Comprehensive Detection of Candidate Pathogens in the Lower Respiratory Tract of Pediatric Patients with Unexpected Cardiopulmonary Deterioration using Next-Generation Sequencing

Suguru Takeuchi, M.D., Ph.D.[1], Jun-ichi Kawada, M.D., Ph.D.[1*], Kazuhiro Horiba M.D., Ph.D. [2,3], Makoto Yamaguchi M.D.[1], Toshihiko Okumura M.D.[1], Takako Suzuki M.D., Ph.D.[1], Yuka Torii M.D., Ph.D.[1], Shinji Kawabe M.D., Ph.D.[4], Sho Wada M.D.[5], Takanari Ikeyama M.D.[5] and Yoshinori Ito M.D., Ph.D.[1]

[1] Department of Pediatrics, Nagoya University Graduate School of Medicine, 65 Tsurumai-cho, Showa-ku, Nagoya, Aichi 466-8550, Japan

[2] Department of Genetics, Research Institute of Environmental Medicine, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8601, Japan

[3] Department of Human Genetics and Molecular Biology, Nagoya University Graduate School of Medicine, 65 Tsurumai-cho, Showa-ku, Nagoya, Aichi 466-8550, Japan

[4] Departments of Infection and Immunity, Aichi Children's Health and Medical Center, 7-426 Morioka-machi, Obu, Aichi 474-8710, Japan

[5] Division of Pediatric Critical Care Medicine, Aichi Children's Health and Medical Center, 7-426 Morioka-machi, Obu, Aichi 474-8710, Japan

***Corresponding Author:** Jun-ichi Kawada, M.D., Ph.D.
Department of Pediatrics, Nagoya University Graduate School of Medicine
65 Tsurumai-cho, Showa-ku, Nagoya, Aichi 466-8550, Japan
Tel: +81-52-744-2294; Fax: +81-52-744-2974

E-mail: kawadaj@med.nagoya-u.ac.jp

Funding statement: This work was supported by a Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan [17K10107 to J.K.].

Disclosure Statement: All authors have no conflicts of interest to declare.

Key Words: next-generation sequencing, NGS, unexpected cardiopulmonary deterioration, cardiopulmonary arrest, bronchoalveolar lavage fluid, pathogen

Abstract

***Objective*:** Next-generation sequencing (NGS) has been applied to the investigation of microorganisms in several clinical settings. We investigated the infectious etiologies in respiratory specimens from pediatric patients with unexpected cardiopulmonary deterioration using NGS.

***Design*:** Retrospective, single-center, observational study.

***Setting*:** Tertiary-care, a children's hospital.

***Subjects*:** The study enrolled a total of 16 pediatric patients with unexpected cardiopulmonary deterioration who were admitted to the pediatric intensive care unit. Ten bronchoalveolar fluid (BALF) and six transtracheal aspirate (TTA) samples were analyzed.

***Interventions*:** None.

***Measurements and Main Results*:** RNA libraries were prepared from specimens and analyzed using NGS. One or more bacterial/viral pathogens were detected in the BALF or TTA specimens from ten patients. Bacterial and viral co-infection was considered in four cases. Compared to the conventional culture and viral antigen test results, an additional six bacterial and four viral pathogens were identified by NGS. Conversely, among 18 pathogens identified by the conventional methods, nine pathogens were detected by NGS. Candidate pathogens (e.g., coxsackievirus A6 and *Chlamydia trachomatis*) were detected by NGS in four of ten patients in whom no causative pathogen had been identified by conventional methods.

***Conclusions*:** Our results suggest that viral and bacterial infections are common triggers in unexpected cardiopulmonary deterioration in pediatric patients. NGS has the potential to contribute to clarification of the etiology of pediatric critical illness.

INTRODUCTION

Unexpected cardiopulmonary deterioration events are rare in pediatric patients but can be caused by any of several etiologies, including infection, respiratory disease, and cardiac failure (1, 2). In pediatric patients, cardiopulmonary arrest (CPA) is caused primarily by progressive tissue hypoxia and acidosis due to respiratory failure or circulatory shock (3). Young *et al.* reported that the most common cause of out-of-hospital CPA in children ≤ 12 years old was sudden infant death syndrome (SIDS), whereas infectious diseases were responsible for approximately 10% of the CPA cases (4). However, the role of infection and accompanying inflammatory responses may have been underestimated as triggers of SIDS or CPA. More than half of SIDS cases have been reported to have a presumed recent viral infection (5-7). However, causative viral pathogens were not determined in most cases because comprehensive methods to detect viral pathogens have not been established.

Next-generation sequencing (NGS) has been applied to the investigation of pathogens in the field of infectious diseases (8-12). NGS is a culture-free method that can comprehensively analyze a wide range of microorganisms in a single assay without the need for specific primers. We have demonstrated the utility of NGS for the detection of candidate pathogens in bronchoalveolar lavage fluid (BALF) from pediatric patients with severe respiratory failure (13). In this study, we utilized NGS technology to investigate infectious etiologies in BALF and transtracheal aspirates (TTA) from pediatric patients with unexpected cardiopulmonary deterioration.

MATERIALS AND METHODS

Patients and Samples

BALF or TTA specimens were obtained from 16 pediatric patients admitted to the pediatric intensive care unit with unexpected cardiopulmonary deterioration from December 2018 to October 2019. A summary of the clinical characteristics of the patients is provided in Table 1. TTA specimens for NGS were obtained on admission at the same time as the clinical samples for conventional culture. Whereas, BALF specimens were obtained within 24 h after admission independently of clinical samples. The collected respiratory specimens were cryopreserved at -80°C until use in library preparation.

Library Preparation and Data Analysis

RNA was extracted from the respiratory samples and converted to cDNA. NGS libraries were prepared and sequenced on a HiSeq 2500 system (Illumina, San Diego, CA, USA). Sequence data were analyzed as previously described (13). In this study, we defined several positive cutoff values to exclude common nonpathogenic contaminants based on previous research (8, 9, 13-15). More detailed information about the process of library preparation and data analysis is provided in the supplemental section.

