

## Article

# IDV Typer: An Automated Tool for Lineage Typing of Influenza D Viruses Based on Return Time Distribution

Sanket Limaye <sup>1,†</sup> , Anant Shelke <sup>1,†</sup>, Mohan M. Kale <sup>2</sup>, Urmila Kulkarni-Kale <sup>1,\*,‡,§</sup>   
and Suresh V. Kuchipudi <sup>3,\*</sup> 

<sup>1</sup> Bioinformatics Centre, Savitribai Phule Pune University (Formerly University of Pune), Pune 411007, India; sanket.limaye22@gmail.com (S.L.); anantshelkecell@gmail.com (A.S.)

<sup>2</sup> Department of Statistics, Savitribai Phule Pune University (Formerly University of Pune), Pune 411007, India; mohan.kale80@gmail.com

<sup>3</sup> Department of Infectious Diseases and Microbiology, University of Pittsburgh School of Public Health, Pittsburgh, PA 15261, USA

\* Correspondence: urmila.kulkarni.kale@gmail.com (U.K.-K.); skuchipudi@pitt.edu (S.V.K.)

† These authors contributed equally to this work.

‡ Current address: Department of Natural Sciences and Environmental Health, University of Southeastern Norway, Gullbringvegen 36, 3800 Bø, Norway.

§ Current address: CIS-Citadel Precision Medicine, Hyderabad 500084, India.

**Abstract:** Influenza D virus (IDV) is the most recent addition to the *Orthomyxoviridae* family and cattle serve as the primary reservoir. IDV has been implicated in Bovine Respiratory Disease Complex (BRDC), and there is serological evidence of human infection of IDV. Evolutionary changes in the IDV genome have resulted in the expansion of genetic diversity and the emergence of multiple lineages that might expand the host tropism and potentially increase the pathogenicity to animals and humans. Therefore, there is an urgent need for automated, accurate and rapid typing tools for IDV lineage typing. Currently, IDV lineage typing is carried out using BLAST-based searches and alignment-based molecular phylogeny of the hemagglutinin-esterase fusion (HEF) gene sequences, and lineage is assigned to query sequences based on sequence similarity (BLAST search) and proximity to the reference lineages in the tree topology, respectively. To minimize human intervention and lineage typing time, we developed IDV Typer server, implementing alignment-free method based on return time distribution (RTD) of k-mers. Lineages are assigned using HEF gene sequences. The server performs with 100% sensitivity and specificity. The IDV Typer server is the first application of an RTD-based alignment-free method for typing animal viruses.

**Keywords:** Influenza D virus; lineage typing; IDV Typer server; return time distribution



**Citation:** Limaye, S.; Shelke, A.; Kale, M.M.; Kulkarni-Kale, U.; Kuchipudi, S.V. IDV Typer: An Automated Tool for Lineage Typing of Influenza D Viruses Based on Return Time Distribution. *Viruses* **2024**, *16*, 373. <https://doi.org/10.3390/v16030373>

Academic Editors: Feng Li and Leyi Wang

Received: 30 December 2023

Revised: 24 February 2024

Accepted: 24 February 2024

Published: 28 February 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Influenza D virus (IDV) is the newest member of the *Orthomyxoviridae* family, classified under the *Deltainfluenzavirus* genus [1]. The virus, like other Orthomyxoviruses, is a negative stranded, RNA virus with a segmented genome. IDV was first isolated in 2011 from pigs in Oklahoma [2]. Since then, IDV has been isolated/detected from cattle across many countries, including the United States (US), Mexico [3], Ireland [4], United Kingdom [5], France [6], Italy [7], China [8], Canada [9], Japan [10], Turkey [11], Netherlands [12], Brazil [13], South Korea [14], Australia [15], Argentina [16], Namibia [17] and Denmark [18]. Serological evidence also indicates IDV presence in Morocco, Togo, Benin, Kenya [19], Luxembourg [20] and Ethiopia [21]. While cattle are considered the main reservoir, serological evidence of IDV infection was found in several animals, including pigs, sheep, goats, camels, giraffe, wildebeest, white-tailed deer and horses [3,6–8,17,19,22–27]. IDV's pervasive presence among US cattle is strikingly evident. IDV antibodies also have been detected in US cattle sera from 2003 to 2004 [28]. Further, a study conducted between 2014 and 2015 found IDV seropositive cattle samples in 41 out of 42 states nationwide. An

overall seropositivity rate of 77.5% was recorded among tested cattle sera samples. Notably, this prevalence exhibited considerable regional variation, ranging from 47.7% to 84.6% in different parts of the country. These findings underscore the widespread prevalence of IDV within US cattle populations [29].

Many viral respiratory pathogens have been implicated in Bovine Respiratory Disease (BRD) including Bovine Viral Diarrhea Virus (BVDV), Bovine Herpesvirus type 1 (BHV-1), Bovine Respiratory Syncytial Virus (BRSV), Bovine Parainfluenza Virus 3 (BPIV-3) and Bovine Coronavirus (BCoV) [30]. Other viruses such as bovine adenovirus (BAdV), bovine rhinitis A virus and bovine adeno-associated virus have been found to be associated with BRD less frequently [31,32]. In addition to these existing pathogens, a growing body of evidence suggests a correlation between IDV in the pathogenesis of BRD [3,33–35]. Notably, D/CA2019 lineage, the most recently identified IDV lineage in the US, was isolated from calves with respiratory disease [36]. The presence of serological evidence [37,38] and the detection of IDV nucleic acid in human nasal samples [39,40] collectively constitute compelling proof of IDV infection in humans, underscoring its zoonotic potential.

The IDV genome is segmented, and the seven segments consist of three polymerases, i.e., Polymerase basic protein 2 (PB2), Polymerase basic protein 1 (PB1) and Polymerase protein (PA/P3); Hemagglutinin-Esterase Fusion (HEF) protein, Nucleoprotein (N), Matrix protein (M); and the Non-structural protein (NS). The HEF protein is primarily involved in receptor binding and is the target for neutralizing antibodies [2]. HEF is a type 1 transmembrane protein containing one long ectodomain, a transmembrane region, one small cleavable N-terminal signal peptide and a short cytoplasmic tail. Sialic acid (SA) serves as the main receptor for IDV HEF. IDV utilizes either 9-*O*-acetylated Neu5Ac (Neu5,9Ac<sub>2</sub>) and Neu5Gc (Neu5Gc9Ac), independent of the  $\alpha$ 2,3 or  $\alpha$ 2,6 linkage, as a receptor for entering the cell [41].

