

핵산증폭용 특정 길이의 Primer 검색 프로그램

조영훈¹ · 박기정² · 이대상^{1*}

¹한국폴리텍 바이오대학 바이오생명정보과, ²(주)스몰소프트 정보기술연구소

PCR 방법과 더불어 최근 핵산증폭기법으로 주목받기 시작한 방법이 Rolling Circle Amplification (RCA) 기법을 응용한 것이다. 본 논문에서 RCA 기술을 활용하는 실험에 이용할 수 있는 특정 길이(N-mer)를 가진 primer의 후보 리스트를 vector 서열이나 특정 미생물 염색체 서열에서 검색할 수 있는 프로그램인 RCA-mer를 개발하였다. RCA-mer는 사용자의 염기서열을 입력으로 받아 이 서열을 대상으로 특정 길이의 N-mer에 대한 존재 유무에 대한 리스트의 후보를 웹으로 보여주는 기능을 가지고 있다. 또한 서로 다른 두 개의 염기서열들에 대해 N-mer가 공통으로 존재하는지, 한쪽 서열에만 존재하는지 확인할 수 있는 기능을 가지도록 개발하였다. 사용자는 선별된 primer 후보들로부터 적절한 N-mer 서열을 선택하여 실제 RCA를 이용한 실험 곧바로 응용할 수 있도록 후보 primer를 선별하여 사용할 수 있도록 하였다.

Key words □ Fosmid, N-mer, PCR, primer, rolling circle amplification, vector

Rolling circle amplification mechanism (RCA)은 phage나 virus에서 원형으로 구성되어 있는 DNA를 복제하는 방법으로, 원형의 DNA 주형에 oligonucleotide primer가 annealing되어 특정 가닥의 DNA가 마치 자전거 바퀴를 돌리듯이 복제되어 확장되는 것을 일컫는 말이다(4, 5, 8). PCR과 같이 핵산을 증폭할 수 있는 방법으로 최근 RCA를 응용하는 방법이 SNP genotyping, protein-profiling분야에서 각광을 받고 있다(3, 6, 7). RCA 기법은 최근 미생물 유전체 프로젝트와 같은 대규모의 프로젝트를 수행하는데 있어서도 기존 PCR을 이용한 insert를 증폭하기 위해 plasmid를 준비하고, 특정 colony를 선별하여 insert를 확인하는 과정까지의 단계가 여러 단계일 뿐만 아니라, 노동집약적이라는 단점을 가지고 있기 때문에 이러한 단점을 극복한 multiply-primed RCA 기법의 발전으로 그 활용도가 점차 높아지고 있다(4). RCA를 응용한 유전자 증폭의 경우 각각 중 특이적, vector 특이적 primer를 활용하여 유전자를 증폭할 수 있고, PCR처럼 증폭반응을 수행하는데 있어 온도의 변화를 주는 것이 아니라, RCA 반응에 특이적인 isothermal DNA 중합효소를 활용하여 일정한 온도에서도 증폭 반응을 수행할 있는 등의 실험상의 편의성 덕분에 Fosmid library나 BAC library를 구축하여 대용량으로 염기서열 증폭이 필요한 분야에서의 활용도가 높아지고 있는 상황이다(1, 2).

본 논문에서는 앞으로 응용분야가 더욱 넓어질 것으로 예상되는 RCA 기법을 활용한 연구에서 주로 쓰이게 될 특정 길이를 가진 primer들의 존재 유무를 vector, 미토콘드리아, 미생물 염색체 등과 같은 서열 내에서 검색하고 그 결과로 나온 primer 후보

들을 이용할 수 있는 웹기반으로 운영되는 프로그램인 RCA-mer를 개발하였다.

개발내역

RCA-mer 개발에 있어 Perl을 프로그래밍 언어로 사용하였고, 웹기반으로 모든 검색과 계산이 수행되도록 설계하였다. 본 프로그램은 한 개의 서열에 존재하는 primer 리스트를 검색할 수 있는 모듈과 두 개의 염기서열을 동시에 검색하여 공통으로 존재하거나, 한 쪽에만 존재하는 경우, 양쪽 모두에 존재하지 않는 primer 목록을 검색하는 모듈로 구성되어 있다.

