



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2008-0088541
(43) 공개일자 2008년10월02일

(51) Int. Cl.

G06F 19/00 (2006.01)

(21) 출원번호 10-2008-0073693(분할)
(22) 출원일자 2008년07월28일
심사청구일자 2008년07월28일
(62) 원출원 특허 10-2006-0095841
원출원일자 2006년09월29일
심사청구일자 2006년09월29일

(71) 출원인

고려대학교 산학협력단

서울 성북구 안암동5가1 고려대학교 내

(72) 발명자

이재원

서울특별시 성북구 안암동5가 고려대학교 정경대학

이정복

경기도 안산시 단원구 고잔1동 516

(뒷면에 계속)

(74) 대리인

현종철

전체 청구항 수 : 총 2 항

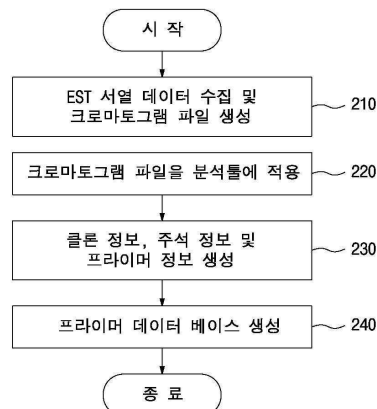
(54) 실험용 프라이머 정보를 제공하는 데이터베이스 구축 방법

(57) 요약

본 발명은 유전체 데이터베이스에 관한 것으로, 특히 실험용 프라이머 정보를 제공하는 데이터베이스 구축 방법에 관한 것이다.

본 명세서에서 개시하는 실험용 프라이머 정보를 제공하는 데이터베이스 구축 방법은 데이터 입력장치를 통해 수신되는 입력신호에 따라 데이터베이스를 검색하여 상기 데이터베이스로부터 검색된 정보를 출력장치를 통해 디스플레이하는 컴퓨터 시스템에서 실험용 프라이머 정보를 제공하는 데이터베이스를 구축하는 방안을 제공한다. 이를 위해 본 발명은 분석툴에 의해 크로마토그램 파일로부터 분석결과를 생성하는 단계; 주석생성툴에 의해 상기 분석결과로부터 주석 정보를 생성하고, 이에스티 클러스터링 툴을 통해 상기 분석결과로부터 프라이머 정보를 생성하는 단계; 및 상기 주석 정보, 상기 프라이머 정보, 및 상기 분석결과로부터 추출되는 클론 정보의 조합에 의해 프라이머 데이터베이스를 생성하는 단계를 포함하여 본 발명의 과제를 해결한다.

대 표 도 - 도2



(72) 발명자

손인석

서울특별시 성북구 안암동5가 고려대학교 정경대학

김철민

부산광역시 서구 아미동1가 부산대학교 의과대학

특허청구의 범위

청구항 1

데이터 입력장치를 통해 수신되는 입력신호에 따라 데이터베이스를 검색하여 상기 데이터베이스로부터 검색된 정보를 출력장치를 통해 디스플레이하는 컴퓨터 시스템에서 실험용 프라이머 정보를 제공하는 데이터베이스를 구축하는 방법에 있어서,

분석툴에 의해 크로마토그램 파일로부터 분석결과를 생성하는 단계;

주석생성툴에 의해 상기 분석결과로부터 주석 정보를 생성하고, 이에스티 클러스터링 툴을 통해 상기 분석결과로부터 프라이머 정보를 생성하는 단계; 및

상기 주석 정보, 상기 프라이머 정보, 및 상기 분석결과로부터 추출되는 클론 정보의 조합에 의해 프라이머 데이터베이스를 생성하는 단계를 포함하는 실험용 프라이머 정보를 제공하는 데이터베이스 구축 방법.

청구항 2

제 1 항에 있어서, 상기 프라이머 정보는

실험용 프라이머 설계를 위한 각종의 파라미터 정보를 포함하는 것을 특징으로 하는 실험용 프라이머 정보를 제공하는 데이터베이스 구축 방법.

명세서

발명의 상세한 설명

기술 분야

- <1> 본 발명은 유전체 데이터베이스에 관한 것으로, 특히 실험용 프라이머 정보를 제공하는 데이터베이스 구축 방법에 관한 것이다.

