Lin et al. BMC Genomics 2015, 16(Suppl 12):S8
http://www.biomedcentral.com/1471-2164/16/S12/S8

BMC Genomics

RESEARCH Open Access

Precise genotyping and recombination detection of Enterovirus

Chieh-Hua Lin[1,2†], Yu-Bin Wang[3,4†], Shu-Hwa Chen[3†], Chao Agnes Hsiung[1,2*], Chung-Yen Lin[1,3,5*]

From Joint 26th Genome Informatics Workshop and Asia Pacific Bioinformatics Network (APBioNet) 14th International Conference on Bioinformatics (GIW/InCoB2015)
Tokyo, Japan. 9-11 September 2015

Abstract

Enteroviruses (EV) with different genotypes cause diverse infectious diseases in humans and mammals. A correct EV typing result is crucial for effective medical treatment and disease control; however, the emergence of novel viral strains has impaired the performance of available diagnostic tools. Here, we present a web-based tool, named EVIDENCE (**E**ntero**V**irus **I**n **DE**ep conception, http://symbiont.iis.sinica.edu.tw/evidence), for EV genotyping and recombination detection. We introduce the idea of using mixed-ranking scores to evaluate the fitness of prototypes based on relatedness and on the genome regions of interest. Using phylogenetic methods, the most possible genotype is determined based on the closest neighbor among the selected references. To detect possible recombination events, EVIDENCE calculates the sequence distance and phylogenetic relationship among sequences of all sliding windows scanning over the whole genome. Detected recombination events are plotted in an interactive figure for viewing of fine details. In addition, all EV sequences available in GenBank were collected and revised using the latest classification and nomenclature of EV in EVIDENCE. These sequences are built into the database and are retrieved in an indexed catalog, or can be searched for by keywords or by sequence similarity. EVIDENCE is the first web-based tool containing pipelines for genotyping and recombination detection, with updated, built-in, and complete reference sequences to improve sensitivity and specificity. The use of EVIDENCE can accelerate genotype identification, aiding clinical diagnosis and enhancing our understanding of EV evolution.

Introduction

The *Enterovirus* (EV) genus (family ***Picornaviridae***) contains twelve species, including *Enterovirus A* to *H* and *J*, and *Rhinovirus A* to *C*. These viruses cause a wide range of diseases in humans and mammals. The single-stranded RNA genome of EV contains a single open reading frame (ORF) flanked by 5' and 3' untranslated regions (UTRs). The ORF encodes a polyprotein, which is further processed into 11 proteins: VP1-4 (structural proteins), and 2A-2C and 3A-3D (non-structural proteins) [1]. The genetic diversity of EVs arises from the accumulation of single-base changes during viral propagation, as well as from recombination events that cause genome segments to be swapped between or within EV genotypes. To date, 308 *Enterovirus* genotypes have been reported (http://www.picornaviridae.com/enterovirus/enterovirus.htm, on 2015/04), and the number is rising.

Different enterovirus genotypes cause different clinical symptoms [1]. Classical serotyping methods, such as serum neutralizing test and immunofluorescent assay, are not sufficient to specify all genotypes. For example, Tsao *et al.* (2010) reported that 15~30% of EV isolates failed to be serotyped in Taiwan [2]. To overcome this problem, many clinicians have turned to sequence-based molecular typing methods, which assign viral genotypes based on nucleotide sequences; such techniques are more successful at resolving EV isolates to the corresponding genotype, and also provide rapid diagnosis [3]. The *VP1* capsid-coding region has been suggested to be the most suitable region for EV genome genotyping [4,5].

* Correspondence: hsiung@nhri.org.tw; cylin@iis.sinica.edu.tw
† Contributed equally
[1]Division of Biostatistics and Bioinformatics, Institute of Population Health Sciences, National Health Research Institutes, Zhunan, Miaoli County, 35053, Taiwan
Full list of author information is available at the end of the article

BioMed Central

© 2015 Lin et al. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

Table 1

Species	GenBank Nomenclature	Assigned by EVIDENCE	Sequences with Congruent Assignment (%*)
Enterovirus A	15875	16057	15794 (99.49%)
Enterovirus B	16432	16502	16238 (98.82%)
Enterovirus C	8224	8202	8146 (99.05%)
Enterovirus D	1031	1132	1031 (100%)
Enterovirus E	73	43	41 (56.16%)
Enterovirus F	5	43	5 (100%)
Enterovirus G	196	216	190 (96.94%)
Enterovirus H	10	10	10 (100%)
Enterovirus J	9	11	5 (55.56%)
Rhinovirus A	2733	5088	2697 (98.68%)
Rhinovirus B	608	1101	592 (97.37%)
Rhinovirus C	2309	3690	2259 (97.83%)
unclassified Enterovirus unclassified Rhinovirus	4957	367	119 (7.4%)

Remark: * The percentage is calculated as follows: $\frac{\text{Sequences with Congruent Assignment}}{\text{Sequences in GenBank nomenclature}} \times 100\%$

In addition, the 5'UTR [6,7], *VP2* [8,9], *VP4* [10,11] and *3D* [11,12] regions, as well as combinations of more than two regions, including the 5'UTR and *VP4/VP2* [13], the 5'UTR and *VP1*[14], and *VP1* and *3D* [15], have been evaluated for their usefulness for improving the sensitivity and specificity of diagnosis. However, incongruent results may be obtained from different typing methods based on either single or multiple coding regions of the genome [16-19].

At present, there are two EV genotyping tools: enterovirus genotyping tool (version 0.1; National Institute of Public Health and the Environment (RIVM), the Netherlands) [20] and the genotyping tool of the NCBI [21]. Both of these resolve genotypes on the *VP1* region, and disregard the rest of the EV genome. This approach limits the ability to distinguish between strains that originated from recombination events. Moreover, EV genotype reference sequences are never updated in these libraries.

A fast, highly sensitive, and specific molecular typing tool is essential for clinical diagnosis and medical treatment. In this study, we developed a web tool, EVIDENCE (EnteroVirus In DEep coNCEption), a workbench for phylogenetic-based genotyping and recombination detection in EV genomes. Up-to-date EV classification data, nomenclature, and GenBank accession numbers for each genotype's prototype sequence were collected from the Picornaviridae Study Group website at http://www.picornaviridae.com/ [22], and these were combined with sequences collected from the NCBI to build the genotyping reference set (GTRefSet). Phylogenetic inference was used to resolve the best-fit genotype of novel EV sequences using single or multiple genomic regions of interest. For detection of recombination events, the closeness between the suspected recombinant and reference sequences was measured as bootscanning supports by the phylogenetic method, and as sequence similarity by the distance method. The pipeline design enables users to seamlessly run recombination analyses with guidance for the choice of references. Furthermore, we revised the EV sequences in GenBank to standardize the nomenclature and to clarify genotype assignments. The collected sequences were built into the database, and can be retrieved in an indexed catalog or be searched for by keyword or sequence similarity.

EVIDENCE is the first web-based tool providing pipelines for genotyping and recombination detection based on both sequence context and phylogenetic inference. Furthermore, EVIDENCE uses the most complete and regularly updated reference sequences to maintain high sensitivity and specificity, thereby accelerating genotype identification in clinical diagnosis and enhancing our understanding of EV evolution. EVIDENCE is available at http://symbiont.iis.sinica.edu.tw/evidence.

