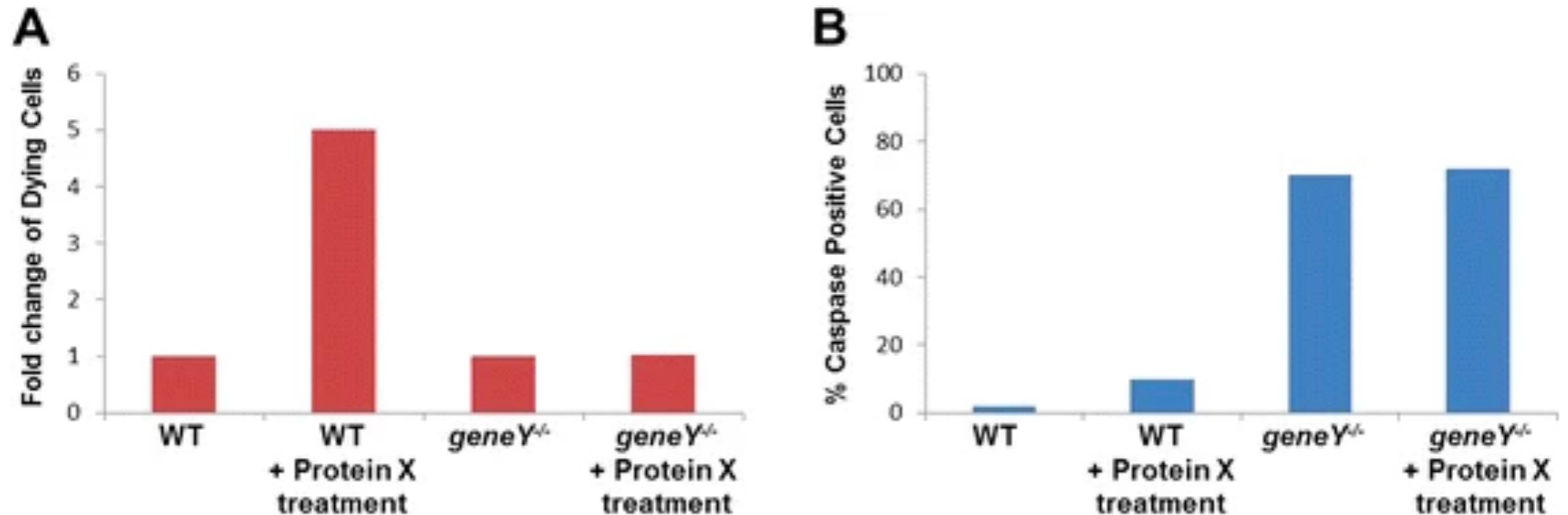
a Fold change of percentage of dying cells in wild-type (*WT*) and mutant samples, either treated with pro-apoptotic factor Protein X or untreated. For each genotype, cells treated with Protein X were compared with the respective untreated control. **b** Percentages of dying cells measured in wild-type and mutant samples, either treated with pro-apoptotic factor Protein X or untreated. Percentages are calculated as the number of Caspase-positive cells per 1000 cells

GOAL 2: Precision. Represent data accurately.
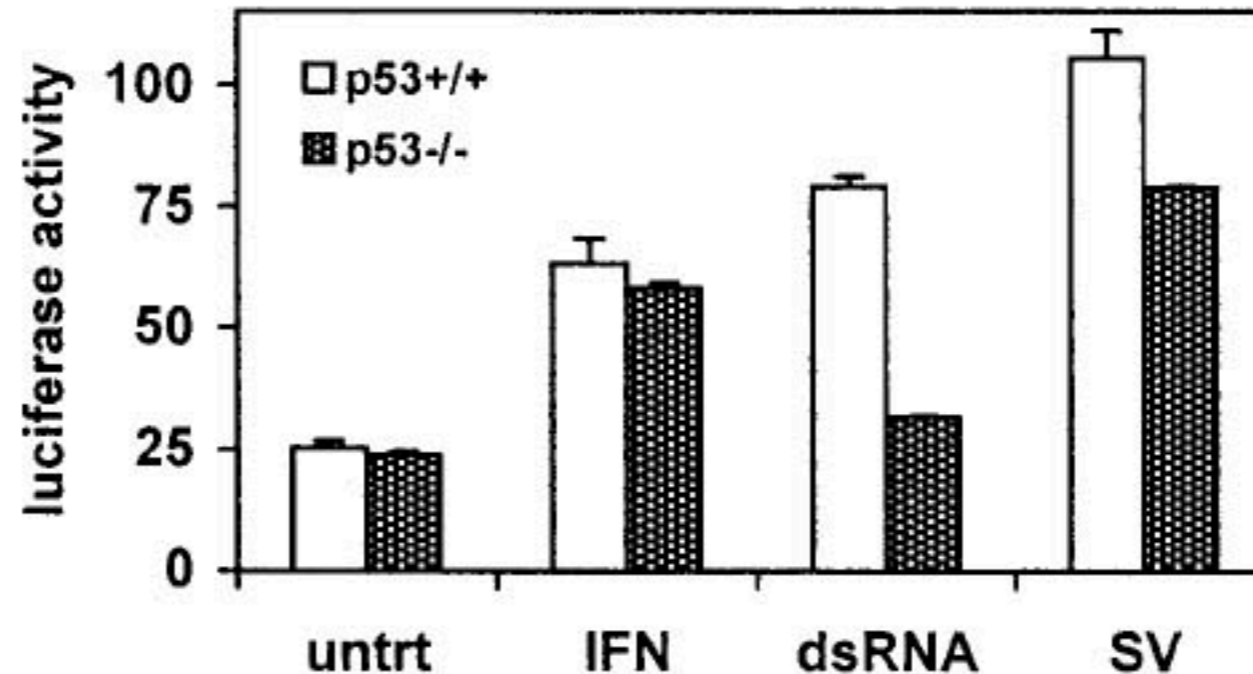
FIG. 4. ISG15 promoter activity mimics endogenous ISG15 mRNA regulation by p53, dsRNA, and virus. Cells (6×10^5 HCT 116) were seeded in 32-mm plates and allowed to attach overnight. Cells were transfected with 500 ng of pGL3/ISG15-Luc, 50 ng of pRL null (Promega), and 450 ng of pcDNA3 for carrier DNA by using Lipofectamine Plus (Life Technologies) following the manufacturer's instructions. Twenty-four hours posttransfection, the medium was aspirated and replaced with medium containing either 1,000 U of IFN-α/ml, 50 μg of dsRNA/ml, or Sendai virus (multiplicity of infection, 10). Cells were incubated for 12 h and then lysed, and luciferase assays were performed. Luciferase activity was assessed on 20 μl of each lysate as directed by the supplier (Dual Luciferase Kit, Promega) using a TD 20/20 luminometer (Turner Designs). Luciferase activity is presented as the ratio of firefly activity to renilla activity to control for differences in transfection efficiency. Each data point is the mean of triplicate samples ± the standard error; the data presented are representative of four independent experiments.

Hummer BT, Li XL, Hassel BA (2001) Role for p53 in gene induction by double-stranded RNA. *J Virol* 75:7774-7777, Figure 4
https://www.biostat.wisc.edu/~kbroman/topten_worstgraphs/

GOAL 3: Ease. Choose data visualizations that are easy to perceive and interpret.

