

Figure_3_B

February 11, 2024

Figure 3 B i,ii,iii,iv We need pandas for reading the excel file and process the information it contains.

We need seaborn and matplotlib for plotting purposes, we choose color palette `n_colors=1` to make the figures exactly same color as the paper

```
[1]: import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
pal = sns.color_palette(n_colors=1)
```

We selected `sheet_name="Fig3 panel b"` and separate them into 4 different data frames to separately plot them.

```
[2]: table_3 = pd.read_excel("Table_3.xlsx", sheet_name="Fig3 panel b")
```

```
[3]: table_3_i = table_3[["TOP 50 ORDERS", "Unnamed: 1"]]
table_3_ii = table_3[["Streptomycetales FAMILIES", "Unnamed: 3"]]
table_3_iii = table_3[["Streptomycetaceae GENERA", "Unnamed: 5"]]
table_3_iv = table_3[["TOP 50 Streptomyces SPECIES", "Unnamed: 7"]]
```

For each figure (for Figure 3 B sub-figure) the order is same; * Setting the title. * Setting the column names. * Reformatting for correct information which is getting rid of redundant information and clear order. * For plotting purposes the data reshaped into the respective scale to configure the figures on the paper.

```
[4]: figure_3_i_title = "Top 50 orders"
table_3_i.columns = table_3_i.iloc[0]
table_3_i = table_3_i.drop(0)
table_3_i = table_3_i.reset_index(drop=True)
table_3_i["Number of GCFs"] = table_3_i["Number of GCFs"]/1000
```

For each figure (for Figure 3 B sub-figure) the order is same for plotting steps; 1. Setting the figure and the axis, the size and dpi(dots per inch) are hyperparameters that should tune for different configurations. 2. Setting the title for the figure. 3. Use Seaborn visualization library for plotting barplot with parameters set by us. 4. Setting `xlabel` & `ylabel` as `None` for matching the figures in paper(reproducibility). 5. Adding the text in correct location and format. 6. Saving the figure.

```

[5]: f, ax = plt.subplots(figsize=(4, 12), dpi=200)

# ax.set_title(figure_3_i_title)

ax = sns.barplot(x="Number of GCFs", y="Taxon", data=table_3_i, palette=pal)

ax.set_xticks([0,2.5,5,7.5])

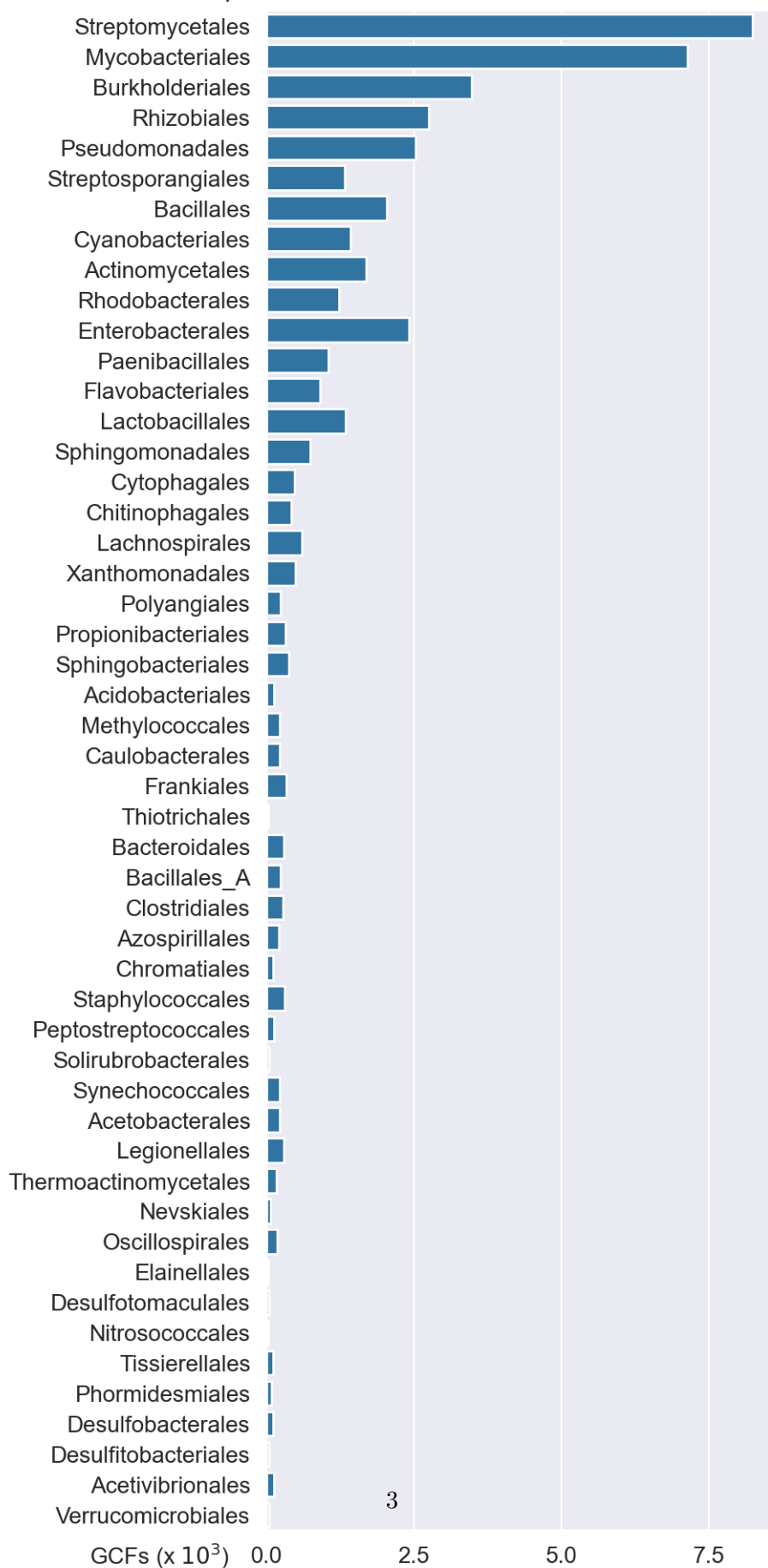
ax.set_xlabel(None)
ax.set_ylabel(None)

ax.text(-3, 50.61, "GCFs (x  $10^3$ )")
ax.text(-1.2, -1, figure_3_i_title, fontsize="large", fontstretch=1000)

ax.figure.savefig('Figures_3_Original_Format/output_figure_3_B_i_Topic_10.png')