IDV is currently classified into five major lineages, namely, D/OK, D/660, D/Yama2016, D/Yama2019 and D/CA2019 based on the nucleotide sequence divergence of the HEF gene [36,42–44]. The D/OK and D/660 lineage isolated from the USA and Italy are reported to undergo reassortment events with each other [42,45]. The two lineages, D/Yama2016 and D/Yama2019 were reported from Japan [43,44] and recently the D/Yama2019 lineage was reported from China, South Korea and Australia [14,15,46]. Even though IDV was only discovered a decade ago, novel genetic clusters and entire lineages are being discovered [44] including multiple and frequent reassortments [42,47]. In Hokkaido, Japan, the viruses isolated from two out of three BRD outbreaks in cattle during 2018–2020 belonged to a new lineage, different from the lineages of Japanese IDVs based on HEF, PB2, P3, NP and NS genes [48]. The viruses in the new lineage were isolated from cattle with severe disease including rapid spread in the herd, debilitation and death due to pneumonia. This study found coinfection of IDV with other bacterial and viral pathogens, highlighting the role of IDV as one of the causative pathogens of BRD. Further, the IDVs of the new lineage had poor cross-neutralization with sera against heterogenous IDV, indicating the circulation of genetically and antigenically distinct IDV lineages in Japan, D/Yama2016 and D/Yama2019. Some IDV isolates from Japan are genetically distinct from IDV from the rest of the world [10].

A new US antigenic lineage was discovered when the first IDV sequences from California were sequenced [36]. These IDVs form a new clade named D/CA2019 in which the P3 gene groups in the D/OK lineage and the other genes cluster with the D/660 lineage. Moreover, sera raised to D/OK or D/660 had poor neutralizing titers against D/CA2019 [36]. The D/OK and D/660 IDV reassortments of P3 and NP gene were also reported from the province of Quebec, Canada [9]. An IDV isolate similar to swine and bovine IDV in the US was sequenced in France in 2012 [6] and reported to form a separate cluster from the clusters of known lineages. This isolate was annotated as lineage D/Fra [49] or D/France-2012 [50] by different groups.

Accurate classification of the circulating lineages could help in tracking the evolution of IDV and the potential impact on animal and human health. In addition, rapid and precise

typing methods help in disease diagnosis and surveillance efforts and facilitate basic and translational research. Owing to the diversification of the IDV into multiple lineages, typing IDV is currently carried out using alignment-based phylogeny methods [36,42–44]. The lineages are assigned based on clustering proximity of the unknown with the known ones, and most often than not, the typing analyses need to be carried out in an iterative mode, which requires considerable amount of time. Similarly, lineage assignment to the query sequence is also carried out using BLAST-based searches wherein the lineage of the sequence sharing highest identity and similarity with end-to-end query coverage is assigned to the query. However, the lack of lineage information in the GenBank entries poses a challenge for assignment of lineage to the query sequence using BLAST-based methods. Evolutionary changes in the IDV genome have resulted in the expansion of genetic diversity and the emergence of multiple lineages that might expand host tropism and potentially increase the pathogenicity to animals and humans. Therefore, there is an urgent need for automated, accurate and rapid typing tools for IDV lineage typing.

This paper reports the development of the IDV Typer webserver using an alignment-free method developed to automate the process of typing for IDV [51,52]. The alignment-free method is based on the return time distribution (RTD) concept and corresponding statistics that enables clustering, molecular phylogeny and typing. This RTD method has been previously implemented for the development of typing servers for viruses such as Mumps virus (MuV), Dengue virus (DENV), West Nile virus (WNV) and Human Rhino virus (HRV) [53–56], which provides a distinctive advantage for typing of the said viruses over alignment-based methods. Typing servers for human enteroviruses, human rhinoviruses, noroviruses, Zika, dengue, chikungunya and yellow fever viruses and SARS-CoV-2 employing alignment-based methods such as BLAST-based searches have also been developed and are available online [57–60].

## 2. Materials and Methods

### 2.1. Datasets

The IDV HEF gene sequences obtained from GenBank and the literature review were curated and compiled as three independent datasets such as reference, true positive and true negative, which were used to develop an alignment-free RTD-based method for lineage typing of IDV.

### 2.2. Reference Dataset

A total of 23 HEF gene sequences of IDV obtained from GenBank with sequence length in the range of 1796–1995 bp ( $\geq 90\%$  query coverage) were compiled, for which the lineage information of the sequences was obtained from the literature [2,15,18,36,45,46,49]. The list of accession numbers, corresponding lineage information and reference (PubMed ID) is provided in Supplementary Materials Table S1.

### 2.3. True Positive Dataset

The HEF gene sequences were extracted from GenBank by performing a BLAST search using Refseq entry (Accession number: NC\_036618.1) as the query. The sequences used in the reference dataset were excluded from the list of significant hits, and thus a total of 143 entries were obtained. The lineage information for 142/143 entries was compiled from the literature review [11,12,15,17,18,36,45,46,50,61].

The lineage information was assigned to the remaining entry using alignment-based Maximum-Likelihood (ML) phylogeny. The 143 entries (142 entries with known lineage and one entry with unknown lineage) of the True Positive dataset were aligned using the MAFFT algorithm implemented in SEED 2 software [62]. Based on the Bayesian Information Criterion (BIC) implemented in the IQ-Tree webserver [63,64], the TVM+F+G4 nucleotide substitution model was used for generating an ML-based phylogenetic tree, and lineage was assigned to the entry (with unknown lineage) based on clustering proximity with other isolates (with known lineage information). The accession numbers, lineage

information and reference (PubMed ID) of the 143 entries used as the True Positive dataset are listed in Table S2. This dataset was used to test the sensitivity of the typing server.

#### 2.4. True Negative Dataset

The negative dataset consisting of 50 entries was generated for testing the specificity and sensitivity of the typing server. The HEF and non-HEF gene sequences from other *Orthomyxoviridae* family members with lengths in the range of 1796–1995 bp were used in the dataset. The non-HEF gene sequences of the IDV were also included in this dataset. The GenBank accession numbers, gene names and virus names of 50 entries used in the True Negative Dataset are listed in Table S3.

