Figure_3_B

March 7, 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 hex code `#9fc5e8` to make the figures exactly same color as the paper.

```
[1]: import pandas as pd
     import seaborn as sns
     import matplotlib.pyplot as plt
     sns.set_style("whitegrid")
```

We selected `sheet_name="Fig3 panel b"` and seperate them info 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=400)
```

1

```python
# ax.set_title(figure_3_i_title)

ax = sns.barplot(x="Number of GCFs", y="Taxon", data=table_3_i, color="#9fc5e8")

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

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

ax.text(-3.3, 50.6, "GCFs (x $10^3$)")
ax.text(-1, -2.1, figure_3_i_title, fontsize="large")

sns.despine(fig=None, ax=None, top=True, right=True, left=False, bottom=True,
 ↪offset=None, trim=False);

# ax.figure.savefig('Figures_3_Original_Format/output_figure_3_B_i_Topic_10.
 ↪png')
```
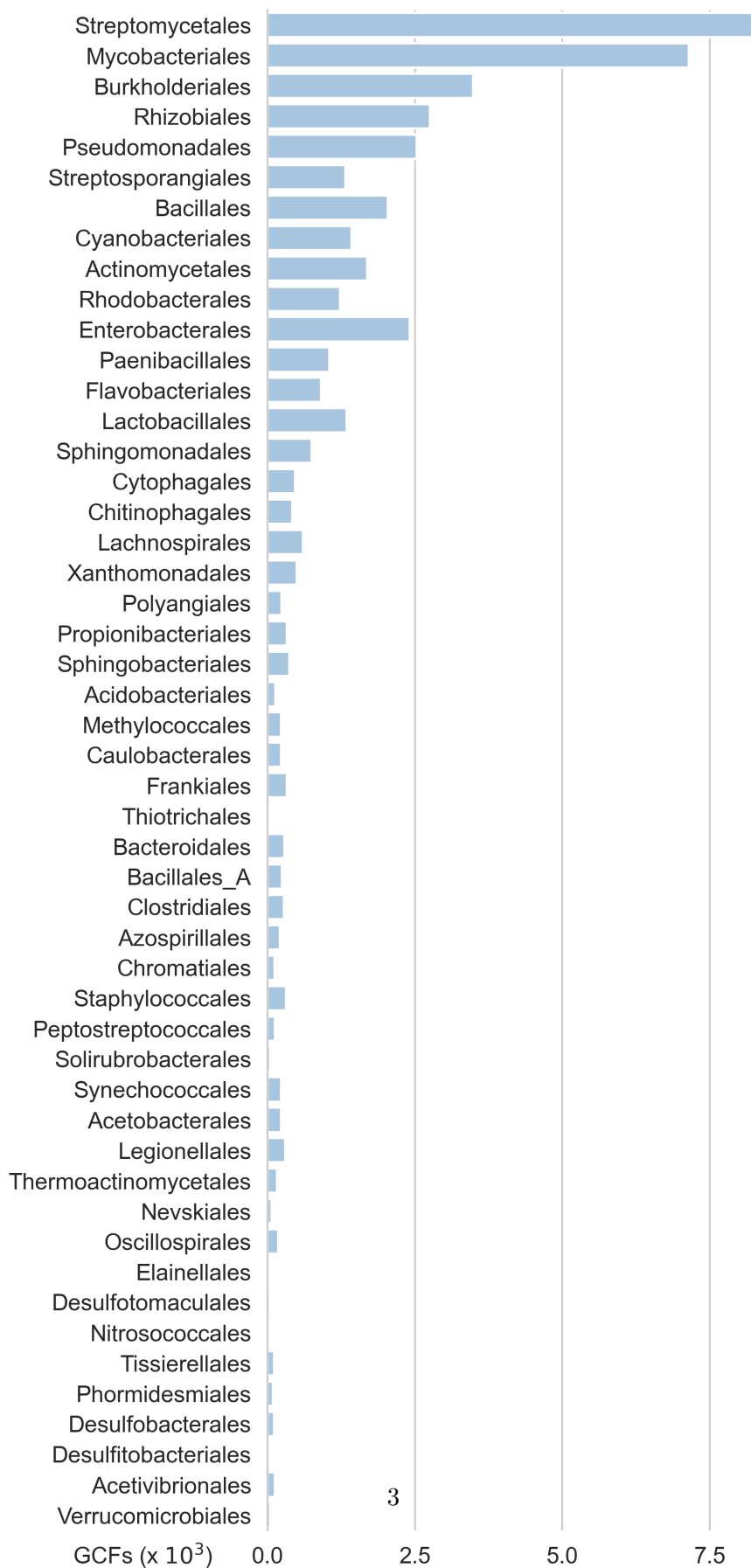
Top 50 orders

Order	GCFs (x 10^3)
Streptomycetales	
Mycobacteriales	
Burkholderiales	
Rhizobiales	
Pseudomonadales	
Streptosporangiales	
Bacillales	
Cyanobacteriales	
Actinomycetales	
Rhodobacterales	
Enterobacterales	
Paenibacillales	
Flavobacteriales	
Lactobacillales	
Sphingomonadales	
Cytophagales	
Chitinophagales	
Lachnospirales	
Xanthomonadales	
Polyangiales	
Propionibacteriales	
Sphingobacteriales	
Acidobacteriales	
Methylococcales	
Caulobacterales	
Frankiales	
Thiotrichales	
Bacteroidales	
Bacillales_A	
Clostridiales	
Azospirillales	
Chromatiales	
Staphylococcales	
Peptostreptococcales	
Solirubrobacterales	
Synechococcales	
Acetobacterales	
Legionellales	
Thermoactinomycetales	
Nevskiales	
Oscillospirales	
Elainellales	
Desulfotomaculales	
Nitrosococcales	
Tissierellales	
Phormidesmiales	
Desulfobacterales	
Desulfitobacteriales	
Acetivibrionales	
Verrucomicrobiales	

GCFs (x 10^3)   0.0   2.5   5.0   7.5

3

