

## 픽프라이머 : 유전자 목표 구간 탐색 모듈을 포함한 프라이머 제작 그래픽 프로그램

정 희<sup>1†</sup> · 문정환<sup>2†</sup> · 이성찬<sup>1</sup> · 유희주<sup>1\*</sup>

<sup>1</sup>가톨릭대학교 생명과학과, <sup>2</sup>국립농업과학원 유전자분석개발과

### Pickprimer: A Graphic User Interface Program for Primer Design on the Gene Target Region

Hee Chung<sup>1†</sup>, Jeong-Hwan Mun<sup>2†</sup>, Seung-Chan Lee<sup>1</sup>, and Hee-Ju Yu<sup>1\*</sup>

<sup>1</sup>Department of Life Sciences, The Catholic University of Korea, Bucheon 420-743, Korea

<sup>2</sup>Department of Agricultural Biotechnology, National Academy of Agricultural Science, Suwon 441-707, Korea

**Abstract.** In genetic and molecular breeding studies of plants, researchers need to design various kinds of primers based on their research purposes. So far many kinds of web- or script-based non-commercial programs for primer design are available. Because most of them do not include user interface for multipurpose usage including gene structure prediction and direct target selection on sequences, it has been a laborious work to design primers targeting on the exon or intron regions of interesting genes. Here we report a primer designing graphic user interface program, Pickprimer, that includes gene structure prediction and primer design modules by combining source codes of the Spidey and Primer3 programs. This program provides simple graphic user interface to input sequences and design primers. Genomic sequence and mRNA or coding sequence of genes can be copy and pasted or input as fasta or text files. Based on alignment of the input sequences using the Spidey module, a putative gene structure is graphically visualized along with exon-intron sequences of color codes. Primer design can be easily performed by dragging mouse on the displayed sequences or input primer targeting position with desirable values of primers. The output of designed primers with detailed information is provided by the Primer3 module. PCR evaluation of 24 selected primer sets successfully amplified single amplicons from six *Brassica rapa* cultivars. The Pickprimer will be a convenient tool for genetic and molecular breeding studies of plants.

**Additional key words:** gene structure prediction, genetic breeding, molecular marker, PCR, target selection

## 서 언

DNA 분자표지는 식물의 유전체 분석과 육종 연구에 필수적으로 활용되는 매우 중요한 도구이다. 다양한 종류의 분자표지가 식물 유전체의 분석과 활용에 광범위하게 적용되고 있는데, 이용의 편리성과 정확성, 다양성 등 다양한 장점 때문에 염기서열 기반 PCR(polymerase chain reaction) 분자표지가 가장 널리 쓰이고 있다. 분자표지가 적용되는 유전 육종 연구 분야는 대단히 많다. DNA에 존재하는 염기서열의 변이를 분석함으로써 유전지도를 작성하고, 이를 이

용하여 중요한 농업형질들의 유전체 상의 위치를 결정할 수도 있고, 특정 형질과 밀접하게 연관된 표지를 이용하여 유전자를 직접 클로닝할 수도 있으며(map-based cloning), 유전자지문(fingerprinting)에 의한 유전자원 탐색에도 활용할 수 있다. 또한 중요 형질과 강하게 연관된 표지 인자를 밝힐 경우 이를 이용하여 여교잡 육종 시 선발 표지로 활용할 수도 있을 뿐만 아니라 야생 종으로부터 유용한 유전자를 도입할 때 도입 유전자를 신속, 정확하게 확인하여 선발할 수 있으며, 양적 유전 형질(quantitative trait loci, QTL)을 분석할 때도 매우 효과적으로 이용할 수 있다.

\*Corresponding author: yuheeju@catholic.ac.kr

※ Received 7 July 2011; Accepted 10 August 2011. <sup>†</sup>These authors are contributed equally to this work. 본 연구는 가톨릭대학교(과제번호: M-2010-B0002-00158) 와 한국연구재단(과제번호: 2011-0013214)의 지원에 의해 수행되었음.

