Update on canine distemper virus (CDV) strains of Arctic-like lineage detected in dogs in Italy

Francesco Mira1*, Giuseppa Purpari1*, Santina Di Bella1, Domenico Vicari1, Giorgia Schirò1, Patrizia Di Marco1, Giusi Macaluso1, Mara Battilani1 and Annalisa Guercio1

*Francesco Mira and Giuseppa Purpari should be considered joint first author

1Istituto Zooprofylattico Sperimentale della Sicilia “A. Mirri”, Via Gino Marinuzzi 3, 90129 Palermo, Italy.
2Dipartimento di Scienze Mediche Veterinarie, Università di Bologna, Via Tolaro di Sopra 50, 40064 Ozzano Emilia (BO), Italy.

*Corresponding author at: Istituto Zooprofylattico Sperimentale della Sicilia ‘A. Mirri’, Via Gino Marinuzzi 3, 90129 Palermo, Italy.
Tel: +39 0916565447, e-mail: dottoremira@gmail.com.

Summary
Canine distemper virus (CDV) is the etiologic agent of distemper in dogs. It exhibits an elevated potential of crossing species barriers, infecting a wide range of wild and domestic carnivores. Of its encoding genes, hemagglutinin (H) shows high heterogeneity, and it was used to determine the relationship between CDV strains due to its variability and key role in determining cell tropism, host shift, and in eliciting a protective immune response. This study analysed the full-length H gene sequence of Arctic-like CDV strains collected from dogs in Italy during a period in which an increased activity of CDV diffusion was observed. The common amino acid changes and features of Arctic-like CDV strains collected from 2011 to 2016 in Europe were described, providing an updated analysis of the genomic features. A comparison with CDV vaccine strains was carried out to evaluate the increased genomic difference with CDV Arctic-like field strains. This study provides a complete and updated analysis of the current spreading strains of Arctic-like lineage and the main amino acid variations in the hemagglutinin gene sequence circulating in Italy. Moreover, it provides novel information regarding the evolution of the most recent CDV Arctic-like lineage strains collected in Europe.

Keywords
Biological evolution, Canine distemper virus, Dogs, Europe, Genomics, Hemagglutinin.

Aggiornamento sui ceppi di canine distemper virus (CDV) del lineage Artico rilevati nei cani in Italia

Parole chiave
Evoluzione biologica, Canine distemper virus, Cani, Europa, Genomica, Emoagglutinina.

Riassunto
Il canine distemper virus (CDV) è l’agente eziologico del cimurro nei cani. Mostra un elevato potenziale di superamento delle barriere di specie, infettando un ampio range di carnivori selvatici e domestici. Dei suoi geni codificanti, l’emoagglutinina (H) mostra alta eterogeneità ed è stata usata per determinare la relazione tra i ceppi di CDV, per via della sua variabilità e il ruolo chiave nel determinare il tropismo cellulare, il passaggio di specie e la capacità di elicitare una risposta immunitaria protettiva. Questo studio ha analizzato l’intera sequenza del gene H dei ceppi di CDV Artici identificati in Italia da cani durante un periodo in cui è stata osservata un’aumentata diffusione di CDV. Sono stati descritti i cambiamenti comuni degli aminoacidi e le caratteristiche dei ceppi CDV Artici collezionati dal 2011 al 2016 in Europa, fornendo un’analisi aggiornata delle caratteristiche genomiche. Per valutare l’aumento della divergenza genomicca rispetto ai ceppi di CDV Artico di campo, è stata effettuata una comparazione con i ceppi vaccinali di CDV. Questo studio restituisce un’analisi completa ed aggiornata della corrente circolazione dei ceppi del lineage Artico e le principali variazioni degli aminoacidi nella sequenza del gene dell’emoagglutinina circolanti in Italia. Presenta, inoltre, nuove informazioni relative all’evoluzione dei più recenti ceppi di CDV del lineage Artico collezionati in Europa.
**Introduction**

Canine Distemper (CD) is a highly contagious and often fatal viral disease, characterised by respiratory, gastrointestinal and nervous signs (Elia et al. 2006, von Messling et al. 2003). The severity of clinical signs is influenced by strain virulence, environmental conditions, host age and immune status (Espinal et al. 2014). Canine distemper virus (CDV), the causative agent of CD, belongs to the *Morbillivirus* genus within the *Paramyxoviridae* family (de Vries et al. 2015). Canine distemper virus was first described by H. Carré in 1905 (Carré 1905) and, over time, it has been reported in a broad range of terrestrial and aquatic carnivores (Carvalho et al. 2012) as well as in non-carnivorous species (Qiu et al. 2011, Yoshikawa et al. 1989). The genome consists in negative-sense, single-stranded RNA encoding for 6 structural proteins: nucleocapsid (N), matrix (M), fusion (F), hemagglutinin (H), polymerase (L), and phosphoprotein (P) (Martella et al. 2008). The H glycoprotein is the most variable integral membrane protein of the viral envelope, and mediates the binding of the virus to receptors on the host cell in the first step of infection (Martella et al. 2008, von Messling et al. 2001). Moreover, it plays an essential role in cell tropism; therefore, sequence variations may affect the virulence, host range, and neutralisation-epitopes of CDV (Ke et al. 2015). The H protein is a transmembrane glycoprotein composed of a short N-terminal cytoplasmatic tail followed by a transmembrane domain and a large C-terminal ectodomain. The ectodomain is structured as a stalk and a 6-blade (B1-B6) beta-propeller fold, and each blade module contains 4-stranded anti-parallel beta-sheets (S1-S4) (Ke et al. 2015).

