Nucleic Acids Research, 2024, **52**, D770–D776
https://doi.org/10.1093/nar/gkad937
Advance access publication date: 1 November 2023
Database issue

RVdb: a comprehensive resource and analysis platform for rhinovirus research

Peng Zhao[1,†], Siyu Zhou[1,†], Panpan Xu[1,†], Haoxiang Su[1], Yelin Han[1], Jie Dong[1], Hongtao Sui[1], Xin Li[1], Yongfeng Hu[1], Zhiqiang Wu[1], Bo Liu[1,*], Ting Zhang[1,*] and Fan Yang[1,2,3,*]

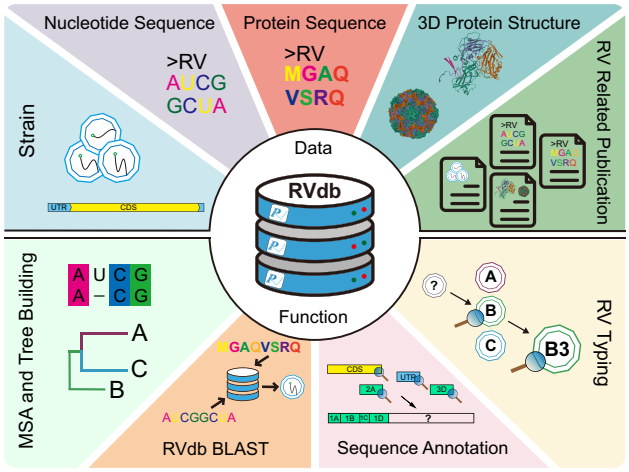
[1]NHC Key Laboratory of Systems Biology of Pathogens, Institute of Pathogen Biology, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 102629, P.R. China
[2]Key Laboratory of Respiratory Disease Pathogenomics, Chinese Academy of Medical Sciences, Beijing 102629, P.R. China
[3]State Key Laboratory of Respiratory Health and Multimorbidity, Beijing 102629, P.R. China

[*]To whom correspondence should be addressed. Tel: +86 10 67875146; Email: obuil@ipbcams.ac.cn
Correspondence may also be addressed to Ting Zhang. Email: zhangting@ipbcams.ac.cn
Correspondence may also be addressed to Fan Yang. Email: ymf129@163.com
[†]The authors wish it to be known that, in their opinion, the first three authors should be regarded as Joint First Authors.

Abstract

Rhinovirus (RV), a prominent causative agent of both upper and lower respiratory diseases, ranks among the most prevalent human respiratory viruses. RV infections are associated with various illnesses, including colds, asthma exacerbations, croup and pneumonia, imposing significant and extended societal burdens. Characterized by a high mutation rate and genomic diversity, RV displays a diverse serological landscape, encompassing a total of 174 serotypes identified to date. Understanding RV genetic diversity is crucial for epidemiological surveillance and investigation of respiratory diseases. This study introduces a comprehensive and high-quality RV data resource, designated RVdb (http://rvdb.mgc.ac.cn), covering 26 909 currently identified RV strains, along with RV-related sequences, 3D protein structures and publications. Furthermore, this resource features a suite of web-based utilities optimized for easy browsing and searching, as well as automatic sequence annotation, multiple sequence alignment (MSA), phylogenetic tree construction, RVdb BLAST and a serotyping pipeline. Equipped with a user-friendly interface and integrated online bioinformatics tools, RVdb provides a convenient and powerful platform on which to analyse the genetic characteristics of RVs. Additionally, RVdb also supports the efforts of virologists and epidemiologists to monitor and trace both existing and emerging RV-related infectious conditions in a public health context.

Graphical abstract



Introduction

Rhinovirus (RV), which belongs to the family Picornaviridae and the genus *Enterovirus*, is a positive-sense, single-stranded RNA (ssRNA) virus with a genome size of approximately 7200 bp. It is classified into three species: RVA, RVB and RVC. The viral genome contains a single coding sequence (CDS) that is translated into several proteins through cleavage by virally encoded proteases. These proteins include four capsid proteins and seven non-structural proteins. The RV capsid is formed by the assembly of 60 copies of each of the four

Received: August 15, 2023. **Revised:** October 8, 2023. **Editorial Decision:** October 9, 2023. **Accepted:** October 10, 2023

capsid proteins (1). RV, the most frequently detected respiratory virus, is responsible for one-half to two-thirds of common colds and one-half of the exacerbations of asthma (2–5). It has been reported that RV infection can also cause bronchiolitis, croup, community-acquired pneumonia, chronic obstructive pulmonary disease, cystic fibrosis, aggravation of chronic lung diseases and other lower respiratory tract infections (6–12). RV imposes a significant economic burden, encompassing costs associated with medical treatment, educational disruptions, work absenteeism and hindrances to social progress (13–16).

