

传染性非典型肺炎可能病原——新型冠状病毒的系统发生学分析*

刘涛 李伍举 范明

军事医学科学院基础医学研究所，北京：100850

摘要 2003年3月以来，一种新型冠状病毒（SARS-CoV）被初步确定为2002年底爆发的致死性传染病——严重急性呼吸综合症（Severe Acute Respiratory Syndrome，即SARS）的病原。该病毒具有其他已知冠状病毒典型的基因组结构。对该病毒进行系统发生学分析对进一步的实验研究具有指导意义。我们首先通过构建SARS-CoV在全基因组水平上的系统发生树来明确其演化位置，然后分别从核酸和蛋白两个水平分析了SARS-CoV的5个主要同源蛋白——复制酶、S蛋白、E蛋白、M蛋白和N蛋白的系统发生树。结果表明，SARS-CoV与目前已知的冠状病毒同源，但具有与其它冠状病毒明显不同的特点——各同源基因的演化历史彼此不同，其中结构蛋白基因的演化历史与基因组的演化历史不同；SARS-CoV与IBV和TGV尤其是IBV的亲缘关系较近，尤其是在E蛋白和M蛋白两水平上的特殊近缘关系在进一步的实验研究中值得注意和参考。

关键词 冠状病毒，严重急性呼吸综合症，系统发生学，基因组，系统发生树

一种新的致死性传染病——严重急性呼吸综合症SARS（Severe Acute Respiratory Syndrome）自2002年末首发于中国广东省始，目前已波及中国多省和全世界几十个国

* 国家自然科学基金（30080017,30270315）资助项目。

家。2003年3月研究人员初步确定该病的病原系一种全新的冠状病毒（SARS-CoV）。目前关于该病毒的起源、动物宿主和毒性来源并不甚清楚¹⁻³。

冠状病毒是一类被膜的正义链RNA病毒，常导致人类以及其他脊椎动物的呼吸道和肠道疾病；冠状病毒分为三大类，第1和第2类特异感染哺乳动物，而第3组则特异感染鸟类。根据宿主范围、抗原性和基因组结构，冠状病毒被进一步划分为不同的种。冠状病毒具有较窄的宿主范围和严格的细胞嗜性。以往已知的人冠状病毒共两种，HCoV-229E 属于第1类，而HCoV-OC43属于第2类，它们与大约三成的轻微上呼吸道疾病有关¹。

SARS-CoV具有其他已知冠状病毒典型的基因组结构：5'短的非翻译区，复制酶蛋白区[rep]，钉蛋白[S]，包膜蛋白[E]，膜蛋白[M]，核壳蛋白[N]，短的非翻译区3'。rep基因约占整个基因组的2/3，可编码随翻译而被裂解为特定多肽片段的蛋白。在rep基因下游的4个基因编码病毒的结构蛋白S、E、M和N¹。

材料与方法

SARS 冠状病毒与其它冠状病毒的基因组序列：从 GenBank 数据库中获取共 16 个基因组序列，它们分别是：禽传染性支气管炎病毒 2 个（IBV 和 IBV BCK），牛冠状病毒 4 个（BCoV、BCoV Q、BCoV ENT、BCoV LUN），人冠状病毒 1 个（HCoV 229E），小鼠肝炎病毒 4 个（MHV、MHV2、MHV Penn 97 1 和 MHV ML10），猪流行性腹泻病毒 1 个（PEDV），猪传染性胃肠炎病毒 1 个（TGV），SARS 冠状病毒 3 个（SARS Urbani、SARS CUHK W1 和 SARS canada），具体如表 1 所示。

软件与系统发生树构建方法：主要运用 ClustalX(V1.8)程序⁴进行多序列比较和系统发

生树的构建分析(Bootstrap Neighbor Joining method) ,系统发生树绘图由 Tree-Explore 程序⁵完成。分别从核酸和蛋白两个水平构建了复制酶、S 蛋白、E 蛋白、M 蛋白和 N 蛋白的系统发生树；另外，为了更好地利用基因组信息来考察 SARS 病毒的系统发生位置，构建了冠状病毒在全基因组水平上的系统发生树。

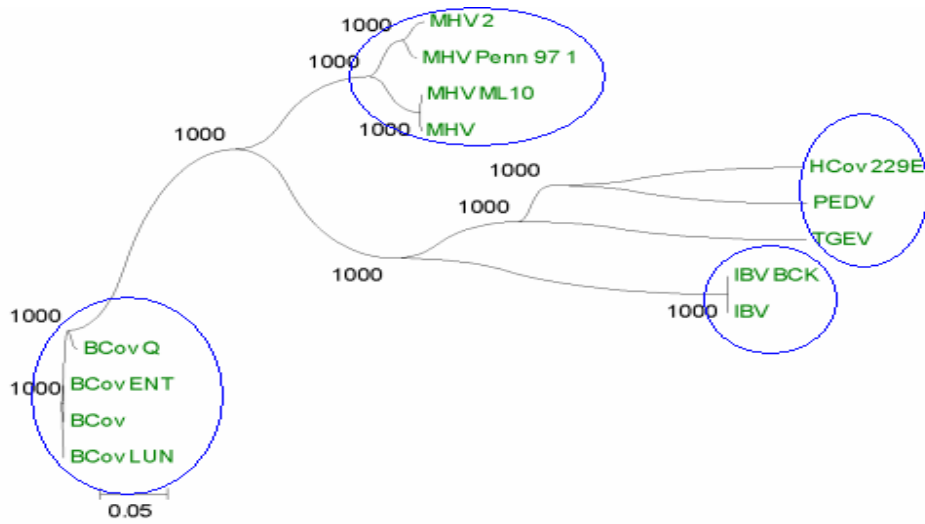
表 1：SARS 冠状病毒与其它冠状病毒的基因组序列

序号	GenBank ACCESSION	本文中的编码	长度 (bp)	描述
01	AJ311317	IBV BCK	27635	Avian infectious bronchitis virus (strain Beaudette CK) complete genomic RNA.
02	NC_001451	IBV	27608	Avian infectious bronchitis virus, complete genome.
03	AF220295	BCoV Q	31100	Bovine coronavirus strain Quebec, complete genome.
04	AF391541	BCoV ENT	31028	Bovine coronavirus isolate BCoV-ENT, complete genome.
05	AF391542	BCoV LUN	31028	Bovine coronavirus isolate BCoV-LUN, complete genome.
06	NC_003045	BCoV	31028	Bovine coronavirus, complete genome.
07	NC_002645	HCoV 229E	27317	Human coronavirus 229E, complete genome.
08	AF201929	MHV 2	31276	Murine hepatitis virus strain 2, complete genome.
09	AF208066	MHV Penn 97 1	31112	Murine hepatitis virus strain Penn 97-1, complete genome.
10	AF208067	MHV ML 10	31233	Murine hepatitis virus strain ML-10, complete genome.
11	NC_001846	MHV	31357	Murine hepatitis virus, complete genome.
12	NC_003436	PEDV	28033	Porcine epidemic diarrhea virus, complete genome.
13	NC_002306	TGV	28586	Transmissible gastroenteritis virus, complete genome.
14	AY278554	SARS CUHK W1	29736	SARS coronavirus CUHK-W1, complete genome.
15	AY278741	SARS Urbani	29727	SARS coronavirus Urbani, complete genome.
16	NC_004718	SARS canada	29736	SARS coronavirus, complete genome.