position

shape

size

color

line width

line type
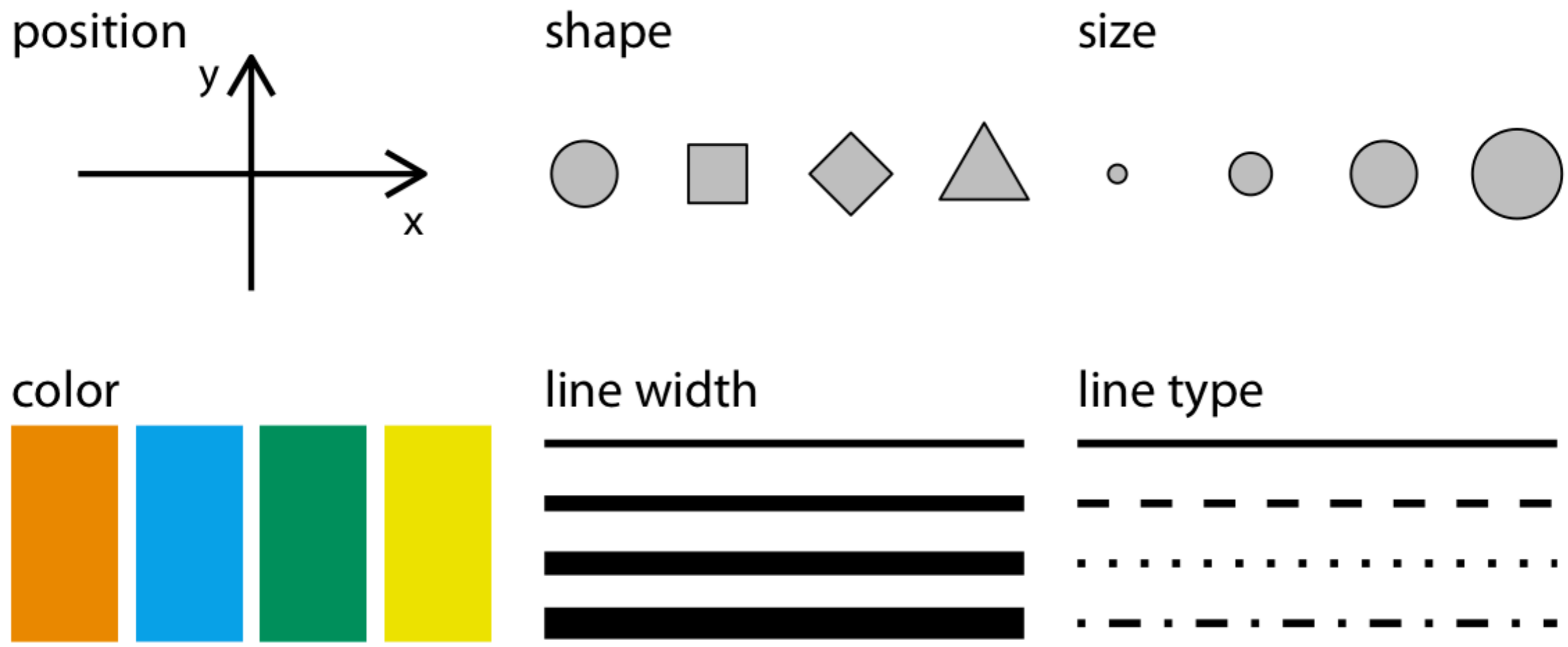
GOAL 3: Ease. Choose data visualizations that are easy to perceive and interpret.

Sequential: one variation of a unique color, used for quantitative data varying from low to high.

Qualitative: rapid variation of colors, used mainly for discrete or categorical data.

Diverging: variation from one color to another, used to highlight deviation from a median value.
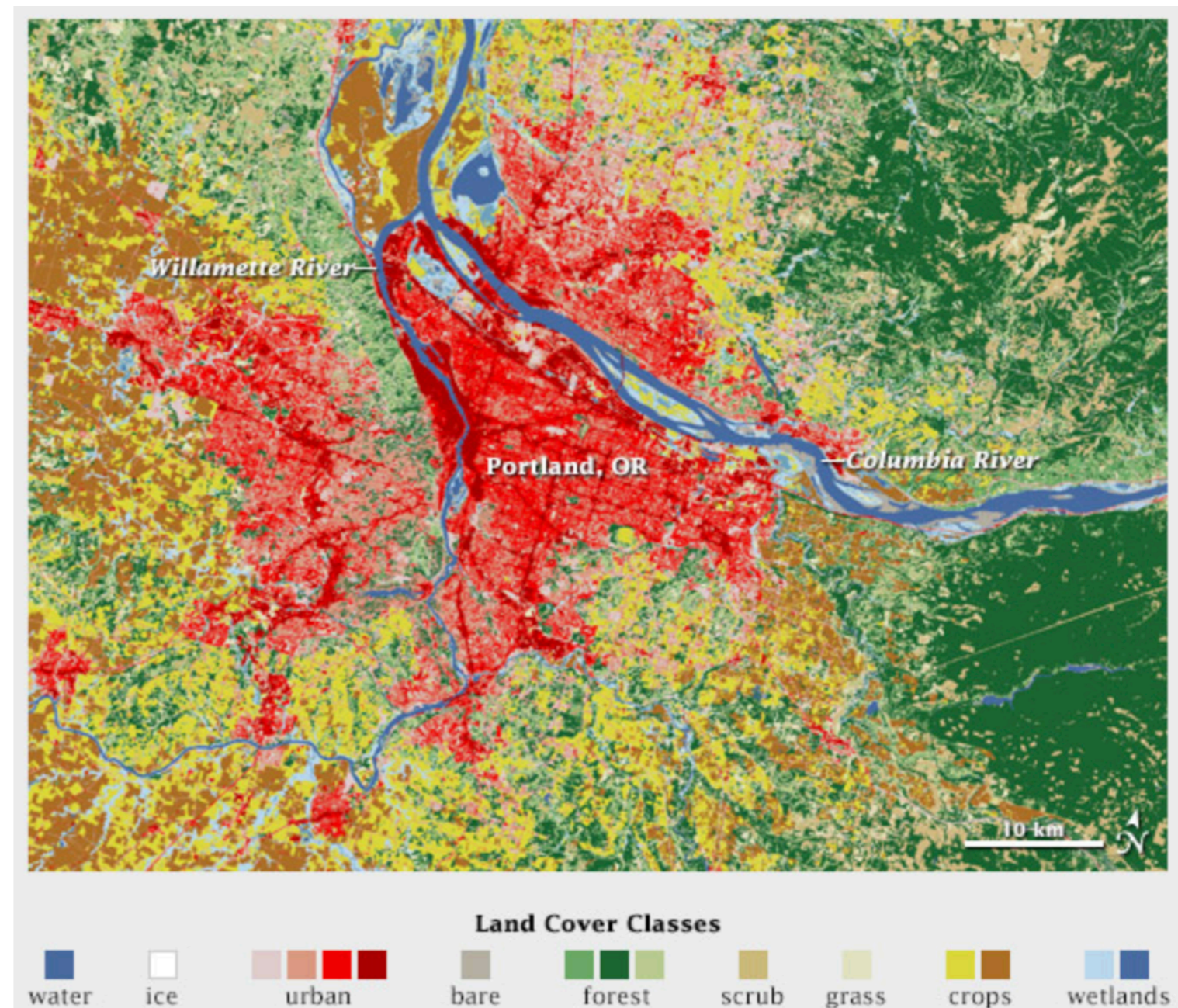
Rougier NP, Droettboom M, Bourne PE (2014) Ten Simple Rules for Better Figures. PLOS Computational Biology 10(9): e1003833. https://doi.org/10.1371/journal.pcbi.1003833
https://earthobservatory.nasa.gov/blogs/elegantfigures/2013/08/12/subtleties-of-color-part-3-of-6/

GOAL 3: Ease. Choose data visualizations that are easy to perceive and interpret.

