IMMUNOME RESEARCH

RESEARCH Open Access

Genotyping of Mumps viruses based on SH gene: Development of a server using alignment-free and alignment-based methods

Pandurang S Kolekar[1], Mohan M Kale[2], Urmila Kulkarni-Kale[1§]

Abstract

Background

Mumps is an acute infectious childhood disease caused by mumps virus (MuV), a member of genus Rubulavirus, family Paramyxoviridae. Based on the genetic variability in small hydrophobic (SH) genes, currently MuVs have been divided into twelve confirmed genotypes designated as A-L and one proposed genotype, M. Despite successful vaccination program, a few genotypes are observed to co-circulate amongst vaccinated population. Furthermore, lack of cross protection between different genotypes is reported and hence, as a part of epidemiological surveillance, WHO has recommended genotyping of MuV. Currently genotyping is carried out using molecular phylogeny analysis (MPA) of SH genes and no genotyping server is available for MuV. The present study reports development of a genotyping server for the same, which employs three independent methods. The server uses two conventional methods viz., BLAST, MPA and a novel method based on Return Time Distribution (RTD), which is developed in-house.

Results

A server for genotyping of mumps virus is developed and made available at http://bioinfo.net.in/muv/homepage.html. RTD-based alignment-free method was initially developed for MPA and is applied for genotyping of MuV for the first time. It is found to have 98.95% of accuracy when measured using leave-one-out cross validation method on reference and test datasets. In addition to RTD, the server also implements BLAST and MPA for genotyping of MuV. All the three methods were found to be highly reliable as evident from consensus predictions.

Conclusions

A server for genotyping of MuV, which implements sequence-based bioinformatics approaches is developed and validated using SH gene sequences of known genotypes. This server will be useful for epidemiological surveillance and to monitor the circulation of MuV genotypes within and across geographic areas. This will also facilitate phylodynamics studies of mumps viruses.

Background

Mumps is an acute infectious viral disease, which is characterized by enlargement of the parotid and salivary glands. It is a common childhood disease and humans are the only natural hosts. The disease can vary from mild upper respiratory illness to viraemia developing aseptic meningitis, which can result in death or disability. The other effects of mumps include permanent deafness, orchitis and pancreatitis.

Mumps is caused by the Mumps virus (MuV), which spreads through respiratory droplets. The MuV belongs to genus *Rubulavirus*, subfamily *Paramyxovirinae*, family *Paramyxoviridae* and order *Mononegavirales*. The virus is enveloped with non-segmented, single-stranded, negative-sense RNA genome of length 15.3kb. It codes for seven proteins viz., nucleocapsid (N), phospo (P), matrix (M), fusion (F), small hydrophobic (SH), hemagglutinin-neuraminidase (HN) and large (L). Besides these, the phospho protein gene (P) also codes for two more proteins, namely V and I [1]. The genotyping is based on extent of similarities and variations in the SH gene sequences as recommended by the International Committee on Taxonomy of Viruses (ICTV) and World Health Organisation (WHO). The phylogenetic tree reconstruction studies of different isolates of mumps virus using SH gene sequences has revealed the existence of 12 genotypes, designated as A-L [2-5] and a newly proposed genotype M along with an unassigned genotype consisting of strains used for vaccination [3, 6, 7].

Live attenuated mumps vaccines are available as monovalent (mumps), bivalent (measles-mumps) and trivalent (measles-mumps-rubella). Mumps vaccination, along with measles and rubella in a triple formulation, is part of regular immunisation schedule of many countries [8-10] and the global vaccination strategies by WHO has resulted into elimination of measles and rubella in some regions [6]. However, the MuV strains are known to co-circulate within highly vaccinated populations [11-18] and lack of cross protection between different genotypes is reported, which is attributed either to failure or to varying accuracy of vaccines [19, 20]. Although, mumps infections are under control due to the implementation of vaccination program, there is a renewed interest to study the antigenic diversity and co-circulation of genotypes in the population due to recent outbreaks of mumps

[1]Bioinformatics Centre, University of Pune, Pune, India.
[2]Department of Statistics, University of Pune, Pune, India.
[§]Corresponding author
Email addresses:
PSK: pandurang@bioinfo.net.in
MMK: mmkale@stats.unipune.ac.in
UKK: urmila@bioinfo.net.in, urmila.kulkarni.kale@gmail.com

epidemics [21]. Hence, WHO has recommended molecular epidemiological studies involving genotyping of circulating strains of MuV as part of the surveillance program [22], which enables to study the pattern of viral circulation and evolution. Furthermore, it is also reported that the distribution of various genotypes may vary among closely related regions within a country [5, 23].