The emergence of the coronavirus disease 2019 (COVID-19) pandemic, caused by severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2), illuminated the potent dangers associated with respiratory RNA viruses (17). Notably, RV exhibits resemblances to SARS-CoV-2 in aspects such as its extensive presence and prevalence within the human population along with high mutation rates. Since its discovery in the 1950s (18), RV has been reported in 58 countries according to RVdb. In the United States, children experience an average of two RV infections per year, while adults typically have one. Furthermore, RVs display a marked level of serotypic diversity. At present, there are 174 identified serotypes, 81 types of species RVA, 33 types of species RVB and 60 types of species RVC. The existence of serotype diversity imposes intricate challenges, from precise typing information and typing annotation to the development of cross-protective therapeutics and vaccines (19,20). This confluence of difficulties provides a compelling rationale for directing attention towards RVs as a subject of significant consideration.

Given the extensive utilization of molecular methodologies in the field of virus identification and functional analyses, the comprehensive retrieval of viral sequences, coupled with associated annotation/meta-information, assumes a pivotal role in facilitating in-depth investigations into the genetic diversity of RVs. However, through our examination of public data, we have observed several issues in RV-related data, including a considerable number of sequences, insufficient meta-information, absence of genomic annotations and inaccuracies or incompleteness in the categorization of species and types, cumulatively constituting formidable barriers that hinder comprehension of the characteristics of RVs, similar to our previous studies on the curation of zoonotic and vector-borne viruses (21). Furthermore, the establishment of interconnections with other important public data resources, such as PubMed, UniProt and PDB, remains notably restricted and inconvenient. Evidently, the challenges mentioned above collectively contribute to the intricate challenge faced by users seeking targeted insights pertaining to known RVs. In this study, we introduce a specialized RV database, designated RVdb (accessible at http://rvdb.mgc.ac.cn). With parallels to the Influenza Research Database (22) for influenza virus and HBVdb (23) for hepatitis B virus, RVdb focuses on the collection and curation of up-to-date knowledge on currently identified RVs. The well-integrated set of data types and analysis tools available on the RVdb platform are actively employed to enhance the comprehension of existing RV populations. Concurrently, these resources are anticipated to assist in the development of vaccines, therapeutics and diagnostics, mitigating the public health impact of respiratory diseases associated with RVs.

Data preparation and database implementation

RVdb, as a sequence-centric database, is dedicated to providing up-to-date and comprehensive information associated with RVs. To retrieve all available sequences from the public domain, we conducted exhaustive searches in the PubMed, Taxonomy and Nucleotide databases (24) of the National Center for Biotechnology Information (NCBI) using the keywords ('Rhinovirus' OR 'RV' OR 'HRV') AND ('virus' OR 'viruses'). The associated GenBank records pertaining to RV within the aforementioned three databases were then downloaded and parsed via a Biopython script to generate human-read tables of the meta-information. Of the original GenBank records inspected, a substantial number of viral sequences, accompanied by preliminary metadata, have been deposited into the GenBank database by submitters. Therein, >83% of the sequences lack necessary information regarding specific sampling locations/times, while > 53% of the sequences are not linked to relevant literature (if it exists). To address the issue of inaccuracies in submitted GenBank records, comprehensive searches were first conducted on mainstream search engines such as Google and Bing for GenBank records that lacked direct links to pertinent literature (if available). These efforts enabled us to establish connections between the sequences and the related literature. Then, we performed meticulous cross-referencing of published records, proofreading for any inconsistencies or gaps in meta-information by carefully examining the details described in the related literature. All these curation processes were undertaken to assure dataset accuracy and comprehensiveness, thereby promoting subsequent analysis and interpretation related to RV. Furthermore, additional data types (e.g. protein sequences, 3D protein structures and RV-related publications) were further collected from GenPept, PDB (25) and PubMed. While molecular methods are predominantly employed for diagnosis and functional studies of viruses, numerous findings have relied on alternative approaches without sequencing (e.g. immunological methods). In addition to the sequence-related publications, our database includes >20 000 sequence-independent publications pertinent to RV research, thereby reinforcing associations between diverse data types and bolstering the comprehensiveness of the RVdb dataset. To facilitate users in efficiently and precisely retrieving individual publications, Medical Subject Headings (MeSH) (26,27) are integrated into the publication retrieval and browsing features of RVdb (Figure 1A).

