UNIVERSIDADE NOVA DE LISBOA

Universidade Nova de Lisboa

Instituto de Higiene e Medicina Tropical

Evaluation of Human *mastadenovirus* molecular diversity in wastewater and environmental water from the Lisbon Metropolitan Area

Joana Filipa Nunes Cavadas

DISSERTATION TO OBTAIN THE MASTER'S DEGREE IN MEDICAL MICROBIOLOGY

OCTOBER 2022

DESDE 1902 INSTITUTO DE HIGIENE E MEDICINA TROPICAL UNIVERSIDADE NOVA DE LISBOA

itqb nova

NOVA SCHOOL OF SCIENCE & TECHNOLOGY

NOVA MEDICAL SCHOOL

Universidade Nova de Lisboa

Instituto de Higiene e Medicina Tropical

Evaluation of Human *mastadenovirus* molecular diversity in wastewater and environmental water from the Lisbon Metropolitan Area

Author: Joana Filipa Nunes Cavadas

Supervisor: Doctor Mónica Nunes, Associated Scientist
Instituto de Biologia Experimental e Tecnológica

Co-supervisor: Professor Doctor Ricardo Parreira, Associate Professor,
Instituto de Higiene e Medicina Tropical,
Universidade NOVA de Lisboa

Dissertation presented to fulfill the necessary requirements to obtain a master's degree in Medical Microbiology

Financial support from AgriWWater: National call 02/SAICT/2017, project number 29586, co-financed by:

Bibliographic Elements

Dissertation publications in peer-reviewed journals:

Cavadas JC, Parreira R, Leonardo I, Crespo MTB, Nunes M. Mastadenovirus Molecular Diversity in Waste and Environmental Waters from the Lisbon Metropolitan Area. Microorganisms. 2022;10(12):2443. https://doi.org/10.3390/microorganisms10122443

Oral Communications:

Cavadas JC, Parreira R, Leonardo I, Crespo MTB, Nunes M. 2022. Caracterização molecular de adenovírus humanos presentes em águas residuais e ambientais da área Metropolitana de Lisboa. Microbiologia 2022 – Congresso Internacional de Microbiologia em Língua Portuguesa. 18th October 2022. Online. Flash presentation.

Acknowledgements

I would like to express my deepest gratitude to my supervisor, Doctor Mónica Nunes, for her patience, unwavering support, advice, and encouragement throughout this dissertation. Without her invaluable guidance, this project would not have been possible.

Also, I'm extremely grateful to my co-supervisor, Professor Doctor Ricardo Parreira, for all knowledge and expertise provided, constant assistance, availability, and helpful suggestions, which were essential for the achievement of the work described in this dissertation.

The development of this study was possible due to a collaboration with Águas de Portugal, whose I'm thankful for granting us access to the several WWTPs across the LMA.

I would like to extend my sincere thanks to all the members of the Food Safety & Microbiology lab from iBET - Instituto de Biologia Experimental e Tecnológica, with whom I had the pleasure of working. In particular, to Doctor Teresa Crespo, for the opportunity to work in her lab. Also, special thanks to Doctor Andreia Silva for all given support during the time we shared in the lab. I would be remiss in not mentioning Inês Leonardo, who helped me countless times. Her patience with all my questions, encouragement, advice, availability to assist me in any way she could, and, especially, her friendship were essential during this process.

I am also grateful to my family for the opportunity, support and mainly for all the sacrifices made so I could achieve this master's degree and do this dissertation work. I'd like to acknowledge my boyfriend for his belief in me, for keeping my motivation high during this course, and for his patience to ear all my concerns and lab stories, even not understanding them.

Lastly but not least, I'd like to mention my friends for the much-needed breaks of work we shared, advice and support provided. Especially, to Jéssica Lopes and Joana Neves, whose friendship made every master's degree challenge so much easier.

Abstract

Many pathogens can be transmitted to humans through contact/consumption of contaminated water, representing a risk to public health. The microbiological contamination is mainly due to the discharge of untreated wastewater, where these agents, including viruses, can remain for long periods. The conventional processes used in wastewater treatment plants (WWTPs) cannot remove/inactivate the totality of microorganisms including some pathogenic, which may contribute to their dissemination in the environment. Water quality is evaluated by faecal indicator bacteria; however, the use of viral indicators has been proposed, due to their wide distribution, high stability in water matrices, and lack of seasonality in their circulation. Among several possibilities, human adenoviruses (HAdV) stand out for their high resistance to the disinfection processes applied in WWTPs, and for their high prevalence in the environment. These viruses are divided into seven species (from A to G), and are responsible for serious pathologies, mainly in vulnerable populations (children, elderly, and immunocompromised individuals).

This work involved the evaluation of HAdV genomic diversity in waters of the Lisbon Metropolitan Area (LMA). For this, WWTP influents and environmental samples were collected between 2018 and 2021. After the concentration of viral particles by organic flocculation and extraction of viral DNA, touch-down PCR protocols were developed for the amplification of part of the coding region of one of the viral capsid proteins (hexon). The amplification products obtained were analysed by a molecular cloning process in which, after random selection of some clones, recombinant vector (plasmid) DNA molecules were sequenced using the Sanger method. Additionally, these products were sequenced by the Illumina method, a Next Generation Sequencing (NGS) approach. The sequences obtained were evaluated by combining the search for homologous sequences available in the public genomic databases, and a phylogenetic analysis, allowing us to associate them to a HAdV species and to identify which ones are circulating in the LMA.

The Sanger sequencing-based approach revealed the presence of viral genomes from the HAdV-F species (37 %) and possible from murine AdV (63 %) in the environmental samples, while species HAdV-A (6 %), -C (18 %), -D (18 %) and -F (58 %) were identified in the WWTPs influent samples. As expected, the analysis of the sequences obtained by NGS allowed the identification of a higher diversity of viral genomes belonging to the species HAdV-A (19 %), -B (1 %), -C (3 %), -D (24 %) and -F (25 %), and possibly murine AdV (30 %) in the WWTPs samples. For the environmental samples, species HAdV-A (19 %), -D (32 %) and -F (36 %) were identified along with possible murine AdV (14 %).

The presence of different HAdV types in the water samples studied demonstrates that the treatments applied in the WWTPs are not efficient in removing viral particles, and the presence of faecal contamination in the environmental waters. Furthermore, the assessment of the diversity of this virus (through the study of its genome) provides important information about the patterns of its molecular epidemiological distribution in the population.

Keywords: Human adenovirus, Lisbon metropolitan area, wastewater, environmental waters, NGS.

Resumo

Muitos agentes patogénicos podem ser transmitidos aos humanos pelo contacto/consumo de água contaminada representando um risco para a saúde pública. A contaminação microbiológica deve-se maioritariamente à descarga de águas residuais (AR) onde esses agentes, incluindo vírus, podem permanecer por longos períodos. Os processos convencionais usados nas estações de tratamento de água residual (ETARs) não conseguem remover/inativar na totalidade os microrganismos, incluindo alguns patogénicos, podendo contribuir para a sua disseminação no ambiente. A qualidade da água é avaliada através de bactérias indicadoras de contaminação fecal, no entanto, tem sido proposto o uso de indicadores virais, devido à sua ampla distribuição, alta estabilidade nas matrizes de água, e ausência de sazonalidade na sua circulação. Entre várias possibilidades, os adenovírus humanos (HAdV) destacam-se pela sua alta resistência aos processos de desinfeção aplicados nas ETARs e elevada prevalência no ambiente. Estes vírus dividem-se em sete espécies (de A a G) e são responsáveis por patologias graves, principalmente em populações vulneráveis (crianças, idosos e indivíduos imunocomprometidos).

Este trabalho envolveu a avaliação da diversidade genómica de HAdV em águas da área Metropolitana de Lisboa (AML). Para tal, foram recolhidas amostras de influentes de ETARs e águas ambientais entre 2018 e 2021. Após concentração das partículas virais por floculação orgânica e extração do DNA viral, foram desenvolvidos protocolos de *touch-down* PCR para a amplificação de parte da região codificante de uma proteína da cápside (hexão). Os produtos de amplificação obtidos foram analisados por um processo de clonagem molecular em que, após seleção aleatória de alguns dos clones, se procedeu à sequenciação das moléculas de DNA plasmídico recombinante pelo método de Sanger. Adicionalmente, esses produtos foram sequenciados pelo método de Illumina, numa abordagem de Sequenciação de Nova Geração (NGS). As sequências obtidas foram analisadas combinando a pesquisa de sequências homólogas disponíveis em bases de dados e análise filogenética, permitindo associar as sequências a uma espécie/tipo de HAdV, e a identificar quais os circulantes na AML.

A sequenciação de Sanger revelou a presença de genomas virais da espécie HAdV-F (37 %) e possivelmente de AdV de murinos (63 %) nas amostras ambientais, enquanto as espécies HAdV-A (6 %), -C (18 %), -D (18 %) e -F (58 %) foram detetadas nas amostras de influentes das ETARs. Como esperado, a análise das sequências obtidas por NGS permitiu a identificação de uma maior diversidade de tipos virais pertencentes às espécies HAdV-A (19 %), -B (1 %), -C (3 %), -D (24 %), -F (25 %), e possivelmente de AdV de murinos (30 %) nas amostras de influente das ETARS. Nas amostras ambientais, foram detetadas as espécies HAdV-A (19 %), -D (32 %), -F (36 %) e provavelmente de AdV de murinos (14 %).

Assim, estes resultados indicam a presença de diferentes tipos de HAdV ao nível das AR e em águas ambientais, demonstrando a ineficiência dos tratamentos aplicados nas ETARs em remover partículas virais, e a presença de contaminação fecal nas águas ambientais. Adicionalmente, a avaliação efetuada da diversidade deste vírus (através do estudo dos seus genomas) fornece informações importantes sobre os padrões da sua distribuição epidemiológica molecular na população.

Palavras-chave: Adenovírus humanos, área metropolitana de Lisboa, águas residuais, águas ambientais, NGS.