#### 2.5. Method

The steps involved in the development of the RTD-based typing server are: (a) reference sequence dataset curation and compilation, (b) optimization of  $k$ -mer, (c) calculate the RTDs and their parameters for the optimum value of  $k$ , (d) obtain a pairwise Euclidean distance matrix based on the parameters of RTD obtained in (c), (e) generation of a Neighbor-joining (NJ) tree using the Euclidean distance matrix, (f) derivation of inter-lineage distance cut-off ranges, (g) assign each lineage to the sequences based on nearest reference lineage and respective distance cut-off range of a specific lineage. Steps (a) and (b) are essential for the development of virus-specific typing servers.

#### 2.6. $k$ Value Optimization for IDV Typing

Optimization of  $k$  is an essential step in the RTD-based typing method to achieve accuracy, sensitivity and specificity [52]. The protocol for optimization of  $k$  for IDV data is as follows. The statistical parameters, mean ( $\mu$ ) and standard deviation ( $\sigma$ ) of each RTDs ( $4^k$ ) for every  $k$ -mer in the range of 1 to 8 were calculated for every sequence in the reference dataset [53] and hence, every gene sequence in the reference dataset constituted a numeric vector of  $2 \times 4^k$  dimensions. Numeric vectors for each value of  $k$  were used for the generation of a Euclidean-based pairwise distance matrix that was used to infer Neighbor-joining (NJ)-based phylogeny trees [52,65] by using the DendroPy library [66] in Python. The optimum  $k$  value was chosen based on the accuracy of clustering in accordance with the lineage classification of the reference dataset.

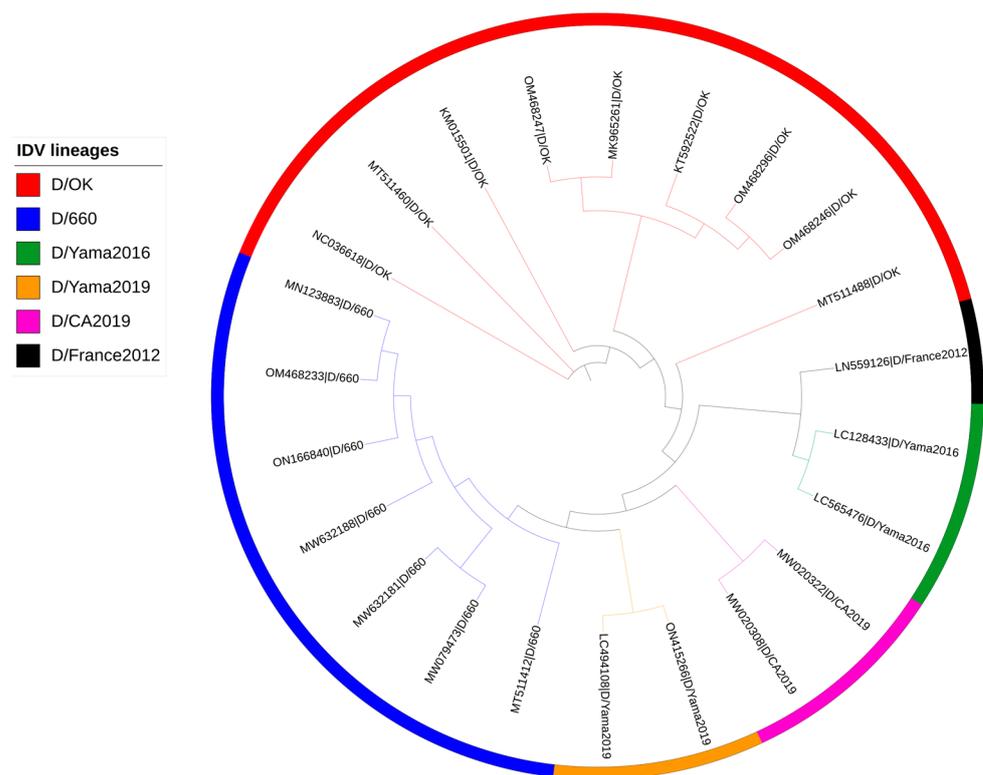
#### 2.7. IDV Typing Server Development

The lineage typing server for IDV, named “IDV Typer” was developed using Apache, CGI and PHP architecture. The server is available in the public domain at <http://bioinfo.unipune.ac.in/IDV/home.html>.

### 3. Results

The IDV reference dataset was compiled to include representative sequences of the five known lineages. The lineages D/Yama2016, D/Yama2019 and D/CA2019 included five, five and three members, respectively, hence two sequences each from D/Yama2016, D/Yama2019 and D/CA2019 were included in the reference dataset. The D/France2012 lineage (tentative sixth lineage) included in reference dataset consisted of a single entry D/bovine/France/2986/2012 (Accession ID: LN559126) as it forms an independent branch outside the other known lineages [49,50]. Thus, a reference dataset of 23 entries was used as a training dataset for optimization of  $k$  and designing of RTD-based lineage typing method for IDV.

In the case of the reference dataset of IDV, RTD-based NJ trees were derived for different values of  $k$  (1 to 8). At  $k = 5$  (Figure 1), lineage-wise clustering was observed and no misclassification was detected, indicative of the optimum  $k$  value ( $k = 5$ ).



**Figure 1.** Alignment-free RTD-based phylogenetic tree of HEF gene reference dataset for  $k = 5$ . The tip labels are divided into two parts based on the pipe (|) character. The first part indicates GenBank accession numbers and the second part indicates lineage. The tree was visualized and annotated using the ITOL server (<https://itol.embl.de/>).

Though complete HEF gene sequence is recommended for lineage typing, analysis of performance of IDV Typer indicated that a minimum coverage of  $\geq 90\%$  ( $\geq 1796$  bp) of sequence length of HEF gene is required.

### 3.1. “IDV Typer” RTD Method-Based Server for Lineage Typing of IDV

“IDV typer” (<http://bioinfo.unipune.ac.in/IDV/home.html>), an RTD-based lineage typing server for IDV was developed and uses the reference dataset along with the derived parameters as a backend knowledgebase. The server takes HEF gene sequences with  $>90\%$  of sequence coverage in FASTA format through a sequence submission form. The CGI scripts at the back end convert sequences into RTD vectors and compute the statistical parameters ( $\mu$  and  $\sigma$ ) for the 1024 pentamer RTDs for the query sequence/s and calculate the Euclidean distance between every query sequence and the sequences of the reference dataset. The lineage assignment to the query sequence is carried out using a 2-step procedure. The first step involved the identification of the lineage of the closest reference sequence based on a distance measure, and the second step involved verifying if the distance of the query sequence from the sequence of respective reference lineage, lay within the pre-computed distance ranges. Successful lineage identification is subject to meeting both these criteria. Failure to meet these criteria for all the lineages results in a message being returned as “No lineage can be assigned to this query sequence using RTD-based method” on the IDV typer page. The time required for the lineage assignment for a dataset consisting of 15–100 sequences is 1–5 s.

