Title: Full genome characterization of two novel Alpha-coronavirus species from Italian bats

Authors: De Sabato Luca, Davide Lelli, Francesca Faccin, Sabrina Canziani, Ilaria Di Bartolo, Gabriele Vaccari, Ana Moreno

PII: S0168-1702(18)30585-9
DOI: https://doi.org/10.1016/j.virusres.2018.11.007
Reference: VIRUS 97524

To appear in: *Virus Research*

Received date: 25 September 2018
Revised date: 7 November 2018
Accepted date: 13 November 2018

Please cite this article as: Luca DS, Lelli D, Faccin F, Canziani S, Di Bartolo I, Vaccari G, Moreno A, Full genome characterization of two novel Alpha-coronavirus species from Italian bats, *Virus Research* (2018), https://doi.org/10.1016/j.virusres.2018.11.007

Full genome characterization of two novel Alpha-coronavirus species from Italian bats.

De Sabato Luca[1,2], Davide Lelli[3], Francesca Faccin[3], Sabrina Canziani[3], Ilaria Di Bartolo[2], Gabriele Vaccari[2*], Ana Moreno[3]

[1]Dept. of Sciences, University Roma Tre, Viale Guglielmo Marconi 446, 00146 Rome, Italy.

[2]Department of Food Safety, Nutrition and Veterinary Public Health, Istituto Superiore di Sanità. Viale Regina Elena 299, 00161 Rome, Italy.

[3]Department of Virology, Istituto Zooprofilattico Sperimentale Lombardia ed Emilia Romagna, Via Antonio Bianchi 9, 25124 Brescia, Italy.

*Correspondence:

Gabriele Vaccari: Istituto Superiore di Sanità, Dept. of Food Safety, Nutrition and Veterinary Public Health, Viale Regina Elena 299, 00161 Rome (Italy);

E-mail: gabriele.vaccari@iss.it

Highlights:

- Three Alpha-CoV strains were fully sequenced by NGS method
- The Italian strains were classified into two novel Alpha-CoV species
- The phylogenetic analysis on RdRp fragment sequences showed correlation to European strains

Summary

Coronaviruses (CoVs) have been detected worldwide in several bat species, which are considered the main reservoir. The attention to the high diversity of CoVs hosted by bats has increased during the last decade due to the high number of human infections caused by two zoonotic Beta-CoVs, SARS-CoV and MERS-CoV, that cause several respiratory diseases. Among coronaviruses, two Alpha-CoV strains (HuCoV-229E and HuCoV-NL63) cause mild respiratory disease that can change to severe disease in children, elderly and individuals affected by illnesses. Phylogenetic analysis

conducted on bat Alpha-CoV strains revealed their evolutive correlation to human strains, suggesting their origin in bats. The genome of CoVs is characterized by a high frequency of mutations and recombination events, increasing their ability to switch hosts and their zoonotic potential. In this study, three strains of Alpha-CoV genera detected in Italian bats (*Pipistrellus kuhlii*) were fully sequenced by Next Generation Sequencing (NGS) and characterized. The complete genome analysis showed the correlation of the Italians strains with a Chinese strain detected in 2013 and, based on CoV molecular species demarcation, two new Alpha-CoV species were established. The analysis of a fragment of the *RNA-dependent RNA polymerase* (*RdRp*) showed the correlation of the Italian strains with CoVs was only detected in the bat *Pipistrellus* genera (*Pipistrellus kuhlii* and *Pipistrellus Pipistrellus*) in European countries.

Key words: Bats; Full genome sequencing; Italy; Alpha-CoV viruses

1. Introduction

Bats are considered the natural reservoirs of several emerging and re-emerging viruses, such as Nipah virus, Marburg virus, rabies virus and coronaviruses, that have caused outbreaks in both humans and animals (Shi, 2013; Smith and Wang, 2013). The ecological features of bats, including their ability to fly long distances, their longevity, their large social colonies and their potential interactions with humans or livestock animals, facilitate virus maintenance and transmission, increasing the risk of intraspecies or interspecies jumping (Calisher et al., 2006). Among bat viruses, in the last decade, a large diversity of coronaviruses has been detected, exceeding the diversity seen in other mammalian hosts (Drexler et al., 2014). Coronaviruses (CoVs) (order *Nidovirales*, family *Coronaviridae*, subfamily *Coronavirinae*) are enveloped viruses characterized by a positive-sense single-stranded RNA genome of approximately 26 to 32 kilobases and classified into four genera (Weiss and Leibowitz, 2011). Alphacoronavirus (Alpha-CoV) and Betacoronavirus (Beta-CoV) infect several mammal species, including humans, bats and pigs, while Gammacoronavirus (Gamma-CoV) and Deltacoronavirus (Delta-CoV) infect birds, wild felines, pigs and some marine mammal species (Woo et al., 2009b; Woo et al., 2012). The genome of coronaviruses is characterized by high frequency recombination and a high mutation rate, which increases their potential for interspecies and intraspecies jumping (Lai, 1992; Holmes, 2009). Six CoV strains are recognized to infect humans. Two Alphacoronaviruses (HuCov-229E, -NL63) and two Betacoronaviruses (HuCoV-OC43, -HKU1) are responsible for the common cold and severe respiratory pathologies in infants, elderly people and immunocompromised patients and are characterized by human-to-human transmission (Hu et al., 2015). The other two Betacoronaviruses species, the Severe Acute Respiratory Syndrome virus (SARS-CoV in 2002–2003) and the Middle East Respiratory Syndrome virus (MERS-CoV in 2012) caused severe respiratory pathologies with case fatality rates of 9% and 35%, respectively (WHO, www.who.int). Phylogenetic analysis on strains detected in bats, humans and other mammals suggested that the origin of these CoVs

was in bats. The Rhinolophus bat species are considered the main reservoir for SARS-related CoVs. Bat MERS-related CoVs were also detected in African, Chinese and Italian bats (Annan et al., 2013; Ithete et al., 2013; Lau et al., 2013; Corman et al., 2014; Moreno et al., 2017), supporting the hypothesis of the bat origin. In three recent studies, related strains of the HuCoV-229E were detected in Hipposideros bats and strains of HuCoV-NL63 were detected in the American tricoloured bat (*Perimyotis subflavus*) and Kenyan *Triaenops afer* species, suggesting bats as potential reservoir host of Alphacoronavirus human strains (Pfefferle et al., 2009; Huynh et al., 2012; Corman et al., 2015). In addition, relatives of HuCoV-NL63 can be grown in immortalized bat cell lines, suggesting their potential association with bats (Huynh et al., 2012). This has led to speculations about an evolutionary origin of all mammalian CoVs in bat hosts (Woo et al., 2009a; Woo et al., 2009c). However, how humans become exposed to remote wildlife viruses is not always clear (Wolfe et al., 2007).