Materials and methods

Reference sequence sets

Nucleotide sequences of EV prototype strains (Additional file 1: Table S1) listed in the Picornaviridae Study Group website (http://www.picornaviridae.com/) were fetched from GenBank database. If the complete genome of a prototype strain was not available, we collected sequences of all the other available regions instead. For a genotype without a reference prototype assignment, the longest and/or earliest reported sequence was selected as the genotype's representative reference. Finally, 396 nucleotide sequences are selected to build

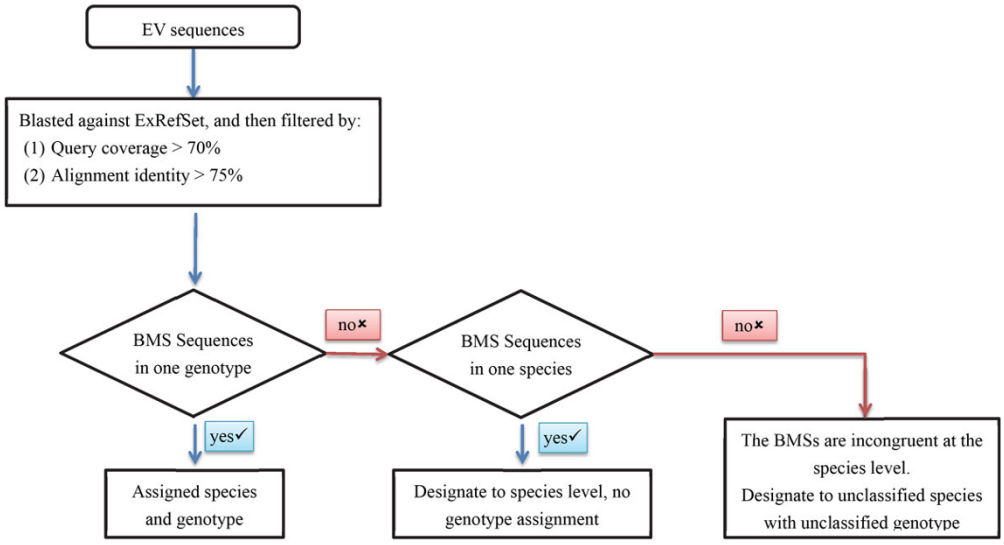

Figure 1 Genotype re-classification procedures.

the three hundred and eight prototype models (the genotyping reference set, GTRefSet). Furthermore, we expanded the prototype GTRefSet to the extended reference set (ExRefSet, Additional file 1: Table S2) by including the sequences that is highly similar (sequences identity >75%) [23] to the representative prototype sequence of the same genotype. GTRefSet was the core reference set for phylogenetic analysis and recombination analysis, and ExRefSet was used for automatic re-assignment of EV sequence genotypes.

Re-classification of all EV sequences in GenBank

Nucleotide sequences from the *Enterovirus* genus deposited in GenBank (release 206) were collected, with the exception of sequences that are denoted as being from environmental samples. To unify the sequence taxonomical nomenclature in accordance with the decisions of the International Committee on Virus Taxonomy (ICTV) 2014 [24,25] and to identify sequences with potentially misassigned genotypes, the collected EV sequences were run through the following procedures (Figure 1). First, we performed a BLAST search (dc-megablast, BLAST+ package, version 2.2.28) [26] using a query sequence q against reference sequence x in ExRefSet (E), and extracted the following information: BLAST raw score ($braw_{qx}$), % of query coverage per reference sequence (cov_{qx}), and % identity ($identity_{qx}$). Cq is a subset of E with all high blast-scored alignments to q,

$$Cq = \{x \in E \mid cov_{qx} > 70\%,\ identity_{qx} > 75\%\}.$$

Let $rank(braw_{qx})$,$rank(cov_{qx})$,$rank(identity_{qx})$ be the ascending sorting rank of BLAST raw score, alignment coverage, and identity between q and x in Cq, respectively. We build a set of the rank sums rS_{Cq} where

$$rs_{Cq} = \{rank(braw_{qx}) + rank(cov_{qx}) + rank(identity_{qx}) \mid x \in Cq\}.$$

Finally, the best mixed-ranking score (**BMS**) for the query q is defined as the largest rank sum r in rS_{Cq},

$$BMS_q = \max(r \mid r \in rs_{cq}).$$

Sequences are assigned to the genotype/species of their BMS references if the BMS is only reached by one reference sequence or multiple references with consistent genotype/prototype assignment. If the BMS of a query sequence is reached by two or more different genotypes in one species, the nomenclature assignment of this sequence is set to the species level. Query sequences are assigned as "unclassified" species with "unclassified" genotype if a BMS from a highly related sequence is not detected in ExRefSet (e.g., ***Cq*** is an empty set), or multiple references reach the BMS but are incongruent at the species level.

The rationales of using three parameters in the mixed-ranking score function are described below. The expectation value (E-value) given by BLAST is often used to indicate the significance of an alignment, and is often inferred to the homology/similarity relationship of the hit to the query sequences. Calculation of E-value is affected by the content of searching database and the length of the matching segment. For example, a small E-value may be granted to a short region in high sequence similarity, leading to false positive results if we use an E-value cut-off for selecting sequence matches.

In contrast, the BLAST raw score is directly derived from alignment segments can eliminate this artifact [27]. Thus, we use BLAST raw score instead. The BLAST raw score is also depended on the scoring parameters being used (*i.e.*, reward for matching base and penalty for mismatching base/gap) [28,29]. In our case, the scoring parameters are using the defaults in discontiguous megablast (dc-megablast), says, match = 2, mismatch = −3, gap open = −5, and gap extension = −2. This empirical setting is optimized for catching highly similar sequences with gap allowance. Then, we calculated the BLAST raw score for the query sequence q to the reference x in the reference set, and granted the rank score $Braw_{qx}$ to each q-x pair by the sorting order of the score.

Furthermore, we adopted the percentage of identity and the coverage of the alignment in the algorithm to address the importance of the conservation of the base components and the overall alignments of any two sequences in comparison. These two indexes are necessary for identifying the closest homologous sequences. The numeric values of the identity and coverage were transformed to ranked scores, in which we can sum up for the importance evaluated by three different scoring scheme in a normalized scale. Therefore, the mixed-ranking score is based on a measure of sequence similarity, adding weights on the importance of the quality of alignment.

Phylogenetic-based genotyping procedure

EV prototypes in GTRefSet were further segmented into the thirteen genomic regions (5'UTR, 11 segments corresponding to the mature peptide-coding regions, and 3'UTR) according to the coordinates described in their original GenBank documents. For sequences without this information, we used NetPicoRNA [30] to predict the polyprotein cleavage sites and then manually curated ambiguous boundaries. The phylogenetic-based genotyping procedure consists of two steps: ***scan region*** and ***phylogenetic analysis***. In the ***scan region*** step, an input query sequence is compared to the GTRefSet members using BLAST. The index of the mixed-ranking score of each genome region of each reference to the query is then calculated (Figure 2). The region scan result table is sorted by the sum of mixed-ranking score $\boldsymbol{r}$ of all available genome regions; the reference sorting order is changed dynamically according to the genome regions selected.

Executing the ***phylogenetic analysis*** requires one query sequence and at least two reference strain prototype sequences. The genome regions of interest are cropped from the query sequence and concatenated for phylogenetic analysis. Multiple sequence alignments of the concatenated query and reference fragments are performed by Clustal Omega version 1.2.0 [31]. The phylogenetic inference was done in by PhyML 3.0 [32] (model GTR + G + I) with user defined bootstrap iterations. The circular phylogenetic tree topology is generated by jsPhyloSVG library [33].

Recombination event detection

We implemented Bootscan [34] and SimPlot [35] methods in EVIDENCE to detect possible recombination events and graphically present the results. Briefly, multiple sequence alignments between the query and reference genotype sequences (built into GTRefSet or custom uploaded) were performed using Clustal Omeg. The distance and similarity between sequences are calculated using DNADIST in PHYLIP (package version 3.5c) [36] and phylogenetic inference is analyzed using the Neighbor-joining (NJ) method with the Kimura two-parameters substitution model. Parameters, including sliding window size, step size, bootstrapping iterations, and the use of trimA1 (version 1.2) [37] to remove gaps, are adjustable. For each sliding window, the bootstrap value of the reference that was the first neighbor clustered to the query was used to derive the percentage of bootstrap support. Finally, profiles of query-reference closeness, evaluated by similarity or percentage of bootstrap support of each query-reference pair in each sliding window segment, were plotted along the EV genome. A clear crossing-over of two query-reference profiles with sharp slopes suggests a swap of best-fit reference and the presence of a nearby recombination breaking point.

System framework

EVIDENCE (http://symbiont.iis.sinica.edu.tw/evidence/) is constructed on an open-source Linux Arch (version), Nginx (version 0.12.4), and SQLAlchemy and SQLite relational database (version 3.8.4.3) structure. Graphical visualization was provided using Canvas and SVG library. Scripts for joining software packages to seamless pipelines were written in Perl and Python. The whole system is run in a virtual machine (CPUs of 2.27GHz, 8 cores, 16 GB RAM) located in the Institute of Information Science, Academia Sinica, Taiwan.