The aforementioned data entries were refined with a controlled and consistent vocabulary and stored in a well-structured MySQL database. The web service layer was executed by a set of Python scripts, employing the Flask framework (http://flask.pocoo.org/). To augment the user experience and interactivity within the interface, we incorporated a suite of web programming libraries, utilizing the Bootstrap framework (https://getbootstrap.com/), the cross-platform JavaScript library jQuery (http://jquery.com), the interactive charting library HighCharts (https://highcharts.com) and DataTables, a comprehensive JavaScript library equipped with a variety of plugins for grid data manipulation (https://www.datatables.net). Additionally, we also integrated a set of bioinformatic tools to facilitate online data analysis and visualization, such as an efficient JavaScript component for the visualization and analysis of multiple sequence alignment (MSA) data known as MSAViewer (28), a

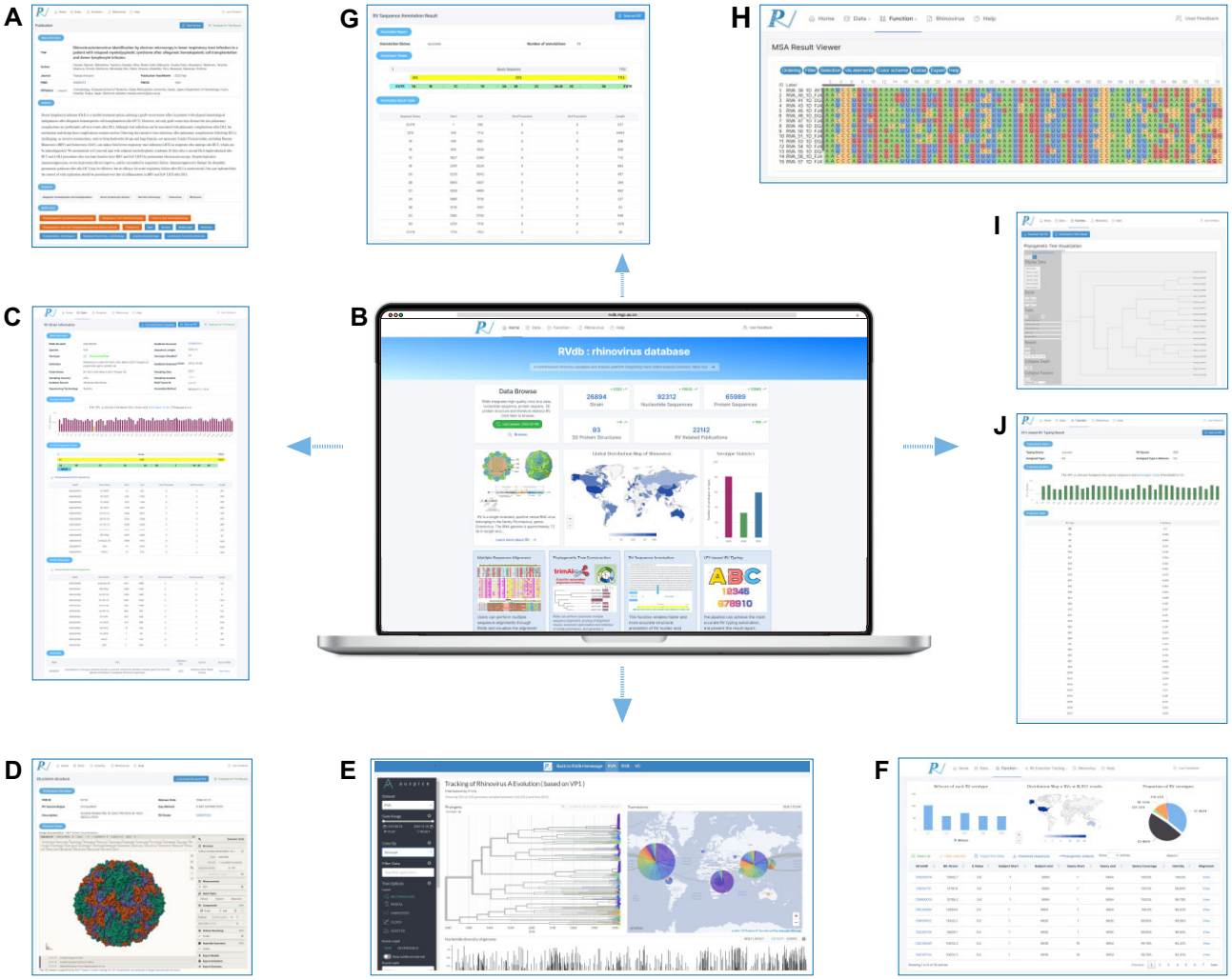


Figure 1. The web interface of RVdb. (**A**) A query example for RV-related publications. (**B**) A visual overview of the RVdb utilities. (**C**) An example of the detailed information of an RV strain. (**D**) The panel displaying 3D protein structures. (**E**) The panel of tracking of RVA evolution. (**F**) The BLAST result panel. (**G**) The report of RV sequence annotation results. (**H**) The panel for MSA result visualization. (**I**) The result visualization panel for phylogenetic tree construction. (**J**) The report showcasing RV typing results.

cutting-edge web application for 3D visualization and analysis of substantial biomolecular structures known as Mol* Viewer (29), a reusable and open-source sequence feature viewer known as PDB ProtVista (https://github.com/PDBeurope/protvista-pdb), a JavaScript library for the visualization of phylogenetic trees known as Archaeopteryx.js (https://github.com/cmzmasek/archaeopteryx-js) and an open-source project for a real-time view of the evolution and spread of RV known as Nextstrain (30).