WHO has standardised scheme for genotyping of MuV with the help of experts [22], which is based on sequencing and sequence analysis of SH gene. Characterisation of at least two identical strains (to avoid sequencing and technical errors) and the divergence of more than 5% from the reference strains are recommended to designate a new genotype [3]. Although, no specific procedure for sequence analysis is prescribed, the virologists have been using N-J analysis with Kimura 2p model and 100 replicate for bootstrap [6] as well as Bayesian method [7] for genotyping of MuV.

Various genotyping servers employing BLAST and/or Molecular phylogeny analysis (MPA) have been developed for individual viruses viz., HepSEQ [24] for Hepatitis B virus, HIV STAR [25] for Human Immunodeficiency Virus; as well as for a group of viruses viz., BioAfrica, which is a general purpose viral genotyping tool [26]. However, no such genotyping server is currently available for MuVs. In the view of the above discussions, an attempt is made to design and develop a server for genotyping of MuV by employing sequence-based bioinformatics approaches. It implements two alignment based methods (BLAST and MPA) as well as a novel alignment-free method based on return time distribution (RTD), which is developed in house for clustering and phylogeny [27]. Its applications for genotyping are relatively new. It has been successfully applied for subtyping of Dengue viruses (http://dx.doi.org/10.1038/npre.2011.5590.1, manuscript communicated). The methods, datasets and benchmarking studies for mumps virus genotyping are reported here.

Results

Mumps, being an important infectious disease with reported insurgence irrespective of effective vaccination, is under surveillance. As per WHO and ICTV recommendations, the typing of co-circulating strains is carried out using sequences of SH gene of MuV. Currently virologists are using the MPA to assign the genotype to the field isolates.

The MuV server is developed (http://bioinfo.net.in/muv/homepage.html) for genotyping of mumps virus using SH gene sequences that implements three independent strategies. The strategies are broadly classified into alignment-free (RTD-based) and alignment based (BLAST and MPA) methods and are implemented at backend (see Figure 1). Of these, BLAST and MPA are routinely used for genotyping of mumps and other viruses whereas RTD based method is applied and optimised for genotyping of MuVs for the first time.

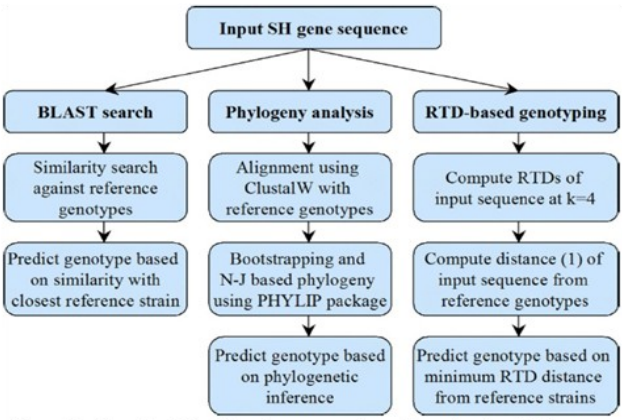

Figure 1 - Flowchart depicting the process flow of genotyping of Mumps virus using BLAST, Molecular Phylogeny Analysis and a novel method based on Return Time Distribution.

Compilation of reference and test datasets:

The 'reference dataset' consists of 28 sequences of SH gene from the confirmed genotypes A to L (see Table 1). The reference dataset contains multiple strains for each genotype except for genotype K, where data is available for only one strain. The newly proposed genotype M was not explicitly used for evaluation of methods. However, the reference dataset for prediction of genotypes that is made available on the server consists of 32 sequences from all the genotypes A to M. Current version does not include data of unassigned strains.

The 'test dataset 1' was curated and compiled to evaluate accuracy of prediction of all the three methods. This dataset consists of 96 entries of SH sequences that are available in GenBank with genotype mentioned explicitly [28]. The dataset is linked from the server (http://bioinfo.net.in/muv/homepage.html).

The 'test dataset 2' was curated to compile all remaining sequences of SH gene for which genotype information was not available in GenBank. It provided a larger set to check consistency of prediction using all the three methods, BLAST, MPA and RTD.

Genotype	Reference dataset	Test dataset 1	Test dataset 2
A	4	-	22
B	4	-	63
C	2	3	9
D	3	-	32
E	2	-	1
F	2	49	8
G	2	20	158
H	2	11	26
I	2	-	15
J	2	13	44
K	1	-	-
L	2	-	2
Total	28	96	380

Table 1 - Genotype-wise distribution of entries of MuVs in reference and test datasets.