The H gene represented a suitable target for the molecular typing of CDV strains since it showed the greatest variability in the viral genome. In fact, on the basis of the genetic divergence in the H gene (Bonami et al. 2007), distinct genetic lineages were recognised. The CDV strains belonged to the same lineage if they clustered together in the same clade and showed an amino acid (aa) divergence of less than 4% (Martella et al. 2006, Mochizuki et al. 1999). Therefore, on the bases of these criteria, 6 lineages (America-1 and -2, Asia-1 and -2, Europe, and Arctic) were recognised for many years (Martella et al. 2008). In recent years, intense and extensive molecular studies have led to the evidence of new, additional lineages. To this end, twelve genetic lineages have now been described (America-1 and -2, Arctic-like, Asia -1, -2, and -3, Europe WildLife, Europe-1/South America-1, South America-2, South America-3, South Africa, Rockborn-like) (Balboni et al. 2014, Espinal et al. 2014). More recently, in addition to these molecular typing criteria, it has been proposed that, within each genotype, subgenotypes were distinguishable on the basis of additional criteria: the nucleotide (nt) identity of at least 98% for the strains belonging to the same clade, and a clear separation in the phylogenetic tree, with high bootstrap values (Budaszewski et al. 2014).

The circulation of 3 distinct lineages have been reported in Italy: Europe (also called Europe-1/ South America-1), Europe WildLife, and Arctic-like lineages (Balboni et al. 2014, Di Francesco et al. 2012, Di Sabatino et al. 2014, Maracci et al. 2014, Martella et al. 2006, Martella et al. 2002, Monne et al. 2011). Arctic-like lineage CDV strains have also been detected in animal hosts other than dogs, such as wolves (Di Sabatino et al. 2014), badgers (Di Sabatino et al. 2016), and tigers (Seimon et al. 2013).

An adequate host immune response against H protein may prevent CDV infections (Martella et al. 2008). However, despite the fact that the incidence of CD in dogs has been reduced by the use of modified live virus (MLV) vaccines (Blixenkrone-Møller 1993), outbreaks of CD have also been reported in vaccinated dogs in Italy as well as in other countries (Balboni et al. 2014, Decaro et al. 2004, Di Francesco et al. 2012, Martella et al. 2007, Martella et al. 2006, Scagliairini et al. 2003). Genetic and antigenic heterogeneity between the vaccine and field strains was hypothesised to be responsible for CD in vaccinated dogs (Bae et al. 2013, Di Francesco et al. 2012, Zhao et al. 2010). However, as has also been hypothesised, some cases of CD in dogs vaccinated shortly before the onset of disease were more likely because of the presence of residues of maternally derived antibodies, and/or to the incorrect handling or management of the immune prophylaxis (Blixenkrone-Møller et al. 1993, Zhao et al. 2010).

In this study we analysed and characterised the H gene sequences of the CDV strains collected from dogs during a period in which increased activity of CDV was observed in Italy. We also reconstructed the geographic and temporal evolutions of the CDV Arctic-like lineage. For this purpose, the recent field strains were sequenced and their sequences were compared with those available in the GenBank in order to elucidate the genetic divergences of circulating CDV strains. The findings in this study reveal features in common with the most recently collected strains in Europe and provide novel information regarding the evolution and diffusion of the CDV Arctic-like lineage.

**Materials and methods**

**Samples**

The CDV strains analysed in this study were identified from samples of 91 dogs in southern Italy that exhibited...
suspected clinical signs of distemper (neurological, respiratory, or enteric signs). The samples were collected from February 2015 to October 2016 and were sent to the Istituto Zooprofilattico Sperimentale della Sicilia ‘A. Mirri’ (Palermo, Italy) for diagnostic purposes. The specimens included conjunctival, nasal, and rectal swabs or organs (brain, lungs, spleen, kidneys, liver, intestine). The swabs and autopsic specimens were both processed using homogenisation (10% w/v) in Eagle’s Minimum Essential Medium (EMEM, Lonza, BioWhittaker, Slough, UK) supplemented with antibiotics and antymycotic (1,000 U/ml penicillin G sodium salt, 1 mg/ml streptomycin sulfate, 2.5 μg/ml amphotericin B) (PAA Laboratories GmbH, Pasching, Austria). The homogenates were centrifuged at low speed (1,500 x g for 15 minutes at 4 °C); the supernatants were then collected, incubated at 37 °C for 1 hour, and were finally stored at -80 °C until processed.