Ethical Considerations

This study was approved by the Institutional Review Board of Aichi Children's Health and Medical Center and Nagoya University Graduate School of Medicine. Written informed consent was obtained from all patients or their guardians.

RESULTS

A mean of 9,901,324 total reads per sample was obtained (Supplementary Table 1). In ten of 16 patients, one or more bacterial/viral pathogens were detected in the BALF or TTA specimens using NGS. Compared to the conventional culture and viral antigen test results, an additional six bacterial and four viral pathogens were identified by NGS (Table 1). Among 18 pathogens identified by the conventional methods, reads of 14 pathogens were detected by NGS. However, five of 14 pathogens did not meet the threshold for positive detection because other bacterial reads were detected more frequently (Supplementary Tables 2 and 3).

In patient 15, several types of streptococcal species met the positive cutoff of NGS, which is consistent with culture results suggesting infections by multiple species. *Chlamydia trachomatis*, which cannot be isolated by routine culture procedures, was detected by NGS in a 17-day-old neonate presenting with cardiac failure and pneumonia (patient 13).

Although the presence of a virus had been confirmed in only one patient (patient 12) in conventional tests, NGS detected significant viral sequences in five individuals (patients 2, 5, 9, 12, and 14). Notably, sequencing permitted us to define genotypes for all of the detected viruses in a single NGS assay per patient. The coverage plots of the detected viruses against each reference sequence with the read mapping approach are shown in Figure 1. Phylogenetic trees for the full genome of the detected human rhinovirus A49 (HRV-A49) and for VP1-encoding sequences of the detected coxsackievirus A6 (CA6) are shown in Supplementary Figures 1 and 2, respectively. Based on this phylogenetic analysis, the CA6 strain detected in patient 9 belongs to lineage E2 and harbors an amino acid change (T283A) in the predicted VP1 protein

domain, a substitution that was first described in the literature in 2017 (16). The consensus sequences of detected viruses that were obtained with high coverage (HRV-A49, CA6, human respiratory syncytial virus A, and human respirovirus 3) were deposited in the DNA Data Bank of Japan as accession numbers LC530048, LC530049, LC530050, and LC530051, respectively.

Conversely, pathogen-derived reads, which were considered contaminants introduced during library preparation or sequencing, were also detected. Stacked bar plots illustrating the composition of pathogen reads at the genus level of the taxonomic hierarchy are shown in Supplementary Figure 3. Details of sequenced reads annotated by microorganism are provided in Supplementary Tables 2 and 3.

DISCUSSION

We investigated infectious etiologies in the lower respiratory tract in pediatric patients with unexpected cardiopulmonary deterioration. NGS offers a comprehensive and unbiased analysis of the nucleotide sequences present in a sample, enabling the simultaneous detection of multiple microorganisms, including unexpected pathogens (8-10, 17, 18). Furthermore, NGS permitted us to determine genotypes for the detected viruses in a single test, facilitating use of the sequences in molecular epidemiological studies.

In four of the five cases in which viruses were detected, one or more bacterial species were detected simultaneously, suggesting the possibility of co-infection. Mixed infection by viruses and bacteria may lead to an aggravation of the clinical symptoms of viral infection (19). Additionally, the presence of underlying disease might lead to a severe outcome. However, it is possible that some of the detected microorganisms were just bystanders in the respiratory tract or represented contaminants introduced by resuscitation or other medical procedures (20).

The present study included six patients with CPA, and pathogens with significant reads were detected in four of these individuals. Considering that these detected pathogens may cause severe infections, some of these organisms may have been the etiologies of CPA. For example, CA6, which is a predominant circulating enterovirus strain causing hand-foot-and-mouth disease worldwide (21), was detected in patient 9. Furthermore, a previously described amino acid change (T283A) was observed in the predicted VP1 protein domain encoded by this virus. A recent study showed that CA6 encoding this amino acid substitution has the potential to cause severe

illness (16). It is possible that severe forms of CA6 infection, such as acute myocarditis or encephalitis, may serve as the etiology of CPA.

Although NGS is a powerful tool for identifying pathogens, some limitations should be discussed. First, standard methods and positive cutoff values for the detection of pathogens by NGS have not been established, and some discrepancies between NGS and conventional culture were observed. It has been shown that RNA sequencing can detect active microorganisms more efficiently than resting ones (22). Therefore, active bacterial species might have been more readily detected by NGS, while some of the species isolated by culturing did not meet the threshold for positive detection by NGS. Furthermore, the difference in sampling timing between clinical samples and BALF might be another reason for the discrepancy between NGS and conventional culture. Second, although RNA sequencing may be suitable for detecting active microorganisms, NGS-positive results do not necessarily indicate the presence of viable or pathogenic microorganisms. Therefore, the results will need to be interpreted carefully in each case.

In conclusion, we demonstrated the utility of NGS in the investigation of etiologies in pediatric patients with unexpected cardiopulmonary deterioration. Our results suggest viral and bacterial infections may be common triggers of cardiopulmonary deterioration in pediatric patients, and that NGS has the potential to contribute to the clarification of the mechanisms of pediatric critical illness.

