



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

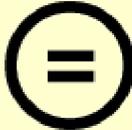
다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

보건학 석사 학위논문

바이오인포매틱스 기법을 활용한
코로나 바이러스의 면역 특이성 연구

**INVESTIGATION ON THE IMMUNOLOGICAL
SPECIFICITY OF CORONA VIRUS UTILIZING
BIOINFORMATICS METHOD**

2016 년 8 월

서울대학교 보건대학원
보건학과 생명정보학 전공
김 미 란

바이오인포매틱스 기법을 활용한
코로나 바이러스의 면역 특이성 연구

지도교수 손 현 석

이 논문을 보건학 석사학위논문으로 제출함

2016년 5월

서울대학교 보건대학원
보건학과 생명정보학 전공
김미란

김미란의 석사학위논문을 인준함

2016년 6월

위원장	정효지 (인)
부위원장	조성일 (인)
위원	손현석 (인)

국 문 초 록

바이오인포매틱스 기법을 활용한 코로나 바이러스의 면역 특이성 연구

서울대학교 보건대학원 보건학과
생명정보학전공

김 미 란

새로운 감염병의 창궐은 공중 보건에 심각한 위협이고 해결해야 할 과제이다. 2015년 한국에 출현한 중동 호흡기 증후군을 일으킨 코로나 바이러스(MERS-CoV)는 국민과 보건당국에 혼란을 주었고, 사회경제적으로 심각한 손실을 초래 하였다. 결과적으로, 한국의 감염병 대응 체계와 의료서비스 환경의 문제점을 인식하게 되었으며, 고위험성 바이러스 감염에 대처하는 한국의 보건 현황과 한계점을 보여 주었고, 치료제와 백신의 개발이라는 과제를 남겨 주었다. 60년전에 처음 발견된 코로나 바이러스는 2003년 사람에게 치명적인 호흡기 감염을 일으키는 SARS-CoV가 유행 되면서 세상의 주목을 받게 되었다. 그 후 코로나 바이러스의 병원성 및 숙주의 면역반응에 관한 이해가 필요함을 재조명하게 되었다. 전 세계적으로 코로나 바이러스에 대한 연구가 진행이 되고 있어서, 코로나 바이러스의 분자 생물학적인 결과와 면역 기전이 밝혀지고는 있지만, 아직 정보들이 산재되어 있고, 기존에 구축된 데이터베이스들 또한 접근과 이용이 용이하지 않다. 따라서 본 연구에서는 코로나 바이러스와 면역 관련 단백질들의 생물학적 특징 및 유전자 정보를 제공하는 HCoV-IMDB를 구축하여 연구자들

에게 유용한 정보를 제공해 주고자 한다. 구축된 데이터베이스의 자료를 기반으로 계통 발생 분석을 시행하여 기존에 알려지지 않았던 코로나 바이러스의 위험성을 예측한 결과, HCoV-OC43의 S 단백질이 MERS-CoV의 S 단백질과 진화적으로 유사하였으며, 이를 통하여 HCoV-OC43의 변이 여부의 지속적인 감시가 필요함을 제시하였다. 본 연구는 새로운 바이러스의 창궐과 기존의 바이러스의 감시 등에 이용 할 수 있는 바이오인포매틱스적인 분석 방법을 제공함으로써 예방을 위한 보건학적 방법을 제시하였다. 또한, 코로나 바이러스의 면역 특이성 분석을 위하여, 네트워크 분석을 시행한 결과, 사람 단백질의 IC1 단백질과 SARS-CoV의 비구조 단백질들 사이의 상호 작용을 확인하였다. IC1은 면역기전 중 보체계의 활성화, 혈액응고, 피브린의 분해와 키닌 합성 등의 물리적인 기전에 관여하는 단백질로써 앞으로 연구가 더 진행이 된다면, 아직 밝혀지지 않은 코로나 바이러스의 면역 기전의 이해에 대한 실마리를 제공해 주고 이는 항생제나 백신의 개발에 도움을 줄 수 있을 것이다. 더 나아가서는 개인의 유전적 차이에 적합한 개인 맞춤형 치료제의 개발에 기반이 되는 연구 자료가 될 것이다. 본 연구에서 이용한 네트워크 분석 방법은 기존의 면역학적 연구 결과들을 통합하여 보여 줌으로써 간과하고 넘어갈 수 있는 부분의 해석이 가능하며, 이러한 연구는 시스템적인 면역 기전의 이해에 기초로 사용될 수 있을 것이다.

.....

주요어 : 코로나 바이러스, 바이오인포매틱스, 면역, 데이터베이스,
보건, 네트워크 분석

학 번 : 2009-21997

목 차

국 문 초 록.....	i
목 차.....	iv
표 목 차.....	vii
그 립 목 차.....	viii
약 어.....	ix
제 1 장. 서 론.....	1
1.1 연구배경.....	1
1.1.1 코로나 바이러스의 특징.....	1
1.1.2 코로나 바이러스의 유행.....	5
1.1.3 코로나 바이러스의 연구현황	8
1.2 코로나 바이러스와 숙주의 상호작용.....	11
1.2.1 바이러스와 면역시스템의 작용.....	11
1.2.2 바이러스와 상호작용하는 면역 단백질	12
1.3 코로나 바이러스에 대한 생명정보학적 분석.....	17
1.3.1 바이러스와 면역 데이터베이스 구축 현황.....	17
1.3.2 계통수 분석법.....	19
1.3.3 네트워크 분석법.....	20
1.4 연구의 필요성	23
1.5 연구의 목적	26

제 2 장. 연구 방법	36
2.1 데이터 수집 및 가공	36
2.2 시스템 개발 환경 및 BLAST 서버 구축	38
2.3 데이터베이스 구현	39
2.4 계통수 작성	41
2.5 네트워크 작성	45
제 3 장. 연구 결과	50
3.1 면역 특이성 데이터베이스 구축	50
3.2 데이터베이스 기반 계통수 분석	52
3.3 데이터베이스 기반 네트워크 분석	53
제 4 장. 고 찰	63
4.1 연구 결과의 고찰	63
4.2 연구의 한계점 및 향후 연구의 제안	69
4.2.1 연구의 한계점	69
4.2.2 향후 연구의 제안	71
4.3 기대성과	73
4.4 보건학적 활용방안	74
제 5 장. 결론 및 총론	76
5.1 결 론	76
5.2 총 론	79

참고문헌	80
Abstract	95
부 록 : 계통수 분석 결과	98

표 목 차

Table 1.1 Classification of Coronaviridae	30
Table 1.2 Function of coronavirus structural proteins	30
Table 1.3 Function of coronavirus non-structural proteins	31
Table 1.4 Function of SARS-CoV and MERS-CoV innate immune	32
Table 1.5 List of phylogenetic tree tools	33
Table 1.6 List of Influenza virus related databases	33
Table 1.7 List of virus-host interaction databases	34
Table 1.8 Criteria for the assessment of visualization tools	35
Table 2.1 Database for virus research	49
Table 2.2 Schema of protein in the database	49
Table 3.1 The frequency of collected viral proteins and immune proteins	61
Table 3.2 The result of pairwise distance matrix of S protein and N protein	62

그림 목 차

Figure 1.1 GenBank data before processing	28
Figure 1.2 GenBank data after processing using Java scripts	29
Figure 2.1 Structure of tree	46
Figure 2.2 Process of research	47
Figure 2.3 Main page of Clustal Omega	48
Figure 2.4 Result of multiple sequence alignment sequence	48
Figure 3.1 Main page of the constructed database	54
Figure 3.2 HCoV protein page of the database	54
Figure 3.3 Immune Related Gene page of the database	55
Figure 3.4 Interleukins and Receptors Gene page of the database	55
Figure 3.5 Standalone BLAST web interface page	56
Figure 3.6 Resulted of homology search by BLAST	57
Figure 3.7 Evolutionary relationships of Spike and Nucleocapsid protein	58
Figure 3.8 Resulted of Virus-Host interaction network by Cytoscape	59
Figure 3.9 Detailed Result of Virus-Host interaction network by Cytoscape	60

약 어

API	activator protein 1
APCs	antigen presenting cells
B2CL1	apoptosis regulator Bcl-2
B2LA1	Bcl-2-related protein A1
BCL2	Bcl-2-like protein 1
BLAST	basic Local Alignment Search Tool
E	envelope
HCoV	human corona virus
HE	hemagglutinin-esterase glycoprotein
IC1	plasma protease C1 inhibitor
IFN	Interferon
IFNAR	Interferon-alpha/beta receptor alpha
IRF	Interferon regulatory factor`12
ISREs	IFN-stimulated response element
Jak1	janus kinase 1
M	membrane
MCL1	Induced myeloid leukemia cell differentiation protein Mcl
MDA-5	melanoma differentiation associated antigen 5
MERS-CoV	middle east respiratory syndrome coronavirus
ML	maximum Likelihood
MP	maximum parsimony
MSA	multiple sequence alignment
N	nucleocapsid
NF-kB	nuclear factor-Kb

NJ	neighbor-joining
NK cell	natural killer lymphocytes
NSP	non structural proteins
ORFs	open reading frames
OTUs	operational taxonomic units=neighbors
PAMPs	pathogen associated molecular patterns
PLP	papain-like protease
PPIA	peptidyl-prolyl cis-trans isomerase A
PPIG	peptidyl-prolyl cis-trans isomerase G
PPIH	peptidyl-prolyl cis-trans isomerase H
RBD	receptor binding domain
RIG	retinoic acid-inducible gene
S	spike
SARS-CoV	severe acute respiratory syndrome coronavirus
SGTA	small glutamine-rich tetratricopeptide repeat-containing proteinalpha
STING	stimulator of interferon genes
TLR	toll like receptor
Tyk2	tyrosine kinase 2
UPGMA	unweighted Pair Group Method with Arithmetic Mean

제 1 장. 서 론

1.1 연구 배경

1.1.1 코로나 바이러스의 특징

1966년에 설립된 ICTV(International Committee on the Taxonomy of Viruses)에서는 바이러스의 보편적인 분류체계를 정립하여, 매년 새로운 바이러스의 정보나 기존에 있던 바이러스의 재 분류 등의 정보를 제공하고 있다. ICTV의 분류에 따르면 코로나 바이러스들은 ‘Nidovirales’목에 속하는 ‘Coronaviridae’과의 바이러스 종이다. ‘Coronaviridae’과는 ‘coronavirinae’와 ‘torovirinae’의 2개의 아과로 분류하며 ‘coronavirinae’는 alpha, beta, gamma, delta coronavirus 4개의 속으로 분류한다(Table 1.1). 인간에게 감염을 일으키는 종으로는 ‘alpha coronaviruses’속의 ‘HCoV-229E’와 ‘HCoV-NL63’이 있고, ‘beta coronaviruses’속에는 ‘HCoV-HKU1’, ‘HCoV-OC43’, ‘MERS-CoV(Middle East respiratory syndrome coronavirus)’, ‘SARS-CoV(Severe Acute Respiratory Syndrome coronavirus)’종이 포함된다(Chan et al., 2015). ‘Nidovirales’목에 있는 바이러스들은 직경이 80~120nm 정도 이고, 외피를 가지고 있는 (+) sense 단일 가닥 RNA 바이러스로써 그 형태가 구형이거나 다형성이며, 유전체의 길이는 27~32kb 정도이고, RNA 바이러스 중 가장 큰 유전체를 가지고 있다(Masters, 2006). 코로나 바이러스 입자는 다섯 가지 주요 구조 단백질인 S(spike), M(membrane), E(envelope), N(nucleocapsid), HE(hemagglutinin-esterase)을

포함하고 있으며, 바이러스 유전체의 3' 말단 부위에 단백질들의 유전자를 가지고 있다. 'HCoV-229E', 'HCoV-NL63', 'SARS-CoV'는 S, M, N, E 네 개의 구조 단백질 유전자를 가지고 있고, 'HCoV-OC43'과 'HCoV-HKU1'는 HE 단백질도 포함한 다섯 개의 구조 단백질 유전자를 포함하고 있다. 바이러스의 복제나 생존력에 관여하지 않는 보조 단백질들의 유전자도 구조단백질 유전자들의 사이에 존재한다. 'MERS-CoV'는 5개의 보조 단백질이 존재하는 반면, 'SARS-CoV'의 경우에는 8개의 보조 단백질이 존재한다(Raj et al., 2013). 'S' 단백질은 바이러스의 외 막에 돌출되어 있으며, 코로나 바이러스가 왕관 모양으로 보이게 하는 단백질이다. 이 단백질은 동중 삼량체의 구조로 당화가 되어 있으며, 숙주 세포막의 수용체에 결합하여 융합이 일어나도록 하는 역할을 수행 한다. 또한 주요 항원으로 작용하여 중화항체뿐만 아니라, 세포 독성 림프구를 자극 하기도 한다(Enjuanes et al., 1995). 'M' 단백질은 25-30kDa의 작은 단백질로 비리온의 막 단백질을 구성하고 있으며, 바이러스의 모양을 형성하고, 세 개의 막통과 도메인 (transmembrane domain)을 가지고 있다(Armstrong et al., 1984). 'N'말단에 당화된 엑토 도메인 (ectodomain)이 존재하고, 'C'말단에 엔도 도메인(endodomain)이 바이러스 입자의 안쪽으로 6-8nm정도 연장되어 존재한다(Nal et al., 2005). 바이러스의 재조합과 감염성에 중요한 역할을 하고, 'N' 단백질과의 결합 뿐만 아니라, 비리온의 막이 만곡이 가능하게 해준다(Neuman et al., 2011; Haan et al., 2003). 'N' 단백질은 두 개의 도메인을 가지고 있으며, 각각의 도메인이 서로 다른 기전으로 RNA에 결합하여 RNA의 합성과 조절에 관여한다(Chang et al., 2006; Kuo et al., 2002). 세포독성 T 림프구가 'N' 단백질의 일부를 인지한다(Perlman, 1998). 이 단백질은 세포막뿐만 아니라, 핵 내에도 존재하며, 세포의 주기 중에서 G2/M단계를 지연을 시키거나, 다음 단계로 넘어가지 못하게

하여 사이토카인의 생성을 억제 시킨다(Wurm et al., 2001). ‘HE’ 단백질은 ‘beta coronavirus’인 ‘HCoV-OC43’과 ‘HCoV-HKU1’에 존재한다. 이 단백질의 절반이 숙주세포의 세포막에 있는 뉴라민산(neuraminic acid)에 결합하여, 세포막에 바이러스가 부착하는 역할을 한다. 에스테르가수분해효소(esterase)는 뉴라민산으로부터 아세틸기를 떼어낸다(Luytjes et al., 1998). 이러한 기능은 ‘S’ 단백질이 세포 내로 침입하고, 바이러스가 점막에서 퍼져 나갈 수 있도록 도와주는 역할을 한다(Kato et al., 2005). ‘E’ 단백질은 외 막에 존재하고, 안쪽에 C 말단이 구부러져 존재하며, N 말단은 돌출되어 있다. 기능은 알려져 있지 않으나, ‘SARS-CoV’의 경우에는 바이러스의 M, N 단백질과 함께 재조합과 방출의 역할을 하는 것으로 알려져 있다(Siu et al., 2008). 코로나 바이러스 구조 단백질의 종류와 특징을 Table 1.2에 정리 하였다. RNA의 복제효소는 유전체의 5' 쪽에 두 개의 ‘ORF1a’와 ‘ORF1b’의 ORFs(open reading frames)이 존재한다. ORFs의 전사가 이루어지면, ‘pp1a’와 ‘pp1ab’ 2개의 다 단백질이 만들어지고, 2개의 바이러스성 단백질효소에 의하여 16개의 비구조 단백질(Non structural proteins, Nsp)들이 만들어진다. 이러한 단백질들은 바이러스 유전체의 복제와 ‘subgenomic messenger RNAs’의 전사에 관여한다(Ziebuhr, 2005). 코로나 바이러스 비구조 단백질의 종류와 특징을 Table 1.3에 정리 하였다.

사람에게 질환을 일으키는 대부분의 코로나 바이러스들은 호흡기 감염을 일으키는 가벼운 감기 정도의 바이러스로 알려져 있었으나, SARS의 유행 후에, 코로나 바이러스들의 중요성을 인식하고 연구가 이루어지기 시작하였다(Susan et al., 2005). ‘alpha coronavirus’에 속하는 ‘HCoV-229E’와 ‘beta coronavirus lineage A’에 속하는 ‘HCoV-OC43’은 50년 전에 분리되어 명명이 되었고(McIntosh et al.,1967), ‘alpha coronavirus’에 속하는 ‘HCoV-NL63’과

‘beta coronavirus lineage A’에 속하는 ‘HCoV-HKU1’은 SARS의 유행 후에 동정이 되었다(Van et al., 2004; Woo et al., 2005). 이러한 바이러스들은 사람에게 호흡기 감염을 일으키는 원인의 15~30%를 차지하며, 신생아나 고령의 노인, 기저질환을 가진 환자에게서는 하부 호흡기 감염의 발병률이 좀 더 높고, 특히 ‘HCoV-NL63’은 급성 후두염과 연관이 높다(Van et al., 2005). ‘HCoV-229E’는 세계 각국에서 분리된 서열들의 차이가 거의 존재하지 않지만, ‘HCoV-OC43’은 같은 지역에서 분리되더라도 연도별로 서열의 유전적 다양성이 존재한다(Chibo et al., 2006; Vijgen et al., 2005). 이것은 ‘HCoV-229E’는 종간의 감염이 일어나지 않는 반면, ‘HCoV-OC43’은 ‘BCoV(bovine coronavirus)’와 유사하며, 반추 동물끼리의 전염이 가능함을 설명할 수 있다. ‘SARS-CoV’는 ‘beta coronavirus lineage B’에 속하며, 2002~2003년 중국의 광둥 지방에서 중증 급성 호흡기 증후군의 원인 바이러스이다. ‘SARS-CoV’는 폐의 상피세포를 감염시키고, 대식세포와 수지상 세포에 침입하여 세포의 사멸을 유도한다(Peiris et al., 2003). 감염된 세포들은 사이토카인과 케모카인을 분비하고, 질환을 일으키는데 중요한 역할을 한다(Spiegel et al., 2006). ‘MERS-CoV’는 ‘beta coronavirus lineage C’인 박쥐 코로나 바이러스인 ‘HKU4’와 ‘HKU5’와 연관성이 높아, 바이러스의 근원지를 박쥐로 여겨졌으나 혈청 학적인 연구에서는 중동지역의 낙타의 항체가 발견되고, 낙타의 세포에서 ‘MERS-CoV’의 증식이 가능한 것을 발견하여 자연 숙주로 낙타를 지목하게 되었다(Von et al., 2015; Eckerle et al., 2014).

코로나 바이러스들의 감염은 숙주세포 내로 침입하면서 시작된다. ‘SARS-CoV’와 ‘MERS-CoV’의 외 막에 존재하는 ‘S’ 당단백인 ‘RBD(receptor binding domain)’이 숙주세포의 세포막에 존재하는 수용체에 결합을 한다. ‘SARS-CoV’는 숙주 세포의 수용체

‘ACE2(angiotensin converting enzyme 2)’와 결합을 하고 ‘MERS-CoV’는 ‘DPP4’의 수용체에 각각 결합한다(Li et al., 2003). 대부분 코로나 바이러스들의 ‘S’ 단백질들은 숙주세포의 세포막에 존재하는 ‘Ecto-peptidase’라는 단백 분해 효소에 의하여 ‘S1’과 ‘S2 RBD’로 분해된다(Yamada et al., 2009). ‘S1’이 수용체에 결합하면, ‘S2 RBD’에 구조적인 변화가 일어나 숙주세포의 세포막에 침입하기 위해 감추어져 있던 결합 펩티드가 노출된다. 이 펩티드를 이용하여 숙주세포의 이중 지질 층에 결합이 이루어지고, 바이러스의 뉴클리오캡시드가 세포질 내로 이동한다. 탈피 과정이 이루어진 후에, 바이러스의 RNA 게놈은 복제에 관여하는 단백질들의 번역을 위한 mRNA의 역할을 한다(Masters, 2006). 숙주 세포막을 재배열 하면서 이중의 막으로 구성된 낭(double membrane vesicles, DMVs)을 형성하고, 그 안에서 코로나바이러스들은 전사와 복제를 진행한다. 이중낭 속에서 새롭게 형성된 RNA는 N 단백질로 싸여지고, 소포체와 골지체의 중간 부분(endoplasmic reticulum-golgi intermediate compartment, ERIC)으로 이동하여 다른 단백질들과 조립을 한 후 새로운 바이러스가 되어 세포막 쪽으로 이동 후 세포 밖으로 방출된다(Azhar et al., 2014).

1.1.2 코로나 바이러스의 유행

바이러스성 병원체들이 급성 호흡기 질환을 발생시키는 주된 요인이며, 바이러스성 호흡기 질환으로 인한 입원이나 진료비용, 업무의 차질, 학교나 보육 시설의 휴교 등은 사회경제적인 영향이 상당하다(Fendrick et al., 2003). 모든 연령 군에서 바이러스성 호흡기 감염을 경험하고 있고, 소아, 노인, 면역력이 저하된 환자들에게서는 치명적이다(Cate, 1998). 역학적 연구들에서 보면 호흡기 바이러스들

중에서도 RS 바이러스(respiratory syncytial virus, RSV), 인플루엔자 바이러스(influenza viruses, flu), 메타뉴모바이러스(human metapneumo virus, HMPV), 리노 바이러스(rhinoviruses, RV), 파라 인플루엔자 바이러스(parainfluenza virus, PIV), 아데노바이러스(adenovirus, AV), 코로나 바이러스(human coronavirus, hCoV) 등이 주된 요인으로 밝혀져 있다. 일반적으로 이러한 바이러스들은 기도의 상피세포를 감염시키고, 숙주세포의 단백질과 상호작용하면서 감염을 촉진시키고, 숙주의 선천면역과 후천면역을 활성화 시켜, 염증을 유발하거나 질병을 일으키는 역할을 한다(Ralph, 2013). 과거에 대부분의 코로나 바이러스들은 감기 정도의 가벼운 증상을 보였지만, 최근에는 급성 호흡기 증후군(SARS), 중동아시아 호흡기 증후군(MERS)의 예에서 볼 수 있었던 것처럼, 심각한 폐 질환과 중증 호흡기 증후군을 동반하면서, 치사율이 다른 호흡기 바이러스보다 높아 코로나 바이러스가 재조명 되고 있다(Zumla et al., 2015). ‘SARS-CoV’는 2002~2003년 중국의 광둥성 지방에서 처음 발생하여 29개의 나라로 빠르게 전파되어, 8,000명 이상이 감염이 되었고, 774명의 사망자가 발생하여, 치사율은 9~10% 정도이다(Hui et al., 2014). 2004년 이후로 SARS의 감염은 보고되지 않고 있지만, 여행객들의 증가로 인하여 주변국으로부터 신종 병원체의 해외 유입 가능성 및 변이 병원체로 인한 유행병의 가능성은 잠재되어 있다. ‘SARS-CoV’가 유행하는 동안 히말라야 판, 사향고양이와 너구리 등의 동물로부터 바이러스가 분리되었다. 이러한 동물들에서 사람의 수용체로 알려진 ‘ACE2’가 확인되었고, ‘SARS-CoV’의 매개체가 박쥐라고 알려졌다. 가금류 시장에 종사하는 몇몇 사람들은 바이러스의 유행이 발생하기 이전에 이미 감염이 되었던 혈청 학적인 소견을 보이지만, 증상을 동반하지는 않았다. 이러한 결과는 이미 오래전부터 가금류 시장에는 코로나 바이러스가

존재하고 있었으나, 사람 집단으로의 확산을 위한 일련의 요소들의 부족으로 유행을 일으키지 못 했던 것이라고 추정할 수 있다(Guan et al., 2003). ‘SARS-CoV’는 감염된 환자로부터의 직접 접촉에 의해서 전파가 일어나므로, 바이러스의 유행은 가정과 병원에서의 관리 및 검역의 강화 등으로 감염을 제어할 수 있었고, 2003년 6월 이후에 작은 수의 감염을 일으키고는 있으나 대단위의 감염은 일어나지 않고 있는 상황이다(Peiris et al., 2003). 2012년에 중동아시아 지역에서 중동아시아 호흡기 증후군 (Middle East respiratory syndrome, MERS)이라 불리는 치사율이 높은 새로운 호흡기 질환이 출현하였다. 이 질환의 원인 바이러스는 HCoV-EMC/2012(human coronavirus-Erasmus Medical Center/ 2012)였으며, 후에 MERS-CoV(Middle East respiratory syndrome coronavirus)라고 재명명 되어졌다(Van et al., 2012). ‘MERS-CoV’는 2012년 6월에 사우디아라비아 에서 처음 발견되고, 주로 중동아시아 지역인 요르단, 카타르, 오멘, 아랍 에미리트 지역에서 발생하였다. 그러나, 교통수단의 발달과 각종 물자교역 및 인적 교류 등으로 인하여, 유럽, 아프리카, 아시아 등으로 퍼져나갔다(Oboho et al., 2015). WHO 따르면, 2014년에는 2012년 9월 이후로 2014년까지 857명의 감염자를 발생시키고 334명의 사망자가 발생하여 치사율이 35%정도 달했으나(WHO, 2014), 2016년 6월까지의 통계를 보면, 전 세계적으로 27개의 나라에서 감염을 보고 하였고, ‘MERS-CoV’에 의한 감염자 수는 1,769명 사망자 수는 630명으로 치사율이 35% 정도라고 보고 하였다(WHO, 2016).

2015년 5월 한국에서 유행을 일으킨 ‘MERS-CoV’는 2016년 1월 바이러스의 종식이 있을 때까지, 186명의 감염자가 발생 하였고, 36명의 사망자, 16,000명 이상의 일반인과 의료종사자가 격리 되었다(Korea centers for disease control and prevention, 2015). 이러한

감염의 발생은 중동아시아를 제외한 지역에서 가장 많은 감염자가 발생한 것으로 기록되었다(WHO, 2015). 한국에서의 ‘MERS-CoV’의 발생은 코로나 바이러스의 감염전파에 대한 새로운 특징을 보여주었으며, 고위험성 바이러스 감염에 대처하는 한국의 보건 현황과 한계점을 보여 주었다. 이는 앞으로 유입될 수 있는 신 변종 바이러스에 대한 분석과 이에 효과적으로 대응할 수 있는 대비책의 마련이 시급함을 알려 주었다. 국내에서 발생한 ‘MERS-CoV’의 특징은 중동지역의 감염 증상과는 차이가 있어, 코로나 바이러스의 위험성에 관심이 집중되고 있다. 중동지역 감염의 특징은 주로 낙타 등의 동물에 의해 감염이 이루어졌고, 35~40%의 높은 사망률에 비하여 전파력은 낮았다. 주로 근접 접촉감염에 의하여 감염이 이루어졌고, 전파가 진행됨에 따라 병원성이 감소되는 경향을 보였다(Butler, 2015; Banik, 2015). 한국에서 발생한 ‘MERS-CoV’감염의 특징은 사람 간의 전파가 이루어지고, 3차 감염이 발생하고, 슈퍼 전파자까지 등장하는 높은 전파력을 보여 주었으나, 사망률은 약 20% 정도로 낮았다(Je et al., 2015).

1.1.3 코로나 바이러스의 연구현황

코로나 바이러스에 관한 연구들을 살펴보면, 2016년 수행된 연구로 한국에서 유행을 일으킨 8개 ‘MERS-CoV’의 S 단백질과 131개 ‘Reference MERS-CoV’의 S 단백질의 유전자 서열을 이용하여 계통수 분석 연구를 시행하였다(Kim Dw et al., 2016). 분석 결과 한국에서 분리된 8개의 균주의 S단백질 서열은 2015년 사우디아라비아 Riyadh에서 분리되었던 ‘MERS-CoV’ S 단백질과 서열과 99.68~99.9%로 높은 동질성을 보였다. 한국에서 분리된 ‘MERS-CoV’의 S 단백질의 ‘RBD’ 시퀀스에서 T3177C와 C3267T의

돌연변이를 확인하였다. 이러한 돌연변이가 한국에서 ‘MERS-CoV’의 급속한 전파에 영향을 미쳤는지의 여부는 더 연구가 진행이 되어야 할 것이라고 제안하였다(Kim Dw et al., 2016). 다른 연구에서는 2015년 ‘MERS-CoV’에 감염된 13명 환자에서 분리된 S 단백질에서 2개의 점 돌연변이를 발견하였고, 이러한 돌연변이를 가진 바이러스들은 사람의 ‘CD26’ 수용체에 결합하는 친화력이 떨어져 바이러스가 세포 내로의 침입이 어려워지므로, MERS-CoV의 전체적인 사망률인 35%보다 우리나라의 사망률이 20%로 낮은 이유라고 보고 하였다. 또한 ‘CD26’에 중화항체 처리를 하면 친화력이 떨어지는 결과도 얻어 이를 이용한 추후 연구를 제안하고 있다(Kim Y et al., 2016). 2015년 질병 관리 본부에서 시행한 연구 결과에 따르면, 2015년 5월 국내 ‘MERS-CoV’ 환자로부터 검출된 바이러스와 2015년 2월에 사우디아라비아에서 검출된 코로나 바이러스와의 염기서열을 분석한 결과, 염기서열의 동질성이 99.61%로 매우 유사하였으며, 두 코로나 바이러스는 2013년 카타르와 사우디아라비아(Hafr-Ai-batin) 지역에서 발생한 유행의 원인 코로나 바이러스와 유사한 서열을 가지고 있었다(Korea Centers for Disease Control and Prevention., 2015). 또 다른 연구에서는 한국에서 검출된 ‘MERS-CoV’, 중국에서 검출된 바이러스, 사우디아라비아에서 검출된 6개 바이러스의 염기서열을 이용하여, 계통수 분석을 실시한 결과, 염기 서열의 동질성이 99.90~99.96%로 매우 높았고, 유전적 재조합 분석을 시행한 결과, 6개의 바이러스에서 공통적으로 유전적 재조합이 일어난 것을 알 수 있었다. 현재까지는 바이러스의 재조합이나 변이가 병원성의 증가와 전파속도의 증가 등에 영향을 미친다는 연구 결과는 없지만, 코로나 바이러스의 유전자 변이 여부의 감시 등은 꾸준히 진행해야 할 과제로 남아 있다고 보고하였다(Wang et al., 2015).