Contents

List of figures

List of tables

Abbreviations

µL – microliters

µm – micrometres

ADB – Agarose Dissolving Buffer

AdV – Adenoviruses

AIDS – Acquired Immunodeficiency Syndrome

aLRT – Approximate Likelihood-Ratio Test

AVP – Adenovirus Protease

AW1 – First Washing Buffer

AW2 – Second Washing Buffer

BLASTn – Nucleotide Basic Local Alignment Search

bp – Base pairs

BR – Broad-Range

CCL – Contaminant Candidate List

Cq – quantification cycle

DBP – DNA Binding Protein

DNA – Desoxyribonucleic acid

dsDNA – double-stranded DNA

EDTA – Ethylenediaminetetraacetic acid

EV – Enteroviruses

FIB – Faecal Indicator Bacteria

FS&M – Food Safety & Microbiology lab

FW – Forward

g – gravitational force equivalent, or g-force

h – hours

HAdV – Human Adenoviruses

HAV – Hepatitis A Virus

HIV – Human immunodeficiency virus

HS – High Sensitivity

ICTV – International Committee for the Taxonomy of Viruses

Id - Identity

IPTG – Isopropyl β-D-1-thiogalactopyranoside

L – Litres

LB – Lysogeny Broth

LMA – Lisbon Metropolitan Area

M - Molar

MAdV – Murine adenoviruses

min – minutes

ML – Maximum Likelihood

mL – millilitres

MLTU – Major Late Transcriptional Unit

mM – millimolar

mRNA – messenger RNA

NCBI – National Center for Biotechnology Information

ng – nanogram

NGS – Next Generation Sequencing

NJ – Neighbour-Joining

nm – nanometre

nM – nanomolar

NoV – Noroviruses

PCR – Polymerase Chain Reaction

pM – picomolar

pTP – preTerminal Protein

PyV – polyomaviruses

QC - Query cover

qPCR – Quantitative (real-time) PCR

RNA – Ribonucleic acid

RoV – Rotaviruses

rpm – Revolutions per minute

RV – reverse

s – seconds

SOC – Super Optimal Broth with Catabolite repression

ssDNA – single-stranded DNA

TAE – Tris-Acetate-EDTA

TE – Tris-EDTA

Tm – melting temperature

TP – Terminal Protein

USEPA – United States Environmental Protection Agency

UV – Ultraviolet

VLP – Viral-like particles

WHO – World Health Organization

WWTP – Wastewater Treatments Plants

X-gal – 5-Bromo-4-chloro-3-indolyl-β-D-galactopyranoside

1. Introduction

1.1. Viral transmission through contaminated water

Water contamination with pathogenic agents represents a health risk for human populations, as many of them are responsible for infectious diseases and can be transmitted by contact with, or consumption of, water or water-contaminated food items. Not surprisingly, around the world, several outbreaks associated with the presence of waterborne pathogens have been reported (1,2).

The quality of water is determined by the levels of faecal indicator bacteria (FIB), which comprises faecal coliforms as *Escherichia coli* (*E. coli*), enterococci such as *Enterococcus faecalis, E. faecium*, and also *Clostridium perfringens* (3,4). To be an ideal assessor of faecal contamination, the FIB should fulfill as many possible of the criteria established by the World Health Organization (WHO) and the United States Environmental Protection Agency (USEPA). These criteria include 1) presence in all types of water; 2) maintenance of viability/infectiousness under environmental conditions, 3) presence whenever enteric pathogens are present, 4) inability to grow in water, 5) presence in faeces and the gastrointestinal tract of warm-blooded animals, and 6) easy detection (5,6). Although the levels of FIB are monitored, not all pathogens are completely removed and/or inactivated by the conventional processes applied in water and Wastewater Treatments Plants (WWTP) (7). Thus, pathogens other than bacteria can remain in WWTP and, consequently, can be introduced into the aquatic environment. Among these we find viruses. Since the main cause of faecal pollution in the aquatic environment is the discharge of raw/untreated sewage into the water (8), and since viruses can remain infectious in the water at high levels regardless of the treatment the water matrix is subjected to (4,8,9), these agents represent a high risk to human populations.

Generally, viruses are more infectious, occur in higher concentrations (3), and can withstand longer periods of environmental circulation than most intestinal bacteria (10), because they are more resistant to environmental degradation (11). For this reason, it is not safe to depend only on bacteriological standards to evaluate the water quality (10), as they underestimate the public health risk that virus-contaminated water poses to the human population (3). As a result, to minimise the risk of infection to humans it is important to understand the environmental distribution of these pathogens, taking into

account the risk of waterborne virus-associated illnesses, as well as the need to monitor viral contamination through sewage and/or environmental analysis (7,9). Some authors have proposed the use of a viral indicator of faecal water contamination. Among the viruses suggested as possible water quality indicators are human adenoviruses (HAdV) (4,12,13), due to their high resistance to water treatment and disinfection processes, high prevalence in all geographic areas/in the population surveyed, and host-specificity (13).

The most studied waterborne pathogens are human enteric viruses, which are responsible for sporadic or outbreak-related diseases throughout the world involving thousands of cases of human infection (10,11). Enteric viruses are frequently related to subclinical infections but also to multiple symptomatic illnesses, like acute gastroenteritis, acute hepatitis, central nervous system infections (meningitis, encephalitis, and paralysis), conjunctivitis, or respiratory diseases (14). These viruses are more problematic for vulnerable groups such as young children (11). The group of enteric viruses frequently studied include polioviruses, enteroviruses (EV), coxsackieviruses, hepatitis A virus (HAV), echoviruses, adenoviruses (AdV), noroviruses (NoV), caliciviruses, astroviruses, reoviruses, rotaviruses (RoV) and polyomaviruses (PyV) (15). Regardless of whether the infections they cause evolve to become symptomatic or not, they replicate in the gastrointestinal tract and are excreted for weeks in large quantities in the faeces and urine of infected individuals (10,14). Thereby, these viruses are transmitted via the faecal-oral route, occur in high concentrations in wastewater and circulate in the community sewage system. Thus, they are frequently introduced into the environment through the discharge of treated or raw sewage, causing disease by the consumption of/contact with contaminated water sources (9,10,16). In addition, humans and/or other animals can directly contaminate the water (11) through their secretions and body fluids, such as faeces and urine.

Due to their stability and persistence, AdV, HAV, NoV, RV, EV, and PyV are, among all enteric viruses, the ones most frequently detected in the environment (9) leading to possible human infections through the contamination of drinking water sources and recreational waters, making them important in terms of public health (17).

1.2. Adenovirus (Adv)

At the moment, the International Committee for the Taxonomy of Viruses (ICTV) recognizes 87 AdV species and six genera (*Mastadenovirus, Aviadenovirus, Atadenovirus, Siadenovirus, Ichtadenovirus* and *Testadenovirus*) in the *Adenoviridae* family (https://ictv.global/report/chapter/adenoviridae/adenoviridae, accessed on 20/09/2022) (18). The AdV placed within the *Mastadenovirus* genus infect a wide range of mammalian species including humans, bats, bovines, canines, deer, dolphins, equines, different murine species, non-human primates, ovine, swine, sea lions, skunks, squirrels, and tree shrews (19). The AdV which only infects birds is placed within the *Aviadenovirus* genus belong, while Atadenoviruses include viruses that infect various hosts, including reptiles, birds, opossums, and ruminants. Siadenoviruses include bird, reptile, and amphibian AdV. The *Ichtadenovirus* genus includes fish AdV, while the AdV that comprise the *Testadenovirus* genus include viruses of turtles (20).

AdV are included in the Contaminant Candidate List (CCL) from USEPA, which is a list of drinking water contaminants that are known to occur in public water systems (https://www.epa.gov/ccl/draft-ccl-5-microbial-contaminants, accessed on 13/07/2022).

1.2.1. Physical characteristics of the agent

The AdV virions correspond to medium-sized (90-100 nm) non-enveloped infectious particles containing a linear double-stranded desoxyribonucleic acid (dsDNA) genome, which contributes to their resistance to environmental degradation compared to many other enteric viruses (11) since dsDNA is more stable than single-stranded DNA (ssDNA) or ribonucleic acid (RNA) genomes.

Structurally, AdVs are composed by the external icosahedral capsid and the internal core (Figure 1) which enclose the genome (21). The capsid contains 240 capsomers with six-fold symmetry, the hexon proteins, which form the icosahedron sides of the virion and account for 60% of the viral particle's total mass. The capsid is also constituted by 12 capsomers with five-fold symmetry, the penton base proteins, which form the vertices of the particle (22,23). From each vertex is projected a fiber, which corresponds to a glycoprotein constituted by three domains: the tail, the shaft, and the knob (19). The tail, which corresponds to the fiber's N-terminus, forms a non-covalent interaction with the

penton base protein and is responsible for the attachment of the fiber to the capsid. Depending on the virus type, the shaft varies in length. The knob corresponds to the fiber's C-terminus and is responsible for the interaction with cell receptors, allowing viruses to attach to susceptible cells. As a result, the knob of the fiber determines viral tropism. The penton is the non-covalent complex formed by the penton base protein and the fiber. This complex is fundamental in the initial stages of infection, as they interact with cell receptors of the host and contain all the necessary information for virus attachment and internalization (19,23).

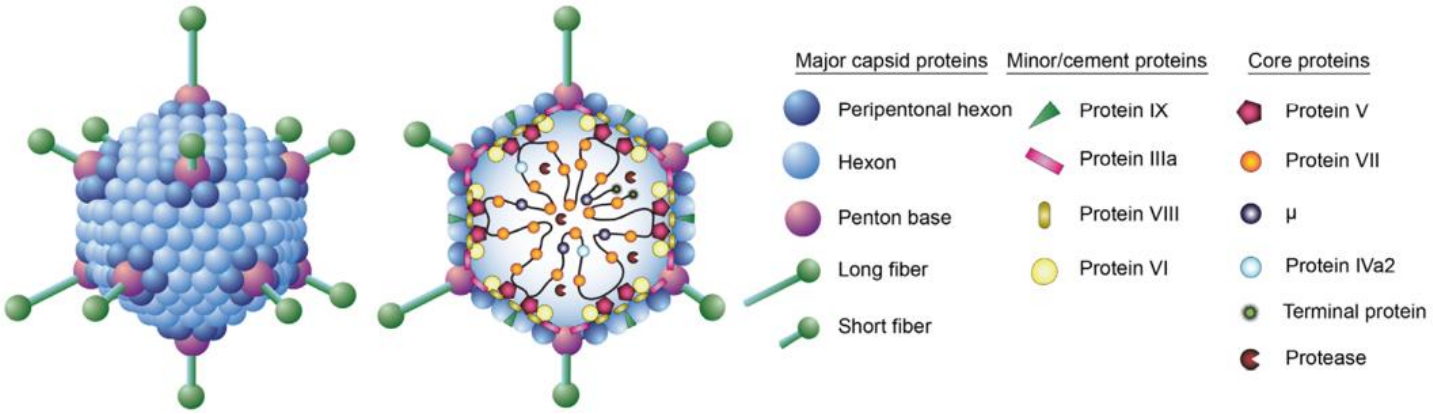

Figure 1. Schematic representation of the external side (left) and the internal side (right) of AdV virion. Adapted from Rafie et. al. (24).

The viral particles also contain several internally located proteins (Figure 1), some of which interact with the penton and hexon subunits, while others are linked to the viral genome (20). These coat minor proteins implanted in the capsid, such as IIIa, VI, VIII, and IX, help the morphogenesis and contribute to maintaining viral stability (25). The number of minor coat proteins and their organization can vary between genera and species. Specifically, protein IX is only found among the members of the *Mastadenovirus* genus, whereas proteins IIIa, VI, and VIII are preserved throughout the whole *Adenoviridae* family and seem to be essential for viral assembly (23). There are also many virus-encoded DNA-binding proteins (DBP) packed inside the virion alongside the genome, forming an internal non-icosahedral core. Some of the core proteins such as V, VII and protein X (also called μ peptide) are thought to be DNA-condensing agents, as they are responsible for the

organization of the genome (23,25). Additionally, there is covalently linked a terminal protein (TP) to each 5' end of the viral genome, and this protein is essential for genome replication. Finally, the other core proteins present a role in genome packing (IVa2) and maturation (adenovirus protease, AVP).

