

바이러스 핵산중합효소의 아미노산 서열에 의한 바이러스 분류

남지현¹ · 이동훈¹ · 이건명² · 이찬희^{1*}

¹충북대학교 자연과학대학 생명과학부, ²전기전자컴퓨터공학부

볼티모어의 분류체계에 의하면 바이러스는 복제 및 단백질합성 전략에 따라 6개의 집단으로 나눌 수 있다. 몇 종류의 작은 DNA 바이러스를 제외한 대부분의 바이러스는 게놈 복제를 위한 자신의 핵산중합효소를 유전자로 암호화하고 있다. 바이러스 핵산중합효소에는 DNA-의존 DNA 중합효소, RNA-의존 RNA 중합효소, RNA-의존 DNA 중합효소 세 종류가 있으며, 이들은 모두 4개의 공통된 모티프(motif)를 가진다. 우리는 볼티모어의 분류체계와 바이러스의 핵산중합효소와의 관계를 아미노산 서열을 통해 분자 계통분류학적 분석을 통해 알아보려고 하였다. NCBI GenBank에서 얻은 바이러스 중합효소의 아미노산 서열을 CLUSTAL X 프로그램으로 다중서열하고, Neighbor-joining, Maximum-likelihood, Bayesian의 세 가지 방법으로 계통도를 그려 보았다. 미세한 차이는 있었으나, 세 가지 방법 모두에서 볼티모어의 분류법과 일치하는 결과를 보였고, 특이하게도 두 가닥 RNA 바이러스는 숙주의 종류에 따라, (-)RNA 바이러스는 게놈의 절편화에 따라 각각 2개의 소집단으로 나뉘어지는 것을 볼 수 있었다.

Key words □ Baltimore classification, phylogeny, viral polymerase

바이러스는 모양, 입자의 크기, 캡시드(capsid)의 구조, 숙주(host) 생물의 종류, 면역학적 특징, 게놈(genome)을 구성하는 핵산과 같은 성질에 의해 분류된다. 대표적인 바이러스 분류체계에는 계층적 바이러스 분류체계(Hierarchical virus classification system)와 볼티모어 분류체계(Baltimore classification system)가 있다. 계층적 바이러스 분류체계는 게놈의 구성, 캡시드의 대칭성, 외피의 존재유무, 비리온과 캡시드의 크기와 같은 형태학적 분류인 반면, 볼티모어의 분류체계는 바이러스 게놈의 구성, 발현, 복제 기작에 따른 분자생물학적 분류로 현재 바이러스를 분류할 때 보편적으로 이용되는 방법이다(<http://www.nlv.ch/Virologytutorials/Classification.htm>). 볼티모어의 분류체계에 의해 바이러스는 두 가닥 DNA(double-strand DNA; dsDNA, Class I), 단일 가닥 DNA(single-strand DNA; ssDNA, Class II), 두 가닥 RNA(double-strand RNA; dsRNA, Class III), 양성 단일 가닥 RNA[positive single-strand RNA; (+)ssRNA, Class IV], 음성 단일 가닥 RNA[negative single-strand RNA; (-)ssRNA, Class V], 레트로이드바이러스(retrovirus, Class VI)의 6종류로 분류할 수 있다(17).

지금까지 바이러스는 과(Family) 이상의 계통분류를 할 수 없었으나(20), 생명정보학의 이용이 증가함에 따라 보다 높은 수준의 바이러스 분류에 계통분류학적 분석이 적절하다고 생각되고 있다. 몇몇 바이러스 분류학자들은 표면상으로 밀접한 그룹의 RNA 바이러스는 공통 조상으로부터 유래되었고, RNA 복제의 주된 방법은 동일할 것이라는 진화적 연관설(20)을 내세웠다. 복

제과정과 계통도 사이의 연관성을 유추하기 어렵지만, 계통분류를 통해 유사한 복제과정을 갖는 바이러스의 관계를 더욱 발전시켜야 하며(14), 바이러스의 다양한 진화적 사건을 식별하여 문(Phylum)의 분류에 이용하고, 바이러스의 분리된 계(Kingdom)를 만들어야 한다(25). 특히, 단일 가닥 RNA 바이러스의 경우, 집단이 매우 클 뿐만 아니라 대부분의 정보가 이용 가능하기 때문에(9), ssRNA 바이러스의 계통도의 구성과 과(Family) 보다 높은 단계의 바이러스 분류가 필요하다.