대규모 게놈 및 단백질체학 방법은 바이러스 감염으로 인한 바이러스와 숙주 사이의 복잡한 상호작용 등의 네트워크를 밝히는데 이용되어 왔다. 예를 들면, 효모를 이용한 실험과 유전자 발현의 전체 게놈 분석을 이용하여 사람과 인플루엔자 ‘H1N1’ 사이의 상호작용의 네트워크를 작성하였으며, 결과 1735개의 후보 유전자를 동정하였다(Shapira, 2009). 마찬가지로, 213개의 숙주세포 인자와 ‘HIV-1’을 인코딩하는 단백질을 이용하여 네트워크 분석을 시행하여 ‘HIV-1’의 DNA 합성의 개시에 관여하는 40개의 새로운 후보 유전자를 동정하였다(Konig et al., 2008). ‘SARS-CoV’의 단백질들 간의 네트워크 분석을 시행한 결과 ‘Nsp2’, ‘Nsp8’과 ‘ORF9b’ 단백질이 다른 단백질들과 다양하게 상호작용하는 결과를 얻었고, 실험을 통하여 ‘Nsp8’은 바이러스의 복제에 관여를 하지만, ‘Nsp2’와 ‘ORF9b’는 바이러스의 복제에는 관여하지 않고, 바이러스의 병원성과 관련이 있다는 결과를 보고하였다(Von et al., 2007). 또 다른 연구에서는 기저질환을 가지고 있는 환자가 ‘SARS-CoV’와 ‘HIV-1’에 감염이 되었을 때 일어나는 상호작용의 결과를 질환 별로 분류하여 네트워크 분석을 시행한 연구도 보고 되었다(Moni et al., 2014). 이러한 바이오인포매틱스 기법을 활용한 바이러스와 숙주 간의 상호작용의 분석 방법들은 새로운 바이러스의 창궐이나 기존에 존재하는 바이러스의 재조합이나 변이 등으로 인한 유행의 여부를 예측하는데 유용한 도구로 이용될 수 있을 것이다.

1.2 코로나 바이러스와 숙주간의 상호작용

1.2.1 바이러스와 면역 시스템의 작용

바이러스는 살아있는 세포 내에서만 살아가는 기생체로 숙주가 소멸된다면, 바이러스도 살아갈 수 없기 때문에, 숙주의 면역체계와 바이러스는 적절하게 균형을 유지하며 상호 작용하면서 진화하여 왔다(Zinkernagel, 1996). 바이러스에 감염이 되면, 두 단계의 면역반응이 일어난다. 첫 번째 반응은 숙주가 바이러스에 신속히 대항하는 ‘선천적 면역반응(비특이적인 반응)’ 으로 인터페론을 합성하여, 자연살해세포(natural killer lymphocytes)와 수지상 세포 (dendritic cell)를 자극하는 등 항바이러스성 상태를 유발하고, 세포사멸, 식균작용, 보체계의 작용 등에 의하여 일차적인 방어가 일어난다. 바이러스의 복제 등이 일어나면서 감염이 진행이 되면 ‘후천적 면역반응(특이적인 면역반응)’이 시작된다. 후천적 면역 반응에는 B 림프구에 의한 바이러스성 특이 항체들을 생성하는 체액성 면역반응과 감염된 세포를 죽이고, 세포독성 T림프구를 생성시키는 세포 매개성 면역반응이 있다. 바이러스가 표적세포에 침입하기 전에는 항체도 효과적인 항바이러스성 기능을 발휘한다. 감염에 대하여 대항하는 항체는 IgA, IgG, IgM이 관여를 한다. IgM이 환자의 혈청 검체에서 발견이 되면, 이는 최근에 또는 현재 감염이 되어 있는 상태를 의미한다. IgG는 혈청에 가장 많은 비율로 존재하며, 오랜 기간 동안 생성되고, 그 결과 과거에 감염이 되었던 병원체에 대하여 빠르게 면역반응이 일어날 수 있다. IgA 항체는 점막에서 숙주 방어에 중요한 역할을 하며, 보체를 활성화시키지 않으므로, 점막 부위에 심한 염증을 유발하지 않고 바이러스 감염을 막아내는 역할을 한다(Lamm, 1997). 세포 독성

T림프구는 확립된 바이러스의 감염을 제거하고 완결시키는데 결정적인 역할을 한다. T 림프구는 항원 제공 세포(Antigen presenting cells, APCs)의 표면에 존재하는 ‘MHC I’과 ‘II’에 의해 바이러스의 침입을 인지한다. 이때 사이토카인들이 B 세포와 T 세포의 상호 작용이 일어나게 하고, 두 세포의 성장을 촉진시킨다. 이러한 후천적 면역반응의 결과 오랫동안 지속되는 기억 세포를 생성하고, 같은 바이러스에 재감염 이 일어나면 보다 더 빠르게 면역반응이 일어나는 것을 가능하게 한다(Teri et al., 2013). 하지만, 바이러스는 이와 같은 단계별 숙주의 방어 기전에 대항하는 기전을 가지고 있어, 급성으로 감염을 일으키고 사라지거나, 평생 숙주에게서 기생하며 잠복 또는 지속 감염을 일으키거나, 면역반응이 생성된 후에 재감염이 이루어졌을 경우 오히려 증상을 악화 시키는 등 바이러스의 종류에 따라 다양한 면역반응이 나타난다(Hwang et al., 2004).

1.2.2 바이러스와 상호 작용하는 면역 단백질

항바이러스성 선천면역에 관여하는 세포들에는 NK세포, 수지상세포, 대식세포, I형 인터페론, 케모카인, 사이토카인과 같은 수용성 매개체들이 있다. 자연 살해 세포는 감염된 세포를 살해함으로써 바이러스의 복제를 직접적으로 막아내거나, II형 인터페론 ‘IFN- γ ’를 이용하여 간접적으로 막아낸다. 수지상 세포와 대식세포는 패턴 인식 수용체(pattern-recognition receptors, PRRs)가 존재하며, 병원체마다 고유하게 가지고 있는 특정한 패턴(pathogen-associated molecule pattern, PAMP)을 면역세포의 PRRs이 인식하여, 바이러스 감염 시 처음 나타나는 ‘IFN- α/β ’, 사이토카인, 케모카인의 분비를 촉진시킨다. 또한, 항원 제공 세포로서의 역할을 하며, NK 세포, B

세포, T 세포의 활성도를 조절한다(Laiosa et al., 2006). 인터페론들은 선천면역계가 바이러스 감염을 처리하는 주된 방법이며, I형 인터페론을 분비하여 항바이러스성 방어, 면역계의 활성화, 세포의 성장과 사멸에 관여한다(Goodbourn et al., 2000). 인터페론은 9번 염색체에 단일 집단으로 존재하며, 아미노산의 서열과 수용체 인식에 기초하여 'IFN- α ', ' β ', ' α/β '로 분류한다. 'IFN- α '는 바이러스 감염에 직접적으로 반응하고, 'IFN- α/β '는 감염된 세포에서 분비되며, 주변 세포의 바이러스 감염을 막는다(Stark et al., 1998). 'IFN- α/β '는 'I형 IFN 수용체'에 인식이 되면, 수용체 관련 키나제인 'Jak1(Janus kinase 1)'과 'Tyk2(Tyrosine kinase2)'가 활성화 되고, 전사 분자인 'Stat1'과 'Stat2'의 분자가 활성화 되면서 'Stat1-Stat2' 복합체가 형성이 된다. 이 복합체에 인터페론 조절인자 9번(IRF9)이 결합한 후, 핵 내의 'ISREs(IFN-stimulated response elements)'의 프로모터에 결합이 이루어지면, 항바이러스성 반응과 후천적 면역반응의 조절에 관여하는 'ISGs(IFN-stimulated genes)'를 생성한다(Aronson et al., 2002). 형질세포 모양 수지상 세포가 'INF- α '의 주 생산 세포이며 대식세포에서도 생성된다. 'INF- β '는 여러 종류의 세포에서 생성이 된다. 'I형 IFN' 합성의 가장 강력한 자극은 바이러스의 핵산이다. 세포질 내의 'RIG(retinoic acid-inducible gene)' 유사 수용체인 'RIG-I', 'MDA-5(melanoma differentiation associated antigen 5)'나 엔도솜(Endosome)에 존재하는 패턴인식 수용체 중의 하나인 'TLR(toll like receptor)-3, 7, 8, 9'에 의해 바이러스의 핵산을 인식하게 되면, 인터페론 조절인자가 활성화되는 신호전달을 통하여 'I형 IFN' 유전자 발현을 조절한다(Heil et al., 2004 ; Kato et al., 2005). 코로나 바이러스 RNA는 숙주세포의 세포질에 들어오면 '5' capping'을 하고 있고, 전사과정 동안에는 이중의 막으로 된 낭(double membrane vesicles, DMVs)안에서 이루어지므로 숙주세포의 세포질에 존재하는

‘RIG-I’와 ‘MDA5’와 같은 이중나선 센서들에 의해 인지가 되지 않으므로, 바이러스의 단백질 합성이 가능하다(Van et al., 2008). 감염된 숙주가 만들어낸 바이러스의 단백질인 ‘Nsp1’, ‘Nsp3’, ‘Nsp7’, ‘Nsp16’, ‘N protein’, ‘M protein’, ‘PLP(papain-like protease)’와 ‘SARS-CoV’의 ‘ORF6’, ‘ORF3b’와 ‘MERS-CoV’의 악세서리 단백질(accessory protein)인 ‘4a’, ‘4b’단백질이 인터페론의 유도과 신호전달을 억제한다(Chan et al., 2015; Raj et al., 2013). ‘Nsp1’은 IFN의 생성을 저해하는 세가지의 기능을 가지고 있다. 첫번째 기능은, ‘STAT1’의 인산화를 저지하여 감염된 세포의 인터페론 신호전달을 방해하고, 인터페론 생산을 저지 시킨다(Wathelet et al., 2007). 두번째, 숙주세포의 40S 리보솜에 결합하여 리보솜의 번역 활동을 저하시킴으로써 숙주세포의 유전자 발현을 방해한다(Raj et al., 2013). 세번째, 숙주세포의 mRNA에 결합하여 숙주세포가 자신의 단백질을 만드는 것을 저해한다(Kamitani et al., 2009). ‘PLP(papain-like protease)’도메인을 통해 생성된 ‘Nsp3’는 ‘IRF3’의 탈유비퀴틴을 일으켜 핵 전위를 방지하고, ‘IFN I’의 생성을 저해한다(Chen et al., 2014). ‘Nsp7’은 ‘RIG-I’ 신호전달 기전을 방해하여, ‘IFN I’의 생성을 저해시킨다 (Totura, 2014). ‘Nsp16’은 코로나 바이러스의 RNA에 ‘20-O 메틸화’를 통하여 숙주세포의 RNA와의 구별이 불가능하게 만든다(Menachery et al., 2014). 구조 단백질인 ‘N’ 단백질은 ‘API(activator protein 1)’의 신호전달과 ‘PKR(protein kinase R)’의 기능을 저해 시킬 뿐만 아니라, ‘NF-kB(nuclear factor-kB)’의 활성을 방해하여 ‘IFN β ’의 생성을 저해한다(Raj et al., 2013; Kopecky-Brimberg et al., 2007). ‘M’ 단백질은 RIG-I, MAVS, IKK ϵ , TBK1을 포함하는 ‘RIG-I’ 신호기전을 방해하여 ‘IFN β ’의 생성을 저해 시킨다(Siu et al., 2008). ‘SARS-CoV’의 ‘ORF6’는 전사에 관여하는 인자인 ‘STAT1’의 핵 전위를 방해하여 ‘IFNAR’가 비활성화 되고, ‘JAK-STAT’신호 전달

기능이 저해됨으로써 ‘IFN I’의 생성이 억제된다(Frieman et al., 2007). ‘ORF3b’는 ‘MAVS’와 ‘RIG I’을 통한 인터페론의 신호 전달을 방해한다(Kopecky et al., 2007). ‘MERS-CoV’의 ‘4a’는 ‘MDA5’와 결합하여, 바이러스의 dsRNA의 인식을 방해 함으로써 ‘IFN’의 합성을 저해한다(Niemeyer et al., 2013). ‘4b’는 ‘IFN- β promoter’의 활성을 저해하여 ‘NF-kB’의 신호전달을 방해 함으로써 인터페론의 활성화를 막는다(Matthews et al., 2014). 숙주세포의 인터페론 생성을 억제하는 ‘SARS-CoV’와 ‘MERS-CoV’의 단백질의 종류와 기능을 Table 1.5에 정리하였다. 대식세포 중에서도 단핵구에서 유래된 대식세포가 ‘MERS-CoV’에 감염이 일어나면 세포 내에서 바이러스가 복제를 할 수 있다. 감염 후 48시간 내에 다른 사람에게 감염을 일으키며 ‘MERS-CoV’에 감염된 단핵구 유래 대식세포에서 ‘viral RNA’가 2~4배 정도 현저하게 증가한다. 이때 감염된 대식세포에서 ‘MERS-CoV’는 ‘IP-10/CXCL-10’, ‘MCP-1/CCL-2’, ‘MIP-1 α /CCL-3’, ‘RANTES/CCL-5’, ‘IL-8’, ‘IL-12’와 같은 염증성, 화학주성 사이토카인과 케모카인의 발현을 유도한다(Zhou et al., 2014). SARS-CoV의 경우에는 사람의 대식세포를 감염 시킬 수는 있으나, 세포 내에서 바이러스의 복제는 일어나지 않고, 바이러스의 입자도 생성하지는 않지만, 감염된 세포에서 ‘IP-10/CXCL10’, ‘MCP-1/CCL2’와 같은 염증성 케모카인을 분비하게 한다. 그러나 선천면역의 주된 역할을 하는 항바이러스성 사이토카인 ‘IFN- α ’와 ‘IFN- β ’는 유도하지 않는다(Cheung et al., 2005). 수지상 세포는 병원체가 침입 시에 면역계의 필수적인 보조 역할을 하는 세포이며, 선천면역과 적응면역의 다리 역할을 한다. 단핵구에서 유래된 수지상 세포가 ‘MERS-CoV’에 감염이 되면, 바이러스성 RNA나 바이러스성 항원이 증가가 된다. 감염된 세포에서는 ‘IFN- β ’와 ‘IFN- α ’는 발현되지 않는 반면, ‘IFN- γ ’, ‘IL-12’, ‘IP-10/CXCL-10’,

‘RANTES/CCL-5’의 분비가 ‘SARS-CoV’에 비하여 훨씬 높았다(Chu et al. 2014). ‘SARS-CoV’에 감염된 세포에서 바이러스의 복제 등이 일어나지는 않지만, 세포 표면에 ‘MHC class II’와 보조 자극 분자 등의 발현과 같은 표현형 및 기능적 성숙을 통하여 T 세포를 자극할 수 있는 능력과 ‘TNF- α ’, ‘MIP-1 α /CCL-3’, ‘RANTES/CCL-5’, ‘IP-10/CXCL-10’, ‘MCP-1/CCL-2’와 같은 사이토카인과 케모카인의 분비를 유도한다(Law et al., 2005).

‘SARS’나 ‘MERS’ 환자에서 보이는 림프구 감소증 특히, T 림프구의 감소는 아직 정확한 기전은 현재 알려지지 않았다. ‘MERS-CoV’의 유행 후 이루어진 최근 연구에서는, 사람의 ‘DPP4’로 형질전환 시킨 T 세포 결핍 쥐의 경우에, 폐에서 ‘MERS-CoV’가 존재하였고, 대조군으로 실험한 B 세포 결핍의 쥐에서는 바이러스가 검출되지 않았다. 이는 T 세포가 바이러스의 병원성에 관여하는 중요한 역할을 하고 있음을 알 수 있다(Zhao et al., 2014). 또한 ‘MERS-CoV’에 감염된 T 세포들은 caspase 의존성 자기 사멸 기전을 유도한다(Chu et al., 2015). ‘MERS-CoV’가 T 세포를 감염시키고, 사멸을 유도하는 이러한 특이적인 능력이 바이러스의 높은 병원성을 유발하는 것을 증명해 준다. ‘SARS-CoV’의 감염의 병원성을 조절하는 T 세포의 역할은 완전하게 증명되지 않았다. 쥐를 이용한 실험을 통하여, ‘CD4+ T 세포’의 결핍은 간질성 폐렴과 폐 조직에서 ‘SARS-CoV’의 제거가 지연이 되는 것을 관찰하였고, 이는 중화항체의 생성 감소와 림프구를 폐에 모여들게 하는 사이토카인의 감소를 초래하였다(Chen et al., 2010).

1.3 코로나바이러스에 대한 생명정보학적분석

1.3.1 바이러스와 면역 데이터베이스 구축 현황

분자생물학적인 기술과 바이오인포매틱스 기법들의 발달로 인하여 유전체의 분석 결과와 실험 결과들은 점점 늘어나고 있다. 이러한 정보를 잘 분류하고 저장하여 2차 연구를 가능하게 해주는 데이터베이스의 구축이 필요하며, 국내외의 많은 분야에서 데이터베이스가 구축되고 있다. 하지만, 코로나 바이러스의 유행이 최근에 일어나기 시작하고 주목을 받기 시작하면서 관련 데이터베이스의 구축은 다른 바이러스들에 비하여 많이 부족한 실정이다. 국외의 연구에서도 코로나 바이러스의 유전자와 단백질을 단독으로 저장해 놓은 데이터베이스는 2007년에 개발한 ‘CoVDB’(http://covdb.microbiology.hku.hk)가 유일하지만 현재는 업데이트가 이루어지고 있지 않은 실정이다. 개발 당시, 3000개 정도의 코로나바이러스 서열 정보와 전체 계통은 264개 정도를 GenBank에서 수집하여 정보를 제공하였다. ‘MERS-CoV’가 유행 전의 구축이라서 ‘MERS-CoV’에 관한 정보는 포함되어 있지 않다(Huang et al., 2008). 국내에서는 2007~2008년 질병관리본부에서 연구 사업의 일환으로 코로나 바이러스의 데이터베이스가 구축이 되었고, 현재는 ‘VSD’(virus sequence database, http://vsd.cdc.go.kr/vsd)에 다른 바이러스와 같이 편재되었다. 2008년 당시 데이터베이스에는 4,965건의 코로나 바이러스 관련 정보를 참고문헌과 함께 제공해주었으며, 현재는 ‘VSD’에서 관련 정보를 다운로드할 수 있다.

코로나 바이러스의 데이터베이스가 단독으로 구성된 유용한 데이터베이스는 부족하지만, 여러 바이러스와 통합이 되어 정보를

제공해 주는 데이터베이스가 구축되어 있다. 대표적인 데이터베이스는 ‘VIPR’(http://www.viprbrc.org)이 있으며, 다양한 숙주에 따른 바이러스의 서열정보, 단백질구조, 항원결정기, 임상/감시 메타데이터 등을 제공해 주며 상동성 검색, 서열 정렬, 계통수 분석 등이 가능하도록 구축이 되어 있다(Pickett et al., 2012). 코로나 바이러스 이외에 인류에게 오랜 기간 동안 지속적으로 감염을 일으키고 있는 바이러스들에 관한 연구는 많이 되어 있는 상황이고, 데이터베이스의 구축도 많이 되어있다. 예를 들면 호흡기 감염의 대표적인 바이러스인 인플루엔자에 관한 데이터베이스는 외국에 11개 정도가 구축이 되어 있다. 인플루엔자 바이러스의 서열 정보, 관련 분석 도구, 역학적 정보, 숙주와의 상호작용 등 여러가지의 정보를 제공하여 연구자들에게 2차적인 연구를 유용하게 할 수 있도록 제공을 해준다. 각각의 데이터베이스에 저장되어 있는 계놈의 수와 서열 정보의 수, 제공하는 정보들의 특징과 참고문헌을 Table 1.7에 정리하였다.

코로나 바이러스와 숙주 간의 상호작용 결과를 제공해 주는 데이터베이스는 국내에는 존재하지 않으며, 외국의 경우에는 ‘VirHostNet’(http://virhostnet.prabi.fr) 과 ‘virusmentha’(http://virusmentha.Roma2.it)에서 ‘SARS-CoV’와 숙주 단백질 사이의 상호작용 데이터를 제공해 준다. 아직까지 코로나 바이러스의 면역이나 상호작용 등의 연구가 진행중에 있어 데이터도 부족하고, 데이터베이스의 구축도 부족한 실정이다. 두 데이터베이스도 코로나 바이러스에 관한 정보는 다른 바이러스들에 비하여 현저히 부족하였고, 제공해 주는 정보도 상호작용이 있는지 없는지의 여부만 확인이 가능하며, 코로나 바이러스의 면역특이성을 가지는 바이오 마커 등의 후보물질을 발굴하는 데에는 한계가 있었다. ‘VIPR’에서 코로나 바이러스 관련 항원 결정기의 정보를 제공해 주고 있으나, 숙주와의

상호작용과 관련된 데이터는 저장되어 있지 않다. 기존에 연구가 많이 이루어진 바이러스들과 숙주 간의 상호작용 데이터를 저장해 놓은 데이터베이스들의 특징과 참고문헌을 Table 1.6에 정리하였다.

1.3.2 계통수 분석법

유전자나 단백질의 서열 정보를 이용하여 생물 개체 간의 연관성을 연구하거나 진화 연구를 수행할 때 다양한 바이오인포매틱스 기법을 이용한다. 계통수의 작성은 이러한 연구들에서 가장 많이 이용되고 있으며, 계통의 분자 진화와 기능을 추론하는데 있어 점점 더 중요해 지고 있다(Felsenstein, 2003). 계통수는 생물체의 종(Organism)이나 속(Species)의 진화적인 관계를 보여주기 위한 유용한 방법이며, 관계를 보여주는 각각의 그룹을 분류군(Taxon)이라고 표현한다. 계통수는 두가지의 구성요소인 노드와 가지로 구성되어 있으며, 가지는 두 개의 노드를 연결해 준다. 노드는 외부 노드(External node)와 내부 노드(Internal node)로 구분되며, 외부 노드는 고려되는 분류 군의 일부를 의미하고 내부 노드는 서로 다른 두 노드가 공통 조상을 가지고 있음을 의미한다(Barry, 2001). 이는 각 개체 및 그룹들의 염기서열을 정렬하고 모든 자리에서 서열들 간의 유전자를 비교하여 계산된 유전적 거리를 바탕으로 줄기와 가지의 관계로 나타내어 그 연관 관계를 보여준다. 하나의 가지에 있거나 가까운 가지에 있을수록 진화적인 관계가 가까운 것을 보여준다. 계통수 작성을 위하여 아미노산이나 DNA의 서열 정렬이 필요하며, 이용되는 방법으로는 'Pairwise Alignment', 'MSA(Multiple Sequence Alignment)', 'Local Alignment', 'Global Alignment' 방법 등이 있다. 계통수를 작성하는 방법은 크게 'rooted' 계통수와 'unrooted' 계통수로 나눌 수 있고,

유전적 거리 기반 방법과 서열 내의 진화를 추적하는 특성 기반 방법으로 나눌 수 있다. ‘Rooted’ 계통수는 공통 조상이 잘 드러나는 반면 ‘unrooted’ 계통수는 공통 조상이 뚜렷하지 않다. 거리 기반 방법은 계산은 빠르지만 염기서열 간의 차이를 거리로만 환산함으로써 염기서열 자체의 정보가 사라지는 단점이 있으며, 대표적인 방법으로는 ‘NJ(Neighbor-Joining)’와 ‘UPGMA(Unweighted Pair-Group Method with Arithmetic Mean)’ 방법이 있다. 거리 기반 방법은 다중 정렬을 기반으로 하여, 분류 군들 간의 시퀀스의 차이를 거리 행렬(Distance matrix)로 작성하여 계산을 한다(Barry, 2001). ‘UPGMA’는 계층적 방법으로 분류 군끼리 군집(Cluster)을 형성하면서 거리 행렬을 작성하는 방법을 이용한다. 근원(Root)으로부터의 거리가 모든 분류 군이 같게 작성이 된다(Sokal et al., 1958). NJ방법은 거리 행렬을 작성하는 면에서는 ‘UPGMA’와 유사하지만, 군집을 형성하지 않고 내부 노드의 거리를 직접적으로 계산하여 거리 행렬을 작성하는 차이점이 있다(Barry, 2001). 특성 기반 방법으로는 ‘MP(Maximum Parsimony)’, ‘ML(Maximum Likelihood)’, ‘Bayesian Markov chain Monte Carlo’ 방법 등이 있다. 이들 방법들은 진화적 역사를 잘 예측은 할 수 있지만 매우 긴 계산 시간을 필요로 하는 단점이 있다. 이러한 다양한 분석 알고리즘을 이용하여 개발된 상용화된 계통수 분석 프로그램을 Table 1.4에 정리하였다.

1.3.3 네트워크 분석법

생명정보학은 지속적으로 확대되고, 다양한 분야에서 접목이 되고 있으며, 특히 생명 과학 연구에서 데이터의 해석과 예측 등의 방법으로 다양하게 적용이 되면서 성장하고 있다. 수많은 방법의 연구와 실험에서 얻어지는 결과물들은 서로 다른 특징의 데이터를

가지게 되고, 이러한 데이터들은 복잡성과 역동성이 증가하고 있다. 현재 이러한 서열 단백질 구조 및 단백질체학, 온톨로지, 유전자 발현 및 기타 실험 데이터들은 별개의 데이터베이스에 저장된다. 이러한 데이터베이스는 데이터의 특징에 적합한 포맷으로 정보를 분류하고 저장하기 때문에, 생물학적 시스템의 복잡성을 이해하기 위해서는 각각의 데이터베이스에 저장된 정보를 효율적인 방식으로 결합할 필요가 있다. 이러한 문제를 해결하기 위한 방법으로 네트워크 분석법이 많이 이용이 되고 있다(Lenzerini, 2002). 네트워크 분석법은 생명체를 구성하는 세포의 내부에서 일어나는 분자들 간의 조절 관계나, 세포 간의 상호작용 등을 시스템적으로 해석하는 방법이다. 복잡한 생명 현상을 네트워크의 구조적 성질과 연관 지어 분석이 가능하다. 네트워크의 종류는 크게 신호전달 네트워크, 대사 네트워크, 단백질 상호작용 네트워크, 유전자 조절 네트워크로 나눌 수 있다. 신호전달 경로는 특정 자극에 대하여 순차적으로 활성화 되는 단백질들의 나열을 나타내는 것으로, 이러한 경로는 독립적이지 않고 단백질들간의 상호작용을 통하여 복잡하게 연결되어 있다. 이러한 상호작용의 연결을 신호전달 네트워크라고 표현하며, 노드는 단백질이나 분자와 같은 신호전달 물질이며, 링크는 물질들 간의 신호전달 과정을 의미한다. 대사 네트워크란 생체 내에서 일어나는 대사작용을 통합하여 나타내는 네트워크로 대사물질을 노드로 나타내고 화학반응을 촉진하는 효소를 링크로 나타낸 네트워크이다. 단백질 상호작용 네트워크는 단백질들 간의 결합관계를 나타낸 네트워크로, 노드는 단백질을 링크는 단백질들 간의 결합을 의미한다. 단백질들 간의 결합의 방향성이 없으므로, 단순 무 방향성 네트워크가 되는 것이다. 유전자 조절 네트워크는 복잡한 유전자들 간의 조절 관계를 단순화시켜 표현하는 네트워크이다. 이러한 네트워크 분석은 바이러스스

감염 동안에 세포에서 일어나는 단백질이나 유전자들 간의 상호작용의 이해를 도울 수는 있지만, 이들이 병원성에 어떠한 영향을 미치는지에 여부를 정확하게 확인할 수는 없다. 하지만, 바이오인포매틱스적인 분석을 이용한다면, 바이러스와 숙주 사이의 상호작용에 관한 실험 결과들을 동적이고, 다변량으로의 해석을 가능하게 하여 생물학적 기전에 근접한 결과의 도출이 가능 한 것이다. 네트워크 분석을 위한 다양한 응용 프로그램이 개발이 되어 있으며, Table 1.7 에 정리하였다. 이를 이용한 다양한 분석은 바이러스의 감염의 범위를 넓혀주어 바이러스의 병원성의 이해의 폭을 넓혀줄 것이며, 항 바이러스 치료를 위한 새로운 표적의 제시가 가능할 것이다.

1.4 연구의 필요성

코로나 바이러스의 대유행이 2000년부터 시작이 되어 연구기간이 충분하지 못했고, 코로나 바이러스가 고병원성의 특징 때문에 연구의 여러 가지 제약과 한계점이 있어 실험 등의 연구가 활발히 이루어지지 못하고 있는 실정이다. 한국에서도 코로나 바이러스의 연구는 미진하였으나, 2015년에 ‘MERS-CoV’의 유행이 이루어지면서 여러 연구들이 진행되고 있다. 하지만, 대부분이 코로나 바이러스의 감시 체계나 대응 방안 등의 역학적인 연구가 주를 이루고 있는 실정으로 바이오인포매틱스 기법을 활용한 코로나 바이러스의 감염원성이나 면역기전 등의 연구가 많이 이루어져야 할 것이다. 백신의 연구 또한 주로 S 단백질을 중심으로 이루어지고 있어, 다른 비구조 단백질들의 연구도 병행되어야 할 것이다. 이러한 연구는 아직까지 개발하지 못한 치료제와 백신의 개발에 기초 자료로 이용이 될 수 있을 것이다. 인체에서 일어나는 면역반응은 다양한 생화학 분자들이 작용하여 만들어 내는 거대한 네트워크이다. 면역반응을 이해하기 위하여 연구자들은 실험을 통하여 많은 데이터들을 축적해 오고 있다. 컴퓨터를 활용한 바이오인포매틱스 기법이 면역학에 도입이 되면서 면역반응의 많은 실험 자료들을 저장하여 분석이 가능하게 되었고, 이를 바탕으로 전체 면역체계를 이해하기 위한 다양한 분석 도구들이 개발이 되고 있다. 아직 우리나라는 이러한 부분에서 외국에 뒤쳐지고 있는 것이 사실이다. 하지만, 우리나라의 강력한 IT 인프라를 이용하여 연구에 박차를 가한다면 외국의 연구결과들에 뒤지지 않을 결과가 보고될 수 있을 것이다.