Results and discussion

The usage of EVIDENCE

EVIDENCE is a searchable database for updated and re-classified EV sequences and a workbench for EV genotyping and recombination detection (Figure 2).

Database search

A total of 54,790 up-to-date EV sequences were curated and built into an indexed category of virus species, genotype, and epidemiological annotations (host, continent, country, and the reported year). This database can be browsed through the hierarchical structure or searched by keywords (Figure 2A, panel 1) or sequence similarity (e.g., BLASTN, TBLASTN, or TBLASTX) (Figure 2A, panel 2).

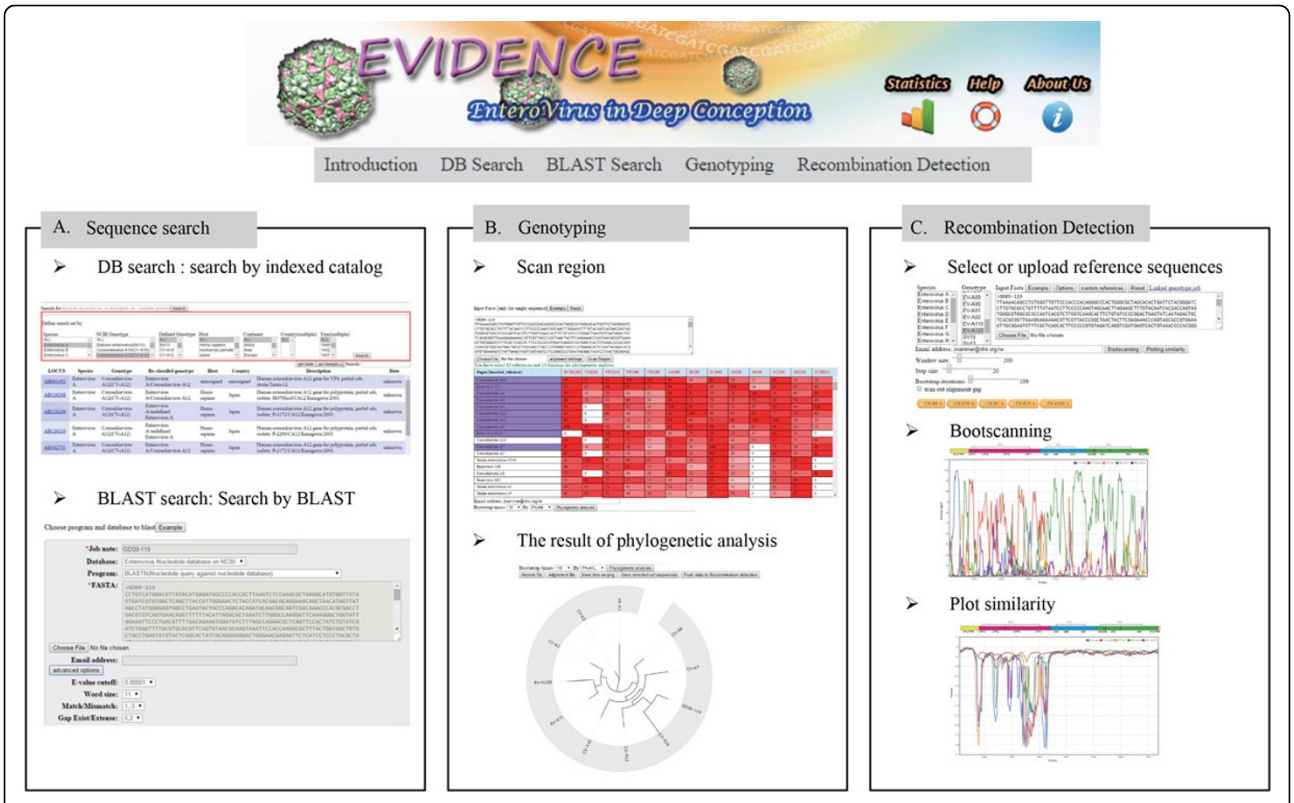

Figure 2 EVIDENCE interface. (A) **Sequence Search**. In DB search, keywords or multiple conditions can be used to search for related sequences. In BLAST search, single or multiple sequences are used to BLAST against all EV sequences to search for similar sequences.
(B) **Genotyping**. First, the query sequence is compared to the GTRefSet (the scan region step). The result table contains indices of mixed-ranking scores presenting segment similarity. References for phylogenetic analysis are selected from this table. (C) **Recombination Detection**. The potential recombination events can be detected through observing swaps of the most similar reference (plot similarity) or changes of the relatively closest neighbor (bootscanning) in the genome segments.

For both sequence search functions, the basic features extracted from GenBank records and reclassified genotypes are shown in table format. The search results, including the brief table of reported entries and the FASTA file of hits, are made available for download.

Genotyping

EVIDENCE matches query sequences to references in GTRefSet using the BLAST algorithm (Figure 2B, panel 1). To perform genotyping, the query nucleotide sequence in FASTA format is pasted or uploaded through the sequence input interface, and then submitted by clicking on 'scan region'. Scores for the query to each EV genomic region of each prototype are calculated and displayed in the results table. The table functions like a flexible input interface for the next phylogenetic analysis step. Clicking on the table column (the genome region) or table row (sequence title) will select the region or the references, respectively. The table is decreasingly sorted by the mixed-ranking score of a single column, or by the sum of mixed-ranking scores of all the selected columns (Figure 2B). The sorting order implies the relative fitness of each query-reference pair with respect to the region(s) of interest.

The phylogenetic analysis step calculates the relatedness between the query and the selected references. A tree topology is generated by PhyML with adjustable bootstrap iterations (default: 100 iterations). Phylogenetic analysis outputs, including the tree topology as a newick file and a png file, multiple sequence alignment, and the selected reference sequences, are made available for download (Figure 2B).

Recombination

The basic principle for detecting potential recombination events is to segment the whole EV genome into small overlapping segments, in order to identify swaps of the most similar reference (plot similarity) or changes in the relatively closest neighbor (bootscanning) in the successive segments; similarity is derived from the sequence distance for each query-reference pair, and sequence neighbors are determined based on the percentage of bootstrap iterations supporting the reference as the closest neighbor to the query.

The recombination analysis takes a single sequence query, accompanied with three or more references to give adequate estimations. The reference strains are selected from those built into GTRefSet and/or from uploaded 'custom references'. The parameters for phylogenetic inference include sliding window (default: 200), step size (default: 20), bootstrapping (default: 100 iterations), and whether or not gaps in the multiple sequence alignment are trimmed (default: no trimming); all of these parameters are adjustable (Figure 2C, panel 1).

Plotting of the similarity or bootstrapping results is optional. The EV genome diagram is plotted on top to help visualize the location of recombination events. Dynamic figures are used to present the data and allow the user to zoom into/out of the plot. The value of each sliding window on the plot is shown via mouseover events, and the user may zoom in on a specific region of the plot by cropping the region through mouse dragging (Figure 2C, panel 2). The bootscanning/similarity plot, multiple alignment table, and bootstrap/similarity value table can be downloaded.

The choice of reference strains for recombination detection is crucial. Using inappropriate reference strains to identify the recombination region may eliminate the significance of the bootscanning result, and increase the noise of recombination breakpoint determination. In EVIDENCE, users can start the analysis from genotyping. References in GTRefSet are evaluated for the fitness of each genome region, which can help users select appropriate regions. The data, including the query sequence and at least three selected references, are redirected by clicking "push data to recombination detection" after the phylogenetic analysis step.

It is worth noting that using too many reference sequences in a recombination detection analysis may return an insignificant bootscanning plot or a messy similarity plot. We suggest that analysis should begin with less than ten references, and then the non-informative ones should be removed to improve the resolution.

Statistics of GTRefSet and ExtRefSet

The genotyping reference set (GTRefSet) of 308 reported prototype models includes 204 full genomes, 17 partial genomes, and 69 partial segments (data in 2015/04). Eighteen genotypes are not related to any reference sequence in the database and will not be included in the present EVIDENCE reference sets. Figure 3 show statistics for GTRefSet, by species and by genome region. Of the 13 genome regions, VP1 is the most sequenced; the VP1 regions of almost every prototype were reported. By selecting sequences with an identity >75% to the representative prototype sequence of the same genotype, 143 full genomes, 26 partial genomes, and one fragment were appended to form ExtRefSet (Supplementary Table S2, See Additional file 1).