모든 RNA 바이러스는 복제를 위해 RNA-의존 RNA 중합효소(RdRp)가 필요하고, RdRp의 아미노산 서열은 16S ribosomal RNA 염기서열처럼 RNA 바이러스 사이의 기본적인 분류에 이용할 수 있다(6, 10, 11, 16, 22). 또한, DNA 중합효소 복합체의 단백질 서열도 DNA 바이러스의 기본적인 계통분류학적 분석에 가능하다는 보고가 있다(2). 대부분의 바이러스들은 RNA나 DNA 중합효소를 하나의 단백질(protein)로 암호화하며, 각 중합효소의 공통된 기본 구조와 모티프를 가진다. 모든 핵산중합효소는 효소반응의 중심부위(catalytic center)로 palm domain에 네 부분의 공통 모티프를 갖고 있으며, 약 120~210 아미노산 서열로 이루어져 있다(19). 이러한 핵산중합효소의 공통 모티프를 통한 서열 분석은 바이러스의 진화와 계통분류에 대한 우리의 지식을 재정립하고 확장시킬 수 있다고 생각한다. 본 연구에서는 RdRp, RdDp, DdDp(DNA-의존 DNA 중합효소)를 가진 바이러스의 중합효소 아미노산 서열을 통하여 계통을 분석하고 볼티모어 분류와의 관계를 알아보려고 하였다.

*To whom correspondence should be addressed.
Tel: 82-43-261-2304, Fax: 82-43-273-2451
E-mail: chlee@chungbuk.ac.kr

재료 및 방법

바이러스 핵산중합효소 아미노산 서열의 획득

바이러스 분류에 관한 국제 위원회 데이터베이스(<http://www.ncbi.nlm.nih.gov/ICTVdb/index.htm>)로부터 볼티모어 분류체계에 의한 대표적인 Family (-viridae)에 속하는 바이러스를 기준으로 바이러스를 선정하였다. 볼티모어의 분류 중 전적으로 숙주의 핵산중합효소를 이용하는 ssDNA (ClassII)를 제외하고, 111 바이러

스의 RdRp, 12 바이러스의 RdDp, 22 바이러스의 DdDp 아미노산 서열을 NCBI의 GenBank (<http://www.ncbi.nlm.nih.gov/>)로부터 얻었다(Table 1). 선정된 바이러스는 각 바이러스의 참고서열과 웹 상으로 구조가 제공되는 바이러스를 기준으로 하였다.

아미노산 서열의 정렬

분석 대상 바이러스가 가지고 있는 세 종류 핵산중합효소의 아미노산 서열을 CLUSTAL X program v.1.81 (23)을 사용하여

Table 1. The list of viral polymerase and GenBank accession number used in the phylogenetic analysis

Baltimore Class	Family or Genus*	GenBank accession number
Class I (dsDNA)	Poxviridae	AAB19092, NP_039057, NP_659611, NP_570196, NP_064832
	Iridoviridae	NP_078724, NP_149500
	Phycodnaviridae	NP_048532, AAB66713, AAB49740, NP_077578
	Baculoviridae	NP_054095, NP_663268
	Herpesviridae	NP_044632, NP_040151, NP_066862, 2115435A, NP_039988, NP_042931, NP_047983
	Adenoviridae	NP_108658, NP_050281
Class III (dsRNA)	Reoviridae	AAL36027, AAA87364, P22678, NP_690891, AAM92745, AAK40249, NP_620544, NP_620541
	Birnaviridae	NP_047197, NP_690839, NP_690806
	Totiviridae	NP_620495, NP_659390, NP_041191
	Partitiviridae	NP_624349, AAM95601, AAB27624
	Hypoviridae	NP_051710
	Varicosavirus	BAC16226
Class IV (+RNA)	Narnaviridae	NP_660178, NP_660174
	Leviviridae	NP_040650
	Picornaviridae	NP_740478, NP_740525, NP_740411, NP_740514, NP_740737
	Dicistroviridae	NP_733845
	Sequiviridae	NP_734447, NP_734463
	Comoviridae	NP_734057, NP_733962, NP_734007
	Potyviridae	NP_734199, NP_734327, AAB02823, NP_734292, NP_697039, NP_734273
	Caliciviridae	NP_786896, NP_740332, AAP83352, BAB03262
	Hepatitis E-like virus	NP_056786
	Astroviridae	NP_751905
	Nodaviridae	NP_077730, NP_689433
	Sobemovirus	NP_736584
	Luteoviridae	NP_840014, NP_056748, NP_620026
	Umbravirus	NP_054006
	Tombusviridae	NP_835253, NP_051882, NP_619751, CAA60596, NP_056825, NP_619711, NP_619718, NP_068342
	Coronaviridae	NP_740629, NP_835352, NP_742138, AAD32995, NP_828869, P18458
	Arteriviridae	NP_705590
	Flaviviridae	NP_776009, NP_777493, NP_751928
	Togaviridae	NP_818936, NP_803428
	Tobamovirus	NP_597747
Tobravirus	AAC02066	

Table 1. Continued

Baltimore Class	Family or Genus*	GenBank accession number
Class IV (+RNA)	Hordeivirus	NP_604481
	Furovirus	BAA94797
	Pomovirus	NP_612628
	Pecluvirus	NP_620047
	Bromoviridae	NP_620768, P03594, NP_049324, NP_620043
	Idaeovirus	NP_620465
	Closteroviridae	NP_733949, NP_733943
	Vitivirus	NP_619662
	Tymovirus	NP_733819
	Barnaviridae	NP_042510
	Marafivirus	NP_734076
Class V (-RNA)	Bornaviridae	NP_042024
	Filoviridae	NP_690587
	Paramyxoviridae	NP_604442, NP_056924, NP_054714, NP_056866, YP_012613
	Rhabdoviridae	NP_041716, NP_056797, NP_065409, NP_042681, NP_597914, NP_042286
	Orthomyxoviridae	NP_056657, AAF89738, O41353
	Bunyaviridae	NP_047211, AAG28740, NP_690576, NP_049341, NP_049362
Arenaviridae	NP_694845	
Class VI (Retroid)	Caulimoviridae	NP_127504, NP_056728, NP_068729, NP_056848, NP_056762, NP_039820
	Retroviridae	NP_040550, NP_056790, NP_045937, NP_044450
	Hepadnaviridae	AAO12692, NP_039821

*Some viruses are classified into genus level only.

