**The command line and tip of BGDMdocker**

**Guide for workflow usage (running inside Container)**

**Running Prokka genome annotation in Container of BGDMdocker in Command line interaction patterns (if you have your own genome sequences, you need this step to generate** **“\*.gbff” annotation files):**

1. Check out help documentation and command parameters:

$ prokka –help

2. Copy the following commands to run the analysis for genome annotation of Ba\_xx strains from the command-line interface of Container (for boldface text, please enter your data):

$ prokka --kingdom Bacteria --gcode 11 --genus Bacillus \

--species **Amyloliquefaciens** \

--strain **Ba\_xx** --locustag **Ba\_xx** --prefix **Ba\_xx** --rfam \

--rawproduct --outdir **/home/manager/PRJNA291327** \

**/home/manager/Ba\_xx.fasta**

“Ba\_xx.fasta” is the genome sequence, “PRJNA291327” is the output folder of results.

**Running panX analysis on pan-genome in Container** **of BGDMdocker in Command line** **interaction patterns:**

PanX starts with a set of annotated sequences files, \*.gbff (.gbk) (e.g., NCBI RefSeq or GenBank), from a bacterial species genome.

1. Check out help documentation and command parameters of panX:

$ sh run.sh

2. Copy the following commands to run the analysis of the pan-genome of 44 *B. amyloliquefaciens* strains from the command-line interface of Container (for boldface text, please enter your data). If using your own GenBank files (or if you have downloaded these files), parameter “2” can be skipped when running “run-pipeline.py”. For detailed parameters see [here](https://github.com/neherlab/pan-genome-analysis).

$ cd /pan-genome-analysis

$ python ./scripts/run-pipeline.py -fn ./data/B\_amy -sl B\_amy-

RefSeq.txt -st 1 3 4 5 6 7 8 9 10 11

\*.gbff (GenBank files) and B\_amy-RefSeq.txt (accession list for strains) should reside in “./data/B\_amy” folder. The result will also be output to the “./data/B\_amy” folder.

3. Visualization of the pan-genome of 44 *B. amyloliquefaciens* strains (run in Container):

$ python link-to-server.py B\_amy

$ add-new-pages-repo.sh B\_amy

$ gulp

On you host ,open http://localhost:8000/**B\_amy** with a web browser to access the visualization of the pan-genome immediately.

4. Create a new Image for saving changes in Container data (running in host):

$ sudo docker commit <ID of Container > <name of new Image >

**Running** **antiSMASH to search for gene clusters in each strain in the Container** **of BGDMdocker in Command line interaction patterns:**

1. Check out help documentation and command parameters:

$ run\_antismash.py --help

2. Copy the following commands to run the analysis of biosynthetic gene clusters of Y2 strain from the command-line interface of Container (Y2.gbff) (for boldface text, please enter your data):

$ cd /antismash-3.0.5/antismash/

$ run\_antismash.py /home/manager/input/Y2.gbff \

--outputfolder /home/manager/output/Y2 \_out \

--dbgclusterblast ./generic\_modules/clusterblast \

--pfamdir ./generic\_modules/fullhmmer --input-type nucl --knownclusterblast \

--clusterblast --subclusterblast --inclusive --full-hmmer --smcogs --verbose --asf \

--borderpredict

\*.gbff (GenBank files) should reside in “input” folder. “Y2 \_out” is the output folder for results.

**Building workflow using standalone Dockerfile (recommendation):**

In order to meet the needs of different users, we also provide a standalone Dockerfile for Prokka, panX, and antiSMASH. You can build images and run Container separately.

$ git clone https://github.com/cgwyx/debian\_prokka\_panx\_antismash\_

docker.git

Or: [download](https://github.com/cgwyx/debian_prokka_panx_antismash_biodocker/archive/master.zip) “.zip” file

$ unzip debian\_prokka\_panx\_antismash\_biodocker-master.zip

**Building Image and** **run** **Container of Prokka standalone:**

$ cd ./debian\_prokka\_panx\_antismash\_biodocker/debian\_prokka\_standalone\_dockerfile

$ sudo docker build -t debian\_prokka: latest .

1. Run a Container from the image of debian\_prokka:

$ sudo docker run -it --rm -v home:home --name Prokka-latest prokka:latest

2. Copy the following commands to run the analysis of the genome annotation of Ba\_xx strains from the command-line interface of Container (for boldface text, please replace with your own data if applicable):

$ prokka --kingdom Bacteria --gcode 11 --genus Bacillus \

--species Amyloliquefaciens \

--strain Ba\_xx --locustag Ba\_xx --prefix Ba\_xx --rfam \

--rawproduct --outdir /home/manager/PRJNA291327 \

/home/manager/Ba\_xx.fasta

“Ba\_xx.fasta” is the sequence of the genome; “PRJNA291327” is the output folder of the results.