최근 들어 다양한 식물 종에 대한 유전체 연구가 광범위하게 진행됨에 따라 방대한 데이터로부터 유용한 정보를 효과적으로 발굴, 가공하여 분자표지로 개발하기 위해서는 적절한 생물 정보 프로그램의 활용이 반드시 필요하다. 특히 유전자 구조 예측, 엑손-인트론 경계 설정, 프라이머 제작 등은 유전자 클로닝 또는 분자표지 개발 등 식물의 육종과 유전 연구를 위해 매우 기본적이고 중요한 과정이기 때문에 이와 관련된 다양한 상용 또는 공용 프로그램과 툴이 개발되어 사용되고 있다. 그러나 상용 프로그램들의 경우 비싼 가격 때문에 이용이 제한적이며, 공용 프로그램은 인터넷 상에서 누구나 쉽게 사용할 수 있음에도 불구하고 많은 경우 사용자 편의적인 환경을 제공하지 못하기 때문에 활용도가 낮은 문제점이 있다. 따라서 공용 프로그램을 연구 목적에 따라 조합하거나 소스 코드를 사용자 편의에 맞게 가공하면 그 활용도를 크게 높일 수 있다.

PCR 분자표지를 개발하기 위해 프라이머를 디자인할 때 가장 널리 이용되고 있는 공용 프로그램 중 하나는 Primer3이다(Koressaar and Remm, 2007). 이 프로그램은 소스 코드를 내려 받아 서버나 개인용 컴퓨터에서 웹 기반 또는 스크립트 툴로 다양하게 활용할 수 있으나 유전자의 엑손-인트론 경계 설정이나 염기서열의 정렬이 불가능하기 때문에 인트론 길이 다형성 마커(intron length polymorphism marker) 또는 COS 마커(conserved ortholog set marker) 등 유전자 기반 분자표지(gene-based marker)를 개발하기 위한 프라이머의 제작에는 불충분하다. 따라서 기존에 이용된 방법은 ClustalW(Thompson et al., 1994) 또는 Spidey(Wheelan et al., 2001)와 같은 공용 서열 정렬 프로그램을 이용하여 발현 유전자(expressed sequence tags, EST)나 mRNA 또는 유전자 암호화 서열(coding sequence, CDS)을 유전자에 정렬하여 유전자의 엑손-인트론 영역을 먼저 결정하고 이에 관한 정보를 기록한 후 Primer3에서 프라이머를 작성해야 했다. 이와 같이 유전자의 구조를 고려하여 프라이머를 디자인하기 위해서는 많은 노력과 시간이 소모되기 때문에 상기 과정을 일원화하여 단일 프로그램으로 처리할 수 있는 그래픽 프로그램이 개발되어 왔다.

Rutherford et al.(2000)은 유전자 서열을 시각화하여 염기서열과 아미노산을 함께 볼 수 있으며 단일 유전자부터 전체 유전체 범위까지 유전자의 구조 분석과 프라이머 제작을 동시에 수행할 수 있는 프로그램을 제작하였다. 그러나 이 프로그램은 미생물과 같은 원핵 생물의 유전체 해석에 적합하다고 보고하였다. 염기서열 정보를 시각적으로 정렬하고 유전자의 구조를 해석하는 프로그램인 GATA를 Nix and Eisen(2005)이 개발하였으나, 이 프로그램은 유전체