배경 기술

- <2> 도 1에서와 같이, 이에스티(Expressed Sequence Tag, EST)는 발현된 유전자 단편을 가리키며, 유전자의 발현에 관련된 cDNA(complementary DNA)의 일부를 의미한다. 또한, EST는 불특정 서열의 일부가 각 세포에서 발현되는 것으로 전장(full sequence)이고, 유전자를 찾는데 유용하게 사용된다. EST는 각 세포에서 발현되는 30억에 달하는 염기쌍 서열의 일부로서 여러 인체조직과 다른 미생물의 유전자를 비교하는 연구를 통하여 유전자의 기능을 해석할 수 있게 해준다. 즉, cDNA(complementary DNA)를 클론화하여 염기 서열화한 EST(Expressed Sequence Tag) 데이터는 여러 생물체들의 염기서열 정보들과 비교를 통해 유사점을 찾거나 기능적 부위 검색을 통해 유전자 기능을 추정할 수 있어 기능 유전체 연구에 기여를 하고 있다.
- <3> EST는 도 1과 같이 cDNA 클론 분석을 통해 얻어진 짧은 서열이다. cDNA는 mRNA를 역전사에 의해 이중나선 구조의 DNA로 다시 만든 DNA로, 세포안의 mRNA가 단백질로 발현되어지는 유전자를 나타내기 때문에 cDNA 역시 세포 안에서 발현되는 유전자의 서열을 나타낸다. EST는 cDNA 각 클론의 염기서열을 3' 또는 5' 끝에서 단 한차례 자동화된 기계로 읽어서 서열을 결정된 것으로 실험으로 검증되지 않아 그 염기서열은 다소 부정확할 수 있다. 하지만, 서열을 자동화된 기계로 빠르고 값싸게 얻을 수 있다는 점에서 차차 여러 가지 연구에 응용되기 시작하면서 각광을 받기 시작하였다.
- <4> EST는 새로운 유전자의 발견, 질병에 관련된 새로운 유전자 집단의 발견, 유전자 발현의 정성 및 정량적 분석, 조직간에 차등적으로 발현되는 유전자의 클로닝, EST의 3'UTR (Untranslated region)을 새로운 STS(Sequence tagged site)로 활용한 유전자 발현 지도의 작성, 유전체 서열과의 비교를 통한 복잡한 유전체 서열의 분석 등 다양한 연구 분야의 중요한 재료로써 이용되고 있다.
- <5> EST 분석은 도 1에서와 같이, 어떤 유전자 부위(EST)가 어떤 기능으로 발현되는가를 분석하는 것이고, EST 분석 기술은 DNA 칩 데이터 분석, SNP 분석, Motif 분석, Proteomics, 신약 후보물질 탐색등에 폭넓게 적용할 수 있다. EST 분석은 제약점이 있음에도 불구하고 생명공학 연구 여러분야에서 아주 중요하게 널리 사용되고 있다.
- <6> 그러나, 종래의 유전체 데이터베이스는 프라이머 정보를 검색함에 있어 지나치게 많은 조작을 요구하고, 검색결

과를 사용자가 용이하게 파악하기 어렵고, 실험을 검증하기 위해 필요한 프라이머 정보를 체계적으로 제공할 수 없는 문제점이 있다.

발명의 내용

해결 하고자하는 과제

- <7> 본 발명이 해결하고자 하는 과제는 이에스티의 프로브셋을 검색함에 있어 사용자의 조작을 최소화하여 편리성을 제공하고, 검색결과를 사용자가 용이하게 파악하기 쉽게 테이블 및 이미지로 검색결과를 디스플레이할 수 있으며, 실험을 검증하기 위해 필요한 프라이머 등의 정보를 체계적으로 제공할 수 있는 실험용 프라이머 정보를 제공하는 데이터베이스 구축 방법을 제공하는 것이다.

과제 해결수단

- <8> 상기와 같은 과제를 해결하기 위해 본 명세서에서 개시하는 실험용 프라이머 정보를 제공하는 데이터베이스 구축 방법은 데이터 입력장치를 통해 수신되는 입력신호에 따라 데이터베이스를 검색하여 상기 데이터베이스로부터 검색된 정보를 출력장치를 통해 디스플레이하는 컴퓨터 시스템에서 실험용 프라이머 정보를 제공하는 데이터베이스를 구축하는 방안을 제공한다. 이를 위해 본 발명은 분석틀에 의해 크로마토그램 파일로부터 분석결과를 생성하는 단계; 주석생성틀에 의해 상기 분석결과로부터 주석 정보를 생성하고, 이에스티 클러스터링 툴을 통해 상기 분석결과로부터 프라이머 정보를 생성하는 단계; 및 상기 주석 정보, 상기 프라이머 정보, 및 상기 분석결과로부터 추출되는 클론 정보의 조합에 의해 프라이머 데이터베이스를 생성하는 단계를 포함하여 본 발명의 과제를 해결한다.

효 과

- <9> 본 발명에 의하면, 실험 조건, 기능별 카테고리 등에 따라 DNA 칩 상의 유전자 정보를 분류하여 데이터베이스를 구축하고, 간단하고 직관적인 인터페이스를 이용하여 검색을 하도록 함으로써, 프라이머 정보를 검색함에 있어 조작정도를 최소화하여 편리성을 제공하고, 검색결과를 사용자가 용이하게 파악하기 쉽게 테이블 및 이미지로 검색결과를 디스플레이할 수 있고, 실험을 검증하기 위해 필요한 프라이머 정보를 체계적으로 제공할 수 있는 효과가 있다.

