

B세포항원결정기자료기지의 구축

김대성, 리석준

경애하는 김정은동지께서는 다음과 같이 말씀하시였다.

《과학연구부문에서는 나라의 경제발전과 인민생활향상에서 전망적으로 풀어야 할 문제들과 현실에서 제기되는 과학기술적문제들을 풀고 첨단을 돌파하여 지식경제건설의 지름길을 열어놓아야 합니다.》

면역정보학은 생명정보학의 한 분파로서 현재까지 면역원성이 확증된 수많은 항원 결정기자료들과 면역관련분자들에 대한 자료기지[1, 2]에 기초하여 전염병의 진단과 예방, 치료를 과학화하는데 리용되고있다.[3]

우리는 본문방식의 B세포항원결정기관련자료기지를 구축하고 망을 통한 검색체계를 개발할수 있는 기초연구를 하였다.

재료와 방법

연구에는 IEDB(<ftp://www.immuneepitope.org>)에서 본문형태로 제공되는 B세포항원결정기 화일이 리용되였다. 총 2 281종의 생물들에서 면역원성이 확증된 540 006개에 달하는 B세포항원결정기자료가 리용되였다.

자료기지관리체제로서 MySQL 5.0, 프로그램작성언어로서 Java(JDK1.7)언어, 봉사기용 언어로서 PHP 4.2를 리용하였다.

결과 및 논의

기초자료로 리용한 B세포항원결정기자료는 본문방식의 자료로서 CSV형식(매 자료는 1개의 행에 표시되며 반점에 의하여 항목들이 구분되는 형식)으로 항원결정기관련정보들을 서술하고있다. B세포항원결정기자료의 주요내용을 부류별, 항목별로 보면 표 1과 같다.

표 1에서 보는바와 같이 항원결정기관련정보에는 하나의 B세포항원결정기에 관하여 항원결정기배열과 해당 항원분자, 생물종에 대한 구체적인 자료와 함께 생체면역측정결과와 특이항체에 대한 자료들, 참고문헌 등 수많은 정보들이 기록되어있다.

우리는 Java언어를 리용하여 위에서 언급한 형식의 B세포항원결정기자료들로부터 항목별로 자료들을 자동적으로 추출하여 MySQL자료기지를 구축하기 위한 프로그램을 작성하고 면역원성이 확증된 모든 B세포항원결정기정보를 포함하는 B세포항원결정기자료기지를 구축하였다.

본문에서는 B세포항원결정기표를 중심으로 면역측정방법표, 참고문헌정보표, 검증된 항원결정기표, 항원결정기표, 생물종표 등 총 15개의 표로 구성된 관계형자료기지를 구축함으로써 자료의 중복을 없애고 자료기지에 대한 검색을 효율적으로 진행할수 있게 하였다. 자료기지를 구성하고있는 표들사이 련관은 그림과 같다.

표 1. B세포항원결정기자료의 주요내용

부류	자료항목	자료의 레
기본정보	항원결정기의 자료기지등록번호	1004748
	참고문헌종류	Literature
	의학정보자료기지련결번호	9557954
참고문헌 관련정보	저자	A. Horv, G. K. Toacut
	잡지명	Immunol Lett
	출판날자	1998
B세포항원결정 기관관련정보	논문제목	A hemagglutinin-based multi-peptide construct elicits enhanced protective immune response in mice against influenza A virus infection.
	B세포항원결정기의 종류	Linear peptide
	B세포항원결정기배열	VTGLRNIPSIQSR
B세포항원결정 기관관련정보	항원분자에서 B세포항원결정기 배열의 시작위치	332
	항원분자에서 B세포항원결정기 배열의 끝위치	344
	항원분자명	Hemagglutinin precursor
류사B세포항원 결정기관련정보	항원원천생물종명	Influenza A virus (A/Puerto Rico/8/1934(H1N1))
	B세포항원결정기에 대한 설명	The HA0 317-341 epitope analog, HA1 317-329 peptide was tested in three forms
	B세포항원결정기와의 관계	analog
생체면역측정관 련정보	류사B세포항원결정기의 종류	Fragment of a Natural Sequence Molecule
	류사B세포항원결정기배열	VTGLRNIPSIQSR
	1차 생체면역측정방법	Administration in vivo
면역측정 관련정보	1차 생체면역측정법에 리용된 면역원의 종류	Linear peptide
	1차 생체면역측정법에 리용된 면역원배열	VTGLRNIPSIQSR
	측정방법	ELISA
특이 항체 관련정보	측정방법종류	qualitative binding
	측정결과	Negative
	특이 항체원천	Serum