Distribution of All TFBS Regions

Pseudogene/ambiguous 17%

5' to known gene 22%

Novel 24%

Within or 3' flanking to a known gene 36%

866 Total TFBS Regions
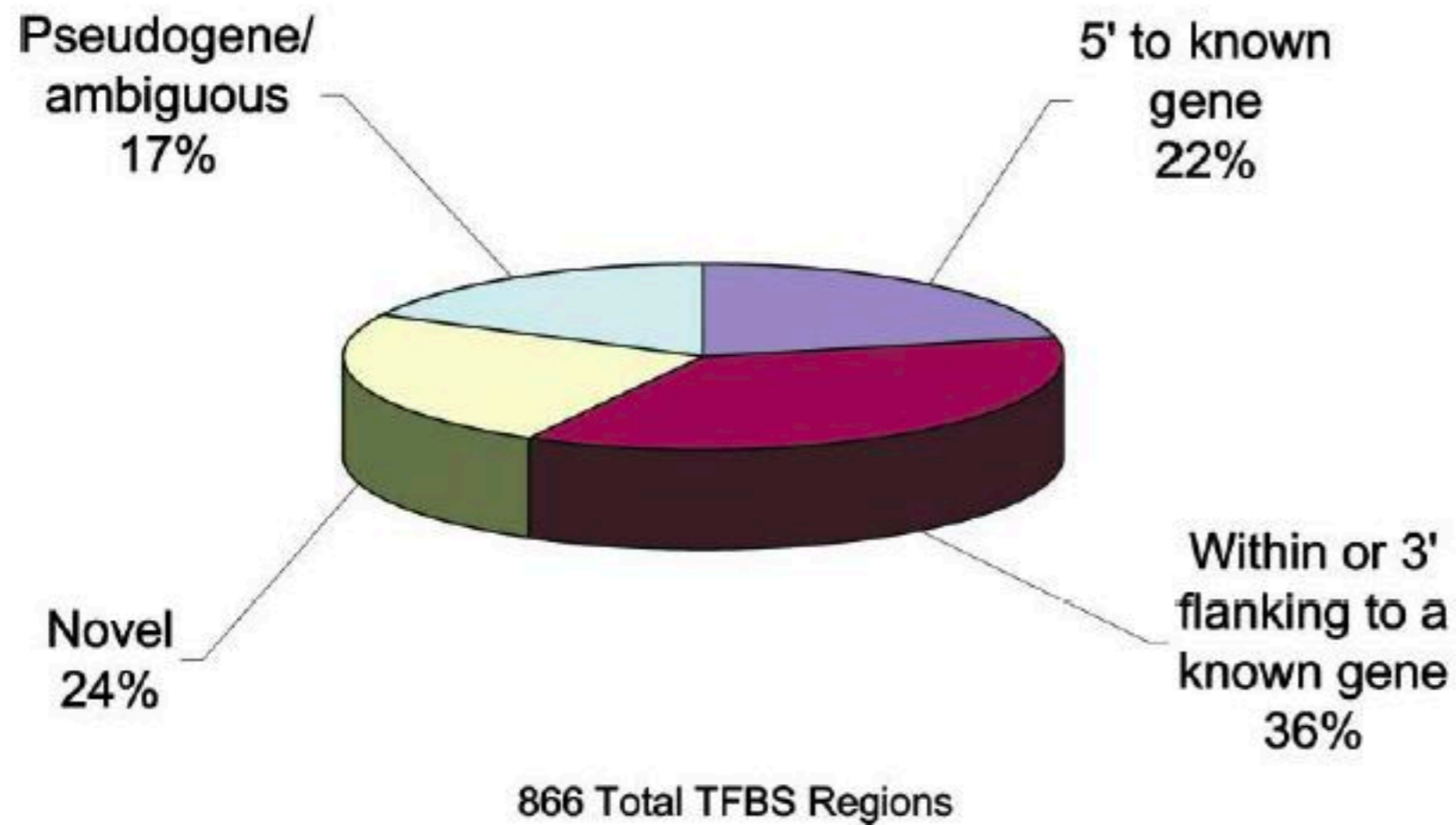
Figure 1. Classification of TFBS Regions

TFBS regions for Sp1, cMyc, and p53 were classified based upon proximity to annotations (RefSeq, Sanger hand-curated annotations, GenBank full-length mRNAs, and Ensembl predicted genes). The proximity was calculated from the center of each TFBS region. TFBS regions were classified as follows: within 5 kb of the 5' most exon of a gene, within 5 kb of the 3' terminal exon, or within a gene, novel or outside of any annotation, and pseudogene/ambiguous (TFBS overlapping or flanking pseudogene annotations, limited to chromosome 22, or TFBS regions falling into more than one of the above categories).

Cawley S, et al. (2004) Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. Cell 116:499-509
https://www.biostat.wisc.edu/~kbroman/topten_worstgraphs/

GOAL 3: Ease. Choose data visualizations that are easy to perceive and interpret.
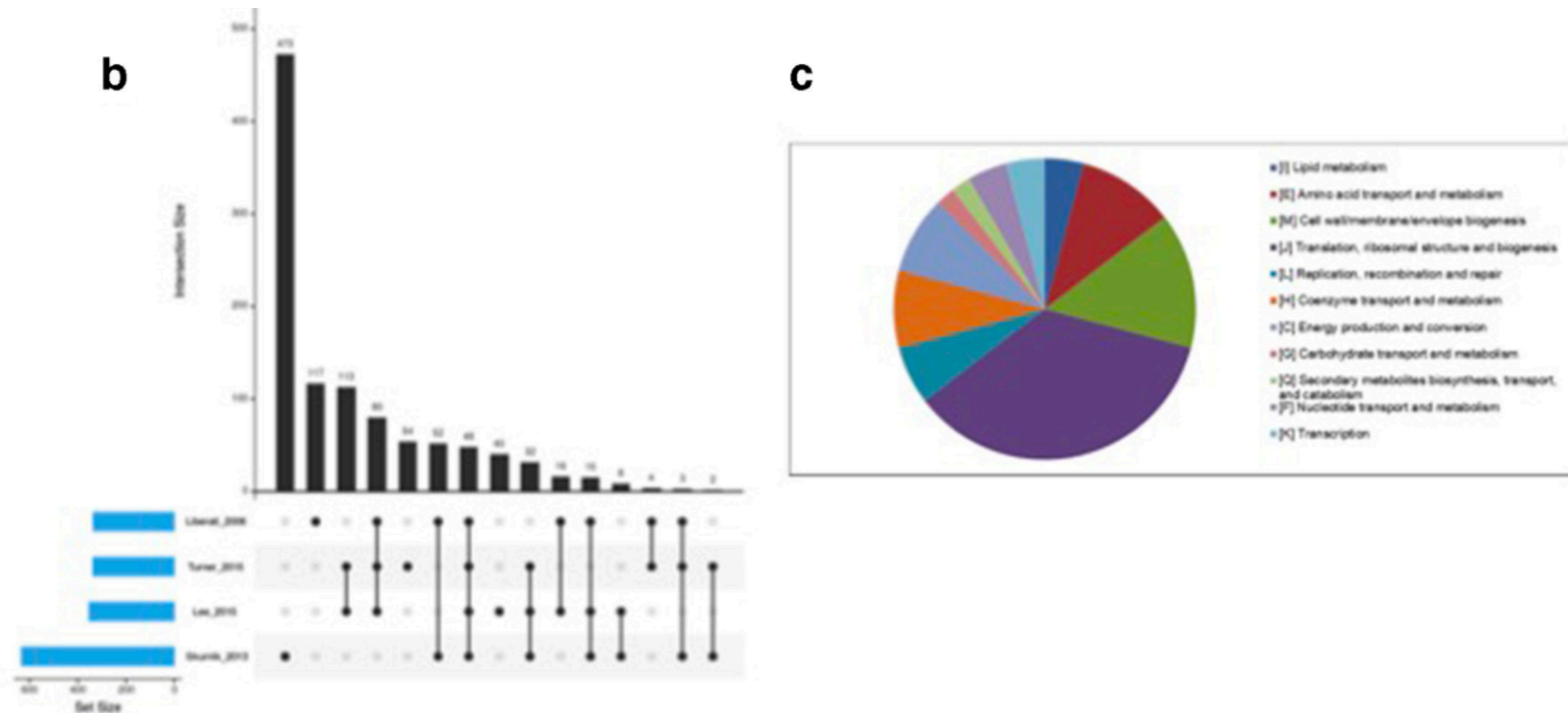
Fig. 1. Reported variability of *P. aeruginosa* essential genes. (a) Timeline of studies made to determine essential genes with different aims and with several experimental approaches in *P. aeruginosa* PAO1 and PA14. (b) UpsetR image of the essential genes reported by Liberati *et al.* [15], Skurnik *et al.* [16], Turner *et al.* [18] and Lee *et al.* [19]. These comparisons show that there are 48 genes that were reported as essential in the four studies. (c) Analysis of the function of the 48 genes that are shared by the four analysed articles using COG classification.

GOAL 3: Ease. Choose data visualizations that are easy to perceive and interpret.

Do Not Trust the Defaults
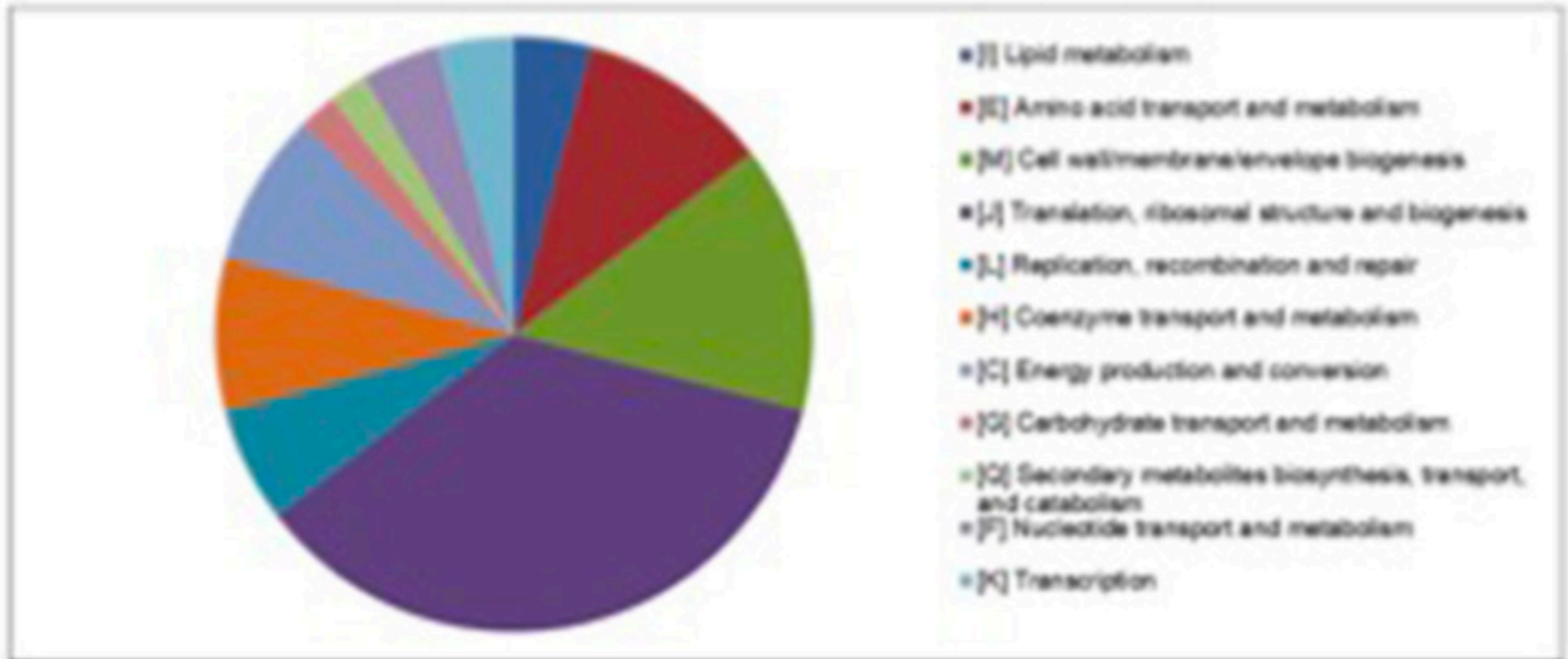
GOAL 4: Efficiency. Present only that which needs presenting. Minimize "chartjunk"

The chart area border: Does it aid information gathering ?