References

1. Benneyworth BD, Bennett WE, Carroll AE. Cross-sectional comparison of critically ill pediatric patients across hospitals with various levels of pediatric care. *BMC Res Notes* 2015; 8:693.
2. Crow SS, Undavalli C,Warner DO, et al. Epidemiology of pediatric critical illness in a population-based birth cohort in Olmsted County, MN. *Pediatr Crit Care Med* 2017; 18:e137–e145.
3. Nadkarni VM, Larkin GL, Peberdy MA, et al. First documented rhythm and clinical outcome from in-hospital cardiac arrest among children and adults. *JAMA* 2006; 295:50-57.
4. Young KD, Gausche-Hill M, McClung CD, et al. A prospective, population-based study of the epidemiology and outcome of out-of-hospital pediatric cardiopulmonary arrest. *Pediatrics* 2004; 114;157.
5. Weber MA, Klein NJ, Hartley JC, et al. Infection and sudden unexpected death in infancy: a systematic retrospective case review. *Lancet* 2008; 371:1848-1853.
6. Blood-Siegfried J. The Role of Infection and Inflammation in Sudden Infant death Syndrome. *Immunopharmacol Immunotoxicol* 2009; 31:516–523.
7. Blackwell C, Moscovis S, Hall S, et al. Exploring the risk factors for sudden infant deaths and their role in inflammatory responses to infection. *Front Immunol* 2015; 6:44.
8. Chiu, C.Y., Miller, S.A. Clinical metagenomics. *Nat Rev Genet* 2019; 20:341–355.
9. Miao Q, Ma Y, Wang Q, et al. Microbiological diagnostic performance of metagenomic next-generation sequencing when applied to clinical practice. *Clin Infect Dis* 2018; 67(S2):S231-S240.

10. Houldctoft CJ, Beale MA, Breuer J. Clinical and biological insights from viral genome sequencing. *Nat Rev Microbiol* 2017; 15:183-192.
11. Kawada J, Okuno Y, Torii Y, et al. Identification of viruses in cases of pediatric acute encephalitis and encephalopathy using next-generation sequencing. *Sci Rep* 2016; 6:33452.
12. Suzuki T, Kawada J, Okuno Y, et al. Comprehensive detection of viruses in pediatric patients with acute liver failure using next-generation sequencing. *J Clin Virol* 2017; 96:67-72.
13. Takeuchi S, Kawada J, Horiba K, et al. Metagenomic analysis using next-generation sequencing of pathogens in bronchoalveolar lavage fluid from pediatric patients with respiratory failure. *Sci Rep* 2019; 9:12909.
14. Simner PJ, Miller HB, Breitwieser FP, et al. Development and Optimization of Metagenomic Next-Generation Sequencing Methods for Cerebrospinal Fluid Diagnostics. *J Clin Microbiol* 2018; 56:e00472-18.
15. Miller S, Naccache SN, Samayoa E, et al. Laboratory validation of a clinical metagenomic sequencing assay for pathogen detection in cerebrospinal fluid. *Genome Res* 2019; 29:831-842.
16. Yang X, Li Y, Zhang C, et al. Clinical features and phylogenetic analysis of severe hand-foot-and-mouth disease caused by Coxsackievirus A6. *Infect Genet Evol* 2020; 77:104054.
17. Horiba K, Kawada J, Okuno Y, et al. Comprehensive detection of pathogens in immunocompromised children with bloodstream infections by next-generation sequencing. *Sci Rep* 2018; 8:3784.
18. Takeuchi S, Kawada J, Okuno Y, et al. Identification of potential pathogenic viruses

in patients with acute myocarditis using next-generation sequencing. *J Med Virol* 2018; 90:1814-1821.

19. Nolan VG, Arnold SR, Bramley AM, et al. Etiology and impact of coinfections in children hospitalized with community-acquired pneumonia. *J Infect Dis* 2018; 218:179-188.
20. Alfelali M, Khandaker G. Infectious causes of sudden infant death syndrome. *Paediatr Respir Rev* 2014; 15:307-311.
21. Zeng H, Lu J, Zheng H, et al. The epidemiological study of coxsackievirus A6 revealing hand, hoot and mouth disease epidemic patterns in Guangdong, China. *Sci Rep* 2015; 5:10550.
22. Ren L, Zhang R, Rao J, et al. Transcriptionally active lung microbiome and its association with bacterial biomass and host inflammatory status. *mSystems* 2018; 3:e00199-18.

Figure legends

Figure 1. Coverage plots of viral genomes detected in critically ill children.

Sequencing reads detected in respiratory specimens of each patient were mapped to the reference genome of human rhinovirus A49, human coronavirus NL63, coxsackievirus A6, human respiratory syncytial virus A, and human respirovirus 3. Light blue, blue, and dark blue colors in the viral genome alignments represent minimal, mean, and maximal coverage in the aggregated 10-bp regions, respectively.

Table 1. Patient characteristics and detected pathogens

Pt No.	Age	Sex	Underlying disease	Diagnostic category	Sample type	NGS results (RPM)		Conventional test results[d]
						Bacteria	Virus	
1	6y0m	M	-	CPA	BALF	-	-	-
2	9y8m	F	-	Acute cardiac failure Arrhythmia	BALF	*S. pneumoniae* (101)	HRV-A49 (742)	-
3	0y5m	M	-	Acute respiratory failure	BALF	-	-	-
4	0y2m	F	-	Acute respiratory failure	BALF	*A. guillouiae* (99)	-	-
5	1y3m	F	Emanuel syndrome TOF[a]	Acute cardiac failure	BALF	***M. catarrhalis*** (111) ***H. influenzae*** (78)	HCoV-NL63 (79)	***H. influenzae*** (2+) *S. pneumoniae* (1+) ***M. catarrhalis*** (±)
6	0y2m	M	-	Acute cardiac failure	BALF	-	-	-
7	0y5m	M	-	CPA	BALF	***S. pneumoniae*** (16,417)	-	*H. influenzae* (2+) ***S. pneumoniae*** (2+) *α-Streptococcus sp.* (2+)
8	0y6m	M	-	CPA	BALF	***M. catarrhalis*** (230)	-	*S. pneumoniae* (3+) *H. influenzae* (3+) ***M. catarrhalis*** (3+)
9	1y4m	F	West syndrome[b]	CPA	BALF	-	Coxsackievirus A6 (86)	-
10	9y11m	F	-	Fulminant myocarditis	BALF	-	-	-
11	4y7m	M	-	CPA Fulminant myocarditis	TTA	-	-	-

