Plascad User Manual

Contact: You CHE, hkucheyou@gmail.com

Introduction:

The plasmid classification, antibiotic resistance genes (ARGs) annotation and plasmid visualization (Plascad) is a computationally efficient tool to classify the plasmids into three categories (i.e., conjugative, mobilizable and non-mobilizable) based on the protein machinery associated with DNA transfer, including relaxase, type IV coupling protein (T4CP) and type IV secretion systems (T4SSs), and to identify the plasmid-borne ARGs and construct a plasmid graph with key genes associated with plasmid transfer and ARGs (Fig. 1).

We compared Plascad to another similar tool (mob_typer, https://github.com/phac-nml/mob-suite; updated in 09/2020) (1-2), which also allows for automatic plasmid typing, but offering no information regarding the ARGs and plasmid maps, and the result shows that Plascad performs better than mob_typer in terms of the running time and the accuracy (Table 1). Our tool achieves an average of 97.6% sensitivity and 98.8% specificity for the classification of the three categories of plasmids.

Outline of the Plascad Workflow

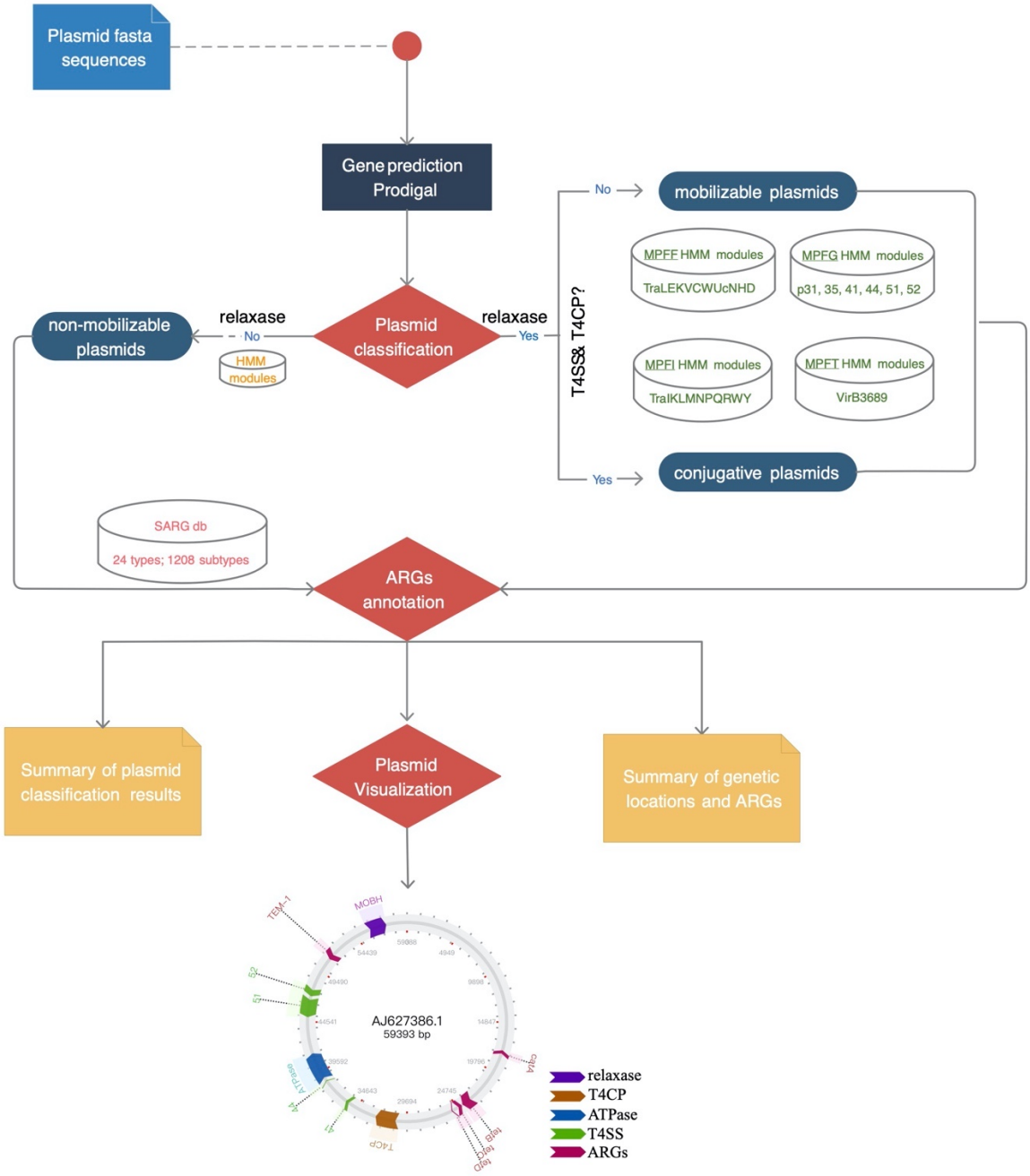


Figure 1. Outline of the Plascad workflow.

Table 1. Comparison of Plascad (blue) and mob_typer (green) performance on plasmid classification

Plasmid types	Sensitivity (Sn)	Specificity (Sp)	Positive predictive value (PPV)
Conjugative	95.5% | 93.2%	100% | 98.4%	100% | 97.6%
Mobilizable	100% | 90.9%	96.3% | 95.2%	88.0% | 83.3%
Non-mobilizable	97.4% | 92.3%	100% | 95.5%	100% | 92.3%
Running time Intel(R) Xeon(R) CPU E5-2683 v4 @ 2.10GHz, 20 threads		5.9 min | 10.2 min	

Note: the plasmid datasets (including 44 conjugative plasmids, 22 mobilizable plasmids and 39 non-mobilizable plasmids) (3) used for the performance evaluation of Plascad and mob_typer could be downloaded from: https://doi.org/10.6084/m9.figshare.13107758.v1.

Pipeline development

1) HMM protein profiles construction

