Plasmid 61 (2009) 47–51

Contents lists available at ScienceDirect

Plasmid

journal homepage: www.elsevier.com/locate/yplas

Classification of plasmid vectors using replication origin, selection marker and promoter as criteria

Zhijun Wang [a,*], Li Jin [a,b], Zhenghong Yuan [c], Grzegorz Węgrzyn [d], Alicja Węgrzyn [e]

[a] CAS-MPG Partner Institute for Computational Biology, Shanghai Institute for Biological Sciences, Chinese Academy of Sciences, Yueyang Road 320#, Shanghai 200031, PR China
[b] MOE Key Laboratory of Contemporary Anthropology and Center for Evolutionary Biology, School of Life Sciences and Institutes of Biomedical Sciences, Fudan University, Shanghai 200433, PR China
[c] Key Laboratory of Medical Molecular Virology, Shanghai Medical College, Fudan University, Shanghai 200032, PR China
[d] Department of Molecular Biology, University of Gdańsk, Kładki 24, Gdańsk 80-822, Poland
[e] Laboratory of Molecular Biology (affiliated with the University of Gdańsk), Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Gdańsk 80-822, Poland

ARTICLE INFO

Article history:
Received 20 June 2008
Revised 22 August 2008
Available online 15 October 2008

Communicated by Eva Top

Keywords:
Plasmid classification
Promoter
Replication origin
Selection marker

ABSTRACT

Although plasmid DNA vectors have been extensively applied in biotechnology, there is still a lack of standard plasmid vector classification. Here, we propose a classification method for commonly used plasmid vectors. Plasmid vectors were classified into different classes based on their replication origin, selection marker and promoter information. The replication origins of plasmid vectors were classified as: prokaryotic replication origin, eukaryotic replication origin and viral replication origin. Selection markers of plasmid vectors were mainly classified as ampicillin, kanamycin, neomycin, chloramphenicol, gentamycin, tetracycline, erythromycin, streptomycin, vancomycin and spectinomycin resistance gene markers. Promoter sequences were also classified as prokaryotic, eukaryotic and viral promoters. Finally, the nomenclature of common plasmid vectors has three determinants. We believe that the classification of plasmid vectors can provide useful information for researchers employing molecular cloning procedures. A web service of the plasmid classification was established and it is available from http://www.computationalmedicalbiology.org/plasclas.aspx.

1. Introduction

Plasmids are extrachromosomal genetic elements able to replicate autonomously and to be maintained in a host cell (Ebersbach and Gerdes, 2005; Węgrzyn, 2005; Ghosh et al., 2006). These replicons are commonly used as cloning vectors in genetic engineering.

Plasmids had been commonly classified as F plasmids (Kline and Palchaudhuri, 1980; Seelke et al., 1982), colicinogenic (Col) plasmids (Zverev et al., 1984) and R plasmids over 30 years ago (Datta, 1977). Currently, there are several different methods for plasmid classification, for example, based on replication mechanism, plasmids were classified into rolling-circle replicating (RCR) plasmids, theta replicating plasmids, and plasmids that use the strand-displacement mechanism of replication (del Solar et al., 1998; Espinosa et al., 1995). However, this kind of classification provides only a very limited information. Classification based on plasmid incompatibility was developed in the early 1970s (Chabbert et al., 1972; Richards and Datta, 1979; Sagai et al., 1976; Sasakawa et al., 1980) and is based on possibility of stable simultaneous maintenance of two tested plasmids in one host. Introduction of a plasmid into a strain carrying another plasmid is crucial for this classification. The strain is examined for the presence of the introduced plasmid after selection. If the introduced plasmid is eliminated, these two plasmids

* Corresponding author. Fax: +86 21 64183281.
E-mail address: zjwang@picb.ac.cn (Z. Wang).

0147-619X/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.plasmid.2008.09.003

are incompatible, and they are assigned to the special incompatibility group. Incompatibility has been (and still is) used as a main feature to classify plasmids, but it has the disadvantage that it does not take into account plasmids with high homology that do not share the replication control genes. So far, there are about 30 plasmid incompatibility groups (Couturier et al., 1988). Nevertheless, currently there is no common classification of plasmid-based cloning vectors.