结果与讨论

1. SARS-CoV 全基因组序列的系统发生学分析。

(a)



(b)

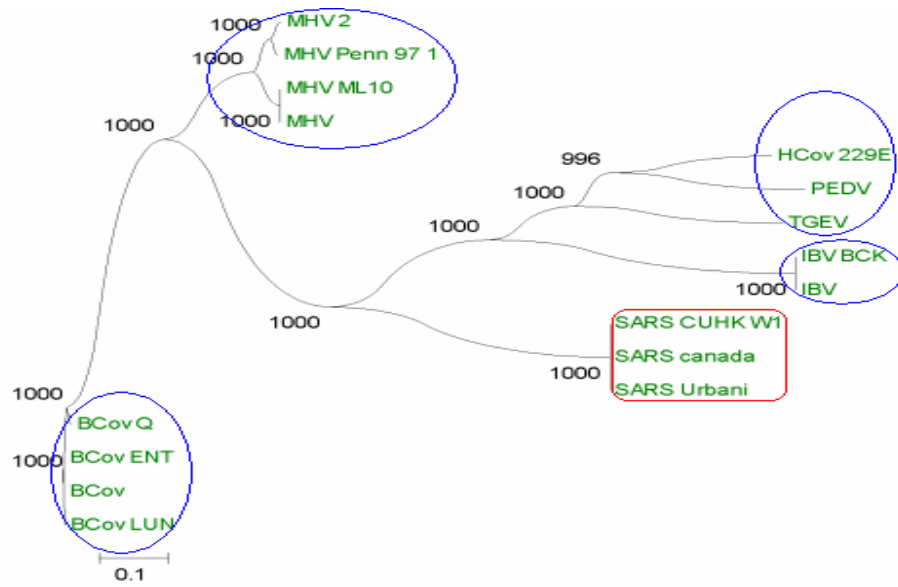


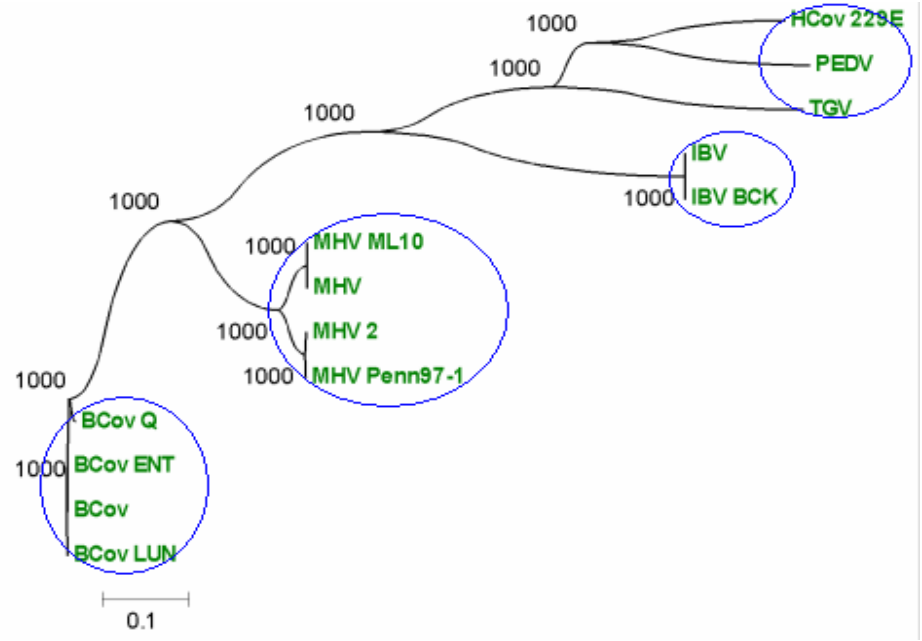
图 1 冠状病毒全基因组序列的系统发生树。(a)不包含 SARS-CoV ; (b)包含 SARS-CoV。

一般认为，全基因组序列包含了最为原始和丰富的物种演化信息，因此，全基因组序列的系统发生学分析结果应该最能反映病毒整体的演化关系。对除 SARS-CoV 之外的 13 个冠状病毒的全基因组序列进行系统发生学分析的结果（图 1-a）表明，同属传统分类法第 2 组的 BCoV 和 MHV 分立为不同的组，同属传统分类法第 1 组的 HCoV-229E、PEDV 和 TGV 仍聚于同一组，传统分类法第 3 组的 IBV 也仍独成一组。当三株 SARS-CoV 的全基因组序列参与分析后，结果（图 1-b）表明，SARS-CoV 独成一组，与 IBV 具有较近的演化关系。

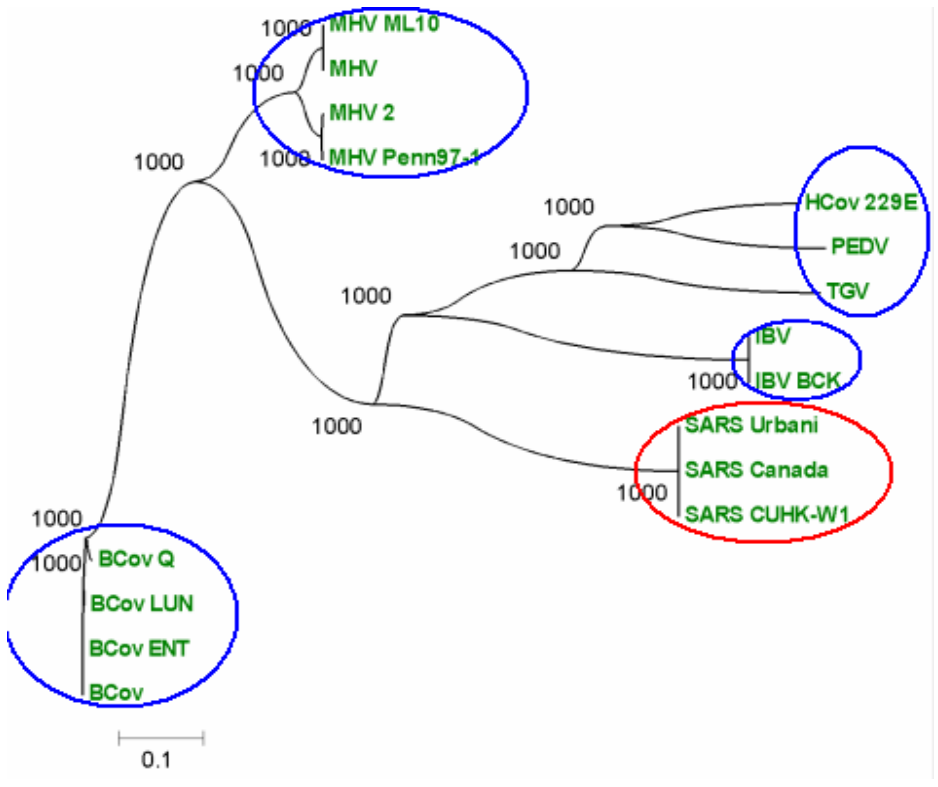
2. SARS-CoV 同源基因的系统发生学比较分析。

一般情况下，核酸序列比蛋白质序列包含更为丰富和原始的关于系统演化关系的信息，因此基因序列的系统发生学分析结果最能反映同源基因的演化关系。为探究 SARS-CoV 同源基因的演化关系，我们构建了 5 个 SARS-CoV 基因的系统发生树，并分别与其它冠状病毒同一基因的系统发生树进行比较（图 2）。