```

Top 50 orders



```
[6]: figure_3_ii_title = "Streptomycetales families"
table_3_ii.columns = table_3_ii.iloc[0]
table_3_ii = table_3_ii.drop(0)
table_3_ii = table_3_ii.reset_index(drop=True)
table_3_ii = table_3_ii.dropna()
table_3_ii["Number of GCFs"] = table_3_ii["Number of GCFs"]/1000
```

```
[7]: f, ax = plt.subplots(figsize=(12, 4), dpi=400)

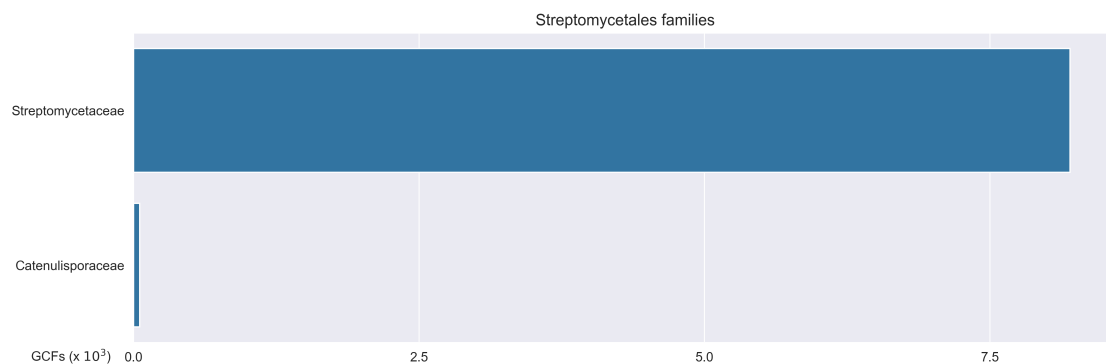
ax.set_title(figure_3_ii_title)

ax = sns.barplot(x="Number of GCFs", y="Taxon", data=table_3_ii, palette=pal)

ax.set_xticks([0, 2.5, 5, 7.5])

ax.set_xlabel(None)
ax.set_ylabel(None)

ax.text(-0.9, 1.62, "GCFs (x  $10^3$ )")
f.tight_layout()
ax.figure.savefig('Figures_3_Original_Format/output_figure_3_B_ii_Topic_10.png')
```



```
[8]: figure_3_iii_title = "Streptomycetaceae genera"
table_3_iii.columns = table_3_iii.iloc[0]
table_3_iii = table_3_iii.drop(0)
table_3_iii = table_3_iii.reset_index(drop=True)
table_3_iii = table_3_iii.dropna()
table_3_iii["Number of GCFs"] = table_3_iii["Number of GCFs"]/1000
```

```
[9]: f, ax = plt.subplots(figsize=(12, 4), dpi=400)
```

```

ax = sns.barplot(x="Number of GCFs", y="Taxon", data=table_3_iii, palette=pal)

ax.set_title(figure_3_iii_title)

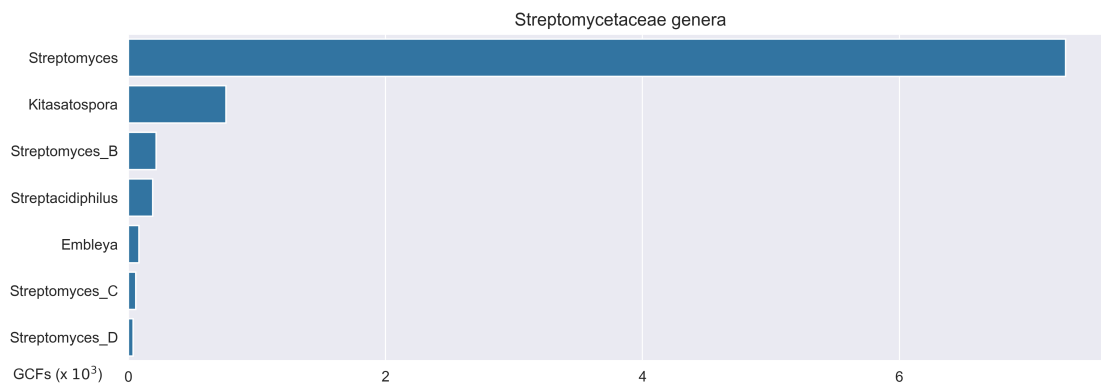
ax.set_xlabel(None)
ax.set_ylabel(None)

ax.set_xticks([0,2,4,6])

ax.text(-0.9, 6.9, "GCFs (x  $10^3$ )")

ax.figure.savefig('Figures_3_Original_Format/output_figure_3_B_iii_Topic_10.
↳png')

```



```

[10]: figure_3_iv_title = "Top 50 Streptomyces species"
table_3_iv.columns = table_3_iv.iloc[0]
table_3_iv = table_3_iv.drop(0)
table_3_iv = table_3_iv.reset_index(drop=True)
table_3_iv = table_3_iv.dropna()
table_3_iv["Number of GCFs"] = table_3_iv["Number of GCFs"]