Database content and features

The RVdb datasets consists of five datasets: serotyping library, virus strain/sample meta-information, the available nucleotide/protein sequences of each virus, 3D protein structure and RV-related publications. Based on these datasets, RVdb offers a highly intuitive and responsive web platform (http://rvdb.mgc.ac.cn) for users to browse and search the up-to-date aforementioned information about presently identified RVs (Figure 1B). Based on the genomic characterization and genetic diversity of RV, we further developed an automatic pipeline for MSA, phylogenetic tree construction and sequence annotation, integrated Basic Local Alignment Search Tool (BLAST) search and VP1- and VP4/2-based RV typing for the identification and characterization of novel RVs or associated variants.

Browse and search the database

The primary utilities of the database, including datasets browsing and searching, real-time statistics, and a detailed help document, are accessible on the homepage by clickable icons, each providing direct access to the respective main page. Within the browsing and searching section, RVdb initially presents an introduction page designed to help users retrieve fundamental information about RV, encompassing genomic organization and genetic diversity. Users can further browse and search curated meta-information pertaining to viruses, sequences, and samples. This includes details, such as strain, species, sequence length and completeness, sampling data and location, and related literature (if available) (Figure 1C). RVdb also offers online browsing of RV 3D protein structures, along with a 1D protein feature viewer module that allows users to visualize annotation information

for structural proteins (Figure 1D). The 3D protein structure data are cross-referenced to the serotyping information by integrating the annotated strain data. To provide an intuitive overview of the geographic trends in the available data, a global map is provided with colour-coded markers denoting the presence of known RVs (Figure 1B). Each section corresponding to a specific country provides a direct link to a panel showcasing detailed information regarding RVs identified within that country. This geographical context is invaluable to epidemiologists and public health researchers, facilitating the monitoring and source tracking of current respiratory diseases caused by RV. Furthermore, RVdb employs Nextstrain, a sophisticated phylogenetic visualization toolset, to describe of the spread and evolution patterns of RVA, RVB and RVC, respectively. This tool enables simultaneous exploration of phylogenetic and spatial/temporal relationships, with different serotypes/countries represented through distinct colours (Figure 1E). Additionally, RVdb incorporates the standalone NCBI BLAST tool (31), enabling users to perform sequence similarity searches against all available sequences. This integration encompasses a set of BLAST programs, including blastn, blastp, blastx, tblastn and tblastx (Figure 2). Users have the flexibility to select the BLAST program most suitable for their specific analytical requirements. Utilizing a set of in-house Python scripts, a concise and legible profile, with statistics for serotype and geographic distribution, is presented within the BLAST result tab (Figure 1F). For comprehensive and detailed guidance on the utilization of the RVdb database, a step-by-step instructional resource is available in the accompanying help document.

Analysis utilities

Genomic annotation, MSA and phylogenetic tree construction play a pivotal role in enabling users to investigate the genomic structure of viruses and to conduct in-depth viral relationship analyses. Genomic annotation constitutes tasks that must be completed prior to submitting sequences to public databases. Therefore, RVdb offers an automatic sequence annotation pipeline based on well-annotated prototype strains (Figures 1G and 2). In addition to CDS, annotations for untranslated regions (UTRs) and mature peptides, such as 1A, 1B, 1C, 1D, 2A, 2B, 2C, 3A, 3B, 3C and 3D, are also provided if present. Users can submit a complete or draft RV genome sequence to obtain detailed annotation reports. Furthermore, given that some original GenBank records do not provide sufficient sequence annotations, we utilized this pipeline to re-annotate all available strains within RVdb. With thorough validation, this pipeline demonstrated significant increases in the number of nucleotide sequence annotations from 32 727 to 92 312 and protein sequence annotations from 25681 to 65989. To provide improved visualization of the genomic organization, we also developed an interactive and user-friendly sequence viewer specifically designed for RV (Figure 1C and G) that behaves similarly to seqviz and NCBI Sequence Viewer (32). Second, MSA forms an essential first step in various sequence analysis methodologies, such as the investigation of phylogenetic relationships. Accordingly, RVdb provides an automatic MSA module, allowing users to upload and calculate private data and supporting online visualization of the results (Figures 1H and 2). Based on the the phylogenetic pipeline, RVdb can auto-selects the best model and infer the phylogenetic tree via maximum likelihood. It also offers online visualization of the generated phylogenetic tree, with options to download associated files and export images (Figures 1I and 2). Automated phylogenetic tree construction utilizes MAFFT for MSA (33,34), trimAl software for the automatic removal of spurious sequences or poorly aligned regions (35), and IQ-TREE2 for the construction of phylogenetic trees (36).