In Europe, several studies described the presence of CoVs in bat populations detecting both Alpha-CoVs and Beta-CoVs in Germany, Spain, Luxembourg, Italy, The Netherlands, the United Kingdom, France and Hungary (Gloza-Rausch et al., 2008; Reusken et al., 2010; Falcon et al., 2011; August et al., 2012; Lelli et al., 2013; Kemenesi et al., 2014; Goffard et al., 2015; Monchatre-Leroy et al., 2017; Pauly et al., 2017) from more than 20 different bat species. The detection of the same CoV strains (100% nucleotide identity) in different colonies of the same bat species or the circulation of different genera of CoVs (Alpha-CoVs and Beta-CoVs) in the same bat species confirm the high heterogeneity of CoVs in bats and that bat-CoV diversity depends more on the species-specificity than the geography and sampling location.

However, these studies were based on the analysis of a fragment of the *RNA-dependent RNA polymerase* (*RdRp*) gene that allows the assignment of the strains to the genera and not to the species. The International Committee on Taxonomy of Viruses (ICTV) established a molecular demarcation method for species assignment using the conserved domains of replicase polyprotein and the pairwise amino acid distance of 90% as threshold value. The Alpha-CoVs are classified into 11 species, 6 of which detected in bats: *Miniopterus bat coronavirus 1*, *Bat coronavirus CDPHE15*, *Miniopterus bat coronavirus HKU8*, *Rhinolophus bat coronavirus HKU2*, *Bat coronavirus HKU10*, and *Scotophilus bat coronavirus 512*, and some strains that to date are not assigned. However, the number of bat species that host CoVs is still unknown and increases proportionally with the increasing of surveillance. In this study, we describe the full genome sequencing by Next Generation Sequencing (NGS), the characterization and the classification of two novel Alpha-CoV species detected from three Italian *Pipistrellus kuhlii* bats (Lelli et al., 2013).

2. Materials and methods

2.1 Sampling

Two bat faecal samples and one carcass from three bat *Pipistrellus kuhlii* species were provided by a rehabilitation centre from Northern Italy between 2010 and 2015 and the bats species were identified according to the European bat identification keys based on their morphologic characteristics (Dietz, 2013). Faecal and organ samples positive for Alpha-CoV genera by a pan-coronavirus one-step RT-PCR (Lelli et al., 2013) were chosen for NGS analysis.

2.2 Whole-genome sequencing

Libraries were prepared following the sequence independent single primer amplification method (SISPA) (Djikeng et al., 2008). The RNA, extracted as previously described by Lelli et al. (2013), was retro-transcribed using the SuperScript IV Reverse Transcriptase (Invitrogen, Monza, Italy), starting with 9 µl of RNA and following the manufacturer's instructions. Twenty microlitres of cDNA were used to synthesize the second strand of cDNA by DNA Polymerase I Large (Klenow) Fragment (Promega, Milan, Italy) and then amplified by the Expand High Fidelity PCR System (Sigma Aldrich S.R.L., Milan, Italy). The PCR amplicons were purified using one volume of Agencourt AMPure XP beads (Beckman, Milan, Italy) following the manufacturer's instructions and eluted in 40 µl of nuclease-free water. Five hundred nanograms of purified DNA, quantified with the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, Rodano, MI, Italy) were digested with the EcoRV enzyme (New England BioLabs, Pero, MI, Italy) and then purified with a 1.8x volume of Agencourt AMPure XP beads (Beckman, Milan, Italy). The libraries were prepared by NEBNext Fast DNA Library Prep Set for Ion Torrent following the standard protocol for 100 ng of DNA. The barcoded libraries were mixed, and the pool was used for the Emulsion PCR performed by the Ion PGM Hi-Q OT2 Kit. The sequencing run was performed according to the manufacturer's instructions (Ion PGM Hi-Q Sequencing Kit) (Thermo Fisher Scientific) by Ion Personal Genome Machine (PGM) on the Ion 318 Chip v2.

2.3 Genome structure and phylogenetic analyses.

NGS, previously described by Moreno et al. (2017) was applied in order to obtain the complete genome. Data obtained by the Ion Torrent sequencer were analysed by the online portal Galaxy Aries (https://aries.iss.it). The reads were checked, cleaned up, and trimmed, and the sequences shorter than 50 nt were filtered. Host sequences were removed by mapping the reads against the Megabat and Microbat complete genomes downloaded from Genome Browser (https://www.genome.ucsc.edu) using the Bowtie 2 tool. The reads aligned by the BLASTn tool to the bacterial non-redundant nucleotide database RefSeq (https://www.ncbi.nlm.nih.gov/refseq/; E-value $>10^{-05}$) were removed by Galaxy Aries. The remaining reads were aligned with the viral non-redundant nucleotide database RefSeq (https://www.ncbi.nlm.nih.gov/refseq) and were parsed with the MEGAN6 software. The reads that showed no significant hits to the reference database were assigned to the unclassified reads.

The sequences classified into the *Coronaviridae* family were extracted and assembled into contigs by a *de novo* assembling method, using the default parameters, and excluding those shorter than 1000 bases using SPAdes tool (Galaxy Aries). The closest viral sequences were chosen by a BLASTn analysis and used to map the reads by the online tool Bowtie2 (Galaxy Aries). The output was visualized by the Integrative Genomics Viewer (IGV) software (http://software.broadinstitute.org/software/igv/), and the consensus sequence was extracted. Nucleotide and amino acids sequences were aligned, and the pairwise identity values were calculated with MEGA7 software (www.megasoftware.net). The open reading frames (ORFs) were predicted using the online tool ORF Finder (NCBI, http://www.ncbi.nlm.nih.gov/gorf/gorf.html). The potential cleavage sites in the orf1ab polyprotein were predicted by amino acid sequence alignment with other CoV strains and by using the online tool NetCorona 1.0 Server (http://www.cbs.dtu.dk/services/NetCorona/) (Kiemer et al., 2004). Comparison of the sequence distances of BatCoV-Ita4 and the closest Alpha-CoV sequences were confirmed using SSE v1.2 (Simmonds, 2012).