Here, we classify plasmid vectors based on sequence information. We selected replication origin, selection marker and promoter sequence information as three elements for plasmid classification and developed a web service for plasmid classification.

2. Methods

2.1. Web service for the plasmid classification

Replication origin is a crucial characteristic for any plasmid (Paulsson and Chattoraj, 2006; Wang et al., 2004). Selection markers are extensively used in plasmid DNA vectors for selection of cells bearing the plasmid (Velten and Schell, 1985). Promoter sequences in plasmid vectors control efficiency of recombinant gene expression (Dekhtyar et al., 2008). We employed a combination of these three components for plasmid classification.

A web-based plasmid vector classification service was developed. It was named PLASCLAS, and it is available from the following website www.computationalmedicalbiology.org/plasclas.aspx. It currently runs on the Microsoft Windows 2003 server platform. PLASCLAS differs from other plasmid classification methods and allows scientists to distinguish different vectors. Once plasmid sequence is pasted into the input textbox, the server can output the classification result to the output textbox. A legend is provided below the classification results, thus, the user can quickly check what do particular symbols denote.

2.2. Input sequence

In the web service, only uppercase or lowercase of strings containing A, G, C, T can be recognized as input sequence. If the input sequence contains other characters, the web service will give information "Please input the right sequence", and the user has to input the right sequence again. However, sequences copied directly from GenBank, which include nucleotide numbers and spaces after every 10 nucleotides, can be accepted by the service.

2.3. Reverse-complementary sequence in the plasmid

The replication origin, selection marker or promoter sequences are often reverse-complementary in different plasmid vectors. In order to recognize the reverse-complementary sequences, the inputted plasmid vector sequence is screened two times. (1) The plasmid sequence is searched for replication origin, selection marker and promoter sequences using the forward sequence. (2) The plasmid sequence is searched for the replication origin, selection marker and promoter sequences using the reverse-complementary sequence. Finally, the obtained plasmid element components from steps 1 and 2 are added together.

2.4. Output style

In the output textbox, the final plasmid nomenclature has three determinants separated by dots. The first determinant denotes the replication origin, the second determinant denotes the selection marker, and the third determinant denotes the promoter sequence. If the plasmid has two or several elements of the same category (e.g. two selection markers), they are expressed as the same elements separated by slash(es).

2.5. Determination of classification efficiency

In order to determine the classification efficiency of the web server PLASCLAS, plasmid DNA vectors with different sizes were used for testing. After plasmid DNA sequences were classified into different replication origins, selection markers and promoter types using the web server PLASCLAS, the classification results were compared with the actual component in the plasmid DNA vector. Classification efficiency was determined using following equation:

$$CE(\%) = \frac{CC}{TP} \cdot 100\%, \quad (1)$$

where *CE* means the classification efficiency, *CC* means the number of correctly classified plasmids, and *TP* means the total number of plasmids. If the classification results from PLASCLAS are identical with the actual replication origin, marker and promoter in the plasmid sequence, according to our selection standards (Tables 1–3), the classification efficiency is defined as 100%.

3. Results and discussion

Searching of the NCBI nucleotide database with the keyword "cloning vector" produced 78,346 items (on June 18, 2008). However information about some cloning vectors is not complete. Moreover, a formal reference to plasmid DNA would be useful, because in most cases the names of plasmid DNA vectors are randomly selected by their constructors.

Plasmids have an essential region which contains the replication origin involved in replication and its control. Plasmid replication origin can be defined as the minimal *cis*-acting region that can support autonomous plasmid replication, the region where DNA strands are melted to initiate the replication process (del Solar et al., 1998). In order to select plasmid-containing cells, selection marker(s) is/are necessary in the plasmid. Moreover, promoter sequences are important for recombinant gene expression. Here, we describe a web server PLASCLAS, which can recognize plasmid DNA sequences for classification based on the replication origin, selection marker and promoter sequences. Our classifier automatically recognizes these three components, solves the problems of reverse-complementary sequences in the plasmid, and automatically recognizes a wrong sequence information.