Note that the genotypes of sequences in test dataset 2 were predicted using BLAST search against reference dataset and confirmed using NJ based molecular phylogeny analysis with bootstrap support.

RTD, a new alignment-free method for genotyping of Mumps viruses

The "return time (RT)" is defined as the time required for the reappearance of particular state without its appearance in between. In a given nucleotide sequence, RT for a particular nucleotide, say Adenine (A), will be defined as the number of non-A nucleotides between the successive appearances of A. The frequency distribution of all such observed RT is termed as RTD of that nucleotide.

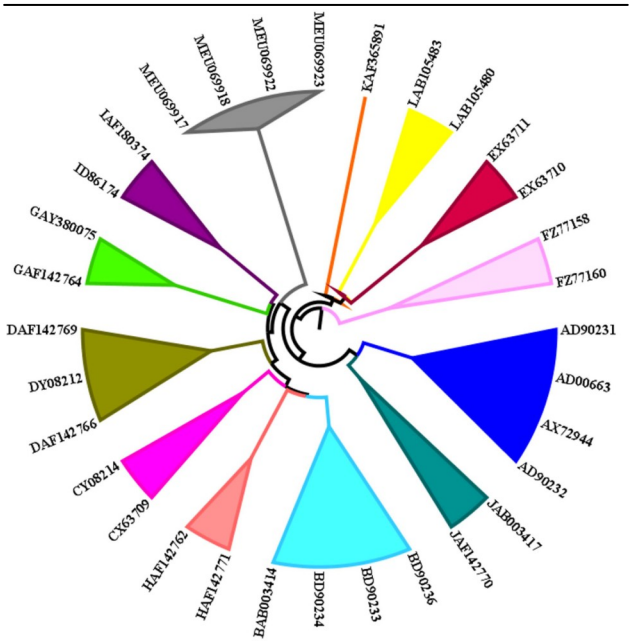

Figure 2 - The phylogenetic tree obtained using SH gene sequences for reference dataset of Mumps viruses using RTD-based method at k=4. Note that the known genotypes (A to L) as well as newly proposed genotype M are included in the analyses and are shown to form genotype-specific clades.
Note: The clades representing various genotypes are highlighted using various colours. The first letter in tip labels indicates the genotype and remaining letters constitute the GenBank accession number of the reference sequence. Tree is drawn using FigTree [http://tree.bio.ed.ac.uk/software/figtree/]

Each RTD will be statistically summarized using its two parameters viz. mean (μ) and standard deviation (σ). In similar way, the RTDs and their parameters for mononucleotide (where k=1) can be computed. In general, for chosen value of k (where k is a positive integer), there are 4^k possible k-mers. Hence, 4^k RTDs corresponding to 4^k possible k-mers with two statistical parameters (μ and σ) can be computed for every sequence. The pair wise distances between nucleotide sequences are then computed using equation (1) described in method section.

The distance matrix, thus obtained is given as an input to any distance based clustering method viz. Unweighted Pair Group Method with Arithmetic mean (UPGMA) [29], Neighbor-joining (NJ) [30] etc. to infer phylogenetic tree. Major steps in RTD-based genotyping includes (a) Computation of RTDs and their parameters of a query sequence(s) for an optimised value of k. (b) Derivation of distance of query sequence from each of the reference genotypes using the distance function. (c) Prediction of genotype of the query sequence based on the closest reference genotype.

Optimisation of the value of k for genotyping of mumps virus:

The optimisation of value of k is an important step in the RTD based method for clustering, phylogeny and genotyping. The reference dataset was used to optimise the value of k, which was varied from 1 to 5. For every value of k, the distance matrix and the clustering pattern in the resultant tree was analysed. It was found that for k=4, the entries in the reference dataset of MuV were found to cluster in accordance with respective genotypes (see Figure 2). At k=4, RTD-based method is able to group the members that belong to the same genotype as well as is able to resolve the various genotypes distinctly. Thus, for the composition of SH genes of MuVs, k=4 is chosen for purpose of genotyping. At k=4, for every SH gene sequence, 256 (4^4) RTDs were computed along with respective μ and σ. For k=4, it took about 2 seconds to infer the phylogenetic tree of reference dataset using proposed alignment-free method on a 32-bit operating system with 2.80 GHz processor and 4 GB of RAM using PERL scripts. It was noted that further increase in value of k does not bring any improvement in the results but adds to computational memory consumption and time.