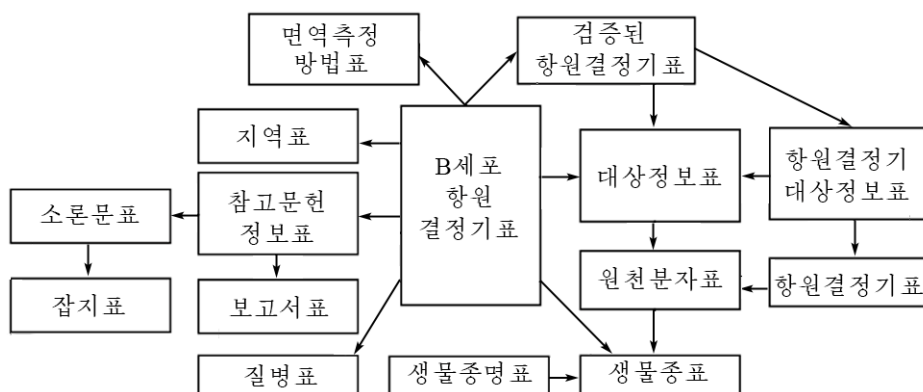


그림. 자료기지를 구성하고있는 표들사이 련관

B세포항원결정기표(표 2)는 B세포항원결정기자료기지의 기본표로서 여기에는 B세포항원결정기관련정보들에 대한 구체적인 자료와 함께 각이한 표들을 참고할수 있는 자료마당들이 포함되어있다.

표 2. B세포항원결정기표의 기본마당구조

마당이름	자료류형	자료의미
BCELL_ID	NUMBER	B세포항원결정기ID
REFERENCE_ID	NUMBER	참고문헌ID
CURATED_EPITOPE_ID	NUMBER	검증된 항원결정기ID
AS_TYPE_ID	NUMBER	측정방법ID
AS_NUM_SUBJECTS	NUMBER	검사받은 대상의 수
AS_NUM_RESPONDED	NUMBER	검사받은 대상중에서 면역응답이 일어난 대상의 수
AS_RESPONSE_FREQUENCY	NUMBER	면역응답을 일으킨 대상의 빈도
AS_IMMUNIZATION_COMMENTS	VARCHAR2(2000)	측정에 리용된 면역법에 대한 해설
AS_COMMENTS	VARCHAR2(4000)	측정법에 대한 해설
H_ORGANISM_ID	NUMBER	숙주생물ID
H_GAZ_ID	VARCHAR2(255)	숙주생물의 지역ID
IV1_ADJUVANTS	VARCHAR2(400)	1차 생체면역법에 리용된 면역보조제
IV1_ROUTE	VARCHAR2(350)	1차 생체면역법에 리용된 면역경로
IV1_DOSE_SCHEDULE	VARCHAR2(250)	1차 생체면역법에 적용된 항원의 량
IV1_DISEASE_ID	NUMBER	1차 생체면역과 관련된 질병ID
IV1_DISEASE_STAGE	VARCHAR2(85)	1차 생체면역과 관련된 질병의 상태
IV1_IMM_TYPE	VARCHAR2(50)	1차 생체면역법의 종류
IV1_IMM_OBJECT_ID	NUMBER	1차 생체면역법에 리용된 대상정보ID
IV2_ADJUVANTS	VARCHAR2(400)	2차 생체면역법에 리용된 면역보조제
IV2_ROUTE	VARCHAR2(35)	2차 생체면역법에 리용된 면역경로
IV2_DOSE_SCHEDULE	VARCHAR2(250)	2차 생체면역법에 적용된 항원의 량
IV2_DISEASE_ID	NUMBER	2차 생체면역과 관련된 질병ID
IV2_DISEASE_STAGE	VARCHAR2(85)	2차 생체면역과 관련된 질병의 상태
IV2_IMM_TYPE	VARCHAR2(50)	2차 생체면역법의 종류
IV2_IMM_OBJECT_ID	NUMBER	2차 생체면역법에 리용된 대상정보ID
AB_NAME	VARCHAR2(200)	특이항체의 이름
AB_TYPE	VARCHAR2(35)	특이항체의 종류
AB_OBJECT_ID	NUMBER	특이항체의 대상정보ID
ANT_TYPE	VARCHAR2(50)	항원의 종류
ANT_OBJECT_ID	NUMBER	항원의 대상정보ID