Updating the classification of sequences using a BLAST-based method

To date, 54,790 EV sequences are deposited in GenBank, but the genotype assignment of these sequences in GenBank may be inaccurate. One possible reason for any inaccuracies is that the genotype of a sequence is assigned by individual researchers upon submission. However, differences in the typing methods used may lead to different conclusions. As mentioned above, serological assays and molecular typing may generate conflicting results if the examined strain has undergone a recombination event that changed the viral genome and the correlation of epitopes with the reference virus strain; such inconsistency is independent of the specificity of the antibody used to resolve strains at the genotype level. A second possible reason is that the nomenclature and the virus classification system are not synchronized between reference databases. Virus classification for NCBI taxonomy is based on that of the International Nucleotide Sequence Database Collaboration (INSDC) [38]. However, classification of viruses is less stable than that of other organism kingdoms due to the rapidly evolving nature of these infectious life forms. Irregular updates have led to the inconsistent nomenclature and taxonomic classification used by GenBank and the International Committee on Virus Taxonomy (ICTV). Moreover, the updated viral classification scheme has not been applied to sequences previously deposited in the database.

In order to unify the nomenclature and taxonomic scheme of EV in accordance with the release by the ICTV (2014), we revised the classification of EV sequences collected from GenBank. To include the intragenic genetic variation of a genotype, we used ExRefSet instead of GTRefSet to enhance the sensitivity of the BLAST-based method. The re-classification results showed high congruence with GenBank classification (Table 1; details on the congruence of each genotype are provided in Additional file 1: Table S3). For example, sequences in the well-studied and largest genotype, EV-A71, comprised 16.12% of the EV-related sequences in GenBank, and over 99% sequences in this genotype remained in their original assignment after re-classification. In addition, 633 EV sequences were newly assigned to EV-A71. About 1/3 of these re-classified sequences were isolated from the EV-A71 strain or identified as EV-A71 [39-49], but were assigned to the species level (Enterovirus A) or misassigned to other genotypes; 68.8% of the remaining sequences (308 sequences) were typed using the amplicons derived from universal 5'UTR primer for all enteroviruses (Table 2) [14,50-55]. It have been reported that the 5'

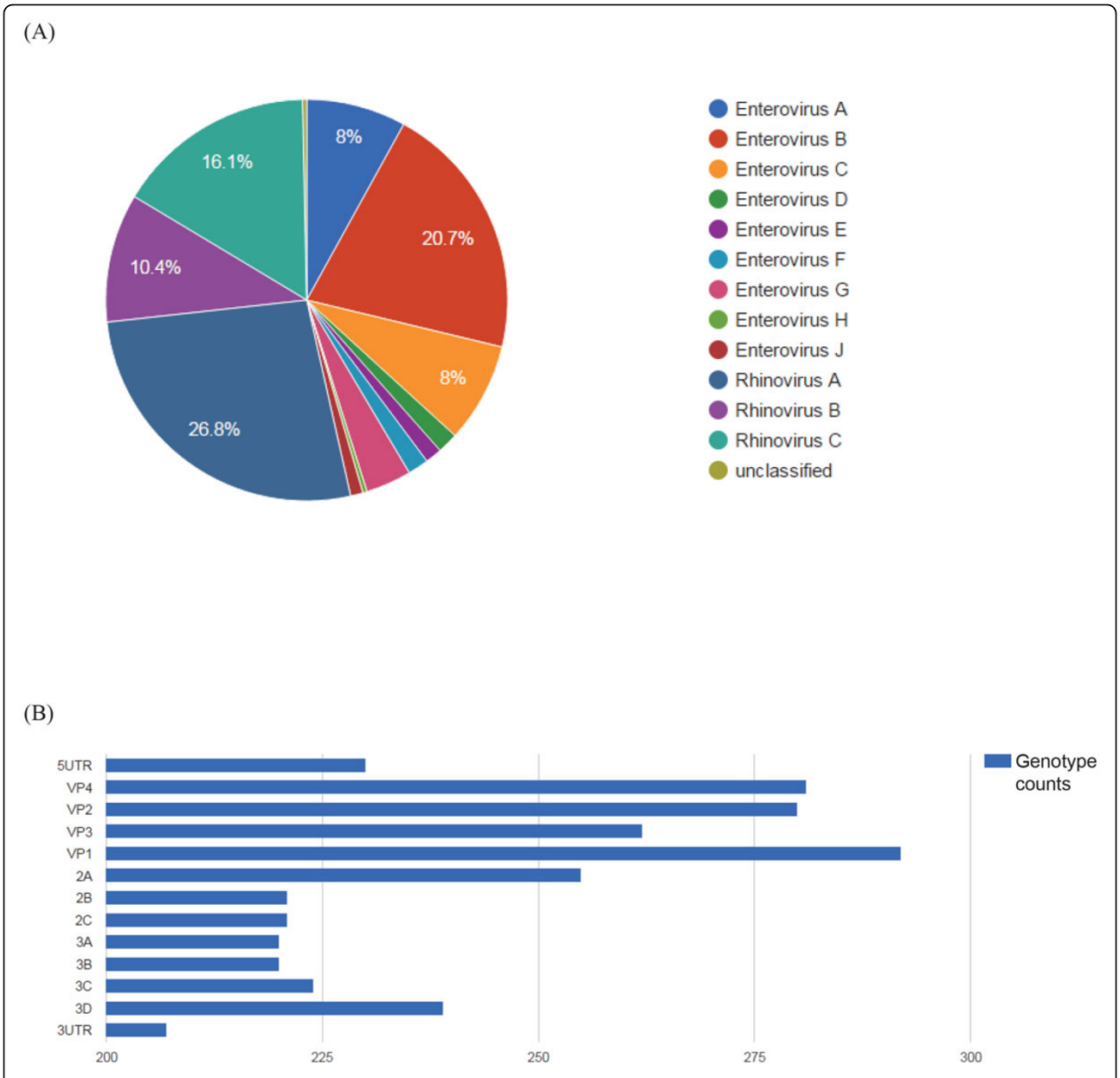

Figure 3 Statistics of GTRefSet. (A) The species composition of the database. Enterovirus B and Rhinovirus A collectively contribute to about half of the total known genotypes. (B) Genome region view.

UTR regions of *Enterovirus A* and *Enterovirus B* are indistinguishable due to their highly conserved secondary structures [7,56,57]. Therefore, the use of certain genotyping techniques may generate inadequate genotype assignments, resulting in the inconsistencies in assignment observed between those of the original records and those obtained with EVIDENCE.

We further compared the performance of our reclassification procedure with that of the RIVM EV genotyping tool (http://www.rivm.nl/mpf/enterovirus/typingtool), a phylogenetic-based typing tool which uses a partial VP1 region and the neighbor-joining (NJ) method with HKY85 or TamNei model, using the sequences identified and typed by traditional neutralization methods [58-61] as a gold standard. As shown in Table 3 EVIDENCE generated more consistent results with antigenic typing. For example, one untyped human rhinovirus (HRV) partial VP1 sequence (GenBank accession number AF152281) has 89.37% similarity to HRV-A31; RIVM genotyping tools assigned it to the HRV species, whereas EVIDENCE returned HRV-A31 as the genetically closest genotype.