다중정렬 하였다. 정렬 후 계통수의 표식(phylogenetic marker)으로 이용되는 RNA 바이러스의 RdRp palm domain (26)을 대상으로 RdRp와 RdDp의 공통 모티프(19), NCBI의 CDD (Conserved Domain Database, <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=cdd>)와 ExPASy (<http://cn.expasy.org/>)의 PROSITE에 제공되는 MotifScan tool (7)을 이용하여 모티프를 재정렬하였다.

계통수 작성

Neighbor-joining (NJ) 방법으로 재정렬한 palm domain 모티프 (562 아미노산)의 거리행렬(distance matrix)을 작성하였다. 거리행렬은 PHYLIP package (8)의 PROTDIST 프로그램 중 JTT (Jones-Taylor-Thornton) 모델로 계산하였다. 이 거리행렬을 바탕으로 Neighbor-joining 프로그램을 사용하여 계통수를 작성하였다. Maximum-likelihood (ML) 방법 역시 PHYLIP package의 PROML 프로그램의 JTT 모델을 사용하여 계통수를 작성하였다. Bayesian (BA) 방법으로 작성한 계통수는 MrBayes 프로그램(12)의 Jones 모델을 사용하였으며, 4 Markov chain으로 각각 1,000,000회 반복하였다. 1,000회 반복마다 파일을 받았고, 300개의 계통수를 버리고, 700개의 계통수를 얻었다. 이 700개의 계통수를 가지고

가장 많은 수의 공통 계통수로부터 하나의 계통수를 작성하였다.

결 과

아미노산 서열의 정렬 및 특징

바이러스 핵산중합효소의 아미노산 서열은 모든 바이러스가 공통적으로 가지고 있는 palm domain의 모티프 위치가 매우 다양하기 때문에 CLUSTAL X로 다중정렬 결과 정렬이 제대로 이루어지지 않았다. 이미 보고된 RdRp와 RdDp의 공통 모티프와 NCBI의 CDD와 ExPASy의 PROSITE를 이용하여 palm domain의 네 부분(A, B, C, D)의 공통 모티프 부분을 재정렬하였다. Figure 1은 볼티모어 분류체계의 Class 중 대표적인 바이러스의 아미노산 서열을 공통 모티프로 재정렬한 것이다. 모든 핵산중합효소는 네 부분의 공통 모티프를 가지고 있었으며, DdDp (Class I), RdRp (Class III, Class IV, Class V) 그리고 RdDp (Class VI) 별로 각각의 다른 공통 서열을 가지고 있음을 알 수 있었다.