Benchmarking & assessment of performance of methods for genotyping of mumps virus:

The reference dataset was used as a knowledgebase for all the three methods. The overall accuracy of BLAST and RTD was found to be 98.95% when tested using test dataset 1. The compilation of results for test dataset 1 is made available on the server under the link 'validation' from the navigation toolbar. As can be seen from the table listing the test dataset, the entry with GenBank accession FJ770566 has been annotated as genotype 'J' in GenBank, whereas it was found to be predicted as genotype 'G' by all the three methods. A closer look at the entry revealed that the strain information "MuVs-THA08-28-G", as per WHO naming convention indicates that it belongs to genotype 'G'. Hence, the GenBank annotation 'J' could possibly be a typographical error.

In order to compute sensitivity and specificity of RTD-based genotype predictions, dataset of true negatives was compiled. It consists of 40 entries which includes sequences of non-MuV SH, non-SH genes of MuV and non-SH non-MuV. Non-MuV SH sequences were included to account for putative false positives due to homology whereas non-SH genes from MuV help to eliminate compositional bias in the genome of MuVs. Non-MuV SH as well as non-SH non MuV were extracted from other members of the family *Paramyxoviridae*. When server was tested using the combined dataset of reference, test dataset 1 and dataset of true negatives, it was found to have 98.95% sensitivity (100% if annotation error in GenBank is taken into account) and 100% specificity. The higher sensitivity and specificity values can be attributed to carefully selected stringent cut-offs for both, BLAST and RTD-based approach. The genotype prediction using BLAST requires query coverage of 100% with sequence similarity ≥ 95% between the query and reference dataset sequence(s). These criteria are based on

sequence similarities between the strains of MuVs that belong to same genotype. The RTD-based method predicts the genotype of a query sequence based on two conditions. The first one is to identify the closest genotype from reference dataset based on minimum RTD distance. The next condition checks whether the computed distance satisfies the empirical cut-off precomputed for every genotype based on distance variations within respective genotype. These cut-offs were introduced to avoid false positive predictions and could be revised periodically upon availability of data on new strains/ genotypes.

Since test dataset 1 was devoid of the SH gene sequences for the genotypes A, B, D, E, I, K and L; it doesn't serve the purpose of all-inclusive comprehensive test dataset to evaluate the performance of newly proposed RTD-based approach. Considering these facts, the test dataset 2 was compiled where genotypes were predicted using BLAST and confirmed using MPA (using NJ) for the entries in GenBank with no genotype information available. Since test dataset 2 does not contain information of experimentally known genotypes, as the case for reference and test dataset 1, test dataset 2 was not intended to be used as the benchmarking dataset. However, test dataset 2 was used to check consistency of prediction of RTD-based genotyping method with MPA and BLAST as it serves a larger pool of SH genes consisting of 380 entries. The genotype-wise distribution of entries in test dataset 2 is also given in Table 1.

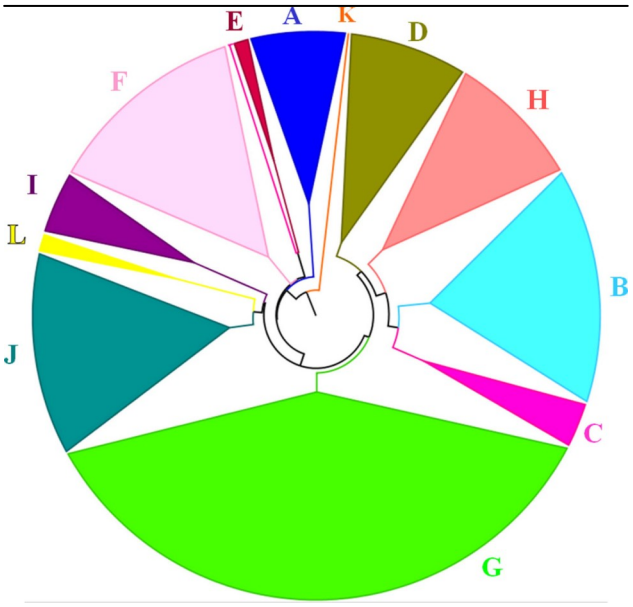

Figure 3 - The phylogenetic tree obtained for combined dataset consisting of reference and both the test datasets of Mumps viruses using proposed RTD-based method at k=4.