1.2.2. Genome and replication cycle

Even though all AdVs share the same virion morphology (Figure 1), their genomic organization differs according to their distribution into different genera (20). AdV genome ranges between 26,000 and 48,000 base pairs (bp) depending on the virus genus. For example, while the ovine adenovirus 7 (*Atadenovirus*) genome has approximately 30,000 bp, human adenovirus 2 (*Mastadenovirus*) has 35,000 bp, fowl adenovirus 1 (*Aviadenovirus*) displays the largest genome with more than 40,000 bp, and frog adenovirus 1 (*Siadenovirus*) has the smallest genome with approximately 26,000 bp (20).

The genome is organized into early, intermediate and late transcriptional units, each one comprising multiple genes which are expressed before and after replication of the viral genome (21,26). Each one of these regions is controlled by a single promoter that is used by the transcriptional machinery of the host cell (20). Figure 2 represents the general viral replication cycle of AdVs. To begin, the fibre knobs attach to the receptor present on the surface of the host cell. This interaction allows the entry of the particle into the cell by receptor-mediated endocytosis. After the uncoating of the viral particle, the genome is transported to the nucleus. There the early region, which encodes five genes (E1A, E1B, E2, E3, and E4), is transcribed by the host cell RNA polymerase II (27). After transcription, the early messenger RNAs (mRNAs) are processed, exported to the cytoplasm, and translated. The early proteins include the viral replication machinery (DNA polymerase, pre-Terminal Protein - pTP, and DBP) (28), which are imported into the nucleus and cooperate with cellular proteins to ensure viral DNA replication. The latter proceeds from both ends by a strand displacement mechanism, using the pTP as a primer. Additionally, other early proteins are necessary for inducing the host cell to enter the S phase of the cell cycle (to create a more conducive environment for viral DNA replication), and for protection against host antiviral immune mechanisms (20). The viral

DNA molecules already replicated can serve as a template for either additional replication rounds or transcription of the major late transcriptional unit (MLTU). The MLTU comprises five genes (L1 – L5) and is activated by viral DNA replication, however, maximally efficient transcription requires the transcription of intermediate genes. After transcription, the late mRNAs are exported from the nucleus and translated into structural proteins required for the assembly of viral particles and packaging of the viral genome (20,27). Near the end of the replication cycle, the newly synthesized structural proteins and viral DNA molecules are assembled within the nucleus into non-infectious immature virus particles. After cleavage of the precursors of some proteins by the AVP, the mature virions are released, typically upon cell lysis (27).

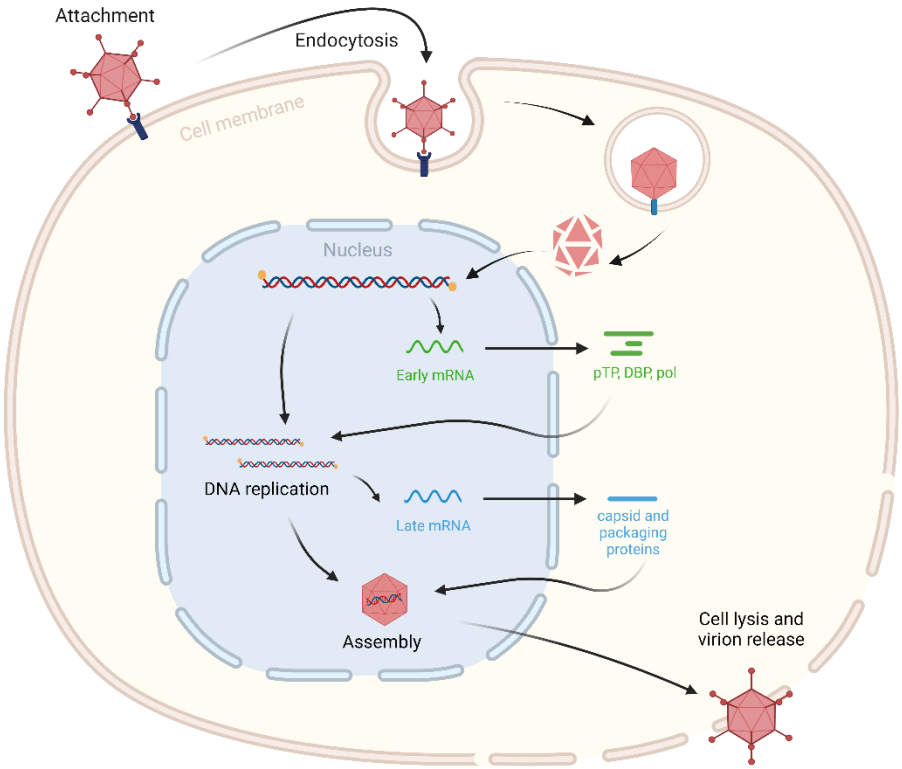

Figure 2. Schematic representation of AdV replication cycle. The viral particle enters the host cell by endocytosis. After uncoating of the particle, the viral genome is transported to the nucleus, where the early transcriptional unit (green) is transcribed. Afterwards, the late proteins (blue) are produced and begin the assembly stage. The mature virions usually exit the cell upon its lysis.

1.2.3. Human adenoviruses (HAdV)

Considering the epidemiological importance, easy detection, and significant occurrence in water, a particular focus has been devoted to HAdV (1). As mentioned above, these viruses have been suggested as a surrogate virological water quality indicator, i.e., HAdVs are among the potential indicators for assessing the presence of human-specific viral pathogens in the environment. In addition, these viruses can be useful for the determination of the efficiency of the disinfection and/or water purification treatments used (4,29,30).

More specifically, HAdVs can be used as a viral indicator of wastewater contamination due to their characteristics such as 1) wide distribution (31), 2) high stability in multiple water matrices from wastewater to surface water (including rivers, seawater, coastal water, swimming pools and water supplies worldwide) (32), 3) environmental persistence revealed by the detection of infectious particles in waters treated by chlorine and in the ultraviolet (UV) disinfected wastewaters (1,33), 4) persistence capacity, since these viruses can withstand longer in the environment than FIB (4), 5) the lack of seasonality in their distribution, shown by its continual detection in sewage (31), 6) higher abundance in waters compared to other enteric viruses (31), 7) specificity to humans hosts (31), and 8) high stability of their genome when compared to other enteric viruses, especially RNA viruses (1).

In the human body, HAdV resist to gastric and biliary secretions (22) and all types (enteric and non-enteric) may be excreted in high concentrations in faeces from infected individuals, even asymptomatic ones (34–36). Under these circumstances, it is important to monitor the disinfection efficacy of WWTP processes to evaluate the risk that the discharge of HAdV-contaminated effluent into watercourses might have on public health. Considering that these watercourses, which receive treated wastewater, often provide drinking water, and can be used for recreational activities and irrigation of parks, sports fields, fruits, and vegetables, if HAdVs are introduced and remain virulent in water, they may be *bona fide* agents of disease for the exposed communities. Additionally, Farkas et al. (37) suggested that AdV may indicate how persistent are enteric viruses, so the adequate control of AdV in water may suggest that other enteric viruses are also controlled (3).

1.2.4. Taxonomy and types of HAdV

HAdVs are categorized into seven species (A through G) based on their physical, chemical, genetic and biological properties (19,26). Species A, B, C, D, E, and F circulate globally and have been associated with limited periodic outbreaks of infection in humans (38,39). On the other hand, the G species is rarely detected (40).

Historically, HAdV serotypes were determined by classical methods such as viral neutralization assays, and hemagglutination properties. However, HAdVs are currently classified mostly into genotypes by sequence analysis (21). Consequently, there are the initial 51 recognized serotypes, and more than 100 genotypes have already been described (http://hadvwg.gmu.edu/ accessed on 20/09/2022) by bioinformatics analysis. Considering that individual HAdV "type" nomenclature has not reached a global consensus within the scientific community, the format 'human adenovirus dash species type' (e.g. HAdV-C5) will be used in this work for sero- and genotypes.

As mentioned, to date more than 100 HAdV types are known (Table 1) (41). The different types do not exhibit the same tissue tropism, and this is correlated with the heterogeneity of the clinical manifestations (34,39). These manifestations range from mild and self-limited infections, which involve the upper or lower respiratory tract, the gastrointestinal tract or the conjunctiva, to severe manifestations such as haemorrhagic cystitis, hepatitis, haemorrhagic colitis, pancreatitis, nephritis or meningoencephalitis (21,38,41). Vulnerable populations (children, elderly, and immunocompromised individuals) have a higher risk of developing severe disease (21,34) and, in some cases, infections may be fatal (19). On the other hand, HAdV infections may also be asymptomatic (34) or are characterized by clinical undifferentiated symptoms that are not easily recognizable as associated with infection by these viruses in a healthy population (19).

The diseases associated with each HAdV species are listed in Table 1. Of all seven species, HAdV-A, -F and -G viruses have selective tropism for the gastrointestinal tract (38,42). The HAdV-B species can cause respiratory infections, as well as urinary tract infections (21). It has been reported that the HAdV-C species cause a major percentage of acute respiratory tract infections in children. Out of the seven species, the largest is the HAdV-D which is mainly responsible for moderate or asymptomatic diseases, except for a few types (HAdV-D8, -D37, -D53, -D54, -D64) that are responsible for epidemic

keratoconjunctivitis (21,39,41). Also, some viral types of HAdV-D have been suggested to display an opportunistic pathogenicity profile, being detected in Human Immunodeficiency Virus (HIV) infected individuals. Regarding the HAdV-E, which is only composed by one type (HAdV-E4), they were found to be usually involved in outbreaks of acute respiratory disease in military units (21), but can also be responsible for conjunctivitis.

Table 1. Associated diseases or infections with each HAdV species. Adapted from Crenshaw et al. (26) and Heim et al. (41).

Species	Type	Associated Disease
A	12, 18, 31, 61	gastrointestinal, respiratory, urinary, cryptic enteric infection, linked to obesity, meningoencephalitis
B	3, 7, 11, 14, 16, 21, 34, 35, 50, 55, 66, 68, 76–79	conjunctivitis, gastrointestinal, respiratory, urinary, pneumonia, meningoencephalitis, cystitis
C	1, 2, 5, 6, 57, 89	respiratory, gastrointestinal, obesity, pneumonia, hepatitis
D	8–10, 13, 15, 17, 19, 20, 22–30, 32, 33, 36–39, 42–49, 51, 53, 54, 56, 58-60, 63-67, 69–75, 80–88, 90–103	epidemic keratoconjunctivitis, gastrointestinal, linked to obesity, meningoencephalitis
E	4	conjunctivitis, respiratory, pneumonia
F	40, 41	gastrointestinal, infantile diarrhoea
G	52	gastrointestinal

Although both healthy children and adults can be infected, it seems that these infections are more prevalent in young children (38) with 5% of childhood respiratory tract infections being caused by HAdV (21).

Since different HAdV types can replace/change the dominance of other types, the most frequently disease-associated viral types may fluctuate over time and vary among countries (26). Until now, HAdV-B3, -B7, -B21, -C1, -C2, -C5, -E4, and -F41 types are frequently reported as being associated with human disease (22,38). Given that HAdV-B3, -B7, -B14, -B55, -D8 and -E4 types are mainly responsible for respiratory diseases, they are more likely to spread and, consequently, are commonly related to outbreaks (26). In addition, some of these types (e.g. HAdV-B7) appear to be more virulent than others (38). For example, a mutation on a viral packaging protein coding gene of the HAdV-B7 type was associated with faster viral replication, where the authors concluded that perhaps this was related to its increased clinical virulence (39).