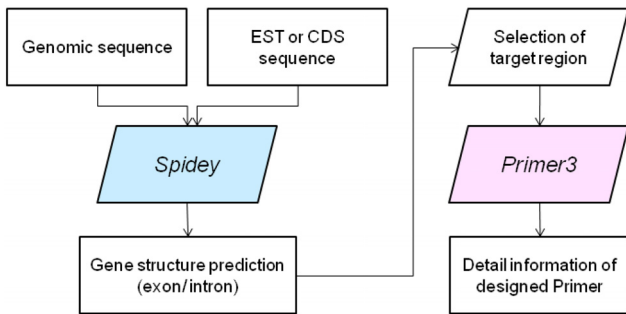
비교에 중점을 둔 것으로서, 분자표지 개발을 위한 프라이머 제작에는 충분치 않았다. 한편, Loraine and Helt(2002)는 사람 유전체에 대한 유전자 주석달기(annotation)가 가능한 그래픽 프로그램을 보고하였다. 식물을 대상으로 유전자의 구조를 분석하고 프라이머를 디자인할 수 있는 프로그램과 데이터베이스 또한 개발된 바 있으나, 이들은 중간 특이적인 인트론 길이 다형성 분자표지 개발에 중점을 둔 것으로서(Fredslund et al., 2006; Yang et al., 2007) 유전자 클로닝 또는 기능 해석을 위한 프라이머 디자인에는 적합하지 않다.

따라서 본 연구에서는 공용 프로그램인 Spidey와 Primer3의 소스 코드를 이용하여 식물 유전자의 엑손과 인트론 구역을 예측하고 작업 창에서 유전자 영역을 끌어오므로 쉽고 간편하게 유전자의 특정 구역에서 프라이머를 제작할 수 있는 일원화된 그래픽 프로그램을 개발하고자 하였다. 새로 개발된 프로그램에서는 전체 유전체 염기서열과 발현 유전자 서열 또는 유전자 암호화 서열을 입력하여 엑손과 인트론으로 이루어지는 유전자의 구조를 인실리코(*in silico*) 상에서 예측하고 2차원 그래픽으로 도식화시켰다. 이와 동시에, 엑손과 인트론의 염기서열 정보가 배열되어 동일한 작업 창에서 볼 수 있도록 하고, 증폭시키고자 하는 유전자의 구역을 마우스로 쉽게 지정할 수 있게 하였다. 또한, 유전자 상의 위치, 프라이머 길이, 용융점(melting temperature:  $T_m$ ), GC 함량과 같은 제작할 프라이머의 상세 정보를 염기서열이 배열된 동일한 창에서 입력하고 확인할 수 있게 하였다. 한편 최근 해독 완료된 배추 유전체 서열을 대상으로 프라이머를 제작하고 PCR과 전기영동 실험을 수행하여 이 프로그램의 실효성을 검증하였다.

## 재료 및 방법

### 프로그램의 개발

프로그램 개발의 기본 전략은 Fig. 1과 같다. C# 프로그램을 이용하여 공용 소프트웨어인 Spidey와 Primer3의 소스 코드를 결합시켜 사용자 친화적인 그래픽 프로그램(Graphic User Interface, GUI) Pickprimer를 완성하였다. 전체 염기서열, 발현 유전자 서열, 유전자 구조, 엑손-인트론 서열, 그리고 프라이머 디자인 영역 등 5구역이 작업 창에서 함께 보일 수 있도록 화면을 설정하여 컴퓨터 화면에서 프라이머를 디자인할 부분을 쉽게 확인할 수 있게 하였다. 염기서열은 데이터베이스나 웹에서 제공되는 정보를 직접 이용할 수 있도록 서열을 복사하여 붙여 넣거나 fasta 또는 text 파일을 불러올 수 있게 하였다. 이때 NCBI 등 공공 데이터베이스 서버로부터 서열을 복사하여 붙여 넣게 되면 숫자 등의 불필요한



**Fig. 1.** The Pickprimer program development strategy. The rectangles indicate input or output values and the parallelograms denote source programs for nucleotide sequence analysis and primer design. The two open source programs, Spidey and Primer3, are fundamental modules of the Pickprimer program. The input format of nucleotide sequences can be a fasta or a text file.