Note: The clades are labelled with genotypes. Tree is drawn using FigTree [http://tree.bio.ed.ac.uk/software/figtree/]

The accuracy of genotype prediction using RTD-based method was then estimated using leave-one-out cross-validation (LOOCV) for combined dataset consisting of reference, test datasets 1 and 2 (total 504). The method was found to be highly reliable with 99.80% accuracy. The RTD-based method failed to assign genotype to only one entry [GenBank: AF365891], which is the reference entry for genotype 'K', since LOOCV method was employed. The phylogenetic tree drawn using RTD-based distance measure for 504 entries is shown in Figure 3. As can be observed, all the sequences are clustering with respective genotypes except for genotype C. Two entries [GenBank: EU370206 and AJ272364] of genotype C cluster at the base of genotype E (see Figure 3), which has also been reported earlier [7]. Thus, LOOCV analysis and RTD based MPA reconfirm suitability of RTD as a method for genotyping of MuVs.

MuV genotyping server: How to use?

The server provides an easy to use common interface for the three independent methods for MuV genotyping. The BLAST and RTD based methods directly provide the predicted genotypes whereas MPA method provides the phylogenetic tree obtained using NJ method and prediction of genotype is subject to careful interpretation of the resultant tree by the user. The genotype prediction results using BLAST and RTD should be taken into consideration to interpret MPA results, especially if the query sequence is non-MuV SH, as MPA will return a phylogenetic tree for the any input data provided.

In case of BLAST-based genotyping, users are suggested to note the percent similarity and the query coverage with reference strain used for prediction of genotype. The server will not predict the genotype of the query sequence using the BLAST, if cut-offs for sequence similarity and query coverage are not met and the message "Genotype cannot be assigned" will be displayed.

The RTD-based method predicts the genotype for a query sequence only when the RTD distance cut-off criteria are satisfied. The server will otherwise flash the message "Genotype cannot be assigned". If the query sequence is predicted to be the genotype E, K or L using RTD-based method, the users are requested to confirm the genotype using BLAST and/or MPA. This is due to limited availability of experimentally validated data for these genotypes (see Table 1). Furthermore, it is a good idea to predict the genotype of a query sequence using all the three methods as consensus predictions provide higher confidence.

Discussion

The genotyping of viruses has become a routine procedure in virology laboratories due to medical and epidemiological importance of viruses. Hence different virus-specific automated genotyping tools have been developed to monitor genotypes circulating in populations worldwide, which also help to design strategies for controlling viral epidemics. The clinical relevance of genotypes with the type of disease and its severity also makes the genotyping an important task for devising treatment and control methods. The outbreaks of Mumps are still being reported even among highly vaccinated populations because of vaccine failure due to lack of cross protection between different genotypes [7], which is supported by the evidences of relationship between MuV genotype and neurotropism [17, 31-32].

Although sequence based genotyping is carried out routinely for MuV, there is no server or online utility for the genotyping of MuV whereas such servers are available

for other viruses viz., Hepatitis B, Hepatitis C, HIV etc. Thus, this study, for the first time reports development of genotyping server for MuVs as well as development of an alignment-free RTD based approach for genotyping of MuVs.

To re-confirm the proposition of new genotype M, the SH gene sequences of genotype M (GenBank accession numbers: EU069917, EU069918, EU069922 and EU069923) reported in [7] were added to reference dataset, which were then given as an input to RTD-based method to infer phylogenetic tree at k=4. The phylogenetic tree, thus obtained also supports the proposal of new genotype M in addition to existing A-L genotypes, as strains that belong to genotype M, form a distinct monophyletic clade from the other reference genotypes (see Figure 2). This proves that the RTD-based method has the capacity to detect the newly emerging genotypes of MuVs as well.

Conclusions

The present study, for the first time, provides a rapid, accurate and statistically robust novel alignment-free method for genotyping of MuVs. A web-based server for genotyping of mumps viruses is developed which provides prediction of genotypes using BLAST, MPA and RTD-based methods with a common interface. The RTD -based approach is relatively simple yet reliable and hence can be effectively used for the genotyping of other viruses as well.

Methods

Datasets

Reference dataset of Mumps viruses

The information of reference genotypes (A-L) of Mumps viruses was obtained from [7]. The complete SH gene sequences (316/318 bp) of the 28 reference strains were retrieved from GenBank database using the cited accession numbers [28]. Thus, this compilation contains multiple representative of each genotype except for genotype K, where only a single entry is available.