A dataset of complete genome references for Alpha-CoVs and Beta-CoVs species from the ICTV taxonomy report (https://talk.ictvonline.org) were obtained including the full genome sequences that displayed the highest nucleotide similarity to strains sequenced in this study, which resulted in 59 CoV sequences. A second dataset was used to build a ML tree using the partial sequence of the RdRp gene (409 nt) sequenced worldwide, excluding identical strains from the same study and bat species and resulting in 226 CoVs sequences. In both ML trees Beta-CoVs from different species were used as an outgroup.

To test the presence of recombination by RDP4 (Martin et al., 2015), six different methods were applied: GENECONV, BootScan, MaxChi, Chimaera, 3Seq, and SiScan, using the default settings.

The Maximum likelihood (ML) phylogenetic trees were built using MEGA7 software, applying, as a substitution model, a general time-reversible (GTR) model with a gamma-distributed (G) rate variation across sites, a proportion of invariant sites (I) (GTR+G+I) and a bootstrap analyses of 1000 pseudo-replicates.

The Bayesian phylogenetic trees were carried out using MrBayes ver. 3.1.2 (Huelsenbeck and Ronquist, 2001) using the sequences of the predicted proteins and excluding the most divergent strains. The Metropolis-coupled Markov chain Monte Carlo (MCMC) was used, starting from a random tree, run for 500 thousand heuristic search generations, sampling every 1,000 generations and discarding 25% of the samples as burn-in.

Analysis of the protein families of spike proteins and the prediction of the secondary structure were performed by the online tools: PFAM, InterProScan, the TMHMM program (http://www.cbs.dtu.dk/services/TMHMM/) (Apweiler et al., 2001; Bateman et al., 2002), Phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index) and the Swiss model (https://swissmodel.expasy.org).

3. Results

3.1 NGS data analysis

The NGS run produced approximately 4 million reads. A total of 1,376,444 reads were obtained for Bat-CoV/P.kuhlii/Italy/206645-41/2011 (BatCoV-Ita3). Of these reads, 8% were host reads, 78% were bacterial reads and 84,069 (6%) were viral reads, of which 81,069 (5.8%) were classified as *Coronaviridae* and 8% were unclassified. For Bat-CoV/P.kuhlii/Italy/3398-19/2015 (BatCoV-Ita4), 970,190 reads were retrieved, of which 6.8% were host reads, 38% were bacterial, 72,127 (7.4%) were viral, of which 5,484 (0.5%) were *Coronaviridae,* and 52% were unclassified. For Bat-CoV/P.kuhlii/Italy/206679-3/2010 (BatCoV-Ita5), 1,602,274 reads were obtained. Of these, 0.3% were host reads, 98% from bacteria, and 22,621 were viral (1,4%), of which 19,053 (1.1%) were *Coronaviridae* sequences and 0.3% were unclassified.

The reads classified into the *Coronaviridae* family were used to assemble the contigs obtaining 1 contig of approximately 27,000 nt for BatCoV-Ita3, 6 contigs > 4,000 nt for BatCoV-Ita4 and BatCoV-Ita5 contigs >2,800 nt for BatCoV-Ita5. The three assembled full genomes showed an average coverage of 751x, 50x and 181x, for BatCoV-Ita3, BatCoV-Ita4, and BatCoV-Ita5, respectively. The RdRp sequence of BatCoV-Ita4 was not obtained by the Sanger method used by Lelli et al. (2013). The other two Italian bat RdRp sequences showed 99% nucleotide identity with the RdRp region of the complete genomes obtained by NGS.

3.2 Genome organization

The complete genome sizes were 27,862 nt for BatCoV-Ita3, 28,129 nt for BatCoV-Ita4 and 28,146 BatCoV-Ita5, with a G+C content of 42%, 40.3% and 40.4%, respectively. The first sequence analysis was performed by BLASTn, comparing the Italian strains with those available online (https://www.ncbi.nlm.nih.gov). The BLASTn search showed similarities with an unclassified strain BtNv-AlphaCoV/SC2013 (KJ473809), and with those viruses classified into HKU10 bat and Porcine epidemic diarrhoea virus (PEDV) species. BatCoV-Ita4 and BatCoV-Ita5 shared a 97% nucleotide identity (nt. id.), and BatCoV-Ita3 shared 70.9% and 71% nt. id. with BatCoV-Ita4 and BatCoV-Ita5, respectively, at the full genome level. The same differences were observed when the ORF nucleotide sequences were aligned separately. BatCoV-Ita4 and BatCoV-Ita5 shared >97% nt. id. in all the ORFs. BatCoV-Ita3 showed the highest differences in the S, ORF3 and N genes with <65% nt. id.. Fewer differences were observed at the ORF1ab, M and E genes (>70% nt. id.) compared to the other two Italian strains. Their genome organization was similar to other Alpha-CoV species, comprehending 6 ORFs and two non-translated termini in the order of 5' terminus-ORF1ab-spike-ORF3-envelope (E)-membrane (M)-nucleocapsid (N)-3' terminus (Table 1). In the ORF1ab, it has been observed that the predicted slippery sequence "UUUAAAC" is involved in the synthesis of the replicase pp1ab polyprotein by ribosomal