1.2.5. Transmission of HAdV

HAdV outbreaks usually occur in overpopulated/overcrowded areas (such as nosocomial and long-term care facilities, military bases, schools, and childcare centres) due to their quick spread resulting from facilitated transmission in close communities (21,26,43).

The most common mode of transmission of HAdV is via the respiratory tract (26), due to the exposure of the mouth/nose mucosa to aerosolized droplets from a cough or a sneeze from infected individuals. However, these viruses can also be transmitted by the faecal-oral route (22), which occurs by ingestion of contaminated food and/or consumption of/contact with contaminated water, such as public swimming pools, recreational marine water or even tap water (21,22,26). Occasionally, transmission can also occur during organ transplantation (21), or through poor hygiene practices, such as improper handwashing (26) whereby HAdV may be directly inoculated into the conjunctiva (22). Although HAdVs are excreted in high concentrations in faeces, their presence in urine is only associated with acquired immunodeficiency syndrome (AIDS) patients, and that is why this excretion is considered insignificant for viral transmission (22).

1.2.6. Surveillance diagnostic tools of HAdV

Traditionally, HAdV identification has been based on virus isolation in cell culture, antibody or antigen detection, and electron microscopy visualization of the viral particles due to their distinctive morphology (44). However, the most widely used methods for the detection of enteric viruses from environmental samples involve the use of molecular techniques such as Polymerase Chain Reaction (PCR). PCR-based detection assays have several advantages in revealing the presence of viral pathogens from environmental samples. This technique is sensitive (allows the detection of viruses present in low quantity), specific, as well as less laborious, expensive and time-consuming than other methods (e.g. virus isolation) (15). Among the disadvantages of this approach stand out the possibility of false-positive (due to cross-contamination) or false-negative amplifications (due to the presence of inhibitors of the amplification reaction). Nevertheless, several variations of this basic method have been performed to improve

their specificity, sensitivity, and efficiency but also to quantify the number of viruses detected. Some variations include nested, touch-down, multiplex, and quantitative (real-time) PCR (qPCR), and their multiple possible combinations. Specifically, for HAdV detection by PCR and sequencing techniques, the main targets are usually the fiber or the hexon-coding genes (38).

The identification and quantification of HAdV in any given water source are critical to assess viral diversity and titer, as well as the potential hazards associated with viral contamination of water. In addition, knowledge of enteric virus diversity and concentration can aid in determining the treatments required, for example, for effective pathogenic virus inactivation in recycled water (42).

Considering that the traditional approaches (including fragment-directed cloning and recombinant vector sequencing methods) may be incapable to provide a precise impression of the genetic viral diversity in wastewater systems, more reliable methods for the detection of enteric viruses are needed. Over time, the multiple improved versions of the alternative Next Generation Sequencing (NGS) technologies have allowed the determination of HAdV and other viruses in complex water samples. In particular, NGS can provide a more reliable picture of the viruses that circulate in the population (42). Considering pathogen surveillance and discovery, in both research and the public health sector, the Illumina MiSeq system has become the most popular platform used (45).

1.2.7. Worldwide surveillance of HAdV

Around the world, several studies have focused on the detection and quantification of HAdVs in multiple water matrices. Thus, these viruses have been widely detected in wastewater (influent and effluent sewage), surface water, and marine and freshwater samples, but they can also be found in treated wastewater (14), disinfected drinking water (36) and tap water (46), showing that these viruses cannot be completely removed by some of the treatments used in the WWTP. Additionally, these viruses have been reported as some of the most common contaminants of surface water all over the world (44). For example, some randomly selected epidemiological scenarios demonstrated the presence of HAdV in 100 % of the river samples from Barcelona and Rio de Janeiro (8). Likewise, in Italy, these viruses have also been detected in rivers and seawaters, along with the

influent and treated effluent of a WWTP, with a decrease in frequency between their presence in the influent (100 %, and the effluent 94 %), river (72 % to 92 %) and seawater (21 %) (1). Dias et al. (47) detected these viruses in all recreational coastal marine environments analysed in Brazil, even in sampling sites considered suitable for bathing according to the bacterial standard counts, as also reported by Kundu et al. (48) in southern California. Additionally, Elmahdy et al. (49) detected a high number of genome copies of HAdV in the final-treated wastewater effluent and Nile River water, and Opere et al. (50) in 5 % of lake water samples in Kenya, corroborating previous findings in South America and Europe. However, since environmental data is frequently obtained as a result of using molecular methods for viral detection, there is a lack of information about the infectious status of HAdV found in those samples (35).

Regarding the characterization and distribution of the HAdV types around the world, some viral types of HAdV-A, -C and -F species have been reported as those most frequently found. Table 2 summarizes some of the HAdV types and species reported in waters from randomly selected studies. Data on the presence of HAdV in 64 % of all effluent samples of East Cape (New Zealand) indicated the circulation of HAdV-C and -F in the country (51). In contrast, Quintão et al. (52) detected HAdV in WWTPs in Brazil with a lower prevalence (27 %), demonstrating the presence of HAdV-F41, which was in accordance with previous studies that reported this enteric viral type as the most commonly found in wastewater and surface water worldwide (35). For example, the genome of this viral type has been detected in 92-96 % of the wastewater samples in Italy (10) and Japan (53,54), as well as in 100 % of wastewater samples in Tunisia (55) and Luxembourg (56). On the other hand, this viral type is also prevalent in surface water. It was detected in 94-100 % of river samples from Luxembourg (56), Italy (57) and Japan (53). Furthermore, it represents 12 to 77 % of all HAdVs described in lake samples collected in the USA (54) and was also detected in hot springs in Taiwan (58). More recently, Nour et al. (59) reported that the HAdV-F41 type was present in all investigated water sources (WWTP, irrigation and lakes water) in Saudi Arabia. Nagarajan et al. (60) investigated the HAdV prevalence and distribution in environmental water samples from Taiwan, and again, HAdV-F41 was the most prevalent HAdV type in the river water, along with HAdV-A12 and -A31, although with a lower prevalence.

Table 2. Summary of HAdV species and types reported in some studies worldwide.

Country	Sample	Species	Type	Reference
Spain	WWTP	F	40, 41	(12)
		A	31, 12	
		B	34, 35, 11	
Japan	Lake	F	40, 41	(54)
Spain	Swimming pool	E	4	(32)
New Zealand	Sewage influent, river	C, F	n.d.	(61)
Italy, Spain, Portugal	Marine and freshwater	A	12, 31	(29)
		D	19	
		F	40, 41	
European countries	Marine and freshwater	F	41, 40	(11)
		A	12, 31	
		C	1	
		D	19	
Chad	River, borehole, open wells, underground water	D	n.d.	(62)
Luxembourg	WWTP	A	12, 31	(56)
		B	3	
		C	1, 2, 6	
		F	40, 41	
Worldwide	n.d.	A	31	(38)
		B	3, 7, 11, 14, 21, 55	
		C	1, 2	
		D	8, 37	
		E	4	
		F	40, 41	
Italy	Tiber River	F	41	(57)
Taiwan	Hot springs	F	41	(58)
Italy	Sewage, river and drinking water	F	41	(10)
		A	12	
Italy	Sewage	A	12, 31	(34)
		B	3	
		C	1, 2, 5	
		D	9, 17, 24,26, 37, 38, 42, 44, 48, 70	
		E	4	
		F	40, 41	
Spain	River e groundwater	C	2	(22)
Brazil	Water and sediment	C	n.d.	(63)

		F	n.d.	
Tunisia	WWTP	F	41	(55)
Japan	Raw sewage	F	41	(64)
France	River	F	41	(35)
Australia	WWTP	A	12, 18, 31	(42)
		B	3, 16	
		C	1, 2, 5, 6	
		D	8, 17, 19, 23, 36, 51, 58	
		E	4	
		F	40, 41	
Brazil	WWTP	F	40, 41	(52)
Republic of Kenya	Lake	F	40, 41	(50)
Saudi Arabia	WWTP, irrigation water and lake	F	41	(59)
Taiwan	River and its tributaries	F	41	(60)
		A	12, 31	

"n.d." stands for "non-determined", and "WWTP" stands for "wastewater treatment plants".

1.3. Justification for this dissertation and objectives of this work

Since the levels of FIB cannot predict the presence of viruses in water, and these agents can be problematic to the exposed communities (22), HAdV can be used as a faecal contamination indicator agent because they are highly resistant to water treatment and disinfection processes, are highly prevalent worldwide and are host-specific (29).

Considering that these viruses are enteric and directly excreted into the sewage treatment systems, if the WWTPs are ineffective in removing/inactivating these agents or if there's a sewage leakage, their dissemination into the environment can occur, representing a public health problem. So, in order to prevent possible risks to the population, is essential to understand the presence and distribution of these pathogens. In Portugal, epidemiological data on HAdV appear to be scarce, with inconsistent data reporting their geographical distribution. Consequently, the main purpose of this work regarded the evaluation of the molecular diversity of HAdV in the Lisbon Metropolitan Area (LMA). For that, raw sewage (i.e., influent) and environmental water samples collected between 2018 and 2021 in the LMA were analysed. To assess the presence of HAdV in each sample, hexon-coding sequences of HAdV types found in Europe, particularly those most frequently associated with gastrointestinal infections, were used for primer design. Then three nested touch-down PCR protocols were developed. The amplified products were analysed by a molecular cloning approach where, after a random selection of bacterial clones containing recombinant plasmid DNA with viral inserts, these molecules were sequenced by the Sanger method. In addition, these amplified products were also sequenced using the Illumina method. The obtained sequences were tentatively identified by combining the search for homologous sequences available in the public genomic databases and phylogenetic analysis, allowing us to associate the obtained sequences to a specific HAdV species or, in most cases, to a viral type.

2. Materials and methods

2.1. Sample collection

In this study, eight environmental and nine wastewater samples collected in 14 different locations in the LMA were selected from a previous study from the Food Safety & Microbiology (FS&M) laboratory (65). In what concerns the wastewater samples, the solid particle-free influent of six different WWTP distributed in the LMA was collected in different periods. Depending on the different collection sites, the influent (1 L) was collected either in October 2018, April 2019, July, October, or November 2020. Some of the sites were sampled more than once and in total nine wastewater samples were collected.

The WWTPs were classified from A to F, where WWTP A treats a population equivalent of 756,000 habitants, WWTP B and WWTP C was projected to treat the influent of approximately 210,698 and 30,000 habitants, respectively; WWTP D treats a population equivalent of 253,000 habitants, and WWTP F has the higher treatment capacity, treating the influent of 903,069 habitants.

On the other hand, eight environmental water samples (10 L) were collected in October and November of 2020, six of which were from rivers. Two samples from the margins of the Tagus River, two samples from the margins of the Sado River, one sample from the mouth of the Lizandro River, and one sample from the Trancão River, a tributary of the Tagus River. The other two samples were taken from two different creeks (Ribeira do Carenque and Ribeira das Lajes).