Table 2

Original Species	Original Genotype	#Sequences	Region[2] (sequence number)
Enterovirus A	Coxsackievirus A2	11	5'UTR (1), 2C (2), 3D (8)
Enterovirus A	Coxsackievirus A3	2	5'UTR (1), 3D (1)
Enterovirus A	Coxsackievirus A4	13	2C (2), 3D (11)
Enterovirus A	Coxsackievirus A5	2	3D (2)
Enterovirus A	Coxsackievirus A6	74	5'UTR (69), 2C (2), 3D (3)
Enterovirus A	Coxsackievirus A7	7	2C (2), 3D (5)
Enterovirus A	Coxsackievirus A8	6	2C (1), 3D (5)
Enterovirus A	Coxsackievirus A10	94	5'UTR (85), VP4-VP2 (2), 2C (3), 3D (4)
Enterovirus A	Coxsackievirus A14	3	VP1 (1), 2C (1), 3D (1)
Enterovirus A	Coxsackievirus A16	1	3D (1)
Enterovirus A	Enterovirus A76	1	3D (1)
Enterovirus A	NA[1]	54	5'UTR (28), 2BC (2), 3D (24)
Enterovirus B	Echovirus 4	3	5'UTR (3)
Enterovirus B	Echovirus 9	5	5'UTR (5)
Enterovirus B	NA[1]	10	5'UTR (10)
unclassified Enterovirus	UE[1]	22	5'UTR (10), VP4 (3), VP1 (3), 3D (6)

Remarks:

1. Abbreviations in use: NA, Not assigned; UE, including all classes under unclassified Enterovirus.

2. The sequenced region is a partial or complete segment.

Table 3

Species	#Sequence	Region	#Genotype	Number (%) of Genotype Discrepancies		References
				EVIDENCE	RIVM Genotyping Tool	
Enterovirus A	49	5'UTR	2	0	49 (100%) (49 only typed to species level)	[72]
	59	VP1	8	0	0	[27,28,52,53]
	3	CG	3	0	0	[53]
Enterovirus B	101	VP1/ VP2	22	6 (5.9%) (4 seqs typed to species level, 2 mistyped)	6 (5.9%) (4 seqs typed to species level, 2 mistyped)	[27,28,52]
Enterovirus C	13	VP1	6	1 (7.7%) (mistyped)	2 (15.4%) (1 seq typed to species level, 1 mistyped)	[27,28]
Rhinovirus A	1	VP1	1	0	1 (100%) (1 seq typed to species level)	[28]
Untypeable	5	VP1	NA	5 (100%) (5 successfully typed)	5 (100%) (1 seq typed to species level, 4 successfully typed)	[28]

A case study

Here, we demonstrate a genotyping and recombination pipeline using two coxsackievirus A16 (CV-A16) strains with distinct pathogenesis.

Coxsackievirus strains CV-A16 GD09/24 (GenBank accession KC117317) and GD09/119 (GenBank accession KC117318), exhibiting differing levels of clinical virulence, were isolated in Guangdong, China, in 2009 [62]. The authors performed phylogenetic analysis with 28 CV-A16 homologous strains and one EV-A71 prototype strain to assign GD09/24 and GD09/119 to the CV-A16 genotype. To detect recombination events, the authors compared two novel CA-A16 strains with two EV-A71 strains and one CV-A16 prototype strain. Bootscanning results indicated that GD09/24 and GD09/119 underwent homologous recombination with EV-A71 in the P2 and P3 regions.

In this study, we used our EVIDENCE analysis pipeline to study GD09/24 and GD09/119. First, the sequences were submitted for analysis by the genotyping tool. After the scan region step (alignment parameters at the default settings), reported references were sorted by the sum of mixed-ranking scores of all 13 genome regions. We selected all genome regions and the top 10 ranked references to perform phylogenetic analysis with 500 bootstraps. The phylogenetic tree topology indicated

that the closest genotype to GD09/119 is CV-A16 (Figure 4A; Additional file 2: Figure S1 for GD09/24), which is consistent with the findings of the previous study. The five selected reference strains and the query were piped to the recombination detection page by clicking on the "push data to recombination detection" button. The bootscanning (Figure 4B) and similarity (Figure 4C) results revealed that GD09/119 is closely related to CV-A4 in the 5'UTR region, to CV-A16 in the P1 region, and to EV-A71, CV-A6, and CV-A7 in

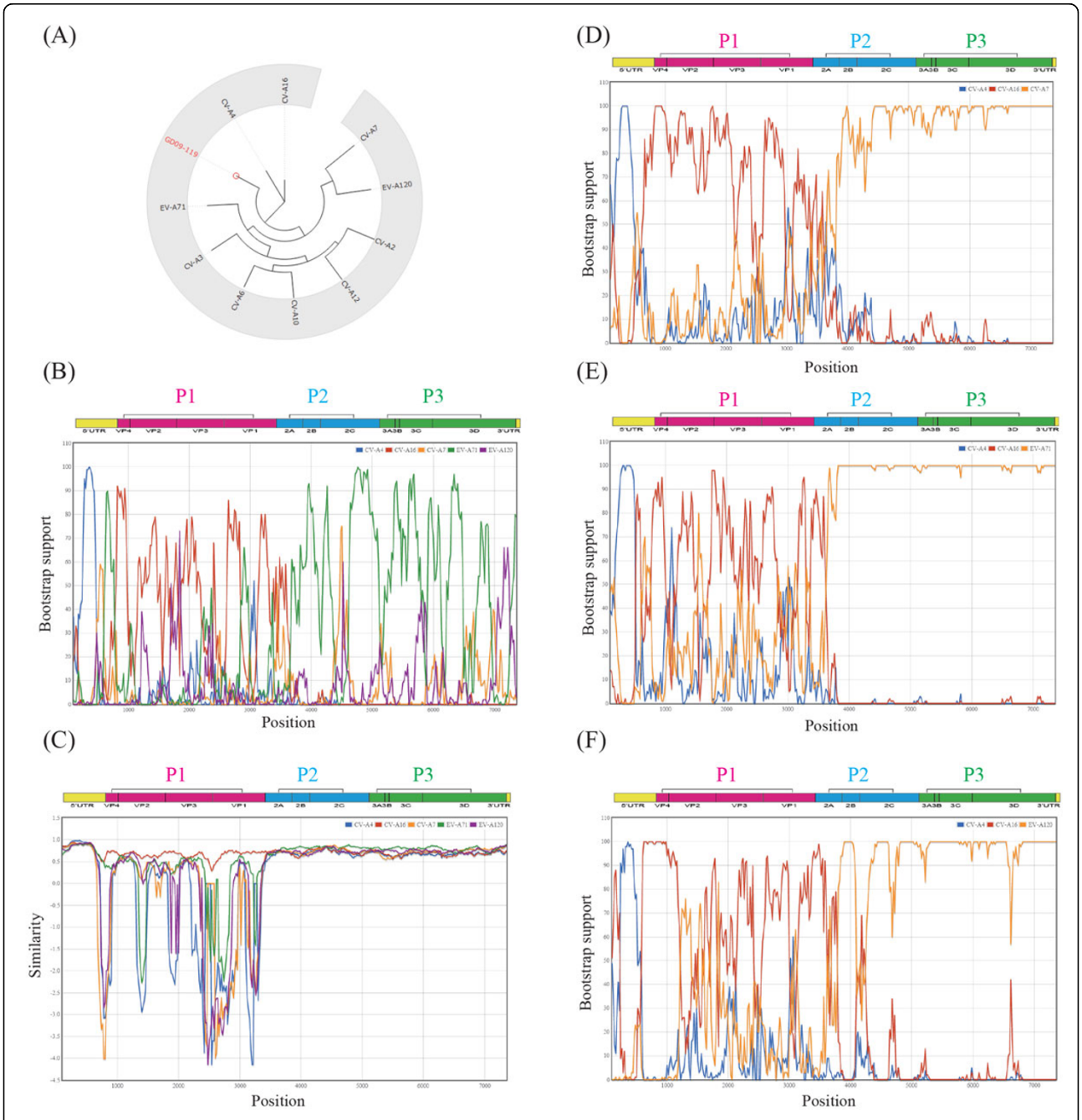

Figure 4 Detection of recombination events in a highly virulent enterovirus strain CA16/GD09/119. The ML tree of the GD09/119 complete genome sequence to the ten top ranked prototype references in GTRefSet is shown in panel A. The five closest references (CV-A4, CV-A16, CV-A7, EV-A71, and EV-A120) were selected for recombination detection, which was plotted using Bootscanning (panel B) and similarity (panel C). The results of the potential recombination analyses of GD09/119 to CVA4, CVA16, and one of CV-A7 (panel D), EV-A71 (panel E), or EV-A120 (panel F), are shown as bootscanning plots.

the P2 and P3 regions. GD09/24 showed the same recombination pattern (Additional file 2: Figure S1). In the earlier study [50], the authors used two references, EV-A71 and CV-A16, to detect recombination events. Thus, they did not observe the high correlation of 5'UTR to CV-A4 described here. In addition, it is difficult to determine the origin of the P2 and P3 regions of GD09/119 and GD09/24. EV-A71, CV-A6, and CV-A7 all showed high bootstrap values when individually subjected to bootstrapping with CV-A4 and CV-A16, supporting the hypothesis that recombination occurred in the P2 and P3 regions (Supplementary Figure S1, See Additional file 2). In fact, the P2 and P3 regions are highly conserved in EV-A71, CV-A6, and CV-A7. The results emphasize the limitations of phylogenetic-based recombination-detecting methods. The non-structural regions play major roles in viral replication, protein processing, virulence, and virus shedding [63-66], and thus influence host immune responses [67]. This may be the reason for the distinct pathogenicity of these two isolates. In addition, it has been reported that CV-A16 co-circulated and/or co-infected with EV-A71, CV-A6, or CV-A4 [68-71] in China from 2008 to 2014. Therefore, the novel epidemic strains of EV isolates GD09/24 and GD09/119 may have originated from recombination of CV-A16 to EV-A71, CV-A4, CV-A6, CV-A7, or EV-A120.