**Building Image and run Container of panX standalone:**

$ cd ./debian\_prokka\_panx\_antismash\_biodocker/debian\_panx\_standalone\_dockerfile

$ sudo docker build -t debian\_panx:latest .

1. Run a Container from the image of debian\_panx

$ sudo docker run -it --rm -v home:home --name debian\_panx debian\_panx:latest

2. Copy the following commands to run the analysis of the pan-genome of 44 *B. amyloliquefaciens* strains from the command-line interface of Container (for boldface text, please replace with your own data if applicable):

$ cd /pan-genome-analysis

$ python ./scripts/run-pipeline.py -fn ./data/B\_amy -sl B\_amy-RefSeq.txt -st 1 3 4 5 6 7 8 9 10 11

\*.gbff (GenBank files) and B\_amy-RefSeq.txt (accession list for strains) should be in the “./data/B\_amy” folder; output results will be also in “./data/B\_amy” folder.

3. Visualize the pan-genome of 44 *B. amyloliquefaciens* strains (run in Container):

$ python link-to-server.py B\_amy

$ add-new-pages-repo.sh B\_amy

$ gulp

Open [http://localhost:8000/**B\_amy**](http://localhost:8000/B_amy)with a web browser to access the visualization of the pan-genome immediately.

4. Create a new Image for saving changes in Container data (running in host):

$ sudo docker commit <ID of Container > <name of new images >

**Building Image and run Container of antiSMASH** **standalone:**

$ cd ./debian\_prokka\_panx\_antismash\_biodocker/debian\_antismash\_standalone\_dockerfile

$ sudo docker build -t debian\_antismas:latest .

1. Run a Container from the image of debian\_antismash

$ sudo docker run -it --rm -v home:home --name antiSAMSH debian\_antismash:latest

2. Copy the following commands to run the analysis of biosynthetic gene clusters of Y2 strain from the command-line interface of Container (Y2.gbff) (for boldface text, please replace with your own data if applicable):

$ cd /antismash-3.0.5/antismash/

$ run\_antismash.py /home/manager/input/Y2.gbff \

--outputfolder /home/manager/output/Y2 \_out \

--dbgclusterblast ./generic\_modules/clusterblast \

--pfamdir ./generic\_modules/fullhmmer --input-type nucl --knownclusterblast \

--clusterblast --subclusterblast --inclusive --full-hmmer --smcogs --verbose --asf \

--borderpredict

\*.gbff (GenBank files) reside in “input”folder; “Y2 \_out” is the output folder for the results.

**Tip: How can I download all “genomic.gbff.gz” of a specified species from the RefSeq or GenBank databases? Replace boldface text with your species if applicable:**

1. Installing script on your host

$ wget

ftp://ftp.ncbi.nlm.nih.gov/entrez/entrezdirect/versions/current/edirect.zip

$ unzip -u -q edirect.zip

$ export PATH=$PATH:$HOME/edirect

$ ./edirect/setup.sh

2. Download “\*genomic.gbff.gz” of all strains of *Bacillus amyloliquefaciens* from GenBank

$ esearch -db assembly -query "Bacillus amyloliquefaciens [ORGN]" | efetch -format docsum | xtract -pattern "DocumentSummary" -element FtpPath\_GenBank | sed 's/$/\/\*genomic.gbff.gz/' |xargs wget -c -nd;sleep 3s;

Or:

3. Download “\*genomic.gbff.gz” of all *Bacillus amyloliquefaciens* strains from RefSeq

$ esearch -db assembly -query "Bacillus amyloliquefaciens[ORGN]" | efetch -format docsum | xtract -pattern "DocumentSummary" -element FtpPath\_RefSeq | sed 's/$/\/\*genomic.gbff.gz/' |xargs wget -c -nd;sleep 3s;

**Visualizing results (local host)**

**For visualizing the pan-genome of 44 *B. amyloliquefaciens* strains using Docker (Docker must be installed):**

Access the web download page at <http://pangenome.zggskj.com/home> and download the file “B\_amly\_44\_strans\_pan\_genome\_panx\_vis.tar”,store in a home directory of your host. Copy the following commands to visualize the pan-genome of 44 *B. amyloliquefaciens* strains on the local host (Docker must be installed):

$ sudo docker load < B\_amly\_44\_strans\_pan\_genome\_panx\_vis.tar

$ sudo docker run -d --rm -p 8000:8000 busybox\_nodejs: nodejs\_v7.3.0

**For visualizing** **biosynthetic gene clusters** **of 44 *B. amyloliquefaciens* strains:**

Access the web download page at <http://pangenome.zggskj.com/home> and download the file “B\_amyl\_44\_strains\_antiSAMSH\_genecluster.zip”. Extract it into any directory then into the strain folder. Use the browser to open “index.html” to visualize the clusters of the strains.