```
[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=(9, 4), dpi=400)

     ax.set_title(figure_3_ii_title)

     ax = sns.barplot(x="Number of GCFs", y="Taxon", data=table_3_ii,␣
      ↪color="#9fc5e8")

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

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

     sns.despine(fig=None, ax=None, top=True, right=True, left=False, bottom=True,␣
      ↪offset=None, trim=False)

     ax.text(-1.48, 1.62, "GCFs (x $10^3$)");

     # ax.figure.savefig('Figures_3_Original_Format/output_figure_3_B_ii_Topic_10.
      ↪png')
```
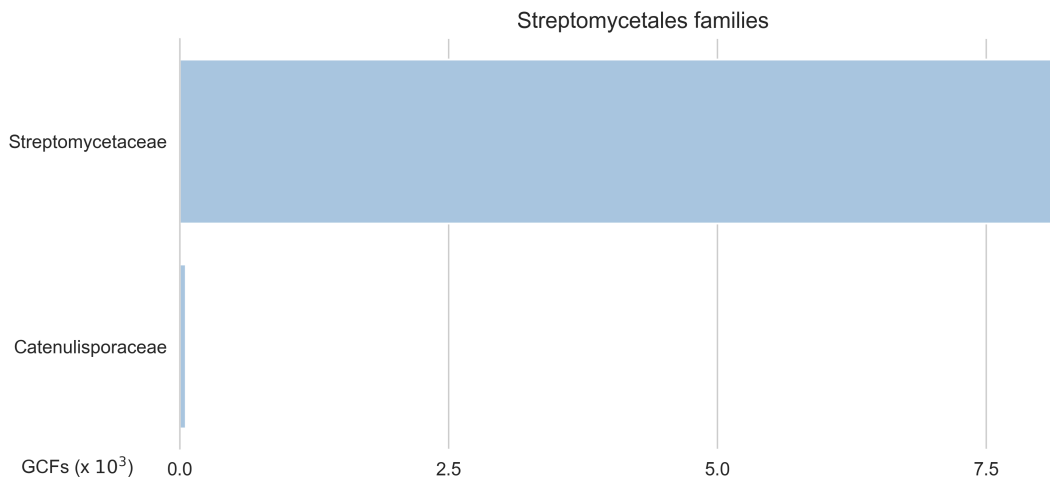
```
[8]: figure_3_iii_title = "Streptomycetaceae genera"
     table_3_iii.columns = table_3_iii.iloc[0]
     table_3_iii = table_3_iii.drop(0)
     table_3_iii = table_3_iii.reset_index(drop=True)
     table_3_iii = table_3_iii.dropna()
     table_3_iii["Number of GCFs"] = table_3_iii["Number of GCFs"]/1000
```

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

     ax = sns.barplot(x="Number of GCFs", y="Taxon", data=table_3_iii,␣
       ↪color="#9fc5e8")

     ax.set_title(figure_3_iii_title)

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

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

     sns.despine(fig=None, ax=None, top=True, right=True, left=False, bottom=True,␣
       ↪offset=None, trim=False)

     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"]
```