볼티모어 분류와 계통분류학적 분석의 연관성

재정렬한 모티프 부분만을 가지고 거리행렬을 작성하고, NJ 방법으로 계통도를 구성한 결과 몇몇을 제외하고 대부분 볼티모

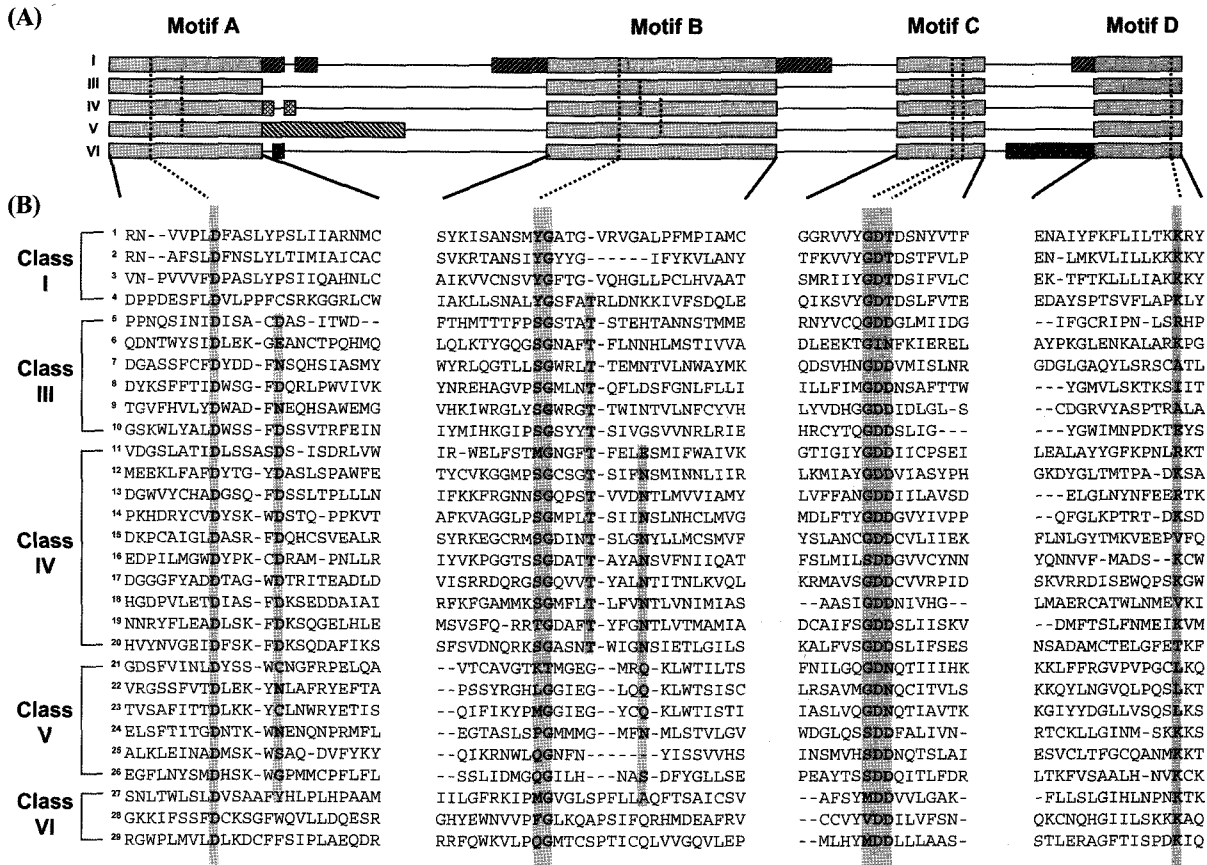


Fig. 1. Multiple sequence alignment of the four conserved motifs of viral polymerases. (A) Schematic representation. (B) Amino acid sequences of the motif A-D. Identical and similar residues are highlighted with bold letters in gray shade. The GenBank accession numbers for the sequences used in this figure are: ¹NP_149500, ²NP_054095, ³NP_044632, ⁴NP_108658, ⁵AAL36027, ⁶NP_047197, ⁷NP_620495, ⁸NP_624349, ⁹AAM95601, ¹⁰AAB27624, ¹¹NP_040650, ¹²NP_740478, ¹³NP_734199, ¹⁴NP_786896, ¹⁵NP_835253, ¹⁶NP_740629, ¹⁷NP_776009, ¹⁸NP_818936, ¹⁹P03594, ²⁰NP_733949, ²¹NP_042024, ²²NP_690587, ²³NP_056924, ²⁴NP_041716, ²⁵NP_047211, ²⁶NP_694845, ²⁷AAO12692, ²⁸NP_056728, ²⁹NP_040550

어 분류체계의 Class별로 클러스터를 형성함을 알 수 있었다. 하지만 bootstrap값이 너무 낮기 때문에, 신뢰성의 문제를 해결하기 위하여 ML과 BA 방법으로도 계통도를 구성하였고, 여기에서도 역시 몇몇을 제외하고 대부분 볼티모어 분류체계의 Class별로 클러스터를 형성하였다(Fig. 2).

Koonin과 Dolja (15)의 보고와 같이, Class III (dsRNA 바이러스)는 계통분류학적 분석 시 Class IV (+ssRNA 바이러스) 내에 그룹이 형성됨을 모든 계통도에서 확인해 볼 수 있었다(Fig. 2). 나아가 Class III 바이러스들은 원핵세포인 세균에서부터 고등 진핵세포인 동물에 이르기까지 숙주가 매우 다양하고 숙주가 고등 생물(동물, 식물)이나 하등미생물(세균, 곰팡이, 원생동물)이나에 따라 두 개의 소집단으로 분리되고 있다. 더욱이 *Birnaviridae*는 NJ 계통수에서는 가지가 멀리 떨어져 있고, ML과 BA 계통수에서는 Class III 집단에 속한다기보다는 Class IV에서 그룹화 되었다. *Birnaviridae*의 경우, RdRp의 가장 중요한 모티프인 GDD 모티프(C)가 A 앞부분에 존재하여, 모티프가 A-B-C 순서가 아닌 C-A-B의 순서이므로 C를 정렬할 수 없을 뿐 아니라, palm domain의 비교 모티프 부분이 결여되어 있었다(21).