발명의 실시를 위한 구체적인 내용

- <10> 유전체 데이터는 복잡하고 그 양과 자료의 범위가 넓다. 따라서 유전체 데이터베이스는 그 각각의 자료 타입과 값들을 유연하게 처리해야 한다. 유전체 자료는 특성상 빠르게 변화하므로 데이터베이스의 빠른 스키마의 변경을 지원할 수 있도록 해야한다. 유전체 데이터베이스는 어느 특정한 분야가 아닌 의학, 약학, 화학, 생물학 등의 다양한 분야와 관계가 있는 만큼 그 각각의 분야에서 원하는 자료의 표현 방법들이 틀리다. 따라서 비록 하나의 유전체 데이터베이스를 사용하더라도 그들 각각의 요구사항들을 충족시킬 수 있는 다양한 스키마를 제공해야 한다.
- <11> 유전체 데이터베이스 시스템을 구축하는데 있어서 해결해야 할 제약 사항들은 아래와 같다.
- <12> 첫째, 특정 순서에 대한 상이한 관점들이 적절히 표현 가능해야 한다. 둘째, 개념들이 항상 변하고 유동적이므로 유동성을 지원할 수 있는 환경이 필요하다. 셋째, 자료가 계속적으로 변화함으로 자료의 불일치와 모호성이 야기됨으로써 구축과정 중에 동적인 스키마를 지원해야 한다. 즉, 개념과 스키마가 계속적으로 변하기 때문에 강력한 질의어가 필요하며, 염색체의 적절한 순서와 간격 등을 식별할 수 있는 강력한 패턴 매칭 기술 등도 함께 제공해야 한다.
- <13> 이하, 본 발명이 해결하고자 하는 과제의 해결 방안을 명확하게 하기 위한 발명의 구성을 본 발명의 바람직한 실시예에 근거하여 첨부 도면을 참조하여 상세히 설명한다.
- <14> 도 2는 본 발명에 따른 실험용 프라이머 정보를 제공하는 데이터베이스 구축 방법의 흐름도이다.
- <15> 먼저, EST 서열 데이터(EST sequence data)를 수집하고, EST 서열 데이터를 이용하여 크로마토그램 파일을 생성한다(210 과정). 이때, EST 서열 데이터는 농업과학기술원(NIAST) 등으로 부터 수집될 수 있다.
- <16> 다음, 크로마토그램 파일을 분석틀에 적용하여 분석결과를 생성한다(220 과정). 이때, 분석틀로 Finch-Suite를

사용할 수 있다.

- <17> 분석결과가 생성되면, 분석결과로부터 클론 정보를 추출하고, 분석결과를 주석생성물에 적용하여 주석 정보를 생성하며, 분석결과를 EST 클러스터링 툴에 적용하여 프라이머 정보를 생성한다(230 과정). 이때, 주석생성물은 Pedant-Pro를 사용하고, 클러스터링 툴은 StackPack을 사용할 수 있다. 또한 프라이머 정보를 생성하기 위한 툴로써 EST 클러스터 정보를 이용하는 Prime을 사용할 수 있다.
- <18> 마지막으로, 클론 정보, 주석 정보 및 프라이머 정보를 조합하여 프라이머 데이터베이스를 생성한다(240 과정). 바람직하게는, 프라이머 데이터베이스를 복수의 클라이언트 컴퓨터와 네트워크로 연결된 서버용 컴퓨터에 저장하고, 클라이언트 컴퓨터에서 키워드를 입력하면 입력된 키워드와 관련된 정보를 프라이머 데이터베이스로부터 추출하여 클라이언트 컴퓨터로 출력하도록 하는 시스템을 설계할 수 있다. 바람직하게는, 이 과정(240 과정)은 프라이머를 이용하여 실험적으로 증폭하고 염기서열을 분석 및 확인한 후 프라이머의 정보를 데이터베이스에 기록하는 과정을 포함할 수 있다.
- <19> 이때, 프라이머 정보는 실험용 프라이머 설계를 위한 각종의 파라미터 정보를 포함할 수 있다.
- <20> 도 3a는 상용 소프트웨어를 이용하여 도 2의 데이터베이스를 구축하는 과정의 흐름도이다.
- <21> Finch Suite와 같은 Finch Genome Analysis System은 염기서열 품질 분석 및 유전체 분석 시스템으로서, On line이 가능하다. Pedant Pro는 유전체 기능 정보 예측 시스템으로서, 역시 On line이 가능하다. StackPACK은 EST를 분석하기 위한 툴이다.
- <22> 도 3a에서, 크로마토그램은 크로마토그래피에서 시간에 따라 분리되어 나오는 성분들의 용리(熔離) 곡선 또는 그 가운데 색을 띤 반점을 의미한다.
- <23> 도 3b는 도 3a의 데이터베이스를 적용하여 프로브셋을 검색하기 위한 소프트웨어인 큐 프로브의 화면 구성을 도시한 것이다.
- <24> 큐 프로브(Q-Probe)의 제작을 위해서, GCG Wisconsin 패키지의 "Prime+"과 유닉스 스크립트를 이용한 자동화 모듈을 분석 개발하고, 베타 EST 주석 데이터베이스와 연계 인터페이스를 개발한다.
- <25> 또한, GCG 패키지의 "Prime"과 유닉스 스크립트를 이용한 자동화 모듈을 분석한다. EST clone 에 대해서는 다음의 설계 조건으로 분석을 수행한다.
- <26> 즉, 설계 파라미터(Design Parameters)는 Product MIN size=100 sets the minimum products size (bp), Product MAX size=500 sets the maximum products size (bp), DNA concentration=50.0 sets the primer DNA concentration (nM), Salt concentration=50.0 sets the salt concentration (mM), Clamp=s specifies primer 3' clamp (using IUB ambiguity codes), GC MIN primer=40.0 sets the minimum primer % G+C, GC MAX primer=55.0 sets the maximum primer % G+C, TM MIN primer=50.0 sets the minimum primer melting temperature (Celsius), TM MAX primer=65.0 sets the maximum primer melting temperature (Celsius) 으로 설정할 수 있다.
- <27> 큐 프로브는 DNA 칩 분석에서 얻어진 결과의 실험적 증명을 위한 Q-RT-PCR용 프라이머(primer) 제공한다. 데이터베이스화된 프라이머(primer) 셋들은 실제 실험적으로 증폭 및 염기서열을 분석하여 확인하고 데이터베이스에 기록되어 실제 연구에 이용될 수 있는 프라이머(primer) 셋을 제공한다.
- <28> 도 3b의 큐 프로브는 rice EST primary consensus contig에 대해서 분석 수행한 결과를 포함한다. 콘티그(contig)는 DNA의 세그먼트를 완전히 구성하는 일련의 클론들의 위치를 나타내는 맵을 의미하기도 하고, 개별적인 서열 정보들이 통합되어 더 큰 DNA의 서열을 나타낼 때 이와같이 조합된 서열을 의미하기도 한다.
- <29> 도 3b의 검색창에 키워드를 입력하거나 버튼을 클릭하면 도 9c와 같은 화면이 출력된다.
- <30> 도 3c는 도 3b에서 검색창에 키워드를 입력한 후의 화면이다.
- <31> 도 3c는 EST 주석에 대한 검색결과로서, "Cluster ID" 또는 "Clone ID"를 클릭하면 해당 상세정보를 볼 수 있다. 도 9c에서, 각 항목의 상세 정보의 일례는 다음과 같다.
- <32> 즉, 클론 및 서열 정보의 최적 검색 결과(Best Hit of Clone and Sequence Information), rice EST 주석 정보(rice EST annotation Information), 프라이머 설계의 상세 정보(Detail View of primer Design)이다. 각각은 도 3d 내지 도 3f에 도시되어 있다.