5

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

      ax = sns.barplot(x="Number of GCFs", y="Taxon", data=table_3_iv,␣
       ↪color="#9fc5e8")

      # 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)

      sns.despine(fig=None, ax=None, top=True, right=True, left=False, bottom=True,␣
       ↪offset=None, trim=False)

      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')
```
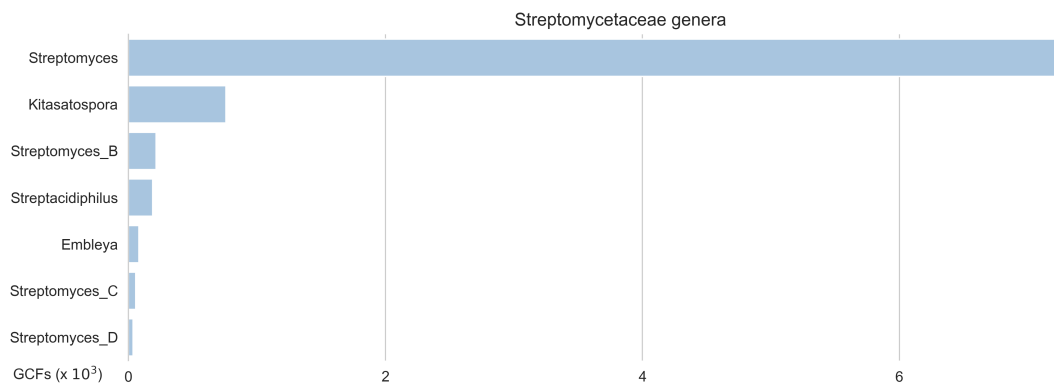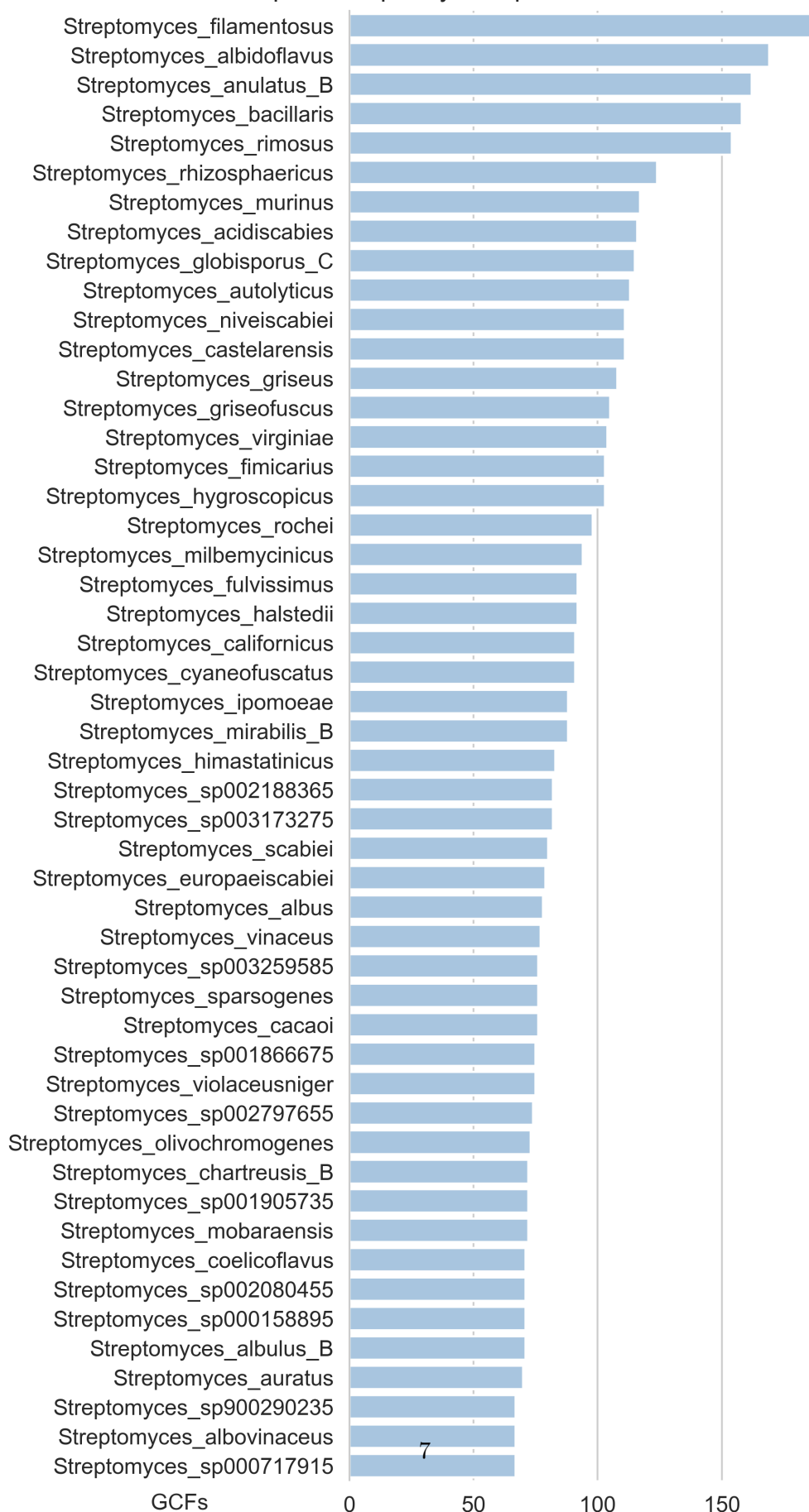
Top 50 Streptomyces species

Species	
Streptomyces_filamentosus	
Streptomyces_albidoflavus	
Streptomyces_anulatus_B	
Streptomyces_bacillaris	
Streptomyces_rimosus	
Streptomyces_rhizosphaericus	
Streptomyces_murinus	
Streptomyces_acidiscabies	
Streptomyces_globisporus_C	
Streptomyces_autolyticus	
Streptomyces_niveiscabiei	
Streptomyces_castelarensis	
Streptomyces_griseus	
Streptomyces_griseofuscus	
Streptomyces_virginiae	
Streptomyces_fimicarius	
Streptomyces_hygroscopicus	
Streptomyces_rochei	
Streptomyces_milbemycinicus	
Streptomyces_fulvissimus	
Streptomyces_halstedii	
Streptomyces_californicus	
Streptomyces_cyaneofuscatus	
Streptomyces_ipomoeae	
Streptomyces_mirabilis_B	
Streptomyces_himastatinicus	
Streptomyces_sp002188365	
Streptomyces_sp003173275	
Streptomyces_scabiei	
Streptomyces_europaeiscabiei	
Streptomyces_albus	
Streptomyces_vinaceus	
Streptomyces_sp003259585	
Streptomyces_sparsogenes	
Streptomyces_cacaoi	
Streptomyces_sp001866675	
Streptomyces_violaceusniger	
Streptomyces_sp002797655	
Streptomyces_olivochromogenes	
Streptomyces_chartreusis_B	
Streptomyces_sp001905735	
Streptomyces_mobaraensis	
Streptomyces_coelicoflavus	
Streptomyces_sp002080455	
Streptomyces_sp000158895	
Streptomyces_albulus_B	
Streptomyces_auratus	
Streptomyces_sp900290235	
Streptomyces_albovinaceus	
Streptomyces_sp000717915	

7

GCFs   0   50   100   150

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