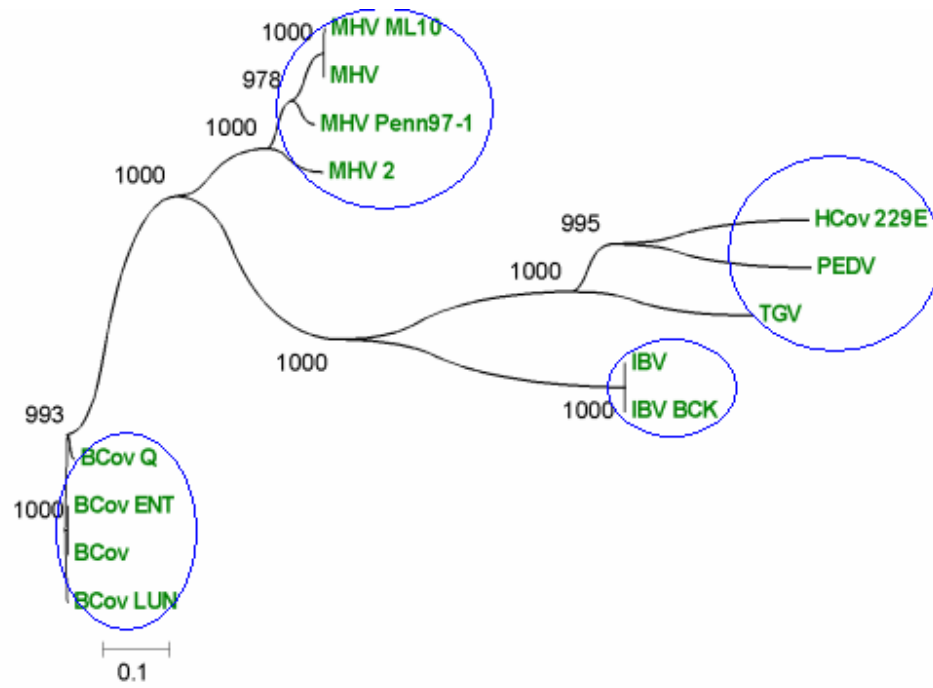
(a1)



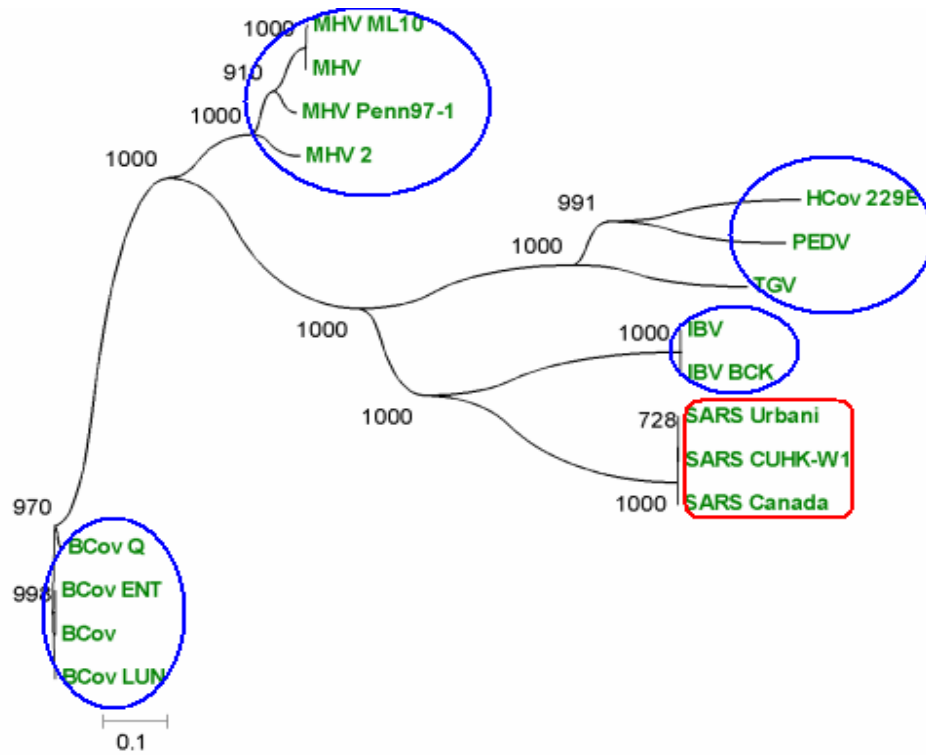
(a2)



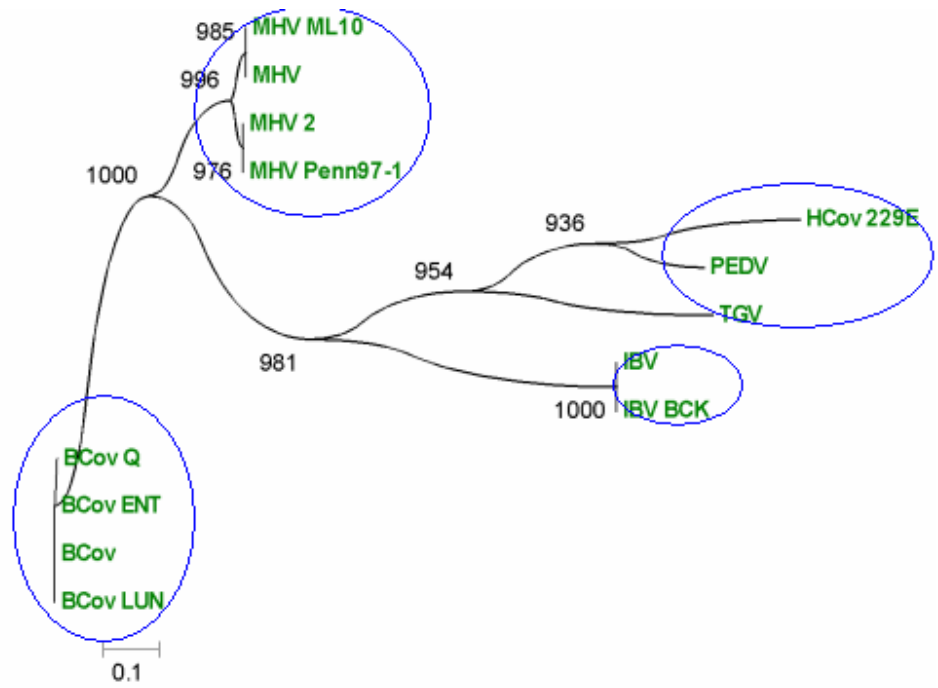
(b1)