모든 분석 방법에서 Class V (-ssRNA 바이러스)의 경우 계통의 구성에 따라 절편화 계통(segmented genome)과 비절편화 계통(non-segmented genome)의 두 소집단으로 분리됨을 알 수 있었다. Class V 바이러스만을 대상으로 분석해본 결과 세 가지 방법에서 모두 동일하게 두 개의 소집단을 형성하였기에 BA 계통수만을 결과로 제시하였다(Fig. 3). 절편화 계통을 가지는 Va 바이러스의 계통수 형태를 보면, 하나는 *Orthomyxoviridae*로, 또 다른 하나는 *Bunyaviridae/Arenaviridae*로 두 개의 주요 계열로의 독립적 진화를 볼 수 있다. Vb 바이러스의 경우 계통도의 형태가 *Paramyxoviridae*가 paraphyletic taxon인 것을 제외하고는 과와 아과로 taxonomic grouping됨을 알 수 있었다. *Paramyxoviridae*의 경우 아과인 *Pneumovirinae*가 다른 아과인 *Filoviridae*와 더욱 가까운 관계를 보였고, *Bornaviridae*, *Rhabdoviridae*와 *Paramyxoviridae* 사이의 관계는 알 수 없었다.

고 찰

일부 바이러스 분류학자들은 바이러스의 조상을 다윈 발생

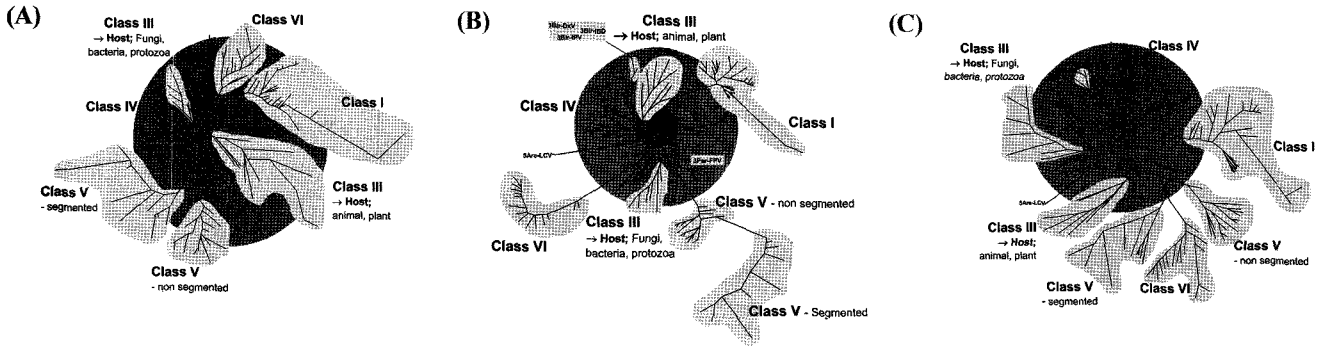


Fig. 2. Phylogenetic trees derived from amino acid sequence of viral polymerases with neighbor-joining (A), maximum-likelihood (B) and Bayesian (C) methods. Amino acid sequences of viral polymerases were multiple-aligned by CLUSTAL X and phylogenetic trees were constructed using JTT (Jones-Taylor-Thornton) model of sequence evolution. The viruses used in this study and their NCBI GenBank accession numbers are listed in Table 1.