Serotyping of RV is significant for clinical identification, precise traceability and effective prevention and control measures. Using our up-to-date prototype strain dataset, we also build an automatic RV typing pipeline based on calculating the *p*-distance of VP1 between an untyped strain and prototype strains. Users need only submit the VP1 sequence to RVdb, after which the pipeline determines the RV species, aligns it with the corresponding prototype strains, calculates the *p*-distance and generates a detailed typing report (Figures 1J and 2). Because of the revision of serotyping standards, incorrect or incomplete typing information is available in the original GenBank records and publications. To address this issue, RVdb systematically performs serotype validation for all strains with VP1 sequences available over 90% integrity. It further provides intuitive visualization to present validation specifics on individual strain information page (Figure 1C). Therein, the typing information for 716 RV strains, which were previously untyped, has been added to the dataset. Typing information for a total of 1847 RV strains have been verified as correct, while 337 errors have been rectified. Furthermore, this methodology has contributed to the identification of four previously unreported serotypes, B107, C58, C59 and C60, which are detailed in our previous work (37).

Furthermore, the VP4/2 segment of RV, which is approximately 435 nucleotides in length, is well conserved and easily amplified. This region can serve as a predictive marker for discerning RV species and types, offering an alternative means of type identification in epidemiological studies when VP1 sequences are either absent or display low completeness (38). Consequently, RVdb also includes a typing module based on VP4/2 (Figure 2). This module is designed to fulfil the requisites for RV type prediction or identification in epidemiological studies, particularly in scenarios where VP1 sequences exhibit suboptimal completeness. Since this region is not classified as an official mature peptide, it lacks a corresponding annotation in GenBank. In contrast, RVdb includes comprehensive annotations specific to this region, providing an optional solution for types and species identification. Due to certain typing limitations for the VP4/2 region, including its shorter length, higher sequence conservation and possible divergence from VP1-based typing outcomes, we refrained from conducting VP4/2-based serotype verification for RV strains in the database.

Discussion and future directions

RVdb serves as an advanced and versatile platform, centralizing data resources crucial for current RV research. As of August 2023, a total of 26909 RV strains along with their respective sequences, meta-information, 3D protein structures and publications had been added. Within this framework, a set of analytical utilities, tailored exclusively for RV investigation, have been integrated. These utilities are designed to deliver essential data and functional support across multiple domains: RV-related identification, classification, characterization and epidemiological surveillance.