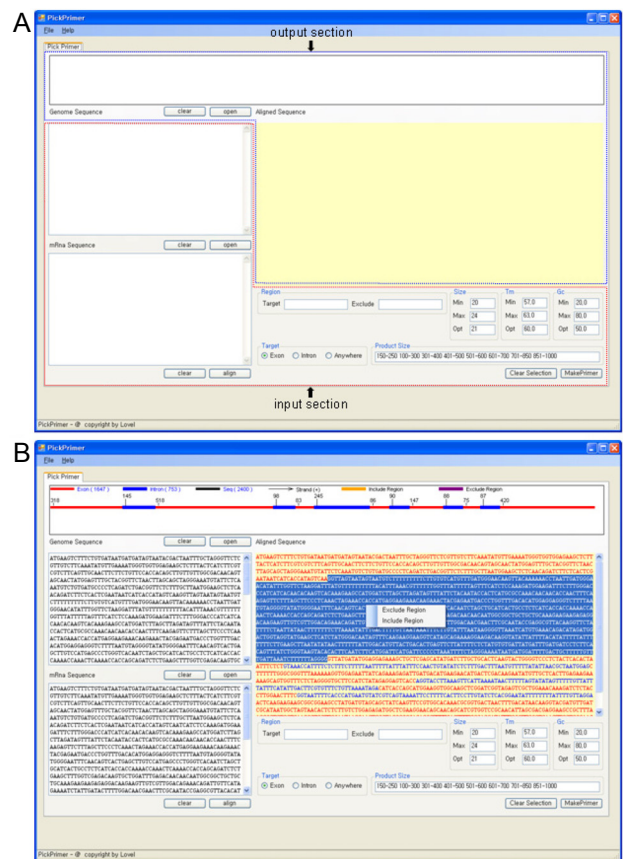
데이터가 입력되는데 이러한 값들을 자동으로 제거하게 함으로써 사용자의 수고를 덜어줄 수 있도록 고안하였다. 입력된 염기 서열은 Spidey 모듈에 의해 전체 염기서열의 엑손 구간이 예측되며, 이렇게 얻어진 엑손-인트론 영역과 서열을 사용자가 쉽게 알아볼 수 있도록 C#의 그래픽 라이브러리(Graphic Library)를 이용하여 그래픽으로 구현하였다. 또한, Primer3을 이용한 프라이머 제작 시 필요한 조건인 프라이머 길이, Tm 값, GC 함량 등을 실험 목적에 맞게 입력하도록 작업창의 일부분에 설정하였다. Pickprimer 프로그램은 Windows XP 이상의 운영체제와 Microsoft .Net Framework 버전 2.0 이상에서 작동하며, 설치 프로그램의 다운로드와 설치 방법, 그리고 사용설명서는 [www.brassica-rapa.org/Pickprimer](http://www.brassica-rapa.org/Pickprimer)에서 볼 수 있다.

### 프라이머 제작 및 PCR 실험

VC1, SR5, Super CR, BP079, IT033820, Kenshin 등 6품종의 배추를 파종하여 한 달간 온실에서 재배하였다. 품종별로 약 2.5g의 잎을 채취하여 액체 질소 하에서 동결 분쇄한 후 CTAB 방법(Doyle and Doyle, 1990)으로 genomic DNA(gDNA)를 추출하였다. 약 50ng의 gDNA에 Pickprimer로 제작한 각각 5μM의 상류와 하류 프라이머, 10× PCR buffer, 2.5mM MgCl<sub>2</sub>, 2.0mM의 dNTP 혼합액, 0.5U의 DNA polymerase로 만든 PCR 혼합액을 넣고 ddH<sub>2</sub>O로 최종 반응 부피를 20μL로 하였다. PCR은 프라이머의 조건에 따라서 Tm을 조절했으며 총 35 cycles로 실시하였다.