Molecular detection of CDV

Viral RNA was extracted from 140 μl of homogenate using a QIAamp® Viral RNA Mini Kit (QIAGEN S.r.l., Milan, Italy) according to the manufacturer’s instructions. The presence of the CDV genome was screened using a reverse transcription PCR (RT-PCR) assay, which amplified a 429-bp fragment of the P gene (Barrett et al. 1993). The RT-PCR was carried out using the QIAGEN® OneStep RT-PCR Kit (QIAGEN S.r.l., Milan - Italy) in a 50 μl reaction mix consisting of 10 μl of 5x Buffer, 2 μl of deoxynucleotide (dNTP) Mix, 0.3 μl of RNase Inhibitor (30 U/μl), 1 μl of each primer DMV1-DMV2 (20 μM) (Table I), 2 μl of Enzyme Mix, 31.2 μl of nuclease-free water, and 2.5 μl of RNA extract. CDV RNA (strain Bussell) and nuclease-free water were used as positive and negative controls, respectively. Reverse transcription and amplification were carried out in a single-step protocol under the following thermal conditions: 50 °C for 30 minutes to synthesise the first cDNA, 95 °C for 15 minutes to inactivate the reverse transcriptase, followed by 35 cycles each of 94 °C for 45 seconds, 56 °C for 45 seconds, 72 °C for 1 minute, and a final extension of 72 °C for 10 minutes.

Sequencing and phylogenetic analysis

Positive samples from 10 screened dogs (Table II), selected on the basis of different geographical areas of collection and origin, underwent a second previously described RT-PCR using primer pair C (Demeter et al. 2007). This process amplified the entire CDV H gene sequence (Table I). The CDV RNA was amplified in a single-step PCR protocol using the QIAGEN® OneStep RT-PCR Kit (QIAGEN S.r.l.) in a 100 μl reaction mix consisting of 20 μl of 5x Buffer, 4 μl of dNTP Mix, 0.6 μl of Rnase Inhibitor (30 U/μl), 1.6 μl of each primer C-forward/C-reverse (20 μM) (Table I), 4 μl of Enzyme Mix, 63.2 μl of nuclease-free water, and 5 μl of RNA extract. Reverse transcription and amplification were carried out under the following thermal conditions: 50 °C for 30 minutes, 95 °C for 15 minutes, followed by 40 cycles each of 94 °C for 45 seconds, 52 °C for 45 seconds, 72 °C for 2 minutes, and a final extension of 72 °C for 10 minutes.
A representative PCR product of each of the 10 dogs was purified using an Illustra™ GFX™ PCR DNA and Gel Band Purification Kit (GE Healthcare Life Sciences, Amersham, Buckinghamshire, UK) and was sent to BMR Genomics srl (Padova, Italy) for direct Sanger sequencing with a set of 4 primers (primer pairs B-forward/reverse and C-forward/reverse) (Table I). According to an overlapping strategy, the sequences were assembled using the ABI Chroma Align web-based programme. Assembled nt sequences were entered into the BLAST (www.blast.ncbi.nlm.nih.gov/Blast.cgi) and the ENA (www.ebi.ac.uk/ena) web-based programmes to search for related sequences in the public domain sequence databases. A dataset of 122 CDV H gene sequences from worldwide CDV strains was retrieved from the GenBank/EMBL databases. The dataset included 59 H gene sequences of Arctic-like CDVs, collected from domestic and wild animals in Italy (DQ226087-88, KF914669, KC966928-29, KM115532-36, KX024708-09, KF184985-87, KF184989-91, KC992186-87, HM443706, HM443710-19, HM443720-22, HM443724), Hungary (DQ889178-86), Switzerland (KR002657-61), the U.S.A. (AY964108-12), China (AF172411, EF445052), Greenland (Z47760), Russia (X84998), and Austria (GQ214373) from 1988 to 2015. The H gene sequences of America I (Onderstepoort-AF305419; Convac-Z35493; Lederle-EF418782; Snyder Hill-AF259552) and Rockborn-like (Rockborn Candur-GU266280; Rockborn 46th-GU810819; Lesser Panda-AF178039; Vanguard-EF095750) lineage vaccine strains were also included in the dataset. Nucleotide and deduced amino acid sequences were aligned and analysed using BioEdit ver. 7.0. 0 software (Hall 1999). The prediction of potential N-linked glycosylation sites was carried out using NetNGlyc 1.0 web-based programme (www.cbs.dtu.dk/services/NetNGlyc/).