최근에는 분자생물학적인 기술의 발달로 인하여 바이러스의 분자구조, 단백질의 종류와 그 기능 등을 분석한 정보가 쏟아져

나오고 있다. 면역관련 유전자와 단백질들도 그 서열이나 기능이 밝혀지고 있으며, 관련 유전자들의 신호전달경로 등도 많이 증명되고 있으나 그 정보들이 산재 되어 있다. 물론 연구자들이 ‘NCBI’의 ‘GenBank’나 ‘Uniprot’과 같은 데이터베이스를 이용하여 코로나 바이러스 관련 유전자와 면역관련 인자들의 유전 정보나 서열 정보 등을 얻을 수 있다(Figure 1.1). 하지만, 이는 모든 생물종의 유전자 서열이나 비슷한 유전자들의 서열이 혼합되어 산재되어 있기 때문에, 원하는 유전자의 서열이나 단백질을 수집하는 데에는 상당한 노력과 시간이 소모된다. 또한 현재 구축된 코로나 바이러스 관련 데이터베이스는 모든 바이러스를 모아 놓은 데이터베이스의 한 부분을 차지하고 있기 때문에, 코로나 바이러스의 연구를 위해서는 서열 정보 및 특징 등을 확인하고 수집하는데 여러 단계의 작업을 수행해야 한다(Figure 1.2). 이러한 노력과 시간을 절약하기 위해서는 코로나 바이러스의 구조적인 특징 및 서열 정보의 특징 등에 따라, 분류하고 저장하여 2차 연구에 쉽게 이용할 수 있는 유용한 데이터베이스의 구축이 필요하다. 또한, 바이러스의 면역관련 데이터베이스는 그 수가 제한적이다. 면역관련 단백질과 유전자는 인종 별, 지역 간의 차이가 존재하기 때문에 우리나라의 특성에 적합한 데이터를 수집하고 분류하여 저장하는 데이터베이스의 구축은 코로나 바이러스의 면역 특이성 연구를 함에 있어 선행되어야 할 연구이다. 더 나아가서는 우리나라에서 유행하는 주요 병원체에 대하여 한국인에 적합한 면역 특이성 인자 등의 발굴을 위해서도 필요한 연구이다. 앞으로 바이러스와 면역기전의 이해에 있어 바이오인포매틱스 기법을 활용한 접근은 반드시 필요한 연구 방법이 될 것이다. 면역 정보학은 전통적인 면역학에 컴퓨터 과학, 수학, 통계, 화학, 생화학, 유전체학, 단백질체학등의 연구를

결합시킨 분야이다. 아직까지는 면역 정보학이 발달 초기 단계에 있지만, 이러한 연구 방법은 명확한 면역학 연구를 가속화 시킬 수 있을 것이다. 면역학 연구에 전산 모델의 제시는 실험실에서 적절한 실험 설계 등에 이용이 되고, 전통적인 방법으로는 증명할 수 없었던 가설들의 검증에 유용한 방법으로 이용이 될 수 있을 것이다. 이러한 연구는 면역기전의 대단위 분석을 가능하게 해주고, 시스템 생물학과 연관되어 증명되지 않은 많은 생명현상을 이해하기 위한 기초 자료로 반드시 필요한 연구이다.

1.5 연구의 목적

한국은 2015년 ‘MERS-CoV’의 전파를 막기 위하여 수천 명의 보건 당국자들이 노력을 하였음에도 불구하고, 186명의 감염자와 36명의 사망자, 166,000명 정도의 격리자가 발생이 되면서 막대한 사회경제적인 손실을 야기하였다(Choi, 2015). 이는 현재까지 사우디아라비아 지역을 제외하고 가장 많은 감염자와 사상자수를 발생 시킨 사례로 보고 되었다(WHO, 2015). ‘MERS’의 전파 양상이 다른 나라와는 차이가 있어 이에 대한 다양한 연구가 진행되고 있으나, 우리나라의 특징에 적합한 연구가 많이 부족하며, 이에 관한 연구가 활발히 진행되어야 할 것이다. 감염된 사람에게서 다른 사람으로 전파가 가능하게 된 심각한 코로나 바이러스의 유행은 계속 출현을 할 것이며, 이러한 유행을 조절하기 위해서는 ‘MERS-CoV’ 질환과 면역반응의 이해가 있어야 한다(Rahul et al., 2016). 우리나라에서의 ‘MERS-CoV’ 전파 속도가 신속했던 이유가 바이러스의 변이가 아닌 여러 가지 환경적인 요인이나 환자의 영양상태, 건강 상태, 개개인이 가지고 있는 유전적 소인의 차이로 인한 면역반응의 차이가 원인일 것이라 예상할 수 있다. 또한, 같은 환경 조건에서도 바이러스에 대한 대처 능력이 다르게 나타나고, 감염의 완급에 차이가 있었던 것은 개개인의 면역반응의 차이라고 생각된다. 따라서 코로나 바이러스에 관한 면역 기전의 이해가 아직까지 개발되지 못한 치료제나 백신의 연구에 도움이 될 것이다. 이에 본 연구에서는 코로나 바이러스의 유전정보와 함께 숙주세포의 면역에 관여하는 산재되어 있는 유전정보를 수집하고, 체계적으로 분류하고 저장하여 사용자가 손쉽게 이용할 수 있는 데이터베이스를 구축함으로써, 이를 기반으로 한 2차 연구에 유용하게 사용할 수 있도록 하는 것에 목적이 있다. 바이오인포매틱스 방법을 이용한

코로나 바이러스의 면역 특이성 연구를 통하여 바이러스의 유전자와 단백질의 변이 여부 등을 분석함으로써, 새롭게 출현하거나 유행을 일으킬 수 있는 가능성이 있는 신변종 바이러스를 예측하여, 지속적으로 감시하고, 유행이 발생하였을때를 대비한 대책을 미리 수립함으로써 ‘MERS’ 사태와 같은 상황이 일어나지 않도록 하고, 백신이나 치료제의 개발 등에 기여하는 것에 그 목적이 있다. 또한 전통적인 면역학 연구에서 소모되는 시간과 노력을 단축 시켜주고 실험 등으로 증명하지 못하는 부분의 해석을 가능하게 해줌으로써, 아직 밝혀지지 않은 수많은 인체내 면역반응의 이해에 유용한 기초 자료를 제공해 주는데 목적이 있다. 더 나아가서는 이러한 연구 결과를 기반으로 개인 간의 유전적 소인에서 비롯되는 바이러스의 감수성 여부의 차이 등을 이해하고 규명하는 기반을 제공함으로써 개인 맞춤형 면역치료제 등의 개발에 기초자료의 역할을 할 것이다.

연구의 목적을 요약하면, 첫째 코로나 바이러스와 면역관련 단백질과 유전자의 정보들을 제공하는 데이터베이스를 구축하여 연구자들에게 용이한 정보를 제공해 주고자 하며, 둘째 바이오인포매틱스 방법을 활용한 연구 방법을 통하여 기존에 알려지지 않았던 코로나 바이러스의 위험성을 예측하고, 예측된 바이러스에 대한 변이 여부의 지속적인 감시 등의 필요성을 제안하고자 하고, 셋째, 코로나 바이러스의 면역 특이성 분석을 통하여, 코로나 바이러스와 면역단백질들을 재조명 함으로써 아직 밝혀지지 않은 면역기전의 이해의 실마리를 제공해 주고, 개발되지 않은 치료제와 예방백신의 개발에 도움이 되고자 한다. 더 나아가서는 개인의 면역반응의 차이에 적합한 개인 맞춤형 치료제의 개발에도 도움을 주고자 한다.

GenBank - Send -

Change region shown ▾

Customize view ▾

Middle East respiratory syndrome coronavirus isolate MERS-CoV/KOR/CNUH_SNU/122_06_2015 spike protein (S) gene, partial cds

GenBank: KT717706.1
[FASTA](#) [Graphics](#) [PopSet](#)

[Go to:](#) [\(0\)](#)

LOCUS KT717706 615 bp RNA linear VRL 03-OCT-2015
 DEFINITION Middle East respiratory syndrome coronavirus isolate
 MERS-CoV/KOR/CNUH_SNU/122_06_2015 spike protein (S) gene, partial
 cds.
 ACCESSION KT717706
 VERSION KT717706.1 GI:930625839
 KEYWORDS .
 SOURCE Middle East respiratory syndrome coronavirus (MERS-CoV)
 ORGANISM Middle East respiratory syndrome coronavirus [MERS-CoV]
 Viruses; ssRNA viruses; ssRNA positive-strand viruses, no DNA
 stage; Nidovirales; Coronaviridae; Coronavirinae; Betacoronavirus;
 unclassified Betacoronavirus.
 REFERENCE 1 (bases 1 to 615)
 AUTHORS Kim, Y.-S. and Cho, N.-H.
 TITLE Outbreak of Middle East Respiratory Syndrome in Two Hospitals in
 South Korea
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 615)
 AUTHORS Kim, Y.-S. and Cho, N.-H.
 TITLE Direct Submission
 JOURNAL Submitted (07-SEP-2015) Department of Microbiology and Immunology,
 Seoul National University College of Medicine, 103 Daehak-ro,
 Jongno-gu, Seoul, Seoul 151-747, South Korea
 COMMENT ##Assembly-Data-START##
 Sequencing Technology: :: Sanger dideoxy sequencing
 ##Assembly-Data-END##
 FEATURES
 Location/Qualifiers
 source 1..615
 /organism="Middle East respiratory syndrome coronavirus"
 /mol_type="genomic RNA"
 /isolate="MERS-CoV/KOR/CNUH_SNU/122_06_2015"
 /isolation_source="sputum"
 /hosts="Homo sapiens"
 /db_xref="taxon:32339526"
 /country="South Korea"
 /collection_date="10-Jun-2015"
 /PCR_primers="fwd_name: vs904, fwd_seq:
 tcagggtgacccaactcaatg, rev_name: vs905, rev_seq:
 ttggatgaagccaccacgct"
 .-.-615
 ##Name
 /gene="S"
 .-.-615
 /gene="S"
 /note="MERS-CoV spike protein"
 /codon_start=1
 /product="spike protein"
 /protein_id="BL308424.1"
 /db_xref="GI:930625839"
 /translation="SYFLSLPTNFSQVYTGVIQTITQIVTVDCIQVVCNFPQICEQ
 LLREYQFCSTENGLNANRQDQRRFLASVIVSQGSPZPFDQNLTLLEPV
 S1STGSRARSATZEDLLFDVITADPSVHQVDQVQQPQASARDLICQVVAQVYVLL
 PPLFDVNEAAVYSSLGSIADVQVITAGLSFAATPFAQGSFVRL"
 ORIGIN
 1 agttattttt aattaaagat acctactaat ttttctcttg gtgtgactca ggaatcatt
 61 cagcaaccca ttccagaagt tccgtgtgat gttcaacagt agttttgcaa tggtrttccg
 121 agttgtgagc aattacagcc cagatctgag cagttttgct ccaaatataa ccaggtcttc
 181 catggtgcca atttaagcca ggaatgattc tgaatgatt tgtttcgag cgtgaaagc
 241 tctctctcat cctctctacc accaggtttt ggaatgact tcaattgag actttctgaa
 301 cctctctcca tctctctacc cagttgtag gctctctgag ctattctgag ttgtctctt
 361 gacaaagtca ctatagctga tccctgttat atgcaaggtc agatgattg tatcagcaa
 421 gttctccgac cagctctgca tccctattgt gctctctatg tgtctggtta taagctata
 481 cctctctctt tggatgttaa tatgagacc gactacactt catctttgct tggcagata
 541 gctgggtgtg cctggaatgc tggcttatcc tcccttgctg ctattctcatt tgcacagat
 601 attctctata ggtta
 //

Analyze this sequence ▾
 Run BLAST
 Pick Primers
 Highlight Sequence Features
 Find in this Sequence

MERS Coronavirus Resource ▾
 Retrieve, view, and download MERS coronavirus
 genomic and protein sequences.

Related information ▾
 PopSet
 Protein
 Taxonomy

LinkOut to external resources ▾
 VPR - Virus Pathogen Resource [Virus Pathogen Resource]
 Order S cDNA clone/Protein/Antibody/RNA [OrfGene]

Recent activity ▾
 Middle East respiratory syndrome
 coronavirus isolate MERS-
 Nucleotide
 See more...

Figure 1.1 GenBank data before processing. The form of initial data obtained from the GenBank before carrying out this study. Gene sequences contain various kinds of information.

a) FASTA format

```

1 nsp1|coronavirus_nsp1_FL1_PRO_FL2_PRO_HD|KF923918|Human|China|Betacoronavirus_1*
2 GGACGGCGTGTACATTAAAGAACAGCCTACAGTAAGAGAGATTAAAGCAATCTTAAGATTATTAAGG
3 TTTTTATGAGCTTGACAAATGATTTAAATACTATTTAAATACTGGCGTGGAGTGTTTGAGGTGGATCA
4 TACTGTGGATTTGAGGAAATTTATGCTGTGGTGTATGATCCATAGAAAGGAAACTTCTCCATTTAG
5 GAGCTTGAAGGTGTAGGTGCTAAAGTTAGTGCCTTTTACAGAAATTAGAGGATAATCCCTATTTTAT
6 TGTATGAGCTGGCAGAGAGTTTTCCTCCTAAATGTATTTGGCTTTACAGCCTCTGAAGATGATGA
7 CTTTCTTTGAGGAAAGTATGTTGAGAGAGATGATGTAGAAAGTGTAGGAAACTGTATTTACTATACAA
8 GCTGGACACCTTTGTTTGTCTAGTGAACAGGAGAACTTCTGAAGTCTTAGAGGACACTTGTGATGAT
9 GTCCAAAGTGTGGAGACATCTGATTCACAAGTGAAGAGAGTGAAGAAATGTCCGATTTTGTGATCTGA
10 ATCTGTGATTCAGGATTAAGAAATGTTTGTGAGTTTTACTACAGAGCCAGAAATTTGTTAAAGTT
11 TTGGCTCTGATTTGGCTAAAGCAACTCGCAACAATTTGCTGGTTCGATGATTTGGCAGTGTATGCGA
12 AATGGCTTGTCAATTTAAAGATAAAATTTGAGAGTCTTTGGGTGTTATACAGCAACAGTATGCTCA
13 GTTGTGTTGATACCTTGGTAAAGATACCTGCTATATTTGACTTCCACAAAGTGGTATGTTGCT
14 GATTTTGCATATTTGTTTTAACCTTATGTGATTTGGCAGTGTGTCATACTGGAAATGCATTTAAAT
15 ATTAGCTCTTAAGCTTAAAGCTTGGATGCTAGTGTCTTTTATGGTGAATGTGCTCACATATATGCA
16 GTGGGTGAGTCTAAGTACTTATGATGTGTGATGTCGCAATTTACAGCCACTTGTCTCTTAAAGAT
17 TTGTTTGTGCTATTTTACTAGAGGCTATGTGTGATTAAGCACTTGTGTTGGTGTAAATGATGAT
18 ATCTATGCTTGTGTTGATGGTAAACAAATGATGATCACTGATCACTAGTATTACTAGTAAAGT
19 TGATTTTATTTTGGCAGGATGATGCTATTTCAATGACTACTTTTGAATTTGCCCAATTTGATGGT
20 TGTATAAGCACTTATGTTGTTTTTGTAAAGTGTGATTAATTAAGTGTCTTAAGCTTGTTTAAAG
21 TTGTTTAAACCCGTAATGGCCATATGGCAATGGTACATGGTGGTGGTGTGCAAAAGCTATGCAAG
22 TGGACAGCAGTTTGTAAAGAGACTACCAATATGGTAAAGTCTAAAGGAGTGTGCTACTGGAGAT
23 TATCTCTACAGGGGGCAAAATTTGTTAAACTGTGCTTAAATGTGTTGGACCTGATGGCAGACACAG
24 GTAACAAAGTATGATTTTGTAGAGCTGTTTAAACATTTTAAACACTATGACTGTGTTTACAC
25 TTGATCTCAGCTGGATATTTAGTGTGCTCTGATGATGCTTTAACATACTACTTGGTACTGCTAAG
26 AAGCAAGTGTCTGTGTTGCTAATAACAGAGATTTTGAATCTTATTTCTAAAGTCTGAGATGCTG
27 TTGAGGCACTAAGAAATTTGGCAGCGCTCTTCCCTTAAATGTTGGACGTTCCATGTTTACGGAA
28 TGCTAATAGTGTATTTAATCAATGATGATTTGCAATTTGTCGACATTTAAGTGTTCAGAGTGT
29 TCTTAAAGACTGATTAAGCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG
30 CTGAGCTTGGCGTGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG
31 TACTGGGCACTAGATATTTGAGCCAAAGATAAAGTTTTGGTATGATACAGCAGGATTTTAAATA
32 GCTACTGTGACTTAAAGTAAAGGCTTGTGTTTGGATTAAGTGGACACTCTGCTAAGTGTGATG
33 ATTTGCTATTAAGTGTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTG
34 AGTTAAGTCCAGGAAGTAAAGTGTGATATAAATATAAAGTAAAGCTTTTCCAGTTTGAATAACT
35 TCTAGTGAAGATTTAAAGGCTTAAGAAAGTTCCTTAAATTTGATCAGAAAGAAATGCTTGGCTAT
36 ACAAGTCTTAAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT
37 TAACAATGTTTTTAAAGTGTGCTTAAAGTGTGCTTAAAGTGTGCTTAAAGTGTGCTTAAAGT
38 TGGCAAGGACAGTGGCTTAAATTCGTTTGGCCGCTGCTAGATTTGATGATTTGGTTTGGCTA
39 GTGGTTTTAAATTTGGAGTGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG
  
```

b) SQL format

id	name	accession	organism	protein	sequence
1	nsp1	coronavirus_nsp1_FL1_PRO_FL2_PRO_HD_KF923918	Betacoronavirus_1	Human: China	GGACGGCGTGTACATTAAAGAACAGCCTACAGTAAGAGAGATTAAAGCAATCTTAAGATTATTAAGG
2	nsp1	coronavirus_nsp1_FL1_PRO_FL2_PRO_HD_KF923918	Betacoronavirus_1	Human: China	TTTTTATGAGCTTGACAAATGATTTAAATACTATTTAAATACTGGCGTGGAGTGTTTGAGGTGGATCA
3	nsp1	coronavirus_nsp1_FL1_PRO_FL2_PRO_HD_KF923918	Betacoronavirus_1	Human: China	TACTGTGGATTTGAGGAAATTTATGCTGTGGTGTATGATCCATAGAAAGGAAACTTCTCCATTTAG
4	nsp1	coronavirus_nsp1_FL1_PRO_FL2_PRO_HD_KF923918	Betacoronavirus_1	Human: China	GAGCTTGAAGGTGTAGGTGCTAAAGTTAGTGCCTTTTACAGAAATTAGAGGATAATCCCTATTTTAT
5	nsp1	coronavirus_nsp1_FL1_PRO_FL2_PRO_HD_KF923918	Betacoronavirus_1	Human: China	TGTATGAGCTGGCAGAGAGTTTTCCTCCTAAATGTATTTGGCTTTACAGCCTCTGAAGATGATGA
6	nsp1	coronavirus_nsp1_FL1_PRO_FL2_PRO_HD_KF923918	Betacoronavirus_1	Human: China	CTTTCTTTGAGGAAAGTATGTTGAGAGAGATGATGTAGAAAGTGTAGGAAACTGTATTTACTATACAA
7	nsp1	coronavirus_nsp1_FL1_PRO_FL2_PRO_HD_KF923918	Betacoronavirus_1	Human: China	GCTGGACACCTTTGTTTGTCTAGTGAACAGGAGAACTTCTGAAGTCTTAGAGGACACTTGTGATGAT
8	nsp1	coronavirus_nsp1_FL1_PRO_FL2_PRO_HD_KF923918	Betacoronavirus_1	Human: China	GTCCAAAGTGTGGAGACATCTGATTCACAAGTGAAGAGAGTGAAGAAATGTCCGATTTTGTGATCTGA
9	nsp1	coronavirus_nsp1_FL1_PRO_FL2_PRO_HD_KF923918	Betacoronavirus_1	Human: China	ATCTGTGATTCAGGATTAAGAAATGTTTGTGAGTTTTACTACAGAGCCAGAAATTTGTTAAAGTT
10	nsp1	coronavirus_nsp1_FL1_PRO_FL2_PRO_HD_KF923918	Betacoronavirus_1	Human: China	TTGGCTCTGATTTGGCTAAAGCAACTCGCAACAATTTGCTGGTTCGATGATTTGGCAGTGTATGCGA
11	nsp1	coronavirus_nsp1_FL1_PRO_FL2_PRO_HD_KF923918	Betacoronavirus_1	Human: China	AATGGCTTGTCAATTTAAAGATAAAATTTGAGAGTCTTTGGGTGTTATACAGCAACAGTATGCTCA
12	nsp1	coronavirus_nsp1_FL1_PRO_FL2_PRO_HD_KF923918	Betacoronavirus_1	Human: China	GTTGTGTTGATACCTTGGTAAAGATACCTGCTATATTTGACTTCCACAAAGTGGTATGTTGCT
13	nsp1	coronavirus_nsp1_FL1_PRO_FL2_PRO_HD_KF923918	Betacoronavirus_1	Human: China	GATTTTGCATATTTGTTTTAACCTTATGTGATTTGGCAGTGTGTCATACTGGAAATGCATTTAAAT
14	nsp1	coronavirus_nsp1_FL1_PRO_FL2_PRO_HD_KF923918	Betacoronavirus_1	Human: China	ATTAGCTCTTAAGCTTAAAGCTTGGATGCTAGTGTCTTTTATGGTGAATGTGCTCACATATATGCA
15	nsp1	coronavirus_nsp1_FL1_PRO_FL2_PRO_HD_KF923918	Betacoronavirus_1	Human: China	GTGGGTGAGTCTAAGTACTTATGATGTGTGATGTCGCAATTTACAGCCACTTGTCTCTTAAAGAT
16	nsp1	coronavirus_nsp1_FL1_PRO_FL2_PRO_HD_KF923918	Betacoronavirus_1	Human: China	TTGTTTGTGCTATTTTACTAGAGGCTATGTGTGATTAAGCACTTGTGTTGGTGTAAATGATGAT
17	nsp1	coronavirus_nsp1_FL1_PRO_FL2_PRO_HD_KF923918	Betacoronavirus_1	Human: China	ATCTATGCTTGTGTTGATGGTAAACAAATGATGATCACTGATCACTAGTATTACTAGTAAAGT
18	nsp1	coronavirus_nsp1_FL1_PRO_FL2_PRO_HD_KF923918	Betacoronavirus_1	Human: China	TGATTTTATTTTGGCAGGATGATGCTATTTCAATGACTACTTTTGAATTTGCCCAATTTGATGGT
19	nsp1	coronavirus_nsp1_FL1_PRO_FL2_PRO_HD_KF923918	Betacoronavirus_1	Human: China	TGTATAAGCACTTATGTTGTTTTTGTAAAGTGTGATTAATTAAGTGTCTTAAGCTTGTTTAAAG
20	nsp1	coronavirus_nsp1_FL1_PRO_FL2_PRO_HD_KF923918	Betacoronavirus_1	Human: China	TTGTTTAAACCCGTAATGGCCATATGGCAATGGTACATGGTGGTGGTGTGCAAAAGCTATGCAAG
21	nsp1	coronavirus_nsp1_FL1_PRO_FL2_PRO_HD_KF923918	Betacoronavirus_1	Human: China	TGGACAGCAGTTTGTAAAGAGACTACCAATATGGTAAAGTCTAAAGGAGTGTGCTACTGGAGAT
22	nsp1	coronavirus_nsp1_FL1_PRO_FL2_PRO_HD_KF923918	Betacoronavirus_1	Human: China	TATCTCTACAGGGGGCAAAATTTGTTAAACTGTGCTTAAATGTGTTGGACCTGATGGCAGACACAG
23	nsp1	coronavirus_nsp1_FL1_PRO_FL2_PRO_HD_KF923918	Betacoronavirus_1	Human: China	GTAACAAAGTATGATTTTGTAGAGCTGTTTAAACATTTTAAACACTATGACTGTGTTTACAC
24	nsp1	coronavirus_nsp1_FL1_PRO_FL2_PRO_HD_KF923918	Betacoronavirus_1	Human: China	TTGATCTCAGCTGGATATTTAGTGTGCTCTGATGATGCTTTAACATACTACTTGGTACTGCTAAG
25	nsp1	coronavirus_nsp1_FL1_PRO_FL2_PRO_HD_KF923918	Betacoronavirus_1	Human: China	AAGCAAGTGTCTGTGTTGCTAATAACAGAGATTTTGAATCTTATTTCTAAAGTCTGAGATGCTG
26	nsp1	coronavirus_nsp1_FL1_PRO_FL2_PRO_HD_KF923918	Betacoronavirus_1	Human: China	TTGAGGCACTAAGAAATTTGGCAGCGCTCTTCCCTTAAATGTTGGACGTTCCATGTTTACGGAA
27	nsp1	coronavirus_nsp1_FL1_PRO_FL2_PRO_HD_KF923918	Betacoronavirus_1	Human: China	TGCTAATAGTGTATTTAATCAATGATGATTTGCAATTTGTCGACATTTAAGTGTTCAGAGTGT
28	nsp1	coronavirus_nsp1_FL1_PRO_FL2_PRO_HD_KF923918	Betacoronavirus_1	Human: China	TCTTAAAGACTGATTAAGCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG
29	nsp1	coronavirus_nsp1_FL1_PRO_FL2_PRO_HD_KF923918	Betacoronavirus_1	Human: China	CTGAGCTTGGCGTGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG
30	nsp1	coronavirus_nsp1_FL1_PRO_FL2_PRO_HD_KF923918	Betacoronavirus_1	Human: China	TACTGGGCACTAGATATTTGAGCCAAAGATAAAGTTTTGGTATGATACAGCAGGATTTTAAATA
31	nsp1	coronavirus_nsp1_FL1_PRO_FL2_PRO_HD_KF923918	Betacoronavirus_1	Human: China	GCTACTGTGACTTAAAGTAAAGGCTTGTGTTTGGATTAAGTGGACACTCTGCTAAGTGTGATG
32	nsp1	coronavirus_nsp1_FL1_PRO_FL2_PRO_HD_KF923918	Betacoronavirus_1	Human: China	ATTTGCTATTAAGTGTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTG
33	nsp1	coronavirus_nsp1_FL1_PRO_FL2_PRO_HD_KF923918	Betacoronavirus_1	Human: China	AGTTAAGTCCAGGAAGTAAAGTGTGATATAAATATAAAGTAAAGCTTTTCCAGTTTGAATAACT
34	nsp1	coronavirus_nsp1_FL1_PRO_FL2_PRO_HD_KF923918	Betacoronavirus_1	Human: China	TCTAGTGAAGATTTAAAGGCTTAAGAAAGTTCCTTAAATTTGATCAGAAAGAAATGCTTGGCTAT
35	nsp1	coronavirus_nsp1_FL1_PRO_FL2_PRO_HD_KF923918	Betacoronavirus_1	Human: China	ACAAGTCTTAAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT
36	nsp1	coronavirus_nsp1_FL1_PRO_FL2_PRO_HD_KF923918	Betacoronavirus_1	Human: China	TAACAATGTTTTTAAAGTGTGCTTAAAGTGTGCTTAAAGTGTGCTTAAAGTGTGCTTAAAGT
37	nsp1	coronavirus_nsp1_FL1_PRO_FL2_PRO_HD_KF923918	Betacoronavirus_1	Human: China	TGGCAAGGACAGTGGCTTAAATTCGTTTGGCCGCTGCTAGATTTGATGATTTGGTTTGGCTA
38	nsp1	coronavirus_nsp1_FL1_PRO_FL2_PRO_HD_KF923918	Betacoronavirus_1	Human: China	GTGGTTTTAAATTTGGAGTGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG
39	nsp1	coronavirus_nsp1_FL1_PRO_FL2_PRO_HD_KF923918	Betacoronavirus_1	Human: China	GTGGTTTTAAATTTGGAGTGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG

Figure 1.2 GenBank data after processing using JavaScript. GenBank data are changed using JavaScript in the a) FASTA format of nsp1 and in the SQL format for the table, which has information on coronavirus non-structural proteins.

Table 1.1 Classification of coronaviridae.

Subfamily	Genus	Species	
Coronavirinae	Alpha corona virus	HCoV-229E* HCoV-NL63*	
	Beta corona virus	Lineage A	HCoV-HKU1*
		Lineage B	HCoV-OC43*
		Lineage C	SARS*
		Lineage D	MERS*
	Gamma corona virus		Btcov-HKUg† Avian corona virus†, Swine corona virus†
		Delta corona virus	BuCoV- HKU11†, ThCoV- HKU12†, MunCoV-HKU13†
Torovirinae	Bafinivirus		
	Torovirus		

Coronaviridae is divided two subfamilies and six genera. Beta coronavirus is divided four lineages (a, b, c, d).

*Host is human

† Host is animal

Table 1.2 Functions of coronavirus structural proteins.

Protein	Function	References
S	Receptor interaction, cell-to-cell fusion, viral entry	Weiss et al.,(2005)
M	Virion membrane, viral assembly, infectivity	Haan et al., (2003)
N	Bind the viral genome in a beads on a string conformation	Chang et al.,(2006)
	Induces a cell cycle delay or arrest.	Wurm et al.,(2001)
E	Viral assembly, forms virus-like particle with M protein	Vennema et
HE	Acetyl-esterase activities enhance cell entry and virus	Kazi et al., (2005)

Coronavirus particles contain five main structural proteins. These are the S(spike), M(membrane), N(nucleocapsid), and HE(hemagglutinin-esterase) proteins, all of which are encoded within the 3' end of the viral genome.

Table 1.3 Functions of coronavirus non-structural proteins (nsps)

Protein	Function.	References
nsp1	mRNA degradation of host cell	Huang et al.,(2011)
nsp2	Blocking the binding of the protein	Graham et al.,(2005)
nsp3	Interact with N protein cytokine expression Blocks host innate immune response	Eriksson et al.,(2006)
nsp4	Important for proper structure of DMVs	Clementz et al.,(2008)
nsp5	Cleaves viral polyprotein	Lu et al.,(1995)
nsp6	Potential transmembrane scaffold protein	Oostra et al.,(2008)
nsp7	Forms complex with nsp8, may act RNA polymerase	Zhai et al.,(2005)
nsp8	Forms complex with nsp7, may act RNA polymerase	Imbert et al.,(2006)
nsp9	RNA binding protein	Egloff et al.,(2004)
nsp10	Cofactor for nsp16 and nsp14	Bouvet et al.,(2010)
nsp12	RdRp	Xu et al.,(2003)
nsp13	RNA helicase, 5' triphosphatase	Ivanov et al.,(2004)
nsp14	3'-5' exoribonuclease, ExoN; proofreading of viral	Eckerle et al.,(2007)
nsp15	Viral endoribonuclease, NendoU	Bhardwaj et al.,(2006)
nsp16	2'-O-MT; shields viral RNA from MDA5 recognition	Decroly et al.,(2006)

For a list of non-structural proteins and their proposed function. Ribonucleases nsp15(NendoU) and nsp14 (ExoN) are unique function of nidovirales(Snijder et al.,2003)

Table 1.4 Function of SARS-CoV and MERS-CoV innate immune antagonists.