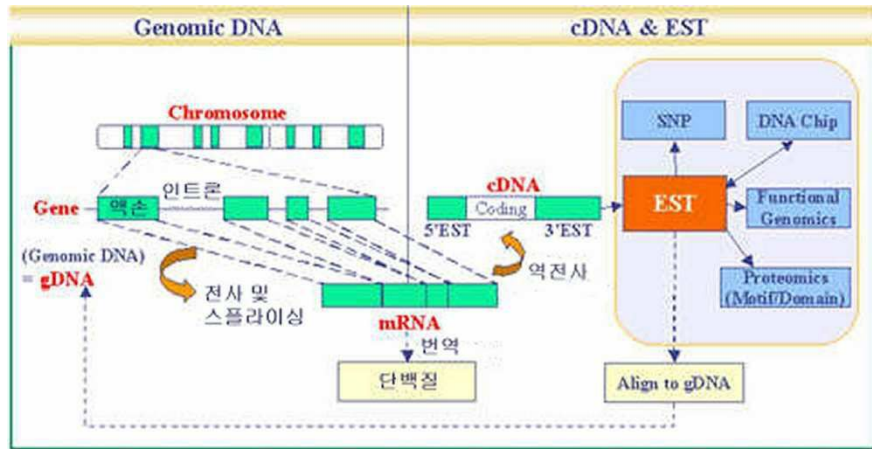
- <33> 도 3d는 도 3c에서 상세 정보를 클릭한 경우의 일 예를 보여주는 화면이다. 도 3e는 도 3d와 관련한 주석 정보의 일 예를 도시한 것이다. 도 3f는 도 3a에서의 검색과 관련하여 실제 실험에 사용될 프라이머 셋을 제공하는 화면이다.
- <34> 본 발명에 따르면, 복제 실험 및 염료 교환 실험의 신뢰도 검증 방법, 유효 유전자 검색 방법, EST 기능 검색 방법, 실험용 프라이머 정보를 제공하는 데이터베이스 구축 방법을 통합한 하나의 소프트웨어인 레드 뷰(REDVIEW)를 제작할 수 있다.
- <35> REDVIEW는 rice EST 분석 데이터와 벼 레퍼런스(rice reference) DNA 칩을 연동한 데이터베이스 및 뷰어 시스템이다. REDVIEW는 벼 레퍼런스(rice reference) DNA 칩 활용을 위한 데이터베이스 및 뷰어시스템이다. rice EST를 분석한 주석정보와 DNA 칩정보를 연계시켜 웹으로 테이블 및 이미지 형태로 연구자들에게 정보를 제공한다. REDVIEW는 벼 레퍼런스(rice reference) DNA 칩 이미지를 보여주는 부분, 실험 조건에 따른 유효한 유전자를 찾는 부분, 기능별 카테고리에 의한 유전자를 찾는 부분, DNA 칩 분석에서 얻어진 결과의 실험적 증명을 위한 Q-RT-PCR용 프라이머 세트를 찾는 부분으로 구성된다. REDVIEW는 rice EST를 분석한 주석정보와 DNA 칩정보를 연계시켜 웹으로 테이블 및 이미지 형태로 연구자들에게 정보를 제공한다.
- <36> REDVIEW를 이용하면, 수행된 연구 결과를 일반 연구자들에게 효율적으로 제공하고, 레퍼런스 칩의 활용도를 증대시킬 수 있고, 레퍼런스 칩 정보의 관계형 데이터베이스를 구축할 수 있으며, 칩 레이아웃(Chip layout)과 클론(clone) 정보를 포함한 레퍼런스 칩 데이터베이스를 구축할 수 있고, 주석 정보의 데이터베이스 구축 및 가상 레퍼런스 칩과 연동한 정보를 제공할 수 있다.
- <37> 또한, REDVIEW를 이용하면, 클러스터링 툴(Clustering tool)을 이용한 클론(clone)의 클러스터링(clustering) 구현 및 인터페이스를 구축할 수 있고, 프라이머(primer) 데이터베이스 구축 완료 및 각 데이터베이스의 관계형 스키마를 구현할 수 있으며, 레퍼런스 칩 레이아웃(layout)을 가시화할 수 있다.
- <38> 또한, REDVIEW를 이용하면, 반점 레이아웃 뷰어(Spot layout viewer)를 구현할 수 있고, 레이아웃 뷰어(layout viewer)와 주석 데이터베이스의 링크 및 인터페이스를 구현할 수 있으며, DNA 칩의 복제, 염료 교환의 신뢰성 판단을 위한 이미지 뷰어(image viewer) 모듈을 구현할 수 있다.
- <39> 또한, REDVIEW를 이용하면, 기능별 주석으로 부터 레이아웃을 확인할 수 있는 인터페이스를 제공하고, rice EST의 블래스트(BLAST) 검색을 위한 포맷 데이터베이스를 구축할 수 있으며, BLAST 검색 인터페이스 모듈을 구성할 수 있다.
- <40> 이제까지 본 발명에 대하여 그 바람직한 실시예를 중심으로 살펴보았다. 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자는 본 발명이 본 발명의 본질적인 특성에서 벗어나지 않는 범위에서 변형된 형태로 구현될 수 있음을 이해할 수 있을 것이다.
- <41> 그러므로 개시된 실시예들은 한정적인 관점이 아니라 설명적인 관점에서 고려되어야 한다. 본 발명의 범위는 전술한 설명이 아니라 특허청구범위에 나타나 있으며, 그와 균등한 범위 내에 있는 모든 차이점은 본 발명에 포함된 것으로 해석되어야 할 것이다.