Additionally, seven other environmental samples (13 L) were collected in October of 2021 also within the LMA. These included two samples from creeks in Cascais (Ribeira das Vinhas and Ribeira de Caparide), two samples from Seixal ditches (Fogueteiro and Corroios), and one sample from Almada (Sobreda), Cascais and Moita (Alhos Vedros) ditches. All samples were brought to the laboratory at room temperature, immediately stored at 4 ºC, and processed the day after collection.

All sample collection sites are shown in Table 3 and illustrated in Figure 3. The environmental sampling sites were selected based on their proximity to possible faecal pollution sources like WWTP and sewage discharges.

Table 3. Geographical coordinates and date of all the water samples analysed in this study.

Sample type	Collection site (district)	Coordinates	Date of collection
Environmental	Alhos Vedros ditch (Moita)	38º 39' 12.7" N 9º 1' 57" W	October 2021
	Cascais ditch (Cascais)	38º 42' 55" N 9º 25'39" W	October 2021
	Corroios ditch (Seixal)	38º 38' 26" N 9º 9' 25" W	October 2021
	Fogueteiro ditch (Seixal)	38º 36' 45.8" N 9º 6' 25.7" W	October 2021
	Lizandro River (Mafra)	38º 56' 24" N 9º 24' 47" W	November 2020
	Ribeira das Lajes (Oeiras)	38º 41' 9" N 9º 18' 51" W	October 2020
	Ribeira das Vinhas (Cascais)	38º43' 6" N 9º 25' 51" W	October 2021
	Ribeira de Caparide (Cascais)	38º 41' 42" N 9º 22' 32" W	October 2021
	Ribeira do Carenque (Amadora)	38º 45' 24" N 9º 14' 55" W	November 2020
	Sado River (Industrial Area of Setúbal)	38º 30' 22" N 8º 50' 52" W	November 2020
	Sado River (Port of Setúbal)	38º 31' 14" N 8º 53' 14" W	November 2020
	Sobreda ditch (Almada)	38º 38' 38.6" N 9º 10' 23.9" W	October 2021
	Tagus River (Algés)	38º 41' 41" N 9º 13' 48" W	October 2020
	Tagus River (Barreiro)	38º 39' 55" N 9º 4' 39" W	October 2020
	Trancão River (Loures)	38º 47' 47" N 9º 5' 57" W	October 2020
Wastewater (influent)	WWTP A1	n.a.	October 2018
	WWTP A2		April 2019
	WWTP A3		July 2020
	WWTP B1	n.a.	October 2018
	WWTP B2		April 2019
	WWTP C	n.a.	November 2020
	WWTP D	n.a.	November 2020
	WWTP E	n.a.	November 2020
	WWTP F	n.a.	November 2020

º indicates degrees,' minutes and '' seconds. "n.a." stands for "not applied" (the wastewater samples were collected under a confidentiality agreement).

Figure 3. Geographic localization of the wastewater and environmental sampling sites in the LMA. The total number of samples is indicated between brackets.

2.2. Viral-like particles (VLP) concentration by skimmed milk flocculation

The viral-like particles (VLP) present in all samples were concentrated by skimmed milk flocculation as previously described by Calgua et al. (66). As represented in Figure 4, a pre-flocculated 1 % (w/v) powdered milk solution was prepared using skimmed milk powder (Conda Pronadisa, Spain) dissolved in synthetic seawater (Paragon Scientific Ltd, United Kingdom). Then, approximately 130 mL of this solution was directly added to each environmental water sample to obtain the final concentration of skimmed milk (in the sample) of 0.01 % (w/v). Finally, the pH of the solution was adjusted to 3.5 by adding HCl 1 M. All samples were stirred for 8 h at room temperature to allow the VLP to be adsorbed into the skimmed milk aggregates and then incubated for more 8 h at room temperature without agitation to allow the sedimentation of the flocs. After that, for each sample, the supernatant was carefully removed with a vacuum pump (Masterflex) and the sediment (approximately 470 mL) was transferred to a centrifugation bottle. The sediment was concentrated by centrifugation at 5,500 *g* for 45 min at 12 ºC (J-26 Avanti® XPI Beckman Coulter™ centrifuge). Finally, the supernatant was rejected and the

concentrate, which contained the VLP, was resuspended in 8 ml of phosphate buffer (1:2, v/v of 0.2 M Na_2HPO_4, 0.2 M NaH_2PO_4, pH 7.5), transferred into a 15 mL Falcon tube, and stored at -80 ºC until DNA extraction.

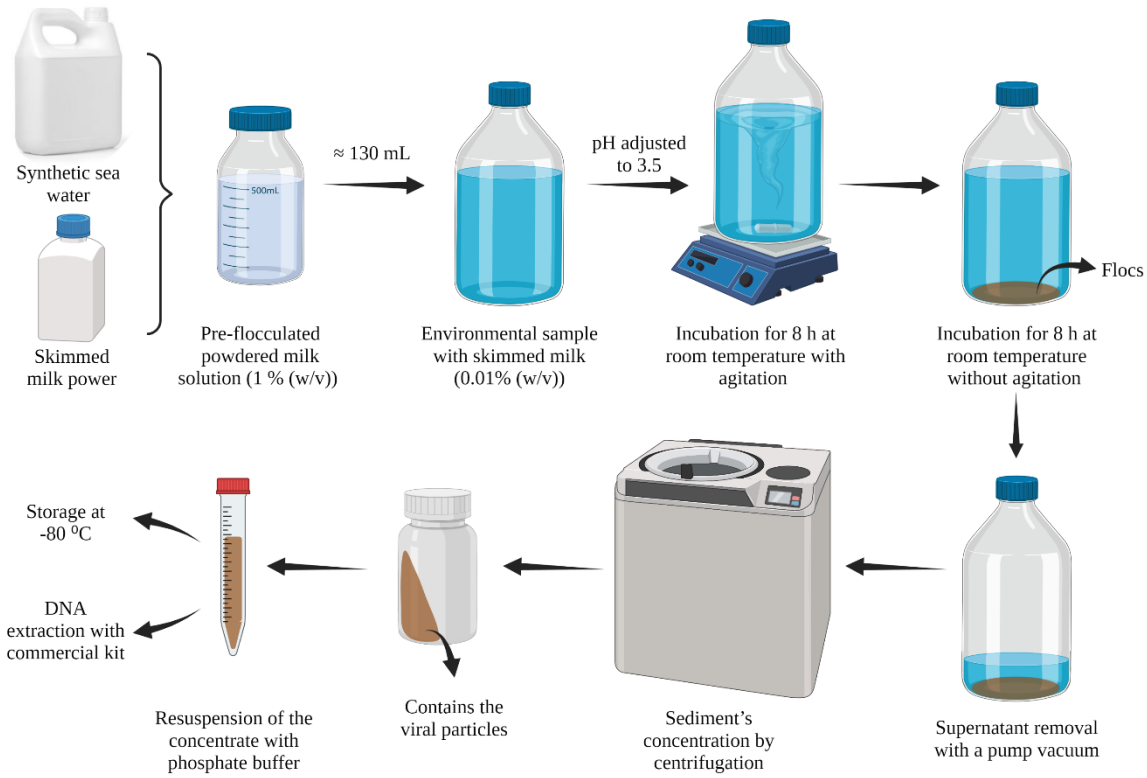

Figure 4. Schematic representation of viral-like particles concentration by skimmed milk flocculation.

2.3. DNA extraction from the VLP concentrates

For DNA extraction, the QIAamp® DNA Mini Kit (Qiagen, Germany) and its suggested protocol were used following the manufacturer's instructions. Briefly, 200 µL of buffer AL was added to 200 µL of each sample, the solution was blended by vortexing for 15 s, followed by a 10 min incubation at 70 ºC. Then, 200 µL of ethanol (98 %) was added, and the solution was homogenized by vortexing and a short centrifugal spin. This mixture (600 µL) was transferred onto the QIAamp Mini spin column, previously placed in a 2 mL collection tube. After centrifugation at 6,000 *g* for 1 min, the flow-through and the collection tube were rejected. The column was placed in a new collection tube and 500 µL of the first washing buffer (AW1) was added. After centrifugation using the same conditions mentioned above, the flow-through and the collection were discarded again.

The extraction column was then placed in another collection tube and 500 µL of the second washing buffer (AW2) was added. This time, the mixture was centrifugated at 20,000 *g* for 3 min. After discarding the flow-through and the collection tube, the column was transferred to a new collection tube and centrifugated at 20,000 *g* for 1 min to eliminate the chance of DNA contamination with buffer AW2. Afterwards, the column was placed in a 1.5 mL microcentrifuge tube, and 200 µL of AE elution buffer was added. After 1 min of incubation at room temperature, total DNA was collected by centrifugation at 6,000 *g* for 1 min.

DNA concentration and purification ratios were determined with Nanodrop® 1,000 (ThermoFisher Scientific, USA) and the extracts were stored at -20 ºC until further analysis.

2.4. Nucleotide sequence dataset construction for primer design

The different datasets used were constructed by compiling HAdV hexon and penton nucleotide coding sequences of the most abundant proteins of the capsid, corresponding to the viral types described (26,41). These sequences were downloaded from the National Center for Biotechnology Information (NCBI) Virus database (available at https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/) and, when possible, six sequences of each type were selected and included in each sequence dataset. The length of the sequence (more than 600 bp) and the geographic region (preferably belonging to Southern Europe and Portuguese-speaking African countries) were used as minimal inclusion criteria. Also, rarely represented types (some of HAdV-D and -G) for which only was available one sequence were not included in these datasets. HAdV-C species was also excluded from these datasets because these viruses are mainly associated with respiratory infections. Since this study focused on the analysis of HAdV in water, the primer design essentially targeted the detection of the viral types mostly associated with gastrointestinal infections.

Viral sequences were aligned using the iterative G-INS-I method implemented in MAFFT (version 7) (available at https://mafft.cbrc.jp/alignment/server/) (67). Then, the sequences were edited with Gblocks (available at http://phylogeny.lirmm.fr/phylo_cgi/)

(68), and, for each gene, a phylogenetic tree was constructed using the Neighbour-Joining (NJ) method and the Mega software (version 6.0).

After the inspection of the obtained trees (hexon and penton-specific), the remaining analysis proceeded only with the HAdV hexon-coding sequences, i.e., only the latter were selected for primer design. Lastly, according to the phylogenetic distance of the species in the hexon tree, some of them were compiled and three secondary datasets (A+F, B+E, D) were constructed with only the sequences from the viral types found in Europe. These datasets were used for primer design (described in section 2.5).