The values: Can we remove them if they are self evident ?

The Gridlines: Can we do without them ?

Background color : Is it necessary ?

Tick marks: Are they needed ?

Legend: Does it present additional information ?

Rainfall (in cm)

Jan, Feb, Mar, Apr, May, Jun
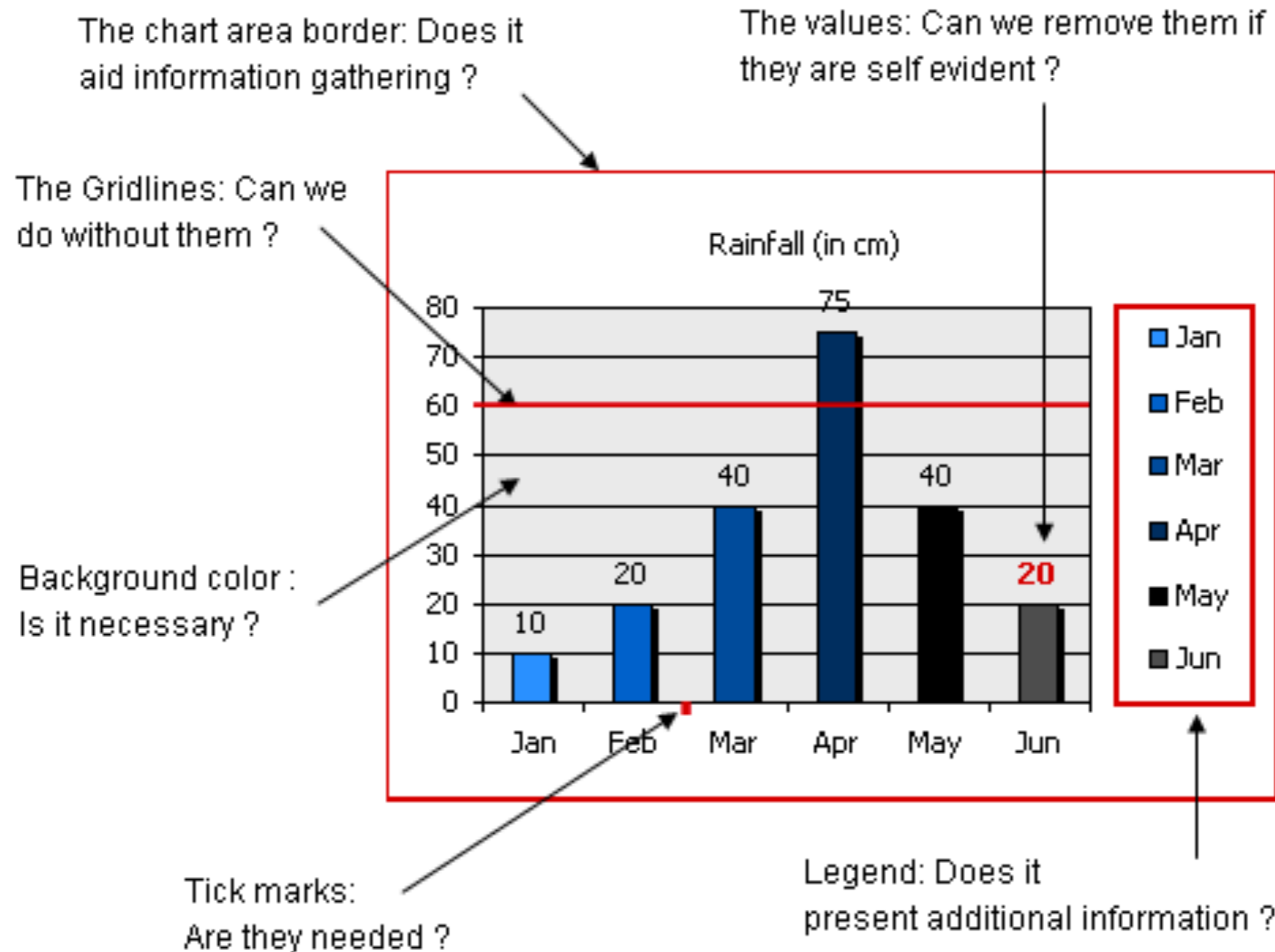
GOAL 4: Efficiency. Adapt figures to the medium (presentation vs. publication)
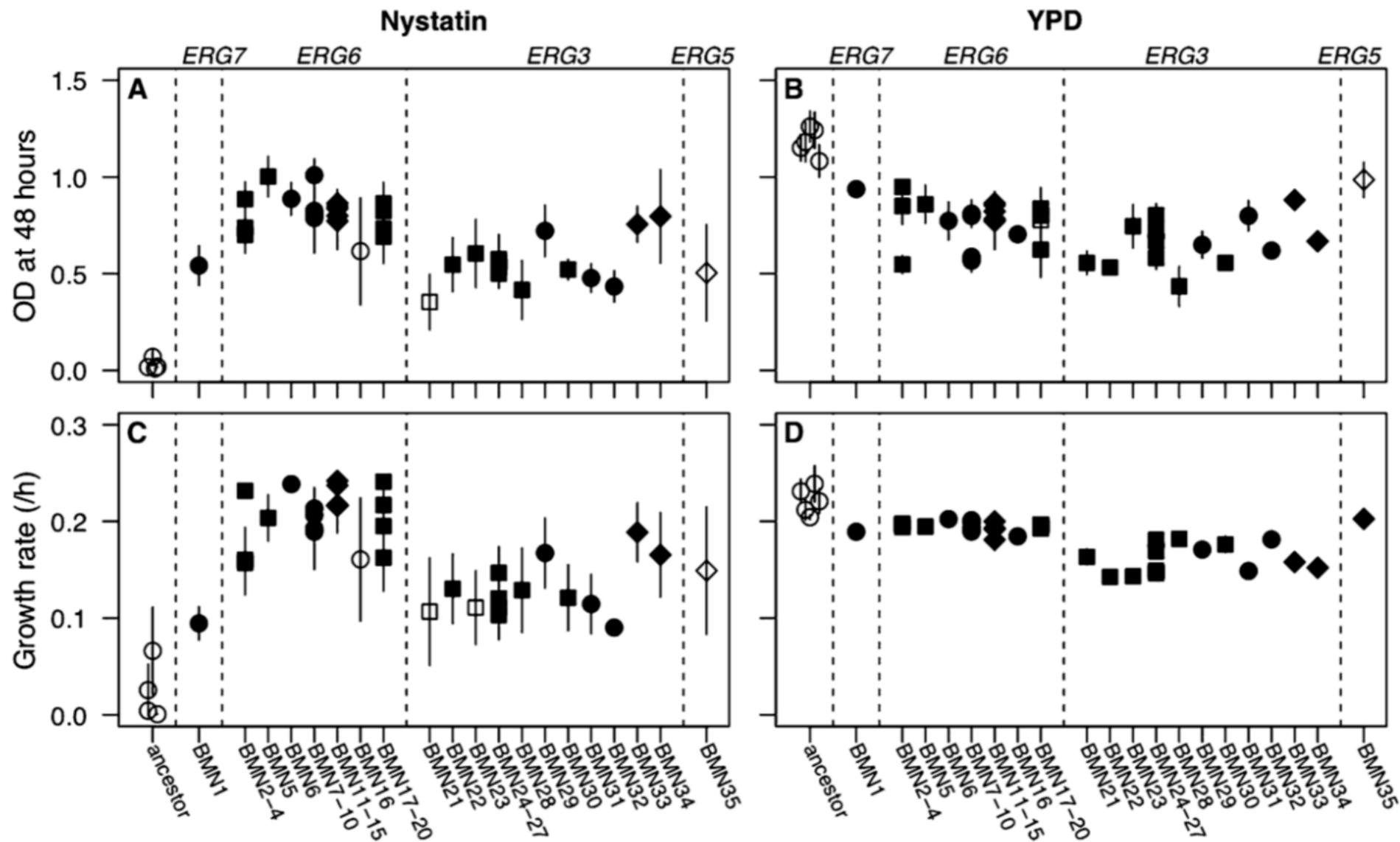
Figure 4 Two different fitness assays, OD at 48 hr (A, in nystatin; B, in YPD, an unstressful environment) and maximum growth rate (C, in nystatin; D, in YPD), show that BMN lines have increased growth in YPD + 4 μM nystatin and reduced growth in YPD relative to the ancestor. Lines that are significantly different from the ancestral colonies are plotted with solid symbols (*t*-test results presented in Table S4, Table S5, Table S6, and Table S7). Error bars depict the standard error of four bioscreen well replicates for each line. Mutation lines are arranged on the *x*-axis as in Figure 3. For the ancestral lines, five independent colonies were isolated and used to establish the lines, each of which was assayed four times (bars again show SE based on the replicates for each of these five lines). Mutations in the same gene are grouped together, and genes that are closer to producing ergosterol (the end product of the common pathway) are farther to the right. When multiple mutations are present in the same gene, mutations closer to the start codon are numbered lower.