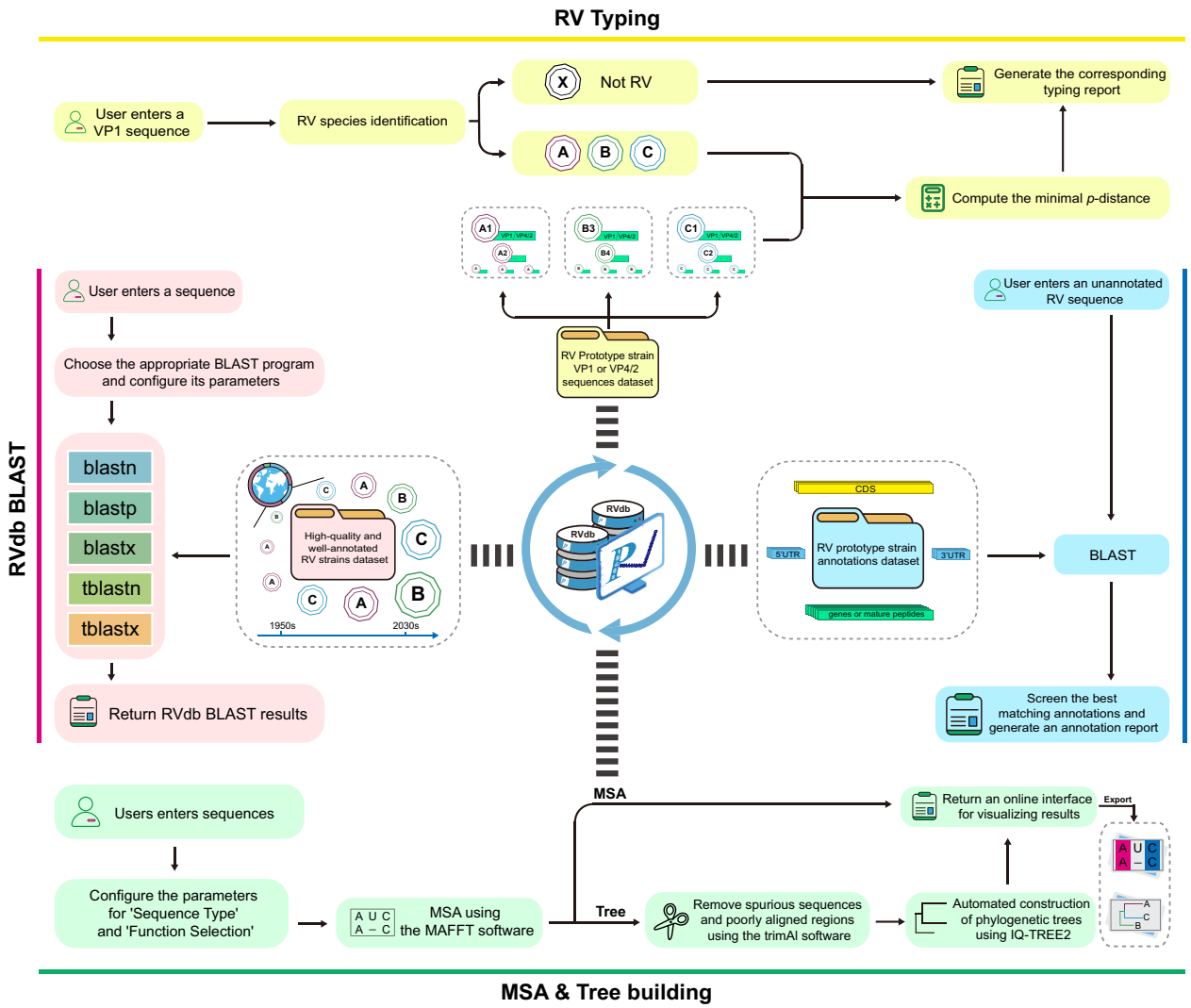


Figure 2. Flowchart outlining the online analysis functions of RVdb. Sub-flowcharts correspond to RV typing coloured in yellow (top), RV sequence annotation coloured in blue (right), MSA and tree building coloured in green (bottom), and RVdb BLAST coloured in pink (left).

Despite our rigorous efforts in data collection and validation, it is imperative to acknowledge several potential limitations. First, the inclusion of all presently identified RV records cannot be guaranteed through only the application of retrieval keywords. In specific cases, the possibility of overlooked RV strains is inevitable, owing to misclassification by the submitters or historical alterations in taxonomy. Consequently, we earnestly appreciate and encourage users to communicate with us regarding any sequences or literature that may have inadvertently evaded inclusion in the RVdb database. Second, RV-related meta-information, including the publication, benefits our understanding of viral characteristics and acts as a key foundation for linkage with other datasets. Nevertheless, it is noteworthy that >25% of the RV strains documented in the database remain unpublished, lacking associated literature. Given this situation, we will continue trying to determine whether there is relevant literature available to address possible inconsistencies or gaps in the follow-up study. Third, the molecular typing standard for RVs relies on the genetic distances within the VP1 or VP4/2 region. This criterion, however, fails to maintain absolute consistency with the cross-reactivity observed in serological experiments, exemplified by cases such as RVA29 and RVA44 (39). Consequently, we recommend users exercise careful consideration when interpreting biological implications, given that computer-generated results invariably necessitate subsequent experimental validation for conclusive confirmation. Additionally, the necessity for sequence integrity and specific gene fragments complicates the serotyping of RV strains. As more RVs are identified and published, the focus will shift towards establishing more precise and efficient RV molecular typing standards and methodologies.

Given the widespread presence of RV in the human population, it is plausible that numerous undiscovered novel strains or serotypes may exist within publicly available high-throughput sequencing (HTS) datasets. Future endeavours will include the development and implementation of a specialized RV identification pipeline tailored for HTS metadata analysis. This advancement will enable RVdb to thoroughly scan clinical metadata in the public domain while identifying new RV strains or serotypes, further aiding in the understanding of the prevalence and coexistence of RVs and thereby providing valuable insights into the impact of different RV strains or serotypes on human health. Addition-

ally, various machine/deep-learning-based methods are now used in the classification of genes or pathogens. Compared to homology-based identification, alignment-independent deep-learning-based prediction could leverage large auxiliary data to provide new avenues for biological research on the classification of RV strains or serotypes (40). Therefore, while providing in-depth analysis methodologies, we will continue to update existing datasets every two months to provide users with the most accurate information. Meanwhile, our research subject will be gradually expanded to provide genomic and clinical datasets of other significant human-related respiratory viruses, establishing a comprehensive global surveillance network for respiratory viruses for efficient prevention and control of future emerging respiratory diseases.

Data availability

RVdb, accessible at http://rvdb.mgc.ac.cn, is freely available to all users without any login requirements.

Funding

National Key Plan for Scientific Research and Development of China [2021YFC2302003 to T.Z.]; CAMS Innovation Fund for Medical Sciences [2021-I2M-1-039 to F.Y.]; CAMS Innovation Fund for Medical Sciences [2021-I2M-1-038 to Y.H.]. Funding for open access charge: the National Key Plan for Scientific Research and Development of China.

Conflict of interest statement

None declared.