### 결 과

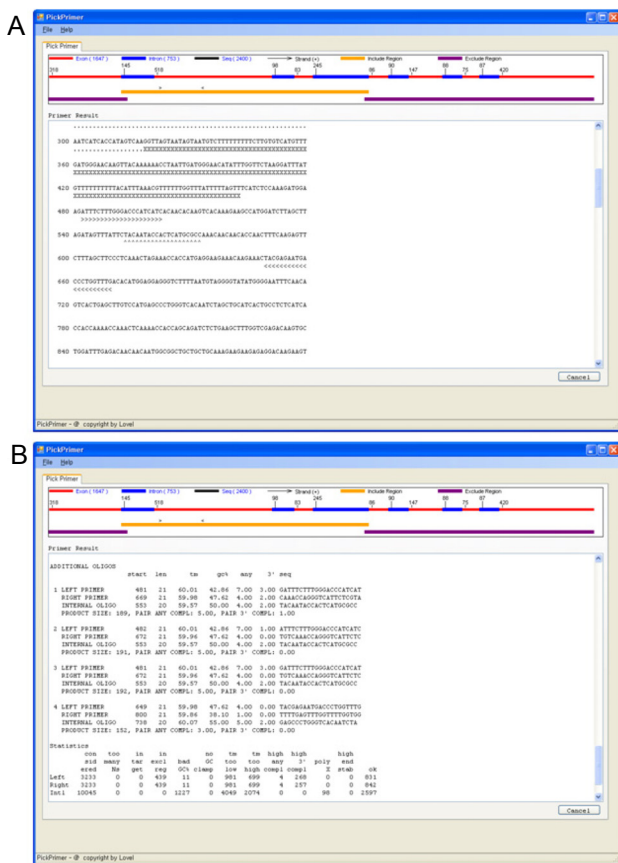
본 연구에서는 유전체 염기서열 정보(genome sequences)



**Fig. 2.** Screen shots of the Pickprimer used for primer design. A. Initial window of the Pickprimer consisting of five sections. Input sections (red dots) are genome sequence, mRNA sequence, and primer condition windows. Output sections (blue dots) are gene structure and sequence alignment. B. Working windows of the program. Genomic and mRNA sequences should be input in the designated sections and alignment can be performed by clicking the 'align' icon. Then the predicted gene structure is presented in the top section of the same window. The nucleotide sequences are also identified with red (exon) and blue colors (intron) in the 'Aligned Sequence' section. The values of primer length, Tm, GC contents, and product size can be modified in the bottom section according to the purpose of study. The target region for primer design can be easily specified either by dragging a mouse on the sequences or by entering position to include or exclude specific region in the 'Region' section. Primer design is performed by clicking the 'MakePrimer' icon.

와 발현 유전자 염기서열 또는 유전자 암호화 서열 정보를 바탕으로 유전자의 구조를 예측하고, 동시에 실험 목적에 따라 유전자의 엑손 또는 인트론 구간에서 프라이머를 디자인할 수 있는 사용자 친화적인 그래픽 프로그램 Pickprimer를 개발하였다.