This demonstration shows that the genotyping and recombination pipeline in EVIDENCE can provide suitable candidates as references for recombination detection. Additionally, users can download all output files, and perform analyses using different reference sequences or genomic region(s) with a user-friendly interface.

Conclusions

Classical EV typing is largely dependent on serotyping methods. VP1 has been the subject of extensive research on account of the neutralization potency of its antiserum [23,58,72]. Typing specificity can be improved by using a panel of antibodies against VP1 and other viral proteins [73-77]. Thus, genotypes may be assigned through observing the response of several antibodies raised from epitopes in different genome regions. If a novel EV strain emerged from a recombinant event that joined epitopes of different parent strains, the serological phenotype may fail to reflect clinical virulence. As more EV sequences are reported, it is increasingly apparent that recombination occurs frequently within inter- or intra- genotypes. Moreover, each genomic region is subject to distinct selective pressures, and thus their evolution is independent of one another [78,79]. Increased genetic diversity often leads to phenotypic variation, which is problematic for clinical therapy.

EVIDENCE can be used to perform EV typing based on sequence context. This tool disassociates reference prototypes into functional components of the virus genome, and performs analysis in a modularized manner. However, the correlation of individual genome regions with genome virulence remains unclear. We hope that EVIDENCE can be used to address this question and provide insights into EV evolution, as well as facilitate the diagnosis of clinical specimens to ensure appropriate treatment.

Additional material

Additional file 1: Supplementary data files: Table S1: GTRefSet list, Table S2: ExRefSet list and Table S3: Summary of congruence between GenBank and re-classification in this study.

Additional file 2: Figure S1. Detection of recombination events in a mild virulent enterovirus strain CA16/GD09/24. (*.pdf)

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

CHL and CYL designed the algorithm, conducted the experiments, and drafted the manuscript together with SHC. CHL and YBW worked on EVIDENCE website construction and implemented the tool and phylogenetic analysis workflow. CYL, CAH and SHC participated in discussions and conceptualization as well as revising the draft. All the authors read and approved the manuscript.

Acknowledgements

The authors wish to thank Dr. Min-Shi Lee and Dr. Pao-Yang Chen for valuable discussions on epidemiology of enterovirus and model of molecular tying, respectively. We also thank the editor and anonymous reviewers for their helpful advice.
The research was funded by Ministry of Science and Technology (MOST), Taiwan, for financially supporting this research through MOST 104-2319-B-400-002 to CAH, MOST 103-2311-B-001-033-MY3 to CYL, MOST 101-2321-B-001-043-MY2 and MOST 102-2811-B-001-046 to SHC.

Declaration

The publication costs for this article were funded by Ministry of Science and Technology (MOST), Taiwan.
This article has been published as part of *BMC Genomics* Volume 16 Supplement 12, 2015: Joint 26th Genome Informatics Workshop and 14th International Conference on Bioinformatics: Genomics. The full contents of the supplement are available online at http://www.biomedcentral.com/bmcgenomics/supplements/16/S12.

Authors' details

[1]Division of Biostatistics and Bioinformatics, Institute of Population Health Sciences, National Health Research Institutes, Zhunan, Miaoli County, 35053, Taiwan. [2]Institute of Bioinformatics and Structural Biology, National Tsing Hua University, Hsinchu, 30013, Taiwan. [3]Institute of Information Science, Academia Sinica, 128 Academia Road, Section 2, Nankang, Taipei, 11529, Taiwan. [4]Institute of Zoology, National Taiwan University, No. 1, Sec. 4, Roosevelt Rd., Taipei, 10617, Taiwan. [5]Institute of Fisheries Science, National Taiwan University, No. 1, Sec. 4, Roosevelt Rd., Taipei, 10617, Taiwan.

Published: 9 December 2015

References

1. Tapparel C, Siegrist F, Petty TJ, Kaiser L: **Picornavirus and enterovirus diversity with associated human diseases.** *Infect Genet Evol* 2013, **14**:282-293.
2. Tsao KC, Huang CG, Huang YL, Chen FC, Huang PN, Huang YC, *et al*: **Epidemiologic features and virus isolation of enteroviruses in Northern Taiwan during 2000-2008.** *J Virol Methods* 2010, **165(2)**:330-332.