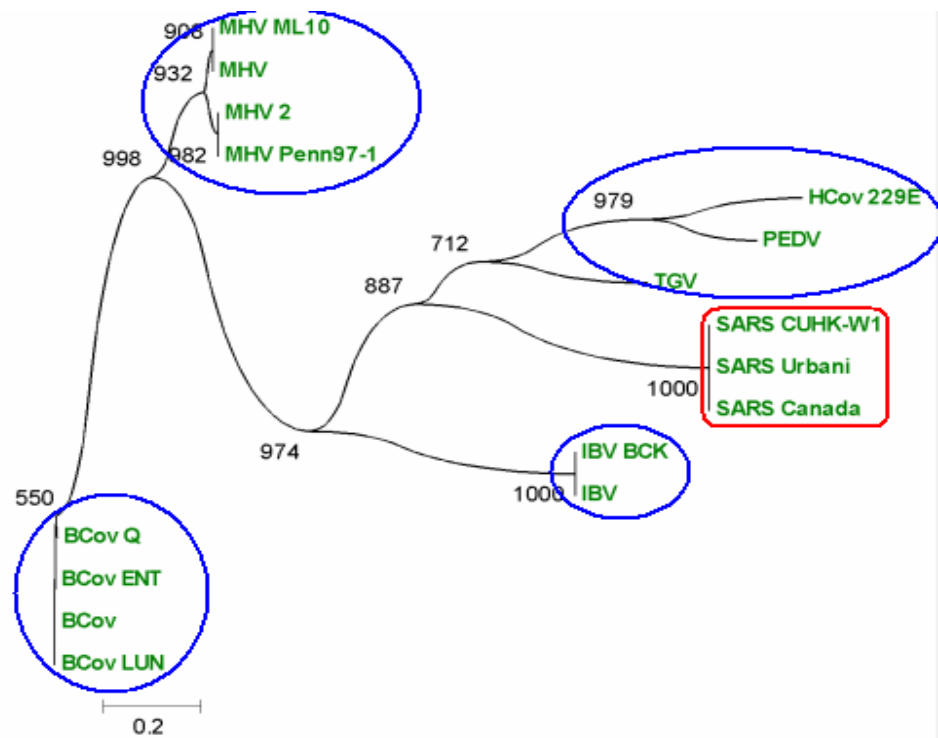
(b2)



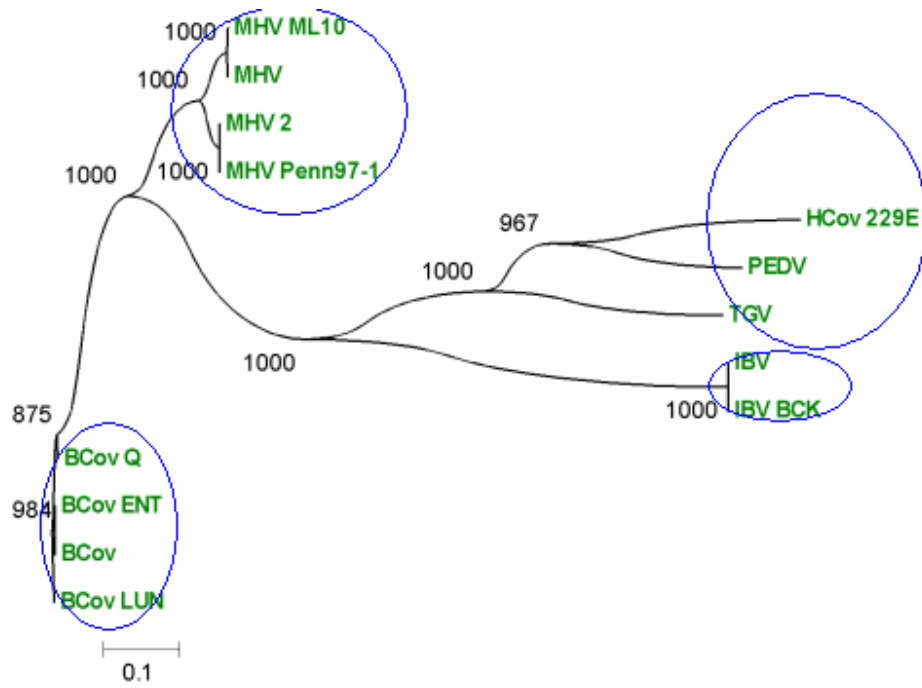
(c1)



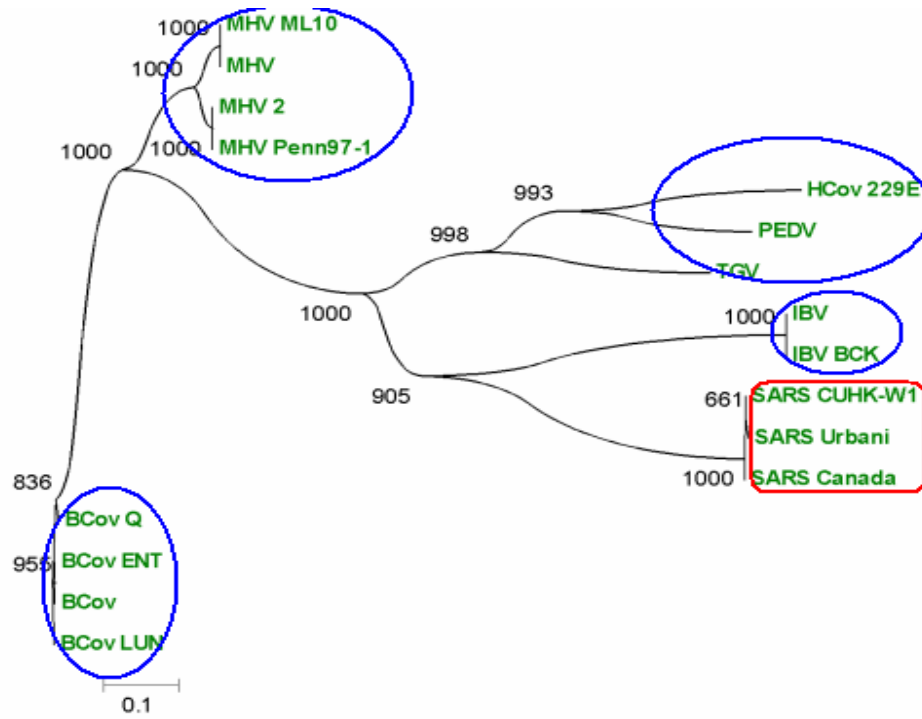
(c2)