GOAL 4: Efficiency. Adapt figures to the medium (presentation vs. publication)
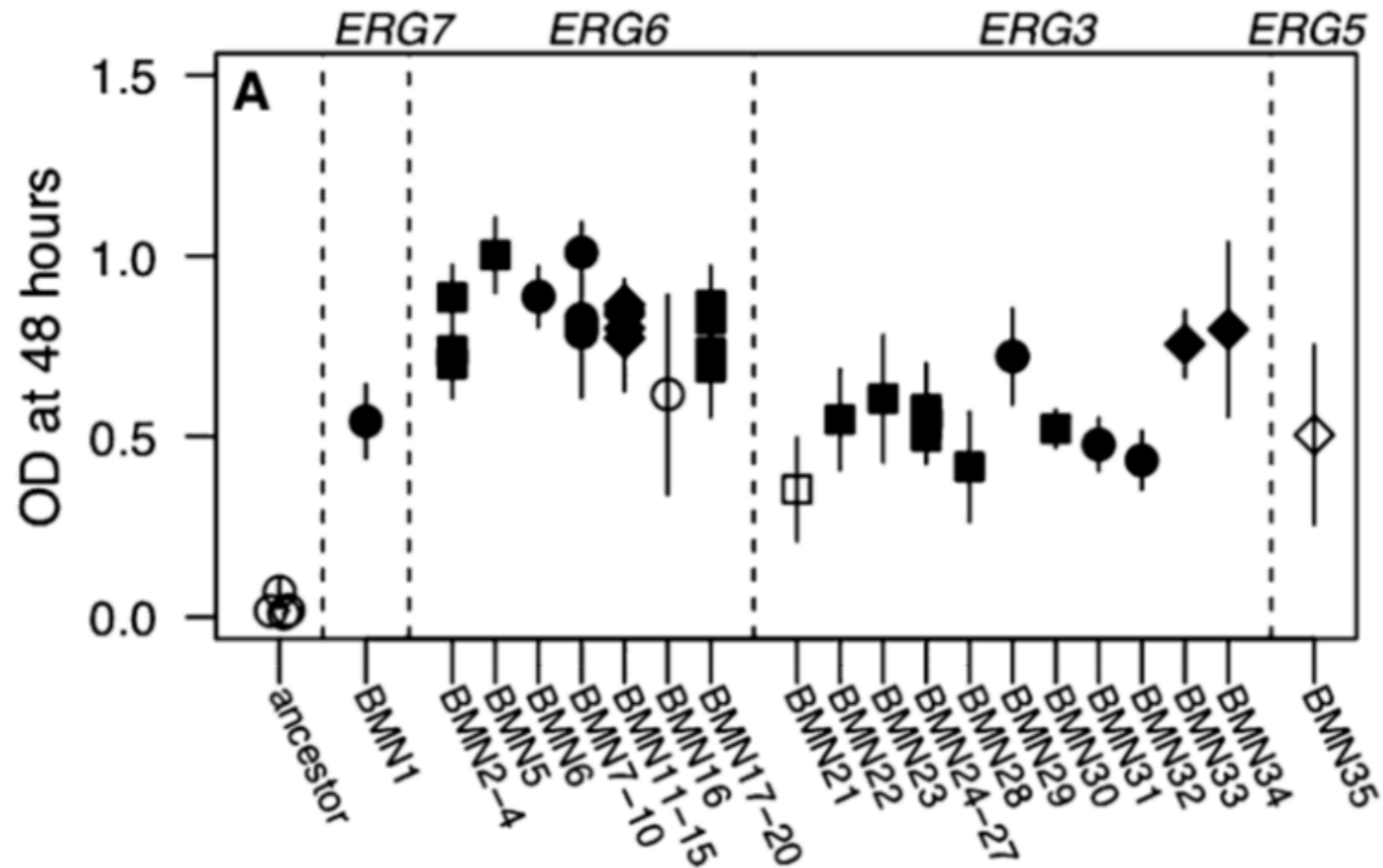
Activity 1:

For each of the following three figures, discuss in your lab group whether they met the four major goals. Can you suggest any improvements the authors could have made to improve these figures?

GOAL 1: Clarity

GOAL 2: Precision

GOAL 3: Ease

GOAL 4: Efficiency

"We isolated 16 ad_P10 bacteriophages [bacteriophage P10 variants adapted to *E. coli* strain MG1655] and tested them on a collection of 105 Enterobacteriaceae strains (Figure 2 and Table S1). Almost half of the ad_P10 isolates had a reduced host range than the wild-type strain (up to 50% reduction), and most others showed marginal difference; one exceptional ad_P10 variant had expanded its host range to 11 more strains than the wild-type P10 (Figure 2)."
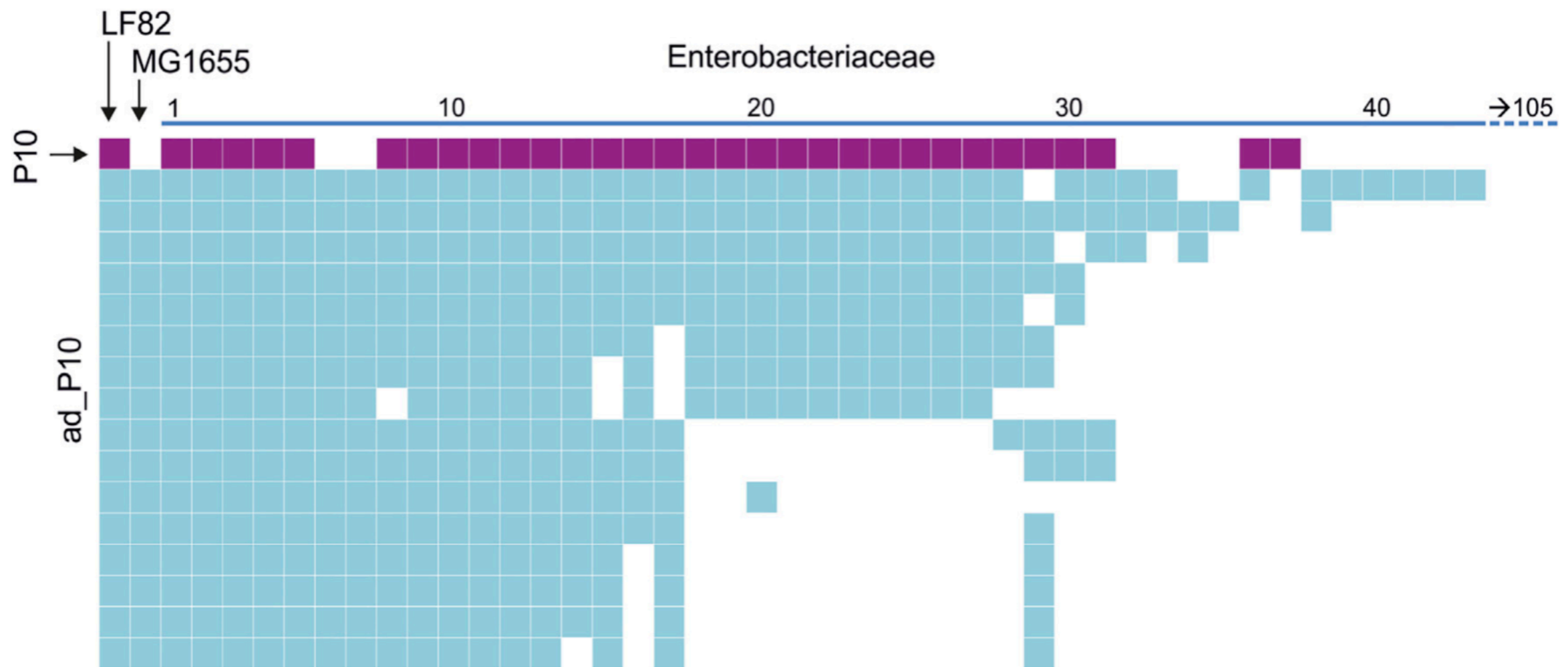
Figure 2. Broad Host Range Variations Are Associated with Bacteriophage Adaptation

Ad_P10 (n = 16) were tested for infectivity (light blue) versus 105 Enterobacteriaceae (see Table S1) and compared with the ancestral P10 bacteriophage (purple). The shared acquired infectivity toward strain MG1655 has a very variable effect on the global host range. Only positive results are shown for clarity.