3. Caro V, Guillot S, Delpeyroux F, Crainic R: **Molecular strategy for 'serotyping' of human enteroviruses.** *J Gen Virol* 2001, **82(Pt 1)**:79-91.
4. Nix WA, Oberste MS, Pallansch MA: **Sensitive, seminested PCR amplification of VP1 sequences for direct identification of all enterovirus serotypes from original clinical specimens.** *J Clin Microbiol* 2006, **44(8)**:2698-2704.
5. Oberste MS, Maher K, Kilpatrick DR, Flemister MR, Brown BA, Pallansch MA: **Typing of human enteroviruses by partial sequencing of VP1.** *J Clin Microbiol* 1999, **37(5)**:1288-1293.
6. Zhou F, Kong F, Wang B, McPhie K, Gilbert GL, Dwyer DE: **Molecular characterization of enterovirus 71 and coxsackievirus A16 using the 5' untranslated region and VP1 region.** *J Med Microbiol* 2011, **60(Pt 3)**:349-358.
7. Zhou F, Wang Q, Sintchenko V, Gilbert GL, O'Sullivan MV, Iredell JR, Dwyer DE: **Use of the 5' Untranslated Region and VP1 Region to Examine the Molecular Diversity in Enterovirus B Species.** *J Med Microbiol* 2014, **63(Pt 10)**:1339-1355.
8. Ibrahim W, Boukhadra N, Nasri-Zoghlami D, Berthelot P, Omar S, Bourlet T, *et al*: **Partial sequencing of the VP2 capsid gene for direct enterovirus genotyping in clinical specimens.** *Clin Microbiol Infect* 2014, **20(9)**: O558-O565.
9. Ibrahim W, Ouerdani I, Pillet S, Aouni M, Pozzetto B, Harrath R: **Direct typing of human enteroviruses from wastewater samples.** *J Virol Methods* 2014, **207**:215-219.
10. Perera D, Shimizu H, Yoshida H, Tu PV, Ishiko H, McMinn PC, Cardosa MJ: **A comparison of the VP1, VP2, and VP4 regions for molecular typing of human enteroviruses.** *J Med Virol* 2010, **82(4)**:649-657.
11. Chan YF, Wee KL, Chiam CW, Khor CS, Chan SY, Amalina WM, Sam IC: **Comparative genetic analysis of VP4, VP1 and 3D gene regions of enterovirus 71 and coxsackievirus A16 circulating in Malaysia between 1997-2008.** *Trop Biomed* 2012, **29(3)**:451-466.
12. Casas I, Palacios GF, Trallero G, Cisterna D, Freire MC, Tenorio A: **Molecular characterization of human enteroviruses in clinical samples: comparison between VP2, VP1, and RNA polymerase regions using RT nested PCR assays and direct sequencing of products.** *J Med Virol* 2001, **65(1)**:138-148.
13. Bochkov YA, Grindle K, Vang F, Evans MD, Gern JE: **Improved molecular typing assay for rhinovirus species A, B, and C.** *J Clin Microbiol* 2014, **52(7)**:2461-2471.
14. Thoelen I, Moes E, Lemey P, Mostmans S, Wollants E, Lindberg AM, *et al*: **Analysis of the serotype and genotype correlation of VP1 and the 5' noncoding region in an epidemiological survey of the human enterovirus B species.** *J Clin Microbiol* 2004, **42(3)**:963-971.
15. Chan YF, Sam IC, AbuBakar S: **Phylogenetic designation of enterovirus 71 genotypes and subgenotypes using complete genome sequences.** *Infect Genet Evol* 2010, **10(3)**:404-412.
16. Lindberg AM, Andersson P, Savolainen C, Mulders MN, Hovi T: **Evolution of the genome of Human enterovirus B: incongruence between phylogenies of the VP1 and 3CD regions indicates frequent recombination within the species.** *Journal of General Virology* 2003, **84(Pt 5)**:1223-1235.
17. Hu L, Zhang Y, Hong M, Zhu S, Yan D, Wang D, *et al*: **Phylogenetic evidence for multiple intertypic recombinations in enterovirus B81 strains isolated in Tibet, China.** *Sci Rep* 2014, **4**:6035.
18. Tang J, Yoshida H, Ding Z, Tao Z, Zhang J, Tian B, *et al*: **Molecular Epidemiology and Recombination of Human Enteroviruses from AFP surveillance in Yunnan, China from 2006 to 2010.** *Sci Rep* 2014, **4**:6058.
19. Yip CC, Lau SK, Lo JY, Chan KH, Woo PC, Yuen KY: **Genetic characterization of EV71 isolates from 2004 to 2010 reveals predominance and persistent circulation of the newly proposed genotype D and recent emergence of a distinct lineage of subgenotype C2 in Hong Kong.** *Virol J* 2013, **10**:222.
20. Kroneman A, Vennema H, Deforche K, van der Avoort H, Penaranda S, Oberste MS, *et al*: **An automated genotyping tool for enteroviruses and noroviruses.** *Journal of Clinical Virology* 2011, **51(2)**:121-125.
21. Rozanov M, Plikat U, Chappey C, Kochergin A, Tatusova T: **A web-based genotyping resource for viral sequences.** *Nucleic Acids Res* 2004, **32(Web Server issue)**:W654-W659.
22. King AM, Adams MJ, Lefkowitz EJ, Carstens EB: **Virus taxonomy: classification and nomenclature of viruses: Ninth Report of the International Committee on Taxonomy of Viruses.** *Elsevier* 2012, **9**.
23. Oberste MS, Pallansch MA: **Enterovirus molecular detection and typing.** *Reviews in Medical Microbiology* 2005, **16(4)**:163-171.
24. Adams MJ, Lefkowitz EJ, King AM, Carstens EB: **Ratification vote on taxonomic proposals to the International Committee on Taxonomy of Viruses (2014).** *Archives of virology* 2014, **159(10)**:2831-2841.
25. Bukreyev AA, Chandran K, Dolnik O, Dye JM, Ebihara H, Leroy EM, *et al*: **Discussions and decisions of the 2012-2014 International Committee on Taxonomy of Viruses (ICTV) Filoviridae Study Group, January 2012-June 2013.** *Archives of Virology* 2014, **159(4)**:821-830.
26. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25(17)**:3389-3402.
27. Rasko DA, Myers GS, Ravel J: **Visualization of comparative genomic analyses by BLAST score ratio.** *BMC Bioinformatics* 2005, **6**:2.
28. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215(3)**:403-410.
29. Korf I, Yandell M, Bedell J: **BLAST.** O'Reilly Media, Incorporated; 2003.
30. Blom N, Hansen J, Blaas D, Brunak S: **Cleavage site analysis in picornaviral polyproteins: discovering cellular targets by neural networks.** *Protein Sci* 1996, **5(11)**:2203-2216.
31. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, *et al*: **Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega.** *Mol Syst Biol* 2011, **7**:539.
32. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O: **New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0.** *Syst Biol* 2010, **59(3)**:307-321.
33. Smits SA, Ouverney CC: **jsPhyloSVG: a javascript library for visualizing interactive and vector-based phylogenetic trees on the web.** *PLoS One* 2010, **5(8)**:e12267.
34. Salminen MO, Carr JK, Burke DS, McCutchan FE: **Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning.** *AIDS Res Hum Retroviruses* 1995, **11(11)**:1423-1425.
35. Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, Novak NG, *et al*: **Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination.** *J Virol* 1999, **73(1)**:152-160.
36. Felsenstein J: **PHYLIP 3.5 (phylogeny inference package).** *Department of Genetics, University of Washington, Seattle* 1993.
37. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T: **trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses.** *Bioinformatics* 2009, **25(15)**:1972-1973.
38. Federhen S: **The NCBI Taxonomy database.** *Nucleic Acids Res* 2012, **40(Database issue)**:D136-D143.
39. Tu PV, Thao NT, Perera D, Huu TK, Tien NT, Thuong TC, *et al*: **Epidemiologic and virologic investigation of hand, foot, and mouth disease, southern Vietnam, 2005.** *Emerg Infecti Dis* 2007, **13(11)**:1733-1741.
40. Sanders SA, Herrero LJ, McPhie K, Chow SS, Craig ME, Dwyer DE, *et al*: **Molecular epidemiology of enterovirus 71 over two decades in an Australian urban community.** *Arch Virol* 2006, **151(5)**:1003-1013.
41. Nasri D, Bouslama L, Omar S, Saoudin H, Bourlet T, Aouni M, *et al*: **Typing of human enterovirus by partial sequencing of VP2.** *J Clin Microbiol* 2007, **45(8)**:2370-2379.
42. Aw TG, Gin KY: **Environmental surveillance and molecular characterization of human enteric viruses in tropical urban wastewaters.** *Journal of Applied Microbiology* 2010, **109(2)**:716-730.
43. Zhong JY, Zhu B, Hua L, Wang CB, Kuang L, Xie JH, Chen Y: **[Complete genomic sequence analysis on human enterovirus 71 strain in Guangzhou, in 2008 and 2010].** *Zhonghua Liu Xing Bing Xue Za Zhi* 2011, **32(7)**:700-704.
44. Bessaud M, Pillet S, Ibrahim W, Joffret ML, Pozzetto B, Delpeyroux F, Gouandjika-Vasilache I: **Molecular characterization of human enteroviruses in the Central African Republic: uncovering wide diversity and identification of a new human enterovirus A71 genogroup.** *J Clin Microbiol* 2012, **50(5)**:1650-1658.
45. Sun Q, Zhang Y, Zhu S, Tian H, Huang G, Cui H, *et al*: **Transmission of human enterovirus 85 recombinants containing new unknown serotype HEV-B donor sequences in Xinjiang Uighur autonomous region, China.** *PLoS One* 2013, **8(1)**:e55480.
46. Lukashev AN, Shumilina EY, Belalov IS, Ivanova OE, Eremeeva TP, Reznik VI, *et al*: **Recombination strategies and evolutionary dynamics of the Human enterovirus A global gene pool.** *J Gen Virol* 2014, **95(Pt 4)**:868-873.
47. Ni H, Yi B, Yin J, Fang T, He T, Du Y, *et al*: **Epidemiological and etiological characteristics of hand, foot, and mouth disease in Ningbo, China, 2008-2011.** *J Clin Virol* 2012, **54(4)**:342-348.

