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.