도면의 간단한 설명

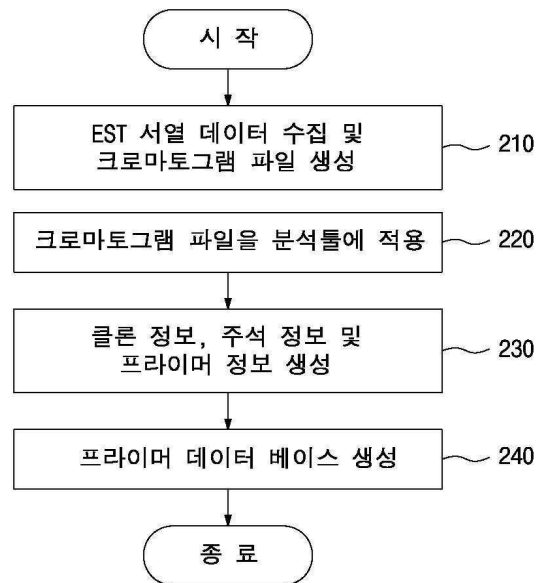
- <42> 도 1은 본 발명이 적용되는 이에스티의 개념과 활용을 도시한 것이다.
- <43> 도 2는 본 발명에 따른 실험용 프라이머 정보를 제공하는 데이터베이스 구축 방법의 흐름도이다.
- <44> 도 3a는 상용 소프트웨어를 이용하여 도 2의 데이터베이스를 구축하는 과정의 흐름도이다.
- <45> 도 3b는 도 3a의 데이터베이스를 적용하여 프로브셋을 검색하기 위한 쿼리 프로브의 화면 구성을 도시한 것이다.
- <46> 도 3c는 도 3b에서 검색창에 키워드를 입력한 후의 화면이다.
- <47> 도 3d는 도 3c에서 상세 정보를 클릭한 경우의 일 예를 보여주는 화면이다.
- <48> 도 3e는 도 3d와 관련한 주석 정보의 일 예를 도시한 것이다.
- <49> 도 3f는 도 3a에서의 검색과 관련하여 실제 실험에 사용될 프라이머 셋을 제공하는 화면이다.