Test dataset 1 of Mumps viruses

All available MuV nucleotide sequences were obtained from GenBank (http://www.ncbi.nlm.nih.gov/genbank/) with the MuV taxonomy ID (txid11161) as a query. This resulted in total 910 nucleotide sequences. The dataset was further curated to compile sequences of complete SH gene (316/318 bp) leading to 476 entries, excluding reference dataset entries. Of these only 96 entries had explicit annotation in terms of genotype information in GenBank. Thus, test dataset 1 for MuV consists of 96 SH gene sequences. This dataset was used to validate the performance of BLAST, MPA and RTD based genotyping. It must be noted that this dataset was devoid of SH entries of genotypes A, B, D, E, I, K and L.

Test dataset 2 of Mumps viruses

The entries of complete SH sequences from GenBank where genotype annotation was not available are (476 – 96 = 380) and termed as Test dataset 2. The genotypes of these entries were predicted using BLAST searches against reference dataset using the criteria of 100% query coverage and ≥ 95% of sequence similarity with reference genotypes, as specified by WHO. The predicted genotypes of these sequences were also confirmed using molecular phylogeny analysis by employing NJ and Maximum likelihood method with 100 bootstrap as implemented in PHYLIP package (http://evolution.genetics.washington.edu/phylip.html). This extended dataset contains entries representing known genotypes A to L and was used to further assess performance of server for applicability of RTD for genotyping of MuV.

Method for Genotyping of MuVs

Following three sequence-based approaches are used for genotyping of MuVs using SH genes. The major steps in prediction of genotype using each of these approaches are shown in Figure 1.

BLAST similarity search:

The genotype of query sequence is predicted on the basis of BLAST similarity search against the reference dataset of MuVs [33]. The query coverage over 316 bases and >95% similarity with the reference strains are used as criteria, based on observed sequence similarity within genotypes.

Molecular phylogeny analysis using NJ method:

The query sequence along with the reference dataset is subjected for molecular phylogenetic analysis. The NJ method with 100 bootstrap as implemented in PHYLIP package is used [30]. The server provides phylogenetic tree, which shows clustering of query sequence with the reference genotypes. The genotype is predicted by the user based on phylogenetic inference supported by clustering proximity of query sequence with the monophyletic group of reference genotype and bootstrap values.

Return Time Distribution of k-mers:

This is a novel alignment-free method, which is developed in house and is reported earlier [27]. The method was suitably modified and optimised for size of k-mers such that the genotype-specific clustering was observed for reference dataset. The SH gene sequences of reference genotypes were represented as return time distributions of possible 4^k k-mers. Two statistical parameters (μ and σ) for each RTD are computed. Thus, for chosen value of k, each nucleotide sequence will be represented as numeric vector of size $2*4^k$ comprising μ and σ of 4^k possible RTDs in definite order. A sample computation of RTD and its parameters at k=1 is shown in Additional File 1. The distance matrix is computed, using the distance function given below (1), which has been reported earlier [27]. The distance matrix is then submitted for clustering using UPGMA method in PHYLIP package.

$$D_{ij} = (\sum [G_{ir\mu} - G_{jr\mu}]^2 + \sum [G_{ir\sigma} - G_{jr\sigma}]^2)^{1/2} \quad (1)$$

where, G_{ir} and G_{jr} represent the RTD of particular k-mer (for chosen value of k, there are 4^k possible RTDs) in two genomic sequences G_i and G_j. The μ and σ are their statistical parameters. Distance of query sequence from each

of the sequences of reference genotypes were computed using equation (1). The genotype of query sequence is then predicted based on closest reference genotype. Based on the observed distances amongst the strains of respective genotypes, cut-off criteria for every genotype, except for genotype K were devised. These are useful in improving accuracy of prediction and to eliminate false positive predictions. The sensitivity and specificity of RTD based prediction was also determined.

Assessment of reliability of RTD based method for genotyping of MuVs

Applicability of RTD-based method for the purpose of genotyping of MuVs was tested using leave-one-out cross validation method. The complete SH gene sequences in the reference and test datasets 1 were combined together and their RTDs were computed for the optimised value of k, which is 4. From this combined dataset, at a time, one SH gene was removed and its distance (1) from remaining sequences was computed. The 'removed' sequence was then assigned the genotype of the closest sequence in the remaining dataset. This exercise led to estimation of accuracy of the RTD-based method using leave-one-out cross validation method.