To first perform MOB typing, PSI-BLAST (2.3.0) (4) was conducted using eight families of N-terminal (~300 aa) relaxase protein sequences (MOBC: $1e^{-4}$, MOBF: $1e^{-8}$, MOBH: $1e^{-2}$, MOBP: $1e^{-4}$, MOBQ: $1e^{-4}$, MOBV: $1e^{-2}$, MOBB: $1e^{-4}$ and MOBT: $1e^{-4}$) as queries against all the Prodigal-predicted ORFs (5) of the complete plasmids with a maximum of 14 iterations. The relaxase protein sequences and the plasmid sequences could be downloaded from (https://doi.org/10.6084/m9.figshare.13107767.v1 and https://doi.org/10.6084/m9.figshare.13109771.v1, respectively).

To further classify T4SSs into four archetypes, the major ATPase (*VirB4* and *TraU*), T4CP (*VirD4*) and four sets of well-described type-specific genes (MPFF:

TraLEKVCWUcNHD; MPFG: *p31*, *p35*, *p41*, *p44*, *p51*, *p52*; MPFI: *TraIKLMNPQRWY*; MPFT: *VirB3689*) were considered. These well-described type-specific genes could be downloaded from (https://doi.org/10.6084/m9.figshare.13118111.v1). We also performed a PSI-BLAST search (maximum of 30 iterations) of each key protein against the plasmid database.

Next, multiple alignments coupled with phylogenetic analyses were performed for each uncovered protein family using MUSCLE (v3.8.31) and FastTree (2.1.10) to remove the spurious hits. Finally, we used HMMER (3.2.1) (6) to build HMM protein profiles for each of the manually curated alignments, since two conserved regions were detected for the alignment of the T4CP family; as a result, two HMM profiles were built for VirD4. All the constructed HMM profiles could be found in (https://github.com/pianpianyouche/plascad/tree/master/plas_cad/database/hmm_module). plasmid classification was performed using hmmsearch against the constructed HMM profiles, and the obtained homologs were further filtered based on the alignment length (covering more than 50% of the protein profile) and c-value (< 0.01). A type of MPF was attributed to a conjugative plasmid if the cluster contained at least 5, 4, 4, and 3 type-specific markers for MPFF, MPFI, MPFG and MPFT, respectively, in addition to the colocalization of ATPase and T4CP.