도면

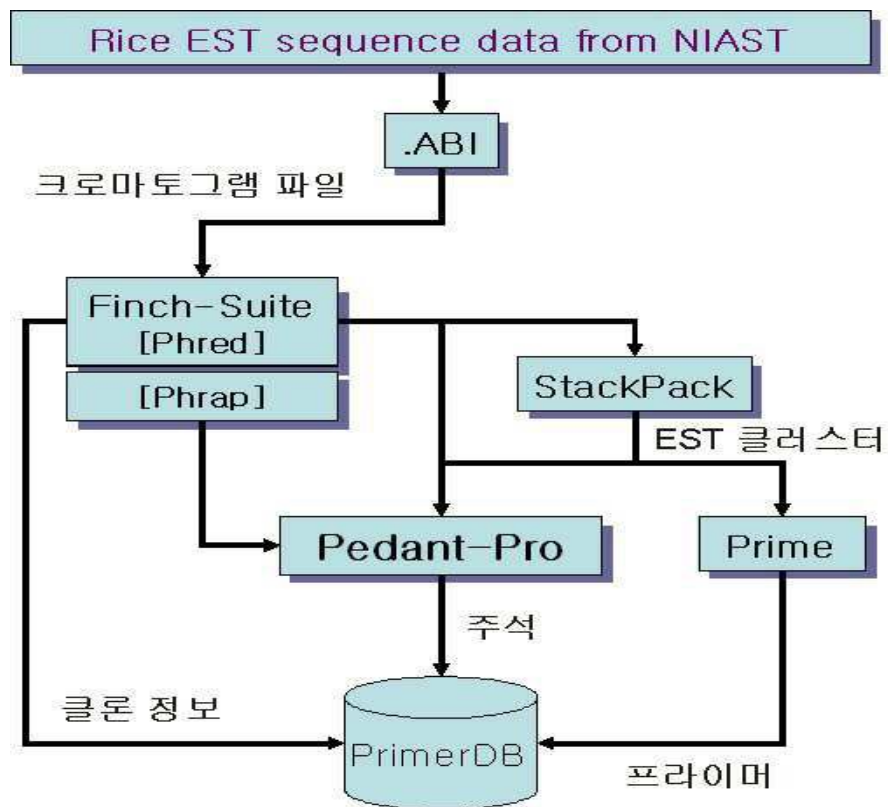
도면1



도면2



도면3a



도면3b

Search probe set of rice ESTs

Busan Genome Center

Usage:

For search all items, please empty and submit

For search K24P clones, enter 'k24p'.

도면3c

Search Results		
Total matched clusters : 780		
Total matched clones : 2839		
Cluster ID	Clone ID	Best Hit
cl1ct1cn1	ks15a11	Yet Annotated
	ks19b07	Yet Annotated
cl1ct2cn2	kc06h06	Yet Annotated
	kcl146a6sk	Yet Annotated
	kcl147h10sk	Yet Annotated
	kcl149b5sk	Yet Annotated
	kcl149g9sk	Yet Annotated
	kcl150d12sk	Yet Annotated
	kcl152d9sk	Yet Annotated
	kcl163a2sk	Yet Annotated
	ks11c01	Yet Annotated
	ks19c03	Yet Annotated
	120502ac02k24p0706	Glycine-rich RNA-binding protein
	120502bb12k24p0816	Glycine-rich RNA-binding protein GRP1A
	k24p16d0412130208	Glycine-rich RNA-binding protein GRP1A
cl1ct3cn3	kc01a01	Yet Annotated
	kc02g01	Yet Annotated
	kc03h01	Yet Annotated
	kc05d02	Yet Annotated
	kc10a06	Yet Annotated
	kcl142d3sk	Yet Annotated
	kcl145g6sk	Yet Annotated
	kcl147c4sk	Yet Annotated
	kcl150d5sk	Yet Annotated
	kcl150g7sk	Yet Annotated
	kcl154b4sk	Yet Annotated
	kcl159f8sk	Yet Annotated

도면3d

Detail View	
[Hit] [Sequence] [Annotations] [Transcript]	
Best Hit of Clone	
ID	120502ac02k24p0706
HIT ID	Q03879 GRP_DAUCA
DESCRIPTION	Glycine-rich RNA-binding protein
METHOD	BLASTX ref
START/STOP	165/323
E-VALUE	3.0E-16
Sequence Information	
//NA SEQUENCE 1.0 120502ac02k24p0706 Chromatid=59356 Length=325 Name=POLIMERID Id=419 120502ac02k24p0706.seq Length: 325 October 26, 2003 11:14 Type: M Check: 243 ...	
1	GTCGGGTCC GATCGATTTC CGGTTTTCG TTCGGTTCG GGGTGGTGG
51	GAAGCATGG GGGCGCGGA TGTGAGTAC CGTCTCTTCG TCGGGGCTC
101	CGCTGGGCG ACCGACGAC GCTGCTCGA AGCGGCTTC TCGAECTAG
151	CGATTTCCT TTAGTCCAG ATCATCAGG ATCGGAGAC GGGGAGGTG
201	CGCGGGTTCG GGTTCGTAC GTCTCGAGC CACCACTCA TCGGGAGCG
251	CATCGAGGG ATGACGGCA AGGAGCTCA CGGCGGAC ATCACTCTA
301	ACGAGGCCA GTGGGCGCG TCGG