### 3.2. Server Validation

The server’s performance was assessed using the True Positive and True Negative datasets compiled as reported in the Materials section. The True Positive dataset con-

sists of 143 entries of IDV HEF gene sequences. The lineage information was assigned to these sequences either based on the literature survey (#142) or clustering proximity (#1) in alignment-based phylogeny with the True Positive dataset. The information regarding number of entries from every lineage used in the Reference and True Positive datasets is provided in Table 1. The one entry for which lineage information was not available was assigned lineage D/660 based on the clustering proximity in the ML-based tree. One entry isolated from Turkey (GenBank ID: OP353622) was reported to form an independent branch and was tentatively assigned a novel lineage called D/Bursa2013 [11]. However, it was observed that this entry formed an independent branch within the D/OK lineage in the ML-based phylogenetic tree of the True Positive dataset. Therefore, this entry was assigned as D/OK lineage. Phylogenetic analysis by a recent study reported independent branching of few isolates and classified them as independent groups. These groups were named as D/Michigan-2019, consisting of two entries (GenBank ID: MW632174 and MW079478) and D/Texas-2017, consisting of one entry (GenBank ID: MT636473) [50]. However, these isolates have been classified under the D/OK lineage by other recent studies [15,46]. Therefore, these isolates have been assigned to D/OK lineage in this study. The lineage assignment of all 143 entries by the RTD-based IDV typer server corroborated with the lineage assignment carried out by literature review and alignment-based ML methods, which confirmed 100% sensitivity of the IDV Typer server.

**Table 1.** Lineage-wise distribution of reference dataset.

Lineage	Reference Dataset	True Positive Dataset
D/OK	9	89
D/660	7	47
D/Yama2016	2	3
D/Yama2019	2	3
D/CA2019	2	1
D/France2012	1	0

The negative dataset consisted of 50 entries from the members of the *Orthomyxoviridae* family. The RTD-based alignment-free method performed with 100% specificity as it did not assign any lineage to the entries of the True Negative dataset.

#### 4. Discussion

Genomic surveillance is a cornerstone in proactively monitoring and managing emerging viral pathogens. In particular, proactive surveillance of emerging pathogens like IDV, which exhibits a propensity for genetic diversity and cross-species transmission, is crucial. Accurate typing of such viral variants facilitates the elucidation of their evolutionary trajectories and the identification of potential virulence factors and determinants of host specificity. Moreover, it enables real-time tracking of viral spread and informs the development of targeted interventions, ranging from vaccine design to surveillance strategies. Therefore, genomic surveillance coupled with precise typing methodologies is indispensable in safeguarding public health and mitigating the impacts of emerging viral threats. Since the isolation of IDV from swine in 2011 in Oklahoma, USA, its widespread prevalence across multiple countries and its ability to infect numerous animal hosts have become evident [22–24]. The spread of the virus on such a large scale, along with developments in sequencing technologies, has led to increased sequencing of virus samples, facilitating the understanding of IDV evolution and genetic diversity.

The threat of IDV to public health looms large due to its ability to infect not only humans but also multiple non-animal hosts. With a global distribution and the capacity to evolve continually, IDV presents a formidable challenge, particularly exacerbated by the lack of regular testing and surveillance efforts. Without ongoing monitoring, the virus can

mutate and spread under the radar, potentially leading to the emergence of lineages with increased pathogenic and/or virulence in humans, leading to epidemics and or pandemics. The global circulation of IDV and its ability to infect a wide range of hosts has prompted the need to detect and map diversifying lineages, which requires developing a system for identifying and classifying the virus into their respective lineages. Automated, accurate and rapid IDV typing tools can help in the quick identification of currently circulating and emerging lineages.

The phylogenetic and evolutionary studies of IDV have been carried out to understand diversification of IDV into known lineages. These studies provide insight into the evolution of IDV and the emergence of new lineages. A study has reported the presence of a lineage-specific HEF amino acid mutation that might impact antibody binding affinity [42]. Consequently, a recent study has established the broader antigenicity of the D/CA2019 lineage over the others as antibodies elicited by D/CA2019 isolates might neutralize isolates of other lineages [36]. The presence of sites under positive selection with lineage specific mutations in HEF has also been reported [18,67]. With the ability to type circulating IDV lineages rapidly, the IDV Typer server can help in studies characterizing the potential impact on host tropism, human health implications and pathogenicity.

The most frequently used approach for IDV lineage typing is based on alignment-based phylogenetic analysis of HEF gene sequences, which involves multiple sequence alignment (MSA). Even though the currently available alignment-based molecular phylogeny analyses (MPA) method is useful, it has two major limitations, the complexity of analysis increases significantly with an increase in the size of the dataset and the need for human intervention for contextual interpretation of tree topologies to assign lineage. Furthermore, with the addition of even a single entry, the whole analysis, beginning with MSA needs to be repeated. The BLAST-based approach is also occasionally used for assignment of lineage to IDV isolates. However, BLAST-based lineage typing is subject to both manual intervention and availability of annotation/metadata in the knowledgebase. Currently, GenBank entries of significant number of isolates lack lineage information and therefore, there is a need to perform downstream analysis involving alignment-based phylogeny or extensive literature survey in order to assign lineages to the query sequences from respective isolates.

To overcome these limitations, a novel alignment-free RTD-based method for fast and accurate lineage typing of IDV is developed. Customization of RTD approach for typing of any virus requires the development of a virus-specific knowledgebase (reference dataset) and the derivation of optimum value of  $k$  [52] that helps to determine the pre-computed empirical cut-off and eliminates the basic step of MSA implied in the alignment-based method for geno/sero/lineage typing. Using 1024 pentamers of RTDs for alignment-free MPA of reference and true positive sequences was completed within a few minutes and needed to be completed only by developers. A comparative analysis using a True Positive dataset was carried out to identify the time required for lineage typing of IDV sequences using RTD-based IDV Typer, BLAST-based search and alignment-based MPA. The IDV typer was able to assign correct lineage information to 143 entries from True Positive dataset in about 12 s. In comparison, the BLAST-based search (online mode) and alignment-based phylogeny (standalone mode) of the True Positive dataset required an execution time of about 42 and 500 s, respectively, which is further extended due to the need for manual interpretation of results for lineage assignment. Thus, the RTD-based IDV Typer required significantly less amount of time for correct lineage assignment to 143 entries which indicates the high speed and accuracy of RTD-based typing as compared to other methods. The knowledgebase is updated periodically by the developers and therefore, in case of identification of a novel lineage and/or reclassification isolates to different lineage, the whole process of updating the knowledgebase,  $k$ -mer optimization and deriving genetic distance cut-off will be carried out by the developers and users need not worry about it and can carry out lineage assignment without having to compile reference datasets and perform alignment-based phylogenetic analysis.