2) ARGs annotation

We embedded a comprehensive structured ARG database (7) (24 types, 1208 subtypes, 12306 sequences), which was developed in our group into this tool. ARGs were

identified by aligning the coding sequences of the plasmids to this database using BLASTP at E value ≤ 10^{-7} with a minimum similarity of 80% over 70% query coverage. The ARG database could be found in (https://github.com/pianpianyouche/plascad/tree/master/plas_cad/database/ARGsdb).

3) Plasmid visualization

To construct the plasmid map with key genes associated with plasmid transfer and ARGs, a plasmid visualization component (https://github.com/vixis/angularplasmid) (8) using AngularJS was integrated into this tool.

Required Dependencies

Linux

Python >= 3.6 (https://www.python.org/downloads/)

Biopython (https://biopython.org/)

Prodigal >= 2.6.3 (https://github.com/hyattpd/Prodigal)

BLAST >= 2.7.1 (ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/)

Hmmer >= 3.2.1 (http://hmmer.org/)

Installation

Plascad can be installed either through conda or pip, though we advise to use Conda.

1) Conda

Use miniconda (https://docs.conda.io/en/latest/miniconda.html) or anaconda

(https://www.anaconda.com/) to install Plascad.

conda create -n Plascad -y -c bioconda -c pianpianyouche plascad

conda activate Plascad

2) pip3

If you have the dependencies (Python >=3.6, blast >=2.7.1, prodigal >=2.6.3, hmmer >=3.2.1) in your PATH, you can install with pip3.

pip3 install Plascad

Usage

Usage: Plascad [-h] [-i] [-n] [-cMOBB, default=75] [-cMOBC, default=75] [-cMOBF, default=75] [-cMOBT, default=75] [-cMOBPB, default=75] [-cMOBH, default=70] [-cMOBP, default=65] [-cMOBV, default=60] [-cMOBQ, default=55]

Help:

-h, Show this help message and exit

-i, FASTA file of plasmid sequences

-n, Prodigal normal mode (default anonymous mode)

-cMOB[B,C,F,T,PB,H,P,V,Q], alignment coverage for MOB HMM profiles

Note: (1), as suggested by the author of Prodigal, short sequences (<100kbp) such as ***plasmids***, phages, and viruses should generally be analyzed using ***Anonymous Mode***.

(2), More potential mobilizable plasmids may be identified by decreasing the alignment

coverage for MOB HMM profiles at the expense of accuracy, thus further validation is required before any conclusion can be drawn.

Input

Plasmid sequences in FASTA format are used as input files.

output

The final output of Plascad includes four files for the given plasmid sequences:

1) Summary file of the plasmid classification results.
2) Genetic location of the genes associated with plasmid transfer and ARGs for both mobilizable and conjugative plasmids.
3) Html files containing the plasmid maps for all the identified conjugative plasmids.

Example

wget https://github.com/pianpianyouche/plascad/raw/master/plas_cad/example/example.fasta

Plascad -i example.fasta

Example output files

1) example_plasmids_classification_sum.txt. Sumary of the plasmid classification results.

Name	Plasmid type	ARGs
AJ627386.1	Conj	chloramphenicol__*catA*;tetracycline__*tetB*; tetracycline__*tetC*;tetracycline__*tetD*;beta-lactam__$bla_{\text{TEM-1}}$
NC_002377.1	Conj	
NC_002483.1	Conj	
NC_005014.1	Conj	tetracycline__*tetD*;tetracycline__*tetC*;tetracycline__*tetB*; aminoglycoside__*aph(3")-I*;aminoglycoside__*aph(6)-I*

note: ARGs are displayed as type_subtype.