2.5. Primer design

For primer design, the online tool Primer Design-M (available at https://www.hiv.lanl.gov/content/sequence/PRIMER_DESIGN.html) (69) was used considering, as a starting point, the three secondary datasets (A+F, B+E, D), containing only sequences from Europe, as previously mentioned. The primers were designed considering default settings for multiple parameters such as their length (between 20 and 25 bp), detection limit (which represents the inclusion of degenerated primer sites to cover genetic diversity) of 5 %, maximum temperature difference of 5 ºC between the two primers in each pair, dimer window (i.e., the dimer window size to be considered) of 10, and dimer max ratio (i.e., the maximum ratio of matched nucleotides within the dimer window) of 0.9. When primers included degenerate positions, the limit of complexity (i.e., the maximum number of primer species caused by degenerated sites) used was 48, which means that, in this case, the most complex primer has five degenerate positions. Additionally, the melting temperature (Tm) of each primer was calculated using the empirical nearest neighbour model (69). The formation of primer dimers was verified with the Multiple Primer Analyzer tool (ThermoFisher Scientific, USA) (available at https://www.thermo-scientific-web-tools/multiple-primer-analyzer.html). All parameters considered, four pairs of primers were obtained (Table 4), one pair for the 1st round of the nested-PCR, that amplifies part of the coding sequence of the hexon-coding region (originating a fragment with 1615 bp) of all HAdV types included in the datasets, as well as three pairs for the 2nd round of the nested-PCR that were designed to tentatively amplify the hexon region of HAdV-A and -F, -B and -E, and -D species, originating fragments with 1108, 702 and 1339 bp, respectively (Figure 5).

Table 4. List of primers designed and used for the HAdV screening in the collected water samples, corresponding genome location, complexity limit, melting temperature (Tm), and size of the PCR product.

PCR round	Primer name	Primer sequence (5' – 3')	Genome location (start..stop)	Complexity	Tm (mean) [ºC]	PCR product (bp)
1st	FW universal	CTRGCYGTGGGYGAYAACMG	18038..18057	32	63.94	1615
	RV universal	GAYTGRTCRTTGGTRTTCRTT	19687..19668	32	56.50	
2nd	FW A+F	TAYCARCCVGARCCKCAAGT	HAdV-A: 18274..18293 HAdV-F: 18168..18187	48	62.11	1108
	RV A+F	AAGTTCCAYTCRTAVGTGTA	HAdV-A: 19394..19375 HAdV-F: 19309..19290	12	54.75	
	FW B+E	GTRGGCGACAACMGHGTGCT	HAdV-B: 18515..18534 HAdV-E: 18509..18528	12	65.69	702
	RV B+E	AAGCCAATGTARTTGGGTCTGTT	HAdV-B: 19230..19208 HAdV-E: 19182..19160	2	61.51	
	FW D	TTCAAACCCTACTCGGGCAC	HAdV-D: 18126..18145	1	61.83	1339
	RV D	TGATGGCAAAGAACTTTTGGGGC	HAdV-D: 19470..19448	1	63.76	

H (A, C or T), K (G or T), M (A or C), R (A or G), S (C or G), Y (C or T), V (A, C or G). "ºC" indicates degrees Celsius; "FW" stands for "forward", "RV" for "reverse", and "bp" for "base pairs". This work is the reference for all the primers listed.

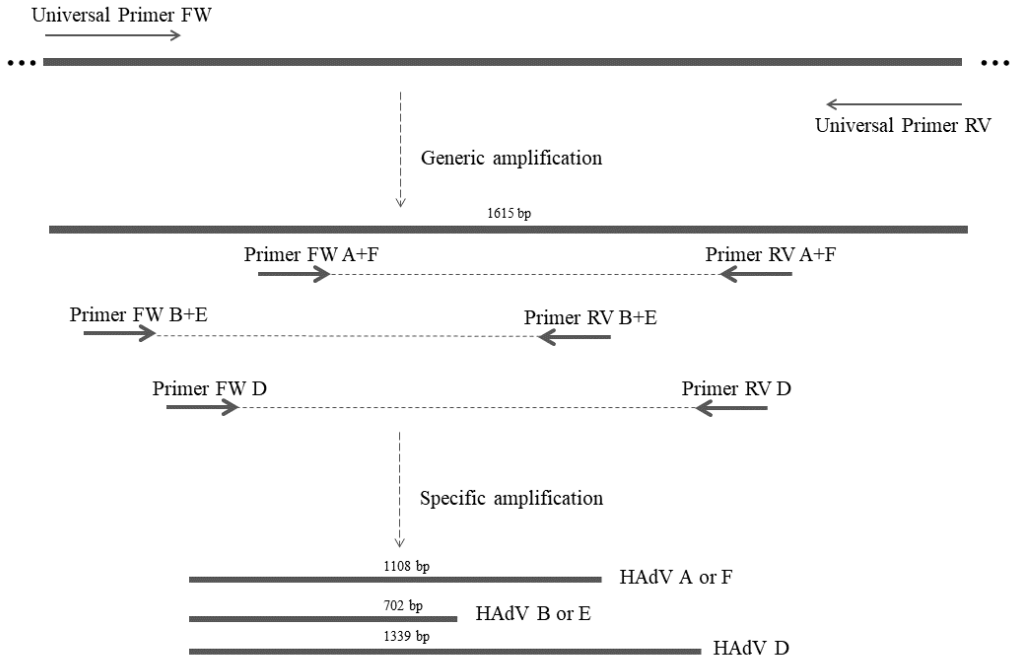

Figure 5. Nested-PCR organization for amplification of HAdV hexon capsid gene.
"bp" stands for "base pairs", "FW" for "forward", and "RV" for "reverse".

2.6. Nested touch-down PCR assays

For the partial amplification of the HAdV hexon gene, three nested touch-down PCR protocols were developed: one for the amplification of HAdV-A and -F species, one for the amplification of -B and -E and another for the amplification of -D (Figure 5 of section 2.5), using the thermal profiles and reaction conditions indicated in Table 5. Both rounds of the nested-PCR included a nucleic acid free control in which H_2O was used instead of the DNA template. All the amplification steps were performed on a T3,000 thermocycler (Biometra®). All the amplification reactions carried out using the DNA extracted from the environmental samples were performed with the Supreme NZYTaq II Green Master Mix (NZYTech, Portugal), while for the amplification reactions with the DNA extracted from the wastewater samples the NZYTaq II 2x Green Master Mix (NZYTech, Portugal) was used.

After the 2nd round of amplification, the PCR products obtained were analysed by gel electrophoresis on a 1 % agarose gel with 1x Tris-Acetate-Ethylenediaminetetraacetic acid (TAE) buffer (40 mM Tris, 20 mM acetic acid and 1 mM EDTA), using GreenSafe Premium (NZYTech, Portugal) as a fluorescent dye, and the NZYDNA ladder III

(NZYTech, Portugal) molecular marker. The electrophoresis was performed on PowerPac™ 300 Electrophoresis Power Supply (Bio-Rad).

Table 5. Nested touch-down PCRs thermal profiles and reaction conditions used.

	1st round	**2nd round**
Total volume (µL)	25	35
NZYTaq Master Mix (µL)	12.5	17.5
Primers (µL)	1 (universal primers)	2 (A+F, B+E or D primers)
DNA template (µL)	3	5
H_2O (µL)	7.5	8.5
PCR conditions	95 ºC for 3 min	95 ºC for 3 min
	10 cycles: 95 ºC 30 s 55 ºC 30 s (-1 ºC per cycle) 72 ºC 1 min 15 s 8 ºC ∞	10 cycles: 95 ºC 30 s 53 ºC (A+F) / 57 ºC (B+E, D) 30 s (-1 ºC per cycle) 72 ºC 1 min 15 s 8 ºC ∞
	30 cycles: 95 ºC 30 s 45 ºC 30 s 72 ºC 1 min 15 s	40 cycles: 95 ºC 30 s 43 ºC (A+F) / 47 ºC (B+E, D) 30 s 72 ºC 1 min 15 s
	72 ºC 7 min 8 ºC ∞	72 ºC 7 min 8 ºC ∞

"ºC" indicates degrees Celsius, "min" minutes, "s" seconds, "µL" microliters, and "∞" stands for undetermined time.

2.7. Purification of PCR products from agarose gels

When the analysis of the agarose gel after electrophoresis revealed a band with a size corresponding to that of the expected amplification product, the band was purified using the Zymoclean™ Gel DNA Recovery Kit (Zymo Research, USA). Briefly, the agarose gel containing the amplified DNA fragment was excised using a scalpel and transferred to a 1.5 mL microcentrifuge tube. Then three volumes of agarose dissolving buffer (ADB) were added. To completely dissolve the agarose, the tube was incubated at 55 ºC for 10 min in a thermo-block (VWR®) and the melted agarose solution was transferred to a Zymo-Spin™ Column. After centrifugation at 10,000 *g* for 1 min, 200 µL of wash buffer was added to clean the DNA that was concentrated in the column from

contaminants. Then the DNA was eluted in 6 µL of elution buffer. The successful purification of each amplicon was verified by analysing 1 µL of the purified PCR product by gel electrophoresis as previously described (section 2.6).

2.8. Molecular cloning of DNA molecules in a plasmid vector

Molecular cloning was carried out using blue-white screening of potential recombinant bacterial colonies starting from an *E. coli lacZΔM15* host. Blue-white screening allows for the identification of recombinant bacteria and depends on the activity of β-galactosidase, an enzyme coded by the *lacZ* gene in *E. coli*, which cleaves lactose into glucose and galactose. When the plasmid vector is taken up by such cells, the product encoded by the *lacZ* gene (α-fragment) in the plasmid vector complements the ω-fragment of β-galactosidase encoded by the deleted version of the host chromosomal *lacZ*, producing the functional enzyme (α-complementation). On another hand, when a recombinant plasmid where an exogenous DNA fragment has been inserted is used to transform *lacZΔM15 E. coli* cells, α-complementation does not occur and a functional β-galactosidase enzyme is not produced (https://www.sigmaaldrich.com/cloning-and-expression/blue-white-screening accessed on 09/03/2022). The cloning vectors include selectable markers (such as antibiotic resistance markers), that allow cells growth on a selective medium. Additionally, 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) is used as a visual indication of the bacterial colonies that have been transformed with non-recombinant vector derivatives (i.e. no insert). This compound is a colourless modified galactose sugar that is hydrolysed by β-galactosidase, resulting in the accumulation of a blue-coloured product only in those colonies. White colonies indicate the loss of the cells' ability to hydrolyse X-Gal due to the insertion of foreign DNA in the vector used (70).

The ligation mixtures were prepared using the pGEM®-T Easy Vector System I Kit (Promega, USA) following the manufacturer's instructions. After a short centrifugal spin of all components and vigorous vortexing of the ligation buffer, 5 µL of ligation buffer, 3 µL of purified PCR product, 1 µL of pGEM®-T Easy vector, and 1 µL of T4 DNA ligase were added to a nuclease-free Eppendorf. This mixture was incubated at 4 ºC for 24 h

since this incubation period resulted in higher numbers of transformants than an equivalent incubation for just 1 h.

Subsequently, 5 µL of each ligation mixture was used to transform 50 µL of NZYStar chemically competent *E. coli* cells (NZYTech, Portugal). After 30 min on ice, the cell suspension was subjected to a heat shock at 42 ºC for 40 s in a water bath, followed by 2 min on ice. Before plating on selective medium, 900 µL of Super Optimal Broth with Catabolite repression (SOC) medium (Sigma-Aldrich®, USA) were added to each cell suspension after heat shock, followed by an incubation period at 37 ºC with agitation at 180 rpm on Innova 44 (New Brunswick Scientific™) orbital shaker for 2 h. After centrifugation at 845 *g* for 1 min, 850 µL were removed and the cells were re-suspended in the remaining volume by gentle pipetting. Afterwards, 100 µL of the mixture was plated on Lysogeny Broth (LB) agar (GriSP, Portugal) plates supplemented with ampicillin (100 µg/mL), X-gal (100 µg/mL), and Isopropyl β-D-1-thiogalactopyranoside (IPTG) (0.5 mM), and were incubated overnight at 37 ºC.