Implementation of method in the form of a server for genotyping of MuVs

The server, employing all the three methods described above (see Figure 1), for genotyping of MuVs is implemented using Apache, PHP and CGI and is available at http://bioinfo.net.in/muv/homepage.html

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

PSK developed, implemented and validated algorithm. UKK and MMK guided the projects. MMK provided statistical expertise. PSK and UKK drafted the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The assistance of Ms Shalika Sudhera, (Guru Nanak Dev University, Amritsar, Punjab, India) in compilation of reference datasets and testing of BLAST based server is greatly acknowledged. It was part of her master's project, carried out at the Bioinformatics Centre, University of Pune, India under the supervision of UKK. PSK acknowledges the BioInformatics National Certification (BINC) fellowship awarded by Department of Biotechnology (DBT), Govt. of India. UKK acknowledges infrastructural facilities and financial support under the Centre of Excellence (CoE) grant of Department of Biotechnology (DBT), Govt. of India.

Additional Files

Additional file 1 - A sample computation of RTD for k=1
www.immunome-research.net/journal/homeimmunomeojsfiles/journals/1/articles/23/supp/23-117-1-SP.ppt

References

1. Paterson RG, Lamb RA: **RNA editing by G-nucleotide insertion in mumps virus P-gene mRNA transcripts**. *J Virol* 1990, **64**(9):4137-4145.
2. Jin L, Beard S, Brown DW: **Genetic heterogeneity of mumps virus in the United Kingdom: identification of two new genotypes**. *J Infect Dis* 1999, **180**(3):829-833.
3. Jin L, Rima B, Brown D, Orvell C, Tecle T, Afzal M, Uchida K, Nakayama T, Song JW, Kang C *et al*: **Proposal for genetic characterisation of wild-type mumps strains: preliminary standardisation of the nomenclature**. *Arch Virol* 2005, **150**(9):1903-1909.
4. Kim SH, Song KJ, Shin YK, Kim JH, Choi SM, Park KS, Baek LJ, Lee YJ, Song JW: **Phylogenetic analysis of the small hydrophobic (SH) gene of mumps virus in Korea: identification of a new genotype**. *Microbiol Immunol* 2000, **44**(3):173-177.
5. Tecle T, Bottiger B, Orvell C, Johansson B: **Characterization of two decades of temporal co-circulation of four mumps virus genotypes in Denmark: identification of a new genotype**. *J Gen Virol* 2001, **82**(Pt 11):2675-2680.
6. Echevarría JE, Castellanos A, Sanz JC, de Aragón MVM, Rey IP, Mosquera M, de Ory F, Royuela E: **Mumps Virus Genotyping: Basis and Known Circulating Genotypes**. *Open Vaccine Journal* 2010, **3**:37-41.
7. Santos CL, Ishida MA, Foster PG, Sallum MA, Benega MA, Borges DB, Correa KO, Constantino CR, Afzal MA, Paiva TM: **Detection of a new mumps virus genotype during parotitis epidemic of 2006-2007 in the state of Sao Paulo, Brazil**. *J Med Virol* 2008, **80**(2):323-329.
8. Folb PI, Bernatowska E, Chen R, Clemens J, Dodoo AN, Ellenberg SS, Farrington CP, John TJ, Lambert PH, Macdonald NE *et al*: **A global perspective on vaccine safety and public health: the Global Advisory Committee on Vaccine Safety**. *Am J Public Health* 2004, **94**(11):1926-1931.
9. Furesz J: **Safety of live mumps virus vaccines**. *J Med Virol* 2002, **67**(3):299-300.
10. Galazka AM, Robertson SE, Kraigher A: **Mumps and mumps vaccine: a global review**. *Bull World Health Organ* 1999, **77**(1):3-14.
11. Afzal MA, Buchanan J, Dias JA, Cordeiro M, Bentley ML, Shorrock CA, Minor PD: **RT-PCR based diagnosis and molecular characterisation of mumps viruses derived from clinical specimens collected during the 1996 mumps outbreak in Portugal**. *Journal of Medical Virology* 1997, **52**(4):349-353.
12. Briss PA, Fehrs LJ, Parker RA, Wright PF, Sannella EC, Hutcheson RH, Schaffner W: **Sustained transmission of mumps in a highly vaccinated population: assessment of primary vaccine failure and waning vaccine-induced immunity**. *J Infect Dis* 1994, **169**(1):77-82.
