**The command line and tip of BGDMdocker**

**1.Guide for BGDMdocker workflow usage：**

**1) installing latest Docker-CE** (Ubuntu, Debian, Raspbian, Fedora, Centos, Redhat, Suse, Oracle, Linux *et al*., all applicable):

$ curl -sSL https://get.docker.com/ | bash -x

or:

$ wget -qO- https://get.docker.com/ | bash -x

adding user to the "docker" group with something like:

$ sudo usermod -aG docker manager

Type the following commands at your shell prompt. If this outputs the Docker version, your installation was successful:

$ docker version

On your host (with Docker), type the following command lines to build a BGDMdocker workflow:

$ git clone https://github.com/cgwyx/BGDMdocker.git

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

$ unzip BGDMdocker-master.zip

**2) Build workflow Images:**

$ cd ./BGDMdocker

$ docker build -t BGDMdocker:latest .

Or:pull Images of BGDMdocker from [DockerHub](https://hub.docker.com/r/cgwyx/bgdmdocker/),such as :

$ docker pull cgwyx/BGDMdocker

**3) Running Container and Prokka to genome annotation**

Copy the following commands to run the analysis for genome annotation of **Ba\_xx** strains (for boldface text, please enter your data): **(if you have your own genome sequences, you need this step to generate** **“\*.gbff” annotation files):**

$ docker run —rm -v home:home BGDMdocker \

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(must be in your host).

**4) Running panX analysis 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),and these data should also reside in a folder within“ /pan-genome-analysis/data/ ”

in Container,we will refer to this folder as run directory below. The name of the run directory is used as a species name in down-stream analysis and visualization.Therefore,you need to enter the Container to run the relevant commands,and commit the Container to save image of visualization at last,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). For detailed parameters see [here](https://github.com/neherlab/pan-genome-analysis).

$ cd /pan-genome-analysis

$ cp -r /home/manager/B\_amy /pan-genome-analysis/data/

$ ./panX.py -fn./data/B\_amy -sl B\_amy-RefSeq.txt -t 4

“/home/manager/B\_amy” is your loclhost annotated sequences files of \*.gbff (GenBank files) and B\_amy-RefSeq.txt (accession list for strains), should copy to Container and reside in “/pan-genome-analysis/data/B\_amy” folder, The result will also be output to the this folder in Container.

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\_amywith a web browser to access the visualization of the pan-genome immediately.

Create a new Image from Container for saving changes of visualization data (running in host):

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

**5) Running Container and antiSMASH to search for gene clusters:**

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):

$ docker run —rm -v home:home BGDMdocker:latest\

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 --asf\

--inclusive --full-hmmer --smcogs --verbose --borderpredict

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

**2. 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.

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

$ git clone https://github.com/cgwyx/prokka\_conda\_docker.git

Or: download [“prokka\_conda\_docker-master.zip](https://github.com/cgwyx/prokka_conda_docker/archive/master.zip)” file

$ unzip prokka\_conda\_docker-master.zip

$ cd ./prokka\_conda\_docker-master

$ sudo docker build -t conda:prokka .

Or: pull Image from [DockerHub](https://hub.docker.com/r/cgwyx/prokka_conda_docker/):

docker pull cgwyx/prokka\_conda\_docker

Run a Container from the image and copy the following commands to run the analysis of the genome annotation of **Ba\_xx** strains (for boldface text, please replace with your own data if applicable):

$ docker run —rm -v home:home prokka:latest \

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,thye are all must be in your host.

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

$ git clone https://github.com/cgwyx/panx\_conda\_docker.git

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

$ unzip panx\_conda\_docker-master.zip

$ cd ./panx\_conda\_docker-master

$ sudo docker build -t conda:panx .

Or:pull Image from [dockerhub](https://hub.docker.com/r/cgwyx/panx_conda_docker/):

$ docker pull cgwyx/panx\_conda\_docker

$docker run -it --rm -v /home:/home cgwyx/panx\_conda\_docker:latest

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),you need to enter the container to run the relevant commands:

First your must copy data from your host to “/pan-genome-analysis/data/”,then run panX in Container;

$ cp -r /home/manager/B\_amy /pan-genome-analysis/data/

$ ./panX.py -fn ./data/B\_amy -sl B\_amy-RefSeq.txt -t 4

\*.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>with a web browser to access the visualization of the pan-genome immediately.

Create a new Image for saving changes in Container data of visualization(running in host):

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

**3) Building Image and run Container of antiSMASH4** **with** **database standalone:**

$ git clone https://github.com/cgwyx/antismash4\_db.git

Or: [download](https://github.com/cgwyx/antismash4_lite_docker/archive/master.zip) “[antismash4\_db-master.zip](https://github.com/cgwyx/antismash4_db/archive/master.zip)” file

$ unzip antismash4\_db-master.zip

$ cd ./antismash4\_lite\_docker-master

$ sudo docker build -t conda:antismash4\_db .

Or: pull Image from [dockerhub](https://hub.docker.com/r/cgwyx/antismash4_db/):

$ docker pull cgwyx/antismash4\_db.

Run a Container from the image 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):

$ docker run -it --rm -v home:home cgwyx/antismash4\_db:latest \

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 --asf\

--inclusive --full-hmmer --smcogs --verbose --borderpredict

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

**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/](ftp://ftp.ncbi.nlm.nih.gov/entrez/entrezdirect/versions/current/edirect.zip)

[current/edirect.zip](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;

**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 on your loclhost like our** [**website**](http://bapgd.hygenomics.com/pangenome/home) **using Docker (Docker must be installed):**

Access the web download page at <http://bapgd.hygenomics.com/pangenome/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):

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

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

Open http:// localhost:8000/bamf\_gbk44with a web browser to access the visualization of pan-genome of 44 *B. amyloliquefaciens* strains immediately.

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

Access the web download page at <http://pangenome.zggskj.com/home> and download data of genecluster of all strains,and you may download the standalone strain genecluster at home page also.Extract it into any directory then into the strain folder. Use the browser to open “index.html” to visualize the clusters of the strains.