frameshift, a characteristic of the *Nidovirales* order. The sizes, the genomic localization and the 15 expected cleavage sites of the nonstructural protein (NSP 1-16) that are encoded by ORF1ab, were predicted by sequence comparison with other Alpha-CoV species (Table 2). BatCoV-Ita4 and BatCoV-Ita5 showed the same sequences of cleavage sites. BatCoV-Ita3, compared to the other 2 strains, showed two amino acid changes in the cleavage sites between NSP1/NSP2 and NSP12/NSP13. A leader predicted transcription regulatory sequences (TRS-L), and the putative body TRSs, representing signals for the discontinuous transcription of subgenomic mRNAs (sgmRNAs), have been identified in the three genomes (Table 1). The TRS-L and TRSs preceded the codon start of all ORFs in BatCoV-Ita3 and suggested the synthesis of 6 monocistronic subgenomic mRNAs. The lack of TRSs before the ORF3 gene codon start in BatCoV-Ita4 and BatCoV-Ita5 suggests the synthesis of 4 monocistronic and 1 polycistronic subgenomic mRNAs. The differences at the nucleotide level were also confirmed at the amino acid level. BatCoV-Ita4 and BatCoV-Ita5 showed high similarities (<97%) and high differences with BatCoV-Ita3 in the spike, ORF3 and nucleocapsid proteins. The ICTV has established the 90% amino acid sequence identity of the seven concatenated domains within the ORF1ab as the threshold value to assign two strains to the same species: NSP3 (ADRP), NSP5 (3CLpro), NSP12 (RdRp), NSP13 (Hel, NTPase), NSP14 (ExoN, NMT), NSP15 (NendoU), and NSP16 (OMT). To classify the Italian strains into known coronavirus species, the ORF1ab concatenated domains were compared with the 11 Alpha-CoV species: *Miniopterus bat coronavirus 1, Bat coronavirus CDPHE15, Miniopterus bat coronavirus HKU8, Rhinolophus bat coronavirus HKU2, Bat coronavirus HKU10, Scotophilus bat coronavirus 512, PEDV, HuCoV-229E, HuCoV-NL63,* and *AlphaCoVs1,* and some strains that to date are not assigned. BatCoV-Ita3 concatenated domains showed sequence identities <83.8% with all the Alpha-CoV strains. BatCoV-Ita4 and BatCoV-Ita5 shared 99.3% identity and had <79.1% with all other Alpha-CoVs, suggesting that the classification of the Italian strains should be into two novel Alpha-CoVs species.

3.3 Phylogenetic analyses

The RDP4 recombination detection methods, applied to the dataset of CoVs complete genomes to detect the occurrence of recombination, supported the absence of recombination between the Italian strains and the Alpha-CoVs strains (P values >0.05). As shown in the ML tree built with complete genomes (Fig. 1), the Italian strains clustered with the Chinese strain BtNv-AlphaCoV/SC2013 (KJ473809) out of the monophyletic clade formed by the complete genomes of the HKU-8, 1A, 1B and HKU10 species. The former cluster is divided into two sub-clusters: one sub-cluster represented by BatCoV-Ita3 and BtNv-AlphaCoV/SC2013, sharing 75% nt. id., and the other sub-cluster represented by BatCoV-Ita4 and BatCoV-Ita5, sharing 71% nt. id. with the Chinese strain.

The Italian strains showed approximately 62% nt. id. with the strains classified into the HKU10 species (Hipposideros bat coronavirus HKU10 isolate LSH5A, Rousettus bat coronavirus HKU10 isolate 183A) and <60% with all other

AlphaCoV strains (Supplementary Fig. 1, Supplementary Fig. 2). Additionally, at the amino acid level, the Italian strains showed the highest identities with the Chinese BtNv-AlphaCoV/SC2013 strain with respect to the other Alpha-CoVs. BatCoV-Ita3 showed high identities in all predicted proteins excepting in the ORF3 and N proteins. BatCoV-Ita4 and BatCoV-Ita5 showed lower identities with respect to BatCoV-Ita3, which showed high identities in the orf1ab polyprotein and M proteins (>75%) and low identities in the other predicted proteins.

The Bayesian trees, built using the predicted protein sequences of E, M and N, confirmed the clustering of the Italian strains with the Chinese strain BtNv-AlphaCoV/SC2013 (data not shown). The tree built with S protein sequences showed a uniquely supported clade, containing the Italian strains, the BtNv-AlphaCoV/SC2013, HKU10, 1A, 1B, and HKU8 species and the unclassified strain BtMr-AlphaCoV/SAX2011, suggesting correlation only between those bat species (Supplementary Fig. 3).

The Italian bat strains showed low identities with the HuCoV-229E (<49%) and HuCoV-NL63 (<45%) strains at the spike protein level and had <45% identity with HuCoV-229E and <35% with HuCoV-NL63 at the Receptor Binding Domain (RBD) level. The prediction structure of the spike protein showed a type I membrane glycoprotein divided into two subunits (S1 and S2), as other Alpha-CoVs Spike proteins with most of the protein exposed on the outside of the virus and two transmembrane domains located at the C terminus. However, the Italian strains did not exhibit significant or supported similarities to the known secondary structure receptor-binding domains (HuCoV-229E, -NL63) using the online tool Phyre2 or the Swiss model due to their high divergences (data not shown).

To investigate the correlation among Alpha-CoV strains previously detected worldwide, a phylogenetic tree of the partial RdRp gene was built (Supplementary Fig. 4). The ML showed that strains detected in the same continent shared >89% nt. id. and were correlated, forming monophyletic clusters while sequences from a different cluster showed a nt. id. <85%. Most of the Alpha-CoVs species were detected in the same continent as the 1A, 1B, Bat-CoV 512, HKU2, and HKU8 species in Asia or the CDPHE15 species in North America. Bat coronaviruses related to human HuCoV-229E were retrieved in Africa and the coronaviruses related to HuCoV-NL63, in Africa and America. The HKU10 CoV strains showed sequences similar to those detected in Asia and Europe. Some strains formed a cluster outside of those classified into known Alpha-CoV species.

The Italian strains formed two clusters with the Chinese strain BtNv-AlphaCoV/SC2013 and some European strains. At the RdRp partial gene level, the BtNv-AlphaCoV/SC2013 strain showed 83% nucleotide identity with BatCoV-Ita3, 83.8% with BatCoV-Ita4 and 82.8% with BatCoV-Ita5. The first cluster is formed by BatCoV-Ita4 and BatCoV-Ita5, one Italian strain and one Spanish strain (P.kuh/Iprima/Spain/2007, HQ184058), collected from the bat *Pipistrellus kuhlii* species in the Southwest Piedmont region in Northern Italy (Pkuh605, KY780383) and in Spain in 2014 and 2007. These strains shared >96.7% nt. id. The second cluster contains the BatCoV-Ita3 with one Italian strain collected

in the centre of the Piedmont region in Northern Italy (Ppip1015C, KY780385), and a French strain (KT345294, Pip1_Cr_FR_2014), both collected in 2014. Those strains formed a monophyletic clade with two European strains, detected in Bulgaria (GU190239, BNM98-30/BGR/2008) and Spain (HQ184057, M.myo/I/Spain/2007), and two strains from South Africa (KF843855, BtCoV/GrNC1/Neo; KF843862, BtCoV/GrNC8/Neo) from the *Nyctalus leisleri*, *Myotis myotis* and *Neoromicia capensis* species, sharing with BatCoV-Ita3 approximately 83% nt. id.