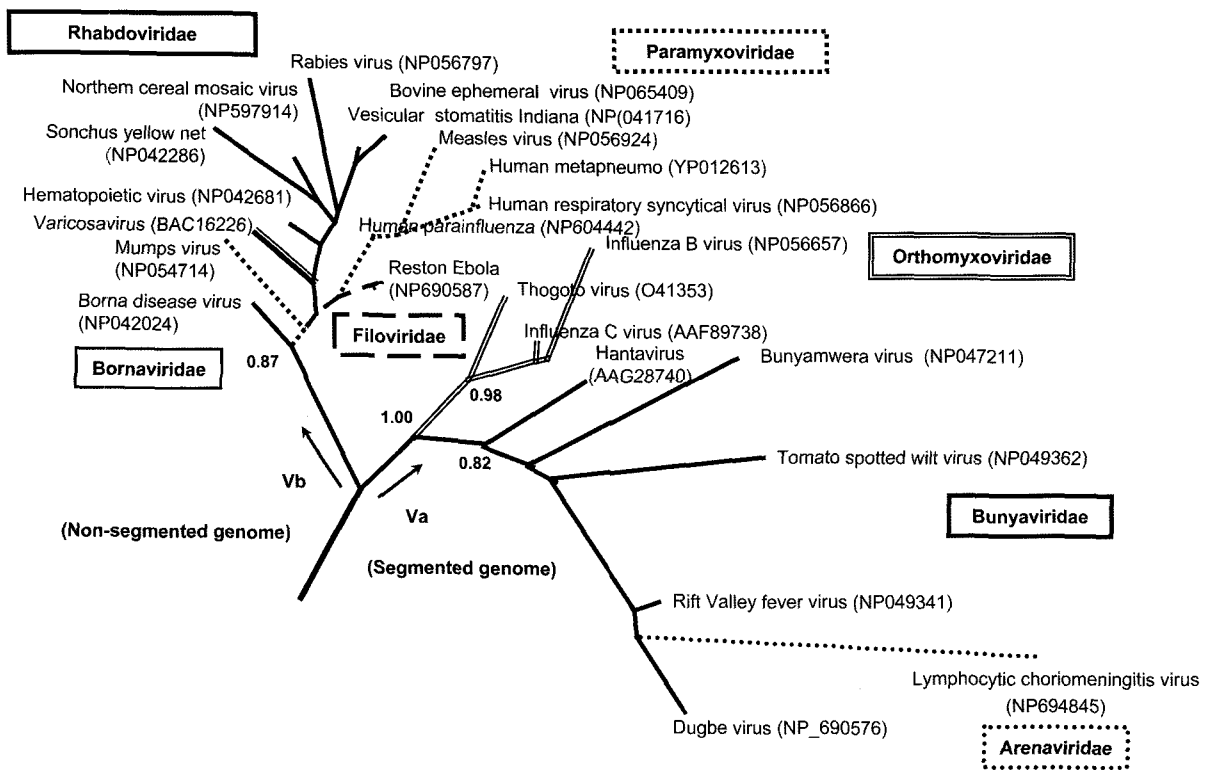


Fig. 3. Bayesian tree of the RdRp of Class V. The Class V was divided into two sub-clusters. The viruses of Class Va (*Orthomyxoviridae*, *Bunyaviridae*, *Arenaviridae*) have segmented genomes and those of Class Vb (*Bornaviridae*, *Rhabdoviridae*, *Paramyxoviridae*, *Filoviridae*) have non-segmented genomes. The names and GenBank accession numbers of the representing viruses for each family are included.

(polyphyletic origin)으로 보고, 계통분류를 통한 모든 바이러스의 진화적 관계를 하나의 계통도로 나타내는 것은 어렵다고 주장하였다(20). 특히, 생명정보학을 적용한 RdRp 아미노산 서열의 비교분석으로는 바이러스의 과(Family)간의 관계를 명백히 밝힐 수 없었고(15), RNA 바이러스 사이에서 유사성과 계통학적 관계가 명확하다 할지라도 RdRp와 RdDp가 공통 조상으로부터 유래되었다고 하기에는 불충분하다고 한다(26). 하지만, 모든 RNA 바이

러스는 복제를 위해서 RdRp를 필요로 하고, RdRp의 아미노산 서열은 16S ribosomal RNA 염기서열처럼 RNA 바이러스의 기본적인 분류에 이용할 수 있으며(6, 10, 11, 16, 22), DNA 중합효소 복합체의 단백질 서열도 DNA 바이러스의 기본적인 계통분류학적 분석에 가능하다는 보고도 있다(2). 핵산중합효소의 공통 모티프를 통한 서열 분석은 바이러스의 진화와 계통분류에 대한 지식과 볼티모어 분류체계의 이론을 재정립하고 확장하기에 타

당하다.

RdRp는 right hand palm, fingers, thumb domain의 공통부위를 갖는다. 효소반응의 중심부위(catalytic center)인 palm domain은 서열이 매우 보존되어있고, palm과 fingers domain 복합체로 고정되어 있다(5). Thumb domain은 구조적 변화로 효소의 활성(catabolic activity)에 관여한다(1, 18). RdRp의 반응에서 중요한 palm domain은 A, B, C, D 네 부분의 공통 모티프를 가지며, 약 120~210 아미노산 서열로 이루어져 있다. 모티프 A는 아스파르트산(D)으로 β strand 구조의 끝 부분에 존재한다. 모티프 B는 글리신(G)으로 회전구조를 형성하고, 핵산중합효소의 구조를 지탱하는데 관여한다. 모티프 C는 GDD로 이루어졌으며, DD에 의해 β turn β 구조를 형성한다. 모티프 D는 α helix 구조로 카복시기말단 부위이다. 모티프 A와 C의 아미노산 서열은 매우 보존된 부위이며, 중합효소의 활성을 위해서 반드시 필요하고, 양이온과의 결합, 주형의 특이성과 효소반응의 과정에 관여한다. DdDp의 경우는 모티프 C 부위가 GDD가 아닌 YGDTD로 이루어지는 차이점을 보인다(27). 또한 바이러스 중합효소 중 가장 많은 바이러스가 가지고 있는 RNA 바이러스의 RdRp는 DNA 바이러스의 DdDp와 레트로바이러스의 RdRp 사이에서도 가장 많은 유사성을 보이는 것으로 알려져 있다(4).

세 핵산중합효소 palm domain의 공통 모티프 부분으로 세 가지 방법의 계통수를 구성한 결과, 몇몇을 제외하고 대부분 볼티모어 분류체계의 Class별로 클러스터를 형성함을 알 수 있었다. Class III (dsRNA 바이러스)의 경우, Class IV (+ssRNA 바이러스) 내에 그룹이 형성되었고 이를 통해, dsRNA 바이러스가 +ssRNA 바이러스로부터 진화되었다고 추측할 수 있다. 또한, Class III는 모든 계통도에서 숙주(고등생물과 하등생물)에 따라 두 소집단으로 분리되므로 숙주에 따른 기원이 다름을 알 수 있다. 이것은 바이러스의 생활 형식이 stable, persistent, noninfectious state 중 어느 것을 택하느냐에 따라 RdRp의 보존서열이 다르게 발현되기 때문이지만, 두 소집단을 형성함에 있어 숙주에 따라 매우 일찍이 분화하였음에도 불구하고 각각 숙주의 선택적 압력에 대한 dsRNA 바이러스의 palm domain의 염기서열은 잃어버리지 않았다고 추측해 볼 수 있다(3).