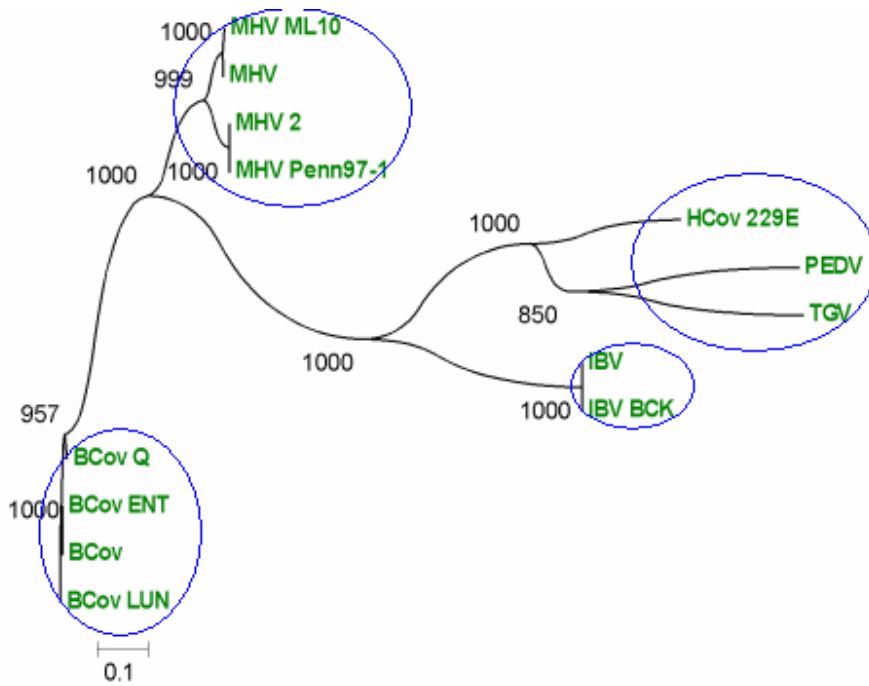
(d1)



(d2)



(e1)



(e2)

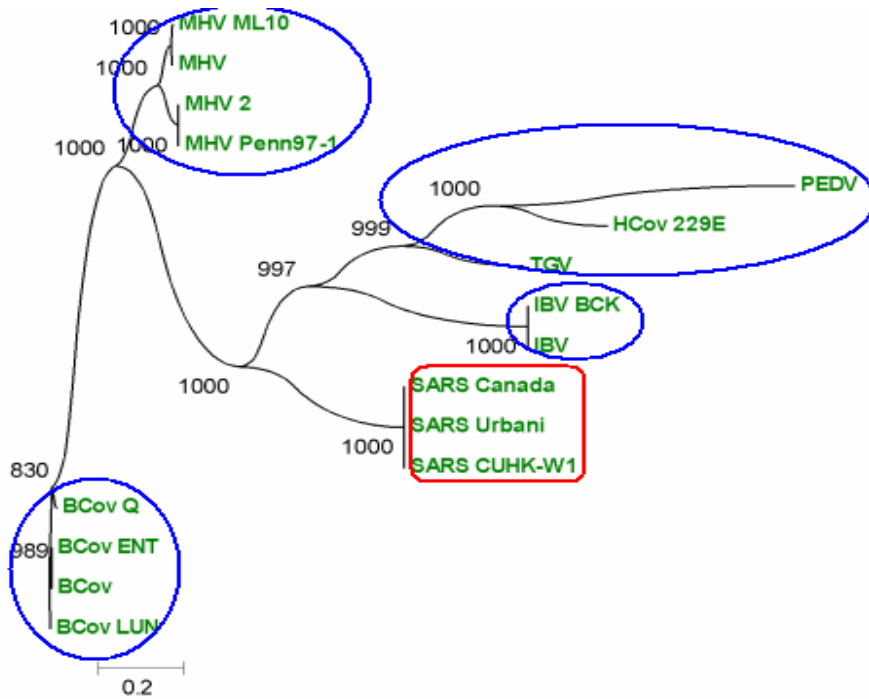


图2 冠状病毒5个主要同源基因的系统发生树。a、b、c、d和e分别表示rep、S、E、M和N基因的系统发生树，其中1表示未包含SARS-CoV，2表示包含SARS-CoV。

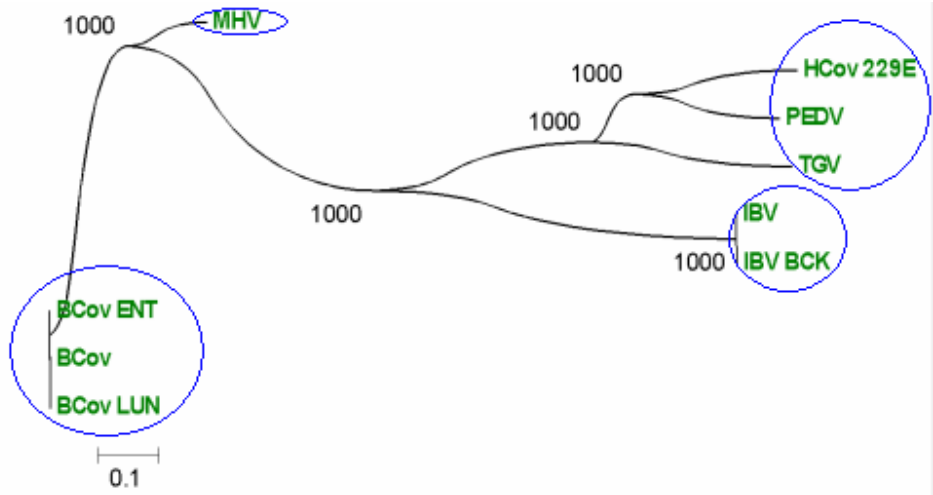
结果表明，其它冠状病毒的 rep、S、E、M 和 N 基因系统发生树与全基因组系统发生树基本一致，只有 N 基因系统发生树在传统分类法第 1 组范围内与全基因组系统发生树略有不同。那么 SARS-CoV 的 rep、S、E、M 和 N 基因系统发生树是否也与全基因组系统发生树基本一致呢？分析结果表明并非如此。S 和 M 基因的系统发生树中，SARS-CoV 与 IBV 的近缘关系较基因组水平进一步加强；E 基因的系统发生树中，SARS-CoV 与 TGV 的近缘关系较基因组水平进一步加强；N 基因的系统发生树中，TGV 的演化位置因 SARS-CoV 的加入而迁向 SARS-CoV 一侧，显示了 TGV 与 SARS-CoV 在 N 基因上相当的接近程度。只有 rep 基因的系统发生树与全基因组系统发生树基本一致。

值得注意的是，5 个同源基因各自的演化关系表明，SARS-CoV 各同源基因的演化历史彼此不同，其中结构蛋白基因的演化历史与基因组的演化历史不同，这一点与其它冠状病毒明显不同。