2.9. DNA extraction by boiling protocol for recombinant plasmid screening by PCR

Ten randomly selected white colonies were picked with a sterile toothpick and inoculated in 4 mL of LB medium (GriSP, Portugal) supplemented with ampicillin (100 µg/mL), allowing the preparation of saturated cell suspensions after overnight incubation at 37 ºC with agitation at 180 rpm on Innova 44 (New Brunswick Scientific™) orbital shaker.

From the saturated bacterial culture, 100 µL was used for DNA extraction by the boiling protocol. First, the cells were sedimented by centrifugation at 13,000 *g* for 2 min. After discarding the supernatant, the cells were re-suspended in 100 µL of H_2O, and heated at 96 ºC for 15 min in a thermo-block (VWR®) followed by a short centrifugal spin. Insert screening was carried out by PCR using 2 µL of the supernatant according to the conditions of Table 5. After the amplification, the PCR products obtained were analysed by gel electrophoresis on a 1% agarose gel as described in section 2.6. When their size matched that of the expected fragments, 1.5 mL of the saturated bacterial culture with which they were associated was used for plasmid DNA extraction (see below) and

the remaining culture was stored in 20 % of glycerol. For this, 1,125 µL of the culture was added to 375 µL of glycerol (80 %) in cryogenic vials and stored at -80 ºC.

2.10. Plasmid DNA extraction from the saturated bacterial culture

For plasmid DNA extraction the NZYMiniprep kit (NZYTech, Portugal) was used following the manufacturer's instructions. First, 1.5 mL of a saturated *E. coli* culture was centrifugated for 30 s at 13,000 *g*. After discarding the supernatant, the pellet was re-suspended in 250 µL of A1 buffer by vortexing. Then, 250 µL of A2 buffer was added to each tube and the mixture was gently mixed by inverting, followed by incubation at room temperature for 4 min. Then, 300 µL of A3 buffer was added and gently mixed (by inversion), and centrifugation at 13,000 *g* for 8 min at room temperature was performed. The supernatant was loaded onto the NZYTech column, previously placed in a 2 mL collection tube, and centrifuged for 1 min at 11,000 *g*. After discarding the flow-through, 500 µL of AY buffer was added and then centrifugated at 13,000 *g* for 1 min. Subsequently, the flow-through was discarded and 600 µL of A4 buffer was added. After centrifugation for 1 min at 13,000 *g*, the flow-through was discarded. To dry the silica membrane, additional centrifugation at 13,000 *g* for 2 min was performed. To elute the plasmid DNA, the column was placed into a 1.5 mL Eppendorf tube and 50 µL of AE buffer was added. After an incubation period of 1 min at room temperature, the tube/column ensemble was centrifuged at 13,000 *g* for 1 min. To increase the yield of DNA extraction, the solution containing the eluted DNA was passed through the column once more (centrifugation at 13,000 *g* for 1 min).

Finally, the plasmid DNA was quantified, and purification ratios were determined with Nanodrop® 1,000 (ThermoFisher Scientific, USA) and stored at -20 ºC to be further sequenced by the Sanger method.

2.11. Edit and preliminary analysis of nucleotide sequences obtained by Sanger sequencing

For each sample, a minimum of one and a maximum of six recombinant plasmids (n = 60 in total) were proceeded by STAB VIDA, Lda and sequenced by the Sanger

method using the forward primer T7 specific for the pGEM®-T Easy Vector used (section 2.8).

The obtained sequences were edited using the Trim Ends tool of the Geneious Prime® program (version 2022.1.1) considering the quality of the sequencing. The error probability limit used was 0.001, meaning that the software trimmed the regions with more than a 0.1 % chance of an error per base.

A preliminary taxonomic identification based on the analysis of the cleaned sequences was performed with the Nucleotide Basic Local Alignment Search (BLASTn) feature incorporated in the Geneious Prime® program, and those sequences identified as not corresponding to AdV (n = 19) were excluded from the remaining analysis.

2.12. Next Generation Sequencing (NGS) analysis

2.12.1. DNA fragmentation

The purified DNA products (section 2.7) were quantified using a Qubit® Fluorometer (Invitrogen, USA) with Qubit® dsDNA Broad-Range (BR) Assay Kit following the manufacturer's instructions. First, the Qubit® working solution was prepared by diluting the Qubit® dsDNA BR reagent 1:200 in Qubit® dsDNA BR buffer, ensuring 200 µL for each assay tube (the two required standards and all desired samples). Then, 190 and 199 µL of this solution were loaded into the standards and individual assay tubes, respectively. Afterwards, 10 µL of the standards and 1 µL of each sample were added to the appropriate tubes. After vortexing, the samples were protected from the light and incubated for 2 min at room temperature. Finally, the read of standards and samples was performed by placing first the standards and then the samples on the equipment.

To achieve a successful fragmentation of the DNA in all samples, their concentration was adjusted, considering Qubit® quantification, to be in the 1 – 20 ng/µL range, using 0.1X Tris-EDTA (TE) buffer (1 mM Tris HCl, 0.1 mM EDTA, pH 8.0). The recommended concentration was 10 ng/µL. The total volume for each sample was 100 µL. Afterwards, all DNA samples were fragmented by sonication using the Bioruptor® Plus (Diagenode, Belgium) by six repetitions of five cycles of 30 sec and 90 sec ON and OFF, respectively.

The obtained fragments of 11 randomly selected samples were analysed using the Agilent High Sensitivity (HS) DNA reagent kit (Agilent Technologies, USA), in the Agilent 2100 Bioanalyzer equipment (Agilent Technologies, USA) following the manufacturer's instructions. The first steps involved the set-up of the chip priming station, which included the adjustment of a syringe and the base plate, and the placement of a new HS dsDNA chip on the station. Subsequently, the gel-dye mix was prepared by pipetting 15 µL of the HS dsDNA dye concentrate to the HS dsDNA gel matrix vial. Next, this mix was transferred to the top receptacle of a spin filter and centrifugated at 2,240 *g* for 10 min. The filter was discarded and 9 µL of the gel-dye mix was loaded into the marked well. Then the priming station was closed, and pressure was applied to the system by pressing the plunger of the syringe. After 60 s, the plunger was pulled to the original position and the priming station was open. The remaining marked wells were filled with 9 µL of the gel-dye mix and 5 µL of the HS dsDNA marker was added into each of the 11 sample wells and the ladder well. Finally, in each of the 11 samples and ladder highlighted wells, 1 µL of the sample or ladder, respectively, were loaded. The chip was vortexed on the MS3 vortexer (IKA®) for 60 s at 2,400 rpm. The last step was setting up the Bioanalyzer, inserting the chip and starting the chip run.

After the run, the size of the obtained DNA fragments (between 128 and 193 bp, as expected, Table A4 of appendix A section), and the DNA concentration (Table A4, appendix A section) were verified. According to the obtained values, it was possible to ensure that the concentration of fragmented DNA (for each sample) was greater than 50 ng (being between 147.5 – 419.8 ng, Table A4 of appendix A section), allowing us to follow with the libraries preparation protocol for size selection (section 2.12.2).

2.12.2. Libraries preparation

The DNA libraries were prepared with the NEBNext® Ultra™ II Library Prep Kit for Illumina® (New England Biolabs, USA) following the manufacturer's instructions. All the amplification steps were carried out using a T-personal thermocycler (Biometra®).

First, the NEBNext End Prep mixture was prepared by adding 3 µL of NEBNext Ultra II End Prep Enzyme Mix and 7 µL of NEBNext Ultra II End Prep Reaction Buffer

to 50 µL of fragmented DNA (section 2.12.1). All tubes were placed in the thermocycler at 20 ºC for 30 min, followed by 30 min at 65 ºC.

The ligation of adaptors to the obtained DNA fragments was carried out by adding 30 µL of NEBNext Ultra II Ligation Master Mix, 1 µL of NEBNext Ligation Enhancer and 2.5 µL of NEBNext Adaptor for Illumina directly to the NEBNext End Prep reaction mixture. This ligation mixture was incubated at 20 ºC for 15 min in the thermocycler with the heated lid off. Afterwards, 3 µL of USER Enzyme was added to the ligation mixture and then incubated for 15 min at 37 ºC in the thermocycler with the heated lid set to 47 ºC.

Afterwards, the protocol for size selection was followed. After the adaptor-ligation step, the obtained DNA was purified using 50 µL of resuspended (by vortexing) beads that were added to the ligation. Then the samples were incubated at room temperature for 5 min and then placed on a magnetic stand (Figure 6). When the solution was clear (Figure 6A), the supernatant (that contained the desired DNA) was transferred to a new tube and the beads were rejected. Later, 25 µL of resuspended purification beads was added, the tubes were placed again on a magnetic stand and when the solution was clear (Figure 6A), the supernatant was discarded. The beads (that contained the desired DNA fragments) were washed twice with 200 µL of 80 % ethanol. Finally, the beads were air-dried on the magnetic stand with the tube lids open (Figure 6B) and the DNA was eluted in 17 µL of 0.1X TE.

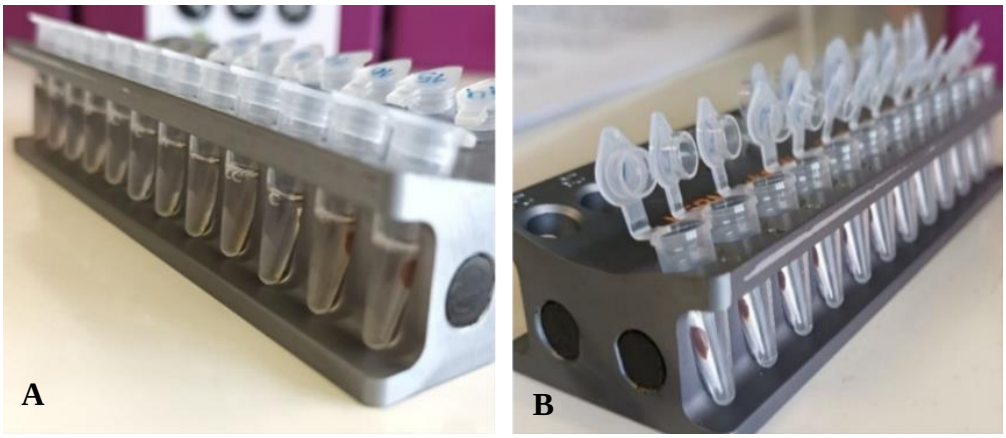

Figure 6. Samples on the magnetic stand for size selection and clean-up of PCR reaction steps from libraries preparation protocol (NEBNext® Ultra™ II Library Prep Kit for Illumina®, New England Biolabs, USA). (**A**) Samples were placed on a magnetic stand until the solution was clear. (**B**) Air drying the beads on the magnetic stand with the Eppendorf lids open.