모든 분석 방법에서 Class V (-ssRNA 바이러스)의 경우 계통의 구성에 따라 절편화 계층과 비절편화 계층의 두 소집단으로 분리됨을 알 수 있었다. 절편화 계층을 가지는 Class Va에서 하나는 *Orthomyxoviridae*로, 또 다른 하나는 *Bunyaviridae/Arenaviridae*로 두 개의 주요 계열로의 독립적 진화를 볼 수 있다. *Orthomyxoviridae*는 하나의 캡시드 내에 10개의 RNA 절편을 가지고 있으며, *Bunyaviridae/Arenaviridae*는 바이러스 입자 내에 2~3개의 캡시드를 가지고 있고, 각 캡시드에는 서로 다른 RNA 절편을 가지고 있다. *Arenaviridae*가 다른 절편화 계층을 가진 바이러스보다 유난히 거리가 멀지만 Class IV와 그룹화 된 이유는 아주 오랜 옛날부터 이차 숙주에 의한 진화 과정으로 즉, "diffuse coevolution" 때문이라 생각된다(13). *Arenaviridae*는 *Bunyaviridae*의 가장 먼 *Narivirus dugbe*와 가까운 관계를 보였으며, 이를 통해 *Arenaviridae*는 *Bunyaviridae*의 중간 조상으로부터 유래되었다고

추측해 볼 수 있다. 또한, *Bunyaviridae*는 *Narivirus*와 *Arenaviridae* 분류군 사이의 밀접한 관계에 의해 paraphyletic taxon임을 알 수 있다. 이는 Vieth 등이 *Lassa virus*의 RNA 염기서열을 가지고 분석하면서, (-)ssRNA 바이러스(Class V)의 핵산중합효소 모티프 부분의 313 아미노산 염기서열로 Dayhoff와 JTT, WAG 치환모델을 사용하여 다양한 방법으로 계통수를 구성한 것과 동일한 결과이다(24).

이상에서 살펴본 바와 같이, 각 핵산중합효소간의 차이점은 A, B, C, D 모티프 외에 공통서열로부터 나타났다고 생각한다. RdRp의 motif B부분의 경우 각각 Class III와 Class IV, Class V의 공통 서열을 가지고 있고, Class IV (+RNA 바이러스)는 이 두 Class (III, V)의 motif B에 속하는 공통 서열을 모두 가지며 mRNA로 곧장 사용이 가능한 특징으로 미루어 볼 때, Class IV는 RNA world에서부터 존재하고 RdRp가 첫 번째의 단백질이었으며, Class III (dsRNA 바이러스)와 Class V (-RNA 바이러스)는 각각 Class IV로부터 진화되었음을 추측할 수 있다. 특히, 아미노산 서열에 기초한 바이러스 중합효소의 계통분류학적 분석으로, 계층의 복제에 따른 바이러스 분류체계인 볼티모어 이론을 지지할 수 있다.

감사의 말

이 논문은 2007년 교육인적자원부의 재원으로 한국학술진흥재단의 지원을 받아 수행된 연구임(지방연구중심대학육성사업/충북 BIT연구중심대학육성사업단).

참고문헌

1. Biswal, B.K., M.M. Cherney, M. Wang, L. Chan, C.G. Yannopoulos, D. Bilimoria, O. Nicolas, J. Bedard, and M.N. James. 2005. Crystal structures of the RNA-dependent RNA polymerase genotype 2a of hepatitis C virus reveal two conformations and suggest mechanisms of inhibition by non-nucleoside inhibitors. *J. Biol. Chem.* 280, 18202-18210.
2. Braithwaite, D.K. and J. Ito. 1993. Compilation, alignment, and phylogenetic relationships of DNA polymerases. *Nucleic Acids Res.* 21, 787-802.
3. Bruenn, J.A. 1993. A closely related group of RNA-dependent RNA polymerases from double-stranded RNA viruses. *Nucleic Acids Res.* 21, 5667-5669.
4. Bruenn, J.A. 2003. A structural and primary sequence comparison of the viral RNA-dependent RNA polymerases. *Nucleic Acids Res.* 31, 1821-1829.
5. Dmitrieva, T.M., A.V. Alexeevskia, G.S. Shatskaya, E.A. Tol-skaya, A.P. Gmyl, E.V. Khitrina, and V.I. Agol. 2007. Significance of the C-terminal amino acid residue in mengovirus RNA-dependent RNA polymerase. *Virology* 365, 79-91.
6. Dolja, V.V. and J.C. Carrington. 1992. Evolution of positive-strand RNA viruses. *Sem. Virol.* 3, 315-326.
7. Falquet, L., M. Pagni, P. Bucher, N. Hulo, C.J. Sigrist, K. Hofmann, and A. Bairoch. 2002. The PROSITE database, its status in 2002. *Nucleic Acids Res.* 30, 235-238.
8. Felsenstein, J. 2002. PHYLIP (Phylogeny Inference Package) Ver-