Thus, the development of virus-specific RTD-based alignment-free typing servers addresses three gaps, namely, the elimination of the step involving MSA for MPA, automation of typing protocol and a significant reduction in computational time required for typing while maintaining 100% accuracy. This proves to be a unique strength of the RTD-based alignment-free method over the conventional alignment-based method; hence, for faster surveillance of circulating lineages of IDV, the RTD-based server was developed.

## 5. Conclusions

The alignment-free RTD-based method with suitable modifications was successfully applied for the lineage typing of IDV. A web server implementing the alignment-free RTD-based method was developed and made available online for the assignment of the IDV into the known lineages without human intervention and thereby automating the lineage assignment process. The IDV Typer server is a highly sensitive, specific and computationally efficient tool that provides an alternative over conventional methods such as alignment-based MPA and BLAST-based searches. The IDV Typer outperforms available methods for lineage assignment in terms of computational efficiency and essential accuracy. It automates the process of lineage typing while providing an easy to use interface, due to which the tool could be deployed in the field for real-time surveillance of the circulating IDV lineages and help to monitor genetic diversity of IDVs isolates. The IDV typer can, therefore, be a handy tool for better understanding the epidemiology of IDV and enabling genomic surveillance.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/v16030373/s1>, Table S1. Reference dataset; Table S2. True positive dataset; Table S3. True Negative dataset.

**Author Contributions:** Conceptualization, U.K.-K. and S.L.; methodology, U.K.-K., M.M.K. and S.L.; software, S.L. and A.S.; validation, U.K.-K., M.M.K. and S.V.K.; formal analysis, S.L. and A.S.; data curation, S.L.; writing—original draft preparation, U.K.-K., S.L., M.M.K., A.S. and S.V.K.; writing—review and editing, U.K.-K., M.M.K. and S.V.K.; visualization, S.L. and A.S.; supervision, U.K.-K., M.M.K. and S.V.K.; project administration, U.K.-K., M.M.K. and S.V.K.; funding acquisition, U.K.-K. and S.V.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** The Department of Biotechnology (DBT), Government of India, New Delhi, for funding the grant BT/PR40171/BTIS/137/21/2021 to establish Bioinformatics and Computational Biology Centre (BIC) at the Savitribai Phule Pune University, Pune, India, (UKK). Chair start-up funds from the University of Pittsburgh School of Public Health (SVK).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The original contributions presented in the study are included in the article and supplementary material.

**Acknowledgments:** U.K.-K. and M.M.K. acknowledge infrastructural and financial support from the Savitribai Phule Pune University. S.L. acknowledges the NET-SRF fellowship awarded by the University Grants Commission (UGC), Govt. of India. A.S. acknowledges the Bioinformatics Centre for the support.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Hause, B.M.; Collin, E.A.; Liu, R.; Huang, B.; Sheng, Z.; Lu, W.; Wang, D.; Nelson, E.A.; Li, F. Characterization of a novel influenza virus in cattle and Swine: Proposal for a new genus in the Orthomyxoviridae family. *mBio* **2014**, *5*, e00031-14. [[CrossRef](#)]
2. Hause, B.M.; Ducatez, M.; Collin, E.A.; Ran, Z.; Liu, R.; Sheng, Z.; Arminen, A.; Kaplan, B.; Chakravarty, S.; Hoppe, A.D. Isolation of a novel swine influenza virus from Oklahoma in 2011 which is distantly related to human influenza C viruses. *PLoS Pathog.* **2013**, *9*, e1003176. [[CrossRef](#)]