The phylogenetic analysis was carried out with MEGAS software (Tamura et al. 2011) using the Neighbor-joining method according to the Tamura 3-parameter model, modelled using a gamma distribution with 5 rate categories (bootstrap 1,000 replicates). The best-fit model of nt substitution was selected using the Find Best DNA/Protein Model function included with the MEGAS software. The Phocine Distemper Virus - strain Ulster/88 (D10371) was used as the outgroup. The H gene sequences of the CDV Arctic-like lineage available from the sequence databases were different lengths. For this reason the phylogenetic analysis was carried out

### Table III. Nucleotide and deduced amino acid (in brackets) variations in H gene sequence in analyzed CDV strains.

<table>
<thead>
<tr>
<th>Nt position</th>
<th>Aa position</th>
<th>114</th>
<th>145 (49)</th>
<th>173 (58)</th>
<th>197 (66)</th>
<th>240</th>
<th>249</th>
<th>290</th>
<th>360</th>
<th>423</th>
<th>516 (172)</th>
<th>528</th>
<th>583 (195)</th>
<th>818 (273)</th>
<th>843</th>
<th>897</th>
<th>929 (310)</th>
<th>1026</th>
<th>1059</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference strains&lt;sup&gt;(1)&lt;/sup&gt;</td>
<td>G</td>
<td>A</td>
<td>T</td>
<td>G</td>
<td>A</td>
<td>C</td>
<td>A</td>
<td>C</td>
<td>G</td>
<td>G</td>
<td>(Val)</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>Strains Italy 2015/16&lt;sup&gt;(2)&lt;/sup&gt;</td>
<td>G</td>
<td>A</td>
<td>T</td>
<td>G</td>
<td>A</td>
<td>C</td>
<td>A</td>
<td>T</td>
<td>(Phe)</td>
<td>G</td>
<td>(Val)</td>
<td>T</td>
<td>A</td>
<td>G</td>
<td>A</td>
<td>(Asp)</td>
<td>A</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>Strains Swiss&lt;sup&gt;(3)&lt;/sup&gt;</td>
<td>.</td>
<td>.</td>
<td>T</td>
<td>.</td>
<td>.</td>
<td>G</td>
<td>T</td>
<td>.</td>
<td>G</td>
<td>(Leu)</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>Strains 2011/13&lt;sup&gt;(4)&lt;/sup&gt;</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>G</td>
<td>.</td>
<td>G</td>
<td>.</td>
<td>G</td>
<td>(Leu)</td>
<td>.</td>
<td>A</td>
<td>(Ile)</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wolves&lt;sup&gt;(5)&lt;/sup&gt;</td>
<td>.</td>
<td>G</td>
<td>.</td>
<td>G</td>
<td>.</td>
<td>G</td>
<td>(Leu)</td>
<td>.</td>
<td>A</td>
<td>(Ile)</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Badgers&lt;sup&gt;(6)&lt;/sup&gt;</td>
<td>A</td>
<td>G</td>
<td>.</td>
<td>A</td>
<td>G</td>
<td>.</td>
<td>G</td>
<td>(Leu)</td>
<td>T</td>
<td>A</td>
<td>(Ile)</td>
<td>C</td>
<td>(Thr)</td>
<td>A</td>
<td>.</td>
<td>.</td>
<td>T</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>(1)</sup>Reference strains (DQ226087-88),<sup>(2)</sup>Strains analyzed in this study,<sup>(3)</sup>Strains Swiss (KR002657-58),<sup>(4)</sup>Strain 2011/13 (KF184989-91, KM115532-36, KJ567090-93), KF914669,<sup>(5)</sup>Wolves (KX96928-29),<sup>(6)</sup>Badgers (KX024708-09).
including only complete CDV H gene sequences (1824 nts), of which 27 sequences were from CDV Arctic-like lineage strains collected in Europe from 2003 to 2015, and 68 sequences of other lineages. In order to identify the relationships with previously available CDV Arctic-like lineage sequences and to compare the higher number of H gene sequences available at the time of this study, an additional phylogenetic analysis including 53 sequences of 861 nts (from nt residue 469 to 1329) was carried out using the same previous statistical model without gamma distribution and outgroup. The H gene sequences of the Italian CDV strains analysed in this study were submitted to the GenBank/EMBL/DDBJ databases under the accession numbers reported in Table II.

**Results**

A 429-bp fragment of the P gene was detected in samples from 33 of the dogs that were tested (36.3%). Amplicons obtained from the samples of 10 dogs that tested positive (Table II) were sequenced, and the full-length H gene sequence (1824 nts) was determined. The deduced aa sequences (607 aas) were also inferred. The reciprocal comparison of the sequences showed complete identity between 3 CDV sequences (CDV_IZSSI_8387cucciolo_2015, CDV_IZSSI_45182_2015, and CDV_IZSSI_36770PA_2016) while the other sequences showed high identity rates (99.94-99.67%). Comparison with related reference sequences showed high reciprocal identity rates with CDV Arctic-like lineage reference sequences (> 96.3%) and, of these, the highest identity rates were obtained with other European strains: strains from dogs (99.78-99.45%) collected in Switzerland in 2013 and from dogs (99.67-99.23%), wolves (99.72-99.45%), and badgers (99.28-99.01%) collected in Italy in 2013 and 2015. The nt and aa identity rates of the CDV strain (CDV_IZSSI_3853CT_2016) with the closely related Onderstepoort (AF305419) and Rockborn (GU266280) vaccine strains were 91-89% and 95-94%, respectively.