Table 1
Replication origin information for plasmid vector classification

Replication origin type	Class	Replication origin
Not known	0	Unclassified origin
Prokaryotic replication origin	1	pMB1 replication origin
	2	pUC replication origin
	3	ColE1 replication origin
	4	p15A replication origin
	5	pSC101 replication origin
	6	R1 replication origin
	7	RK2 replication origin
	8	R6K replication origin
	9	F1 replication origin
	10	M13 replication origin
	11	Lambda replication origin
	12	pA81 replication origin
	13	pRAS3.1 replication origin
	14	pTi replication origin
	15	pBPS1 replication origin
	16	pUO1 replication origin
	17	pKH9 replication origin
	18	pWKS1 replication origin
	19	pCD1 replication origin
	20	pMAK3 replication origin
	21	pBL63.1 replication origin
	22	pTA1060 replication origin
	23	p4M replication origin
	24	pHT926 replication origin
	25	pCD6 replication origin
	26	pJB01 replication origin
	27	pLME300 replication origin
	28	pMD5057 replication origin
	29	pTE44 replication origin
	30	pDP1 replication origin
	31	pT38 replication origin
Eukaryotic replication origin	A	Yeast 2-micron replication origin
	B	pSR1 replication origin
Viral replication origin	a	EBV replication origin (OriP)
	b	HSV type 2 replication origin (OriS)

Table 2
Selection marker information for plasmid vector classification

Class	Resistance marker	Gene
0	Unclassified	None
1	Ampicillin	*bla*
2	Ampicillin	*blaA*
3	Ampicillin	*blaZ*
4	Kanamycin	*aph*
5	Neomycin	*neo*
6	Chloramphenicol	*cat*
7	Chloramphenicol	*cmlA*
8	Chloramphenicol	*catAIII*
9	Chloramphenicol	*catB2*
10	Chloramphenicol	*cmx*
11	Gentamycin	*aacC1*
12	Gentamycin	*aacC2*
13	Tetracycline	*tetA*(*A*)
14	Tetracycline	*tetA*(*C*)
15	Tetracycline	*tetA*(*D*)
16	Tetracycline	*tetA*(*E*)
17	Tetracycline	*tetA*(*G*)
18	Tetracycline	*tetA*(*H*)
19	Tetracycline	*tetA*(*L*)
20	Tetracycline	*tetA*(Q)
21	Tetracycline	*tetA*(*S*)
22	Tetracycline	*tetA*(*Y*)
23	Tetracycline	*tetA*(*Z*)
24	Erythromycin	*erm*
25	Vancomycin	*van*
26	Spectinomycin	*aadA7*
27	Streptomycin	*str*

Table 3
Promoter information for plasmid vector classification

Promoter type	Class	Promoter information
Not known	0	Unclassified promoter
Prokaryotic promoter	1	T7 promoter
	2	T5 promoter
	3	T3 promoter
	4	SP6 promoter
	5	pR promoter
	6	pL promoter
	7	Arabinose inducible promoter (PBAD)
	8	EM-7 promoter
	9	Ptac promoter
	10	Plac promoter
	11	lacUV5 promoter
Eukaryotic promoter	A	Phosphoglycerate kinase (PGK) promoter
	B	Chicken beta-actin promoter
	C	Elongation factor (EF1)-alpha promoter
	D	Human H1 promoter
	E	U6 promoter
Viral promoter	a	Cytomegalovirus CMV promoter
	b	Cauliflower mosaic virus CaMV 35S promoter
	c	Rous sarcoma virus (RSV) promoter
	d	Simian virus (SV40) promoter
	e	Mouse mammary tumor virus (MMTV) promoter

In the web service, the replication origins of plasmid DNA vectors were classified as prokaryotic replication origin, eukaryotic replication origin and viral replication origin (Table 1). Moreover, '0' means an unclassified replication origin; '1–31' mean 31 different prokaryotic replication origins; 'A' and 'B' mean selected eukaryotic replication origins; 'a', 'b' mean selected viral replication origins. The selection markers were classified as 27 classes (Table 2), '0' means an unclassified selection marker. '1–27' mean 27 different resistance gene markers. Promoter sequences were classified as prokaryotic promoter, eukaryotic promoter or viral promoter (Table 3). '0' means an unclassified promoter; '1–11' mean 11 different prokaryotic promoter sequences; 'A–E' mean different eukaryotic promoter sequences, 'a–e' mean different viral promoter sequences. Table 4 shows examples of classification of some commonly used plasmid vectors. For instance, pBR322 was classified as: 1.1/14.0, as this plasmid has a pMB1 replication origin, ampicillin resistance gene, tetracycline resistance gene (*tetA*(*C*)), and an unclassified promoter. pUC19 was classified as: 2.1.11, as it has a pUC replication origin, ampicillin resistance gene and *lacUV5* promoter. pUT18 was classified as: 1.1.9, as it has pMB1 replication origin, ampicillin resistance gene, and Ptac promoter.