12	1y7m	F	-	Fulminant myocarditis	TTA	***S. pneumoniae*** (6,126) *S. oralis* (1,451)	**HRSV- A** (1,651)	***S. pneumoniae*** (2+) *H. influenzae* (2+) **HRSV**[e]
13	d17	F	-	Acute cardiac failure Pneumonia	TTA	*C. trachomatis* (76)	-	-
14	6y0m	F	Epilepsy[c]	Acute cardiac failure	TTA	***P. aeruginosa*** (1,508)	PIV-3 (243)	*P. fluorescens/putida* (±) ***P. aeruginosa*** (±) *α-Streptococcus sp.* (±)
15	0y3m	M	-	CPA	TTA	***S. oralis*** (1,783) ***S. pneumoniae*** (1,602) ***S. mitis*** (1,058) *R. mucilaginosa* (869) *S. salivarius* (768)	-	***S. pneumoniae*** **(2+)** ***α-Streptococcus sp.*** (2+) *M. catarrhalis* (±)
16	5y4m	F	-	Acute myocarditis	TTA	-	-	-

Abbreviation: BALF, bronchoalveolar lavage fluid; CPA, cardiopulmonary arrest; HCoV-NL63, Human coronavirus NL63; HRSV, Human respiratory syncytial virus; HRV-A49, Human rhinovirus A49; PIV-3, Parainfluenzavirus 3 (Human respirovirus 3); RPM, reads per million; TOF, Tetralogy of Fallot; TTA, transtracheal aspirates. Bold letters indicate identical pathogens between NGS and conventional tests.

[a]A patient with Emanuel syndrome and TOF after Rastelli repair.

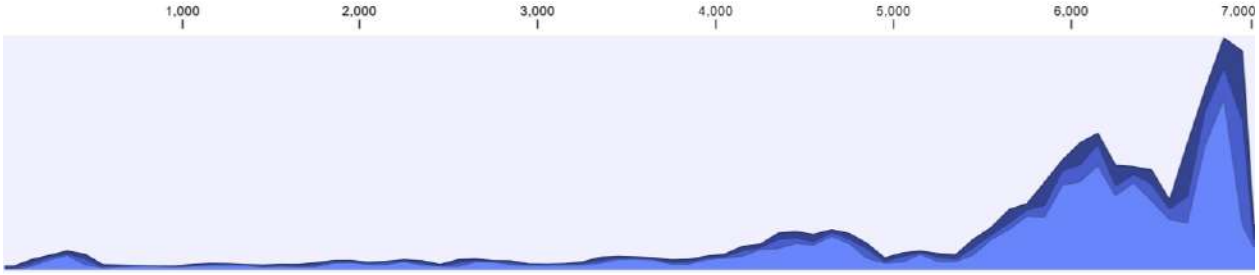
[b]A patient with West syndrome after group B streptococcal meningitis.

[c]A patient with epilepsy after acute encephalopathy.

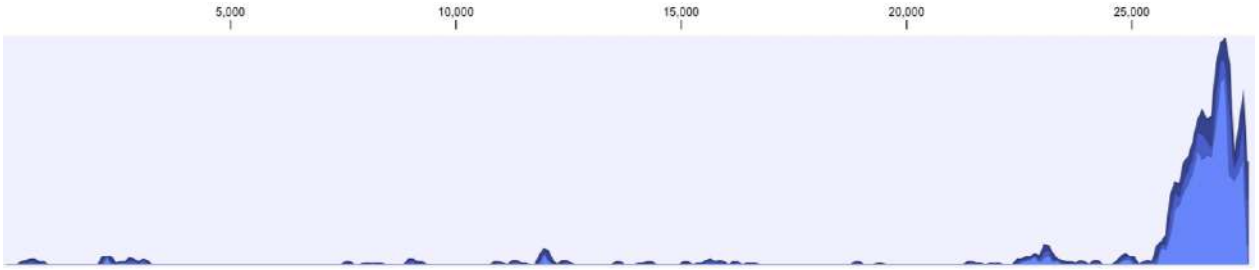
[d]All detected pathogens by conventional methods except for human respiratory syncytial virus in patient 12 reflected bacterial culture test results with transtracheal aspirates.

[e]Positive for antigen test.

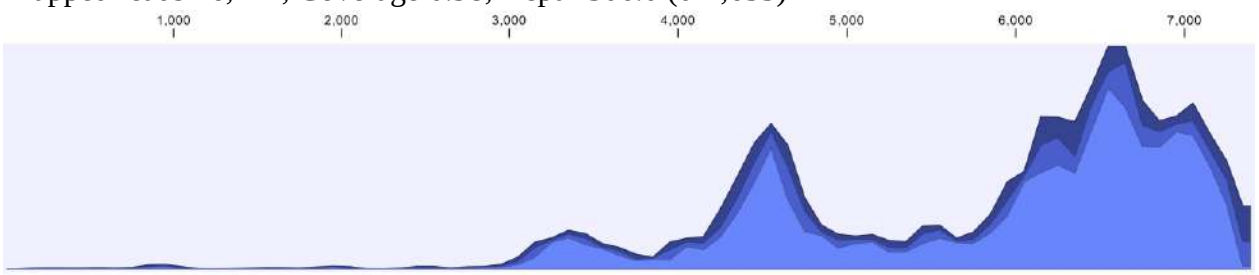
Pt 2
Human rhinovirus A49 (KY369896)
Mapped reads 55,827, Coverage 1.00, Depth 10,726.8 (266-83,752)