- sion 3.6a3. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle, USA.
9. Francki, R.I.B., C.M. Fauquet, D.L. Knudson, and F. Brown. 1991. Virus taxonomy. Classification and nomenclature of Viruses. Fifth report of the international committee on taxonomy of viruses.
 10. Goldbach, R. and J. Wellink. 1988. Evolution of plus-strand RNA viruses. *Intervirology* 29, 260-267.
 11. Habili, N. and R.H. Symons. 1989. Evolutionary relationship between luteoviruses and other RNA plant viruses based on sequence motifs and their putative RNA polymerases and nucleic acid helicases. *Nucleic Acids Res.* 17, 9543-9555.
 12. Huelsenbeck, J.P. and F. Ronquist. Command reference for MrBayes ver. 3.0B4, Serial Version. Department of Biology, University of Rochester, USA.
 13. Hugot, J.P., J.P. Gonzalez, and C. Denys. 2001. Evolution of the old world arenaviridae and their rodent hosts: generalized host-transfer or association by descent? *Infect. Genet. Evol.* 2, 1-8.
 14. Iyer, L.M., E.V. Koonin, and L. Aravind. 2003. Evolutionary connection between the catalytic subunits of DNA-dependent RNA polymerases and eukaryotic RNA-dependent RNA polymerases and the origin of RNA polymerases. *BMC Structural Biol.* 3, 1.
 15. Koonin, E.V. and V.V. Dolja. 1993. Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequences. *Crit. Rev. Biochem. Mol. Biol.* 28, 375-430.
 16. Matthews, R.E.F. 1985. Viral taxonomy for the non-virologist. *Ann. Rev. Microbiol.* 39, 451-474.
 17. Mayo, M.A. and C.R. Fringle. 1998. Virus taxonomy-1997. *J. Gen. Virol.* 79, 649-657.
 18. Ng, K.K., N. Pendas-Franco, J. Rojo, J.A. Boga, A. Machin, J.M. Alonso, and F. Parra. 2004. Crystal structure of Norwalk virus polymerase reveals the carboxyl terminus in the active site cleft. *J. Biol. Chem.* 279, 16638-16645.
 19. Poch, O., I. Sauvaget, M. Delarue, and N. Torodo. 1989. Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *EMBO J.* 8, 3867-3874.
 20. Rybicki, E.P. 1990. The classification of organisms at the edge of life, or problems with virus systematics. *South African J. Sci.* 86, 182-186.
 21. Shwed, P.S., P. Dobos, L.A. Cameron, V.N. Vakharia, and R. Duncan. 2002. Birnavirus VP1 proteins from a distinct subgroup of RNA-dependent RNA polymerases lacking a GDD motif. *Virology* 296, 241-250.
 22. Strauss, J.H. and E.G. Strauss. 1988. Evolution of RNA viruses. *Ann. Rev. Microbiol.* 42, 657-683.
 23. Thompson, J.D., T.J. Gibson, F. Plewniak, F. Jeanmougin, and D.G. Higgins. 1997. The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25, 4876-4882.
 24. Vieth, S., A.E. Torda, A. Marcel, S. Herbert, and G. Stephan. 2004. Sequence analysis of L RNA of Lassa virus. *Virology* 318, 153-168.
 25. Ward, C.W. 1993. Progress towards a higher taxonomy of viruses. *Res. Virol.* 144, 419-453.
 26. Zanutto, P.M., M.J. Gibbs, E.A. Gould, and E.C. Holmes. 1996. A reevaluation of the higher taxonomy of viruses based on RNA polymerases. *J. Virol.* 70, 6083-6096.

(Received November 30, 2007/Accepted December 21, 2007)

ABSTRACT : Classification of Viruses Based on the Amino Acid Sequences of Viral Polymerases
Ji-Hyun Nam¹, Dong-Hun Lee¹, Keon Myung Lee², and Chan-Hee Lee^{1*} (¹School of Life Sciences and ²School of Electrical and Computer Engineering, Chungbuk National University, Cheongju 361-763, Korea)

According to the Baltimore Scheme, viruses are classified into 6 main classes based on their replication and coding strategies. Except for some small DNA viruses, most viruses code for their own polymerases: DNA-dependent DNA, RNA-dependent RNA and RNA-dependent DNA polymerases, all of which contain 4 common motifs. We undertook a phylogenetic study to establish the relationship between the Baltimore Scheme and viral polymerases. Amino acid sequence data sets of viral polymerases were taken from NCBI GenBank, and a multiple alignment was performed with CLUSTAL X program. Phylogenetic trees of viral polymerases constructed from the distance matrices were generally consistent with Baltimore Scheme with some minor exceptions. Interestingly, negative RNA viruses (Class V) could be further divided into 2 subgroups with segmented and non-segmented genomes. Thus, Baltimore Scheme for viral taxonomy could be supported by phylogenetic analysis based on the amino acid sequences of viral polymerases.