"What is the host range of individual virus types? Twenty-four viruses infected only a single cellular clade. In contrast, 23 virus types were detected in >2 host genera within the Sulfolobaceae family. Every previously characterized virus detected was found in at least one new host species. For example, STIV previously shown to infect S. solfataricus [42], was also detected in Acidocryptum cells."
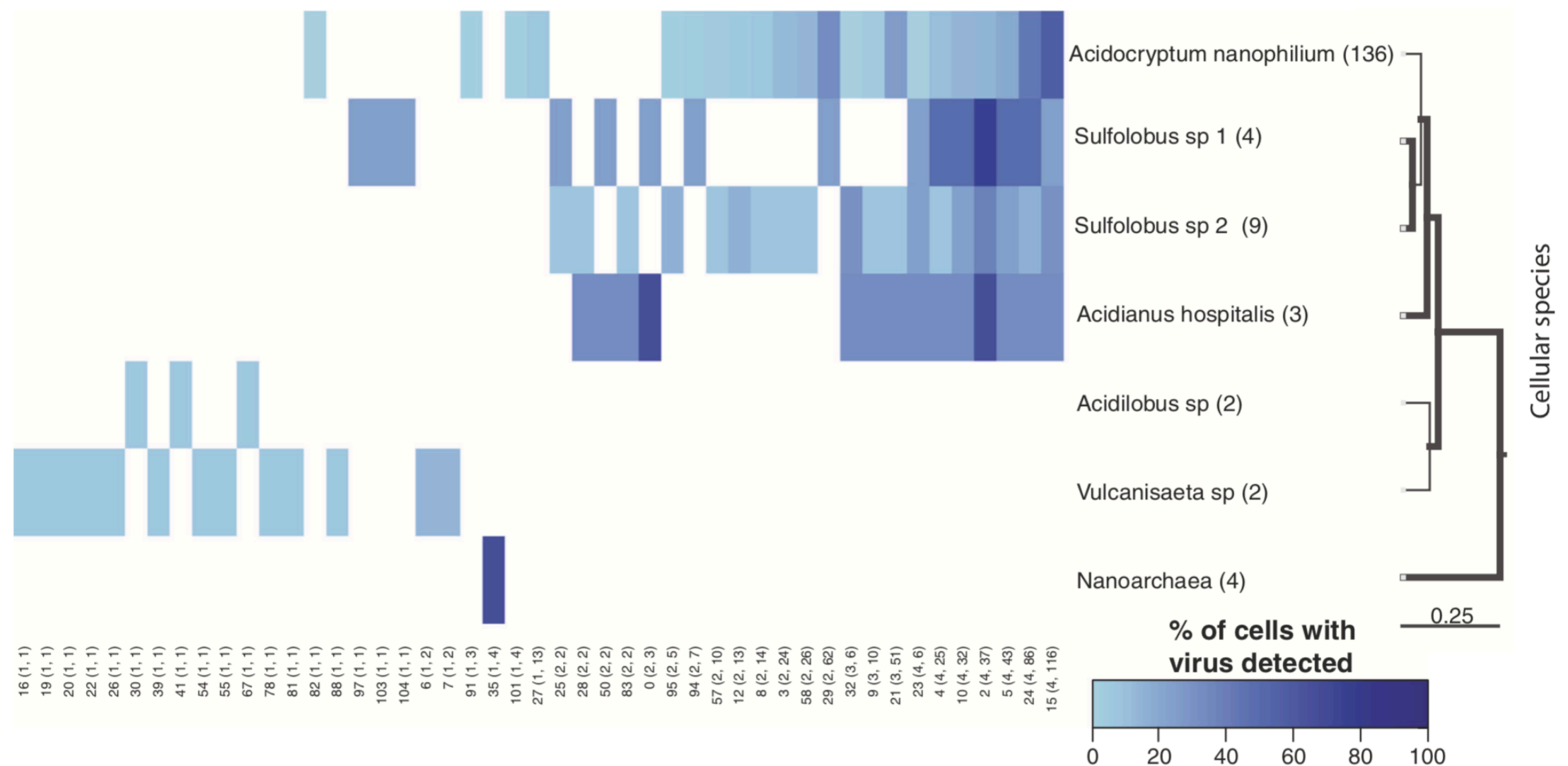
Fig. 3 Ubiquitous interaction of multiple viruses with cells. The heatmap indicates the detection frequency of 47 viral groups detected by BLASTn analysis or the matching of CRISPR spacer sequences. Viral groups are arranged from least frequently detected to the most frequently detected. Numbers below the heatmap are viral group numbers taken from [16] and numbers in parenthesis indicate the number of species and cells that a group was detected in. The number after the species name on the right hand side is the number of cells classified as members of that species. Partial length 16S sequences from representative genomes were used to make a ML tree and nodes with greater than 0.95 posterior probability are bolded. The scale bar is in substitutions per base. Detected viral groups with described members are: group 0 = SIRV1,2, group 23 = ASV1, SSV1,2, 4–9, group 26 = ATV, group 28 = AFV1, group 29 = STIV1,2 and group 32 = STST1,2 and ARSV1

"All of the species examined grew between 5 and 25 °C, with optimal growth varying between 10–25 °C. Bat isolate BL308 was more restricted in its growth, showing an optimum growth of 10°C with no growth at 25°C, similar to *P. destructans*. Strains that were closely related to each other, such as BL578 and *P. pannorum* ATCC 16222, showed similar responses to temperature, but there was otherwise no immediately evident correlation between physiology and strain relationships."

Figure 2. Temperature-dependent growth of environmental isolates of *Pseudogymnoascus*. The plot shows the average growth rate of each taxon over 6 weeks. Error bars indicate the significant difference of each assay (n = 10). *P. pan* = *P. pannorum*. doi:10.1371/journal.pone.0086437.g002

Tables

Table 2 Oligonucleotide primers used in this study

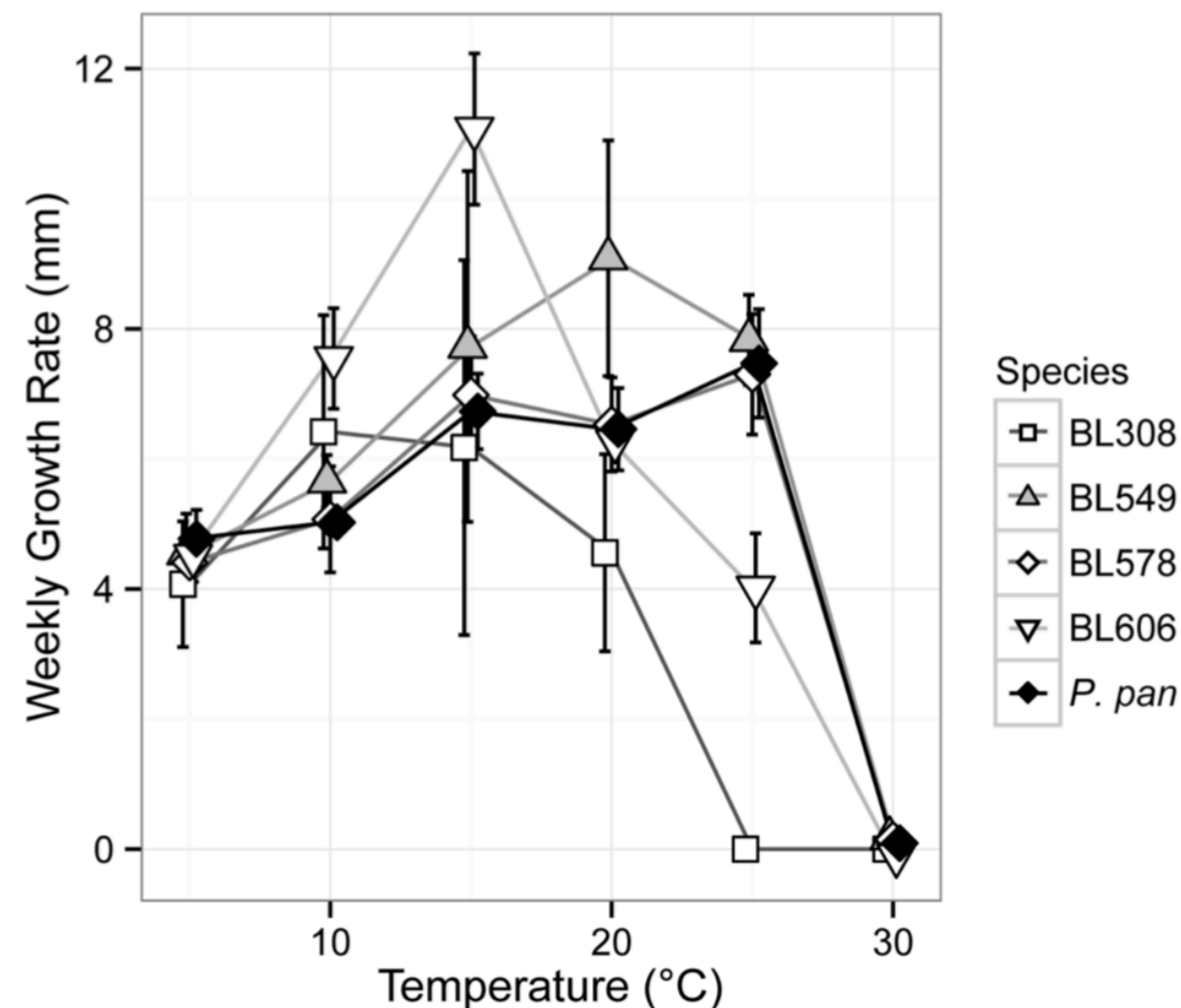| Name	Sequence
hht-FW	TCGGCCAAGATTTCAAGACTG
hht-RV	CGCCACGGAGACGACGAG
hhf-FW	ATTCGCGATGCCGTCACCTA
hhf-RV	TAACCACCGAAACCATAAAT
hta-FW	CTTCGCCGCCGTTTTGGAATA
hta-RV	CGTTACGGATGGCGAGTTGAAGAT
htb1-FW	AGCCATGCGTATCTTGAACTCTTT
htb1-RV	GGTAACGGCGTGCTTGGCTAACT
hht1-FW	CTGTCACCCTTTGATATGTTG
hht1-RV	TAACACATATCCGTTCCCATC
hht2-FW	GCATTGATTGCCTAATATTTTATTTG
hht2-RV	AAATTAATATGCTAAACCCGAC
cdc18-FW	TCCCTCGTTTACGAACAAAG
cdc18-RV	CAGCATGCTGAGATACAAC
cdc22-FW	CAGGTAGAGGGTACATATG
cdc22-RV	TGAGATGTTGAAGCAGTAGGC
mik1-FW	GGGATTATTGCAGGTCATGG
mik1-RV	ATCAACCATCGAGGAGACCGG
ace2-FW	GAATTCCTCCGGAGACAATG
ace2-RV	AAGTCACAGCGATACGGACG
ams2-FW (antisense)	GAGCCTTTATCCGAAATTGGG
ams2-RV (sense)	GCAAACAGGCAGAGTTGGC
act1ChIP5	CCACTATGTATCCCGGTATTGC
act1ChIP6	CAATCTTGACCTTCATGGAGCT