Comparison of the sequences in this study with the first complete H gene sequences of Arctic-like lineage reported in Italy (DQ226087-88) showed 17 single-nucleotide polymorphisms (SNPs), including 9 synonymous and 8 non-synonymous substitutions. The point mutations observed were 13 transitions and 4 transversions. Of these SNPs, G516T was common to the strains collected in Italy in 2000 (HM443714-16-17) while T1754C and A1807C substitutions were characteristic only of the samples in this study, while T1754C and A1807C substitutions were characteristic only of the samples in this study.

**Table IV. Amino acid variations of analysed viral strains shared with Arctic-like lineage and America I/Rockborn-like lineage vaccines sequences.**

<table>
<thead>
<tr>
<th>Strains (accession n.)</th>
<th>H gene amino acid residues</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>172</td>
</tr>
<tr>
<td>2004 Italy(*) (DQ226087)</td>
<td>Leu</td>
</tr>
<tr>
<td>2005 Italy(*) (DQ226088)</td>
<td>-</td>
</tr>
<tr>
<td>CDV_IZSSI_37465CT_2016</td>
<td>Phe</td>
</tr>
<tr>
<td>CDV_IZSSI_36770PA_2016</td>
<td>Phe</td>
</tr>
<tr>
<td>CDV_IZSSI_3853CT_2016</td>
<td>Phe</td>
</tr>
<tr>
<td>CDV_IZSSI_3315CT_2016</td>
<td>Phe</td>
</tr>
<tr>
<td>CDV_IZSSI_980PA_2016</td>
<td>Phe</td>
</tr>
<tr>
<td>CDV_IZSSI_285PA_2016</td>
<td>Phe</td>
</tr>
<tr>
<td>CDV_IZSSI_47586c8_2015</td>
<td>Phe</td>
</tr>
<tr>
<td>CDV_IZSSI_45182_2015</td>
<td>Phe</td>
</tr>
<tr>
<td>CDV_IZSSI_25130_2015</td>
<td>Phe</td>
</tr>
<tr>
<td>CDV_IZSSI_3853cucciolo_2015</td>
<td>Phe</td>
</tr>
<tr>
<td>2013 Swiss(*)</td>
<td>-</td>
</tr>
<tr>
<td>2012-15 Italy(*)</td>
<td>-</td>
</tr>
<tr>
<td>2008 Italy(*) (HM443706)</td>
<td>-</td>
</tr>
<tr>
<td>2000 Italy(*)</td>
<td>Phe</td>
</tr>
<tr>
<td>1988-2006(*) Arctic-like lineage</td>
<td>-</td>
</tr>
<tr>
<td>1994-2007(*) America I lineage vaccines</td>
<td>-</td>
</tr>
<tr>
<td>1999-2009(*) Rockborn-like lineage vaccines</td>
<td>-</td>
</tr>
</tbody>
</table>

Strains were named(*) including year of collection/submission, country and accession number or (*)were grouped (2013 Swiss: KR002265-61; 2012-15 Italy: KX024708-09, KX966928-29, KP144669, KM15552-36, 2000 Italy: HM443714, HM443716, HM443717; 1998-2006 Arctic-like lineage: DQ889178-86, HM443710-13, HM443715, HM443719-22, HM443724, EF446052, EF446051, AF061408, GQ124373, EF122413, Z47769, XM94998; America I lineage vaccines: AF259552, EF418702, Z51949, AF305419; Rockborn-like lineage vaccines: EF095750, AF178039, GU108189, GU266280) and titled including year of collection/submission and country or lineage. (*) except isolate 3148-03 (2003, Austria, GQ124373; Asn603Ser).
and they were all non-synonymous substitutions (Leu172Phe, Ile85Thr, and Asn603His, respectively). Twelve SNPs (A240G, C423T, G843A, G292A, C1026A, A1071G, T1250C, A1278G, G1303A, T1333A, C1524T, C1675T) were common to the sequences of the Arctic-like lineage collected beginning in 2011 (7 synonymous and 5 non-synonymous). Of these, 8 SNPs (at nt residues 240, 423, 843, 929, 1026, 1196, 1250, 1303) were also common in strains collected in Italy in 2008 (HM443706). Compared to the sequences of Arctic-like lineage collected beginning in 2011, 4 SNPs characteristics of sequences collected only in Italy from 2011 to 2015 (3 synonymous – A360G, C1059T, C1626T – and 2 non-synonymous – G583A, A1663G) and 2 characteristics of strains collected only in Switzerland in 2013 (non synonymous – T173C, G1117A) were observed. Nucleotide changes and the predicted amino acid sequence substitutions are shown in Table III.