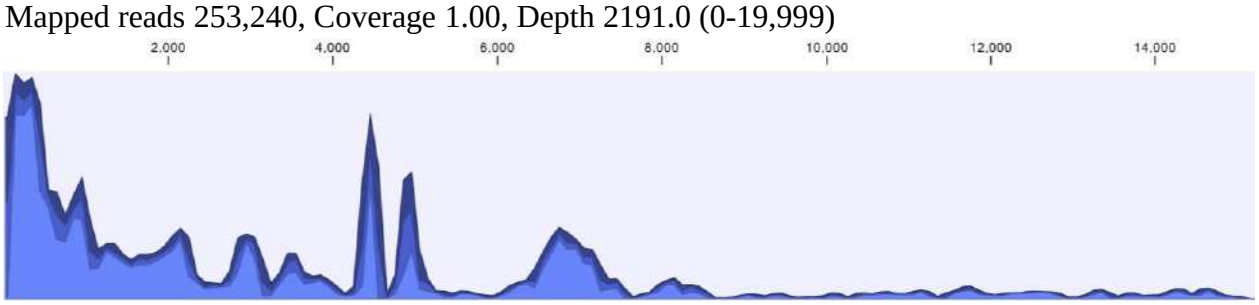
Pt 5
Human coronavirus NL63 (NC_005831)
Mapped reads 1,317, Coverage 0.34, Depth 6.6 (0-170)



Pt 9
Coxsackievirus A6 (MN032612)
Mapped reads 16,417, Coverage 0.98, Depth 306.0 (0-1,633)



Pt 12
Human respiratory syncytial virus A (MG793382)