TABLE 2 Distribution of protein and ATPase in fractions of dialyzed membranes[a]

Membrane	Fraction	ATPase	
		U/mg of protein	Total U
Control	Depleted membrane	0.036	2.3
	Concentrated supernatant	0.134	4.82
E1 treated	Depleted membrane	0.034	1.98
	Concentrated supernatant	0.11	4.6

[a] Specific activities of ATPase of nondepleted membranes from control and treated bacteria were 0.21 and 0.20, respectively.

Table 3. General essential genes reported *P. aeruginosa* PAO1 (19) that we found to be missing in *P. aeruginosa* PA14

Query data sequences						Best hit data		
Gene annotation	Function	COG	Query sequence length (aa)	E-value	Best hit length sequence (aa)	Best hit gene anotation	Sequence proportion of alignment: Query part/ total hit sequence
imm2	pyocin S2 immunity protein/bacteriocin immunity	Adaptation/protection/defense mechanisms	87	1.0	2352	PA14_21020	25/2352
wbpB	UDP-2-acetamido-2-deoxy-ᴅ-glucuronic acid 3-dehydrogenase/O antigen biosynthetic process	Putative enzymes; cell wall/ LPS/capsule	316	0.009	347	PA14_02930	200/347

Tables

Tables should be formatted as follows. Arrange the data so that columns of like material read down, not across. The headings should be sufficiently clear so that the meaning of the data is understandable without reference to the text. Explanatory footnotes are acceptable, but more-extensive table "legends" are not. Footnotes should not include detailed descriptions of the experiment. Tables must include enough information to warrant table format; those with fewer than six pieces of data will be incorporated into the text by the copy editor. Table 2 is an example of a well-constructed table.

TABLE 2 Distribution of protein and ATPase in fractions of dialyzed membranes[a]

Membrane	Fraction	ATPase	
		U/mg of protein	Total U
Control	Depleted membrane	0.036	2.3
	Concentrated supernatant	0.134	4.82
E1 treated	Depleted membrane	0.034	1.98
	Concentrated supernatant	0.11	4.6

[a] Specific activities of ATPase of nondepleted membranes from control and treated bacteria were 0.21 and 0.20, respectively.