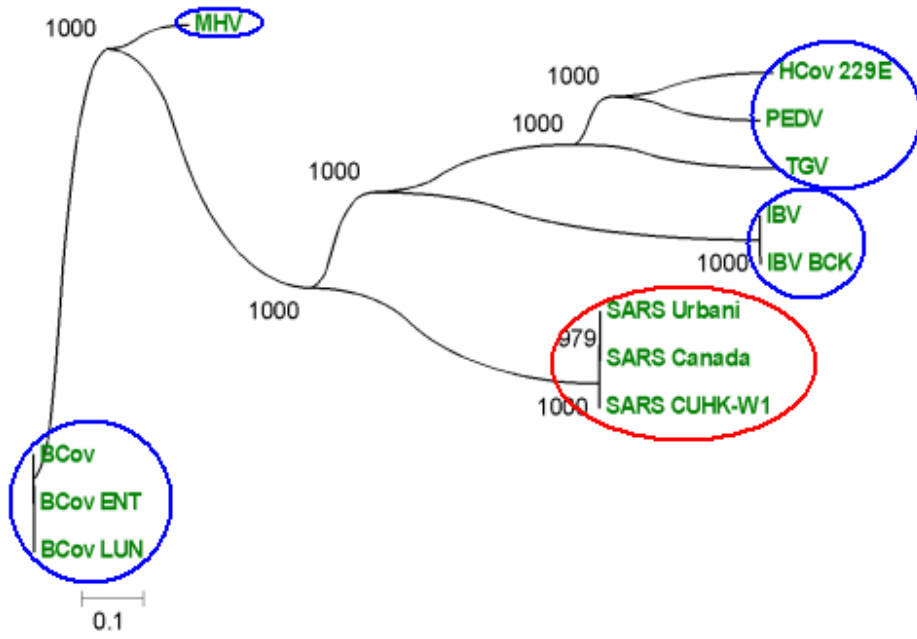
3. SARS-CoV 同源基因的系统发生学比较分析。

一般来说，蛋白序列的系统发生学分析结果有利于获取与蛋白间功能上的可能联系相关的信息。为探究 SARS-CoV 蛋白与其它冠状病毒同源蛋白间在功能上的可能联系，我们构建了 5 个 SARS-CoV 蛋白的系统发生树，并分别与其它冠状病毒同一蛋白的系统发生树进行比较（图 3）。

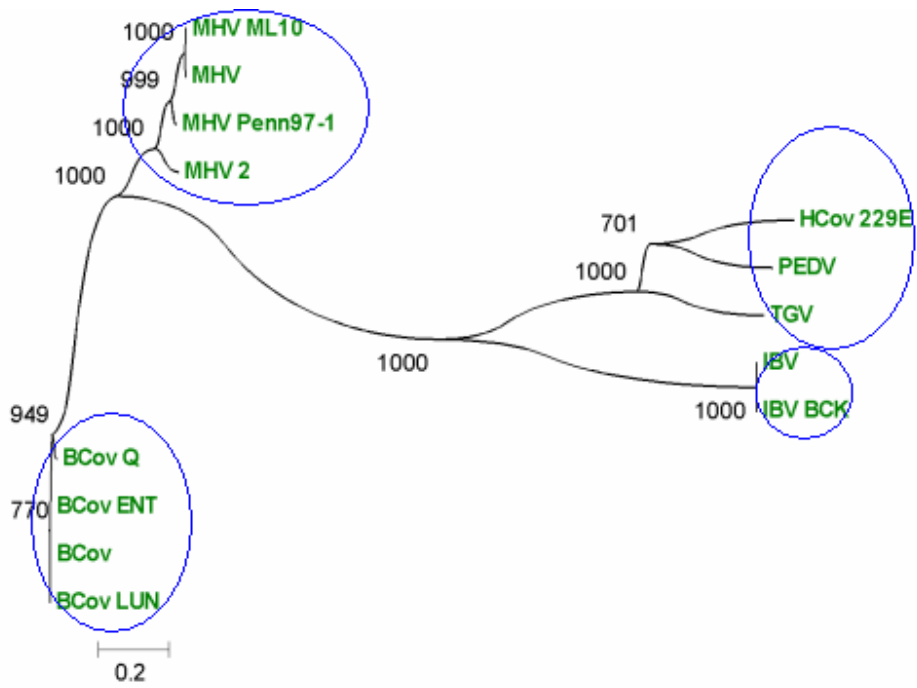
(a1)



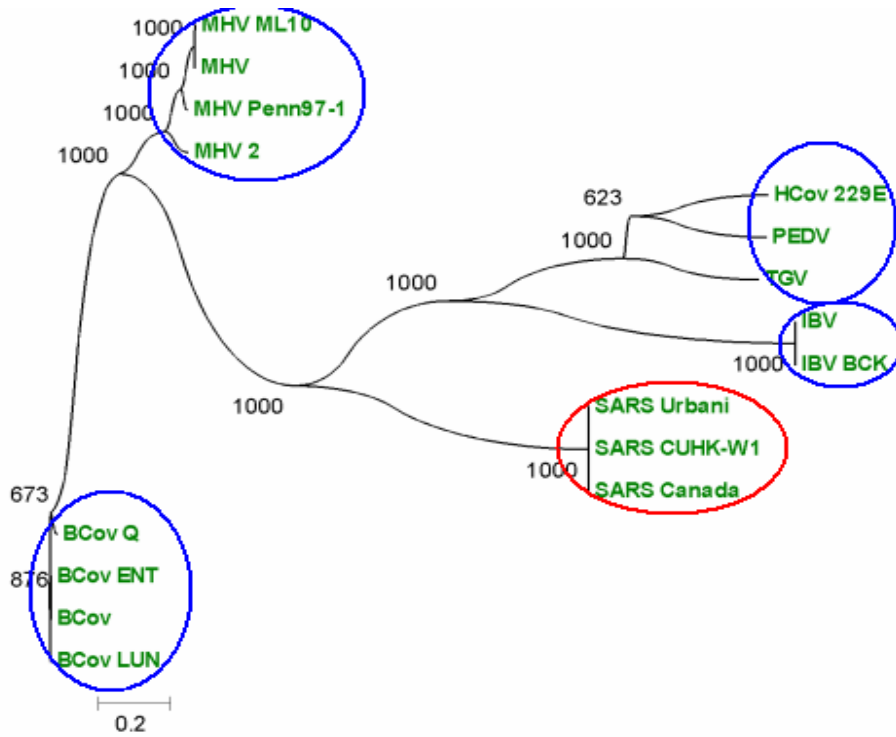
(a2)



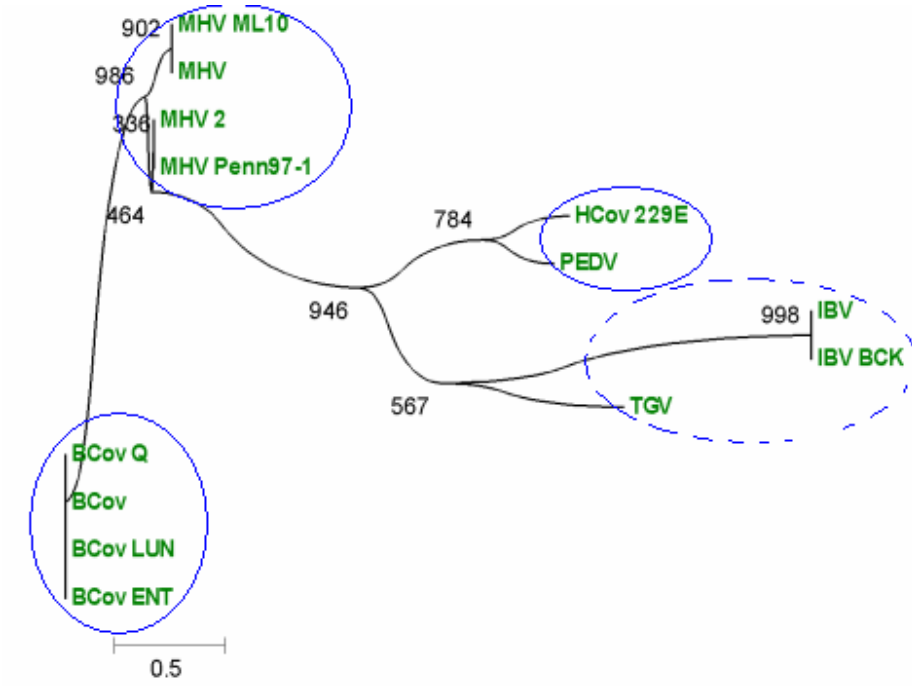
(b1)