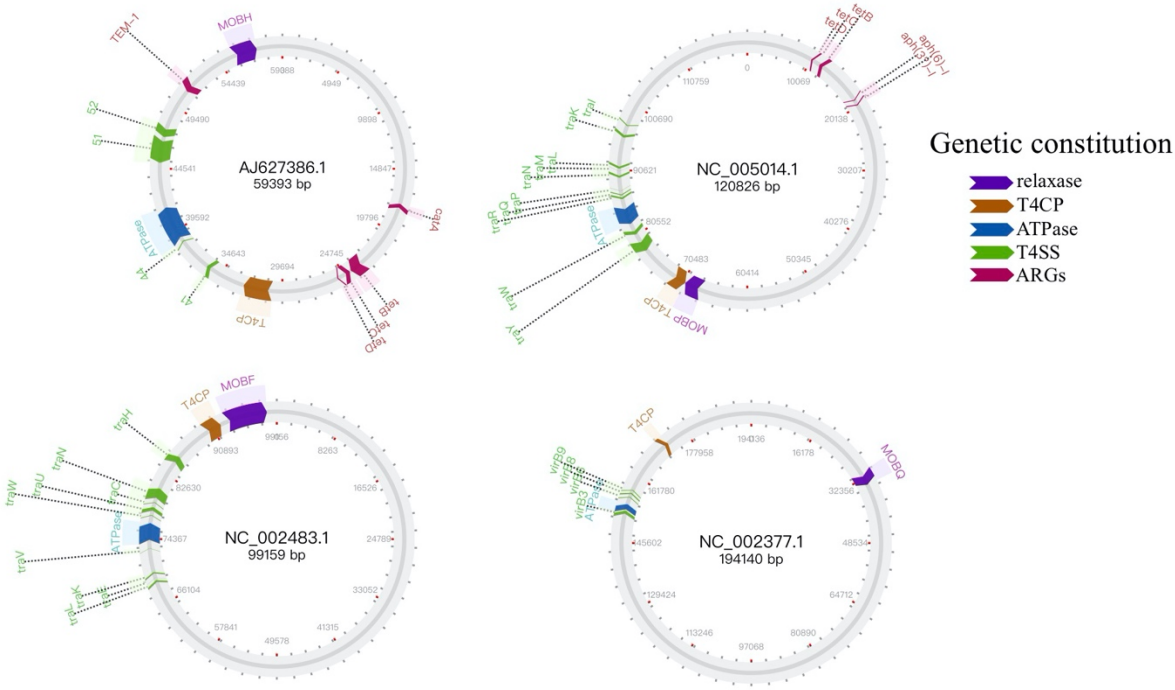
2) example_Conj_plasmids_loc_sum.txt. Genetic location of the genes associated with transfer and the detected ARGs.

AJ627386.1	MPFG_41	1.20E-88	35289-35918	1
AJ627386.1	MPFG_44	9.90E-53	38227-38634	1
AJ627386.1	MPFG_51	8.10E-147	44998-46998	-1
AJ627386.1	MPFG_52	9.00E-123	47013-47954	-1
AJ627386.1	MPFG_ATPase	3.80E-178	38650-41520	1
AJ627386.1	MPFG_T4CP_2	8.80E-54	30481-32730	1
AJ627386.1	MPFG_T4CP_1	4.40E-193	30481-32730	1
AJ627386.1	MOBH	3.90E-67	55497-57404	1
AJ627386.1	chloramphenicol__catA	1.94E-164	17570-18211	-1
AJ627386.1	tetracycline__tetB	0	22897-24102	1
AJ627386.1	tetracycline__tetC	2.36E-166	24215-24883	-1
AJ627386.1	tetracycline__tetD	1.18E-100	24896-25312	1
AJ627386.1	beta-lactam__TEM-1	0	50985-51845	-1
NC_002377.1	MPFT_ATPase	1.20E-276	152657-155026	1
NC_002377.1	MPFT_T4CP_2	3.50E-05	170649-172619	1
NC_002377.1	MPFT_T4CP_1	6.90E-155	170649-172619	1
NC_002377.1	MPFT_virB3	1.60E-29	152331-152657	1
NC_002377.1	MPFT_virB6	4.70E-55	155800-156687	1
NC_002377.1	MPFT_virB8	1.70E-76	156815-157588	1
NC_002377.1	MPFT_virB9	6.30E-97	157585-158466	1
NC_002377.1	MOBQ	5.70E-104	30831-34133	1
NC_002483.1	MPFF_ATPase	0	73741-76368	1
NC_002483.1	MPFF_T4CP_1	0	89804-91957	1
NC_002483.1	MPFF_T4CP_2	3.40E-65	89804-91957	1
NC_002483.1	MPFF_traC	4.20E-134	78378-79016	1