Our plasmid classification style has following important characteristics with regard to GenBank accession number. The GenBank numbers of the plasmid DNA of interest are arranged according to the order of deposit. It is not possible

Table 4
Classification results of selected plasmids

GenBank accession no	Plasmid	Classification
J01749	pBR322	1.1/14.0
M77789	pUC19	2.1.11
J01566	ColE1	3.0.0
D13250	pUT18	1.1.9
X06403	pACYC184	4.14.0
AJ515144	pA81	12.0.0
AY043298	pRAS3.1	13.14.0
AF060155	pTiC58	14.0.0
AJ404864	pBPS1	15.0.0
AB063332	pUO1	16.0.0
AF295336	pKH9	17.0.0
AF482428	pWKS1	18.0.0
AF053946	pCD1	19.0.0
AB366442	pMAK3	20.0.0
AJ748844	pBL63.1	21.0.0
U32380	pTA1060	22.0.0
AF359574	p4M	23.0.0
D43692	pHT926	24.0.0
AY350745	pCD6	25.0.0
AY425961	pJB01	26.0.0
AJ488494	pLME300	27.0.0
AF440277	pMD5057	28.0.0
AY082384	pTE44	29.24.0
AF047696	pDP1	30.0.0
AJ566638	pt38	31.0.0
DQ023608	p426GALL	1/9/10/A.1.1/3/10
X02398	pSR1	B.0.0
U75992	pFlpAB-4	1/a.1.1/11/a/d

to get information about replication origin, marker and promoter only from the GenBank numbers, and to obtain such data it is necessary to search the sequence with the GenBank number or to analyze a plasmid map. However, our plasmid DNA classification style can provide important information about replication origins, markers and promoter sequences from the classification names. In other words, one can directly recognize important information from classification name.

The classification accuracy of PLASCLAS was tested using 3000 different plasmids. The classification efficiencies were determined as described in Section 2. Our results show that classification efficiency is gradually decreased with the increase of plasmid DNA size (Fig. 1). The average classification efficiency is 98% with the plasmid DNA size as low as 4.0 ± 0.5 kb. Average classification efficiency is 80% with inputted plasmid DNA size of 20 ± 1 kb.

In summary, we used sequence comparative algorithm to search different components in plasmid DNA vectors, and then determined the component information. Because we selected only common replication origins, selection markers and promoter sequences as the classification standards, some replication origins, selection markers and promoter sequences were not included as standards for selection. Although PLASCLAS can not distinguish all replication origins, selection markers or promoter sequence information, we believe that this classification may be useful in the practice of genetic engineering. Obviously, PLASCLAS can be extended to include more replication origins, markers, and promoters when new plasmid vectors are constructed and described. We assume that this classification, which is an open system, should grow up and be more

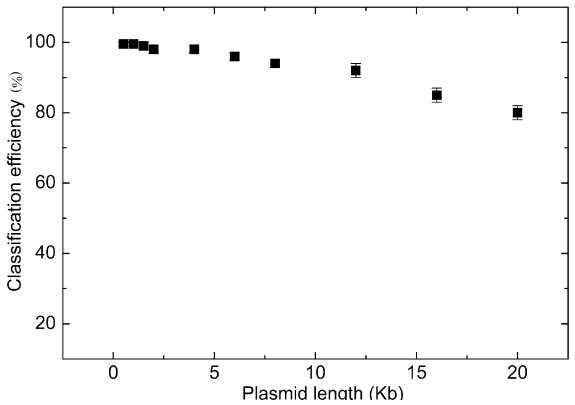

Fig. 1. The classification efficiency of PLASCLAS plotted against plasmid vector size. In total, 3000 plasmids were used in the determination of classification efficiency. Each point in the figure means an average classification efficiency of 300 plasmids, and these 300 plasmids have similar plasmid size with the standard deviation <1.5%.

and more complete when new vectors appear. Moreover, the authors would be grateful for suggestions to include replication origins, selection markers and promoter sequences which were not included in the current version of PLASCLAS (see http://www.computationalmedicalbiology.org/plasclas.aspx) and which are employed in already constructed cloning vectors (the information where to send suggestions of incorporation of new sequences is provided in the web site).