13. Hersh BS, Fine PE, Kent WK, Cochi SL, Kahn LH, Zell ER, Hays PL, Wood CL: **Mumps outbreak in a highly vaccinated population**. *J Pediatr* 1991, **119**(2):187-193.
14. Orvell C, Alsheikhly AR, Kalantari M, Johansson B: **Characterization of genotype-specific epitopes of the HN protein of mumps virus**. *J Gen Virol* 1997, **78 (Pt 12)**:3187-3193.
15. Pons C, Pelayo T, Pachon I, Galmes A, Gonzalez L, Sanchez C, Martinez F: **Two outbreaks of mumps in children vaccinated with the Rubini strain in Spain indicate low vaccine efficacy**. *Euro Surveill* 2000, **5**(7):80-84.
16. Richard JL, Zwahlen M, Feuz M, Matter HC: **Comparison of the effectiveness of two mumps vaccines during an outbreak in Switzerland in 1999 and 2000: a case-cohort study**. *Eur J Epidemiol* 2003, **18**(6):569-577.
17. Tecle T, Johansson B, Jejcic A, Forsgren M, Orvell C: **Characterization of three co-circulating genotypes of the small hydrophobic protein gene of mumps virus**. *J Gen Virol* 1998, **79 (Pt 12)**:2929-2937.
18. Vandermeulen C, Roelants M, Vermoere M, Roseeuw K, Goubau P, Hoppenbrouwers K: **Outbreak of mumps in a vaccinated child population: a question of vaccine failure?** *Vaccine* 2004, **22**(21-22):2713-2716.
19. Kulkarni-Kale U, Ojha J, Manjari GS, Deobagkar DD, Mallya AD, Dhere RM, Kapre SV: **Mapping antigenic diversity and strain specificity of mumps virus: a bioinformatics approach**. *Virology* 2007, **359**(2):436-446.
20. Nojd J, Tecle T, Samuelsson A, Orvell C: **Mumps virus neutralizing antibodies do not protect against reinfection with a heterologous mumps virus genotype**. *Vaccine* 2001, **19**(13-14):1727-1731.
21. CDC: **Update: multistate outbreak of mumps--United States, January 1-May 2, 2006**. *MMWR Morb Mortal Wkly Rep* 2006, **55**(20):559-563.
22. WHO: **Global status of mumps immunization and surveillance**. *Wkly Epidemiol Rec* 2005, **80**(48):418-424.
23. Takahashi M, Nakayama T, Kashiwagi Y, Takami T, Sonoda S, Yamanaka T, Ochiai H, Ihara T, Tajima T: **Single genotype of measles virus is dominant whereas several genotypes of mumps virus are co-circulating**. *J Med Virol* 2000, **62**(2):278-285.
24. Gnaneshan S, Ijaz S, Moran J, Ramsay M, Green J: **HepSEQ: International Public Health Repository for Hepatitis B**. *Nucl Acids Res* 2007, **35**(suppl_1):D367-370.
25. Myers RE, Gale CV, Harrison A, Takeuchi Y, Kellam P: **A statistical model for HIV-1 sequence classification using the subtype analyser (STAR)**. *Bioinformatics* 2005, **21**(17):3535-3540.
26. de Oliveira T, Deforche K, Cassol S, Salminen M, Paraskevis D, Seebregts C, Snoeck J, van Rensburg EJ, Wensing AM, van de Vijver DA *et al*: **An automated genotyping system for analysis of HIV-1 and other microbial sequences**. *Bioinformatics* 2005, **21**(19):3797-3800.
27. Kolekar PS, Kale MM, Kulkarni-Kale U: **`Inter-Arrival Time' Inspired Algorithm and its Application in Clustering and Molecular Phylogeny**. *AIP Conference Proceedings* 2010, **1298**(1):307-312.
28. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW: **GenBank**. *Nucleic Acids Research* 2011, **39**(suppl 1):D32-D37.
29. Sneath PHAS, R. R. : **Numerical Taxonomy: the Principles and Practice of Numerical Classification.** San Francisco: W. H. Freeman.; 1973.
30. Saitou N, Nei M: **The neighbor-joining method: a new method for reconstructing phylogenetic trees**. *Mol Biol Evol* 1987, **4**(4):406-425.
31. Rubin S, Mauldin J, Chumakov K, Vanderzanden J, Iskow R, Carbone K: **Serological and phylogenetic evidence of monotypic immune responses to different mumps virus strains**. *Vaccine* 2006, **24**(14):2662-2668.

32. Saito H, Takahashi Y, Harata S, Tanaka K, Sato H, Suto T, Yamada A, Yamazaki S, Morita M: **Cloning and sequence analysis of genomic RNA from the mumps virus strain with high incidence of aseptic meningitis**. *Microbiology and Immunology* 1998, **42**:133–137.
33. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool**. *J Mol Biol* 1990, **215**(3):403-410.