면역측정방법과 그 결과에 대한 자료는 면역측정방법표에, 항원결정기의 배열과 수식 정보 등과 관련된 구체적인 자료들은 검증된 항원결정기표와 항원결정기표, 항원결정기대상 정보표들에 자료기지화하였다.(표 3)

표 3. 면역측정방법표와 검증된 항원결정기표, 항원결정기표, 항원결정기대상정보표들의 기본마당구조

표이름	마당이름	자료류형	자료의미
면역측정 방법표	ASSAY_TYPE_ID	NUMBER	측정방법ID
	ASSAY_TYPE	VARCHAR2(50)	면역측정방법
	RESPONSE	VARCHAR2(85)	면역응답형태
	UNITS	VARCHAR2(30)	측정단위
검증된 항원 결정기표	CURATED_EPITOPE_ID	NUMBER	검증된 항원결정기ID
	REFERENCE_ID	NUMBER	참고문헌정보ID
	E_OBJECT_ID	NUMBER	검증된 항원결정기의 대상정보ID
	RELATED_OBJECT_ID	NUMBER	류사항원결정기의 대상정보ID
	EPITOPE_ID	NUMBER	항원결정기ID
	DESCRIPTION	VARCHAR2(535)	항원결정기배열
항원 결정기표	LINEAR_PEPTIDE_SEQ	VARCHAR2(4000)	선형항원결정기의 배열
	LINEAR_PEPTIDE_MODIFIED_SEQ	VARCHAR2(4000)	선형항원결정기의 수식배열
	LINEAR_PEPTIDE_MODIFICATION	VARCHAR2(85)	선형항원결정기의 수식잔기
	NON_AA_SOURCE_ID	NUMBER	비펩티드성항원결정기의 원천분자ID
	DISC_SOURCE_ID	NUMBER	불련속항원결정기의 원천분자ID
	MC_REGION	VARCHAR2(4000)	다중사슬단백질의 불련속항원결정기배열
항원결정기 대상정보표	EPITOPE_ID	NUMBER	항원결정기ID
	OBJECT_ID	NUMBER	대상정보ID
	SOURCE_ANTIGEN_ACCESSION	VARCHAR2(50)	단백질자료기지에서 항원분자등록번호
	SOURCE_ORGANISM_ORG_ID	NUMBER	원천생물종ID

배열정보와 관련된 자료는 대상정보표에, 항원결정기를 포함하고있는 원천분자에 대한 자료는 원천분자표에 자료기지화하였다.(표 4)

표 4. 대상정보표와 원천분자표의 기본마당구조

표이름	마당이름	자료류형	자료의미
대상정보표	OBJECT_ID	NUMBER	대상정보ID
	REFERENCE_ID	NUMBER	참고문헌정보ID
	OBJECT_DESCRIPTION	VARCHAR2(535)	대상정보의 배열정보
	ORGANISM_ID	NUMBER	생물종ID
	STARTING_POSITION	NUMBER	원천분자에서 선형항원결정기의 시작위치
	ENDING_POSITION	NUMBER	원천분자에서 선형항원결정기의 마감위치
원천분자표	MOL1_SEQ	VARCHAR2(4000)	원천분자의 배열정보
	MOL_SOURCE_ID	NUMBER	원천분자ID
	SOURCE_ID	NUMBER	원천분자ID
	NAME	VARCHAR2(535)	원천분자의 이름
	CHEMICAL_TYPE	VARCHAR2(85)	화합물질종류
	SEQUENCE	CLOB	원천분자의 배열정보
	ORGANISM_ID	NUMBER	원천생물종ID
	ORGANISM_NAME	VARCHAR2(150)	원천생물종명

또한 참고문헌관련정보들은 참고문헌관련표들(표 5)에, 항원관련생물종들과 질병들, 특정의 지역관련정보들은 표 6에서와 같이 따로 자료기지화하였다.