첫째로, 한 개의 서열에 존재하는 primer 리스트를 검색할 수 있는 프로그램 전체 흐름도는 Fig. 1과 같으며 각 과정에서 주요

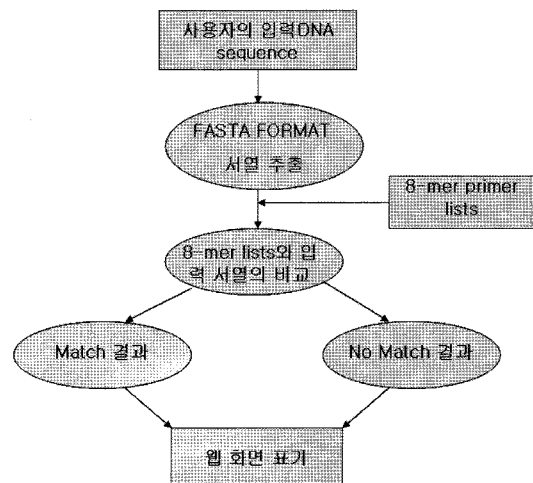


Fig. 1. Flow chart for finding 8-mer primer lists in a single sequence.

*To whom correspondence should be addressed.
Tel: 82-41-746-7374, Fax: 82-41-746-7370
E-mail: gencia@gmail.com

작업은 다음과 같다.

먼저, 검색하고자 하는 목표가 되는 서열인 vector, 미토콘드리아, 미생물 유전체 등의 FASTA format으로 이뤄져 있는 DNA 서열 파일을 입력한다. 사용자가 입력한 서열 파일로부터 DNA 서열 데이터만을 추출하고, 미리 만들어 놓은 8-mer primer 목록들의 조합들과 사용자의 입력서열을 상호 비교하게 된다.

검색 결과를 8-mer primer 목록과 사용자의 입력 서열이 일치하는 경우와 일치하지 않는 경우로 구분하고, 일치와 불일치 각각에 해당하는 서열의 개수를 표시하도록 하였다. 마지막으로 일치하는 8-mer primer 서열을 웹 화면을 통해 출력하여 사용자가 실험에 응용할 수 있도록 하였다. 분석의 각 단계별 진행 상태를 보여주는 screen shot 화면은 Fig. 2와 같다.

두 번째로, RCA-mer는 사용자가 입력으로 사용한 두 개의 서열에 공동으로 존재하거나 한 쪽에만 존재하는 특정 길이의 primer에 대한 검색기능을 가지고 있다.

이 방법은 주로, 특정 vector (Fosmid library, BAC library)에 삽입되어 있는 insert DNA를 RCA기법을 활용하여 증폭할 때, 숙주로 사용한 미생물의 염색체 DNA가 서열결정과정에서 오염되는 것을 막을 수 있는 방법에 활용될 수 있을 것으로 추정된다. 예를 들어, 숙주에는 존재하지 않고 vector 서열에만 존재하는 primer 서열을 찾았다고 가정할 경우, vector 서열에만 존재하는 primer 목록들을 RCA용 multiplex primer로 활용하여 vector

내의 삽입서열을 RCA 방법을 활용하여 증폭할 수 있을 것으로 판단된다.

두 개의 염기서열에서 primer 목록을 검색하는 방법은 다음과 같다. 검색하고자하는 목표 서열이 FASTA Format으로 구성된 두 개의 서열 파일을 각각 입력한다. 입력된 두 개의 서열 파일에서 DNA 서열 부분만을 추출한 후 사용자에게 검색에 사용할 서열이 맞는지 확인 시키는 과정을 거친 후, 미리 작성된 검색용 8-mer primer 목록 파일을 이용하여 검색 프로그램에 서열 정보를 전달하게 된다.

이 단계까지는 기본 프로그램인 한 개의 서열에서 검색하는 것과 동일하다. 이 후 8-mer 검색 부분에서는 일단 입력된 서열 데이터 중의 하나에 대해서 검색을 먼저 진행한다. 예를 들어, 두 개의 서열 데이터를 가지고 검색을 하는 것이라면 검색이 진행될과 동시에 일치하는 것으로 나오는 서열들을 이용하여, 남아 있는 나머지 서열자료에 대해 한 번 더 검색하는 과정을 거친다.

8-mer primer 목록들이 입력한 서열 데이터 두 개에 모두 존재하는 경우, 두 개의 입력 서열들 가운데 한쪽에만 일치하는 경우, 그리고 양쪽 모두 일치하지 않는 경우로 분류된다. 그 결과물들은 웹 화면을 통해 출력되어 사용자에게 보여진다. 두 개의 DNA 서열들로부터 8-mer primer 목록을 검색하는 프로그램의 흐름도와 단계별 screen shot은 Fig. 3과 Fig. 4와 같다.