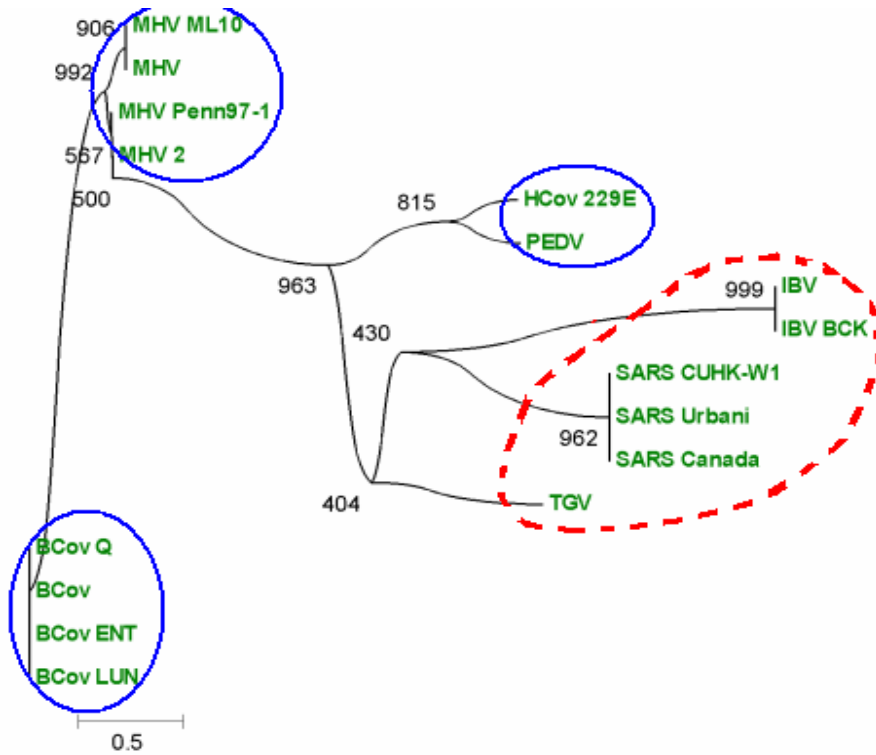
(b2)



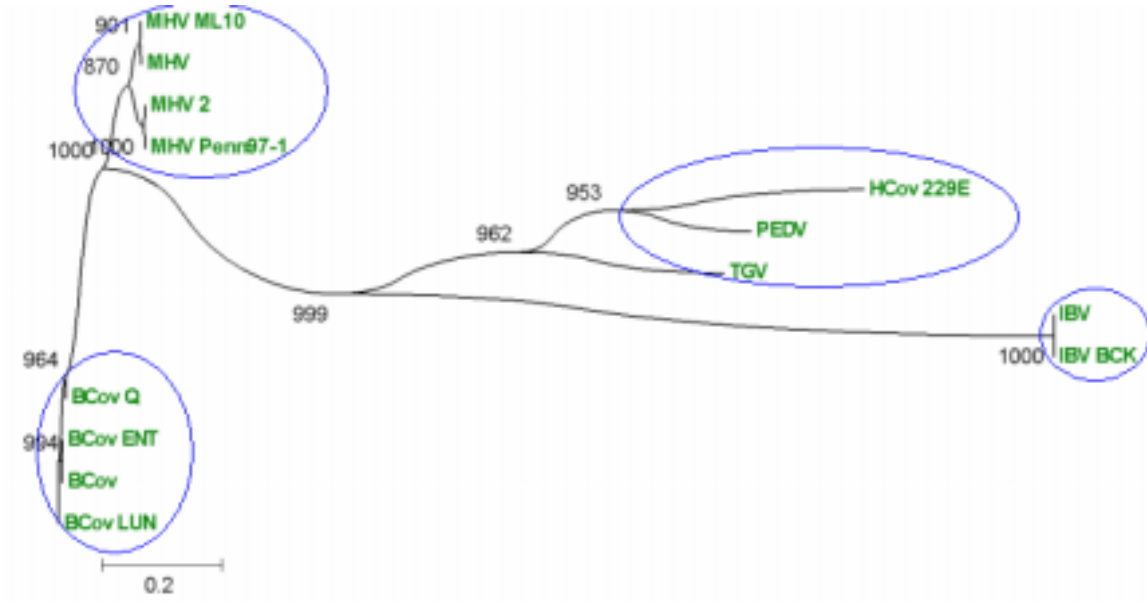
(c1)



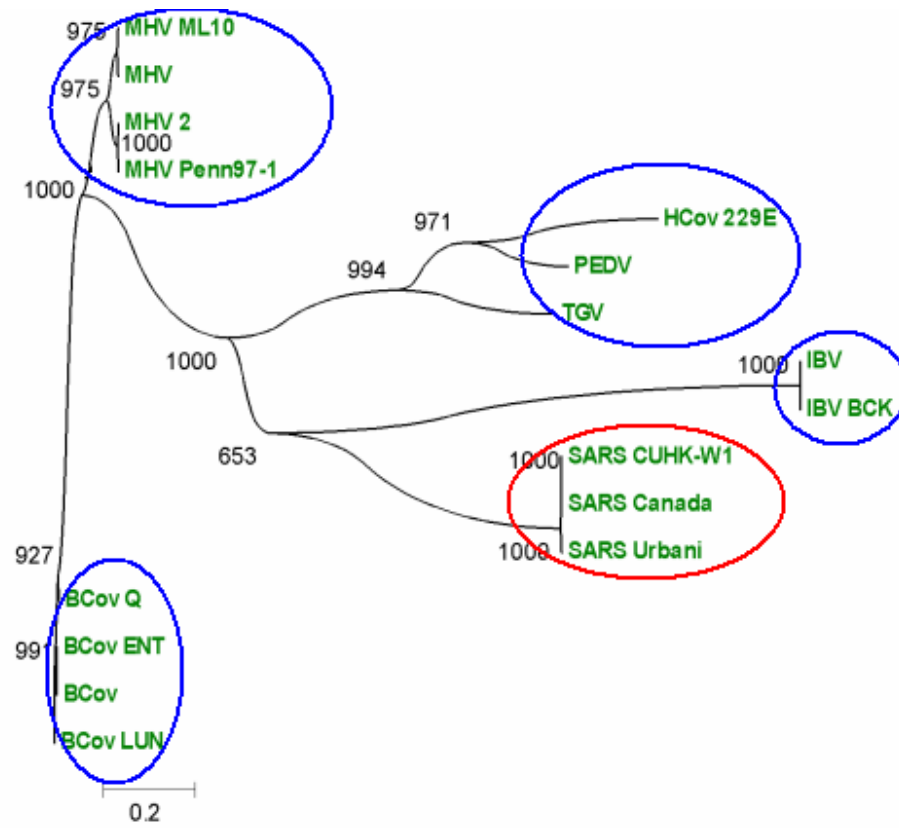
(c2)