NC_002483.1	MPFF_traE	1.30E-128	68978-69544	1
NC_002483.1	MPFF_traH	0	84052-85428	1
NC_002483.1	MPFF_traK	1.30E-162	69531-70259	1
NC_002483.1	MPFF_traL	9.60E-75	68645-68956	1
NC_002483.1	MPFF_traN	0	79013-80821	1
NC_002483.1	MPFF_traU	1.30E-241	77377-78369	1
NC_002483.1	MPFF_traV	1.10E-108	72710-73225	1
NC_002483.1	MPFF_traW	3.50E-139	76748-77380	1
NC_002483.1	MOBF	4.50E-132	92673-97943	1
NC_005014.1	MPFI_ATPase	0	82377-85421	-1
NC_005014.1	MPFI_T4CP_2	6.50E-74	70127-72418	-1
NC_005014.1	MPFI_T4CP_1	0	70127-72418	-1
NC_005014.1	MPFI_traI	9.90E-134	97428-98246	-1
NC_005014.1	MPFI_traK	1.60E-48	95996-96286	-1
NC_005014.1	MPFI_traL	1.80E-65	91208-91555	-1
NC_005014.1	MPFI_traM	1.50E-109	90519-91211	-1
NC_005014.1	MPFI_traN	1.70E-147	89525-90508	-1
NC_005014.1	MPFI_traP	4.70E-98	87529-88233	-1
NC_005014.1	MPFI_traQ	7.70E-83	87002-87529	-1
NC_005014.1	MPFI_traR	5.20E-55	86547-86951	-1
NC_005014.1	MPFI_traW	1.30E-191	80594-81796	-1
NC_005014.1	MPFI_traY	0	77716-79884	-1
NC_005014.1	MOBP	1.50E-117	67391-70090	1
NC_005014.1	tetracycline__tetD	1.18E-100	10300-10716	-1
NC_005014.1	tetracycline__tetC	2.12E-145	10804-11397	1
NC_005014.1	tetracycline__tetB	0	11510-12715	-1
NC_005014.1	aminoglycoside__aph(3")-I	0	18301-19104	1
NC_005014.1	aminoglycoside__aph(6)-I	0	19104-19940	1

note: c-value for hmmsearch, e-value for blastp

3) example_Conjugative_plasmids_map. Folder containing the maps for all the identified conjugative plasmids.



Tips for visualization

A plasmid visualization component using AngularJS is integrated into our pipeline, all the plasmid maps are in **HTML** formats. In order to view the map locally, you need to download the **js folder** in addtion to the HTML files.

Reference:

1. J. Robertson and J. H.E.N., MOB-suite: software tools for clustering, reconstruction and typing of plasmids from draft assemblies. *Microb Genom*. **4**, e000206 (2018).
2. J. Robertson, K. Bessonov, J, Schonfeld, and J. H.E.N. Universal whole-sequence-based plasmid typing and its utility to prediction of host range and epidemiological surveillance. *Microb Genom*. doi: 10.1099/mgen.0.000435. Online ahead of print (2020).
3. X. Li, *et al.*, oriTfinder: a web-based tool for the identification of origin of transfers in DNA sequences of bacterial mobile genetic elements. *Nucleic Acids Res.* **46**, 229-

234 (2018).

4. S. F. Altschul, *et al*., Basic local alignment search tool. *J Mol Biol*. **215**, 403-410 (1990).
5. D. Hyatt, *et al*., Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*. **11**, 119 (2010).
6. S. R. Eddy. Accelerated profile HMM searches. *PLoS Comput Biol*. **7**, e1002195 (2011).
7. X. Yin, *et al*., ARGs-OAP v2.0 with an expanded SARG database and Hidden Markov Models for enhancement characterization and quantification of antibiotic resistance genes in environmental metagenomes. *Bioinformatics*. **34**, 2263-2270 (2018).
8. R. Chawdry. angularplasmid. (https://github.com/vixis/angularplasmid) (2014).