Protein	Function	References
Nsp1	Antagonizes IFN by block phosphorylation of STAT1 inactivation of host translational machinery degrades host mRNA	Huang et al.,(2011) Wathelet et al.,(2007) Kamitani et al.,(2009)
Nsp3	Antagonizes IFN I production by deubiquitinating IRF3 and prevents its nuclear translocation	Chen et al.,(2014)
Nsp7	Antagonizes IFN by block RIG-I signaling pathway	Totura., (2014)
NSP16	Preventing recognition and activation of cell intrinsic pathway by viral 2'-O-methyltransferase (2'-O-MTase)	Menachery et al.,(2014)
PLpro	Antagonizes IFN by block phosphorylation IRF3	Devaraj et al.,(2007)
(PLP)	Antagonizes IFN by interacts with STING. Antagonizes IFN by block NF-κB signaling pathway.	Frieman et al.,(2009) Sun et al.,(2012)
ORF3b (SARS)	Antagonizes IFN(β) by inhibiting RIG-I and MAVS	Freundt et al.,(2009)
ORF6 (SARS)	Antagonizes IFN by blocking JAK-STAT signaling	Frieman et al., (2007)
M	Antagonizes IFN by blocking the RIG-I signaling pathway.	Siu et al., (2009)
N	Antagonizes IFN(β) by block NF-κB signaling pathway Antagonizes IFN(β) inhibits activator protein 1 (AP1) and protein kinase R (PKR) function	Kopecky-Bromberg et al., (2007) Raj et al.,(2013)
ORF4a (MERS)	Antagonizes IFN by blocking MDAs signaling pathway.	Niemeyer et al.,(2013)
ORF4b (MERS)	Antagonizes IFN by blocking NF-κB signaling pathways	Matthews et al., (2014)

Transcription and subsequent signaling of interferon is vital for activating the antiviral response in host cells. Because of this, SARS-CoV and MERS-CoV encode proteins that antagonize the IFN response to viral infection.

Table 1.5 List of phylogenetic tree tools.

Name	Method
PAUP	Maximum parsimony,distance,Maximum likelihood
PHYLIP	Maximum parsimony,distance,Maximum likelihood
PAML	Maximum likelihood
MrBayes*	Bayesian inference
BEAST	Bayesian inference, relaxed molecular clock, Demographic
MEGA	Distance, Parsimony and Maximum composite likelihood

Table 1.6 List of Influenza virus related databases

Database	Num. Of genome	Num. Of sequence	Feature	reference
NCBI-IVR	17,784	297,660	Provides access to influenza virus and analysis tools.	Bao et al., (2008)
IRD	17,988	306,164	Provides influenza virus host-pathogen interaction.	Squires et al., (2006)
FluGenome	7,135	118.810	For genotyping influenza A and analyzing reassortment,	Lu et al., (2007)
IVDB*	~3,000	43,875	Hosts influenza virus sequence and analysis tools	Chang et al., (2007)
EpiFlu	16,144	367,672	Comprehensive of influenza virus sequences containing clinical epidemiological data	Bogner et al., (2006)
OpenFluDB	11,433	232,688	Contains influenza virus type, host, and collection date/place	Liechti et al., (2010)
ISED	-	181,552	Influenza virus sequence and epitope from Asia	Yang et.al., (2009)
IPDR	-	305,990	Designing primers/probes for influenza diagnostics	Bose et.al., (2008)
FluTE†	-	-	Influenza epidemic simulation model	Chao et.al., (2010)
GiRaF†	-	-	Identification of influenza virus reassortments	Nagarajan et.al., (2011)
ATIVS*	-	-	Analysis tools for influenza virus surveillance	Liao et.al., (2009)

*Has not been updated for more than 2 years.

† Download available database.

Table 1.7 List of virus-host interaction databases

Database	Feature	Reference
HBVPathDB	HBV infection-related molecular interaction pathways and network.	Zhang et al., (2005)
VirusMINT	Management and analysis of interactions between viral and human proteins	Chatr-aryamontri et al., (2009)
VirHostNet	Management and analysis of interactions between viral and human proteins	Navratil et al., (2009)
NCBI-HHPID	Catalog all interactions between HIV-1 and human proteins	Fu et al., (2009)
CAPIH	Comparative analyses of host-HIV-1 protein-protein interactions in human and three model animals (chimpanzee, rhesus macaque, and mouse)	Lin et al., (2009)
phEVER	Provides evolutionary and phylogenetic information for exploring virus-host and virus-virus lateral gene	Palmeira et al., (2011)

— Virus-host interaction database is to understand the complex interplay between viral and host cellular proteins.

Table 1.8 Criteria for the assessment of visualization tools

Name	Visualization	Compatibility	Function	Strength	Reference
Medusa	2D medium scale	Own format	Highly interactive	Multi-edge connection	Hooper et al.(2005)
Cytoscape	2D Large-scale	Various data import GO & KEGG,OBO.	Multiple function available	Visualization of molecular interaction	Shannon et al.(2003)
BioLayout Express3D	3D and 2D Large-scale	simple input file format	Markov Clustering algorithm (MCL)	BioLayout Express3D microarray data	Freeman et al. (2007)
Osprey	2D medium scale	connecting directly to several databases	cluster genes by GO Processes	filtering capabilities	Breitkreutz (2003)
ProViz	2D and 3D Large-scale	Tulip, PSI-MI* and IntAct† formats.	merge different datasets	protein – protein interaction network	Hermjakob et al. (2004)
Ondex	2D Large-scale bidirectional	Various data import GO& KEGG, MeSH.	selectable rules type (distance)	combine heterogeneous data types	Köhler et al. (2006)
PATIKA	2D Large-scale	Various data import Entrez Gene, UniProt, PubChem, GO, IntAct , HPRD and Reactome.	database to construct the desired pathway	data integration and pathway analysis	Demir et al. (2006)
PIVOT	2D Large-scale	four different species (human, yeast, drosophila and mouse),	expand the network to display	protein-protein interactions	Nir Orlev et al. (2003)
Pajek	2D and 3D Large-scale	own input file format	detect clusters	facilitate exploration and pattern identification within networks	Batagelj et al. (1998)

*a community standard for the representation of protein interaction data(Hermjakob et. al., 2004)

† IntAct database even though it comes with support of some standard file formats, established in systems biology(Hermjakob et al.,2004).

제 2 장. 연구 방법

2.1 데이터 수집 및 가공

코로나 바이러스의 면역 특이성 연구를 위하여 다양한 데이터를 수집하고, 분류하여 데이터베이스를 구축한 후, 저장된 정보를 이용하여 바이오인포매틱스 기법을 활용한 분석을 시행하였다. 먼저, 연구에 필요한 데이터의 수집은 국내외에 발표된 논문의 검토와 기존에 구축되어 있는 바이러스와 면역 관련 데이터베이스를 참조하여, 데이터의 리스트를 확인하고 수집하였다. 코로나 바이러스 관련 단백질들의 분류와 수집은 문헌 검토를 통하여 코로나 바이러스가 크게 구조단백질과 비구조 단백질로 나누어지는 것을 확인하였다. 각각의 큰 카테고리 안에 해당되는 단백질의 종류, 특징과 서열 정보 등을 ‘GenBank’ ([http:// www.ncbi . nlm.nih.gov/genbank](http://www.ncbi.nlm.nih.gov/genbank))와 ‘VIPR’(<http://www.viprbrc.org>)에서 수집하였다. 코로나 바이러스의 단백질 수집의 기준은 분리된 숙주를 모든 종을 포함하였고, 지역과 분리된 연도와 날짜에 관계없이 등록되어 있는 모든 단백질들을 수집하였다. 면역관련 단백질의 수집은 관련 문헌의 검토를 통하여 바이러스의 면역에 관여하는 단백질의 리스트를 확인하고, 면역 관련 데이터베이스인 ‘IMMPORT’ ([http:// www.immport.org](http://www.immport.org)), ‘Uniprot’ (<http://www.uniprot.org>), ‘GenBank’ ([http:// www.ncbi.nlm.nih.gov/genbank](http://www.ncbi.nlm.nih.gov/genbank))에서 단백질 들을 수집하였다. 단백질 들의 수집은 Fasta 파일의 형식으로 다운을 받아 프로그래밍 언어인 자바를 이용하여 데이터베이스 구축에 필요한 정보만을 선택하여 가공하였다. 이러한 과정은 구축된 데이터베이스에 단백질의 정보와

서열 정보를 이식하기 위해 필요한 과정이다. 코로나 바이러스 관련 단백질은 바이러스의 ‘Virion’ 구조를 이루는 구조단백질인 ‘S’, ‘N’, ‘M’, ‘E’, ‘HE’ 단백질에 대해 수집하였으며, 수집한 정보는 단백질이 분리된 숙주 정보, 지역, 검체 채취일, 단백질 서열등의 정보를 포함하였다. 코로나 바이러스의 전사와 복제 등에 관여하는 비구조 단백질인 ‘NSP1~16’, ‘TM’, ‘L’, ‘L2’등의 단백질에 대해 수집하여 분리된 숙주의 정보, 단백질이 분리된 바이러스의 종, 지역과 단백질의 서열도 함께 수집하였다. 면역관련 단백질은 세 가지의 카테고리로 나누어 단백질의 유전자 기호, 이름, 동의어, 염색체의 위치와 단백질의 서열을 수집하였다. 코로나 바이러스와 숙주 간의 상호작용 관련 데이터는 ‘VirHostNet’과 ‘Viralmentha’에서 코로나 바이러스와 숙주 간의 상호작용, 바이러스 단백질들간의 상호작용을 수집하여 네트워크 분석에 이용하였다. 데이터의 수집 및 가공 과정을 Figure 2.1(a)에 도식화하여 요약하였다.

2.2 시스템 개발 환경 및 BLAST 서버 구축

코로나 바이러스의 면역 특이성 연구를 위한 데이터베이스의 구축을 위하여 ‘HPC cluster systems’를 기반으로 한 8C AMD Opteron-6128 2.0Ghz, 8Gb RAM, 500Gb SATA 7200rpm 3Gbps HDD 사양의 서버를 구축하였다. ‘Operating System’은 Linux v2.6.18을 사용하였다. Linux 서버 환경에 데이터를 저장하기 위하여 ‘Database Management System(DBMS)’으로 MySQL v5.5를 이용하였다. 수집한 단백질 정보들의 가공을 위하여 자바 프로그래밍 언어(JAVA (ver. jdk1.8.0_73))를 사용하였다. 웹에 기반을 둔 데이터베이스를 구축하고, 웹과의 연동을 위하여 ‘Java Server Page (JSP v2.3)’, ‘Hyper Text Makeup Language (HTML)’과 ‘Java Script’를 이용하였다. 웹 서버의 작업환경을 구축하고자 웹 컨테이너 Tomcat v8.0을 이용하여 동적인 웹 페이지가 구현되는 데이터베이스를 구축하였다(Table 2.1). 또한, 사용자가 가지고 있는 쿼리 시퀀스의 상동성 정보를 제공해 주기 위한 웹 인터페이스인 BLAST (Basic Local Alignment Search Tool) 서버를 구축하기 위하여 웹 BLAST 패키지 ‘wwwblast (ncbi-blast-2.2.26, for Linux)’를 설치 하였다. BLAST 검색을 이용하는 방법은 BLAST페이지를 클릭한 후, Program을 클릭하여 쿼리 시퀀스의 종류에 따라서 DNA서열이면 ‘blastn’을 클릭하고, 단백질 서열인 경우에는 ‘blastp’를 클릭한다. ‘Enter sequence in fasta format’에 검색하고자 하는 서열을 직접 붙여 넣거나, 파일 첨부란에 파일을 첨부하고 ‘submit sequence’를 클릭한다(Figure 3.5). 쿼리시퀀스의 상동성 정보의 결과를 ‘Sequence producing significant alignment’에서 확인 할 수 있다(Figure 3.6). 왼쪽에는 각각의 시퀀스의 이름과 파일의 accession no가 나타나며, 파일의 이름을 클릭하면 파일의 내용을 확인할 수 가 있다. ‘bit

score’는 값이 높을수록 내가 가지고 있는 서열과의 관련이 높다는 것을 의미한다. ‘bit score’도 클릭하면 데이터베이스에 연결이 되어 시퀀스들의 정렬된 모양을 확인 할 수 가 있다. ‘E value’는 데이터베이스에 풍부하지 않은 시퀀스와 조합된 수를 의미하는 것으로 수치가 작을수록 상동성이 높으며, 이러한 유사성은 공통 후손임을 의미한다(Barry, 2001).

2.3 데이터베이스 구현

본 연구에서 구축한 ‘Human Corona Virus Immune Database (HCoV-IMDB)’의 구현 목적은 사용자들에게 코로나 바이러스와 면역세포들의 유전자 서열을 같이 확인하고, 분석하기 위한 편리하고 유용한 플랫폼을 제공하고자 하는 것이다. 본 연구에서 구축된 데이터베이스는 코로나 바이러스의 구조단백질과 비구조 단백질의 유전자 서열 데이터와 면역 관련 단백질과 유전자의 서열 데이터를 통합하여 구축함으로써, 면역시스템에 관여하는 단백질들의 서열을 검색하고 분석할 수 있도록 구현하였다. 데이터베이스는 코로나 바이러스 관련 유전자 정보, 면역관련 유전자 정보, 쿼리 유전자로 유전자 서열의 상동성을 검색할 수 있는 BLAST 서버와 바이오인포매틱스 툴을 이용하여 분석할 수 있도록 크게 네 부분으로 나누어 구현하였다. 데이터베이스의 기능은 다음과 같다. 첫째, 코로나 바이러스 관련 유전자의 정보 제공이다. 코로나 바이러스는 크게 구조 단백질과 비구조 단백질로 구분하여 구현하였다. 모든 종에서 분류된 구조 단백질 카테고리는 ‘S’, ‘M’, ‘E’, ‘N’, ‘HE’로 나누어 테이블을 작성하여 유전자 정보를 제공하고, 비구조 단백질 카테고리에는 ‘NSP 1~16’ 단백질이 포함된

테이블을 작성하여 유전자 특징과 서열의 정보를 제공해 준다. 테이블에 저장되어 있는 데이터는 모두 문자형이며 유전자 상징 기호, 분류 기호, 단백질 이름, 종으로 나누어 필드를 구분하고, 분류체계 정보는 문자 데이터 타입으로 저장하였으며, 유전자 서열 정보는 길이가 길어 텍스트 데이터 타입으로 설정하여 유전자 서열을 모두 포함할 수 있도록 설정하였다(Table 2.2). 두 번째, 면역관련 유전자의 정보 제공이다. ‘Cytokines & Receptor’, ‘Natural killer cell’, ‘Antigen processing and presentation’ 세 개의 카테고리로 나누어져 있다. ‘Cytokine’ 카테고리는 ‘Interleukins & Receptor’, ‘Interferons & Receptor’, ‘TNF-Family members & Receptor’, ‘TGF-Family members & Receptor’, ‘Chemokines & Receptor’로 분류하여 유전자의 정보를 테이블 형태로 제공해 준다. 테이블에 저장되어 있는 데이터는 모두 문자 형이며 유전자 상징 기호, 분류 기호, 단백질 이름, 동의어, 염색체 위치정보 등으로 분류하여 칼럼을 나누고, 분류체계 정보는 문자 데이터 타입으로 저장 하였으며, 유전자 서열 정보는 길이가 길어 텍스트 데이터 타입으로 설정하여 유전자 서열을 모두 포함할 수 있도록 설정하였다(Table 2.2). 세 번째, 사용자가 알고자 하는 쿼리의 유전자의 서열 기반 상동성 비교가 가능한 BLAST 서버의 구현이다. BLAST로 연동되는 웹 기반에서 코로나 바이러스의 단백질과 면역 관련 단백질 별로 데이터베이스가 구축되어 있다. 사용자가 상동성 검색을 시행하고자 하는 BLAST 서버의 단백질을 선택하여 쿼리 서열을 넣으면 상동성이 높은 유전자의 정보들을 확인할 수가 있다. 마지막으로, 분석 부분에서 계통수 분석과 네트워크 분석 결과를 통하여 코로나 바이러스의 위험성을 예측하고, 면역 특이성을 찾아낼 수 있다. 완성된 데이터베이스는 <http://147.47.72.71:8088/hcov-im/HCoV-IMDBmain.jsp>에 공개되어 있다.

2.4 계통수 작성

본 연구에서 구축된 데이터베이스를 기반으로 하여 단백질들 간의 계통수를 작성하였다. 계통수 작성은 ‘NJ’ 방법을 기반으로 하였다. 1986년 Saitou와 Nei에 의해 개발된 ‘NJ’방법의 원리는 최소한의 진화 트리(minimum-evolution tree)를 검색하는 프로세스의 알고리즘을 이용하여 거리 행렬(distance matrix)을 작성하여 트리를 작성한다. 작성된 거리 행렬은 모든 쌍간의 진화적 거리를 나타낸다. 가장 유사한 것끼리 쌍을 이루는 ‘OTUs (Operational taxonomic units, neighbors)’에 내부 노드를 삽입하고, 그 거리를 직접적으로 계산하면서 새로운 거리 행렬을 작성하고, 가지들의 총거리가 최소화 되도록 하는 쌍을 찾아 가면서 연결하여, 그 결과 고유한 최종의 topology를 얻을 수 있다. 또한, 최종 트리에서는 가지의 길이도 함께 제공해 준다. 간단히 알고리즘에 대하여 알아보면, (1)의 식을 이용하여 Figure 2.1(a)의 초기 별모양 트리의 가지의 길이의 총합을 구한다.

D_{ij} : OTUs i 와 j 의 거리

L_{ab} : 노드(node) a 와 b 의 거리

$$S_0 = \sum_{i=1}^N L_{ix} = \frac{1}{N-1} \sum_{i < j} D_{ij}, \quad (1)$$

모든 거리들이 첨가될 때 각각의 가지는 $N-1$ 번 수가 세어진다. Figure 2.1(b)의 내부노드 X 와 Y 사이의 (L_{XY}) 가지의 거리를 구하는 식은 (2)와 같다.

$$L_{XY} = \frac{1}{2(N-1)} \left[\sum_{k=0}^n (D_{1K} + D_{2K}) - (N-2)(L_{1x} + L_{2x}) - 2 \sum_{i=3}^N L_{iy} \right]. \quad (2)$$

식(2)에서 대괄호 안의 공식은 총 거리를 구하는 공식이고, Figure 2.1(b)의 내부 노드인 X, Y를 제거하면, 두개의 별모양의 topology가 남게 된다. 남은 topology의 거리를 구하는 방법은 공식(1)에 적용하면 (3a), (3b)와 같다. (3a)는 OTUs 1과2를 가진 topology의 거리를 구하는 공식이고, (3b)는 나머지 N-2 OTUs를 가진 topology를 구하는 공식 이다.

$$L_{1x} + L_{2x} = D_{12} \quad (3a)$$

$$\sum_{i=3}^N L_{iY} = \frac{1}{N-3} \sum_{3 \leq i < j} D_{ij} \quad (3b)$$

위에서 구한 각각의 가지의 길이를 모두 합하면 Figure 2.1(b)의 총 길이를 구할 수 있으며, 그 식은 (4)와 같다.

$$\begin{aligned} S_{12} &= L_{XY} + (L_{1x} + L_{2x}) + \sum_{i=3}^N L_{iY} \\ &= \frac{1}{2(N-1)} \sum_{K=3}^N (D_{1K} + D_{2K}) + \frac{1}{2} D_{12} + \frac{1}{N-2} \sum_{3 \leq i < j} D_{ij}. \end{aligned} \quad (4)$$

또한 식(4)는 가지 길이의 최소 제곱 추정치(least-squares estimates of branch lengths)들의 합으로 보여 줄 수 있다. S_{ij} 모든 값 중에서 S_{12} 값이 가장 작은 것으로 확인 되면, OTUs 1 과 2는 이웃한 쌍(OTUs 1-2)로 지정이 된다. 결합된 OUT와 다른 OUT j 사이의 거리는 (5)의 식으로 구한다.

$$D_{(1-2)j} = (D_{1j} + D_{2j})/2 \quad (3 \leq j \leq N) \quad (5)$$

결과적으로, OTUs의 수는 한 개가 줄어들고, 새로운 거리 행렬을 작성하기 위하여, 위의 과정을 반복하여 새로운 쌍을 찾아낸다. 이 과정이 OTUs가 3개가 될때까지 반복하며, 마지막에 하나의 unrooted 트리가 작성이 된다.

트리의 가지의 길이를 측정은 1967년 Fitch와 Margoliash의 방법을 이용한다. 처음으로 쌍을 이룬 OUTs 1과 2의 가지의 길이인 L_{1X} 과 L_{2X} 의 계산은 (6a), (6b)의 식에 의해 구한다.

$$L_{1X} = (D_{12} + D_{1Z} - D_{2Z})/2 \quad (6a)$$

$$L_{2X} = (D_{12} + D_{2Z} - D_{1Z})/2 \quad (6b)$$

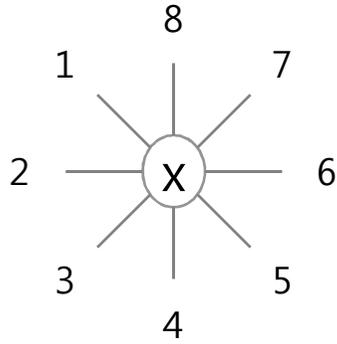
여기에서, $D_{1Z} = (\sum_{i=3}^N D_{1i})/(N-2)$ 이고, $D_{2Z} = (\sum_{i=3}^N D_{2i})/(N-2)$ 이다. Z는 모든 OTUs 그룹을 포함하며, D_{1Z} 와 D_{2Z} 는 1과 Z, 2와 Z사이의 거리를 의미한다. L_{1X} 와 L_{2X} 는 Figure 2(b) 트리의 최소 제곱 추정치를 계산하는 식이다(Sitou et al., 1987).

계통수의 작성은 MEGA7.0 프로그램을 이용하였으며, 트리작성을 하는데 있어 정확성이 높고, 시간이 절약되며, 대용량의 서열의 트리 작성에 적합한 ‘NJ’ 방법을 이용하였다(Mihaescu et al., 2009). 계통수 작성 시 ‘bootstrap value’를 500으로 지정하였고, 이 값은 높을수록 계통수의 정확도가 높다고 할 수 있다. 먼저, 계통수 작성을 위하여, 2015년 한국에서 분리된 ‘MERS-CoV’의 구조 단백질인 S 단백질과, 비구조 단백질인 ‘Nsp 1~16’과 사람에게 질병을 일으키는 6개 종의 바이러스의 서열을 본 연구를 위하여 구축한 ‘HCoV-IMDB’에서 아미노산 서열을 Fasta 파일 형식으로 다운로드 했다. 다운로드한 파일을 다중 서열 정렬 프로그램인 ‘Clustal omega’를 이용하여 계통수 분석을 위한 정렬을 시행하였다(Sievers et al., 2014). 분석 시 default parameter를 사용하였다. 상용화 되어있는 ‘Clustal omega’의 메인 화면에서 step1 또는 step2를 이용하여 서열을 직접 입력 또는 다운로드한 Fasta 파일을 업로드 시킨 후, step3에서 submit을 클릭하면 자동으로 서열을 정렬하여 정렬 상태를 확인이 가능하고, Clustal 파일 형태로

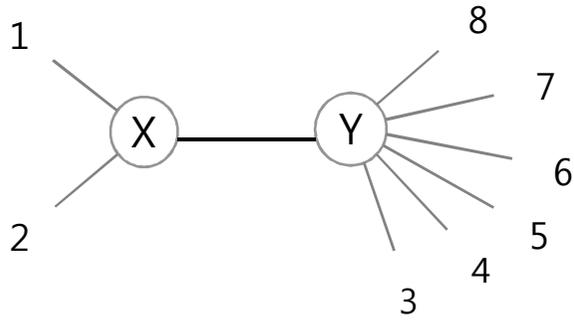
결과를 제공해 준다(Figure 2.1 & Figure 2.2). 계통수 작성을 위하여 정렬된 서열 파일인 Clustal 파일을 MEGA7.0 프로그램에서 제공하는 파일 전환 기능을 이용하여, mega 파일로 전환하여 준다. MEGA7.0 프로그램은 다양한 방법으로 계통수 작성이 가능하다. 본 연구자는 'NJ'를 이용하였고, 분석을 위한 파라미터는 'scope'는 'all selected taxa', 'phylogeny test'는 'Boot Strap method', 'replication No'는 500, 'Substitution type'은 'amino acid', model은 'No.of difference', 'Site of rate'는 'uniform rate'로 지정하여 계통수를 작성하였다.

2.5 네트워크 분석

본 연구에서는 코로나 바이러스와 숙주 간의 면역 특이성을 분석하기 위하여 바이러스의 단백질과 사람의 단백질 간의 상호작용 데이터를 이용하여 네트워크 분석을 실시하였다. 코로나 바이러스와 사람의 단백질들 간의 상호작용 데이터는 상용화되어 있는 데이터베이스인 ‘VirHostNet 2.0’ (<http://virhostnet.prabi.fr>)과 ‘virus mentha’ (<http://virusmentha.uniroma2.it>)에서 ‘SARS’의 구조, 비구조 단백질과 사람의 면역단백질과의 상호작용, 바이러스 간의 상호작용 데이터를 수집하였다. 수집된 데이터는 Cytoscape 3.3.0을 이용하여 네트워크 분석을 수행하였다(Shannon et al. 2003). 먼저 ‘VirHostNet 2.0’에서 바이러스의 단백질과 숙주 간의 상호작용 데이터 80개를 수집하였다. 데이터베이스에 상호작용에 관한 강도의 정보는 포함되어 있지 않아 상호작용 여부만 네트워크 분석을 시행하였다. Cytoscape의 메뉴 바에서 가져오기를 클릭하여 엑셀로 정리된 두 개의 필드를 입력하였다. 입력한 필드에 적합한 노드를 설정하여 준다. 본 연구에서는 바이러스의 데이터를 ‘Target node’로 지정 해주었다. 연구에 이용한 파라미터는 ‘y axis’는 ‘organic method’를 이용하였고, 구분이 쉽도록 두 계열의 노드에 색을 지정하였다. 바이러스 단백질 노드에는 핑크색, 사람 단백질의 노드에는 파란색을 지정하여 구분이 용이한 네트워크를 작성하였다.



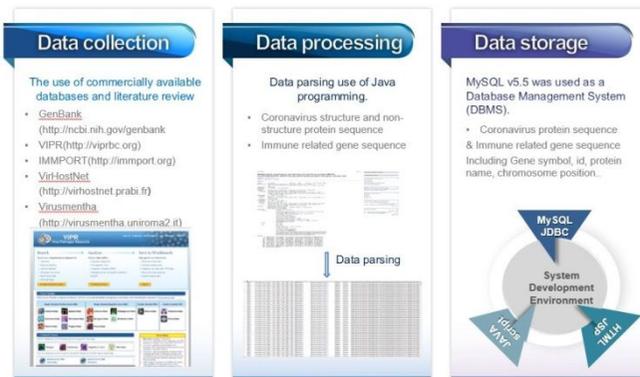
(a)



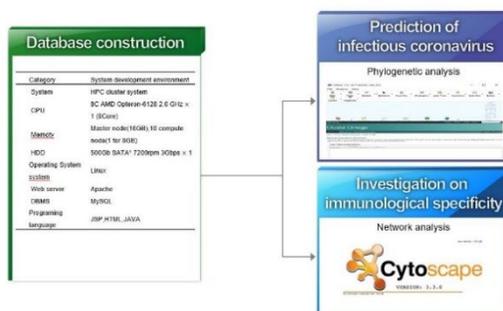
(b)

Figure 2.1 Structure of tree. (a) Initial starlike tree construction and (b) OTUs 1 and 2 are clustered by internal node \textcircled{X} . The other OTUs are clustered by internal node \textcircled{Y} (Sitou et al., 1987).

a) Data processing



b) Database construction and analysis



c) Result of research.



Figure 2.2 Process of research. a) Data processing process is proceeded in collecting, processing and storing of data in three steps. b) After constructing a database, phylogenetic tree analysis conducted to predict of infectious coronavirus infectious, immunological specificity was investigated by the network analysis performed. c) Result of the constructed database and phylogenetic tree analysis and network analysis.

Table 2.1 Databases for virus research

Category	System development environment
System	HPC cluster system
CPU	8C AMD Opteron-6128 2.0 GHz × 1 (8Core)
Memory	Master node(16GB),10 compute node(1 for 8GB)
HDD	500Gb SATA* 7200rpm 3Gbps × 1
Operating System	Linux
Web server	Apache
DBMS	MySQL
Programing language	JSP,HTML,JAVA

The system development environment using the Web-based interface, database construction.