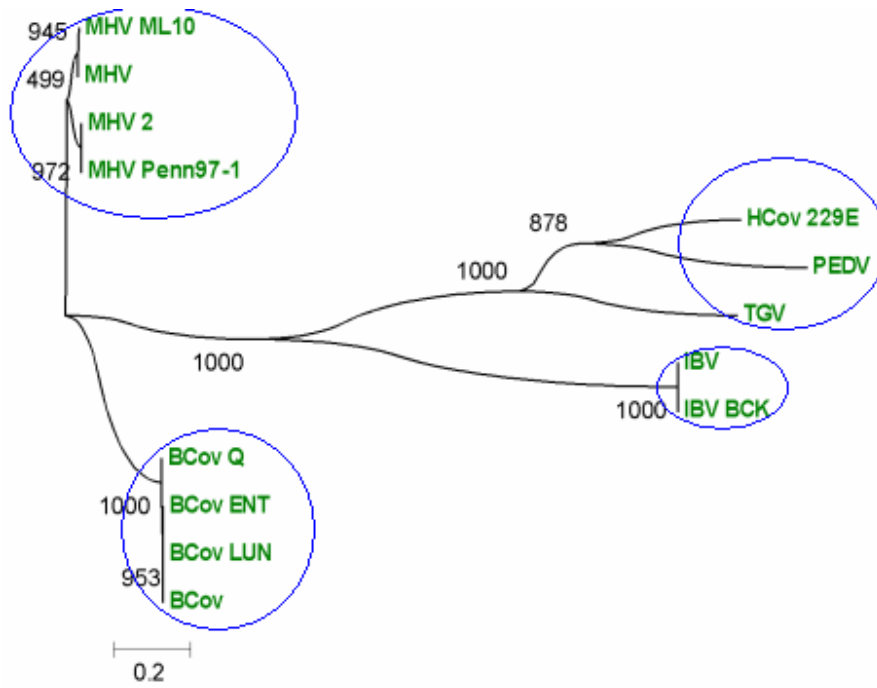
(d1)



(d2)



(e1)



(e2)

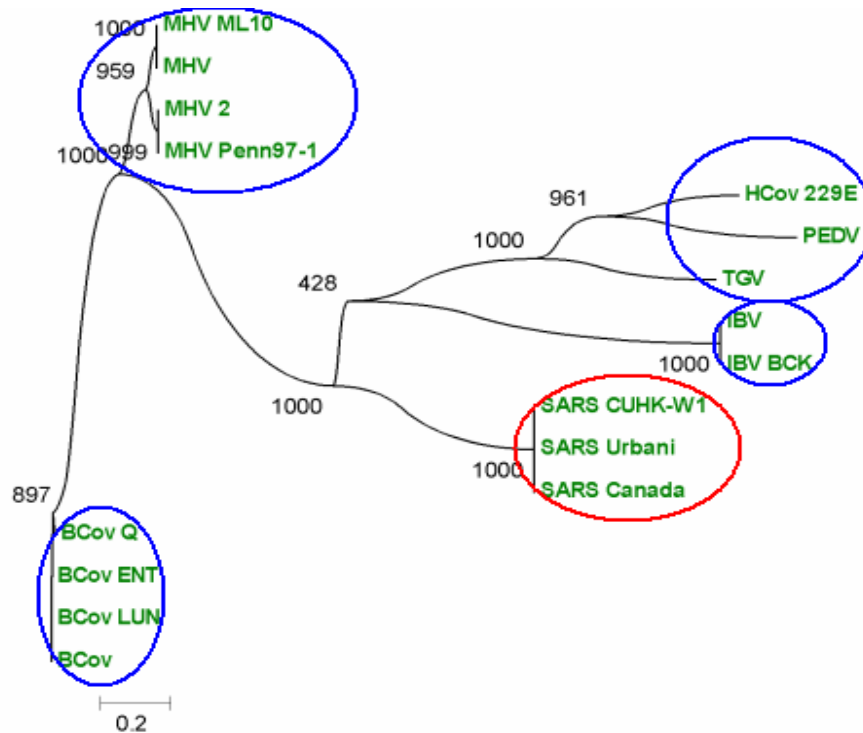


图3 冠状病毒5个主要同源蛋白的系统发生树。a、b、c、d和e分别表示rep、S、E、

M 和 N 蛋白的系统发生树，其中 1 表示未包含 SARS-CoV，2 表示包含 SARS-CoV。

结果表明，其它冠状病毒的 rep、S、M 和 N 蛋白系统发生树彼此一致，与全基因组系统发生树基本一致，只有 E 蛋白系统发生树在传统分类法第 1 组和第 3 组间出现重组现象——第 3 组的 IBV 和第 1 组的 TGV 聚成一组。那么 SARS-CoV 的 rep、S、E、M 和 N 蛋白系统发生树是否也是如此呢？分析结果表明并非完全如此：E 蛋白系统发生树中，IBV、TGV 和 SARS-CoV 的 E 蛋白虽然明确聚于一组，但彼此间演化关系不清；M 蛋白系统发生树中，SARS-CoV 的 M 蛋白与 IBV 的演化关系最近。而 rep、S 和 N 蛋白的系统发生树与全基因组系统发生树基本一致。

5 个主要同源蛋白各自的系统发生学分析结果表明，在分析 SARS-CoV 蛋白可能的毒性功能机制时，SARS-CoV 和 IBV、TGV 在 E 蛋白和 M 蛋白两水平上的特殊近缘关系值得注意和参考。

综合基因组、核酸和蛋白三水平的系统分析结果来看，SARS-CoV 与目前已知的冠状病毒同源，但具有与其它冠状病毒明显不同的特点——各同源基因的演化历史彼此不同，其中结构蛋白基因的演化历史与基因组的演化历史不同；SARS-CoV 与 IBV 和 TGV 尤其是 IBV 的亲缘关系较近，尤其是在 E 蛋白和 M 蛋白两水平上的特殊近缘关系在进一步的实验研究中值得注意和参考。

参考文献

1. P. A. Rota et al. (2003) Characterization of a Novel Coronavirus Associated with Severe Acute Respiratory Syndrome. Published online May 1, 2003; 10.1126/science.1085952

(Science Express Research Articles)

- 2 . S. M. Poutanen et al. (2003) Identification of Severe Acute Respiratory Syndrome in Canada. N. Engl. J. Med. Available April 17 at <http://nejm.org/earlyrelease/sars.asp#4-2>.
- 3 . T. G. Ksiazek, et al. (2003) A Novel Coronavirus Associated with Severe Acute Respiratory Syndrome N. Engl. J. Med. 348(20):1947-1958
- 4 . Thompson,J.D., Gibson,T.J., Plewniak,F., Jeanmougin,F. and Higgins,D.G. (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research, 24:4876-4882.
- 5 . Tree-Explore program was downloaded from http://evolgen.biol.Metro-u.ac.jp/TE/TE_man.html.