Mapped reads 253,240, Coverage 1.00, Depth 2191.0 (0-19,999)



Pt 14
Human respirovirus 3 (NC_001796)
Mapped reads 13,916, Coverage 0.99, Depth 119.8 (0-1,353)

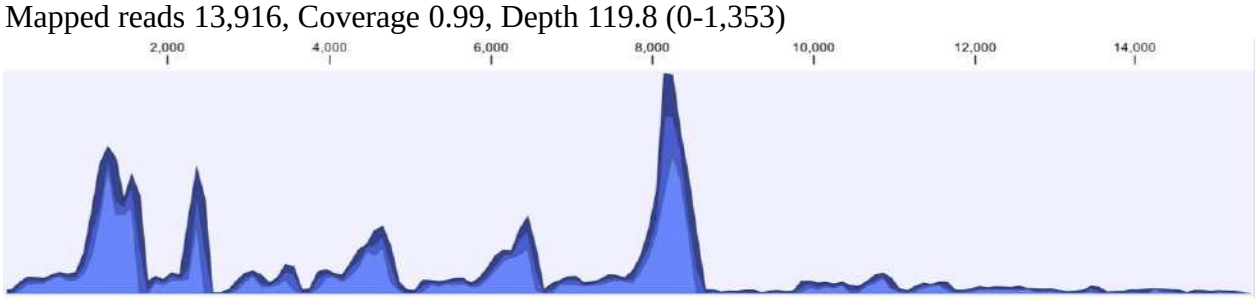
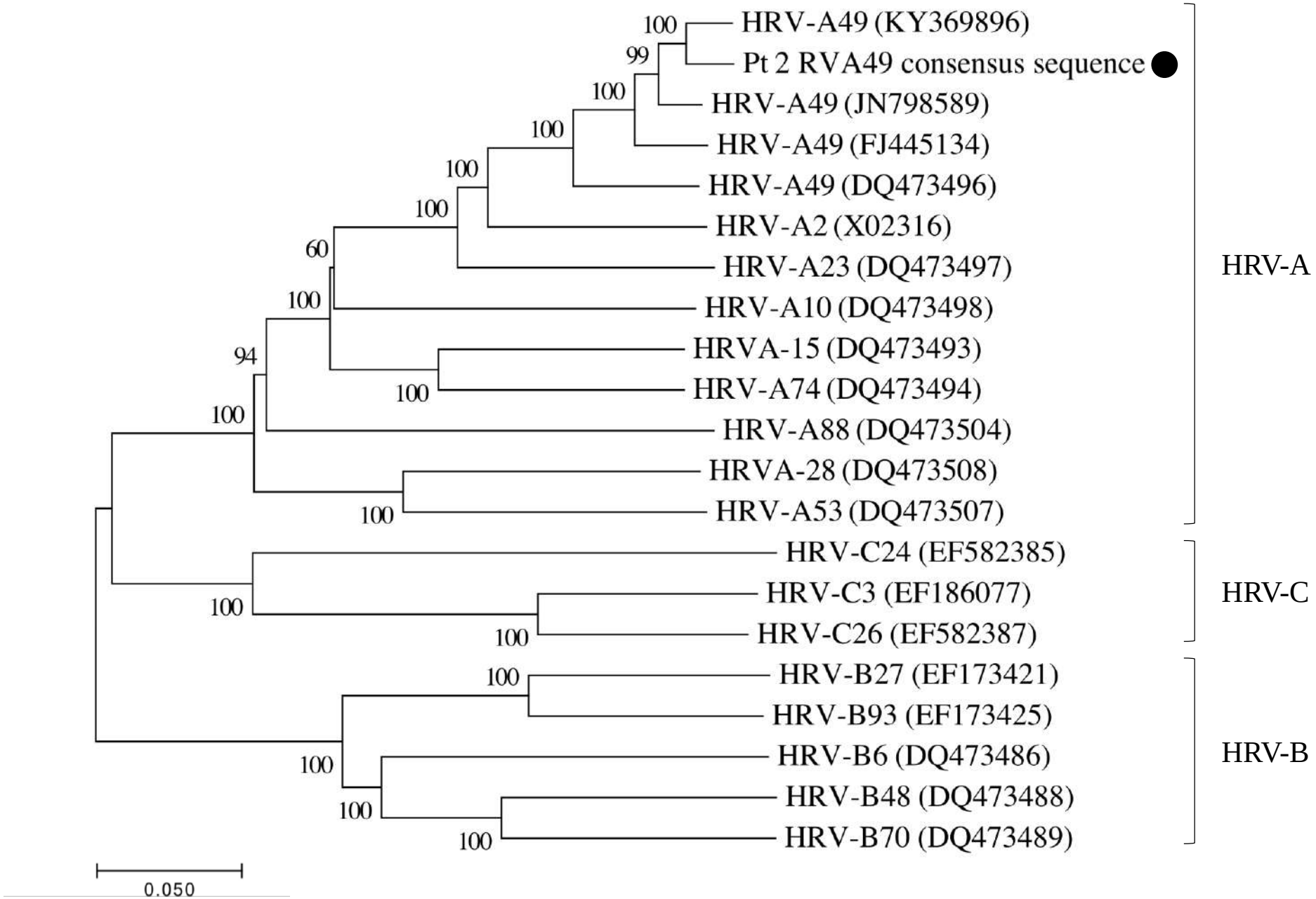
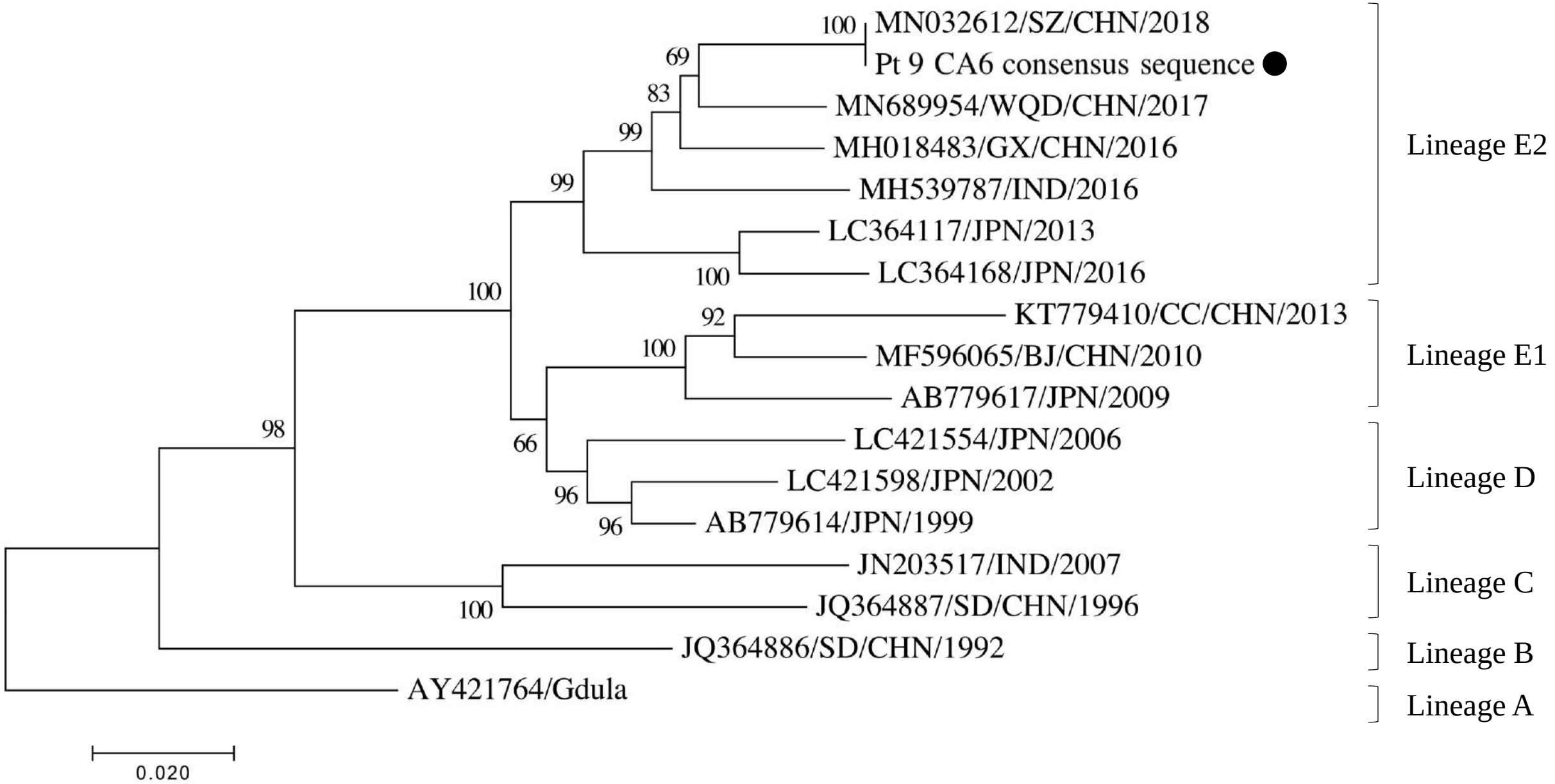