Table 2.2 Schema of protein in the database

Name of field table	System development environment
Corona virus	
Gene symbol	Varchar(50) not null primary key
ID	Varchar(10) null
Protein name	Varchar(50) null
Organism	Varchar(50) null
Sequence	longtext null
Immune	
Gene symbol	Varchar(50) not null primary key
ID	Varchar(10) null
Protein name	Varchar(50) null
Synonyms	Varchar(50) null
Sequence	longtext null
Chromosome	Varchar(50) null

제 3 장. 연구 결과

3.1 면역 특이성 데이터베이스 구축

‘HCoV-IMDB’는 유전자 서열을 얻기 위한 두 개의 주된 페이지와 BLAST 서버 구현, 분석 결과를 나타낸 4개의 페이지로 구성되어 있다(Figure 3.1). 홈페이지에서 ‘HCoV Protein’ 페이지를 클릭하면 코로나 바이러스의 구조단백질과 비구조 단백질을 카테고리별로 분류하여 단백질의 정보와 서열의 정보를 제공해 준다. 바이러스 구조 단백질의 종류는 ‘S’, ‘N’, ‘M’, ‘E’, ‘HE’ 단백질로 분류하여 테이블을 작성하였다. S 단백질은 559개, N 단백질은 448개, M 단백질은 147개, E 단백질은 143개, HE 단백질은 181개의 유전자 서열 및 유전의 특성 정보를 포함하고 있다(Table 3.1). 코로나 바이러스의 비구조 단백질 중 ‘NSP1’ 단백질은 225개, ‘NSP2’ 단백질은 467개, ‘NSP3’ 단백질은 241개, ‘NSP4’ 단백질은 189개, ‘NSP5’ 단백질은 228개, ‘NSP6’ 단백질은 223개, ‘NSP7’ 단백질은 222개, ‘NSP8’ 단백질은 141개, ‘NSP9’ 단백질은 398개, ‘NSP10’ 단백질은 189개, ‘NSP11’ 단백질은 198개, ‘NSP12’ 단백질은 307개, ‘NSP13’ 단백질은 223개, ‘NSP14’ 단백질은 134개, ‘NSP15’ 단백질은 138개, ‘NSP16’ 단백질은 137개, ‘NS5a’ 단백질은 43개, ‘NS2a’ 단백질은 80개, ‘RdRP’ 단백질은 74개, ‘L(leader protein)’은 168개, ‘TM2’은 84개의 유전자 서열 데이터가 포함되었다(Table 3.1). 또한 사용자가 원하는 단백질을 클릭하여 유전자의 서열 정보 등을 Fasta 파일 형식으로 다운로드를 할 수 있도록 제공해 준다(Figure

3.2). ‘Immune Related Gene’ 페이지에서는 면역 관련 단백질들을 8개의 카테고리로 분류하여 정보를 제공해주고, 각각의 페이지에는 면역세포들의 유전자와 그들의 수용체 관련 단백질의 정보와 서열을 포함하고 있다(Figure 3.3). 크게 ‘Cytokines & receptor’, ‘Natural killer cell’, ‘Antigen processing & presentation’ 세 개의 카테고리로 구분되어 있으며, 각각의 카테고리 안에 테이블 형식으로 단백질의 서열과 정보를 제공해 준다. ‘Cytokines & receptor’ 카테고리에는 ‘Interleukins & receptor’ 단백질은 89개, ‘Interferons & receptor’ 단백질은 20개, ‘TNF-Family members & receptor’ 단백질은 31개, ‘TGF-Family members & receptor’ 단백질은 46개, ‘Chemokines & receptor’ 단백질은 155개가 포함되어 있고, ‘Natural killer cell’와 ‘Antigen processing & presentation’ 카테고리에는 각각 112개와 148개의 레코드 포함되어 있다(Table 3.1). 사용자는 원하는 단백질을 클릭하면 Fasta 파일 형식으로 다운로드가 가능하다(Figure 3.4). ‘Analysis’ 페이지에서는 데이터베이스에서 제공하는 바이러스와 면역관련 단백질들의 서열을 기반으로 코로나 바이러스의 단백질들과 면역관련 단백질들의 분석을 시행한 계통수 분석 결과와 네트워크 분석 결과를 확인할 수 있다. BLAST 서버로 연동되는 웹 기반 페이지에서는 바이러스와 면역관련 단백질들을 종류별로 분류하여 구축이 되어 있다(Figure 3.5). 사용자들은 검색하고자 하는 단백질의 BLAST 서버를 클릭하고, 서열의 종류를 선택한 후 서열 정보를 제출하면 데이터베이스에 저장되어 있는 서열들을 기반으로 사용자의 쿼리 시퀀스와 상동성 검색을 진행하고, 그 결과 쿼리 시퀀스의 구조와 기능에 대한 진화의 단서를 얻을 수 있고, 서열의 유사성 영역 등을 찾을 수 있다(Figure3.6).

3.2 데이터베이스 기반 계통수 분석

2015년 한국에서 분리되고 NCBI에 공개되어 있는 13개 ‘MERS-CoV’의 S 단백질을 이용하여 계통수 분석을 시행하였다. 작성 결과 분리된 바이러스의 단백질은 서열이 거의 동일하여, 진화적으로 유의한 차이 등을 구분할 수 없었다(부록1). 다음으로, 숙주와 상호작용하여 면역에 관여한다고 알려진 ‘MERS-CoV’의 비구조 단백질인 ‘NSP1~16’ 단백질도 같은 방법으로 분석을 시행하였다. 분석 결과 S 단백질의 양상과 비슷하였으며, 단백질들 간의 진화적으로 유의한 차이가 없었다(부록2-17). 사람에게 질병을 일으키는 코로나 바이러스 종들 간의 면역의 차이점을 알아보기 위하여, 각각의 종 들에서 분리된 단백질들의 계통수 분석을 시행하였다. 사람에게서 분리된 6종 코로나 바이러스의 계통수를 분석한 결과, S 단백질은 속 간의 서열 차이가 있어서 계통수 분석에서도 속 별로 분류가 이루어졌다(Figure 3.7a). ‘HCov-229E’와 ‘HCoV-NL63’의 거리는 0.407으로 유사한 서열을 가지고 있었고, ‘Beta corona virus genus’ 중에서도 ‘lineage A’에 속하는 ‘HCoV-OC73’과 ‘HCoV-HKU1’의 진화적 거리(distance)도 0.374로 분류가 제대로 이루어졌음을 알 수 있었다. 분석 후 특이한 점은 ‘lineage A’에 속하는 ‘HCoV-OC43’과 ‘lineage C’에 속하는 ‘MERS’의 진화적 거리가 0.946으로 평균 1.067보다는 낮아 다른 바이러스의 S 단백질들보다 진화적으로 가까운 거리에 있음을 알 수 있었다(Table 3.2a). N 단백질의 종별 진화 정도를 확인하기 위하여 계통수 분석을 시행하였다(Figure 3.7 b). 속 별로 분류가 잘 이루어졌으며, S 단백질의 결과와 비슷한 결과를 얻었다. 특이한 점은 ‘SARS’와 ‘MERS’의 N 단백질의 거리가 평균 거리인 1.063보다 낮은 0.637으로 진화적으로 유사한 서열임을 알 수 있었다(Table 3.2b).

3.3 데이터베이스 기반 네트워크 분석

네트워크 분석을 위하여 구축된 데이터베이스에서 ‘SARS-CoV’ 단백질과 숙주의 단백질 간의 상호작용 데이터를 다운로드해서 상용화되어 있는 Cytoscape ver.3.3.0을 이용하여, 분석을 시행하였다 (Figure 3.8). 분석 결과 바이러스의 비구조 단백질을 중심으로 사람의 단백질이 3~5개가 상호 작용하는 결과를 보여 주었다. 상호작용이 일어난 부분을 확대하여 나누어서 그림으로 나타내었다 (Figure 3.9). ‘SARS-CoV’의 ‘Nsp7a’ 단백질과 사람의 ‘BCL2’, ‘B2CL1’, ‘MCL1’, ‘B2LA1’, ‘B2CL2’, ‘SGTA’, ‘SMOC1’, ‘VKOR1’ 단백질이 상호 작용함을 알 수 있었다(Figure 3.9a). ‘SARS-CoV’의 ‘Nsp1-pp1a/pp1ab’와 사람의 ‘LASIL’, ‘RCAN3’, ‘PPIH’, ‘PPIG’, ‘FKB1A’, ‘PPIA’ 단백질의 상호 작용을 확인하였다(Figure 3.9b). 사람 단백질의 ‘IC1’ 단백질과 ‘SARS-CoV’의 ‘Nsp3b’, ‘Nsp7b’, ‘Nsp13-pp1ab(ZD NTPase/HEL)’, ‘Nsp8-pp1a/pp1ab’, ‘Nsp2-pp1a/pp1ab’, ‘Nsp14-pp1ab (nuclease ExoN homolog)’ 단백질들이 상호 작용함을 확인하였다(Figure 3.9c). ‘SARS-CoV’의 ‘Nsp7-pp1a/pp1ab’ 단백질과 사람의 ‘ULA1’, ‘NCOA5’, ‘PLSL’, ‘ARL4D’, ‘BAP1’, ‘RN128’, ‘1433E’, ‘PKHO1’, ‘MI4GD’ 단백질이 상호 작용함을 알 수 있었다(Figure 3.9d). 또한 사람의 ‘DCTN2’ 단백질이 ‘SARS-CoV’의 ‘Nsp6’, ‘Nsp14-pp1ab’, ‘Nsp3b’, ‘Nsp6’, ‘hypothetical protein sars3a’와 상호 작용함을 알 수 있었다(Figure 3.9e). ‘SARS-CoV’ 단백질인 ‘Nsp9-pp1a/pp1ab’와 ‘Nsp12-pp1ab’의 사이에서 사람 단백질이 함께 상호작용 하는 것을 확인하였다(Figure 3.9f).



Figure 3.1 Main page of the constructed database. HCoV-IM ([http:// 147. 47. 72. 71:8088/hcovim/HCoV-IMDBmain.jsp](http://147.47.72.71:8088/hcovim/HCoV-IMDBmain.jsp)) is a database that is specialized for the human coronavirus and immune response proteins.

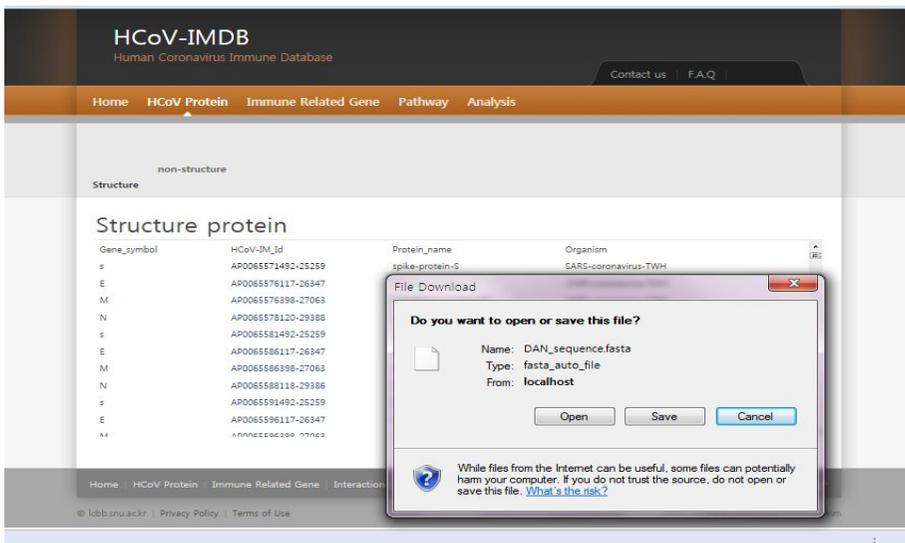


Figure 3.2 HCoV protein page of the database. The HCoV protein page is categorized into structural protein and non-structural protein. Users can access information tables for coronavirus data by clicking each viral species item.

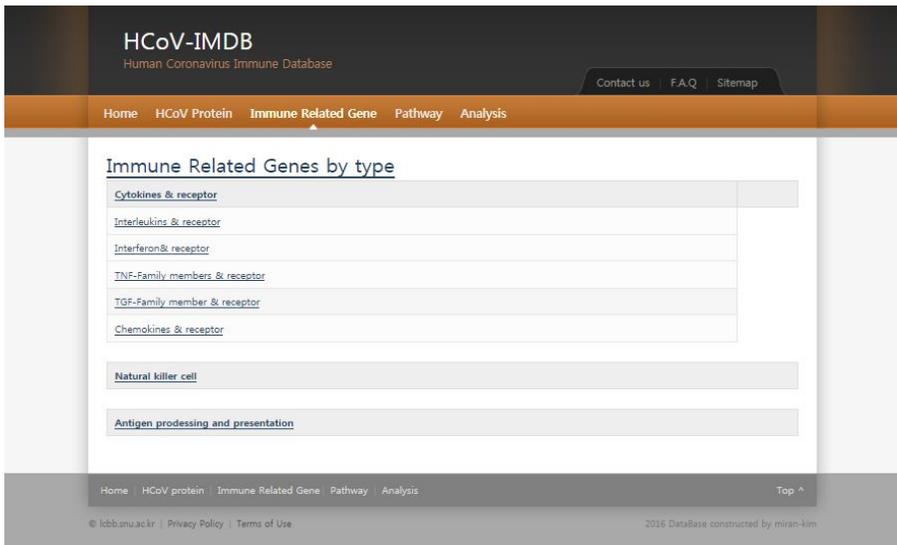


Figure 3.3 Immune Related Gene page of the database. This page is categorized into three proteins and receptors.

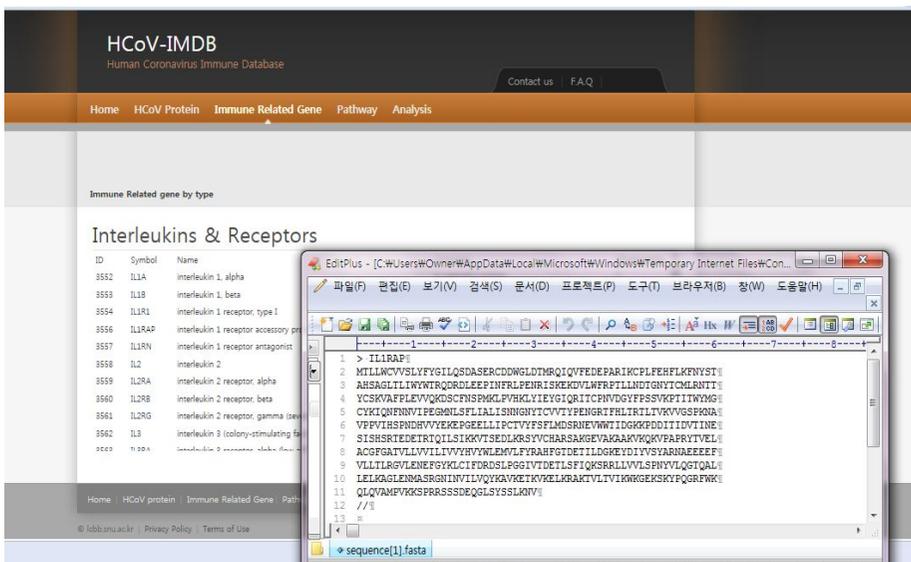


Figure 3.4 Immune Related Gene page of the database. Users can access information tables for immune related gene data by clicking each gene item.

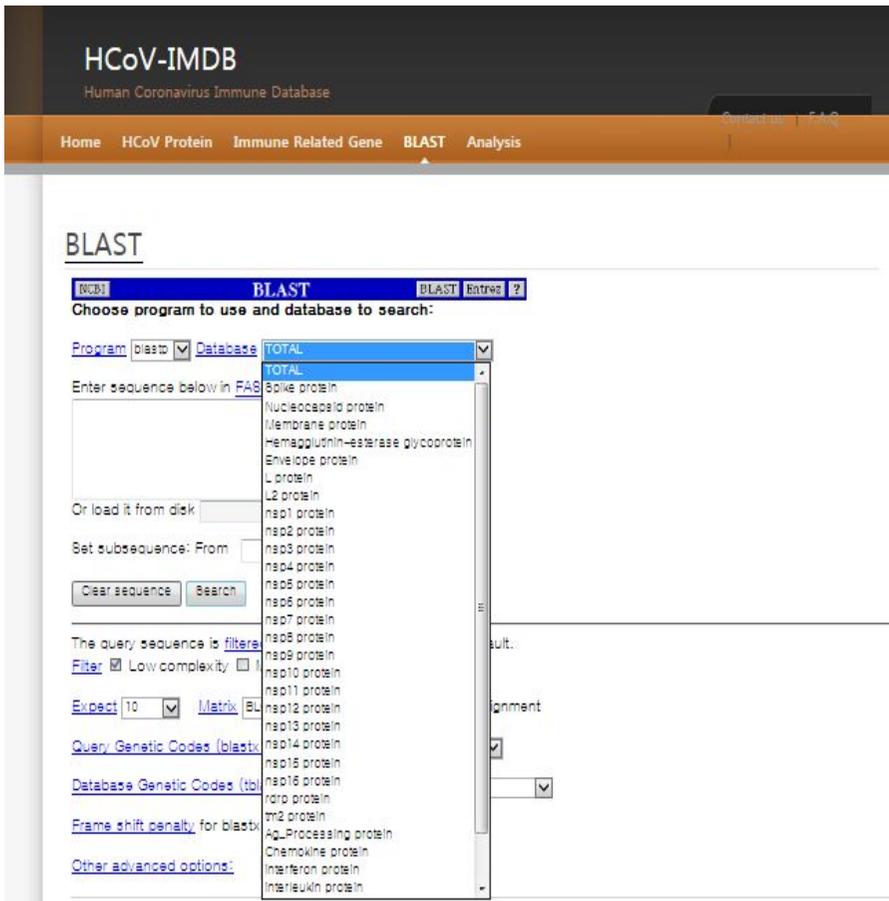


Figure 3.5 Standalone BLAST web interface page. BLAST is constructed on virus HCoV protein and immune related proteins.

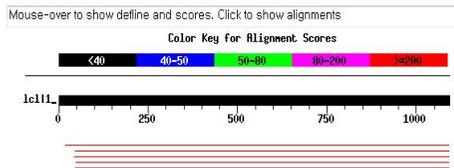
BLAST 2.2.26 [Sep-21-2011]

Reference:
 Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schäffer,
 Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1990),
 "Gapped BLAST and PSI-BLAST: a new generation of protein database search
 programs", *Nucleic Acids Res.* 25:3389-3402.

Database: SPIKE.fasta
 19 sequences; 25,077 total letters

Query=
 (1092 letters)

Distribution of 23 Blast Hits on the Query Sequence



Sequences producing significant alignments:	score (bits)	e- Value
gi 169822561 ref bat coronavirus HKU8 spike protein [Miniopterus]	2147	0.0
gi 971745657 BtR1-AlphaCoV/HuB2 spike glycoprotein [O13]	1035	0.0
gi 971745448 [BtMr-AlphaCoV(BAT)] spike glycoprotein /SAX2011]	979	0.0
gi 971745030 Swine enteric coronavirus spike protein	871	0.0
gi 45555909 Human coronavirus NL63 Spike protein []	859	0.0
gi 971483142 Camel alphacoronavirus spike protein []	785	0.0
gi 594135075 Mink coronavirus spike protein [strain #D1127]	779	0.0
gi 212681380 Thrush coronavirus HKU2-600 spike glycoprotein []	587	e-170
gi 212681390 Munia coronavirus HKU3-3514 spike glycoprotein []	567	e-164
gi 189313870 Turkey coronavirus spike protein []	273	9e-76
gi 187251956 ref Whale coronavirus SW1 spike protein [Beluga]	273	1e-75
gi 15081547 Bovine coronavirus spike structural protein []	205	1e-55
gi 56807326 Human coronavirus HKU1 spike glycoprotein []	205	2e-55
gi 126030134 Pousettus bat coronavirus HKU9 spike glycoprotein []	205	3e-55
gi 394935454 Rabbit coronavirus HKU4 S gene product []	205	3e-55
gi 253750535 [Rat coronavirus Parker spike protein]	201	4e-54
gi 74462656 Betacoronavirus HKU24 HUMAN spike glycoprotein []	200	9e-54
gi 667489389 Human betacoronavirus 2c ENCI spike glycoprotein [/...]	194	7e-52
gi 160707869 Rhinolophus bat coronavirus HKU2 spike glycoprotein []	139	3e-35

```
>gi|169822561|ref|bat coronavirus HKU8| spike protein [Miniopterus ]
    Length = 1375

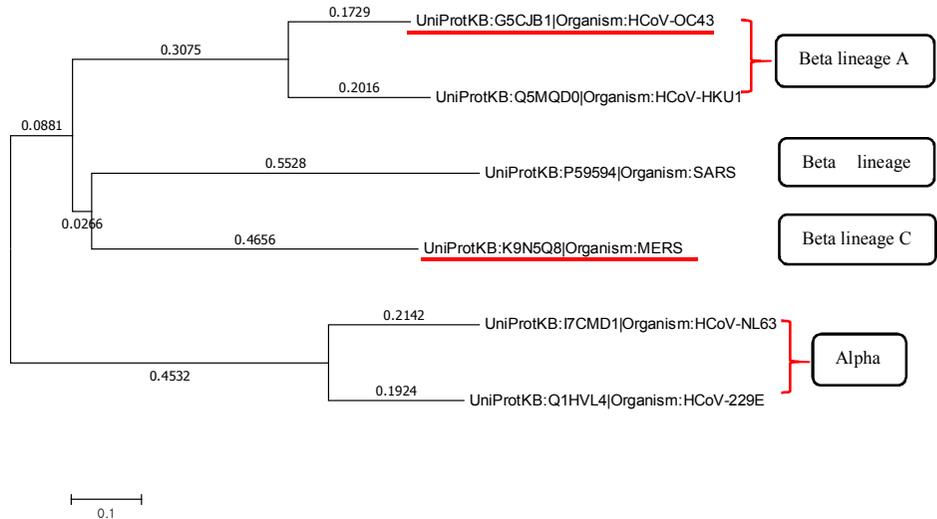
    Score = 2147 bits (5564), Expect = 0.0
    Identities = 1062/1074 (98%), Positives = 1062/1074 (98%)

    Query: 19  NAQVTGEGGYAEGNWWNSCAQHSYFVALKLGFPNASHIVTGVLPKPKGI|CPRFNGA 78
              NAQVTGEGGYAEGNWWNSCAQHSYFVALKLGFPNASHIVTGVLPKPKGI|CPRFNGA
    Sbjct: 19  NAQVTGEGGYAEGNWWNSCAQHSYFVALKLGFPNASHIVTGVLPKPKGI|CPRFNGA 78

    Query: 79  GIYTRNNANAVFVMYRTKALAFELGYSSAGGEGYSAYMAQNGKYLVLRI|CKWQNGTLA 138
              GIYTRNNANAVFVMYRTKALAFELGYSSAGGEGYSAYMAQNGKYLVLRI|CKWQNGTLA
    Sbjct: 79  GIYTRNNANAVFVMYRTKALAFELGYSSAGGEGYSAYMAQNGKYLVLRI|CKWQNGTLA 138
```

Figure 3.6 Result of homology search by BLAST with a query sequence.

a) Phylogenetic tree of spike protein



b) Phylogenetic tree of Nucleocapsid protein

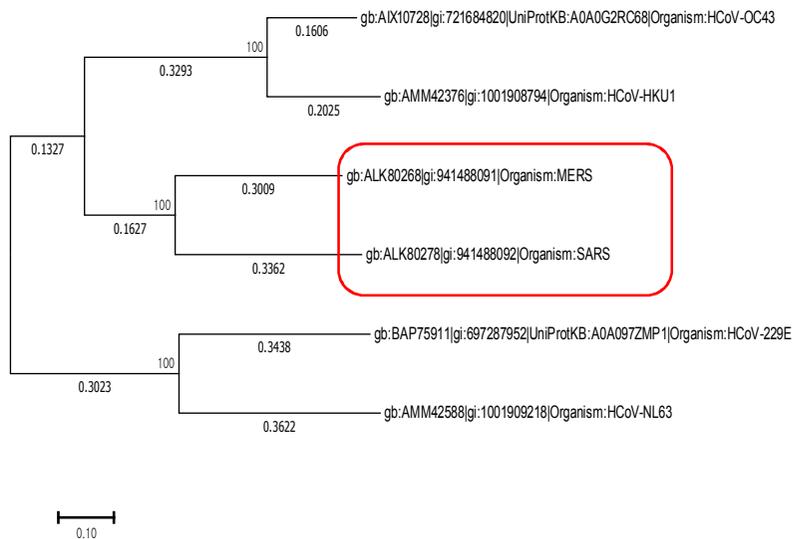


Figure 3.7 Evolutionary relationships of Spike protein and Nucleocapsid protein.
a) the result of phylogenetic tree using neighbor joining method based on spike protein in human. b) the result of phylogenetic tree using neighbor joining method based on Nucleocapsid protein in human.

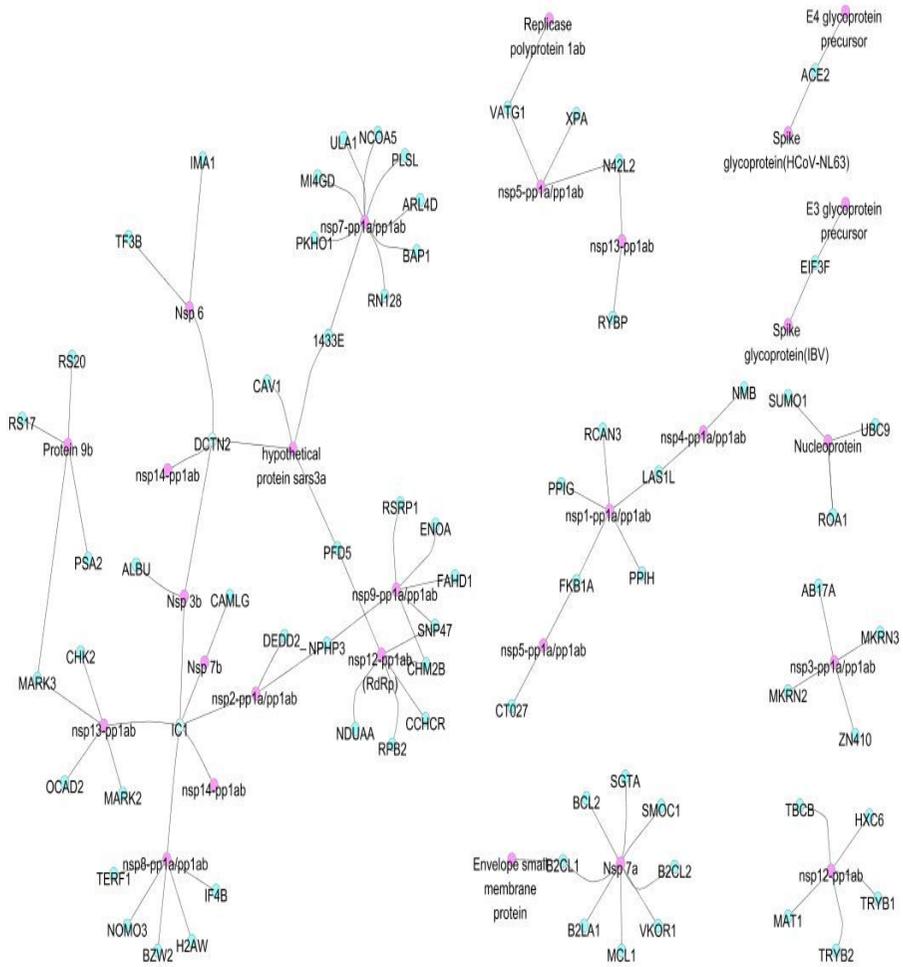


Figure 3.8 Resulted of Virus- Host interaction network by Cytoscape. viral and human proteins are shown. Interaction of viral and human proteins are shown in the different parts. Pink nodes are virus proteins, and green nodes are human proteins.

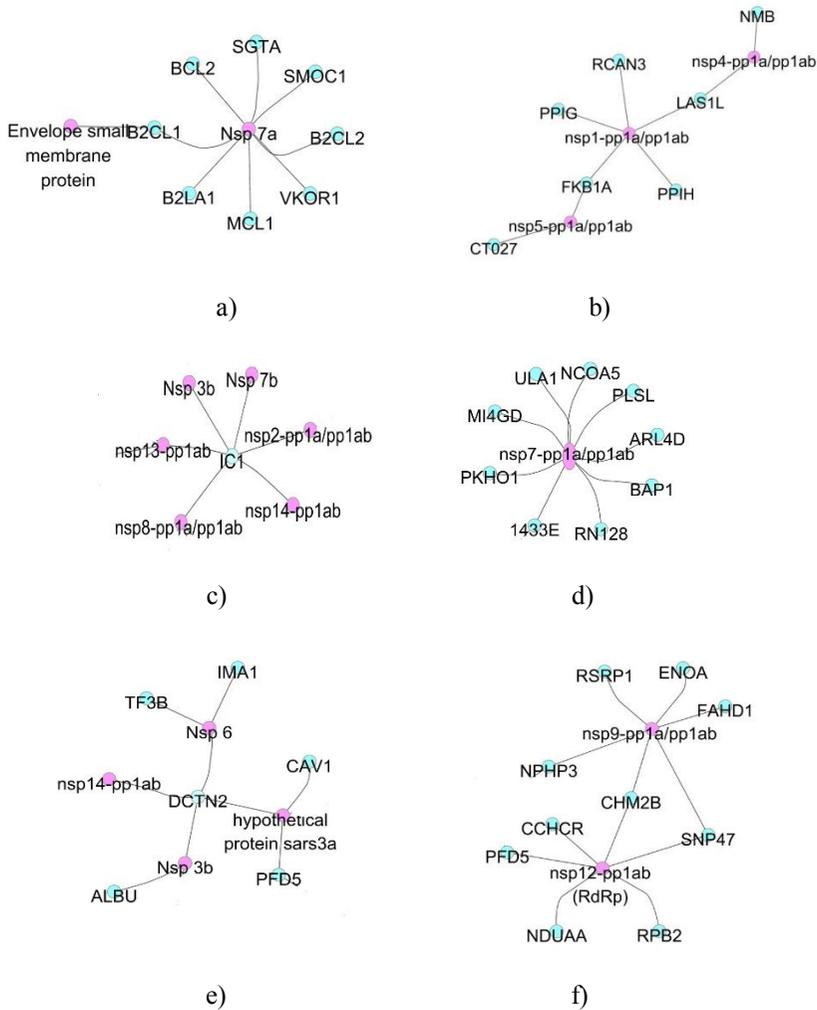


Figure 3.9 Detailed Result of Virus-Host interaction network by Cytoscape. a) Interaction between Nsp7a and human proteins. b) Interaction between Nsp1-pp1a/pp1ab and human proteins. c) Interaction between human protein IC1 and SARS proteins. d) Interaction between Nsp1-pp1a/pp1ab and human proteins. e) Interaction between human protein DCTN2 and SARS proteins. f) Interaction between Nsp9-pp1a/pp1ab and Nsp12-pp1ab and human proteins.

Table 3.1 The frequency of collected viral proteins and immune proteins

Protein	Freq	Protein	Freq
Viral structure protein		NSP12	307
E	143	NSP13	223
S	559	NSP14	134
M	147	NSP15	138
N	448	NSP16	137
HE	181	NS5a	43
Viral non-structure protein		NS2a	80
NSP1	225	RdRP	74
NSP2	467	L	168
NSP3	241	TM2	84
NSP4	189	Cytokine & receptor	
NSP5	228	Interleukin & receptor	89
NSP6	223	Interferon & receptor	20
NSP7	222	TNF-Family member & receptor	31
NSP8	141	TGF-Family member & receptor	46
NSP9	398	Chemokine & receptor	155
NSP10	189	Natural killer cell	112
NSP11	198	Antigen processing	148

This table describes the types and frequency of the viral protein and immune protein, which are included in HCoV-IMDB

Table 3.2 The result of pairwise distance matrix of Spike protein and Nucleocapsid protein

a) Pairwise distance matrix of spike protein

Coronavirus	229E	NL63	OC43	HKU1	MERS	SARS
HCoV-229E						
HCoV-NL63	0.407					
HCoV-OC43	1.236	1.239				
HCoV-HKU1	1.239	1.243	0.374			
MERS	1.236	1.251	0.946	1.015		
SARS	1.285	1.350	1.074	1.087	1.018	

The result of pairwise distance matrix of spike protein (The overall average is 1.063).

b) Pairwise distance matrix of nucleocapsid protein

Coronavirus	229E	NL63	OC43	HKU1	MERS	SARS
HCoV-229E						
HCoV-NL63	0.706					
HCoV-OC43	1.266	1.300				
HCoV-HKU1	1.243	1.386	0.363			
MERS	1.254	1.243	0.939	1.016		
SARS	1.336	1.243	0.998	1.016	0.637	

The result of pairwise distance matrix of nucleocapsid protein (The overall average is 1.067).