4. Discussion

In this study, three Alpha-CoV strains from the *Pipistrellus kuhlii* bat species were fully sequenced. The *P. kuhlii* species is one of the most frequently described bat species in Italy that forages in urban and agricultural areas (Russo and Jones, 2003; Ancillotto et al., 2016).

To fully characterize the three Alpha-CoV strains, the NGS method previously described by Moreno et al. (2017) was applied successfully, obtaining the Alpha-CoV complete genome sequences with high coverage rates. However, the lack of European Alpha-CoV complete genomes make difficult to conduct a comprehensive genetic and phylogenetic analysis. The analysis on the full BatCoV-Ita sequences showed similarities to the Alpha-coronavirus genera and genome organization with 6 open reading frames (ORFs) and the 5' and 3' non-translated sequences. The phylogenetic analysis using the complete genomes showed correlation but with a low nucleotide identity with a Chinese strain detected in 2013 in the *Nyctalus velutinus* species. The phylogenetic analysis on amino acidic sequences also confirmed the correlation with the Chinese strain and supported the hypothesis that bat strains of *Miniopterus bat coronavirus 1*, *Miniopterus bat coronavirus HKU8*, and *Bat coronavirus HKU10* species and some unclassified strains may share a common spike ancestor. However, the analysis of the protein structure was hampered by the lack of similar spike protein structure. Indeed, due to the high genetic divergences with human strains it was impossible to predict the spike structure and the affinity with the human receptor.

The ICTV has established that viruses sharing more than 90% amino acid sequence identity in the conserved concatenated domains of the orf1ab polyprotein can be assigned to the same CoV species (https://talk.ictvonline.org/taxonomy/). The ICTV demarcation criteria for genera and species allowed us to classify the BatCoV-Ita into two novel Alpha-CoVs species. Our results support previous findings about the high heterogeneity of CoVs hosted by bats and support the idea that novel species may be found in the future with increasing surveillance. Several studies described the presence of Alpha-CoV and Beta-CoVs in bats worldwide (Falcon et al., 2011; Gouilh et al., 2011; August et al., 2012; Goffard et al., 2015; Asano et al., 2016; Fischer et al., 2016; Goes et al., 2016; Subudhi et al., 2017; Ar Gouilh et al., 2018; Geldenhuys et al., 2018). However, most of these studies reported phylogenetic analysis on short sequences within the RdRp region, establishing the correlation with other CoV strains but not the assignment to CoV species as established by ICTV.

In Europe, CoV strains were detected in samples from more than 20 different bat species. In Italy, a large variety of CoV strains were detected in *Myotis nattereri, Myotis daubentonii, Myotis myotis, Rhinolophus hipposideros, Hypsugo savii, Pipistrellus kuhlii, Pipistrellus pipistrellus, Nyctalus noctula, Epseticus serotinus, Myotis blythii, Myotis oxygnathus,* and *Plecotus auritus* (Lelli et al., 2013; De Benedictis et al., 2014; Rizzo et al., 2017).

The phylogenetic analysis on the RdRp region showed the correlation of the Alpha-CoV strains detected in the same continent. Interestingly, within each geographic area most of the strains hosted by the same bat genera cluster together, confirming the CoVs -host coevolution. BatCoV-Ita4 and BatCoV-Ita5 strains showed high nucleotide identity with one Italian strain and one Spanish strain detected in 2014 and 2007, respectively, from the bat *Pipistrellus kuhlii* species (HQ184058.1; KY780383.1) (Falcon et al., 2011; Rizzo et al., 2017). The BatCoV-Ita3 result correlated with one Italian and one French strain (KY780385.1; KT345294.1), both collected in 2014, from the bat *Pipistrellus Pipistrellus* species (Goffard et al., 2015; Rizzo et al., 2017). The high identities at the RdRp gene level and the clustering of the European strains with the Italian strains suggest that the two novel Alpha-CoV species detected in this study may infect at least two bat species of the *Pipistrellus* genera (*Pipistrellus kuhlii* and *Pipistrellus Pipistrellus*) from different European countries.

In contrast, some geographical clusters were represented by strains detected in different bat genera, attesting to the capability of the CoVs interspecies jumping that may occur when different species of bats share same roost (Leopardi et al., 2018).

Bat behaviour, including flying long distances, living in large colonies, having social interactions, and cohabitating with different bat species, favour the interspecies or intraspecies transmission of viruses(Calisher et al., 2006). During the last fifteen years, two Beta-CoVs, SARS-CoV and MERS-CoV, have jumped from bats to a mammalian intermediate host to humans (Field, 2009; Omrani et al., 2015). In addition, strains related to human Alpha-CoVs (HuCoV-229E, HuCoV-NL63) have been detected in bats, indicating the importance of the bat as a CoV reservoir (Pfefferle et al., 2009; Corman et al., 2015; Tao et al., 2017; Waruhiu et al., 2017). In this study, we characterized two Alpha-CoVs from Italian bats divergent from human CoVs strains and two new Alpha-CoV species. In addition, the RdRp phylogenetic tree showed that the strains here described were not related to the Alpha-CoV species established so far. This result highlights that the heterogeneity of CoVs in the bat may be higher than what is known to date. Indeed, to better understand the CoV species circulating in bats, their evolution and our understanding of the mechanisms important to cross the species barrier, it is important to have long-term vigilance followed by the complete genome characterization.

Funding

This study was partially funded by the Italian Ministry of Health under the Research projects:

GR-2011-02350591 PGR2011001- “An epizootiological survey of bats as reservoirs of emerging zoonotic viruses in Italy: implications for public health and biological conservation”;

PE-2011-02351681 PRF2011301 “Emerging respiratory viruses: monitoring of coronavirus infections at the human-animal interface”.