3. Mitra, N.; Cernicchiaro, N.; Torres, S.; Li, F.; Hause, B.M. Metagenomic characterization of the virome associated with bovine respiratory disease in feedlot cattle identified novel viruses and suggests an etiologic role for influenza D virus. *J. Gen. Virol.* **2016**, *97*, 1771. [[CrossRef](#)]
4. Flynn, O.; Gallagher, C.; Mooney, J.; Irvine, C.; Ducatez, M.; Hause, B.; McGrath, G.; Ryan, E. Influenza D virus in cattle, Ireland. *Emerg. Infect. Dis.* **2018**, *24*, 389. [[CrossRef](#)]
5. Dane, H.; Duffy, C.; Guelbenzu, M.; Hause, B.; Fee, S.; Forster, F.; McMenamy, M.J.; Lemon, K. Detection of influenza D virus in bovine respiratory disease samples, UK. *Transbound. Emerg. Dis.* **2019**, *66*, 2184–2187. [[CrossRef](#)]
6. Ducatez, M.F.; Pelletier, C.; Meyer, G. Influenza D virus in cattle, France, 2011–2014. *Emerg. Infect. Dis.* **2015**, *21*, 368. [[CrossRef](#)] [[PubMed](#)]
7. Chiapponi, C.; Faccini, S.; De Mattia, A.; Baioni, L.; Barbieri, I.; Rosignoli, C.; Nigrelli, A.; Foni, E. Detection of influenza D virus among swine and cattle, Italy. *Emerg. Infect. Dis.* **2016**, *22*, 352. [[CrossRef](#)]
8. Zhai, S.-L.; Zhang, H.; Chen, S.-N.; Zhou, X.; Lin, T.; Liu, R.; Lv, D.-H.; Wen, X.-H.; Wei, W.-K.; Wang, D. Influenza D virus in animal species in Guangdong Province, southern China. *Emerg. Infect. Dis.* **2017**, *23*, 1392. [[CrossRef](#)] [[PubMed](#)]
9. Saegerman, C.; Gaudino, M.; Savard, C.; Broes, A.; Ariel, O.; Meyer, G.; Ducatez, M.F. Influenza D virus in respiratory disease in Canadian, province of Québec, cattle: Relative importance and evidence of new reassortment between different clades. *Transbound. Emerg. Dis.* **2022**, *69*, 1227–1245. [[CrossRef](#)]
10. Murakami, S.; Endoh, M.; Kobayashi, T.; Takenaka-Uema, A.; Chambers, J.K.; Uchida, K.; Nishihara, M.; Hause, B.; Horimoto, T. Influenza D virus infection in herd of cattle, Japan. *Emerg. Infect. Dis.* **2016**, *22*, 1517. [[CrossRef](#)] [[PubMed](#)]
11. Yesilbag, K.; Toker, E.B.; Ates, O. Recent strains of influenza D virus create a new genetic cluster for European strains. *Microb. Pathog.* **2022**, *172*, 105769. [[CrossRef](#)]
12. Parys, A.; Vereecke, N.; Vandoorn, E.; Theuns, S.; Van Reeth, K. Surveillance and Genomic Characterization of Influenza A and D Viruses in Swine, Belgium and the Netherlands, 2019–2021. *Emerg. Infect. Dis.* **2023**, *29*, 1459–1464. [[CrossRef](#)] [[PubMed](#)]
13. da Silva, M.S.; Mosena, A.C.S.; Baumbach, L.; Demoliner, M.; Gularte, J.S.; Pavarini, S.P.; Driemeier, D.; Weber, M.N.; Spilki, F.R.; Canal, C.W. Cattle influenza D virus in Brazil is divergent from established lineages. *Arch. Virol.* **2022**, *167*, 1181–1184. [[CrossRef](#)] [[PubMed](#)]
14. Lim, E.H.; Lim, S.-I.; Kim, M.J.; Kwon, M.; Kim, M.-J.; Lee, K.-B.; Choe, S.; An, D.-J.; Hyun, B.-H.; Park, J.-Y. First Detection of Influenza D Virus Infection in Cattle and Pigs in the Republic of Korea. *Microorganisms* **2023**, *11*, 1751. [[CrossRef](#)] [[PubMed](#)]
15. Brito, B.P.; Frost, M.J.; Anantanawat, K.; Jaya, F.; Batterham, T.; Djordjevic, S.P.; Chang, W.-S.; Holmes, E.C.; Darling, A.E.; Kirkland, P.D. Expanding the range of the respiratory infectome in Australian feedlot cattle with and without respiratory disease using metatranscriptomics. *Microbiome* **2023**, *11*, 158. [[CrossRef](#)] [[PubMed](#)]
16. Alvarez, I.J.; Fort, M.; Pasucci, J.; Moreno, F.; Gimenez, H.; Näslund, K.; Hägglund, S.; Zohari, S.; Valarcher, J.F. Seroprevalence of influenza D virus in bulls in Argentina. *J. Vet. Diagn. Invest.* **2020**, *32*, 585–588. [[CrossRef](#)] [[PubMed](#)]
17. Molini, U.; Curini, V.; Jacobs, E.; Tongo, E.; Berjaoui, S.; Hemberger, M.Y.; Puglia, I.; Jago, M.; Khaïseb, S.; Cattoli, G. First influenza D virus full-genome sequence retrieved from livestock in Namibia, Africa. *Acta Trop.* **2022**, *232*, 106482. [[CrossRef](#)] [[PubMed](#)]
18. Goecke, N.B.; Liang, Y.; Otten, N.D.; Hjulsgager, C.K.; Larsen, L.E. Characterization of influenza D virus in Danish calves. *Viruses* **2022**, *14*, 423. [[CrossRef](#)]
19. Salem, E.; Cook, E.A.; Lbacha, H.A.; Oliva, J.; Awoume, F.; Aplogan, G.L.; Hymann, E.C.; Muloi, D.; Deem, S.L.; Alali, S. Serologic evidence for influenza C and D virus among ruminants and camelids, Africa, 1991–2015. *Emerg. Infect. Dis.* **2017**, *23*, 1556. [[CrossRef](#)]
20. Snoeck, C.J.; Oliva, J.; Pauly, M.; Losch, S.; Wildschutz, F.; Muller, C.P.; Hübschen, J.M.; Ducatez, M.F. Influenza D virus circulation in cattle and swine, Luxembourg, 2012–2016. *Emerg. Infect. Dis.* **2018**, *24*, 1388. [[CrossRef](#)]
21. Murakami, S.; Odagiri, T.; Melaku, S.K.; Bazartseren, B.; Ishida, H.; Takenaka-Uema, A.; Muraki, Y.; Sentsui, H.; Horimoto, T. Influenza D virus infection in dromedary camels, Ethiopia. *Emerg. Infect. Dis.* **2019**, *25*, 1224. [[CrossRef](#)]
22. Quast, M.; Sreenivasan, C.; Sexton, G.; Nedland, H.; Singrey, A.; Fawcett, L.; Miller, G.; Lauer, D.; Voss, S.; Pollock, S. Serological evidence for the presence of influenza D virus in small ruminants. *Vet. Microbiol.* **2015**, *180*, 281–285. [[CrossRef](#)]
23. Asha, K.; Kumar, B. Emerging influenza D virus threat: What we know so far! *J. Clin. Med.* **2019**, *8*, 192. [[CrossRef](#)]
24. Yu, J.; Li, F.; Wang, D. The first decade of research advances in influenza D virus. *J. Gen. Virol.* **2021**, *102*, jgv001529. [[CrossRef](#)]
25. Guan, M.; Jacobson, O.; Sarafianos, G.; Baroch, J.; Deliberto, T.J.; Wan, X.-F. Exposure of white-tailed deer in North America to influenza D virus. *Virology* **2022**, *573*, 111–117. [[CrossRef](#)]
26. Nedland, H.; Wollman, J.; Sreenivasan, C.; Quast, M.; Singrey, A.; Fawcett, L.; Christopher-Hennings, J.; Nelson, E.; Kaushik, R.; Wang, D. Serological evidence for the co-circulation of two lineages of influenza D viruses in equine populations of the Midwest United States. *Zoonoses Public Health* **2018**, *65*, e148–e154. [[CrossRef](#)]
27. O'donovan, T.; Donohoe, L.; Ducatez, M.F.; Meyer, G.; Ryan, E. Seroprevalence of influenza D virus in selected sample groups of Irish cattle, sheep and pigs. *Ir. Vet. J.* **2019**, *72*, 11. [[CrossRef](#)] [[PubMed](#)]
28. Luo, J.; Ferguson, L.; Smith, D.R.; Woolums, A.R.; Epperson, W.B.; Wan, X.F. Serological evidence for high prevalence of Influenza D Viruses in Cattle, Nebraska, United States, 2003–2004. *Virology* **2017**, *501*, 88–91. [[CrossRef](#)] [[PubMed](#)]
29. Silveira, S.; Falkenberg, S.M.; Kaplan, B.S.; Crossley, B.; Ridpath, J.F.; Bauermann, F.B.; Fossler, C.P.; Dargatz, D.A.; Dassanayake, R.P.; Vincent, A.L.; et al. Serosurvey for Influenza D Virus Exposure in Cattle, United States, 2014–2015. *Emerg. Infect. Dis.* **2019**, *25*, 2074–2080. [[CrossRef](#)] [[PubMed](#)]