Acknowledgments

This work was supported by National Natural Science Foundation of China (Grant no. 30400077), University of Gdańsk (task Grant no. DS/1480-4-114-08) and Institute of Biochemistry and Biophysics of the Polish Academy of Sciences (task Grant 32.1).

References

Chabbert, Y.A., Scavizzi, M.R., Witchitz, J.L., Gerbaud, G.R., Bouanchaud, D.H., 1972. Incompatibility groups and the classification of fi- resistance factors. J. Bacteriol. 112, 666–675.
Couturier, M., Bex, F., Bergquist, P.L., Maas, W.K., 1988. Identification and classification of bacterial plasmids. Microbiol. Rev. 52, 375–395.
Datta, N., 1977. Classification of plasmids as an aid to understanding their epidemiology and evolution. J. Antimicrob. Chemother. 3 (Suppl. C), 19–23.
Dekhtyar, M., Morin, A., Sakanyan, V., 2008. Triad pattern algorithm for predicting strong promoter candidates in bacterial genomes. BMC Bioinformatics 9, 233.
del Solar, G., Giraldo, R., Ruiz-Echevarria, M.J., Espinosa, M., Diaz-Orejas, R., 1998. Replication and control of circular bacterial plasmids. Microbiol. Mol. Biol. Rev. 62, 434–464.
Ebersbach, G., Gerdes, K., 2005. Plasmid segregation mechanisms. Annu. Rev. Genet. 39, 453–479.
Espinosa, M., del Solar, G., Rojo, F., Alonso, J.C., 1995. Plasmid rolling circle replication and its control. FEMS Microbiol. Lett. 130, 111–120.
Ghosh, S.K., Hajra, S., Paek, A., Jayaram, M., 2006. Mechanisms for chromosome and plasmid segregation. Annu. Rev. Biochem. 75, 211–241.
Kline, B.C., Palchaudhuri, S., 1980. Genetic studies of F plasmid maintenance genes. Plasmid 4, 281–291.
Paulsson, J., Chattoraj, D.K., 2006. Origin inactivation in bacterial DNA replication control. Mol. Microbiol. 61, 9–15.
Richards, H., Datta, N., 1979. Reclassification of incompatibility group L (IncL) plasmids. Plasmid 2, 293–295.

Sagai, H., Hasuda, K., Iyobe, S., Bryan, L.E., Holloway, B.W., Mitsuhashi, S., 1976. Classification of R plasmids by incompatibility in *Pseudomonas aeruginosa*. Antimicrob. Agents Chemother. 10, 573–578.

Sasakawa, C., Takamatsu, N., Danbara, H., Yoshikawa, M., 1980. A method of plasmid classification by integrative incompatibility. Plasmid 3, 116–127.

Seelke, R.W., Kline, B.C., Trawick, J.D., Ritts, G.D., 1982. Genetic studies of F plasmid maintenance genes involved in copy number control, incompatability, and partitioning. Plasmid 7, 163–179.

Velten, J., Schell, J., 1985. Selection-expression plasmid vectors for use in genetic transformation of higher plants. Nucleic Acids Res. 13, 6981–6998.

Wang, Z., Yuan, Z., Hengge, U.R., 2004. Processing of plasmid DNA with ColE1-like replication origin. Plasmid 51, 149–161.

Węgrzyn, G., 2005. What does "plasmid biology" currently mean? Summary of the Plasmid Biology 2004 Meeting. Plasmid 53, 14–22.

Zverev, V.V., Kuzmin, N.P., Zuyeva, L.A., Burova, E.I., Alexandrov, A.A., Khmel, I.A., 1984. Regions of homology in small colicinogenic plasmids. Plasmid 12, 203–205.