The funding body was not involved into the design of the study, and collection, analysis, and interpretation of data in the manuscript.

References

Ancillotto, L., Santini, L., Ranc, N., Maiorano, L., Russo, D. Extraordinary range expansion in a common bat: the potential roles of climate change and urbanization. Naturwissenschaften. 2016;103(3–4):15.

Annan, A., Baldwin, H.J., Corman, V.M., et al., 2013. Human Betacoronavirus 2c EMC/2012-related viruses in bats, Ghana and Europe. Emerg. Infect. Dis., 19(3), 456-459.

Apweiler, R., Attwood, T.K., Bairoch, A., et al., 2001. The InterPro database, an integrated documentation resource for protein families, domains and functional sites. Nucleic acids res., 29(1), 37-40.

Ar Gouilh, M., Puechmaille, S.J., Diancourt, L., et al., 2018. SARS-CoV related Betacoronavirus and diverse Alphacoronavirus members found in western old-world. Virology, 517, 88-97.

Asano, K.M., Hora, A.S., Scheffer, K.C., et al., 2016. Alphacoronavirus in urban Molossidae and Phyllostomidae bats, Brazil. Virol. J., 13, 110.

August, T.A., Mathews, F., Nunn, M.A., 2012. Alphacoronavirus detected in bats in the United Kingdom. Vector Borne Zoonotic Dis., 12(6), 530-533.

Bateman, A., Birney, E., Cerruti, L., et al., 2002. The Pfam protein families database. Nucleic acids res., 30(1), 276-280.

Calisher, C.H., Childs, J.E., Field, H.E., et al., 2006. Bats: important reservoir hosts of emerging viruses. Clin. Microbiol. Rev., 19(3), 531-545.

Corman, V.M., Baldwin, H.J., Tateno, A.F., et al., 2015. Evidence for an Ancestral Association of Human Coronavirus 229E with Bats. J. Virol., 89(23), 11858-11870.

Corman, V.M., Ithete, N.L., Richards, L.R., et al., 2014. Rooting the phylogenetic tree of middle East respiratory syndrome coronavirus by characterization of a conspecific virus from an African bat. J. Virol., 88(19), 11297-11303.

De Benedictis, P., Marciano, S., Scaravelli, D., et al., 2014. Alpha and lineage C BetaCoV infections in Italian bats. Virus genes, 48(2), 366-371.

Dietz, C., von Helversen, O. Illustrated identification key to the bats of Europe. Version 1.0, First released 15.12.2004. https://rhonealpes.lpo.fr/images/chiroptere/telecharger/dietz_von_helversen_2004_1.pdf

Djikeng, A., Halpin, R., Kuzmickas, R., et al., 2008. Viral genome sequencing by random priming methods. BMC genomics, 9, 5.

Drexler, J.F., Corman, V.M., Drosten, C., 2014. Ecology, evolution and classification of bat coronaviruses in the aftermath of SARS. Antiviral res., 101, 45-56.

Falcon, A., Vazquez-Moron, S., Casas, I., et al., 2011. Detection of Alpha and Betacoronaviruses in multiple Iberian bat species. Arch. Virol., 156(10), 1883-1890.

Field, H.E., 2009. Bats and emerging zoonoses: henipaviruses and SARS. Zoonoses Public Health, 56(6-7), 278-284.

Fischer, K., Zeus, V., Kwasnitschka, L., et al., 2016. Insectivorous bats carry host specific astroviruses and coronaviruses across different regions in Germany. Infect. Genet. Evol., 37, 108-116.

Geldenhuys, M., Mortlock, M., Weyer, J., et al., 2018. A metagenomic viral discovery approach identifies potential zoonotic and novel mammalian viruses in Neoromicia bats within South Africa. PloS one, 13(3), e0194527.

Gloza-Rausch, F., Ipsen, A., Seebens, A., et al., 2008. Detection and prevalence patterns of group I coronaviruses in bats, northern Germany. Emerg. Infect. Dis., 14(4), 626-631.

Goes, L.G.B., Campos, A.C.A., Carvalho, C., et al., 2016. Genetic diversity of bats coronaviruses in the Atlantic Forest hotspot biome, Brazil. Infect. Genet. Evol., 44, 510-513.

Goffard, A., Demanche, C., Arthur, L., et al., 2015. Alphacoronaviruses Detected in French Bats Are Phylogeographically Linked to Coronaviruses of European Bats. Viruses, 7(12), 6279-6290.

Gouilh, M.A., Puechmaille, S.J., Gonzalez, J.P., et al., 2011. SARS-Coronavirus ancestor's foot-prints in South-East Asian bat colonies and the refuge theory. Infect. Genet. Evol., 11(7), 1690-1702.

Holmes, E.C., 2009. The evolution and emergence of RNA viruses. In: Harvey, P.H., May, R.M. (Eds.), Oxford Series in Ecology and Evolution. Oxford University Press Inc., New York.

Hu, B., Ge, X., Wang, L.F., et al., 2015. Bat origin of human coronaviruses. Virol. J., 12, 221.

Huynh, J., Li, S., Yount, B., et al., 2012. Evidence supporting a zoonotic origin of human coronavirus strain NL63. J. Virol., 86(23), 12816-12825.

Ithete, N.L., Stoffberg, S., Corman, V.M., et al., 2013. Close relative of human Middle East respiratory syndrome coronavirus in bat, South Africa. Emerg. Infect. Dis., 19(10), 1697-1699.