도면3e

Annotation Information		
ID	120502ac02b240704 standard;	
XX		
AC	1	
XX		
DE	BLASTE ORF25	
XX		
XX	General properties	
XX	Length (aa)	52
XX	Molecular weight (kD)	5897.5
XX	Isoelectric point	9.9
XX	Manually edited	no
XX	Cloning name	14661_120502_ac_001_K249_04_001
XX	Position	265-323
XX	GC content (%)	66.6
XX	Start	3106EST_1206-1249_758
XX	Date	10-09-2003
XX	Nucleic code	Standard
XX	Other	W015009
XX	Source	Eukaryotic cDNA/EST
XX	Species	Rice
XX	Protein function	
XX	Cisactin homologue (BLASTP)	sp Q55870 ORF_P202A Glycine-rich RNA-binding protein, 3e-16
XX	Cisactin homologue (BLASTN)	emb14102266.1108A02040 Glycine active mRNA for glycine rich RNA binding protein [GPR 986] 1e-154
XX	Cisactin homologue (BLASTN)	sp P49110 ORF_23MAL Glycine-rich RNA-binding protein ORF2A, 2e-36
XX	Autoconsensually derived Functional	en00000000 95 CLASSIFICATION NOT YET CLEAR-CUT 1e-17 - Aukelins
XX		30.26 chloroplast organization 3e-15 - Aukelins
XX		30.03 organization of cytoplasm 5e-19 - S.eucristace
XX		30.10 nonsex organization 3e-13 - S.eucristace
XX		04.05.09 other mRNA-transcription activators 5e-19 - S.eucristace
XX		04.05.01.04 transcriptional control 1e-19 - Aukelins
XX		03.01 cell growth 1e-14 - S.eucristace
XX		04.02 on-set transcription activities 3e-14 - S.eucristace
XX	COG2	COG074 RNA-binding proteins (RBM domain) 4e-16
XX	PFAM domain	PF00076 RNA recognition motif, (n.k.s. RBM, S.CO006
XX	BLASTS	PF000004 Poly-adenylation binding protein, unbind domain
XX	Protein structure	
XX	Known3D	peb1MAL1MAL hamp selfaggregat: hamp w1 (mb1, abd1):[heterogeneous nuclear ribonucleoprotein m1, hamp w1 1-150] peb120C120C1-A no antigen epitope: f22p RNA-binding domain [cd1b] peb127E127E-A paramorphastic encephalomyelitis antigen selfaggregat: z-terminal two rrm-domains (f2d, zc-antigen) peb102E102E-A paramorphastic encephalomyelitis antigen selfaggregat: z-terminal two rrm-domains (z, p1p2p3p4p5p6p7p8p9p10p11p12p13p14p15p16p17p18p19p20p21p22p23p24p25p26p27p28p29p30p31p32p33p34p35p36p37p38p39p40p41p42p43p44p45p46p47p48p49p50p51p52p53p54p55p56p57p58p59p60p61p62p63p64p65p66p67p68p69p70p71p72p73p74p75p76p77p78p79p80p81p82p83p84p85p86p87p88p89p90p91p92p93p94p95p96p97p98p99p100p101p102p103p104p105p106p107p108p109p110p111p112p113p114p115p116p117p118p119p120p121p122p123p124p125p126p127p128p129p130p131p132p133p134p135p136p137p138p139p140p141p142p143p144p145p146p147p148p149p150p151p152p153p154p155p156p157p158p159p160p161p162p163p164p165p166p167p168p169p170p171p172p173p174p175p176p177p178p179p180p181p182p183p184p185p186p187p188p189p190p191p192p193p194p195p196p197p198p199p200p201p202p203p204p205p206p207p208p209p210p211p212p213p214p215p216p217p218p219p220p221p222p223p224p225p226p227p228p229p230p231p232p233p234p235p236p237p238p239p240p241p242p243p244p245p246p247p248p249p250p251p252p253p254p255p256p257p258p259p260p261p262p263p264p265p266p267p268p269p270p271p272p273p274p275p276p277p278p279p280p281p282p283p284p285p286p287p288p289p290p291p292p293p294p295p296p297p298p299p300p301p302p303p304p305p306p307p308p309p310p311p312p313p314p315p316p317p318p319p320p321p322p323p324p325p326p327p328p329p330p331p332p333p334p335p336p337p338p339p340p341p342p343p344p345p346p347p348p349p350p351p352p353p354p355p356p357p358p359p360p361p362p363p364p365p366p367p368p369p370p371p372p373p374p375p376p377p378p379p380p381p382p383p384p385p386p387p388p389p390p391p392p393p394p395p396p397p398p399p400p401p402p403p404p405p406p407p408p409p410p411p412p413p414p415p416p417p418p419p420p421p422p423p424p425p426p427p428p429p430p431p432p433p434p435p436p437p438p439p440p441p442p443p444p445p446p447p448p449p450p451p452p453p454p455p456p457p458p459p460p461p462p463p464p465p466p467p468p469p470p471p472p473p474p475p476p477p478p479p480p481p482p483p484p485p486p487p488p489p490p491p492p493p494p495p496p497p498p499p500p501p502p503p504p505p506p507p508p509p510p511p512p513p514p515p516p517p518p519p520p521p522p523p524p525p526p527p528p529p530p531p532p533p534p535p536p537p538p539p540p541p542p543p544p545p546p547p548p549p550p551p552p553p554p555p556p557p558p559p560p561p562p563p564p565p566p567p568p569p570p571p572p573p574p575p576p577p578p579p580p581p582p583p584p585p586p587p588p589p590p591p592p593p594p595p596p597p598p599p600p601p602p603p604p605p606p607p608p609p610p611p612p613p614p615p616p617p618p619p620p621p622p623p624p625p626p627p628p629p630p631p632p633p634p635p636p637p638p639p640p641p642p643p644p645p646p647p648p649p650p651p652p653p654p655p656p657p658p659p660p661p662p663p664p665p666p667p668p669p670p671p672p673p674p675p676p677p678p679p680p681p682p683p684p685p686p687p688p689p690p691p692p693p694p695p696p697p698p699p700p701p702p703p704p705p706p707p708p709p710p711p712p713p714p715p716p717p718p719p720p721p722p723p724p725p726p727p728p729p730p731p732p733p734p735p736p737p738p739p740p741p742p743p744p745p746p747p748p749p750p751p752p753p754p755p756p757p758p759p760p761p762p763p764p765p766p767p768p769p770p771p772p773p774p775p776p777p778p779p780p781p782p783p784p785p786p787p788p789p790p791p792p793p794p795p796p797p798p799p800p801p802p803p804p805p806p807p808p809p810p811p812p813p814p815p816p817p818p819p820p821p822p823p824p825p826p827p828p829p830p831p832p833p834p835p836p837p838p839p840p841p842p843p844p845p846p847p848p849p850p851p852p853p854p855p856p857p858p859p860p861p862p863p864p865p866p867p868p869p870p871p872p873p874p875p876p877p878p879p880p881p882p883p884p885p886p887p888p889p890p891p892p893p894p895p896p897p898p899p900p901p902p903p904p905p906p907p908p909p910p911p912p913p914p915p916p917p918p919p920p921p922p923p924p925p926p927p928p929p930p931p932p933p934p935p936p937p938p939p940p941p942p943p944p945p946p947p948p949p950p951p952p953p954p955p956p957p958p959p960p961p962p963p964p965p966p967p968p969p970p971p972p973p974p975p976p977p978p979p980p981p982p983p984p985p986p987p988p989p990p991p992p993p994p995p996p997p998p9991000
XX	SCOP domain	dist1 4.51.7.1.1 (99-180) Nuclear ribonucleoprotein A1, RNP A1, TPI (Human (Homo sapiens)) dist1a 4.51.7.1.4 Splicing factor TAF 95 KDa subunit (Human (Homo sapiens)) dist7a1 4.51.7.1.5 (123-204) Sex-lethal protein (Drosophila melanogaster) dist7a2 4.51.7.1.5 (205-269) Sex-lethal protein (Drosophila melanogaster) dlsy1a1 4.51.7.1.9 (111-90) Poly(A)-binding protein (Human (Homo sapiens)) dlsy1a2 4.51.7.1.9 (91-179) Poly(A)-binding protein (Human (Homo sapiens))