A total of 5 aa changes was shared by our strains and the Arctic-like CDV reference strains collected after 2011: Gly310Asp, Ile417Thr, Asp435Asn, Leu445Met, Pro559Ser. Another 3 aa changes were observed: change Leu172Phe, common to the 3 strains collected in Italy in 2000 (HM443714-16-17), and the aa substitution Ile85Thr and Asn603His, never previously reported. Other critical aa residues were 195Val and 445Leu, in common with the strains in this study and Arctic CDV strains collected in Switzerland in 2013 (KR002657-61). Amino acids at these residues were different from the most recent Italian Arctic CDV strains collected from 2011 to 2013, except for strains CDV_IZSSI_285PA_2016 and CDV_IZSSI_25130_2015, which showed the aa Met (M) at residue 445. None of the sequences in this study showed change Thr555Ala, observed in some Italian strains collected in 2012/13 (KM115532-36). Predicted amino acid sequence changes shared with Arctic-like CDV reference sequences are shown in Table IV.

Alignment with related sequences of CDV vaccine strains belonging to America I and Rockborn-like lineages showed at least 42 and 32 main aa differences, respectively. These aa changes increased the genomic divergence between field Arctic-like strains and vaccine strains (Table IV).

In all the strains in this study, aa Asn (N) and aa Tyr (Y) were observed at residues 530 and 549, respectively. Amino acid residues crucial for receptor interactions (SLAM and Nectin-4) were conserved (aa residues from 454 to 555).

A total of 8 potential glycosilation sites were predicted at aa residues: 19-21, 149-151, 309-311, 391-393, 422-424, 456-458, 587-589, and 603-605, with the exception of strains CDV_IZSSI_285PA_2016 and CDV_IZSSI_980PA_2016, which lacked the glycosilation site in the aa positions 603-605, due to a non-synonymous substitution that changed the first aa (Asn603His). A non-synonymous substitution affects the second aa position (310Asp) of the potential glycosilation site 309-311. Sequences of

---

**Figure 1.** Neighbour-joining tree based on the full-length hemagglutinin (H) gene sequences (1,824 nucleotides) of canine distemper virus (CDV) displaying the genetic relationships between CDV Arctic-like lineage strains and 68 CDV strains of other lineages (bootstrap 1000 replicates; bootstrap values shown greater than 80%). Black dot markings (-) indicate CDV strains analysed in this study. CDV Arctic-like lineage strains were indicated with: year and country of collection, host, and accession number. CDV strains of other lineages were grouped using the following accession numbers in parentheses: Europe/South America I (AF478543, AF478548, Z77671, DQ494317, Z77673, DQ494319, DQ494318, JR10111, DQ889177, AY838315, HM563059, GQ214376, GQ214384, FJ392652, JN215474, JN215476), South America III (KF835420, KF835414, KF835425), South America II (FJ392651), Europe Wildlife (DQ889188, DQ889187, Z77459, GQ214374, GQ214369, D228166, JN153022, JN153021, JN153023), America I (AY498892, AF164967, AY526496, Z77457, Z77464, Z77467, Z77472, Z77476), America II (AF378705, AY466011, AY548109, AY466011, AF259552, AF418782, DQ778941).
vaccine strains belonging to the America I lineage lacked glycosylation site 309-311 due to an aa substitution at residue 309 (aa Ser in Onderstepoort and Convac strains; aa Arg in the Lederle and Snyder Hill strains) while the Onderstepoort strains also lacked the potential glycosylation site 456-458.

In this study the sequence alignment of all the Arctic CDV strains showed a content of 32 Proline residues, as were observed in all the Arctic CDV strains collected after 2011. Previously collected strains showed a higher number (33-35 residues) of Proline residues, with the exception of 3 strains collected in Italy in 2000 (HM443714-16-17), which also showed 32 residues, but in different positions. The main difference among the Proline residues in the strains collected before 2011 was the lack of the aa due to the change at residue 559 (Pro559Ser). Cysteine residues (12) were maintained within the Arctic lineage. The Asn603His substitution described was located immediately adjacent to the cysteine residue at position 602.

In this study the phylogenetic tree showed the relationship between the sequences and the complete H gene sequences retrieved from GenBank (Figure 1). All strains segregated into distinct branches according to 11 of the previously reported lineages (Budaszewski et al. 2014, Espinal et al. 2014): Europe/South America-I, South America-II, South America-III, America-I, America-II, Asia-I, Asia-II, Rockborn-like, Arctic-like, South Africa, Europe Wildlife. All of our Italian strains clustered within the Arctic lineage (intra-lineage variation was < 3.7%) in a clade that included European strains collected after 2011, were clearly distinct from those of the vaccine strains (America I and Rockborn-like lineages) that were used for comparison. The aa variation with respect to the other lineages was constantly > 4%.

In the additional phylogenetic tree of CDV strains only of Arctic-like lineage (Figure 2), the strains in this study clustered together with strains collected in Italy and Switzerland from 2012 to 2016, and strains collected in Italy in 2000 and 2008. Strains collected in Austria, Italy, Hungary, and the USA from 2000-2006 segregated in a separate cluster. Other strains, including the first Arctic-like strains from Russia and Greenland, segregate separately. The selected nt range for this phylogenetic tree involved 5 (residues 172, 195, 310, 417, 435) of the 9 aa residues described. The strains in this study clustered within a main clade including strains collected in China, Europe, and the USA after 1999, and this clade showed an nt identity rate greater than 98 % when compared to the first strains of this lineage collected in Russia (96.7 %) and Greenland (97.3 %).