Kemenesi, G., Dallos, B., Gorfol, T., et al., 2014. Molecular survey of RNA viruses in Hungarian bats: discovering novel astroviruses, coronaviruses, and caliciviruses. Vector Borne Zoonotic Dis., 14(12), 846-855.
Kiemer, L., Lund, O., Brunak, S., et al., 2004. Coronavirus 3CLpro proteinase cleavage sites: possible relevance to SARS virus pathology. BMC bioinformatics, 5, 72.
Lai, M.M., 1992. RNA recombination in animal and plant viruses. Microbiological reviews, 56(1), 61-79.
Lau, S.K., Li, K.S., Tsang, A.K., et al., 2013. Genetic characterization of Betacoronavirus lineage C viruses in bats reveals marked sequence divergence in the spike protein of pipistrellus bat coronavirus HKU5 in Japanese pipistrelle: implications for the origin of the novel Middle East respiratory syndrome coronavirus. J. Virol., 87(15), 8638-8650.
Lelli, D., Papetti, A., Sabelli, C., et al., 2013. Detection of coronaviruses in bats of various species in Italy. Viruses, 5(11), 2679-2689.
Leopardi, S., Holmes, E.C., Gastaldelli, M., et al., 2018. Interplay between co-divergence and cross-species transmission in the evolutionary history of bat coronaviruses. Infect. Genet. Evol., 58, 279-289.
Martin, D.P., Murrell, B., Golden, M., et al., 2015. RDP4: Detection and analysis of recombination patterns in virus genomes. Virus evol., 1(1), vev003.
Monchatre-Leroy, E., Boue, F., Boucher, J.M., et al., 2017. Identification of Alpha and Beta Coronavirus in Wildlife Species in France: Bats, Rodents, Rabbits, and Hedgehogs. Viruses, 9(12).
Moreno, A., Lelli, D., de Sabato, L., et al., 2017. Detection and full genome characterization of two Beta CoV viruses related to Middle East respiratory syndrome from bats in Italy. Virol. J., 14(1), 239.
Omrani, A.S., Al-Tawfiq, J.A., Memish, Z.A., 2015. Middle East respiratory syndrome coronavirus (MERS-CoV): animal to human interaction. Pathog. Glob. Health, 109(8), 354-362.
Pauly, M., Pir, J.B., Loesch, C., et al., 2017. Novel Alphacoronaviruses and Paramyxoviruses Cocirculate with Type 1 and Severe Acute Respiratory System (SARS)-Related Betacoronaviruses in Synanthropic Bats of Luxembourg. Appl. Environ. Microbiol., 83(18).
Pfefferle, S., Oppong, S., Drexler, J.F., et al., 2009. Distant relatives of severe acute respiratory syndrome coronavirus and close relatives of human coronavirus 229E in bats, Ghana. Emerg. Infect. Dis., 15(9), 1377-1384.
Reusken, C.B., Lina, P.H., Pielaat, A., et al., 2010. Circulation of group 2 coronaviruses in a bat species common to urban areas in Western Europe. Vector Borne Zoonotic Dis., 10(8), 785-791.
Rizzo, F., Edenborough, K.M., Toffoli, R., et al., 2017. Coronavirus and paramyxovirus in bats from Northwest Italy. BMC Vet. Res., 13(1), 396.
Russo, D. and Jones, G. Use of foraging habitats by bats in a Mediterranean area determined by acoustic surveys: conservation implications. Ecography. 2003; 26:197–209.
Shi, Z., 2013. Emerging infectious diseases associated with bat viruses. Science China. Life sci., 56(8), 678-682.
Simmonds, P., 2012. SSE: a nucleotide and amino acid sequence analysis platform. BMC Res. Notes, 5, 50.
Smith, I., Wang, L.F., 2013. Bats and their virome: an important source of emerging viruses capable of infecting humans. Curr. Opin. Virol., 3(1), 84-91.
Subudhi, S., Rapin, N., Bollinger, T.K., et al., 2017. A persistently infecting coronavirus in hibernating Myotis lucifugus, the North American little brown bat. J. Gen. Virol., 98(9), 2297-2309.
Tao, Y., Shi, M., Chommanard, C., et al., 2017. Surveillance of Bat Coronaviruses in Kenya Identifies Relatives of Human Coronaviruses NL63 and 229E and Their Recombination History. J. Virol., 91(5).
Waruhiu, C., Ommeh, S., Obanda, V., et al., 2017. Molecular detection of viruses in Kenyan bats and discovery of novel astroviruses, caliciviruses and rotaviruses. Virol. Sin., 32(2), 101-114.
Weiss, S.R., Leibowitz, J.L., 2011. Coronavirus pathogenesis. Adv. Virus Res., 81, 85-164.
Wolfe, N.D., Dunavan, C.P., Diamond, J., 2007. Origins of major human infectious diseases. Nature, 447(7142), 279-283.
Woo, P.C., Lau, S.K., Huang, Y., et al., 2009a. Coronavirus diversity, phylogeny and interspecies jumping. Exp. Biol. Med. (Maywood), 234(10), 1117-1127.
Woo, P.C., Lau, S.K., Lam, C.S., et al., 2009b. Comparative analysis of complete genome sequences of three avian coronaviruses reveals a novel group 3c coronavirus. J. Virol., 83(2), 908-917.
Woo, P.C., Lau, S.K., Lam, C.S., et al., 2012. Discovery of seven novel Mammalian and avian coronaviruses in the genus deltacoronavirus supports bat coronaviruses as the gene source of Alphacoronavirus and Betacoronavirus and avian coronaviruses as the gene source of gammacoronavirus and deltacoronavirus. J. Virol., 86(7), 3995-4008.

Woo, P.C., Lau, S.K., Yuen, K.Y., 2009c. Clinical features and molecular epidemiology of coronavirus-HKU1-associated community-acquired pneumonia. Hong Kong Med. J. = Xianggang yi xue za zhi, 15 Suppl 9, 46-47.

Fig. 1. Maximum phylogenetic tree based on 47 Alpha-CoVs and 12 Beta-CoVs complete genomes. The tree was inferred under the GTR + G + I substitution model and 1000 bootstrap resampling process replications showing values > 70. BatCoV-Ita3, BatCoV-Ita4 and BatCoV-Ita5 are reported in bold and the sequences can be retrieved under accession numbers (MH938448, MH938449, MH938450)