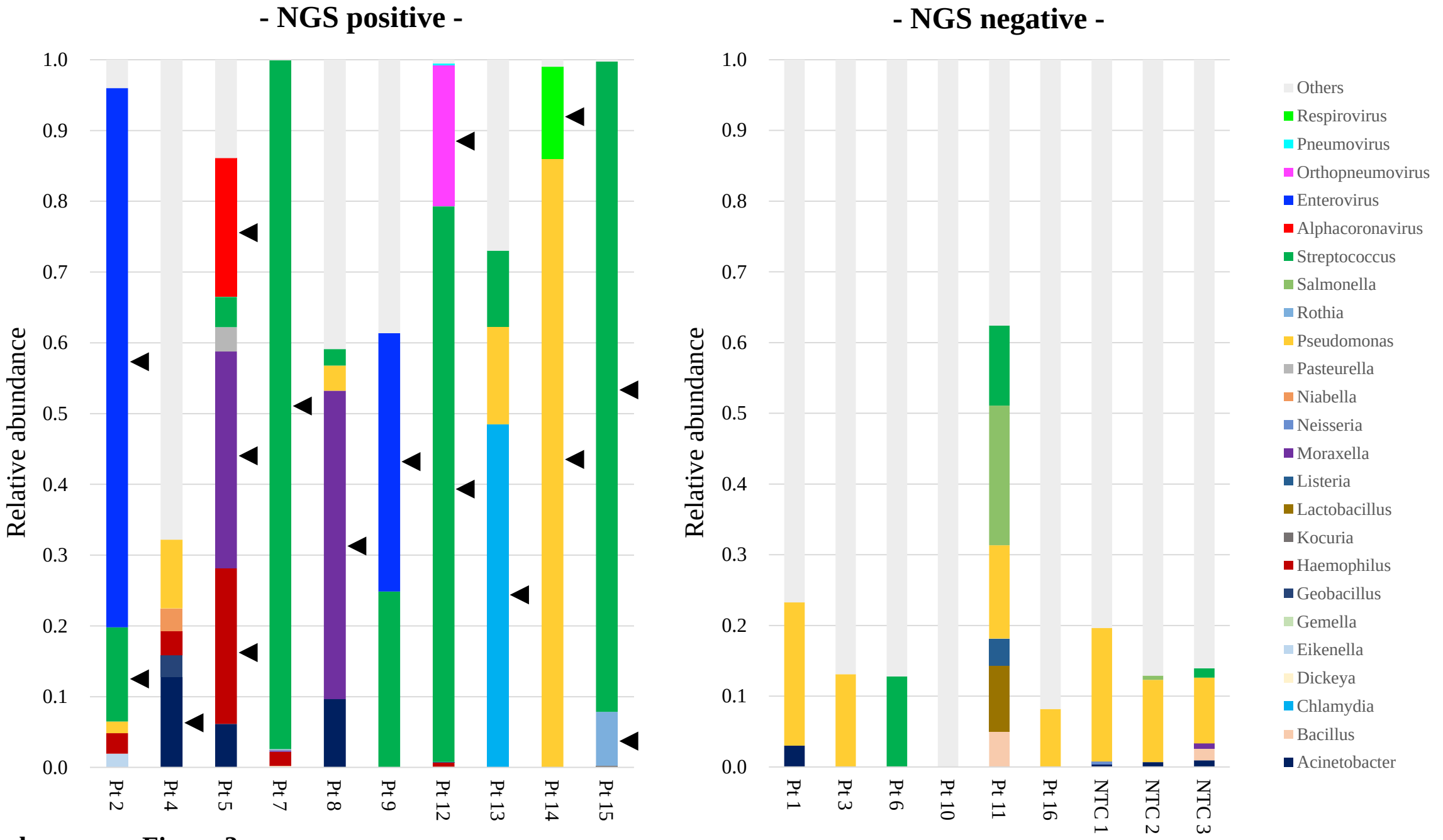
Figure 1



Supplementary Figure 1



Supplementary Figure 2



Supplementary Figure 3

Supplementary Information

Comprehensive Detection of Candidate Pathogens in the Lower Respiratory Tract of Pediatric Patients with Unexpected Cardiopulmonary Deterioration using Next-Generation Sequencing

Suguru Takeuchi, M.D., Ph.D.[1], Jun-ichi Kawada, M.D., Ph.D.[1*], Kazuhiro Horiba M.D., Ph.D.[2,3], Makoto Yamaguchi M.D.[1], Toshihiko Okumura M.D.[1], Takako Suzuki M.D., Ph.D.[1], Yuka Torii M.D., Ph.D.[1], Shinji Kawabe M.D., Ph.D.[4], Sho Wada M.D.[5], Takanari Ikeyama M.D.[5] and Yoshinori Ito M.D., Ph.D.[1]

[1] Department of Pediatrics, Nagoya University Graduate School of Medicine, 65 Tsurumai-cho, Showa-ku, Nagoya, Aichi 466-8550 Japan

[2] Department of Genetics, Research Institute of Environmental Medicine, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8601, Japan

[3] Department of Human Genetics and Molecular Biology, Nagoya University Graduate School of Medicine, 65 Tsurumai-cho, Showa-ku, Nagoya, Aichi 466-8550, Japan

[4] Departments of Infection and Immunity, Aichi Children's Health and Medical Center, 7-426 Morioka-machi, Obu, Aichi 474-8710, Japan

[5] Division of Pediatric Critical Care Medicine, Aichi Children's Health and Medical Center, 7-426 Morioka-machi, Obu, Aichi 474-8710, Japan

***Corresponding Author:** Jun-ichi Kawada, M.D., Ph.D.

1 Department of Pediatrics, Nagoya University Graduate School of Medicine
2 65 Tsurumai-cho, Showa-ku, Nagoya, Aichi 466-8550, Japan
3 Tel: +81-52-744-2294; Fax: +81-52-744-2974
4 E-mail: Kawadaj@med.nagoya-u.ac.jp
5

Supplementary Methods

Library Preparation

RNA was extracted from each respiratory sample using the NucleoSpin RNA Blood kit (MACHEREY-NAGEL, Düren, Germany). The extracted RNA was converted to cDNA and amplified with the REPLI-g WTA Single Cell kit (Qiagen, Hilden, Germany). Then, the synthesized cDNA was used to prepare a sequencing library with the Nextera XT DNA Sample Preparation Kit (Illumina, San Diego, CA, USA). The library quality was analyzed using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA), a Qubit dsDNA HS assay kit (Thermo Fisher Scientific, Waltham, MA, USA), and the QX200TM Droplet Digital PCR System (Bio-Rad, Richmond, CA, USA). The libraries were sequenced on a HiSeq 2500 system (Illumina) using the 2 × 150 bp paired-end protocol. In parallel, a no-template control (NTC), generated from distilled water was also prepared and sequenced using the same protocol as that described above.

Data Analysis

Sequence data were processed and analyzed with the cloud-computing metagenomic pipeline MePIC v2.0 (National Institute of Infectious Disease, Tokyo, Japan) (1). First, unnecessary adapter sequences, low-quality bases (Q-score cutoff, 20), and short reads (length cutoff, 50) were removed by trimming. Second, human-derived reads were detected using the BWA program and removed from the downstream analysis. Third, for the remaining reads, the MEGABLAST program was used to search for sequences similar to those registered in the NCBI nucleotide database (E-value cutoff, 1e-30). Finally, the taxonomic classification of each hit was summarized using

MEGAN6 (University of Tübingen, Tübingen, Germany) (2). Alignment with each viral reference genome was performed with the CLC Genomics Workbench 9.5 (CLC bio; Qiagen). Phylogenetic analysis of the full genome of human rhinovirus A49 and of the VP1-encoding region of coxsackievirus A6 was conducted using the neighbor-joining method in MEGA7 (3).