제 4 장. 고 찰

4.1 연구결과에 대한 고찰

본 연구에서 구축한 ‘HCoV-IMDB’는 유전자 서열을 얻기 위한 두 개의 주된 페이지와 BLAST 서버를 구현한 4개의 페이지로 구성이 되어 있다. ‘HCoV Protein’ 페이지에는 코로나 바이러스의 구조 단백질과 비구조 단백질을 분류하여 단백질의 정보와 서열 정보를 제공해 준다. 또한 사용자가 원하는 단백질을 클릭하여 유전자의 서열 정보 등을 fasta 파일 형식으로 다운로드를 할 수 있도록 제공해 준다. ‘Immune Related Gene’ 페이지에서는 면역 관련 단백질들을 8개의 카테고리로 분류하여 정보를 제공해 주며, 각각의 페이지에는 면역세포들의 유전자와 그들의 수용체 관련 단백질의 정보를 포함하고 있다. ‘Analysis’ 페이지에서는 데이터베이스에서 제공하는 바이러스와 면역관련 단백질들의 서열을 기반으로 코로나 바이러스의 단백질들과 면역 관련 단백질들의 계통수 분석 결과와 네트워크 분석 결과를 확인할 수 있다. BLAST 서버로 연동되는 웹 기반 상동성 검색 페이지에서는 바이러스와 면역관련 단백질들을 종류별로 분류하여 데이터베이스가 구축이 되어 있다. 사용자들은 이를 통하여 시퀀스의 상동성 검색이 가능하여 구조와 기능에 대한 진화의 단서를 얻을 수 있고, 서열의 유사성 영역 등을 찾을 수 있다. 다음으로, 구축된 데이터베이스의 단백질의 유전자 서열을 이용하여 코로나 바이러스의 면역 면역 특이성 연구를 시행하였다. 2015년 5월 한국에서 발생한 ‘MERS’의 유행 양상이 다른 나라에

비해 한국이 치사율은 낮았지만 전파율이 높았다. 이는 한국인의 면역반응이 다른 나라 사람들과는 차이가 있어, 전파율과 유행 양상이 차이를 보인 것이라 생각되어, 그 차이점을 증명해 보고자 연구를 진행하였다. 2015년 한국에서 유행을 일으킨 ‘MERS-CoV’와 ‘Reference MERS-CoV’의 S 단백질 서열을 계통 분석한 결과 사우디아라비아에서 유행한 서열과 99.68~99.9%로 높은 유사성을 보였다는 연구 결과가 있어, 본 연구에서는 그 부분의 의심을 배제하였다(Kim et al., 2016). 한국에서 분리된 ‘MERS-CoV’들 간의 변이가 같이 노출이 되었어도 감염의 완급에 차이를 보여준 것이라 가정하여, 2015년 5월 한국에서 분리된 ‘MERS-CoV’ 13개의 균주들 이용하여 계통수 분석을 시행하였다. 선행 연구에서는 대부분이 S 단백질만이 분석이 되어 있어(Kim DW et al., 2016; Kim Y et al., 2016), 본 연구에서는 비구조 단백질인 ‘Nsp1-16’을 분석에 포함 시켰다. ‘MERS-CoV’의 S 단백질의 변이로 한국인에게 쉽게 감염이 되고, 전파가 되었는지의 여부를 증명하기 위하여, S 단백질의 계통수 분석을 시행하였다. 분석 결과 13개의 서열의 차이가 없는 것을 확인하였다. 이러한 결과는 바이러스가 부착하여 침투하는 단계까지는 큰 차이가 없음을 알 수 있었다. 다음으로, 숙주와 상호작용하여 면역에 관여한다고 알려진 ‘MERS-CoV’의 비구조 단백질인 ‘NSP1~16’ 단백질도 같은 방법으로 분석을 시행하였다. 분석 결과 S 단백질의 양상과 큰 차이를 찾지 못하였다. 한국 병원의 특성과 문화의 차이로 인한 바이러스의 감염도 간과할 수 없는 부분이지만, 같은 환경에 처해 있었던 환자들 중에서 유독 건강이 악화된 환자와 증상이 없이 치유된 환자 등 여러 증상으로 나뉘는 것으로 보아 이는 환경적인 요인보다는 개인의 건강 상태나 유전적 소인의 차이라고 생각되었다. 특히, 개인적 차이 부분에서 개개인의 면역반응의 차이가 가장 큰 요인으로 작용할 것이다.

면역세포들의 빠른 반응과 대처는 개인이 가진 면역력의 차이에서 기인이 될 것이다. 이에 본 연구자는 유전적 소인, 특히 면역 부분에 원인을 찾아 보고자 코로나 바이러스와 면역을 연계한 연구를 설계하게 되었다. 이를 증명하기 위해서는 노출이 되었으나 감염이 되지 않은 사람, 감염이 되었으나 치료가 된 환자, 사망한 환자들의 유전자의 차이를 알아보는 것이 가장 근거 있는 증명일 수 있지만, 현재까지 그러한 연구의 결과는 공개되어 있지 않다. 따라서, 본 연구에서는 코로나 바이러스가 사람의 면역력에 영향을 미칠 수 있는 요인을 찾아 보고자 하였다. ‘MERS’나 ‘SARS’의 창궐은 아무도 예상치 못했던 동물에 감염을 일으켰던 바이러스가 변종이 일어나 사람을 감염 시키면서 혼란을 초래 했었던 부분이다. 따라서 잠재되어 있는 바이러스의 감염성을 찾아내고 그 위험을 미리 인지하여, 예방하는 것이 가장 효과적이라 생각되었다. 사람에게서 분리된 6종의 코로나 바이러스의 유전적 서열의 차이와 진화의 차이가 인체를 감염 시킬 때 영향을 미치고 그 차이가 종들 간의 감염률과 치사율과의 상관성과 연관이 있는지 알아보기 위하여, 사람에게서 분리된 코로나 바이러스의 계통수 분석을 시행하였다. 사람에게 질병을 일으키는 코로나 바이러스 종들 간의 면역의 차이점을 알아보기 위하여, 각각의 종들에서 분리된 단백질들의 계통수 분석을 시행하였다. S 단백질의 계통수 분석을 시행한 결과 코로나 바이러스를 분류 기준인 속 간의 서열 차이가 있어 계통수 분석이 제대로 이루어졌음을 알 수 있었다. 코로나 바이러스의 과인 ‘Coronaviridae’는 두개의 아과인 ‘Coronavirinae’와 ‘Torovirinae’로 나뉘며, ‘Coronavirinae’ 속에 ‘Alpha’, ‘Beta’, ‘Delta’, ‘Gamma corona virus’로 분류되고, 사람에게서 분류된 바이러스들은 ‘Alpha’, ‘Beta corona virus’에 속한다. ‘Alpha coronavirus’에 ‘HCov-229E’, ‘HCoV-NL63’이 속한다. ‘Beta coronavirus’는 바이러스의 구조에

따라서 4개의 lineage A, B, C, D로 나뉜다. lineage A에 ‘HCoV-OC43’과 ‘HCoV-HKU1’이 해당이 되며, lineage B에 ‘SARS’, lineage C에 ‘MERS’가 해당이 된다. 사람에게서 분리된 6종의 계통수를 분석한 결과, ‘HCov-229E’와 ‘HCoV-NL63’의 진화적 거리는 0.407으로 유사한 서열을 가지고 있었고, ‘Beta corona virus genus’ 중에서도 lineage A에 속하는 ‘HCoV-OC43’과 ‘HCoV-HKU1’의 거리도 0.374로 분류가 제대로 이루어 졌음을 알 수 있었다. 분석 후 특이한 점은 lineage A에 속하는 ‘HCoV-OC43’과 lineage C에 속하는 ‘MERS’의 거리가 0.946으로 평균 1.067보다는 낮아 다른 바이러스의 S 단백질들보다 진화적으로 가까운 거리에 있음을 알 수 있었다. 이러한 결과는 Maryam의 연구에서도 ‘HCoV-OC43’ 바이러스가 남아시아 지역의 성인에서 급성 호흡기 감염을 일으키고, 바이러스의 유전자 변이를 유심히 지켜보아야 한다는 연구 결과와 유사한 결과를 보였다(Maryam et al., 2016). 따라서, ‘HCoV-OC43’ 바이러스의 감염과 변이 여부 등의 조사와 감시가 국내에서도 시행이 되어야 할 것이라 사료된다. 다음으로는 ‘P1 (activator protein 1)’의 신호 전달과 ‘PKR (protein kinase R)’의 기능을 저해 시킬 뿐만 아니라, ‘NF-kB (nuclear factor-Kb)’의 활성을 방해하여 인터페론의 생성에 영향을 미치는 N 단백질의 중별 진화 정도를 확인하기 위하여 계통수 분석을 시행하였다. 속 별로 분류가 잘 이루어졌으며, S 단백질의 결과와 비슷한 결과를 얻었다. 결과에서 특이한 점은 ‘SARS-CoV’와 ‘MERS-CoV’의 N 단백질의 거리가 평균 거리인 1.063보다 낮은 0.637으로 진화적으로 유사한 서열임을 알 수 있었다. ‘SARS-CoV’와 ‘MERS-CoV’의 면역반응은 바이러스들의 여러 단백질들이 인터페론 합성을 저해함으로써 성장과 증식을 하게 되고, 숙주 세포는 세포의 기능을 잃거나 사멸하게 되면서 질병이 악화되어 환자를 사망에 이르게 되는 것이 가장 잘 알려진

면역기전이다. 인터페론 형성에 영향을 미치는 N 단백질의 구조의 유사성은 두 바이러스의 유사한 증상이 나타나는 것을 설명할 수 있겠다. 하지만, 계통수 분석 결과만으로 결론을 내리기에는 어려움이 있었고, 이 부분에 대해서는 각각의 바이러스와 면역 세포들 간의 상호작용의 강도나, 단백질의 발현량 등의 연구로 증명이 가능 할 것이다. 네트워크 분석을 위하여 구축된 데이터베이스에서 바이러스 단백질과 숙주 단백질 간의 상호작용 데이터를 다운로드해서 상용화 되어 있는 Cytoscape ver.3.3.0을 이용하여 분석을 시행하였다. 분석 결과 바이러스의 비구조 단백질을 중심으로 사람의 단백질이 3~5개가 상호 작용하는 결과를 보여 주었다. ‘SARS-CoV’의 ‘Nsp7a’ 단백질과 사람의 ‘BCL2’, ‘B2CL1’, ‘MCL1’, ‘B2LA1’, ‘B2CL2’, ‘SGTA’, ‘SMOC1’, ‘VKOR1’ 단백질이 상호 작용함을 알 수 있었다(Figure 3.9a). 이러한 결과는 Marta의 연구에서 ‘Nsp7a’의 단백질이 염증반응과 관계된 ‘NF-kB’ 기전과 ‘type I 인터페론’의 발현의 감소에 영향을 미친다는 연구 결과와 유사한 결과를 얻었다(Marta et al., 2014). ‘BCL2’, ‘B2CL1’, ‘MCL1’은 세포 사멸과 염증 반응에 관여하는 단백질로, 바이러스 단백질인 ‘Nsp7a’가 세포의 사멸 기전에도 연관이 있을 것이라 사료된다(Bruyey et al., 2007; Wei et al., 2008; Maurer et al., 2006). ‘B2LA1’은 ‘IL-3’의 박탈에 의해서 세포사멸을 지연시키고, 감염 기간 동안 내피세포의 유지와 조혈세포들에서 외부 신호전달의 주된 기능을 한다. ‘B2CL2’은 세포 사멸을 촉진하는 단백질을 억제시켜 세포의 생존을 촉진시키는 역할을 한다(Gibson et al., 1996). ‘SGTA’는 단백질의 접힘을 도와주고 응집을 막아 주는 역할을 하는 ‘HSC70’와 ‘HSP70’에 직접 결합하여 샤페론(chaperone)을 도와주고, ‘ATPase’의 활성화에 관여하는 단백질이다(Dutta et al., 2008). ‘Nsp7a’가 이들 단백질과 상호 작용하면서 감염된 숙주 세포는 자신의

단백질을 조절하는 샤페론에 영향을 받아 정상적인 세포의 기능을 할 수 없을 것이다. ‘SARS-CoV’의 ‘Nsp1-pp1a/pp1ab’와 사람의 ‘LASIL’, ‘RCAN3’, ‘PPIH’, ‘PPIG’, ‘FKB1A’, ‘PPIA’ 단백질의 상호작용을 확인하였다(Figure 3.9b). ‘SARS-CoV’의 ‘Nsp1’ 단백질은 40S 리보솜과 결합하여, 숙주 세포의 ‘Capped-mRNA’의 변형을 일으켜 번역 과정이 일어나지 못하고 분해되어 단백질 합성을 저해시킨다(Kamitani et al., 2009). ‘LASIL’ 단백질은 60s 리보솜의 생성과 ‘28S rRNA’의 성장에 관여하고(Castle et al., 2010), ‘PPIH’, ‘PPIG’와 ‘PPIA’는 단백질의 접합, ‘pre-mRNA splicing’, ‘U4/U5/U6 tri-snRNP complex’의 재조합에 관여한다(Teigelkamp, 1998). 이들 단백질과 ‘Nsp1’의 단백질이 상호 작용하면서 사람의 리보솜에 바이러스 단백질의 결합이 가능한 것이라 사료된다. 사람 단백질의 ‘IC1’ 단백질과 ‘SARS-CoV’의 ‘Nsp3b’, ‘Nsp7b’, ‘Nsp13-pp1ab(ZD, NTPase/HEL)’, ‘Nsp8-pp1a/pp1ab’, ‘Nsp2-pp1a/pp1ab’, ‘Nsp14-pp1ab(nuclease ExoN homolog)’ 단백질들이 상호 작용함을 확인하였다(Figure 3.9c). ‘IC1’ 단백질은 보체계 중의 하나인 ‘C1’복합체의 활성화에 관여하는 단백질이다. ‘C1r’ 또는 ‘C1s’ 단백질 분해 효소와 같이 작용하여 복합체를 화학적으로 불활성화시키는데 관여한다. ‘IC1’은 보체계의 활성화, 혈액응고, 피브린의 분해와 근육 속에 생기는 호르몬의 일종인 키닌 합성 등의 물리적인 기전에 관여한다(Aulak, 1993). 이는 ‘SARS-CoV’에 감염된 숙주의 연구에서 면역관련 유전자들의 ‘upregulation’된 패턴과 유사한 결과를 보였다(Lisa et al., 2004). 이 결과를 통하여 코로나 바이러스의 면역 반응에 ‘IC1’ 단백질이 중요한 역할을 하고 있음을 알 수 있었고, 앞으로 이 단백질의 연구가 이루어진다면, 아직 밝혀지지 않은 면역 기전과 항생제나 백신의 연구에 실마리가 될 수 있으리라 사료된다.

4.2 연구의 한계점 및 향후 연구의 제안

4.2.1 연구의 한계점

본 연구를 수행함에 있어서 첫째, 국내 코로나 바이러스 연구의 부족을 확인할 수 있었다. 지난 2015년 ‘MERS’ 사태 이후에 우리나라에서도 코로나 바이러스의 연구가 진행이 되고 있지만, 연구의 많은 부분이 역학적, 임상적 연구들로 이루어져 있었다. 따라서, 현재 우리나라의 코로나 바이러스의 바이오인포매틱스적인 연구가 어디까지 이루어졌는지의 동향을 파악하는데 어려움이 있었다. 또한, 이루어진 연구도 수용체와 관련된 연구이거나, 다른 나라에서 유행했던 코로나 바이러스와의 차이의 연구 등이 대부분이었다. ‘MERS-CoV’의 전파 양상이 우리나라에서 달랐던 경향의 연구가 역학적인 부분도 중요하지만, 우리나라 사람의 면역력과 상관계 등을 규명하는 연구도 굉장히 중요할 것이라 사료된다. 개인별, 인종 별, 대륙 별로 사람의 면역관련 세포의 유전적인 차이가 면역반응의 차이를 불러오고, 이는 유행을 일으키거나 일으키지 않는 원인의 하나로 작용을 한다. 이러한 부분의 면역체계 연구들은 사후 치료가 아닌 사전 예방의 측면에서 중요 하리라 생각한다. 이러한 연구가 활성화되고 많은 기초 자료들이 풍부해진다면, 우리나라의 고유한 특성에 맞는 치료제나 백신의 개발이 가능할 것이며, 미래에는 개인의 면역력에 적합한 치료제와 백신의 개발까지도 가능하고, 국내의 고유성과 차별성을 확보하게 된다면 국제적으로 유용한 연구들로 수용될 수 있을 것이다. 둘째, 코로나 바이러스와 면역 관련 단백질들의 데이터베이스를 구축함에 있어 실질적인 자료의 부족이다. 바이러스의 감염으로 인하여 일어나는 면역반응의 정확한 예측을

위해서는 감염자들의 증상 발현의 정도에 따라서 발견된 바이러스들의 서열 구조와 면역세포들의 유전자 서열의 차이를 분석한다면 보다 정확한 연구가 가능할 것이다. 하지만, 국내에서는 사람에게 질병을 일으키는 코로나 바이러스에 대한 인식이 부족했고, 검출 시도가 미진하여 이를 통한 코로나 바이러스들의 유전자 염기서열이나 관련 면역 데이터 등의 부족으로 신변종 병원체의 출현 및 잠재력을 판단하기에는 가검물의 수가 절대적으로 부족했다. 따라서 한국인의 실정에 적합한 면역 특이성의 연구와 바이오마커의 발굴 등의 연구는 어려운 실정이다. 국가가 일관성 있게 이러한 부분의 기초자료를 수집하고 관리할 수 있는 시스템을 구축한다면, 이를 이용하여 자료들 간의 정보 호환성이 극대화되고, 우리나라의 실정에 맞는 수많은 2차 연구가 가능할 것이라 사료된다. 마지막으로, 분석 부분에서 시스템적인 분석을 진행함에 있어 어려움이 있었다. 이미 국외에서는 선진국을 중심으로 병원체 연구에 있어서 서열 정보뿐만 아니라 면역 작용을 포함한 숙주와의 상호작용에 대한 총체적 분석을 위해서 병원체 정보 데이터베이스 및 분석시스템을 구축해오고 있다. 하지만, 코로나 바이러스와 면역특이성에 적합한 자료는 아직 외국에서도 미비한 상태이므로, 시스템적인 분석을 시행함에 있어서 어려움이 있었다. 향후 연구에서 상호작용에 관한 실질적인 수치와 기전 등을 적용해서 시스템적인 분석이 가능하다면 면역특이성뿐만 아니라 바이오 마커의 발굴, 항원 결정기 예측 등 다양한 분석을 통해서 앞으로 개발해야 할 치료제나 백신의 개발에 도움을 줄 수 있을 것이라 사료된다.

4.2.2 향후 연구의 제안

첫째, 향후 우리나라의 지역적 특성과 인종의 특성, 역학적 특성에 적합한 데이터베이스의 구축이 필요하다. 이러한 국내 특이적 데이터베이스를 기반으로 상호작용의 실질적인 수치나, 강도, 방향성 등의 정보를 고려한 시스템 생물학적 분석이 가능하게 된다면 이는 우리나라 만의 고유성과 다른 나라와의 차별성을 확보하게 되고, 구축된 데이터베이스와 분석법은 많은 연구자들에게 유용하게 사용될 수 있을 것이라고 생각된다. 바이러스와 숙주세포 관련된 단백질이나 유전자의 서열분석이나 상호작용 등을 기반으로 한 시스템적인 면역 기전의 분석은 아직 밝혀지지 않은 면역 회피 기전이나, 병리학적 메커니즘등의 연구에 실마리를 제공하여, 진단과 치료제 연구, 백신 개발 등의 실용적인 연구결과의 도출을 가능하게 할 것이다. 이를 위해서는 코로나 바이러스의 국내 및 국외의 유전자 정보 및 숙주세포의 면역관련 유전자나 단백질의 정보를 지속적으로 업데이트하고, 관리할 수 있는 국가적인 지원과 시스템의 수립이 필요할 것이다. 둘째, 다양한 바이오인포매틱스 방법을 활용하여 코로나 바이러스의 면역 특이성에 대한 분석이 이루어 져야 할 것이다. 향후 연구에서는 바이러스와 숙주 단백질의 아미노산 서열과 3 차원 구조 정보를 이용한 도킹 시뮬레이션을 통해 단백질 간 상호작용을 수치화 하고, 이러한 결과를 이용하여 기존에 알려지지 않은 면역 기전을 예측할 수 있을 것이다. 이러한 연구는 전통적인 면역학을 하는 연구자들에게 실험 설계나 실험으로 설명하지 못하는 부분의 가설 검증 등에 유용하게 이용할 수 있을 것이다. 마지막으로, 바이러스의 감염으로 면역력이 약해진 조직이나 세포에 대한 세균의 기회 감염의 연구가 이루어질 수 있을 것이다. 바이러스와 면역, 세균의 연관성 연구가 이루어지고는

있지만, 아직까지 명확한 기전이 밝혀지지지는 않았다. 이러한 연구를 통하여 병원체와의 면역 기전의 이해가 가능 할 것이며, 병원체의 전파와 감염으로부터 사람을 보호하고 전염을 막을 수 있는 방법을 찾는 데 중요한 역할을 할 수 있을 것이다. 궁극적으로는 감염병으로부터 국민을 보호하고, 국민의 보건 향상에 기여를 할 수 있을 것이다.

4.3 기대 성과

바이러스와 연관된 면역 관련 데이터베이스의 구축은 국외에서는 이미 활발이 진행 중에 있으며, 이를 이용한 다양한 생명정보학적인 분석 방법들이 개발이 되고 있다. 하지만 현재 우리나라는 상용화된 데이터베이스가 아직 미비한 실정이다. 본 연구에서 구축된 데이터베이스의 이용자들은 바이러스와 숙주 간의 면역 특이성 연구를 위해서 코로나 바이러스의 유전자 서열과 생물학적 특성 등의 정보를 쉽게 확인할 수 있고, 항 바이러스성 면역 이전에 관여하는 세포들의 단백질 서열 정보와 생물학적인 정보 등을 체계적으로 이용할 수 있을 것이다. 학술적인 측면에서의 기대 성과는 실험적인 방법이 아닌 바이오인포매틱스 기술을 사용한 코로나 바이러스와 면역 세포들 간의 연구는 축적된 데이터에서 유의미한 정보를 도출하고, 이는 실험 설계 등에 있어 기초 자료로 이용이 될 수 있으리라 기대된다. 기술적 측면에서는 구축된 데이터베이스를 지속적으로 관리하고, 추가적인 정보와 바이오인포매틱스 분석 기법 등을 탑재 한다면 향후 코로나 바이러스와 면역과 관련된 연구자들에게 빠른 분석과 정보 공유의 장을 제공하여 기술적, 시간적 낭비를 줄일 수 있으리라 기대된다. 보건학적 측면에서는, 새로운 감염 병으로 대두되고 있는 사회 문제인 신변종 바이러스의 유행을 예측하고, 바이러스의 변이 등에 따른 인체의 면역반응을 이해하여 사후 조치의 대책이 아닌 사전 예방의 근본적인 대책을 마련 할 수 있는 기초 자료로서의 역할을 하리라 기대된다. 따라서, 본 연구에서 구축한 'HCoV-IMDB'는 코로나 바이러스 관련 면역반응을 이해하는데 유용한 자료를 제공할 수 있을 것이다.

4.4 보건학적 활용방안

본 연구는 코로나 바이러스의 구성 단백질과 면역관련 세포들의 유전자와 단백질을 기반으로 데이터베이스를 구축하여 연구자들에게 편리하고 유용한 정보를 제공하고자 한다. 또한, 코로나 바이러스의 면역 특이성 분석에 바이오인포매틱스 분석 도구를 이용하여 계통수 작성 및 상동성 분석, 바이러스와 숙주 간의 상호작용의 네트워크 분석 등을 시행하여 코로나 바이러스의 노출이나 침입으로 인한 면역 체계의 작용을 이해하고자 하였다. 이러한 분석결과는 아직까지 개발이 진행 중에 있는 코로나 바이러스의 치료제나 백신 개발 등의 연구에 활용이 될 수 있을 것이다. 생물체에 대한 분석법의 발달로 다양한 유전체의 정보가 축적되어 오고 있으며, 최근에는 차세대 염기서열 분석(Next generation sequencing, NGS)등의 기술이 발달하고, 대중화 됨으로써 데이터의 축적 속도와 양이 크게 증가하고 있다. 이렇게 축적된 정보들을 생명정보학을 이용한 총체적 분석이 필요하며, 그 중 데이터베이스의 구축은 다양한 정보들을 체계적으로 저장하여 여러 가지 연구를 진행함에 있어 정보의 수집과 가공 등에 소요되는 시간과 노력을 줄일 수 있으므로 선행되어야 하는 과정이다. 현재 보건학적으로 이슈가 되고 있는 바이러스의 유행은 과거와는 다른 패턴으로 유행이 이루어 지고 있다. 과거에는 인플루엔자 종류의 바이러스들이 유행을 일으키고 대부분의 바이러스들은 특정지역에서 발병하고 사라지는 양상을 보였다. 하지만, 2015년 우리나라를 마비시켰던 ‘MERS’나 2016년 유행을 일으키는 지카 바이러스의 예에서 볼 수 있듯이, 이제 바이러스는 지역을 넘나들며 전염을 시키고 있어서 그 유행 양상을 면밀히 파악하고, 대책을 마련하는 대비책이 시급하다. 기존에 선행된 연구는 대부분이

인플루엔자나, 특정 지역에서 유행을 일으키는 풍토병 관련 바이러스가 주를 이루고 있어, 코로나 바이러스의 연구에 관한 자료는 부족한 실정이다. 이에 본 연구자는 코로나바이러스의 데이터베이스를 구축하였고, 이러한 연구는 새로운 바이러스의 창궐과 변이 바이러스의 예측 등의 기초 자료로 이용이 될 수 있으리라 사료된다. 면역반응은 여러 가지 반응과 다양한 생화학 분자들이 작용하여 만들어 내는 거대한 네트워크이다. 이러한 거대한 네트워크를 이해하기 위해 개별적인 실험 데이터들을 통합하여 시스템적 이해를 돕는 많은 분석 도구들이 개발되어왔다. 암 연구와 비교해서 바이러스와 면역 시스템의 상호작용에 대한 시스템 생물학적 연구는 아직 많이 부족한 실정이다. 네트워크 분석 기법을 포함한 바이오인포매틱스 방법은 우리나라에서 유행하는 주요 병원체에 대한 한국인의 면역 특이성 인자 등의 발굴에 유용하게 사용될 수 있을 것이라 생각한다. 본 연구에서 구축한 'HCoV-IMDB'는 코로나 바이러스 관련 면역반응을 이해하는데 유용한 면역·분자 생물학적 근거자료를 제공할 수 있으며, 특화된 정보를 통해 감염병 연구에서의 활용성을 증대시킬 수 있으리라 사료된다.

제 5 장. 결론 및 총론

5.1 결 론

교통 수단의 발달과 사회 경제적인 발달로 인하여 무역 등이 활발해지고, 내국인의 외국 여행이 증가하고, 외국인의 유입이 활발해짐으로써, 우리의 주변 환경이 상당히 변화하고 있다. 그 대표적인 예로, 과거에는 국한된 지역에서 유행하던 바이러스들이 이동을 하여 전혀 새로운 지역에서 출몰을 하고, 특정 종에서만 감염을 일으키던 바이러스가 다른 종에도 감염을 일으키는 패턴을 보이고 있다. 2003년 유행한 ‘SARS’는 가금류에게서 사람으로의 감염을 일으켜 문제가 되었고, 2015년 한국에서 발생한 ‘MERS’는 중동아시아 지역에서만 유행하던 바이러스였으나, 그곳을 다녀온 사람을 통하여 이동한 바이러스는 한국에서 엄청난 속도로 감염을 일으키고 사상자를 발생 시켰다. 이러한 환경의 변화는 우리에게 바이러스의 대비책에 대한 새로운 방향으로의 과제를 남겨 주었다. 단순히, 감염이 일어나고서 마련하는 후속 대책보다는 보건학 적인 측면으로의 접근과 방향성을 가지고, 새로운 바이러스의 유행의 예측과 변종 바이러스에 대한 대비책과 개개인의 건강을 증진 시킬 수 있는 방안의 마련이 이루어 져야 할 것이다.

본 연구에서는 이러한 필요성에 따라서 코로나바이러스의 면역 특이성 연구를 위한 데이터베이스인 ‘HCoV-IMDB’를 구축 하였으며, 이는 코로나 바이러스의 모든 단백질과 면역세포 단백질들의 유전자 서열과 정보를 포함한 국내 유일의

데이터베이스다. 사람에게 질병을 일으키는 코로나 바이러스들 중에서 우리가 인지하지 못하고 있는 바이러스의 위험성을 예측하기 위하여 계통수 분석을 시행하였다. 계통수 작성의 결과, 코로나 바이러스와 숙주세포와의 부착에 관여하는 S 단백질이 ‘HCoV-OC43’과 ‘MERS-CoV’의 거리가 0.946으로 진화적으로 가까운 거리에 있음을 알 수 있었다. 이러한 결과로 우리에게 감기 정도를 일으키는 감염성이 낮다고 간주했던 ‘HCoV-OC43’ 바이러스가 변이 등이 일어난다면 ‘MERS-CoV’와 유사한 유행이 일어날 수 있으므로 ‘HCoV-OC43’의 지속적인 감시와 변이 여부의 조사가 이루어져야 할 것이다. 인터페론의 생성에 영향을 미치는 N 단백질의 계통수 분석을 시행한 결과, ‘SARS-CoV’와 ‘MERS-CoV’의 N 단백질의 거리가 0.637으로 진화적으로 유사한 서열임을 알 수 있었다. 이는 ‘SARS-CoV’와 ‘MERS-CoV’의 유사한 감염 양상을 설명할 수 있을 것이다. ‘SARS-CoV’ 단백질과 사람 단백질 간의 상호작용을 이용하여 네트워크 분석을 시행 하였다. 분석 결과 사람 단백질의 ‘IC1’ 단백질과 ‘SARS-CoV’의 ‘Nsp3b’, ‘Nsp7b’, ‘Nsp13-pp1ab’, ‘Nsp8-pp1a/pp1ab’, ‘Nsp2-pp1a/pp1ab’, ‘Nsp14-pp1ab’ 단백질들이 상호 작용함을 확인하였다. 이 결과를 통하여 코로나 바이러스의 면역 반응에 ‘Plasma protease C1 inhibitor’ 단백질이 중요한 역할을 하고 있음을 알 수 있었고, 앞으로 이 단백질의 연구가 이루어 진다면, 아직 밝혀지지 않은 면역 기전과 항생제나 백신의 연구에 실마리가 될 수 있으리라 사료된다. 과학 기술의 발달로 수많은 정보들이 축적되고는 있지만 전통적인 방법으로는 복잡한 면역시스템을 이해 하기에는 그 시간과 노력이 상당 시간 필요하다. 따라서 이러한 데이터들을 체계적이고 효율적으로 관리하고, 분석 할 수 있는 생명정보학이 필요하며, 이를 활용한 다양한 연구가 그 시간과 노력을 절약시켜 줄 수 있을 것이다. 본 연구에서는 빠르게

진화하는 바이러스와 면역세포들간의 관계를 이해하고자 전산 기술을 이용하여 산재되어 있는 정보를 수집하고 분류하여 데이터베이스를 구축하였으며, 코로나 바이러스의 단백질과 면역단백질들의 서열 데이터의 다운로드가 가능하도록 하였다. 또한 사용자가 쿼리를 이용하여 상동성 검색이 가능한 BLAST를 구축하였다. 향후 지속적인 데이터의 업데이트와 바이오인포매틱스 분석 기법을 보완한다면, 이는 바이러스와 면역 기전의 연구에 기반이 되는 데이터베이스로 유용하게 이용 될 수 있으리라 기대되며, 아직까지 개발하지 못한 치료제나 백신의 개발에 도움을 줄 수 있으리라 사료된다. 이는 궁극적으로는 국민의 보건 향상에 이바지 할 수 있을 것이다.