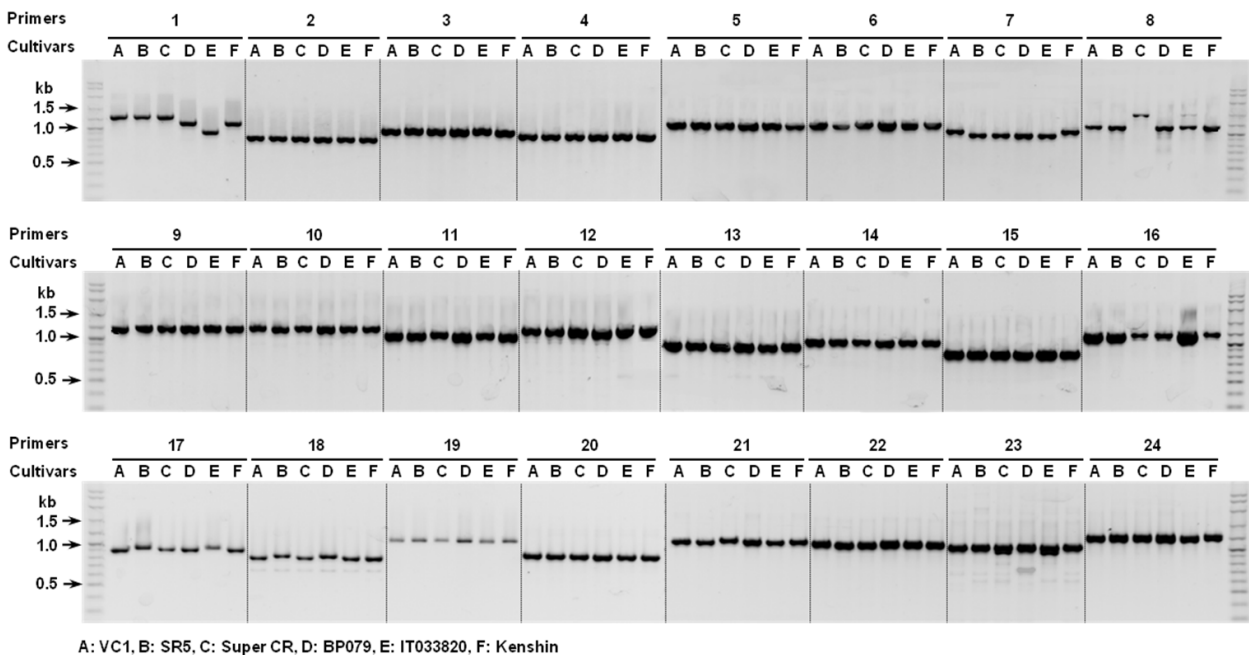
Pickprimer 프로그램의 특징은 원하는 부분의 genomic DNA와 mRNA sequence data를 입력하여 인트론과 엑손의 염기서열을 그래픽으로 확인하면서 목표구간(target region)에서 프라이머를 쉽게 제작할 수 있다는 것이다(Fig. 2). 본 프로그램은 WINDOWS XP 이상의 모든 Windows 기반 운영체



**Fig. 3.** Screen shots of primer design output from the Pickprimer program. The precise information including primer position, sequence, length, Tm, and GC contents of the designed primers are presented. A. Primer position on the target region. B. Detailed information of designed primers.

계에서 사용할 수 있고, 입력 정보는 text 또는 fasta 파일 형태로 이용 가능하며, NCBI에서 내려 받은 정보를 그대로 복사하여 붙일 수 있도록 고안하였다. 목적하는 유전자의 전체 염기 서열을 ‘Genome Sequence’ 창에 파일로 불러오거나 직접 입력하고, 발현 유전자 또는 유전자 암호화 서열을 동일한 방법으로 ‘mRNA Sequence’ 창에 입력한다(Fig. 2A). 입력된 전체 염기서열 정보와 발현 유전자를 정렬하면, Spidey 모듈을 통하여 유전자의 구조가 그림으로 표현되어 엑손과 인트론 구간을 확인할 수 있다(Fig. 2B). 이 때 엑손과 인트론 서열은 각각 붉은색과 파란색으로 표시하여 상세 염기서열 정보가 쉽게 구분되게 하였다. 증폭하고자 하는 유전자의 영역을 마우스로 끌어 지정한 후, 프라이머 길이, Tm 값, GC 함량, 엑손 또는 인트론의 포함 여부, 제외하거나 포함해야 하는 염기서열 구간 등의 조건을 실험 목적에 맞게 조정하여 프라이머를 디자인한다. Fig. 3에서는 이상의 방법을 통해 제작한 프라이머 쌍들에 대한 유전자 상에서의 위치, 길이 및 염기서열, Tm, GC 함량, 예상되는 PCR 증폭 산물의 크기 등 상세정보를 보여준다.