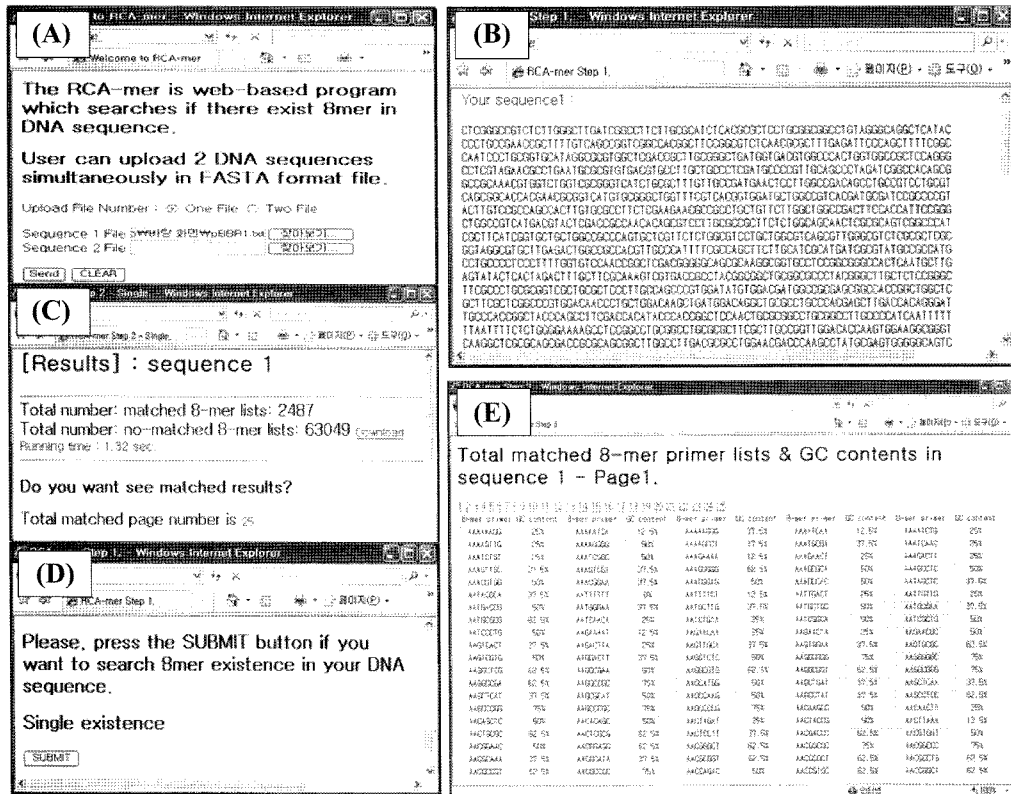


Fig. 2. Screen shots of finding 8-mer primer lists in a single sequence. (A) Main page of RCA-mer. (B) User's sequence information display page. (C) Preliminary results of RCA-mer. (D) Detail viewer button to explore match results of primer. (E) Final results page of matched primer lists and their GC-content.

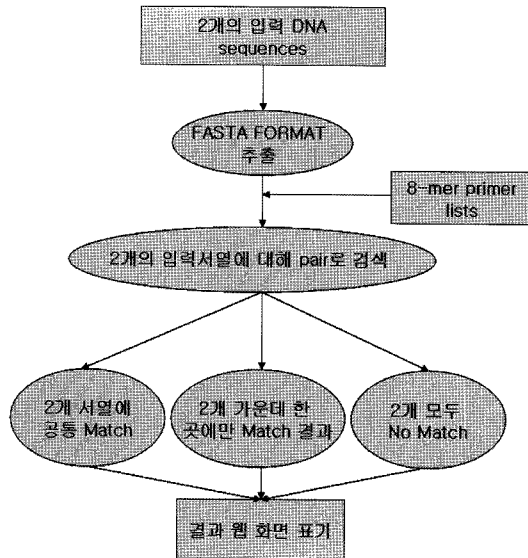


Fig. 3. Flow chart for finding 8-mer primer lists in two sequences.

사례연구

RCA-mer 프로그램을 이용하여 한 개의 DNA 서열에 대해 존재하는 8-mer primer 검색을 위한 사례연구로 크기가 2,736 bp

인 pBBR1 vector (ftp://ftp.ncbi.nih.gov/pub/UniVec/)의 염기서열을 사용하여 개발시스템(Pentium IV 3.0 GHz CPU, 500 MB RAM, Linux Fedora)에서 검색을 수행하여 보았다. 검색에 사용할 8 base 길이를 가진 전체 65,536개의 8-mer들에 대하여 일치하는 것이 2,487개, 일치하지 않는 것이 63,049개 존재하는 것으로 나왔다.