```

```

[11]: f, ax = plt.subplots(figsize=(4, 12), dpi=400)

ax = sns.barplot(x="Number of GCFs", y="Taxon", data=table_3_iv, palette=pal)

# ax.set_title(figure_3_iv_title)

ax.set_xticks([0,50,100,150])

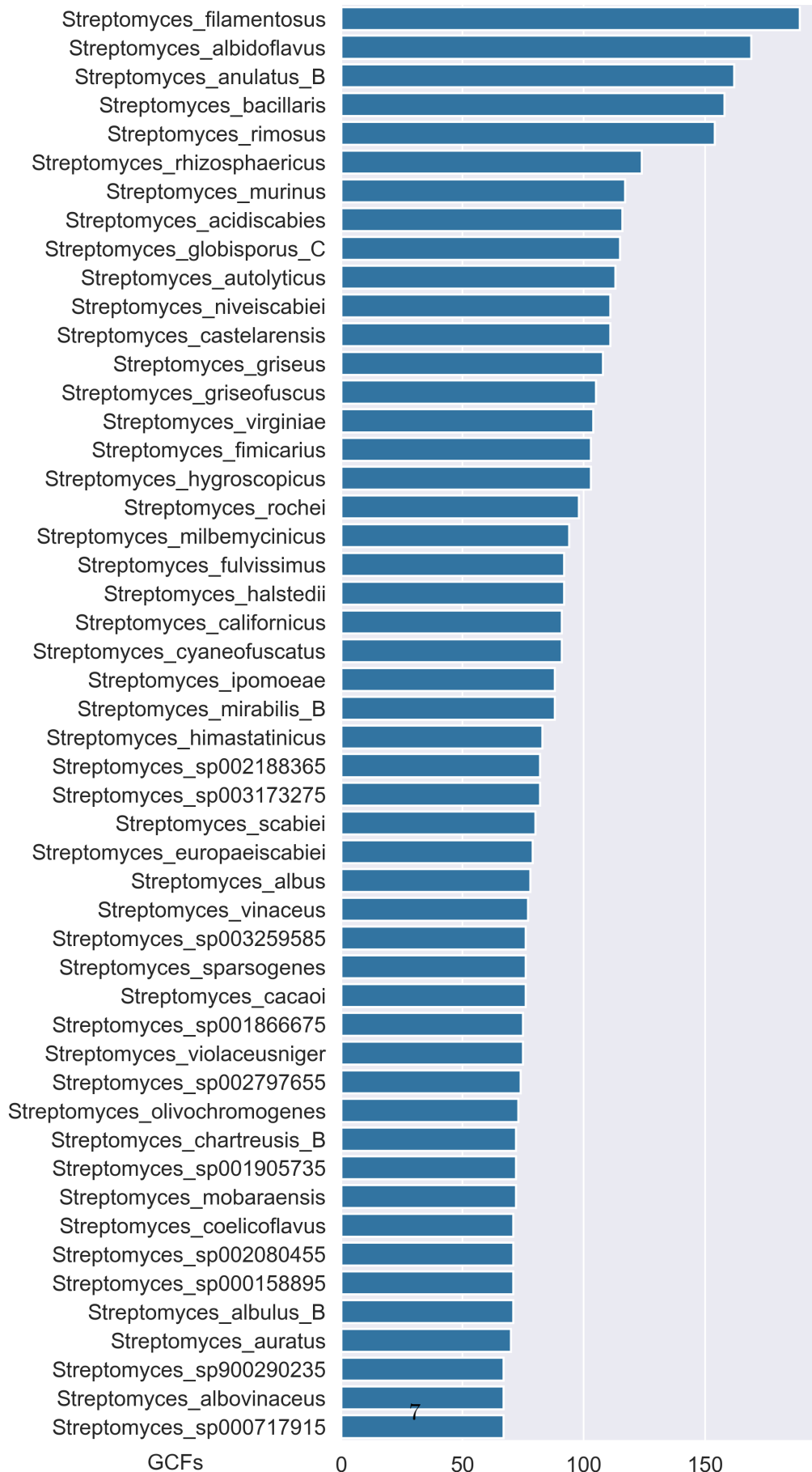
ax.text(-80, 50.5, "GCFs")

ax.set_xlabel(None)
ax.set_ylabel(None)

```

```
ax.text(-42, -1, figure_3_iv_title, fontsize="large", fontstretch=1000)  
ax.figure.savefig('Figures_3_Original_Format/output_figure_3_B_iv_Topic_10.png')
```

Top 50 Streptomyces species



END OF REPRODUCING FIGURE 3 B i, ii, iii, iv .