**Discussion**

Despite the introduction and extensive use of modified live virus vaccines, Canine Distemper Virus still remains one of the most dangerous viruses for dogs. Similar to other members of the genus *Morbillivirus*, CDV is highly contagious, can cause severe disease, and can result in up to 100 % mortality (Sawatsky & von Messling 2010). Of the CDV lineages spreading in Italy, the Arctic-like CDV strains represent an interesting case of transmission among different host species and of spreading through different countries. The first reports of
Arctic-like lineage strains date back to the late 1980s and include the phocine strain PDV-2 identified in 1988 from a Phoca Siberica of Lake Baikal (Visser et al. 1990), and a canine strain GR88 detected in the same year from a sledge dog population in Northern Greenland (Blixenkrone-Möller et al. 1992). The Arctic-like lineage was then reported in the mid-1990s in China (AF172411), in 2003 in Austria (Benetka et al. 2011), in 2004 in the USA (Pardo et al. 2005), and from a fox in 2005 in China (Zhao et al. 2010) – far from the Arctic ecosystem. From 2004-2006, the Arctic-like lineage appeared once again in Europe; closely related CDV strains were reported by Martella and colleagues (Martella et al. 2006) and Demeter and colleagues (Demeter et al. 2007) in Italy and Hungary, respectively.

A possible explanation for the origin of these strains in separate geographic areas is the introduction of dogs imported from Eastern Europe or Northern Asia to Italy, in addition to an intensified phenomenon of uncontrolled trading of low cost and high value breed pets (Martella et al. 2006, Demeter et al. 2007). A retrospective study reported the spread of Arctic-like CDV strains into Italy as early as 2000 (Monne et al. 2011) but it remained unclear whether this evidence represented occasional findings or if these strains had permanently been established in Italy (Balboni et al. 2014, Martella et al. 2006). Almost 10 years later, the first cases of an Arctic-like CDV strain affecting wolves, dogs, and badgers in a central region of Italy were reported (Di Sabatino et al. 2014, Lorusso & Savini 2014, Di Sabatino et al. 2016). A close connection between strains from dogs and wolves in this area was pointed out, suggesting that the domestic dog could act as a reservoir for a virus that was then passed on to wolves (Di Sabatino et al. 2014, Marcacci et al. 2014). The circulation of this particular CDV lineage in this region of Italy (Abruzzo) had not been reported prior to 2009 (Di Francesco et al. 2012), which suggests a more recent introduction than had previously been hypothesised.

In 2014, 6 Arctic-like CDV strains were reported in other central Italian regions (Emilia-Romagna and Lazio) (Balboni et al. 2014) and, as described in this study, a close genomic relationship was observed with the other CDV strains collected from dogs and wolves in central Italian regions (Lazio and Abruzzo) beginning in 2011. The geographic and temporal evidence of CDV Arctic-like lineage in Europe is represented in Figure 3.

Our study described the molecular features of some CDV strains of Arctic-like lineage collected in

---

**Figure 2.** Geographic distribution of the Arctic-like CDV strains from dogs and wild animals in Europe. Each country (red: Switzerland; yellow: Austria; green: Hungary) or Italian region (blue: Trentino-Alto Adige, Friuli-Venezia Giulia, Veneto; pink: Emilia-Romagna; grey: Abruzzo; brown: Lazio; orange: Sicily) and year of collection are indicated using colour coding, according to the following references: (a) Monne et al. 2011; (b) Balboni et al. 2014; (c) Benetka et al. 2011; (d) Martella et al. 2006; (e) Demeter et al. 2007; (f) Di Sabatino et al. 2014; Lorusso & Savini 2014; Marcacci et al. 2014; (g) Willi et al. 2015; (h) Di Sabatino et al. 2016; (i) present study.
2015–2016 from dogs in southern Italy. Due to the different lengths of some sequences in GenBank, it was possible to integrally compare the entire CDV H gene with most, but not all, Arctic-like CDV sequences available. Nonetheless, it was possible to point out common and interesting features.

Compared to strains circulating from 2000 to 2006, the strains in this study showed different H gene aa profiles at specific residues: 172Phe, 195Val, 310Asp, 417Thr, 435Asn, 445Leu/Met, 559Ser, 585Ile/Thr, 603Asn/His. The majority of these aa residues (195, 310, 417, 435, 445, 559) described a pattern shared by all the sequences of Arctic-like lineage found in Italy since 2011 (Balboni et al. 2014, Di Sabatino et al. 2016), and some of these changes (Gly310Asp, Ile417Thr, Asp435Asn, Pro559Ser) have been constantly observed in all recent European sequences of this lineage, suggesting common ancestors.