검색이 완료되기까지 소요된 시간은 약 1.3초 걸렸다. 소요시간의 경우, 검색하고자 하는 서열의 길이를 5 kb, 10 kb, 20 kb로 늘렸을 때, 각각 1.61초, 2.45초, 5.04초가 소요되는 것으로 보아 1 kb당 약 0.3초가 소요되며 서열의 길이가 1 kb가 증가함에 따라 1.5~2배 가량 소요시간이 증가된다는 것을 알 수 있었다.

pBR1 vector의 염기서열에 대해 일치하는 8-mer primer 후보군의 염기비율을 살펴보면 A, T, G, C의 비율은 16.4%, 19.9%, 30.7%, 33.0%로 나타났다. GC-content의 비율이 약 63.7% 정도로 상당히 높게 나타났다.

2개의 서열에 대한 8-mer primer 목록 검색의 사례연구는 크기가 4,458 bp인 pBR322 서열(ftp://ftp.ncbi.nih.gov/pub/UniVec/)과 *E. coli* K12 MG1655 (ftp://ftp.ncbi.nih.gov/genomes/Bacteria/) 염색체 서열을 사용하였다. 검색결과 두 개의 서열 모두에 존재하는 8-mer primer 목록들의 개수는 4,175개, pBR322 cloning vector에만 존재하는 것은 4개, *E. coli* K12 MG1655에만 존재하는 것이 61,185개, 두 개의 서열 모두에 존재하지 않는 것이

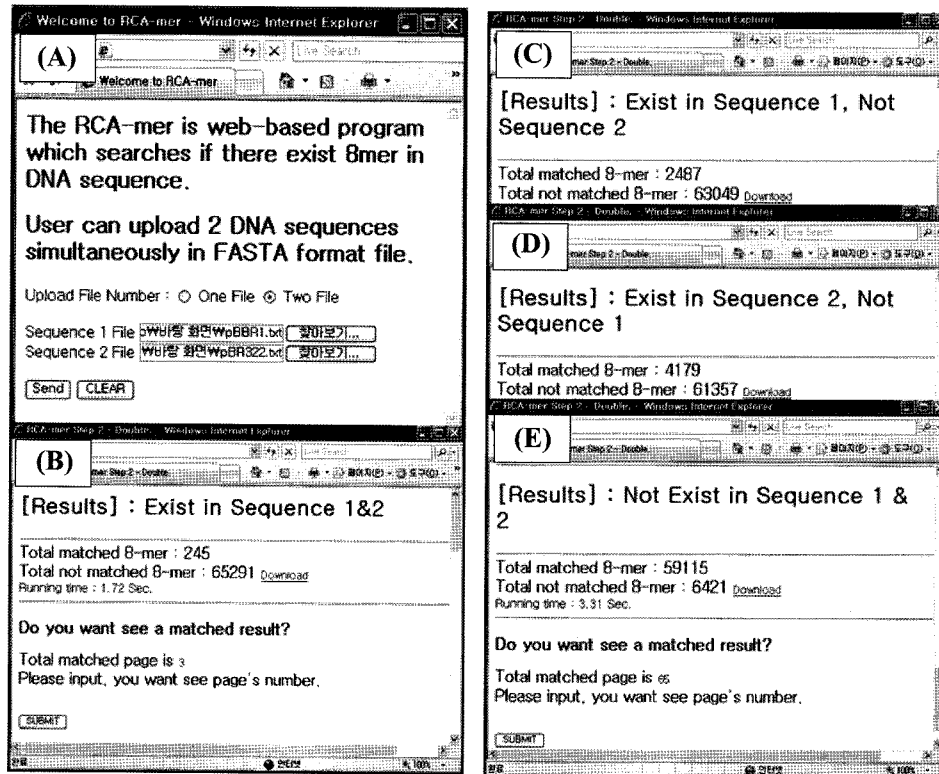


Fig. 4. Screen shots of finding 8-mer primer lists in two sequences. (A) Main page of user interface for two sequences input. (B) Result page of 8-mer coexistence lists in sequence 1 and sequence 2. (C) Result of primer lists which only exist in sequence 1. (D) Result of primer lists which only exist in sequence 2. (E) Result of primer lists which does not exist in sequence 1 as well as sequence 2. User can access the detail 8-mer lists by clicking submit button.

172개였다. 대장균에는 존재하지 않고 pBR322에만 존재하는 4개의 primer 목록들(ACACTAGA, GCTAGAGT, CACCTAGA, CTAGAAGG)의 경우 pBR322에 특이적인 RCA용 primer들로 대장균의 염색체 DNA에 의한 오염의 염려 없이 활용할 수 있을 것으로 사료된다.