30. Ruiz, M.; Puig, A.; Bassols, M.; Fraile, L.; Armengol, R. Influenza D Virus: A Review and Update of Its Role in Bovine Respiratory Syndrome. *Viruses* **2022**, *14*, 2717. [[CrossRef](#)] [[PubMed](#)]
31. Fulton, R.W. Viruses in bovine respiratory disease in North America: Knowledge advances using genomic testing. *Vet. Clin. Food Anim. Pract.* **2020**, *36*, 321–332. [[CrossRef](#)]
32. Padalino, B.; Cirone, F.; Zappaterra, M.; Tullio, D.; Ficco, G.; Giustino, A.; Ndiana, L.A.; Pratelli, A. Factors affecting the development of bovine respiratory disease: A cross-sectional study in beef steers shipped from France to Italy. *Front. Vet. Sci.* **2021**, *8*, 627894. [[CrossRef](#)] [[PubMed](#)]
33. Ng, T.F.F.; Kondov, N.O.; Deng, X.; Van Eenennaam, A.; Neibergs, H.L.; Delwart, E. A metagenomics and case-control study to identify viruses associated with bovine respiratory disease. *J. Virol.* **2015**, *89*, 5340–5349. [[CrossRef](#)] [[PubMed](#)]
34. Zhang, H.; Porter, E.; Lohman, M.; Lu, N.; Peddireddi, L.; Hanzlicek, G.; Marthaler, D.; Liu, X.; Bai, J. Influenza C virus in cattle with respiratory disease, United States, 2016–2018. *Emerg. Infect. Dis.* **2018**, *24*, 1926. [[CrossRef](#)] [[PubMed](#)]
35. Zhang, M.; Hill, J.E.; Fernando, C.; Alexander, T.W.; Timsit, E.; van der Meer, F.; Huang, Y. Respiratory viruses identified in western Canadian beef cattle by metagenomic sequencing and their association with bovine respiratory disease. *Transbound. Emerg. Dis.* **2019**, *66*, 1379–1386. [[CrossRef](#)] [[PubMed](#)]
36. Huang, C.; Yu, J.; Hause, B.M.; Park, J.Y.; Sreenivasan, C.; Uprety, T.; Sheng, Z.; Wang, D.; Li, F. Emergence of new phylogenetic lineage of Influenza D virus with broad antigenicity in California, United States. *Emerg. Microbes Infect.* **2021**, *10*, 739–742. [[CrossRef](#)]
37. White, S.K.; Ma, W.; McDaniel, C.J.; Gray, G.C.; Lednicky, J.A. Serologic evidence of exposure to influenza D virus among persons with occupational contact with cattle. *J. Clin. Virol.* **2016**, *81*, 31–33. [[CrossRef](#)] [[PubMed](#)]
38. Trombetta, C.M.; Marchi, S.; Manini, I.; Kistner, O.; Li, F.; Piu, P.; Manenti, A.; Biuso, F.; Sreenivasan, C.; Druce, J. Influenza D virus: Serological evidence in the Italian population from 2005 to 2017. *Viruses* **2019**, *12*, 30. [[CrossRef](#)] [[PubMed](#)]
39. Leibler, J.H.; Abdelgadir, A.; Seidel, J.; White, R.F.; Johnson, W.E.; Reynolds, S.J.; Gray, G.C.; Schaeffer, J.W. Influenza D virus exposure among US cattle workers: A call for surveillance. *Zoonoses Public Health* **2023**, *70*, 166–170. [[CrossRef](#)]
40. Borkenhagen, L.K.; Mallinson, K.A.; Tsao, R.W.; Ha, S.-J.; Lim, W.-H.; Toh, T.-H.; Anderson, B.D.; Fieldhouse, J.K.; Philo, S.E.; Chong, K.-S. Surveillance for respiratory and diarrheal pathogens at the human-pig interface in Sarawak, Malaysia. *PLoS ONE* **2018**, *13*, e0201295. [[CrossRef](#)]
41. Ferguson, L.; Luo, K.; Olivier, A.K.; Cunningham, F.L.; Blackmon, S.; Hanson-Dorr, K.; Sun, H.; Baroch, J.; Lutman, M.W.; Quade, B.; et al. Influenza D Virus Infection in Feral Swine Populations, United States. *Emerg. Infect. Dis.* **2018**, *24*, 1020–1028. [[CrossRef](#)] [[PubMed](#)]
42. Collin, E.A.; Sheng, Z.; Lang, Y.; Ma, W.; Hause, B.M.; Li, F. Cocirculation of two distinct genetic and antigenic lineages of proposed influenza D virus in cattle. *J. Virol.* **2015**, *89*, 1036–1042. [[CrossRef](#)] [[PubMed](#)]
43. Odagiri, T.; Ishida, H.; Li, J.-Y.; Endo, M.; Kobayashi, T.; Kamiki, H.; Matsugo, H.; Takenaka-Uema, A.; Murakami, S.; Horimoto, T. Antigenic heterogeneity among phylogenetic clusters of influenza D viruses. *J. Vet. Med. Sci.* **2018**, *80*, 1241–1244. [[CrossRef](#)] [[PubMed](#)]
44. Murakami, S.; Sato, R.; Ishida, H.; Katayama, M.; Takenaka-Uema, A.; Horimoto, T. Influenza D virus of new phylogenetic lineage, Japan. *Emerg. Infect. Dis.* **2020**, *26*, 168. [[CrossRef](#)] [[PubMed](#)]
45. Chiapponi, C.; Faccini, S.; Fusaro, A.; Moreno, A.; Prospero, A.; Merenda, M.; Baioni, L.; Gabbi, V.; Rosignoli, C.; Alborali, G.L. Detection of a new genetic cluster of influenza D virus in Italian cattle. *Viruses* **2019**, *11*, 1110. [[CrossRef](#)] [[PubMed](#)]
46. Yu, J.; Li, T.; Wen, Z.; Wu, S.; Wang, Z.; Zheng, J.; Chen, M.; Chen, F.; Wei, W.-K.; Zhai, S.-L. Identification of D/Yama2019 Lineage-Like Influenza D Virus in Chinese Cattle. *Front. Vet. Sci.* **2022**, *9*, 939456. [[CrossRef](#)] [[PubMed](#)]
47. Xu, Y.; Liang, H.; Wen, H. Frequent reassortment and potential recombination shape the genetic diversity of influenza D viruses. *J. Infect.* **2021**, *82*, e36–e38. [[CrossRef](#)] [[PubMed](#)]
48. Hayakawa, J.; Masuko, T.; Takehana, T.; Suzuki, T. Genetic and antigenic characterization and retrospective surveillance of bovine influenza D viruses identified in Hokkaido, Japan from 2018 to 2020. *Viruses* **2020**, *12*, 877. [[CrossRef](#)]
49. Wan, X.-F.; Ferguson, L.; Oliva, J.; Rubrum, A.; Eckard, L.; Zhang, X.; Woolums, A.R.; Lion, A.; Meyer, G.; Murakami, S. Limited cross-protection provided by prior infection contributes to high prevalence of influenza D viruses in cattle. *J. Virol.* **2020**, *94*, e00240-20. [[CrossRef](#)]
50. Gaudino, M.; Chiapponi, C.; Moreno, A.; Zohari, S.; O'Donovan, T.; Quinless, E.; Sausy, A.; Oliva, J.; Salem, E.; Fusade-Boyer, M. Evolutionary and temporal dynamics of emerging influenza D virus in Europe (2009–22). *Virus Evol.* **2022**, *8*, veac081. [[CrossRef](#)]
51. Kolekar, P.S.; Kale, M.M.; Kulkarni-Kale, U. 'Inter-Arrival Time' Inspired Algorithm and Its Application in Clustering and Molecular Phylogeny. *AIP Conf. Proc.* **2010**, *1298*, 307–312.
52. Kolekar, P.; Kale, M.; Kulkarni-Kale, U. Alignment-free distance measure based on return time distribution for sequence analysis: Applications to clustering, molecular phylogeny and subtyping. *Mol. Phylogenetics Evol.* **2012**, *65*, 510–522. [[CrossRef](#)]
53. Kolekar, P.; Kale, M.; Kulkarni-Kale, U. Genotyping of Mumps viruses based on SH gene: Development of a server using alignment-free and alignment-based methods. *Immunome Res.* **2011**, *7*, 1–7.
54. Kolekar, P.; Kale, M.; Kulkarni-Kale, U. Subtyping of Dengue Viruses using Return Time Distribution based Approach. *Nat. Preced.* **2011**, *1*. [[CrossRef](#)]
55. Kolekar, P.; Hake, N.; Kale, M.; Kulkarni-Kale, U. WNV Typer: A server for genotyping of West Nile viruses using an alignment-free method based on a return time distribution. *J. Virol. Methods* **2014**, *198*, 41–55. [[CrossRef](#)]