이렇게 개발된 Pickprimer 프로그램으로 총 24쌍의 프라이머를 제작하여 배추를 대상으로 그 실효성을 검증하였다(Fig. 4). 배추 유전체에서 단일 카피로 존재하는 24개 유전자를 대상으로 상류와 하류 프라이머가 각각 엑손 영역에 위치하도록 하였고, PCR 증폭 구간에는 인트론이 하나 이상 포함되도록 하였다. 이렇게 디자인된 프라이머는 증폭



A: VC1, B: SR5, C: Super CR, D: BP079, E: IT033820, F: Kenshin

**Fig. 4.** PCR evaluation of the designed primers by PickPrimer. Twenty four primer sets were used to amplify PCR products from six *B. rapa* inbred lines (A: VC1, B: SR5, C: Super CR, D: BP079, E: IT033820, F: Kenshin). PCR products were separated on 1% agarose gel.

산물의 길이가 500bp-1.2kb가 되도록 하였으며, 배추 6개의 고정 계통(VC1, SR5, Super CR, BP079, IT033820, Kenshin)의 gDNA를 주형으로 PCR을 수행하였다. 그 결과 24쌍의 프라이머 모두에서 명확한 단일 밴드의 PCR 증폭 산물을 얻을 수 있었다(Fig. 4).

## 고 찰

본 연구에서는 유전자의 구조 정보를 추출하고, 이를 이용하여 원하는 유전자 영역에서 프라이머를 손쉽게 개발할 수 있는 사용자 친화적 그래픽 프로그램을 개발하였다. 기존의 보고된 프로그램들을 이용하여 엑손과 인트론 구역을 고려하여 프라이머를 제작하는 경우, 유전자 구조 예측과 프라이머 디자인 프로그램 등 최소한 두 가지 프로그램을 각각 이용해야 하고 출력된 결과를 수작업으로 통합시켜야 하기 때문에 시간이 오래 걸릴 뿐만 아니라 실험 오류가 발생할 확률도 높았다. 그러나 본 연구에서 개발한 Pickprimer 프로그램은 이러한 이원적 체계를 단일화함으로써 유전자 염기서열과 발현유전자 염기서열만 입력하면 목적하는 유전자의 엑손과 인트론 구역을 도식화 해줌과 동시에 프라이머를 제작할 구간을 쉽게 지정할 수 있게 해준다. 특히 유전체 해독이 진행 중이거나 완료된 작물의 경우 서열 정보를 한 번 입력함으로써 유전자 구조 예측과 프라이머 디자인을 한 인터페이스 상에서 처리할 수 있어 프라이머 개발을 위한 시간과 노력을 크게 줄일 수 있을 뿐만 아니라 예측된 유전자 구조에 대한 결과와 프라이머 디자인 결과를 fasta 또는 text 파일로 저장할 수 있기 때문에 결과를 편집하여 다른 프로그램에서 이용하기도 매우 편리하다. 또한 Pickprimer 프로그램을 이용하여 개발된 프라이머는 유전자 구간을 목표로 할 수 있기 때문에 근연종들에서도 효과적으로 적용시킬 수 있어서 범용 분자표지로 개발할 수 있는 가능성이 매우 크다.

기존의 그래픽에 중점을 둔 프로그램은 사람(Loraine and Helt, 2002)이나 원핵생물(Rutherford et al., 2000), 또는 고리형 DNA(Stothard and Wishart, 2005) 등의 염기 서열을 해석하고 이들의 유전자 구조와 염기서열들을 서로 비교 분석하기 위한 것이 대부분이었다. 반면 본 연구에서 개발된 프로그램은 분자표지 개발뿐만 아니라 유전자를 이용한 다양한 실험에도 매우 유용하게 활용될 것으로 기대된다. 예를 들어 유전자 발현 검정을 위해서는 노던 분석(Northern hybridization analysis)이나, 정량 RT-PCR(reverse-transcription polymerase chain reaction)을 주로 수행하는데 이때 필요한 탐침 또는 유전자의 특정 영역을 목표로 하는 프라이머를 제작해야 하

는 경우 편리하게 구역을 설정할 수 있다. 유전자의 기능 분석에 필요한 형질 전환체 작성의 경우, 유전자 벡터 제작 단계에서 반드시 확인해야 하는 유전자의 방향성을 검정할 수 있어 매우 편리하게 이용될 수 있다. 또한, 유전체 연구의 경우, 발현유전자를 이용한 분자표지나 COS 분자표지, 또는 본 실험에서 활용한 인트론 길이 다형성 분자표지의 개발 등에도 유용하게 이용할 수 있어 그 활용도가 매우 높다.