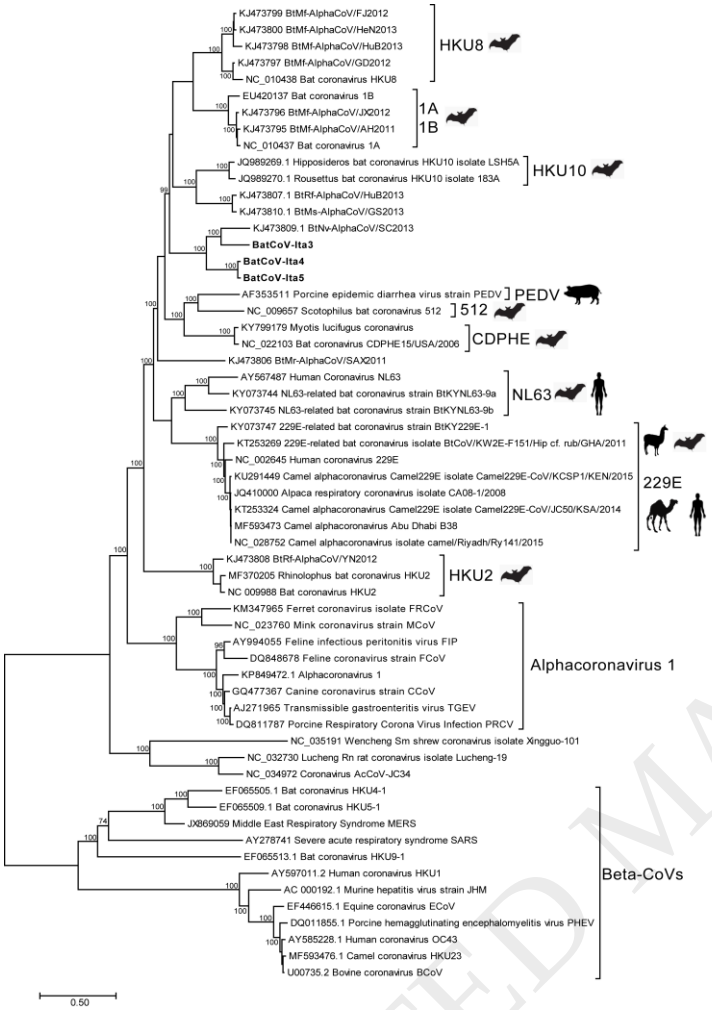

Table 1 Locations of predicted ORFs, protein sequences, putative leader TRS-L and TRS-B

ORF	nt position (start-end)	No. of amino acids	Sequence[a]
BatCoV-Ita3			
ORF1ab (TRS-L)	296-20166	6623	$_{00067}$CTAAAC$_{00073}$
Spike	20153-24274	1373	$_{20100}$−−−−−T−$_{20106}$
ORF3	24274-24924	216	$_{24230}$−−−−−−−$_{24236}$
E	24947-25174	75	$_{24939}$−−−−−−−$_{24945}$
M	25180-25863	227	$_{25170}$−−−−−−−$_{25176}$
N	25870-27174	434	$_{25859}$−−−−−−T$_{25865}$
BatCoV-Ita4			
ORF1ab (TRS-L)	281-20175	6627	$_{00051}$CTAAAC$_{00057}$
Spike	20172-24371	1399	$_{20161}$−−C−−−$_{20167}$
ORF3	24371-25024	217	
E	25103-25330	75	$_{25077}$−−−−−T−$_{25083}$
M	25343-26050	235	$_{25326}$−−−−−−−$_{25332}$
N	26058-27356	432	$_{26046}$−−−−−−−$_{26052}$
BatCoV-Ita5			
ORF1ab (TRS-L)	253-20192	6645	$_{00023}$CTAAAC$_{00029}$
Spike	20189-24388	1399	$_{20178}$−−C−−−$_{20184}$
ORF3	24388-25041	217	
E	25120-25347	75	$_{25094}$−−−−−T−$_{25100}$
M	25360-26067	235	$_{25343}$−−−−−−−$_{25349}$
N	26075-27373	432	$_{26063}$−−−−−−−$_{26069}$

[a] Dashes represent identical nucleotides compared to the leader TRS

Table 2 Prediction of the putative pp1ab cleavage sites

NSP	BatCoV-ITA3	BatCoV-ITA4	BatCoV-ITA5	Putative functional domain(s)[a]
NSP1	M^{1}-A^{107}	M^{1}-G^{107}	M^{1}-G^{107}	
NSP2	G^{108}-G^{771}	G^{108}-G^{771}	G^{108}-G^{771}	
NSP3	G^{772}-G^{2361}	G^{772}-G^{2374}	G^{772}-G^{2389}	ADRP, PL2pro
NSP4	G^{2362}-Q^{2838}	G^{2375}-Q^{2851}	G^{2390}-Q^{2866}	
NSP5	A^{2839}-Q^{3140}	A^{2852}-Q^{3153}	A^{2867}-Q^{3168}	3CLpro
NSP6	S^{3141}-Q^{3419}	S^{3154}-Q^{3432}	S^{3169}-Q^{3447}	
NSP7	S^{3420}-Q^{3502}	S^{3433}-Q^{3515}	S^{3448}-Q^{3530}	
NSP8	$S3^{503}$-Q^{3697}	S^{3516}-Q^{3710}	S^{3531}-Q^{3725}	Primase
NSP9	N^{3698}-Q^{3805}	N^{3711}-Q^{3818}	N^{3726}-Q^{3833}	
NSP10	A^{3806}-Q^{3940}	A^{3819}-Q^{3953}	A^{3834}-Q^{3968}	
NSP11	T^{3941}-L^{3958}	T^{3954}-L^{3971}	T^{3969}-L^{3986}	Short peptide at the end of ORF1a
NSP12	T^{3941}-Q^{4867}	T^{3959}-Q^{4880}	T^{3964}-Q^{4895}	RdRp
NSP13	S^{4868}-Q^{5464}	A^{4881}-Q^{5477}	A^{4896}-Q^{5492}	HEL, NTPase
NSP14	A^{5465}-Q^{5982}	A^{5478}-Q^{5995}	A^{5493}-Q^{6010}	ExoN, NMT
NSP15	S^{5983}-Q^{6321}	S^{5996}-Q^{6330}	S^{6011}-Q^{6348}	NendoU
NSP16	S^{6322}-V^{6623}	S^{6331}-K^{6631}	S^{6349}-V^{6646}	OMT

[a]ADRP ADP-ribose 1-phosphatase, PL2pro papain-like protease 2, 3CLpro coronavirus NSP5 protease, Hel helicase, NTPase nucleoside triphosphatase, ExoN exoribonuclease, NMT N7 methyltransferase, NendoU endoribonuclease, OMT 2' O-methyltransferase