Threshold criteria for detected pathogens

In the present study, we defined several positive cutoff values to exclude common nonpathogenic contaminants based on previous studies, as described below. Bacterial pathogens detected by next-generation sequencing (NGS) were considered significant when all of the following criteria were satisfied: bacterial reads per million of total reads (RPM) ≥50, relative abundance of bacterial reads (RA) ≥0.05, and RPM-ratio (RPM-r) ≥10 (4-9). If more than six bacterial pathogens fulfilled the criteria, the top five were considered NGS positive (4). By contrast, virus-derived reads could be buried in the host genome and overlooked because the viral genome length is much shorter than that of bacteria. Therefore, the threshold of RA was not applied for viruses, and the virus-derived reads detected by NGS were considered significant when all of the following criteria were satisfied: RPM ≥10, RPM-r ≥10, and nonoverlapping reads from ≥3 distinct viral genomic regions were confirmed (4, 5).

In the previous study by Zinter *et al.*, which employed RNA sequencing to investigate pathogens in respiratory specimens, RPM ≥10 and ≥1 were used as the threshold for positive detection of bacteria and viruses, respectively (6). Moreover, we found that a large number of potentially pathogenic bacterial and viral reads were detected in BALF samples obtained from pediatric patients with respiratory failure (7).

Therefore, in the present study, we used the threshold of RPM ≥50 and ≥10 for positive detection of bacteria and viruses, respectively, to avoid making calls based on potentially spurious alignments.

RA ≥0.01was used as the threshold for positive detection of bacteria in the previous study by Ren *et al* (8). Furthermore, our previous study has shown that culture-positive bacteria were readily detectable with RA ≥0.05 by both DNA and RNA sequencing (7). Therefore, in the present study, we used the threshold of RA ≥0.05 for positive detection of bacteria.

RPM-r was defined as "RPM-r = RPM_{sample}/RPM_{NTC}"; that is, RPM-r was calculated as the RPM corresponding to a given species in the clinical sample divided by the RPM in the NTC. Based on previous reports, RPM-r ≥10 was used to exclude common nonpathogenic contaminants that originated from the environment, laboratory reagents, and cross-contamination (4, 5, 9).

Using the above mentioned cutoff values, we confirmed that no significant bacterial or viral reads were detected by NGS in culture-negative BALF specimens obtained from six patients with idiopathic pulmonary fibrosis (data not shown).

Figure legends

Supplementary Figure 1. Phylogenetic analysis of human rhinovirus A49.

A phylogenetic tree of the full genome of the human rhinovirus A49 (HRV-A49) consensus sequence obtained from patient 2 is shown. The phylogenetic relationships of the registered sequences of HRV, including other genotypes, were estimated using the neighbor-joining method with 1,000 replicates in MEGA7 (3). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test is shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the p-distance method and are in units of the number of base differences per site. Scale bars show the genetic distance.

Supplementary Figure 2. Phylogenetic analysis of coxsackievirus A6.

A phylogenetic tree of the VP1-encoding sequence of coxsackievirus A6 (CA6) consensus sequence obtained from patient 9 is shown. The phylogenetic relationships of the registered sequences of CA6 were estimated using the neighbor-joining method with 1,000 replicates in MEGA7 (3). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test is shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the p-distance method and are in units of the number of base differences per site. Lineages were defined previously based on VP1-encoding sequences (10). Scale bars show the genetic distance.

Supplementary Figure 3. The composition of pathogen reads at the genus level of the taxonomic hierarchy.

Stacked bar plots illustrating the composition of pathogen reads at the genus level of the taxonomic hierarchy are shown separately for NGS-positive and -negative samples. Closed arrowheads indicate a pathogen judged as significant by the NGS-based approach.

References

1. Takeuchi F, Sekizuka T, Yamashita A, et al. MePIC, metagenomic pathogen identification for clinical specimens. *Jpn J Infect Dis* 2014; 67:62-65.
2. Huson DH, Meier S, Flade I, et al. MEGAN community edition -intractive exploration and analysis of large-scale microbe sequencing data. *PLoS Comut Biol* 2016; 12:e1004957.
3. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* 2016; 33:1870-1874.
4. Simner PJ, Miller HB, Breitwieser FP, et al. Development and Optimization of Metagenomic Next-Generation Sequencing Methods for Cerebrospinal Fluid Diagnostics. *J Clin Microbiol* 2018; 56:e00472-18.
5. Miller S, Naccache SN, Samayoa E, et al. Laboratory validation of a clinical metagenomic sequencing assay for pathogen detection in cerebrospinal fluid. *Genome Res* 2019; 29:831-842.
6. Zinter MS, Dvorak CC, Mayday MY, et al. Pulmonary Metagenomic Sequencing Suggests Missed Infections in Immunocompromised Children. *Clin Infect Dis* 2019; 68:1847-1855.
7. Takeuchi S, Kawada J, Horiba K, et al. Metagenomic analysis using next-generation sequencing of pathogens in bronchoalveolar lavage fluid from pediatric patients with respiratory failure. *Sci Rep* 2019; 9:12909.
8. Ren L, Zhang R, Rao J, et al. Transcriptionally active lung microbiome and its association with bacterial biomass and host inflammatory status. *mSystems* 2018; 3:e00199-18.
9. Miao Q, Ma Y, Wang Q, et al. Microbiological diagnostic performance of

metagenomic next-generation sequencing when applied to clinical practice. *Clin Infect Dis* 2018; 67(S2):S231-S240.

10. Yang X, Li Y, Zhang C, et al. Clinical features and phylogenetic analysis of severe hand-foot-and-mouth disease caused by Coxsackievirus A6. *Infect Genet Evol* 2020; 77:104054.