향후계획

현재 주목받기 시작한 RCA기법을 활용한 유전자 증폭 방법은 한가지의 primer를 가지고 이것과 일치하는 서열을 가지고 있는 핵산을 모두 증폭시킬 수 있다는 점에서 PCR 반응이 각 핵산마다 다른 primer를 사용해야 한다는 단점을 보완시켜 줄 수 있는 기술이라고 할 수 있다. 이에 특정 서열에 특이적으로 사용 가능한 primer 목록의 후보군을 찾아주는 프로그램인 RCA-mer는 RCA 기술을 응용하여 DNA를 증폭하고자 하는 연구와 바이오 산업계에서 응용성이 클 것으로 생각된다.

현재 검색하고자 하는 N-mer의 길이를 8개로 고정을 시켜 두었으나, 향후 9-mer, 10-mer, 11-mer, 또는 사용자가 원하는 특정 길이의 N-mer에 대해 검색이 가능한 기능을 추가하고, 검색결과 후보로 나오는 primer 목록들에 대해 이들의 위치가 사용자가 입력한 서열의 어느 부분에 존재하는지 객관적으로 확인 할 수 있는 모듈을 추가할 예정이다.

감사의 말

본 연구는 한국폴리텍 바이오대학의 FL 시스템의 일환인 현장 실습 및 프로젝트실습 과제에 의해 수행되었다. RCA-mer 개발

에 대한 조언과 활용에 도움을 주신 솔젠트(주) 고민수 박사에게 감사드린다.

참고문헌

1. 이대상, 박기정. 2007. 미생물 유전체 프로젝트 수행을 위한 Base-Calling 오류감지 프로그램 및 알고리즘 개발. 한국미생물학회지 43, 317-320.
2. 이대상, 태홍석, 박기정. 2003. 유전정보분석시스템. 전자공학회지 30, 68-78.
3. Alsmadi, O.A., C.J. Bornarth, W. Song, M. Wisniewski, J. Du, J.P. Brockman, A.F. Faruqi, S. Hosono, Z. Sun, Y. Du, X. Wu, M. Egholm, P. Abarzua, R.S. Lasken, and M.D. Driscoll. 2003. High accuracy genotyping directly from genomic DNA using a rolling circle amplification based assay. *BMC Genomics* 4, 1471-2164.
4. Dean, F.B., J.R. Nelson, T.L. Giesler, and R.S. Lasken. 2001. Polymerase and multiply-primed rolling circle amplification. *Genome Res.* 11, 1095-1099.
5. Fire, A. and S.Q. Xu. 1995. Rolling replication of short DNA circles. *Proc. Natl. Acad. Sci. USA* 92, 4641-4645.
6. Fumito, M., K. Takehiko, Y. Nobuyasu, T. Katsuji, and N. Masao. 2005. Visualization and enumeration of bacteria carrying a specific gene sequence by *in situ* rolling circle amplification. *Appl. Environ. Microbiol.* 71, 7933-7940.
7. Kingsmore, S.F. and D.D. Patel. 2003. Multiplexed protein profiling on antibody-based microarrays by rolling circle amplification. *Curr. Opin. Biotechnol.* 14, 74-81.
8. Liu, D., S.L. Daubendiek, M.A. Zillman, K. Ryan, and E.T. Kool. 1996. Rolling circle DNA synthesis: Small circular oligonucleotides as efficient templates for DNA polymerases. *J. Am. Chem. Soc.* 118, 1587-1594.

(Received May 14, 2008/Accepted June 4, 2008)

ABSTRACT: RCA-mer: A Web-Based Program Searching for Primer Candidates

Young-Hoon Cho¹, Kiejung Park², and Daesang Lee^{1*} (¹Department of Bioinformatics, Korea Bio Polytechnic, Chungnam 320-905, Republic of Korea, ²Information Technology Institute, SmallSoft Co. Ltd., Daejeon 305-343, Republic of Korea)

Recently, rolling circle amplification (RCA) technique has been widely focused in the field of gene amplification just like PCR method. We have developed RCA-mer, which is a web-based program searching for primer candidates from a given sequence. It can be applied to find primer lists in DNA amplification experiment based on RCA method. The RCA-mer compares 8-mer primer lists with user's input sequence such as vector, mitochondria, and microbial genome sequence. After calculating 8-mers existences in a given sequence, it displays matched and no-matched primer lists with their GC-contents. In addition to it, RCA-mer can search the existence of 8-mer primer lists in two sequences whether they are co-existed or not. Users can apply candidate primer lists to their researches which use RCA techniques.