5.2 총 론

2015년 한국에서 발생한 ‘MERS-CoV’의 유행 후에 코로나 바이러스를 재조명하게 되었다. 연구의 목적은 첫째, 코로나 바이러스와 면역관련 단백질과 유전자의 정보들을 제공하는 데이터베이스를 구축하여 연구자들에게 유용한 정보를 제공해 주고자 하였다. 연구 결과 ‘HCoV-IMDB’를 구축하였으며, 구축된 데이터베이스에서는 코로나바이러스의 구조와 비구조 단백질, 면역관련 단백질들의 서열 정보와 특징을 카테고리 별로 분류하여 저장해 놓았다. 사용자들은 데이터베이스에 손쉽게 접근이 가능하며, 필요한 단백질들의 정보를 다운로드가 가능하고 BLAST 서버를 이용하여 쿼리시퀀스에 대한 상동성 검색이 가능하도록 구축하였다. 둘째, 바이오인포매틱스 방법을 활용한 연구방법을 통하여 기존에 알려지지 않았던 코로나바이러스의 위험성을 예측하여 바이러스에 대한 변이 여부의 지속적인 감시 등의 필요성을 제안하였다. 계통수 분석을 이용한 연구결과를 통해 ‘HCoV-OC43’의 S 단백질이 ‘MERS-CoV’와 진화적으로 유사하였으며, 이는 ‘HCoV-OC43’의 변이 여부의 지속적인 감시가 필요함을 제시하였다. 셋째, 코로나 바이러스의 면역 특이성 분석을 통하여 코로나 바이러스와 면역단백질들을 재조명 함으로써 아직 밝혀지지 않은 면역 기전의 이해의 실마리를 제공해 주고자 하였다. 네트워크 분석을 이용한 연구결과를 통해 사람 단백질의 ‘IC1’ 단백질과 ‘SARS’의 여러 가지 종류의 비구조 단백질들이 상호 작용함을 확인하였다. ‘IC1’은 보체계에 관여하는 단백질로 바이러스의 면역에서 보체계의 역할을 다시 한번 재조명하게 되었으며, 밝혀지지 않은 면역 기전의 이해에 실마리가 될 수 있을 것이다.

참 고 문 헌

- Armstrong J, Niemann H, Smeekens S, Rottier P, Warren G. Sequence and topology of a model intracellular membrane protein, E1 glycoprotein, from a corona virus. *Nature*. 308(5961):751-752(1984).
- Aronson DS, Horvath CM. A road map for those who don't know JAK-Stat. *Science* 296:1653-1655(2002).
- Aulak KS, Davis AE. III, Donaldson VH, Harrison RA. Chymotrypsin inhibitory activity of normal C1-inhibitor and a P1 Arg to His mutant: evidence for the presence of overlapping reactive centers. *Protein Sci*. 2:727-732(1993).
- Azhar EI, El-Kafrawy SA, Farraj SA, Hassan AM, Al-Saeed MS, Hashem AM, Madani TA. Evidence for camel-to-human transmission of MERS coronavirus. *N Engl J Med* 370:2499-2505(2014).
- Banik GR, Khandaker G, Rashid H. Middle East Respiratory Syndrome Coronavirus "MERS-CoV": Current Knowledge Gaps. *Paediatr respir rev*, 16:197-202(2015).
- Bao Y, Bolotov P, Dernovoy D, Kiryutin B, Zaslavsky L, Tatusova T, Ostell J, Lipman D. The influenza virus resource at the National Center for Biotechnology Information. *JVirol*82:596-601(2008).
- Barry GH. *Phylogenetic Trees made easy*. Sinauer, Sunderland, MA(2001).
- Batagelj V, Mrvar A. Pajek – Program for Large Network Analysis. *Connections* 21:47-57(1998).
- Beeler JA, Eichelberger. Influenza and respiratory syncytial virus (RSV) vaccines for infants: safety, immunogenicity and efficacy. *Microb.Pathog*. 55: 9-15(2013).
- Berry DM, Almeida JD. The morphological and biological effects of various antisera on avian infectious bronchitis virus. *J GenVirol* 3:97-102(1968).
- Bhardwaj K, Sun J, Holzenburg A. RNA recognition and cleavage by the SARS coronavirus endoribonuclease. *J Mol Biol* 361:243-256(2006).
- Bogner P, Capua I, Cox NJ, Lipman DJ. A global initiative on sharing avian flu data. *Nature* 442:981(2006).
- Bomsel M, Heyman M, Hocini H, Lagaye S, Belec L, Dupont C, Desgranges C. Intracellular neutralization of HIV transcytosis across tight epithelial barriers by anti-HIV envelope protein IgA or IgM. *Immunity* 9:277-287(1998).

Bose ME, Littrell JC, Patzer AD, Kraft AJ, Metallo JA, Fan J, Henrickson KJ. The Influenza Primer Design Resource: a new tool for translating influenza sequence data into effective diagnostics. *Influenza Other Respir Viruses* 2:23–31(2008).

Bouvet M, Debarnot C, Imbert I. In vitro reconstitution of SARS-coronavirus mRNA cap methylation. *PLoS Pathog* 6:e1000863(2010).

Bradburne AF, Bynoe ML, Tyrell DAJ. Effects of a “new” human respiratory virus in volunteers. *Br Med J*.3:767–769(1967).

Breitkreutz : Osprey: a network visualization system. *Genome Biol* 4(3):R22(2003).

Bruey JM, Bruey-Sedano N, Luciano F, Zhai D, Balpai R, Xu C, Kress CL, Bailly-Maitre B, Li X, Osterman A, Matsuzawa S, Terskikh AV, Faustin B, Reed JC. Bcl-2 and Bcl-XL regulate proinflammatory caspase-1 activation by interaction with NALP1. *Cell* 129:45-56(2007).

Burns JW, Siadat-Pajouh M, Krishnaney AA, Greenberg HB. Protective effect of rotavirus VP6-specific IgA monoclonal antibodies that lack neutralizing activity. *Science* 272:104-107(1996).

Butler D. South Korean MERS outbreak spotlights lack of research. *Nature*, 522:139-40(2015).

Castle CD, Cassimere EK, Lee J, Denicourt C. Las1L is a nucleolar protein required for cell proliferation and ribosome biogenesis. *Mol. Cell. Biol.* 30:4404-4414(2010).

Cate TR. Impact of influenza and other community-acquired viruses. *Semin.Respir. Infect.* 13: 17-23 (1998).

Chan JF, Lau SK, To KK. Middle East respiratory syndrome coronavirus: another zoonotic betacoronavirus causing SARS like disease. *Clin Microbiol Rev.* 28:465 (2015).

Chang CK, Sue SC, Yu TH, Hsieh CM, Tsai CK, Chiang YC, Lee SJ, Hsiao HH, Wu WJ, Chang WL, Lin CH, Huang TH. Modular organization of SARS coronavirus nucleocapsid protein. *Journal of biomedical science.* 13(1):59–72(2006).

Chang S, Zhang J, Liao X, Zhu X, Wang D, Zhu J, Feng T, Zhu B, Gao GF, Wang J, Yang H, Yu J. Influenza Virus Database (IVDB): an integrated information resource and analysis platform for influenza virus research. *Nucleic Acids Res* 35:D376–D38(2007).

Chao DL, Halloran ME, Obenchain VJ, Longini IM, Jr. FluTE, a publicly available stochastic influenza epidemic simulation model. *PLoS Comput Biol*6:e1000656(2010).

- Chatr-aryamontri A, Ceol A, Peluso D, Nardoza A, Panni S, Sacco F, Tinti M, Smolyar A, Castagnoli L, Vidal M, Cusick ME, Cesareni G. VirusMINT: a viral protein interaction database. *Nucleic Acids Res* 37:D669–D673(2009).
- Chen J, Lau YF, Lamirande EW, Paddock CD, Bartlett JH, Zaki SR. Cellular immune responses to severe acute respiratory syndrome coronavirus (SARS-CoV) infection in senescent BALB/c mice: CD4⁺ T cells are important in control of SARS-CoV infection. *J Virol*. 84:1289–301(2010).
- Chen X, Yang X, Zheng Y, Yang Y, Xing Y, Chen Z. SARS coronavirus papain-like protease inhibits the type I interferon signaling pathway through interaction with the STING-TRAF3- TBK1 complex. *Protein Cell* . 5:369-381(2014).
- Chen Y, Cai H, Pan J. Functional screen reveals SARS coronavirus nonstructural protein nsp14 as a novel cap N7 methyltransferase. *Proc Natl Acad Sci U S A* 106:3484–3489(2009).
- Cheung CY, Poon LL, Ng IH, Luk W, Sia SF, Wu MH. Cytokine responses in severe acute respiratory syndrome coronavirus-infected macrophages in vitro: possible relevance to pathogenesis. *J Virol*. 79:7819–26(2005).
- Chibo D, Birch C. Analysis of human coronavirus 229E spike and nucleoprotein genes demonstrates genetic drift between chronologically distinct strains. *J Gen Virol*. 87: 1203–1208(2006).
- Chinese SMEC. Molecular evolution of the SARS coronavirus during the course of the SARS epidemic in China. *Science*. 303:1666-1669(2004).
- Choi EH. Middle East Respiratory Syndrome Outbreak in Korea, 2015 *Pediatr Infect Vaccine* 22:131-135(2015).
- Chu H, Zhou J, Wong BH, Li C, Chan JF, Cheng ZS, Yang D, Wang D, Lee AC, Li C. Middle East Respiratory Syndrome Coronavirus Efficiently Infects Human Primary T Lymphocytes and Activates the Extrinsic and Intrinsic Apoptosis Pathways. *J Infect Dis*. Epub ahead of print(2015).
- Chu H, Zhou J, Wong BH, Li C, Cheng ZS, Lin X. Productive replication of Middle East respiratory syndrome coronavirus in monocyte-derived dendritic cells modulates innate immune response. *Virology*. 454–455:197–205(2014).
- Clementz MA, Kanjanahaluethai A, O'Brien TE.. Mutation in murine coronavirus replication protein nsp4 alters assembly of double membrane vesicles. *Virology* 375:118–129(2008).
- Cooper NR, Jensen FC, Welsh RM, Oldstone MB. Lysis of RNA tumor viruses by human serum: direct antibody-independent triggering of the classical complement pathway. *J Exp Med* 144:970-984(1976).

- Decroly E, Imbert I, Coutard B. Coronavirus nonstructural protein 16 is a cap-0 binding enzyme possessing (nucleoside- 2'O)-methyltransferase activity. *J Virol* 82: 8071–8084(2008).
- Demir OB E, Dogrusoz U, GURSOY A, NISANCI G, CETIN-ATALAY R, OZTURK M: PATIKA: An Integrated Visual Environment for Collaborative Construction and Analysis of Cellular Pathways. *Bioinformatics* 18:996-1003(2002).
- Devaraj SG, Devaraj N, Wang Z, Chen Z, Chen M, Tseng N, Barretto R, Lin CJ, Peters CT, Tseng K, Baker SC. Regulation of IRF-3-dependent innate immunity by the papain-like protease domain of the severe acute respiratory syndrome coronavirus. *Journal of Biological Chemistry*, 282:32208–32221(2007).
- Dutta S, Tan YJ. Structural and functional characterization of human SGT and its interaction with Vpu of the human immunodeficiency virus type 1. *Biochemistry* 47:10123-10131(2008).
- Eckerle I, Corman VM, Muller MA, Lenk M, Ulrich RG, Drosten C. Replicative Capacity of MERS Coronavirus in Livestock Cell Lines. *Emerging infectious diseases*. 20(2):276–279(2014).
- Eckerle LD, Lu X, Sperry SM. High fidelity of murine hepatitis virus replication is decreased in nsp14 exoribonuclease mutants. *J Virol* 81:12135–12144(2007).
- Egloff MP, Ferron F, Campanacci V. The severe acute respiratory syndrome-coronavirus replicative protein nsp9 is a single-stranded RNA-binding sub-unit unique in the RNA virus world. *Proc Natl Acad Sci U S A* 101:3792–3796(2004).
- Egloff MP, Malet H, Putics A. Structural and functional basis for ADP-ribose and poly(ADP-ribose) binding by viral macro domains. *J Virol* 80:8493–8502(2006).
- Enjuanes L, Smerdou C, Castilla J. Development of protection against coronavirus induced diseases. A review. *Adv Exp Med Biol* 380:197(1995).
- Eriksson KK, Cervantes-Barragan L, Ludewig B. Mouse hepatitis virus liver pathology is dependent on ADP-ribose-1"- phosphatase, a viral function conserved in the alpha-like supergroup. *J Virol* 82:12325– 12334(2008).
- Hwang ES, Park CG, Cha CY. Immune Responses to Viral Infection. *Immune Network* .4(2):73-80(2004).
- Felsenstein J. *Inferring Phylogeny*. Sinauer, Sunderland, MA(2003).
- Fendrick AM, Monto AS, Nightengale B, Sarnes M. The economic burden of non-influenza-related viral respiratory tract infection in the United States. *Arch Intern Med* 163:487–494(2003).

Freeman TC, Goldovsky L, Brosch M, van Dongen S, Maziere P, Grocock RJ, Freilich S, Thornton J, Enright AJ: Construction, visualisation, and clustering of transcription networks from microarray expression data. *PLoS computational biology* 3(10):2032-2042(2007).

Freundt EC, Yu L, Park E, Lenardo MJ, Xu XN. Molecular determinants for subcellular localization of the severe acute respiratory syndrome coronavirus open reading frame 3b protein. *Journal of Virology* 83:6631–6640(2009).

Frieman M, Ratia K, Johnston RE, Mesecar AD, Baric RS. Severe acute respiratory syndrome coronavirus papain-like protease ubiquitin-like domain and catalytic domain regulate antagonism of IRF3 and NF- κ B signaling. *Journal of Virology*. 83:6689–6705(2009).

Frieman M, Yount B, Heise M, Kopecky-Bromberg SA, Palese P, Baric RS. Severe acute respiratory syndrome coronavirus ORF6 antagonizes STAT1 function by sequestering nuclear import factors on the rough endoplasmic reticulum/Golgi membrane. *J Virol* 81, 9812–9824(2007).

Fu W, Sanders-Bear BE, Katz KS, Maglott DR, Pruitt KD, Ptak RG. Human immunodeficiency virus type 1, human protein interaction database at NCBI. *Nucleic Acids Res* 37:D417–D422(2009).

Gadlage MJ, Sparks JS, Beachboard DC. Murine hepatitis virus nonstructural protein 4 regulates virus-induced membrane modifications and replication complex function. *J Virol* 84:280–290(2010).

GeneMarta L., DeDiegoa ML, Jose L, Nieto T, Jose M, Jimenez G, Jose A, Regla.N, Carlos C, Rodriguez, Raul FD, Fernando U, Luis E. Coronavirus virulence genes with main focus on SARS-CoV envelope gene. *Virus Research* 194 :124–137(2014).

Gibson L, Holmgren SP, Huang DC, Bernard O, Copeland NG, Jenkins NA, Sutherland GR, Baker E, Adams JM, Cory S. Bcl-w, a novel member of the Bcl-2 family, promotes cell survival. *Oncogene* 13:665-675(1996).

Glezen WP. Influenza surveillance in an urban area. *Can J Infect Dis* 4:272-4.26(1993).

Goodbourn S, Didcock L, Randall RE. Interferon: cell signaling, immune modulation, antiviral response and virus countermeasures. *J. Gen. Virol.* 81:2341-2364(2000).

Graham RL, Sims AC, Brockway SM. The nsp2 replicase proteins of murine hepatitis virus and severe acute respiratory syndrome coronavirus are dispensable for viral replication. *J Virol* 79:13399–13411 (2005).

Guan Y, Zheng BJ, He YQ, Liu XL, Zhuang ZX, Cheung CL, Luo SW, Li PH, Zhang LJ, Guan YJ, Butt KM, Wong KL, Chan KW, Lim W, Shortridge KF, Yuen KY, Peiris JS, Poon LL. Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. *Science*. 302:276–278(2003).

Haan CA, Wit M, Kuo L, Montalto-Morrison C, Haagmans BL, Weiss SR, Masters PS, Rottier PJ. The glycosylation status of the murine hepatitis coronavirus M protein affects the interferogenic capacity of the virus in vitro and its ability to replicate in the liver but not the brain. *Virology* 312:395–406(2003).

Hamre D, Procknow JJ. A new virus isolated from the human respiratory tract. *Proceedings of the Society for Experimental Biology and Medicine Society for Experimental Biology and Medicine*. 121(1):190–193(1966).

Heil F, Hemmi H, Hochrein H, Ampenberger F, Kirschning C, Akira S, Lipford G, Wagner H, Bauer S. Species-specific recognition of single-stranded RNA via toll-like receptor 7 and 8. *Science* 303:1526–9(2004).

Hermjakob H, Montecchi PL, Bader G, Wojcik J, Salwinski L, Ceol A, Moore S, Orchard S, Sarkans U, von Mering C, Roechert B, Poux S, Jung E, Mersch H, Kersey P, Lappe M, Li Y, Zeng R, Rana D, Nikolski M, Husi H, Brun C, Shanker K, Grant SG, Sander C, Bork P, Zhu W, Pandey A, Brazma A, Jacq B, Vidal M, Sherman D, Legrain P, Cesareni G, Xenarios I, Eisenberg D, Steipe B, Hogue C, Apweiler R: The HUPO PSI's molecular interaction format – a community standard for the representation of protein interaction data. *Nat Biotechnol* 22(2):177–183(2004).

Hermjakob H, Montecchi PL, Lewington C, Mudali S, Kerrien S, Orchard S, Vingron M, Roechert B, Roepstorff P, Valencia A, Margalit H, Armstrong J, Bairoch A, Cesareni G, Sherman D, Apweiler R: IntAct: an open source molecular interaction database. *Nucleic Acids Res* D452–455(2004).

Hooper SD, Bork P: Medusa: a simple tool for interaction graph analysis. *Bioinformatics* 21(24):4432–4433(2005).

Huang C, Lokugamage KG, Rozovics JM, Narayanan K, Semler BL, Makino S. SARS coronavirus nsp1 protein induces template-dependent endonucleolytic cleavage of mRNAs: viral mRNAs are resistant to nsp1-induced RNA cleavage. *PLoS Pathogens*, 7, e1002433(2011).

Huang C, Lokugamage KG, Rozovics JM. Alphacoronavirus transmissible gastroenteritis virus nsp1 protein suppresses protein translation in mammalian cells and in cell-free HeLa cell extracts but not in rabbit reticulocyte lysate. *J Virol* 85:638–643(2011).

Huang Y, Lau SK, Woo PC, Yuen KY. CoVDB: a comprehensive database for comparative analysis of coronavirus genes and genomes. *Nucleic Acids Res*. 36:504–511(2008).

Hui DS, Memish ZA, Zumla A. Severe acute respiratory syndrome vs. the Middle East respiratory syndrome. *Curr Opin Pulm Med.* 20:233-241(2014).

Imbert I, Guillemot JC, Bourhis JM. A second, non-canonical RNA-dependent RNA polymerase in SARS corona virus. *EMBO J* 25:4933–4942(2006).

International Committee on Taxonomy of Viruses. [http:// ictvonline. org/ virus Taxonomy.asp](http://ictvonline.org/virusTaxonomy.asp)

Ivanov KA, Thiel V, Dobbe JC. Multiple enzymatic activities associated with severe acute respiratory syndrome corona virus helicase. *J Virol* 78:5619–5632(2004).

Je SM, Bae WJ, Kim JY, Seok SH, Hwang ES. The Preparedness for Re-emerged and Emerging Infectious Diseases: A Lesson Through Outbreak of Middle East Respiratory Syndrome Coronavirus (MERS-CoV) in South Korea. *Journal of Bacteriology and Virology* 45:339–353(2015).

Kaetzel CS, Robinson JK, Lamm ME. Epithelial transcytosis of monomeric IgA and IgG cross-linked through antigen to polymeric IgA. A role for monomeric antibodies in the mucosal immune system. *J Immunol* 152:72-76 (1994).

Kamitani W, Huang C, Narayanan K, Lokugamage KG, Makino S. A two-pronged strategy to suppress host protein synthesis by SARS coronavirus Nsp1 protein. *Nat Struct Mol Biol* 16: 1134–1140(2009).

Kamitani W, Huang C, Narayanan K, Lokugamage KG, Makino S. A two-pronged strategy to suppress host protein synthesis by SARS coronavirus Nsp1 protein. *Nature Structural & Molecular Biology*, 16:1134–1140(2009).

Kato H, Sato S, Yoneyama M, Yamamoto M, Uematsu S, Matsui K, Tsujimura T, Takeda K, Fujita T, Takeuchi O, Akira S. Cell type-specific involvement of RIG-I in antiviral response. *Immunity* 23:19-28(2005).

Kim DW, Yun MR, Yang JS, Kang HJ, Han YW, Lee HS, Kim HM, Kim H, Kim AR, Heo DR, Kim SJ, Jeon JH, Park DB, Kim JA, Cheong HM, Nam JG, Kim K, Kim SS, Kim YJ, Park SH. Variations in Spike Glycoprotein Gene of MERS-CoV, South Korea, 2015. *Emerging Infectious Diseases* 22:100-104(2016).

Kim Y, Cheon S, Min C-K, Sohn KM, Kang Y, Cha Y-J, Kang J-I, Han SK, Ha N-Y, Kim G, Aigerim A, Shin HM, Choi M-S, Kim S, Cho H-S, Kim Y-S, Cho N-H. 2016. Spread of mutant Middle East respiratory syndrome coronavirus with reduced affinity to human CD26 during the South Korean outbreak. *mBio* 7(2):e00019-16(2016).

Köhler JBJ, Jan T, Michael S, Andre S, Alexander R, Chris R, Paul V, Stephan P: Graph-based analysis and visualization of experimental results with ONDEX. *Bioinformatics* 22(11):1383-1390(2006).

Konig R, Zhou Y, Elleder D, Diamond TL, Bonamy GM, Irelan JT, Chiang CY, Tu BP, De Jesus PD, Lilley CE, Seidel S, Opaluch AM, Caldwell JS, Weitzman MD, Kuhen KL, Bandyopadhyay S, Ideker T, Orth AP, Miraglia LJ, Bushman FD, Young JA, Chanda SK. Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication. *Cell* 135:49-60(2008).

Kopecky-Bromberg SA, Martinez-Sobrido L, Frieman M, Baric RA, Palese P. Severe acute respiratory syndrome coronavirus open reading frame (ORF) 3b, ORF 6, and nucleocapsid proteins function as interferon antagonists. *J Virol.*81:548–57(2007).

Kopecky-Bromberg SA, Martinez-Sobrido L, Frieman M, Baric RA, Palese P. Severe acute respiratory syndrome coronavirus open reading frame (ORF) 3b, ORF 6, and nucleocapsid proteins function as interferon antagonists. *Journal of Virology*, 81:548–557(2007).

Korea Centers for Disease Control and Prevention. Middle East Respiratory Syndrome Information. June 2015; [http:// www.mers.go.kr/mers/html/jsp/main.jsp](http://www.mers.go.kr/mers/html/jsp/main.jsp) (Accessed 1 Oct, 2015)

Korea Centers for Disease Control and Prevention. Middle East Respiratory Syndrome Coronavirus Outbreak in the Republic of Korea, Osong Public Health Res Perspect 6:269-78(2015).

Kuo L, Masters PS. Genetic evidence for a structural interaction between the carboxy termini of the membrane and nucleocapsid proteins of mouse hepatitis virus. *J Virol* 76:4987(2002).

Laiosa CV, Stadtfeld M, Graf T. Determinants of lymphoid-myeloid lineage diversification. *Annu.Rev.Immunol.*24:705-738(2006).

Lamm ME. Interaction of antigens and antibodies at mucosal surfaces. *Annu Rev Microbiol* 51:311-340(1997).

Law HK, Cheung CY, Ng HY, Sia SF, Chan YO, Luk W. Chemokine up-regulation in SARS-coronavirus-infected, monocyte-derived human dendritic cells. *Blood.*106:2366–74 (2005).

Lenzerini M. Data Integration: A Theoretical Perspective. *PODS* :243-246(2002).

Li W, Moore MJ, Vasilieva N, Sui J, Wong SK, Berne MA, Somasundaran M, Sullivan JL, Luzuriaga K, Greenough TC. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* .426:450-454(2003).

Liao YC, Ko CY, Tsai MH, Lee MS, Hsiung CA. ATIVS: analytical tool for influenza virus surveillance. *Nucleic Acids Res* 37:W643– W646(2009).

- Liechti R, Gleizes A, Kuznetsov D, Bougueleret L, Le Mercier P, Bairoch A, Xenarios I. OpenFluDB, a database for human and animal influenza virus. Database (Oxford):baq004(2010).
- Lin FK, Pan CL, Yang JM, Chuang TJ, Chen FC. CAPIH: a Web interface for comparative analyses and visualization of host-HIV protein-protein interactions. BMC Microbiol 9:164(2009).
- Lisa EG, Ralph SB. Molecular pathology of emerging coronavirus infections. J Pathol 235:185-195(2015).
- Lu G, Rowley T, Garten R, Donis RO. FluGenome: a web tool for genotyping influenza A virus. Nucleic Acids Res 35:W275–W279(2007).
- Lu Y, Lu X, Denison MR. Identification and characterization of a serine-like proteinase of the murine coronavirus MHV-A59. J Virol 69:3554–3559(1995).
- Luytjes W, Bredenbeek PJ, Noten AF. Sequence of mouse hepatitis virus A59 mRNA 2: indications for RNA recombination between coronaviruses and influenza C virus. Virology.166:415(1988).
- Mazanec MB, Coudret CL, Fletcher DR. Intracellular neutralization of influenza virus by immunoglobulin A anti-hemagglutinin monoclonal antibodies. J Virol 69:1339-1343(1995).
- Marta LD, Jose LN, Regla JM, Jimenez G, Raul FB, Craig F, Carlos CR, Stanley P, Luis E. Inhibition of NF- κ B-Mediated Inflammation in Severe Acute Respiratory Syndrome Coronavirus-Infected Mice Increases Survival. J. Virol. 88:913-924(2014).
- Maryam NA, Kim TN, Xiang YO, Yong KP, Yutaka T, Jack BC, Nik SH, Adeeba K, Kok KT. Molecular epidemiology and evolutionary histories of human coronavirus OC43 and HKU1 among patients with upper respiratory tract infections in Kuala Lumpur, Malaysia. Virol J.13: 33(2016).
- Masters PS, Kuo L, Ye R. Genetic and molecular biological analysis of protein-protein interactions in coronavirus assembly. Adv Exp Med Biol 581:163(2006).
- Masters PS, Perlman S. Coronaviridae. In: Fields Virology, 6th edition, Knipe DM, Howley PM (Eds), Lippincott Williams & Wilkins, Philadelphia(2013).
- Masters PS. The molecular biology of coronaviruses. Adv. Virus Res. 66:193–292(2006).
- Matthews KL, Coleman CM, van der Meer Y, Snijder EJ, Frieman MB. The ORF4b-encoded accessory proteins of Middle East respiratory syndrome coronavirus and two

related bat coronaviruses localize to the nucleus and inhibit innate immune signalling. *J Gen Virol* 95:874–82(2014).

Maurer U, Charvet C, Wagman A.S, DeJardin E, Green D.R. Glycogen synthase kinase-3 regulates mitochondrial outer membrane permeabilization and apoptosis by destabilization of MCL-1. *Mol. Cell* 21:749-760(2006).

Mazanec MB, Lamm ME, Lyn D, Portner A, Nedrud JG. Comparison of IgA versus IgG monoclonal antibodies for passive immunization of the murine respiratory tract. *Virus Res* 23;1-12(1992).

McIntosh K, Becker WB, Chanock RM. Growth in suckling-mouse brain of “IBV-like” viruses from patients with upper respiratory tract disease. *Proceedings of the National Academy of Sciences of the United States of America* 58:2268–2273(1967).

Menachery VD, Yount BL Jr, Josset L, Gralinski LE, Scobey T, Agnihothram S, Katze MG, Baric RS. Attenuation and restoration of severe acute respiratory syndrome coronavirus mutant lacking 20-o-methyltransferase activity. *J Virol* 88:4251- 4264(2014).

Menachery VD, Yount JBL, Josset L, Gralinski LE, Scobey T, Agnihothram S, Katze MZ, Baric BS. Attenuation and restoration of severe acute respiratory syndrome coronavirus mutant lacking 2'-o-methyltransferase activity. *J Virol*, 88:4251–4264(2014).

Mihaescu R, Levy D, Pachter L. Why neighbor-joining works. *Algorithmica* 54 (1): 1–24(2009).

Minskaia E, Hertzog T, Gorbalenya AE. Discovery of an RNA virus 3'->5' exoribonuclease that is critically involved in coronavirus RNA synthesis. *Proc Natl Acad Sci U S A* 103:5108–5113(2006).

Moni MA, Liò P. Network-based analysis of comorbidities risk during an infection: SARS and HIV case studies. *BMC Bioinformatics* 15:333(2014).

Nagarajan N, Kingsford C. GiRaF: robust, computational identification of influenza reassortments via graph mining. *Nucleic Acids Res* 39:e34(2011).