48. Oberste MS, Maher K, Michele SM, Belliot G, Uddin M, Pallansch MA: **Enteroviruses 76, 89, 90 and 91 represent a novel group within the species Human enterovirus A.** *J Gen Virol* 2005, **86(Pt 2)**:445-451.
49. Laitinen OH, Honkanen H, Pakkanen O, Oikarinen S, Hankaniemi MM, Huhtala H, *et al*: **Coxsackievirus B1 is associated with induction of beta-cell autoimmunity that portends type 1 diabetes.** *Diabetes* 2014, **63(2)**:446-455.
50. Chiang PS, Huang ML, Luo ST, Lin TY, Tsao KC, Lee MS: **Comparing molecular methods for early detection and serotyping of enteroviruses in throat swabs of pediatric patients.** *PLoS One* 2012, **7(10)**:e48269.
51. Lu QB, Zhang XA, Wo Y, Xu HM, Li XJ, Wang XJ, *et al*: **Circulation of Coxsackievirus A10 and A6 in hand-foot-mouth disease in China, 2009-2011.** *PLoS One* 2012, **7(12)**:e52073.
52. Abubakar S, Chee HY, Shafee N, Chua KB, Lam SK: **Molecular detection of enteroviruses from an outbreak of hand, foot and mouth disease in Malaysia in 1997.** *Scandinavian journal of infectious diseases* 1999, **31(4)**:331-335.
53. Craig ME, Howard NJ, Silink M, Rawlinson WD: **Reduced frequency of HLA DRB1*03-DQB1*02 in children with type 1 diabetes associated with enterovirus RNA.** *J Infect Dis* 2003, **187(10)**:1562-1570.
54. Garcia J, Espejo V, Nelson M, Sovero M, Villaran MV, Gomez J, *et al*: **Human rhinoviruses and enteroviruses in influenza-like illness in Latin America.** *Virol J* 2013, **10**:305.
55. Sadeuh-Mba SA, Bessaud M, Joffret ML, Endegue Zanga MC, Balanant J, Mpoudi Ngole E, *et al*: **Characterization of Enteroviruses from non-human primates in cameroon revealed virus types widespread in humans along with candidate new types and species.** *PLoS Negl Trop Dis* 2014, **8(7)**: e3052.
56. Oberste MS, Maher K, Pallansch MA: **Evidence for frequent recombination within species human enterovirus B based on complete genomic sequences of all thirty-seven serotypes.** *J Virol* 2004, **78(2)**:855-867.
57. Siafakas N, Papaventsis D, Levidiotou-Stefanou S, Vamvakopoulos NC, Markoulatos P: **Classification and structure of echovirus 5'-UTR sequences.** *Virus Genes* 2005, **31(3)**:293-306.
58. Oberste MS, Maher K, Kilpatrick DR, Pallansch MA: **Molecular evolution of the human enteroviruses: correlation of serotype with VP1 sequence and application to picornavirus classification.** *J Virol* 1999, **73(3)**:1941-1948.
59. Oberste MS, Maher K, Flemister MR, Marchetti G, Kilpatrick DR, Pallansch MA: **Comparison of classic and molecular approaches for the identification of untypeable enteroviruses.** *J Clin Microbiol* 2000, **38(3)**:1170-1174.
60. Wang JR, Tsai HP, Huang SW, Kuo PH, Kiang D, Liu CC: **Laboratory diagnosis and genetic analysis of an echovirus 30-associated outbreak of aseptic meningitis in Taiwan in 2001.** *J Clin Microbiol* 2002, **40(12)**:4439-4444.
61. Chen Y, Li C, He D, Cheng T, Ge S, Shih JW, *et al*: **Antigenic analysis of divergent genotypes human Enterovirus 71 viruses by a panel of neutralizing monoclonal antibodies: current genotyping of EV71 does not reflect their antigenicity.** *Vaccine* 2013, **31(2)**:425-430.
62. Han JF, Yu N, Pan YX, He SJ, Xu LJ, Cao RY, *et al*: **Phenotypic and genomic characterization of human coxsackievirus A16 strains with distinct virulence in mice.** *Virus Research* 2014, **179**:212-219.
63. Junttila N, Leveque N, Magnius LO, Kabue JP, Muyembe-Tamfum JJ, Maslin J, *et al*: **Complete coding regions of the prototypes enterovirus B93 and C95: phylogenetic analyses of the P1 and P3 regions of EV-B and EV-C strains.** *J Med Virol* 2015, **87(3)**:485-497.
64. Riquet FB, Blanchard C, Jegouic S, Balanant J, Guillot S, Vibet MA, *et al*: **Impact of exogenous sequences on the characteristics of an epidemic type 2 recombinant vaccine-derived poliovirus.** *J Virol* 2008, **82(17)**:8927-8932.
65. Oberste MS: **Comparative genomics of the coxsackie B viruses and related enteroviruses.** *Current topics in microbiology and immunology* 2008, **323**:33-47.
66. Arita M, Shimizu H, Nagata N, Ami Y, Suzaki Y, Sata T, *et al*: **Temperature-sensitive mutants of enterovirus 71 show attenuation in cynomolgus monkeys.** *J Gen Virol* 2005, **86(Pt 5)**:1391-1401.
67. Weinzierl AO, Rudolf D, Maurer D, Wernet D, Rammensee HG, Stevanovic S, Klingel K: **Identification of HLA-A*01- and HLA-A*02-restricted CD8+ T-cell epitopes shared among group B enteroviruses.** *J Gen Virol* 2008, **89(Pt 9)**:2090-2097.
68. Mao Q, Wang Y, Yao X, Bian L, Wu X, Xu M, Liang Z: **Coxsackievirus A16: epidemiology, diagnosis, and vaccine.** *Hum Vaccin Immunother* 2014, **10(2)**:360-367.
69. Zhang C, Zhu R, Yang Y, Chi Y, Yin J, Tang X, *et al*: **Phylogenetic analysis of the major causative agents of hand, foot and mouth disease in Suzhou city, Jiangsu province, China, in 2012-2013.** *Emerg Microbes Infect* 2015, **4(2)**:e12.
70. Tan X, Li L, Zhang B, Jorba J, Su X, Ji T, *et al*: **Molecular epidemiology of coxsackievirus A6 associated with outbreaks of hand, foot, and mouth disease in Tianjin, China, in 2013.** *Arch Virol* 2015, **160(4)**:1097-1104.
71. Xu M, Su L, Cao L, Zhong H, Dong N, Xu J: **Enterovirus genotypes causing hand foot and mouth disease in Shanghai, China: a molecular epidemiological analysis.** *BMC Infectious Diseases* 2013, **13**:489.
72. Foo DG, Alonso S, Phoon MC, Ramachandran NP, Chow VT, Poh CL: **Identification of neutralizing linear epitopes from the VP1 capsid protein of Enterovirus 71 using synthetic peptides.** *Virus Res* 2007, **125(1)**:61-68.
73. Kiener TK, Jia Q, Lim XF, He F, Meng T, Chow VT, Kwang J: **Characterization and specificity of the linear epitope of the enterovirus 71 VP2 protein.** *Virol J* 2012, **9**:55.
74. Buttinelli G, Donati V, Ruggeri FM, Joki-Korpela P, Hyypia T, Fiore L: **Antigenic sites of coxsackie A9 virus inducing neutralizing monoclonal antibodies protective in mice.** *Virology* 2003, **312(1)**:74-83.
75. Abed Y, Wolf D, Dagan R, Boivin G: **Development of a serological assay based on a synthetic peptide selected from the VP0 capsid protein for detection of human parechoviruses.** *J Clin Microbiol* 2007, **45(6)**:2037-2039.
76. Dang S, Gao N, Li Y, Li M, Wang X, Jia X, *et al*: **Dominant CD4-dependent RNA-dependent RNA polymerase-specific T-cell responses in children acutely infected with human enterovirus 71 and healthy adult controls.** *Immunology* 2014, **142(1)**:89-100.
77. Varela-Calvino R, Skowera A, Arif S, Peakman M: **Identification of a naturally processed cytotoxic CD8 T-cell epitope of coxsackievirus B4, presented by HLA-A2.1 and located in the PEVKEK region of the P2C nonstructural protein.** *J Virol* 2004, **78(24)**:13399-13408.
78. Lukashev AN: **Role of recombination in evolution of enteroviruses.** *Reviews in medical virology* 2005, **15(3)**:157-167.
79. Lukashev AN, Lashkevich VA, Ivanova OE, Koroleva GA, Hinkkanen AE, Ilonen J: **Recombination in circulating enteroviruses.** *J Virol* 2003, **77(19)**:10423-10431.

doi:10.1186/1471-2164-16-S12-S8
Cite this article as: Lin *et al.*: Precise genotyping and recombination detection of Enterovirus. *BMC Genomics* 2015 **16**(Suppl 12):S8.

Submit your next manuscript to BioMed Central and take full advantage of:

- **Convenient online submission**
- **Thorough peer review**
- **No space constraints or color figure charges**
- **Immediate publication on acceptance**
- **Inclusion in PubMed, CAS, Scopus and Google Scholar**
- **Research which is freely available for redistribution**

Submit your manuscript at
www.biomedcentral.com/submit

BioMed Central