Differences were observed at residues 195 and 445, and these 2 amino acid changes represented critical aa differences between both the strains in this study and the Swiss CDV strains with respect to recent strains observed in 2013 in Italy (Di Sabatino et al. 2016). In specific combination with a few other amino acids that were analysed, residue 195 has relevant and compensational effects in the attachment protein of CDV (Sattler et al. 2014); therefore additional studies regarding the potential role of these changes are necessary.

The Leu172Phe change was only observed in the strains considered in this study, as well as in 3 strains collected in 2000 (Monne et al. 2011). Changes to Ile585Thr and Asn603His were unique for only 4 of our strains. Interestingly, changes at residues 585 and 603 were observed in strains from different geographical areas, while the change at residue 172 was a constant change in all the sequences in this study, irrespective of the collection area, and could therefore be considered a potential synapomorphy.

Similarly, specific aa changes with geographic and temporal patterns were also observed in strains collected in Switzerland (Willi et al. 2015) in 2013 (Val58Ala, Glu373Lys) and in Italy (KM115532-35) in 2012/13 (Thr555Ala). These specific changes could contribute to trace CDV spread among different regions or countries. The temporal evidence of the aa changes of CDV Arctic-like lineage is represented in Figure 4.

According to a previous description (Ke et al. 2015), except for changes at residues 172 and 603, all these changes were located in blades B2, B3, B4, and B6 of the β-propeller fold, and some (417, 445) underwent epistatic interactions. Some changes (310, 559, 603) affected the Proline residues and N-linked glycosilation sites. These changes could have played an evolutionary role, due to their potential impact on the protein structures and the unique characteristics of the N-linked site among the wild-type strains at residues 309-311 (Ke et al. 2015).

In fact, N-linked glycosylation sites showed that they were important for the correct folding, transport, and functioning of other paramyxovirus fusion and attachment glycoproteins (Sawatsky & von Messling 2010); therefore, additional studies are necessary to better address any potential biological role.

The intense sampling in Italy in recent years has confirmed the drift acting on the H gene/glycoprotein, and that the spread was mainly driven by a geographic pattern. This supports the theory that the introduction of these most recent CDV strains were carried by dogs imported from Eastern European countries (Balboni et al. 2014, Demeter et al. 2007, Martella et al. 2006). Given the Hungarian origin of the dogs affected in Switzerland and the high identity rates with the sequences of this study, a potential common origin could be possible. The transportation of dogs still plays an important role in the introduction of novel or re-emerging viral strains, as has already been observed for CDV (Martella et al. 2006) as well as for other canine viruses (Decaro et al. 2007, Mira et al. 2018). This evidence also suggests the potential spread to other nearby European countries and, therefore an additional epidemiological survey is necessary.

Our strains clustered within a clade of strains collected after 2008 in Italy and Switzerland (Figure 2) that showed an nt similarity greater than 98 % when compared to the first strains of this lineage collected in China, Russia, and Greenland before 2000. These strains are clearly separated in the phylogenetic tree and, based on these criteria (Budaszewski et al. 2014), they could be argued to belong to a different subgenotype within the Artic-like lineage. The differences were however consistent with the geographical origins of the strains, which are themselves a result of separate disseminations across the Arctic ecosystem and

---

**Figure 4.** First evidence of H gene sequence amino acid changes of CDV Arctic-like lineage strains, indicated with year, host, and amino acid residue. Positions were based on reference sequences with the accession numbers DQ226087 and DQ226088, according to the following references: 34 Monne et al. 2011; 35 Martella 2013 (unpubl.), Balboni et al. 2014, Di Sabatino et al. 2014, Lorusso & Savini 2014, Maracci et al. 2014; 36 Di Sabatino et al. 2016; 37 present study.
A large body of literature has reported the propensity of CDV to jump species barriers into canine and non-canine hosts, seals, and non-carnivores (Ludlow et al. 2014). In fact, recent cases of CDV infection in wolves (Di Sabatino et al. 2014) and badgers (Di Sabatino et al. 2016) in Italy and tigers in Russia (Seimon et al. 2013) have been reported. Wolves and tigers are subspecies listed as endangered by the International Union for the Conservation of Nature (www.iunc.org). As has previously been suggested (Cleaveland et al. 2000), domestic dogs have been considered to be a likely source of CDV infection; therefore, additional studies are necessary to evaluate any potential CDV spread into wild animals in the same Italian areas.

In conclusion, our findings contribute to knowledge regarding the Arctic-like CDV lineage in Italy and describe CDV strains with specific aa profiles that are spreading in Europe; thus, additional immunisation programmes to improve vaccine coverage among susceptible dogs are necessary to control CDV (Rikula et al. 2007). Even if complete H gene amplification from field samples is difficult due to the scale of the task size and its transcription level (Panzera et al. 2014), complete H gene sequence and related metadata still remain useful in obtaining the complete comparison of circulating strains. This, in turn, facilitates a better understand of the epidemiological features of CDV.
References