Images & Illustrations
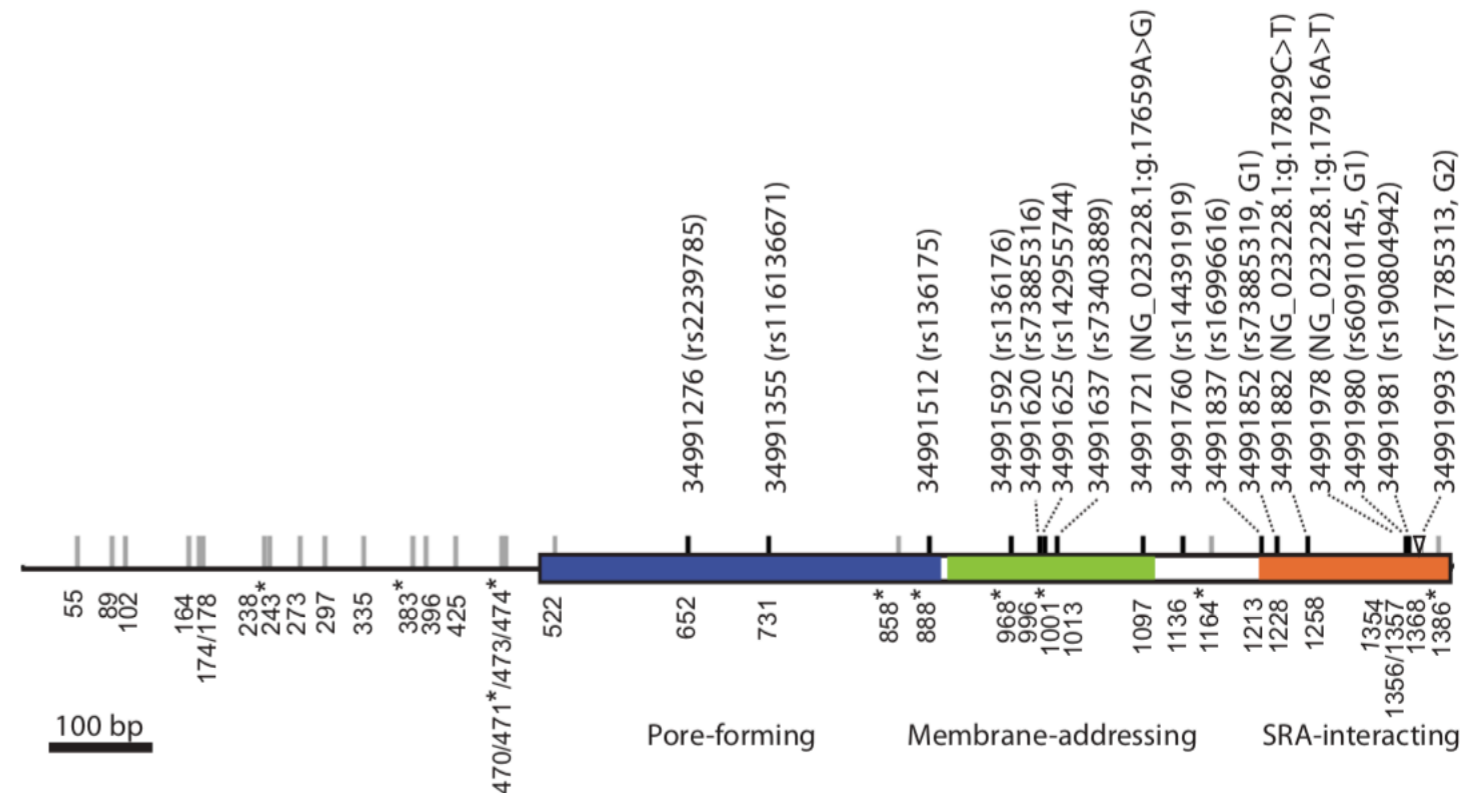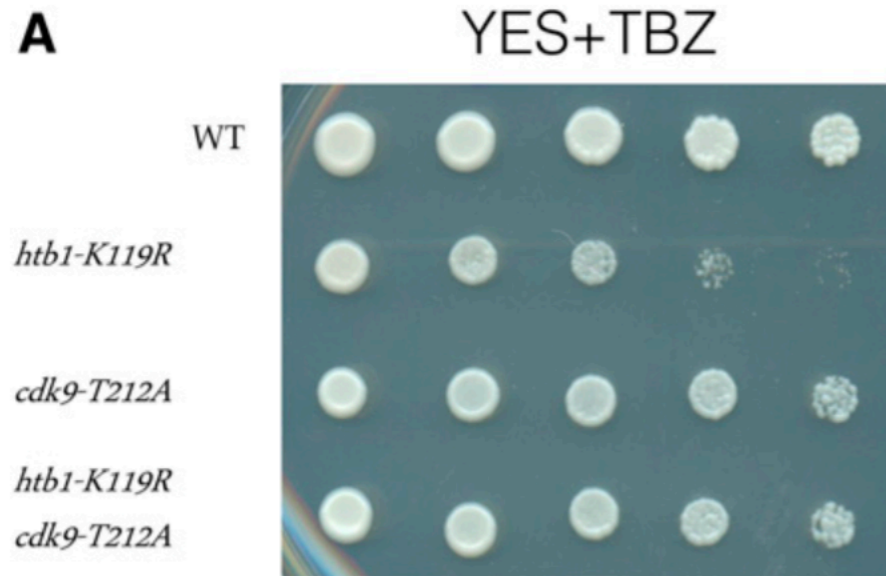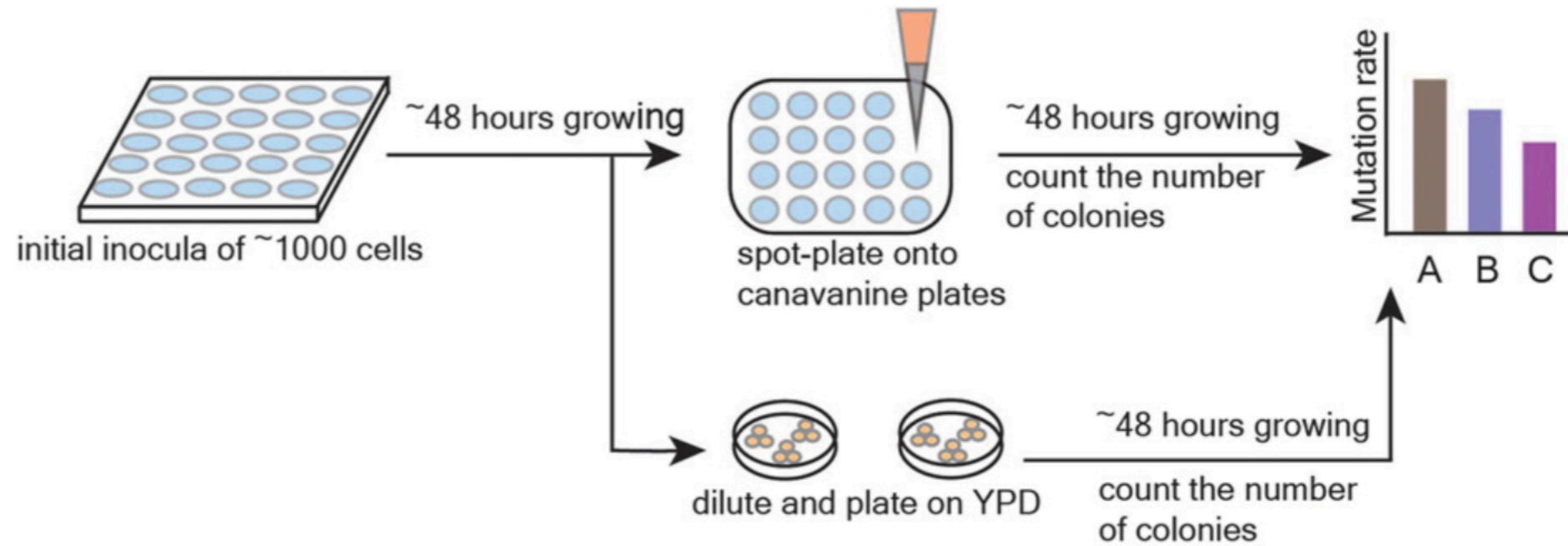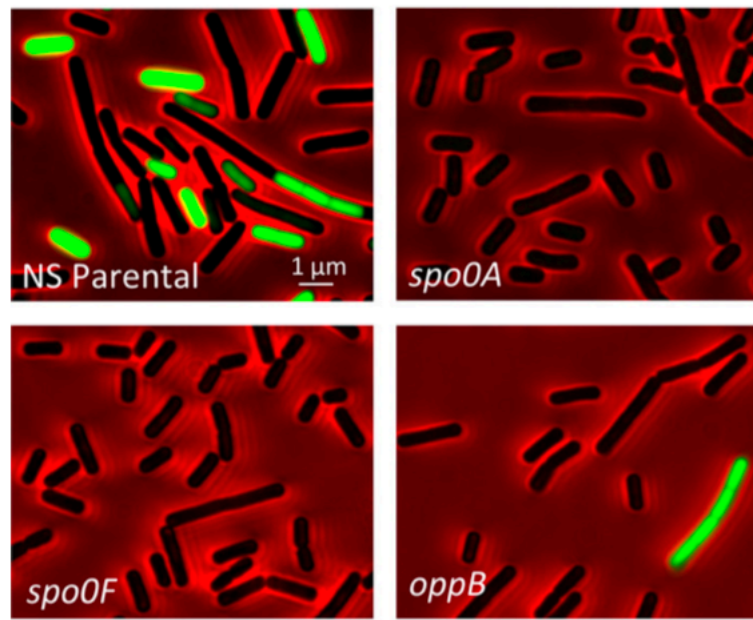
Image manipulation

Linear adjustments to contrast, brightness, and/or color are generally acceptable, as long as the measures taken are necessary to view elements that are already present in the data and the adjustments are applied to the entire image and not just specific areas. Unacceptable adjustments to images include, but are not limited to, the removal or deletion, concealment, duplication (copying and pasting), addition, selective enhancement, or repo- sitioning of elements within the image.

Nonlinear adjustments made to images, such as changes to gamma settings, should be fully disclosed in the figure legends at the time of submission. In addition, images created by compiling multiple files, including noncontiguous portions of the same image, should clearly convey that these multiple files are not a single image. This can be done by "tooling," or inserting thin lines, between the individual images.

Figure and Table Legends

The "Acid Test": Any Table or Figure you present must be sufficiently clear, well-labeled, and described by its legend to be understood by your intended audience without reading the methods or results section, i.e., it must be able to stand alone and be interpretable.

Overly complicated Figures or Tables may be difficult to understand in or out of context, so strive for simplicity whenever possible.

http://abacus.bates.edu/~ganderso/biology/resources/writing/HTWtablefigs.html

Figure Legend Journal Instructions

Canadian Journal of Microbiology

Experimental conditions should not be included

For graphs, captions should not repeat axis labels, but should describe what the data show.

A single caption can be provided for multipart (composite) figures, with necessary details on the separate parts, identified by their individual labels.

If the separate parts require enough information to warrant separate captions, then the composite should be separated into individual figures.

Only essential labelling should be used, with detailed information given in the caption.

mBIO

However, detailed experimental methods must be described in the Materials and Methods section, not in a figure legend.

A method that is unique to one of several experiments may be reported in a legend only if the discussion is very brief (one or two sentences).

Define all symbols used in the figure and define all abbreviations that are not used in the text.

Table Title Journal Instructions

Canadian Journal of Microbiology

The title of the table should be a concise description of the content, no longer than one sentence, that allows the table to be understood without detailed reference to the text.

Column headings should be brief, but may be amplified by footnotes.

Vertical rules should not be used.

Footnotes in tables should be designated by symbols (in the order *, †, ‡, §, ||, ¶, #) or superscript lowercase italic letters.

Descriptive material not designated by a footnote may be placed under a table as a **Note**.

mBIO

Explanatory footnotes are acceptable, but more-extensive table "legends" are not.

Footnotes should not include detailed descriptions of the experiment.

Activity 2: Prepare a figure sketch with the (fake) qRT-PCR data below.

Include all required figure elements including a figure legend.
Tell me in bullet form the choices you made to achieve the four figure goals.
Is there anything missing from the data table that would assist you?

	A	B	C	D	E	F	G	H	I
1		Gene		37 degrees		average			
2		REF	23	22	23	22.7			
3		YFG1	23	23	22	22.7			
4		YFG2	22	21	22	21.7			
5									
6		Gene		30 degrees		average			
7		REF	22	22	23	22.3			
8		YFG1	18	20	19	19			
9		YFG2	17	22	21	20			
10									
11		Gene		42 degrees		average			
12		REF	22	23	22	22.3			
13		YFG1	30	28	26	28			
14		YFG2	27	24	25	25.3			
15									
16	Gene	ΔCT CONT37	ΔCT EXP30	ΔCT EXP42	ΔΔCT (30)	ΔΔCT (42)	Fold Change (30)	Fold Change (42)	Fold Decrease (42)
17	YFG1	0	-3.33	5.67	-3.33	5.67	10.06	0.02	50
18	YFG2	-1	-2.33	3	-1.33	4	2.51	0.06	16.67
19									

= F4-F2

= F14-F12

= D18-B18

$= 2^{-F18}$

= 1/H18