## 초 록

유전 육종 연구를 위해 연구자들은 실험 목적에 따라 다양한 종류의 프라이머를 제작해야 한다. 인터넷 상에서 다양한 공용 프로그램이 이용되고 있으나 많은 경우 사용자 편의성이 낮기 때문에 유전자의 구조를 고려하여 프라이머를 디자인하기 위해서는 시간과 노력이 소요된다. 본 연구에서는 엑손과 인트론 지역을 시각적으로 구별하면서 손쉽게 프라이머를 제작할 수 있는 프로그램인 Pickprimer를 개발하였다. 이 프로그램은 공용 프로그램인 Spidey와 Primer3 프로그램의 소스 코드를 결합한 후 그래픽 인터페이스를 추가하여 사용자가 유전자의 구조를 예측하고 이를 바탕으로 프라이머를 손쉽게 제작할 수 있게 했다. 입력 정보는 공용 데이터베이스에서 내려 받은 서열을 복사-붙임하여 이용할 수 있게 하였으며, 유전자의 구조를 그림으로 표현하고 동시에 엑손과 인트론 서열을 구별할 수 있게 했다. 이 프로그램을 이용하여 배추의 단일 카피 유전자에 대한 24 쌍의 프라이머를 디자인하고 6개 고정 품종을 대상으로 PCR과 전기영동 실험을 수행한 결과 제작한 모든 프라이머 쌍이 명확한 단일 밴드를 성공적으로 증폭시켰다. 이 프로그램은 분자표지의 개발뿐만 아니라 유전자 기능 연구 등 다양한 종류의 유전 육종 실험에 유용하게 이용될 수 있을 것으로 기대된다.

**추가 주요어 :** 유전자 구조 예측, 유전 육종, 분자표지, 핵산 중합연쇄반응, 목표구간 설정

## 인용문헌

- Darling, A.C.E., B. Mau, F.R. Blattner, and N.T. Perna. 2004. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14:1394-1403.
- Doyle, J.J. and J.L. Doyle. 1990. Isolation of plant DNA from fresh tissue. *Focus* 12:13-15.
- Gans, J.D. and M. Wolinsky. 2007. Genomorama: Genome visualization and analysis. *BMC Bioinformatics* 8:204.
- Koressaar, T. and M. Remm. 2007. Enhancements and modifications

- of primer design program Primer3. *Bioinformatics* 23:1289-1291.
- Loraine, A.E. and G.A. Helt. 2002. Visualizing the genome: Techniques for presenting human genome data and annotations. *BMC Bioinformatics* 3:19.
- Nix, D.A. and M.B. Eisen, 2005. GATA: A graphic alignment tool for comparative sequence analysis. *BMC Bioinformatics* 6:9.
- Rutherford K., J. Parkhill, J. Crook, T. Horsnell, P. Rice, M.-A. Rajandream, and B. Barrell. 2000. Artemis: Sequence visualization and annotation. *Bioinformatics* 16:944-945.
- Stothard P. and D.S. Wishart. 2005. Circular genome visualization and exploration using CGView. *Bioinformatics* 21:537-539.
- Thompson, J.D., D.G. Higgins, and T.J. Gibson. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22: 4673-4680.
- Wheelan, S.J., D.M. Church, and J.M. Ostell. 2001. Spidey: A tool for mRNA-to-genomic alignments. *Genome Res.* 11: 1952-1957.
- Yang, L., J. Gulei, X. Zhao, Y. Zheng, Z. Xu, and W. Wu. 2007. PIP: A database of potential intron polymorphism markers. *Bioinformatics* 23:2174-2177.