도면3f

```

Primer Selection Results

PRINT of: 120502ac02K24p0706.seq  ckt: 143  from: 1 to: 325  October 30, 2003 13:21

      INPUT SUMMARY
      -----

Input sequence: ../seq/120502ac02K24p0706.seq

Primer constraints:
  primer size: 10 - 22
  primer 3' clamp: 5
  primer sequence ambiguity: NOT ALLOWED
  primer GC content: 40.0 - 55.0%
  primer Tm: 50.0 - 65.0 degrees Celsius
  primer self-annealing: . . .
    3' end: < 5 (weight: 2.0)
    total: < 14 (weight: 1.0)
  unique primer binding sites: required
  primer-template and primer-repeat annealing: . .
    3' end: ignored
    total: ignored
  repeated sequences screened: none specified

Product constraints:
  product length: 100 - 325
  product GC content: 40.0 - 55.0%
  product Tm: 70.0 - 95.0 degrees Celsius
  duplicate primer endpoints: NOT ALLOWED
  difference in primer Tm: < 2.0 degrees Celsius
  primer-primer annealing: . . .
    3' end: < 8 (weight: 2.0)
    total: < 14 (weight: 1.0)

      PRIMER SUMMARY
      -----



|                                      | forward | reverse |
|--------------------------------------|---------|---------|
| Number of primers considered:        | 1080    | 1057    |
| Number of primers rejected for . . . |         |         |
| primer 3' clamp:                     | 109     | 112     |
| primer sequence ambiguity:           | 0       | 0       |
| primer GC content:                   | 665     | 646     |
| primer Tm:                           | 41      | 40      |
| non-unique binding sites:            | 0       | 0       |
| primer self-annealing:               | 14      | 15      |


```