References

1. Palmenberg,A.C., Rathe,J.A. and Liggett,S.B. (2010) Analysis of the complete genome sequences of human rhinovirus. *J. Allergy Clin. Immunol.*, **125**, 1190–1199.
2. Gwaltney,J.M. Jr, Hendley,J.O., Simon,G. and Jordan,W.S. Jr (1966) Rhinovirus infections in an industrial population. I. The occurrence of illness. *N. Engl. J. Med.*, **275**, 1261–1268.
3. Johnston,S.L., Pattemore,P.K., Sanderson,G., Smith,S., Campbell,M.J., Josephs,L.K., Cunningham,A., Robinson,B.S., Myint,S.H., Ward,M.E., *et al.* (1996) The relationship between upper respiratory infections and hospital admissions for asthma: a time-trend analysis. *Am. J. Respir. Crit. Care Med.*, **154**, 654–660.
4. Nicholson,K.G., Kent,J. and Ireland,D.C. (1993) Respiratory viruses and exacerbations of asthma in adults. *BMJ*, **307**, 982–986.
5. Rankin,D.A., Spieker,A.J., Perez,A., Stahl,A.L., Rahman,H.K., Stewart,L.S., Schuster,J.E., Lively,J.Y., Haddadin,Z., Probst,V., *et al.* (2023) Circulation of rhinoviruses and/or enteroviruses in pediatric patients with acute respiratory illness before and during the COVID-19 Pandemic in the US. *JAMA Netw. Open*, **6**, e2254909.
6. Jacobs,S.E., Lamson,D.M., St George,K. and Walsh,T.J. (2013) Human rhinoviruses. *Clin. Microbiol. Rev.*, **26**, 135–162.
7. Henquell,C., Mirand,A., Deusebis,A.L., Regagnon,C., Archimbaud,C., Chambon,M., Bailly,J.L., Gourdon,F., Hermet,E., Dauphin,J.B., *et al.* (2012) Prospective genotyping of human rhinoviruses in children and adults during the winter of 2009-2010. *J. Clin. Virol.*, **53**, 280–284.
8. Xiang,Z., Gonzalez,R., Xie,Z., Xiao,Y., Liu,J., Chen,L., Liu,C., Zhang,J., Ren,L., Vernet,G., *et al.* (2010) Human rhinovirus C infections mirror those of human rhinovirus A in children with community-acquired pneumonia. *J. Clin. Virol.*, **49**, 94–99.
9. Ritchie,A.I., Farne,H.A., Singanayagam,A., Jackson,D.J., Mallia,P. and Johnston,S.L. (2015) Pathogenesis of viral infection in exacerbations of airway disease. *Ann. Am Thorac. Soc.*, **12**(Suppl. 2), S115–S132.
10. Jackson,D.J., Gangnon,R.E., Evans,M.D., Roberg,K.A., Anderson,E.L., Pappas,T.E., Printz,M.C., Lee,W.M., Shult,P.A., Reisdorf,E., *et al.* (2008) Wheezing rhinovirus illnesses in early life predict asthma development in high-risk children. *Am. J. Respir. Crit. Care Med.*, **178**, 667–672.
11. Kieninger,E., Singer,F., Tapparel,C., Alves,M.P., Latzin,P., Tan,H.L., Bossley,C., Casaulta,C., Bush,A., Davies,J.C., *et al.* (2013) High rhinovirus burden in lower airways of children with cystic fibrosis. *Chest*, **143**, 782–790.
12. McManus,T.E., Marley,A.M., Baxter,N., Christie,S.N., O'Neill,H.J., Elborn,J.S., Coyle,P.V. and Kidney,J.C. (2008) Respiratory viral infection in exacerbations of COPD. *Respir. Med.*, **102**, 1575–1580.
13. Fendrick,A.M., Monto,A.S., Nightengale,B. and Sarnes,M. (2003) The economic burden of non-influenza-related viral respiratory tract infection in the United States. *Arch. Intern. Med.*, **163**, 487–494.
14. Bertino,J.S. (2002) Cost burden of viral respiratory infections: issues for formulary decision makers. *Am. J. Med.*, **112**(Suppl. 6A), 42S–49S.
15. Nichol,K.L., D'Heilly,S. and Ehlinger,E. (2005) Colds and influenza-like illnesses in university students: impact on health, academic and work performance, and health care use. *Clin. Infect. Dis.*, **40**, 1263–1270.
16. Roelen,C.A., Koopmans,P.C., Notenbomer,A. and Groothoff,J.W. (2011) Job satisfaction and short sickness absence due to the common cold. *Work*, **39**, 305–313.
17. Huang,C., Wang,Y., Li,X., Ren,L., Zhao,J., Hu,Y., Zhang,L., Fan,G., Xu,J., Gu,X., *et al.* (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*, **395**, 497–506.
18. Price,W.H. (1956) The isolation of a new virus associated with Respiratory clinical disease in humans. *Proc. Natl. Acad. Sci. U.S.A.*, **42**, 892–896.
19. Williams,G.R., Kubajewska,I., Glanville,N., Johnston,S.L. and McLean,G.R. (2016) The potential for a protective vaccine for rhinovirus infections. *Expert Rev. Vaccines*, **15**, 569–571.
20. McLean,G.R. (2020) Vaccine strategies to induce broadly protective immunity to rhinoviruses. *Hum. Vaccin. Immunother.*, **16**, 684–686.
21. Zhou,S., Liu,B., Han,Y., Wang,Y., Chen,L., Wu,Z. and Yang,J. (2022) ZOVER: the database of zoonotic and vector-borne viruses. *Nucleic Acids Res.*, **50**, D943–D949.
22. Zhang,Y., Aevermann,B.D., Anderson,T.K., Burke,D.F., Dauphin,G., Gu,Z., He,S., Kumar,S., Larsen,C.N., Lee,A.J., *et al.* (2017) Influenza Research Database: an integrated bioinformatics resource for influenza virus research. *Nucleic Acids Res.*, **45**, D466–D474.
23. Hayer,J., Jadeau,F., Deleage,G., Kay,A., Zoulim,F. and Combet,C. (2013) HBVdb: a knowledge database for Hepatitis B Virus. *Nucleic Acids Res.*, **41**, D566–D570.
24. Sayers,E.W., Cavanaugh,M., Clark,K., Ostell,J., Pruitt,K.D. and Karsch-Mizrachi,I. (2020) GenBank. *Nucleic Acids Res.*, **48**, D84–D86.
25. Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.
26. Dhammi,I.K. and Kumar,S. (2014) Medical subject headings (MeSH) terms. *Indian J. Orthop.*, **48**, 443–444.
27. Leydesdorff,L., Comins,J.A., Sorensen,A.A., Bornmann,L. and Hellsten,I. (2016) Cited references and Medical Subject Headings (MeSH) as two different knowledge representations: clustering and mappings at the paper level. *Scientometrics*, **109**, 2077–2091.
28. Yachdav,G., Wilzbach,S., Rauscher,B., Sheridan,R., Sillitoe,I., Procter,J., Lewis,S.E., Rost,B. and Goldberg,T. (2016) MSAViewer:

interactive JavaScript visualization of multiple sequence alignments. *Bioinformatics*, **32**, 3501–3503.
29. Sehnal,D., Bittrich,S., Deshpande,M., Svobodova,R., Berka,K., Bazgier,V., Velankar,S., Burley,S.K., Koca,J. and Rose,A.S. (2021) Mol* viewer: modern web app for 3D visualization and analysis of large biomolecular structures. *Nucleic Acids Res*., **49**, W431–W437.
30. Hadfield,J., Megill,C., Bell,S.M., Huddleston,J., Potter,B., Callender,C., Sagulenko,P., Bedford,T. and Neher,R.A. (2018) Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics*, **34**, 4121–4123.
31. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*., **25**, 3389–3402.
32. Rangwala,S.H., Kuznetsov,A., Ananiev,V., Asztalos,A., Borodin,E., Evgeniev,V., Joukov,V., Lotov,V., Pannu,R., Rudnev,D., *et al.* (2021) Accessing NCBI data using the NCBI sequence viewer and genome data viewer (GDV). *Genome Res*., **31**, 159–169.
33. Katoh,K., Misawa,K., Kuma,K. and Miyata,T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast fourier transform. *Nucleic Acids Res*., **30**, 3059–3066.
34. Rozewicki,J., Li,S., Amada,K.M., Standley,D.M. and Katoh,K. (2019) MAFFT-DASH: integrated protein sequence and structural alignment. *Nucleic Acids Res*., **47**, W5–W10.
35. Capella-Gutierrez,S., Silla-Martinez,J.M. and Gabaldon,T. (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, **25**, 1972–1973.
36. Minh,B.Q., Schmidt,H.A., Chernomor,O., Schrempf,D., Woodhams,M.D., von Haeseler,A. and Lanfear,R. (2020) IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol*., **37**, 1530–1534.
37. Zhao,P., Shao,N., Dong,J., Su,H., Sui,H., Zhang,T. and Yang,F. (2023) Genetic diversity and characterization of rhinoviruses from Chinese clinical samples with a global perspective. *Microbiol. Spectr.*, https://doi.org/10.1128/spectrum.00840-23.
38. McIntyre,C.L., Knowles,N.J. and Simmonds,P. (2013) Proposals for the classification of human rhinovirus species A, B and C into genotypically assigned types. *J. Gen. Virol*., **94**, 1791–1806.
39. Cooney,M.K., Fox,J.P. and Kenny,G.E. (1982) Antigenic groupings of 90 rhinovirus serotypes. *Infect. Immun*., **37**, 642–647.
40. Zheng,D., Pang,G., Liu,B., Chen,L. and Yang,J. (2020) Learning transferable deep convolutional neural networks for the classification of bacterial virulence factors. *Bioinformatics*, **36**, 3693–3702.

Received: August 15, 2023. **Revised:** October 8, 2023. **Editorial Decision:** October 9, 2023. **Accepted:** October 10, 2023