Nal B, Chan C, Kien F, Siu L, Tse J, Chu K, Kam J, Staropoli I, Crescenzo-Chaigne B, Escriou N, van der Werf S, Yuen KY, Altmeyer R. Differential maturation and subcellular localization of severe acute respiratory syndrome coronavirus surface proteins S, M and E. *The Journal of general virology*. 86(5):1423–1434(2005).

Navratil V, deChasse B, Meyniel L, Delmotte S, Gautier C, Andre P, Lotteau V, Rabourdin-Combe C. VirHostNet: a knowledge base for the management and the

analysis of proteome-wide virus-host interaction networks. *Nucleic Acids Res* 37:D661–D668(2009).

Neuman BW, Joseph JS, Saikatendu KS. Proteomics analysis unravels the functional repertoire of coronavirus nonstructural protein 3. *J Virol* 82:5279–5294(2008).

Neuman BW, Kiss G, Kunding AH, Bhella D, Baksh MF, Connelly S, Droese B, Klaus JP, Makino S, Sawicki SG, Siddell SG, Stamou DG, Wilson IA, Kuhn P, Buchmeier MJ. A structural analysis of M protein in coronavirus assembly and morphology. *Journal of structural biology*. 174(1):11–22(2011).

Ng LF, Hibberd ML, Ooi EE, Tang KF, Neo SY, Tan J, Murthy KR, Vega VB, Chia J. A human in vitro model system for investigating genome-wide host responses to SARS coronavirus infection. *BMC Infect Dis* 4: 34(2004).

Niemeyer D, Zillinger T, Muth D, Zielecki F, Horvath G, Suliman T. Middle East respiratory syndrome coronavirus accessory protein 4a is a type I interferon antagonist. *J Virol*. 87:12489–95(2013).

NirOrlev RS, Yosef S. PIVOT: protein interactions visualization tool. *Bioinformatics* (2003).

Oboho IK, Tomczyk SM, Al-Asmari AM, Banjar AA, Al-Mugti H, Aloraini MS, Alkhaldi KZ, Almohammadi EL, Alraddadi BM, Gerber SI. MERS-CoV outbreak in Jeddah — a link to health care facilities. *N Engl J Med*. 372:846-854(2015).

Oostra M, Hagemeijer MC, van Gent M. Topology and membrane anchoring of the coronavirus replication complex: not all hydrophobic domains of nsp3 and nsp6 are membrane spanning. *J Virol* 82:12392–12405(2008).

Palmeira L, Penel S, Lotteau V, Rabourdin-Combe C, Gautier C. PhEVER: a database for the global exploration of virus-host evolutionary relationships. *Nucleic Acids Res* 39:D569–D575(2011).

Peiris JS, Chu CM, Cheng VC, Chan KS, Hung IF, Poon LL, Law KI, Tang BS, Hon TY, Chan CS, Chan KH, Ng JS, Zheng BJ, Ng WL, Lai RW, Guan Y, Yuen KY. Clinical progression and viral load in a community outbreak of coronavirus-associated SARS pneumonia: a prospective study. *Lancet*. 361(9371):1767–1772(2003).

Peiris JS, Guan Y, Yuen KY. Severe acute respiratory syndrome. *Nature Med*. 10, S88–S97 (2004).

Peiris JS, Yuen KY, Osterhaus AD. The severe acute respiratory syndrome. *N Engl J Med* 349:2431–2441(2003).

Perlman S. Pathogenesis of coronavirus-induced infections. Review of pathological and immunological aspects. *Adv Exp Med Biol* .440:503(1998).

Pickett BE, Greer DS, Zhang Y, Stewart L, Zhou L, Sun GG, Gu Z, Kumar S, Zaremba S, Larsen SN, Jen W, Klem EB, Scheuermann RH. Virus Pathogen Database and Analysis Resource (ViPR): A Comprehensive Bioinformatics Database and Analysis Resource for the Coronavirus Research Community. *Viruses*. 4(11): 3209–3226(2012).

Rahul V, Stanley P. Middle East respiratory syndrome and severe acute respiratory syndrome. *Current Opinion in Virology* 16:70–76(2016).

Raj VS, Mou H, Smits SL, Dekkers DH, Muller MA, Dijkman R, Muth D, Demmers JA, Zaki A, Fouchier RA. Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. *Nature* 495:251-254(2013).

Ralph A. Tripp. Respiratory Syncytial Virus (RSV) Modulation at the Virus-Host Interface Affects Immune Outcome and Disease Pathogenesis. *Immune Network*. 13(5): 163-167(2013).

Schmidt AC. Progress in respiratory virus vaccine development. *Semin. Respir. Crit. Care Med*. 32: 527-540(2011).

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13(11):2498-2504(2003).

Shapira SD, Gat-Viks I, Shum BO, Dricot A, de Grace M, Wu L, Gupta PB, Hao T, Silver SJ, Root DE, Hill DE, Regev A, Hacohen N. A physical and regulatory map of host-influenza interactions reveals pathways in H1N1 infection. *Cell* 139:1255-1267(2009).

Sharma D, Priyadarshini P, Vrati S. Unraveling the web of viroinformatics: computational tools and databases in virus research. *J Virol* 89: 1489-1501(2015).

Sievers F, Higgins DG. Clustal Omega, accurate alignment of very large numbers of sequences. *Method mol bio*. 1079:105-16(2014).

Sitou N, Nei M. The Neighbor-joining Method: A new method for reconstructing Phylogenetic trees. *Molecular and Biology and Evolution*. 4:406-425(1987).

Siu KL, Kok KH, Ng MHJ, Poon VKM, Yuen KY, Zheng BJ, Jin DY. Severe acute respiratory syndrome coronavirus M protein inhibits Type I interferon production by impeding the formation of TRAF3.TANK.TBK1/IKK ϵ complex. *Journal of Biological Chemistry*. 284 :16202–16209(2009).

Siu KL, Yeung ML, Kok KH, Yuen KS, Kew C, Lui PY, Chan CP, Tse H, Woo PC, Yuen KY. Middle east respiratory syndrome coronavirus 4a protein is a double-

stranded RNA-binding protein that suppresses PACT-induced activation of RIG-I and MDA5 in the innate antiviral response. *J Virol*, 88; 4866–4876(2014).

Siu YL, Teoh KT, Lo J. The M, E, and N structural proteins of the severe acute respiratory syndrome coronavirus are required for efficient assembly, trafficking, and release of virus-like particles. *J Virol* .82:113-18(2008).

Snijder EJ, Bredenbeek PJ, Dobbe JC. Unique conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *J Mol Biol* 331:991-1004(2003).

Sokal R, Michener C. A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin* 38: 1409–1438(1958).

Spiegel M, Schneider K, Weber F, Weidmann M, Hufert FT. Interaction of severe acute respiratory syndrome-associated coronavirus with dendritic cells. *J Gen Virol*. 87:1953–1960(2006).

Squires RB, Noronha J, Hunt V, Garcia-Sastre A, Macken C, Baumgarth N, Suarez D, Pickett BE, Zhang Y, Larsen CN, Ramsey A, Zhou L, Zaremba S, Kumar S, Deitrich J, Klem E, Scheuermann RH. Influenza research database: an integrated bioinformatics resource for influenza research and surveillance. *Influenza Other Respir Viruses* 6:404–416(2012).

Stark GR, Kerr IM, Williams BR, Silverman RH, Schreiber R.D. How cells respond to interferons. *Annu.Rev.Bioche*.67:227-264 (1998).

Stohlman SA, Baric RS, Nelson GN, Soe LH, Welter LM, Deans RJ. Specific interaction between coronavirus leader RNA and nucleocapsid protein. *Journal of virology*. 62(11):4288–4295(1988).

Sun L, Xing Y, Chen X, Zheng Y, Yang Y, Nichols DB, Clementz MA, Banach BS, Li K, Baker SC. Coronavirus papain-like proteases negatively regulate antiviral innate immune response through disruption of STING-mediated signaling. *PLoS ONE*, 7 p. e30802(2012).

Teigelkamp S, Achsel T, Mundt C, Goethel SF, Cronshagen U, Lane WS, Marahiel M, Luehrmann R. The 20kD protein of human [U4/U6.U5] tri-snRNPs is a novel cyclophilin that forms a complex with the U4/U6-specific 60kD and 90kD proteins. *RNA* 4:127-141(1998).

Totura AL. SARS coronavirus antagonizes innate immune signaling initiated by RIG-I but is recognized by TLR signaling via the adaptor molecule TRIF. *The university of north Carolina at chapel hill* 170; 366856(2014).

Van BS, de Graaf M, Lauber C, Bestebroer TM, Raj VS, Zaki AM, Osterhaus AD, Haagmans BL, Gorbalenya AE, Snijder EJ, Fouchier RA. Genomic characterization

of a newly discovered coronavirus associated with acute respiratory distress syndrome in humans. *mBio* 3:e00473–129(2012).

Van DHL, Pyrc K, Jebbink MF, Vermeulen-Oost W, Berkhout RJ, Wolthers KC, Wertheimvan DPM, Kaandorp J, Spaargaren J, Berkhout B. Identification of a new human coronavirus. *Nat Med.* 10(4):368–373(2004).

Van DHL, Sure K, Ihorst G, Stang A, Pyrc K, Jebbink MF, Petersen G, Forster J, Berkhout B, Uberla K. Croup is associated with the novel coronavirus NL63. *PLoS Med.* 2(8):e240(2005).

Van HMJ, Van WSH, Knoops K, Mommaas AM, Gorbalenya AE, Snijder EJ. SARS-coronavirus replication/transcription complexes are membrane-protected and need a host factor for activity in vitro. *PLoS Pathog.* 4: (2008).

Van JA, vander ME, Miltenburg LA, van KP, Verhoef J. Tumour necrosis factor triggers granulocytes to internalize complement-coated virus particles. *Immunology* 73:77-82(1991).

Vennema H, Godeke GJ, Rossen JW, Voorhout WF, Horzinek MC, Opstelten DJ, Rottier PJ. Nucleocapsid-independent assembly of coronavirus-like particles by co-expression of viral envelope protein genes. *EMBO J.* 15:2020-2028(1996).

Vijgen L, Keyaerts E, Lemey P, Moes E, Li S, Vandamme AM, Van Ranst M. Circulation of genetically distinct contemporary human coronavirus OC43 strains. *Virology.* 337(1):85–92(2005).

Von BA, Teepe C, Simpson JC, Pepperkok R, Friedel CC, Zimmer R, Robert R, Baric R, Haas J. Analysis of Intraviral Protein-Protein Interactions of the SARS Coronavirus ORFome. *PLoS ONE* 2(5): e459(2007).

Wang Y, Liu D, Shi W, Lu R, Wang W, Zhao Y, Deng Y, Zhou W, Ren H, Wu J, Wang Y, Wu G, Gao GF, Tan W. Origin and possible genetic recombination of the Middle East respiratory syndrome coronavirus from the first imported case in China: phylogenetics and coalescence analysis. *mBio* 6(5):e01280-15(2015).

Wathelet MG, Orr M, Frieman MB, Baric RS. Severe acute respiratory syndrome coronavirus evades antiviral signaling: role of nsp1 and rational design of an attenuated strain. *Journal of Virology,* 81:11620–11633(2007).

Wei Y, Pattingre S, Sinha S, Bassik M, Levine B. JNK1-mediated phosphorylation of Bcl-2 regulates starvation-induced autophagy. *Mol. Cell* 30:678-688(2008).

Woo PC, Lau SK, Chu CM, Chan KH, Tsoi HW, Huang Y, Wong BH, Poon RW, Cai JJ, Luk WK, Poon LL, Wong SS, Guan Y, Peiris JS, Yuen KY. Characterization

andcomplete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. *Journal of virology*. 79(2):884–895(2005).

World Health Organization. Middle East respiratory syndrome coronavirus (MERS-CoV): Epicurve of confirmed cases and deaths in Republic of Korea, China, Saudi Arabia and other countries. July 2015; <http://www.who.int/emergencies/merscov/en/> (Accessed 1 Oct, 2015).

Wurm T, Chen H, Hodgson T, Britton P, Brooks G, Hiscox JA. Localization to the nucleolus is a common feature of coronavirus nucleoproteins, and the protein may disrupt host cell division. *J. Virol*. 75:9345-9356(2001).

Xu X, Liu Y, Weiss S. Molecular model of SARS coronavirus polymerase: implications for biochemical functions and drug design. *Nucleic Acids Res* 31:7117–7130(2003).

Yamada Y, Liu XB, Fang SG, Tay FP, Liu DX. Acquisition of cell–cell fusion activity by amino acid substitutions in spike protein determines the infectivity of a coronavirus in cultured cells. *PLoS ONE* 4:e6130 (2009).

Yang IS, Lee JY, Lee JS, Mitchell WP, Oh HB, Kang C, Kim KH. Influenza sequence and epitope database. *Nucleic Acids Res* 37:D423– D430(2009).

Zhai Y, Sun F, Li X. Insights into SARS-CoV transcription and replication from the structure of the nsp7-nsp8 hexadecamer. *Nat Struct Mol Biol* 12:980–986(2005).

Zhang Y, Bo XC, Yang J, Wang SQ. HBVPathDB: a database of HBV infection-related molecular interaction network. *World J Gastroenterol* 11:1690–1692(2005).

Zhao JC, Li K, Wohlford-Lenane C, Agnihothram SS, Fett C, Zhao JX. Rapid generation of a mouse model for Middle East respiratory syndrome. *Proc Natl Acad Sci U S A*. 111:4970–5(2014).

Zhou J, Chu H, Li C, Wong BH, Cheng ZS, Poon VK. Active replication of Middle East respiratory syndrome coronavirus and aberrant induction of inflammatory cytokines and chemokines in human macrophages: implications for pathogenesis. *J Infect Dis*. 209:1331–42(2014).

Ziebuhr J. The corona virus replicase. *Curr. Top. Microbiol. Immunol*. 287:57–94(2005).

Zinkernagel RM. Immunology taught by viruses. *Science* 271;173-178(1996).

Zumla A, Hui DS, Perlman S. Middle East respiratory syndrome. *Lancet*, 386:995-1007(2015).

ABSTRACT

INVESTIGATION ON THE IMMUNOLOGICAL SPECIFICITY OF CORONA VIRUS UTILIZING BIOINFORMATICS METHOD

Mi-Ran kim

**Laboratory of Computational Biology and Bioinformatics
The Graduated School
Seoul National University**

The epidemic of a new infectious disease remains a challenge and a serious threat to public health. MERS emerged in South Korea in 2015, caused confusion to the public and the health authorities, and resulted in severe social and economic losses. As a result, we have discovered the problem of the infectious disease response system in Korea. Additionally, it was aware of the problems of environmental health services, it had to leave the task of the development of therapeutics and vaccines. The coronavirus was first discovered 60 years ago and caused a deadly respiratory infection, SARS-CoV, in people in 2003; the trend drew global attention. Then, the focus became the need to understand the pathogenic and host immune response. Worldwide research has been conducted on the coronavirus. Although the molecular biological and immune mechanisms that result in coronavirus have been found, the information has been scattered, and existing database is not easy to access and use. In this study, by building HCoV-IMDB to provide information on the coronavirus and immune-related proteins and genes, it is hoped that useful information will be provided to researchers. Through a research method utilizing phylogenetic analysis based on the data in the built

database, the risk of unknown coronavirus was predicted. The S protein of HCoV-OC43 was similar to MERS-CoV evolution. This suggests the need for continuous monitoring of the mutations of HCoV-OC43. Furthermore, it provided bioinformatics tools that can be used for monitoring the outbreak of the traditional virus and the new virus. By providing a method for disease prevention, not a method for treatment, it is expected to contribute to the development of Public Health. For the immune specificity analysis of the coronavirus, network analysis was performed. Interaction between the IC1 protein of humans and the non-structural proteins of SARS was confirmed. IC1 is involved in the activity of the complementary system of the immune mechanism. Additionally, it is a protein involved in blood coagulation, the degradation of fibrin, and kinin synthesis. Studying the protein provides insight for understanding the unknown immune mechanisms of coronavirus, which will be helpful in the development of antibiotics and vaccines. Furthermore, it will be helpful in the development of personalized treatments for the genetic predisposition of individuals. In addition, the network analysis method can show results at a glance by integrating the results of traditional immunological studies, and it will be used as the basis for system immunology.

.....

Keywords : Corona virus, Bioinformatics, Immune, Database,

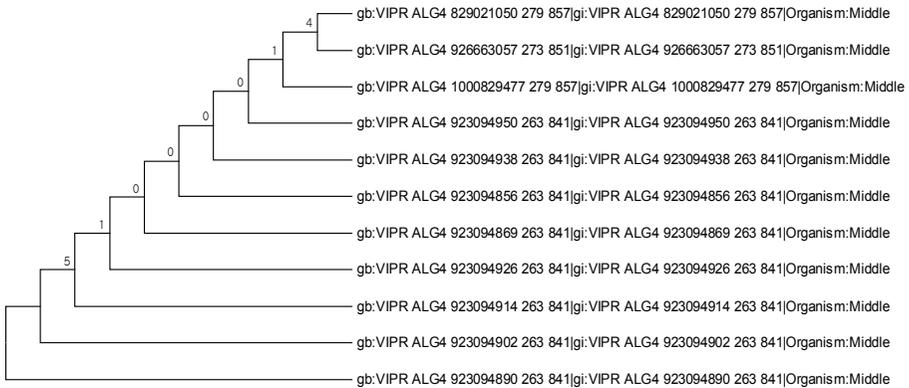
Public health, Network analysis.

Student Number : 2009-21997

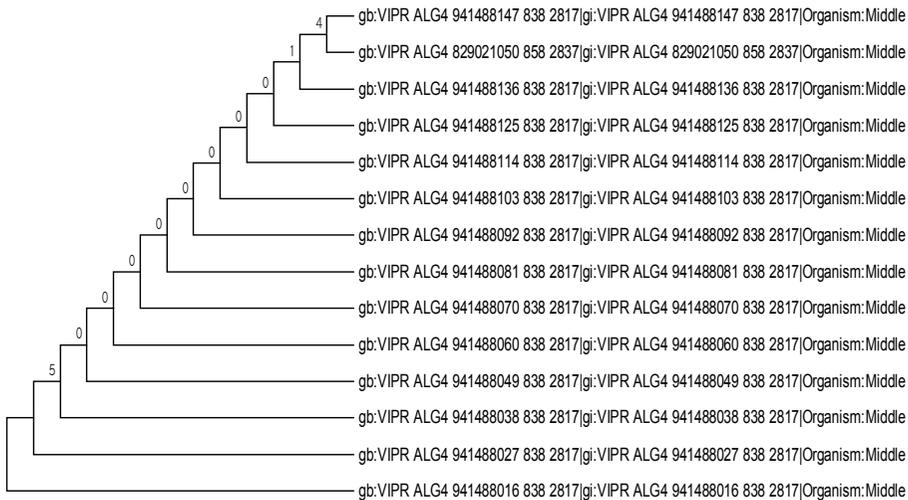
부 록



1. Evolutionary relationships of Spike protein. Spike protein of MERS isolated from Korea.



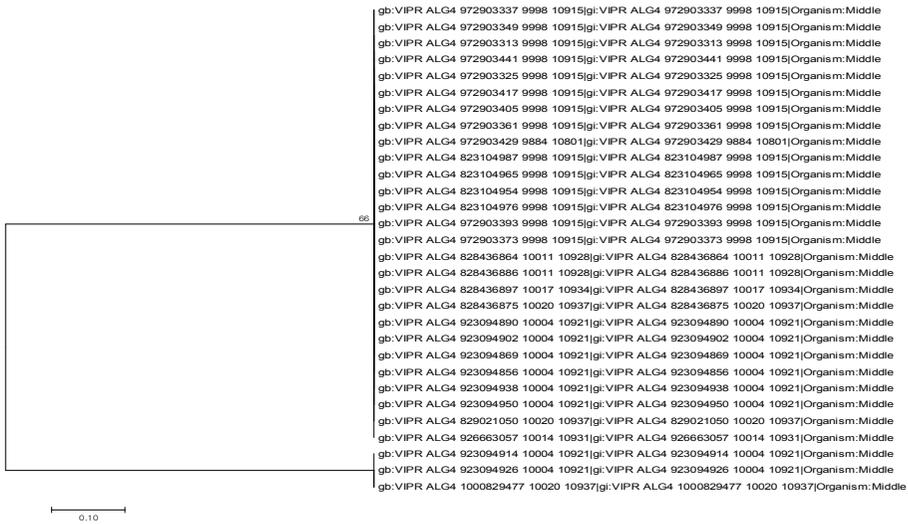
2. Evolutionary relationships of NSP1 protein. NSP1 protein of MERS isolated from Korea.



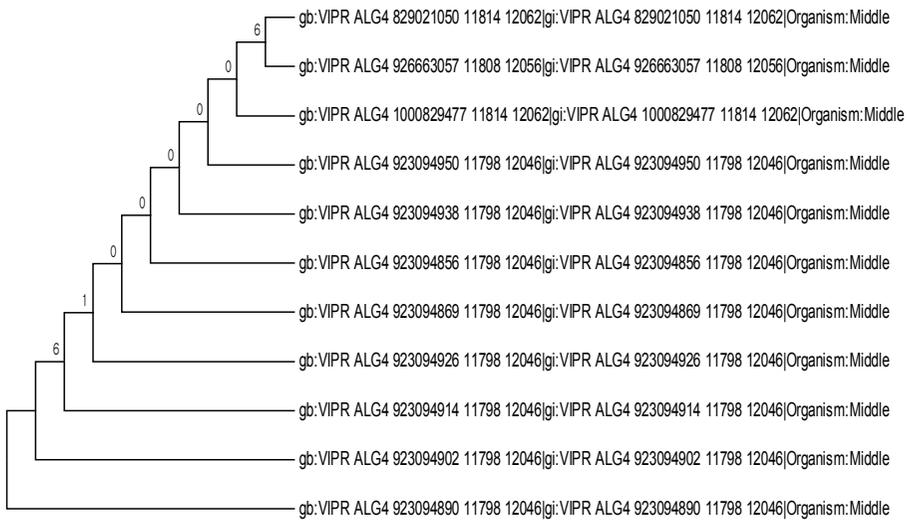
3. Evolutionary relationships of NSP2 protein. NSP2 protein of MERS isolated from Korea.



4. Evolutionary relationships of NSP3 protein. NSP3 protein of MERS isolated from Korea.



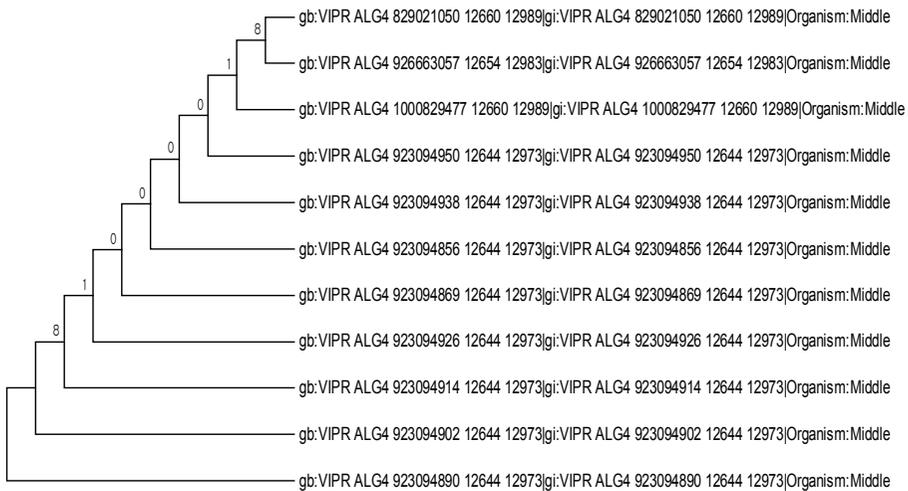
7. Evolutionary relationships of NSP6 protein. NSP6 protein of MERS isolated from Korea.



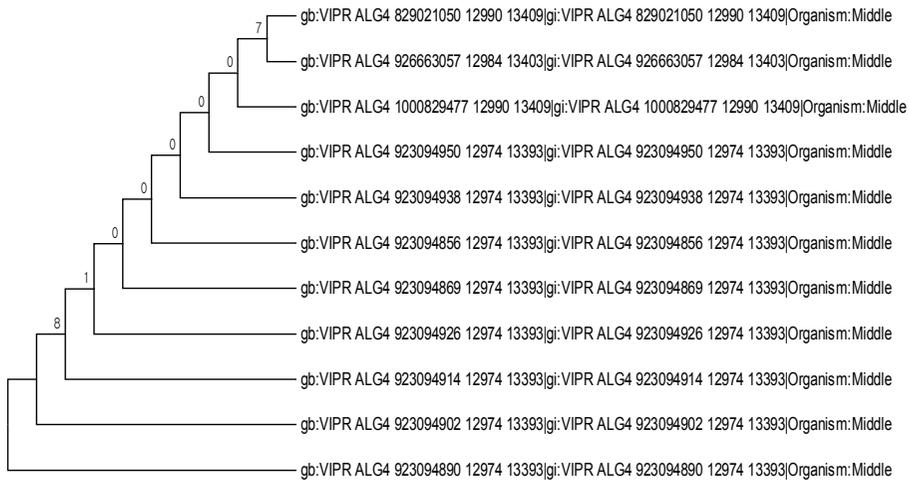
8. Evolutionary relationships of NSP7 protein. NSP7 protein of MERS isolated from Korea.



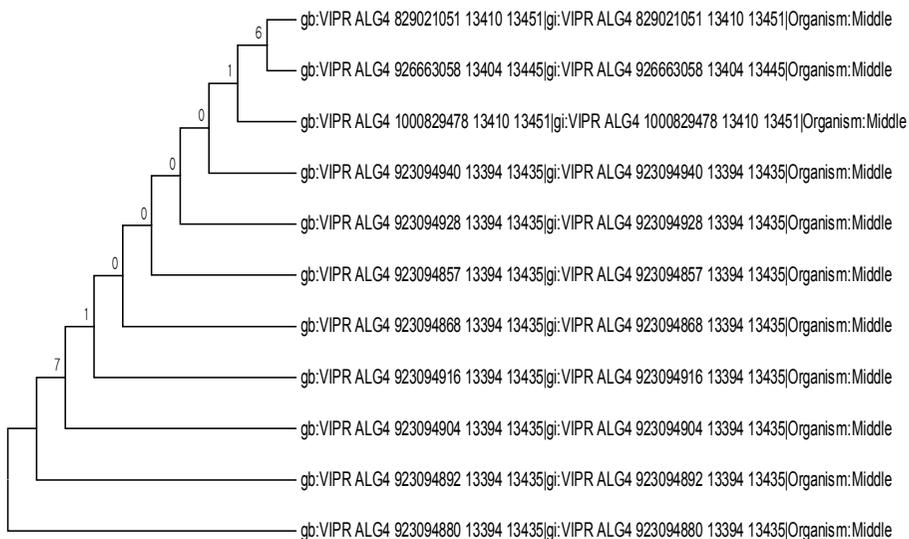
9. Evolutionary relationships of NSP8 protein. NSP8 protein of MERS isolated from Korea.



10. Evolutionary relationships of NSP9 protein. NSP9 protein of MERS isolated from Korea.



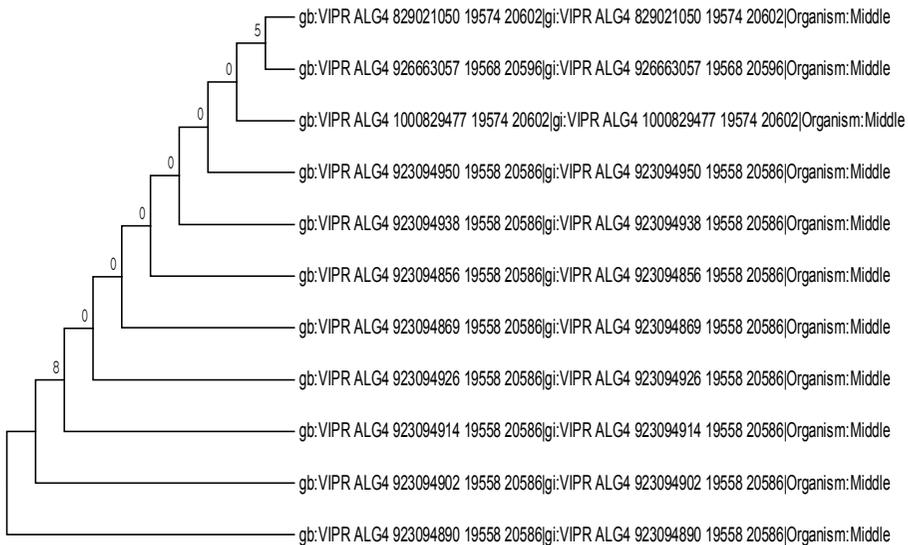
11. Evolutionary relationships of NSP10 protein. NSP10 protein of MERS isolated from Korea.



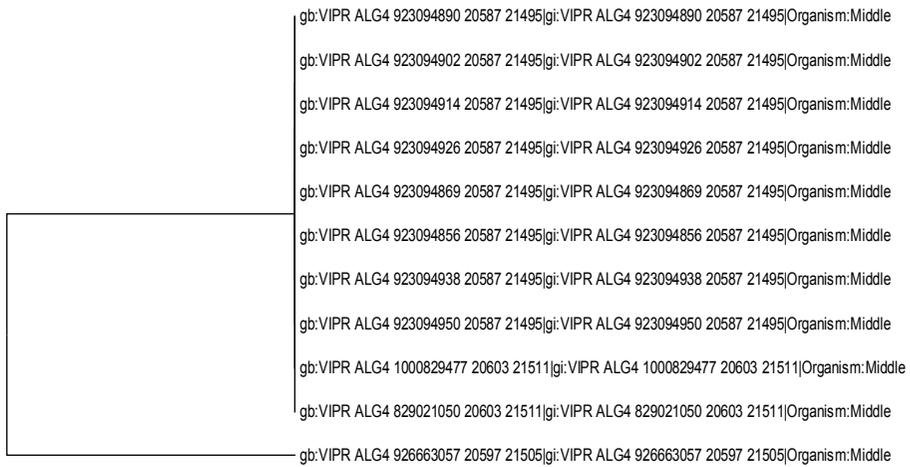
12. Evolutionary relationships of NSP11 protein. NSP11 protein of MERS isolated from Korea.



15. Evolutionary relationships of NSP14 protein. NSP14 protein of MERS isolated from Korea.



16. Evolutionary relationships of NSP15 protein. NSP15 protein of MERS isolated from Korea.



0.10

17. Evolutionary relationships of NSP16 protein. NSP16 protein of MERS isolated from Korea.