표 5. 참고문헌관련표들의 기본마당구조

표이름	마당이름	자료유형	자료의미
참고문헌 정보표	REFERENCE_ID	NUMBER	참고문헌정보ID
	REFERENCE_TYPE	VARCHAR2(15)	참고문헌의 종류
	ARTICLE_ID	NUMBER	소론문ID
	REFERENCE_ID	NUMBER	참고문헌정보ID
소론문표	JOURNAL_ID	NUMBER	잡지ID
	JOURNAL_VOLUME	VARCHAR2(15)	소론문이 실린 잡지의 권수
	JOURNAL_ISSUE	VARCHAR2(20)	소론문이 실린 잡지의 호수
	ARTICLE_TITLE	VARCHAR2(1000)	소론문 제목
	ARTICLE_AUTHORS	VARCHAR2(4000)	소론문의 저자
	ARTICLE_ABSTRACT	VARCHAR2(4000)	소론문에 대한 간단한 설명
	PUBMED_ID	VARCHAR2(20)	의학정보자료기지(Medline)연결정보
	JOURNAL_ID	NUMBER	잡지ID
잡지표	JOURNAL_TITLE	VARCHAR2(2000)	잡지제목
	MEDLINE_TA	VARCHAR2(200)	의학정보자료기지(Medline)에 등록된 잡지제목
	SUBMISSION_ID	NUMBER	보고서ID
	REFERENCE_ID	NUMBER	참고문헌정보ID
보고서표	SUBMITTER_NAME	VARCHAR2(85)	보고서의 제목
	SUBMISSION_AUTHORS	VARCHAR2(2000)	보고서의 저자
	SUBMISSION_TITLE	VARCHAR2(400)	보고서의 제목
	SUBMISSION_ABSTRACT	VARCHAR2(4000)	보고서에 대한 간단한 설명

표 6. 지역표와 질병표, 생물종표, 생물종명표의 기본마당구조

표이름	마당이름	자료유형	자료의미
지역표	GAZ_ID	VARCHAR2(200)	지역ID
	DISPLAY_NAME	VARCHAR2(500)	지역이름
질병표	DISEASE_ID	NUMBER	질병ID
	DISEASE_NAME	VARCHAR2(200)	질병이름
	SYNONYMS	CLOB	질병의 다른 이름
생물종표	ORGANISM_ID	NUMBER	생물종ID
	PATH	VARCHAR2(500)	생물분류체계
생물종명표	ORGANISM_ID	NUMBER	생물종ID
	NAME_TXT	VARCHAR2(150)	생물종명

결과적으로 2 281종의 생물들의 각이한 항원성분자들에서 면역원성이 확증된 540 006개의 B세포항원결정기관련자료가 자료기지로 구축되었으며 자료기지크기는 1.77G를 차지하였다.

자료기지에 대한 검색은 검색마당과 검색조건, 표들사이의 연결정보를 포함하는 SQL문을 작성하여 실현할수 있다.

맺 는 말

프로그램을 리용하여 홈페이지를 통한 자료기지의 검색에 효과적으로 리용할수 있는 면역원성이 확증된 B세포항원결정기자료기지를 구축하였다. 자료기지는 B세포항원결정기표를 중심으로 15개 표로 된 망용관계형자료기지로 되어있다.

참 고 문 헌

- [1] Randi Vita et al.; BMC Bioinformatics, **341**, 7, 1, 2006.
- [2] Randi Vita et al.; Database, **105**, 1, 2018.
- [3] Syed Faraz Ahmed et al.; Viruses, **254**, 12, 1, 2020.

주체110(2021)년 4월 5일 원고접수

Development of B Cell Epitope Database

Kim Tae Song, Ri Sok Jun

Recently a number of B cell epitopes of various pathogens is tested and widely distributed through the internet, and thus this information is variously used in immunological research.

We collected the text type of B cell epitopes on the internet and constructed B cell epitope database, which consists of 15 tables including B cell epitope table, and developed the browsing software of B cell epitope information.

Keywords: B cell, epitope, database