56. Kolekar, P.S.; Waman, V.P.; Kale, M.M.; Kulkarni-Kale, U. RV-Typer: A web server for typing of Rhinoviruses using alignment-free approach. *PLoS ONE* **2016**, *11*, e0149350. [[CrossRef](#)]
57. Kroneman, A.; Vennema, H.; Deforche, K.; Avoort, H.; Peñaranda, S.; Oberste, M.; Vinjé, J.; Koopmans, M. An automated genotyping tool for enteroviruses and noroviruses. *J. Clin. Virol.* **2011**, *51*, 121–125. [[CrossRef](#)] [[PubMed](#)]
58. Vilsker, M.; Moosa, Y.; Nooij, S.; Fonseca, V.; Ghysens, Y.; Dumon, K.; Pauwels, R.; Alcántara, L.C.; Vanden Eynden, E.; Vandamme, A.-M. Genome Detective: An automated system for virus identification from high-throughput sequencing data. *Bioinformatics* **2019**, *35*, 871–873. [[CrossRef](#)]
59. Fonseca, V.; Libin, P.J.; Theys, K.; Faria, N.R.; Nunes, M.R.; Restovic, M.I.; Freire, M.; Giovanetti, M.; Cuypers, L.; Nowe, A. A computational method for the identification of Dengue, Zika and Chikungunya virus species and genotypes. *PLoS Neglected Trop. Dis.* **2019**, *13*, e0007231. [[CrossRef](#)] [[PubMed](#)]
60. Cleemput, S.; Dumon, W.; Fonseca, V.; Abdool Karim, W.; Giovanetti, M.; Alcántara, L.C.; Deforche, K.; De Oliveira, T. Genome Detective Coronavirus Typing Tool for rapid identification and characterization of novel coronavirus genomes. *Bioinformatics* **2020**, *36*, 3552–3555. [[CrossRef](#)] [[PubMed](#)]
61. Ishida, H.; Murakami, S.; Kamiki, H.; Matsugo, H.; Takenaka-Uema, A.; Horimoto, T. Establishment of a reverse genetics system for influenza D virus. *J. Virol.* **2020**, *94*, e01767-19. [[CrossRef](#)]
62. Větrovský, T.; Baldrian, P.; Morais, D. SEED 2: A user-friendly platform for amplicon high-throughput sequencing data analyses. *Bioinformatics* **2018**, *34*, 2292–2294. [[CrossRef](#)]
63. Nguyen, L.-T.; Schmidt, H.A.; Von Haeseler, A.; Minh, B.Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **2015**, *32*, 268–274. [[CrossRef](#)]
64. Trifinopoulos, J.; Nguyen, L.-T.; von Haeseler, A.; Minh, B.Q. W-IQ-TREE: A fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* **2016**, *44*, W232–W235. [[CrossRef](#)] [[PubMed](#)]
65. Saitou, N.; Nei, M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **1987**, *4*, 406–425. [[PubMed](#)]
66. Sukumaran, J.; Holder, M.T. DendroPy: A Python library for phylogenetic computing. *Bioinformatics* **2010**, *26*, 1569–1571. [[CrossRef](#)] [[PubMed](#)]
67. He, W.-T.; Lu, M.; Xing, G.; Shao, Y.; Zhang, M.; Yang, Y.; Li, X.; Zhang, L.; Li, G.; Cao, Z. Emergence and adaptive evolution of influenza D virus. *Microb. Pathog.* **2021**, *160*, 105193. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.