Subsequently, a PCR enrichment of adaptor-ligated DNA was performed by adding 25 µL of NEBNext Ultra II Q5 Master Mix, 5 µL of universal PCR primer and 5 µL of index primer to 15 µL of eluted DNA. The amplification step was performed with the following cycling conditions: initial denaturation at 98 ºC for 30 s, 30 cycles with denaturation at 98 ºC for 10 s, and annealing/extension at 65 ºC for 75 s, the final extension was at 65 ºC for 5 min. Afterwards, the PCR reaction mixture was cleaned (i.e., enzyme, nucleotides, primers, and buffer components were removed) by adding 45 µL of the resuspended purification beads and incubation at room temperature for 5 min. The tubes were then placed on a magnetic stand to separate the beads from the supernatant. When the solution was clear (Figure 6A), the supernatant was discarded, and the beads (that contained DNA targets) were washed twice with 200 µL of 80 % ethanol. After air drying the beads (Figure 6B), the tubes were removed from the magnetic stand, and the DNA was eluted from the beads with 33 µL of 0.1X TE. Afterwards, the tubes were placed again on a magnetic stand until the solution was clear, after which 30 µL of the eluted DNA was transferred into a new tube and stored at -20 ºC until library normalization. The remaining 3 µL (with the beads) were also stored for library quantification and verification of the amplified DNA size distribution (section 2.12.3).

2.12.3. Libraries quantification

For libraries quantification, the NEBNext® Library Quant Kit for Illumina® was used following the manufacturer's instructions. First, a 1:1,000 dilution was prepared for each library by adding 1 µL of each library to 999 µL nuclease-free H_2O. Then, 1:10,000 and 1:100,000 dilutions were prepared. Afterwards, 16 µL of NEBNext Library Quant Master Mix (with primers) was added to all the wells of the qPCR plate and then, 4 µL of DNA standard, library dilution or H_2O was also distributed. After sealing the qPCR plate and centrifugation at 629 *g* for 2 min, the qPCR assay was performed in QuantStudio™ 5 (Thermo Fischer Scientific, USA). The cycling conditions included an initial denaturation at 95 ºC for 1 min, 35 cycles with denaturation at 95 ºC for 15 s and extension at 63 ºC for 45 s. Afterwards, the quantification cycle (Cq) values were obtained for the standards and the mentioned dilutions (Table A5 of appendix A section).

To confirm the distribution of DNA fragments in the sequencing libraries, the Agilent Bioanalyzer HS DNA chip was used following the manufacturers' instructions as mentioned in section 2.12.1 (Table A5 of appendix A section).

According to the sizes of the libraries from Bioanalyzer and the Cq values obtained by qPCR assay, the concentration of each undiluted library (nM) was calculated using the data analysis template of KAPA Library Quantification Kit (version 5.17) (Illumina® platforms). Afterwards, since the sequencing kit used was MiSeq Reagent Kit V2, a 2 nM libraries pool was created using Pooling Calculator (Illumina) (available at: https://support.illumina.com/help/pooling-calculator/pooling-calculator.html). For that, the information was introduced in the required fields, such as the sample plexity (i.e. no. samples), the unit of measure for the library, the pooled library concentration (2 nM), the total pooled library volume (50 µL) and each undiluted library concentration (nM). Then, the calculator indicated the volume (µL) of each library to use (from the tube with 30 µl of the library – section 2.12.2), and the volume (µL) of 10 mM Tris-HCl to be added (Table A6 of appendix A section). To create the 2 nM pool of 50 µL, a certain volume (µL) of each diluted library (also indicated by the calculator, Table A6 of appendix A section) was combined in a new tube.

2.12.4. Libraries normalization

The libraries were normalized according to the MiSeq System Denature and Dilute Libraries Guide (Illumina, Portugal) following protocol A – standard normalization method.

The first step was the denaturation of the 2 nM library, which involved the combination of 5 µL of the pooled library with 5 µL of 0.2 M NaOH. After a brief homogenization by vortexing, this mixture was incubated at room temperature for 5 min, and then 990 µL of prechilled HT1 buffer was added to obtain 1 mL of 10 pM denatured library.

Next, the bacteriophage PhiX174 DNA (the internal control of the reaction) was diluted to 4 nM by combining 2 µL of 10 nM PhiX174 library and 3 µL of 10 mM Tris-Cl (pH 8.5, 0.1% tween). Then, for the denaturation, 5 µL of the 4 nM library and 5 µL of 0.2 M NaOH were combined and the mixture was centrifugated at 280 *g* for 1 min.

After incubation at room temperature for 5 min, 990 µL of prechilled HT1 buffer was added to the 10 µL of the denatured library mentioned, resulting in 1 mL of 20 pM library. Finally, a dilution of 12.5 pM, which produces an optimal cluster density using MiSeq Reagent Kit v2 reagents, was made by adding 35.5 µL of 20 pM PhiX174 library and 22.5 µL of prechilled HT1 buffer.

The pooled library was spiked with 5% of PhiX174 control through the combination of 30 µL of 12.5 pM denatured PhiX174 with 570 µL of 10 pM denatured library.

2.12.5. Illumina sequencing

The MiSeq Reagent Kit V2 (Illumina, Portugal) was used following the manufacturer's instructions. First, a sample sheet was created with the selection of the appropriate library preparation workflow. The index adapters for each one of the 20 samples were introduced, the indexing scheme (single-index libraries) and the paired-end (i.e., the sequencing run includes read 1 and read 2) options were selected, and in the cycles read field, the number of cycles (151 cycles) was introduced. After the cartridge had been inserted, the flow cell was cleaned and the 600 µL of combined denatured pooled library and 12.5 pM PhiX174 control was loaded into the proper reservoir. Afterwards, the flow cell and the reagent cartridge were loaded into the corresponding compartments and the sequencing run was started. The obtained reads for each sequenced sample are presented in Table A7 of appendix A section.

2.12.6. Clean-up of nucleotide sequences obtained by NGS

The sequences obtained by NGS were edited using the OmicsBox program (version 2.1.14). First, in general tools, the FASTQ Preprocessing tool was used, which includes the removal of adapters (Nextera option) and contamination sequences, and the filtering of short and low-quality reads (Sliding Window Trimming was selected along with the default parameters). Then in metagenomic tools, the Metagenomic Assembly was applied using meta-SPAdes. After obtaining the contigs and scaffolds, the Metagenomic Gene Prediction tool with the Prodigal algorithm (which generates nucleotide and protein sequences) was used. A BLASTn search was then performed, and for each result, the GO

Mapping and GO Annotation features were applied. Finally, for each NGS sequenced sample, the annotated and mapped BLASTn result was exported in a Microsoft Excel-compatible format. Using filter options available in Excel, nodes whose sequence description did not contain "hexon" were excluded, and from these were also excluded the nodes which corresponded to sequences smaller than 350 bp. The contigs obtained to each sequenced sample are presented in Table A7 of appendix A section

All the nucleotide sequences obtained during this study (from Sanger and Illumina sequencing methodologies) that were associated with HAdV were deposited in the GenBank/ENA/EMBL public genomic database under the accession numbers OP605769 – OP605897.

2.13. Phylogenetic analysis

The preliminary taxonomic identification of the Sanger sequences (n = 41) was confirmed by phylogenetic analysis. In the case of NGS sequences, only those for which it was not possible to achieve HAdV type classification by BLASTn (n = 12; MegaBLAST option) were subsequently analysed by phylogenetic reconstruction. For both approaches, the sequences whose BLASTn identity values of the matches with their closest homologues were below 95 % (n = 5 from the Sanger sequencing method, and n = 37 from NGS) were also confirmed by phylogeny. Sanger sequences ranged between 396 and 1023 bp, while NGS sequences' ranged between 352 and 1,416 bp.

For the analysis of the obtained Sanger and NGS sequences, nucleotide datasets were constructed using hexon gene sequences available in the GenBank nucleotide sequence database. Each dataset included only sequences with a minimum length of 350 nt. The multiple alignment and edition of the nucleotide sequences were performed as previously described in section 2.4.

Afterwards, the appropriate evolution model was determined for each dataset using the Model Selection feature of IQ TREE Web Server (available at: http://iqtree.cibiv.univie.ac.at/) (71). All obtained models are shown in Table B1 (appendix B section). From the three available criteria, the Akaike Information Criterion was chosen. The phylogenetic tree was constructed using the Tree Inference feature also from the IQ TREE Web Server, which uses a quick and efficient algorithm for inferring

phylogenetic trees by Maximum Likelihood (ML). The topological stability of the obtained trees was assessed using 1,000 bootstraps and 1,000 iterations of the approximate likelihood-ratio test (aLRT), considering 75% and above as relevant values for both tests.

3. Results and discussion

This study aimed for the detection and genetic characterization of HAdV hexon sequences obtained from a total of 24 water samples. These included nine influent wastewater and 15 environmental samples, collected at 21 different sites spread across the LMA at six-time points: October 2018, April 2019, July, October, November 2020, and October 2021. Firstly, the VLP from the water samples were isolated and the viral DNA was extracted, followed by the design of HAdV-specific primers, and by the optimization of nested-PCR protocols for the amplification of the viral DNA. Finally, the amplification products were purified and used for subsequent genetic analyses which included both molecular cloning and sequencing, and direct-unbiased amplicon sequencing using NGS.

3.1. DNA extraction from the VLP concentrates

Before proceeding with the nested-PCR protocols, the VLP in the water samples were concentrated by skimmed milk flocculation (section 2.2). This inclusion of a VLP concentration step before the DNA extraction was deemed essential in the used experimental algorithm due to the low concentration of viruses in most aquatic samples. As a result, before their detection, a viral concentration step was crucial (43).

Then, the DNA was extracted from the obtained VLP sediments (section 2.3) and its concentration and purity ratios were determined. To measure the DNA purity, the ratio of absorbance at 260 nm and 280 nm was used (T042 – Technical Bulletin Thermo Scientific available at: https://dna.uga.edu/Note-on-the-260_280-and-260_230-Ratios.pdf accessed on 17/08/2022). A ratio of 1.8 is generally accepted as “pure” for DNA. However, if the ratio is considerably lower, it might indicate the presence of contaminants that absorb at or near 280 nm (such as proteins, phenol, or others). Additionally, the ratio 260/230 is used as a secondary measure of nucleic acid purity, and nucleic acid is often considered “pure” if this ratio value range between 2.0 and 2.2. However, if the ratio is significantly lower than expected it can indicate the presence of contaminants, e.g. extraction reagents, which absorb at 230 nm. For the studied samples, 260/280 and 260/230 ratios varied between 1.02 – 2.03 and 0.20 – 1.40, respectively

(Table A1 of appendix A section). The values obtained show that the DNA samples were contaminated with extraction reagents or environmental components that the extraction protocol was unable to remove. However, these apparent lower purity ratios did not interfere with the intended nested-PCR amplification (see below).

3.2. Preliminary phylogenetic analysis for primer design

A preliminary phylogenetic analysis was performed to recreate the evolutionary relationships between the hexon and penton nucleotide sequences of the multiple HAdV species. Two datasets and two phylogenetic trees were constructed (Figure 7) with sequences of HAdV types for each gene as described in section 2.4. The "best" marker (or target) was considered that which gave rise to a tree with the highest topological stability or the one with the highest sequence number in the databases.

The analysis of the obtained trees (Figure 7) suggested that the viral sequences almost segregate with the same general branching pattern, regardless of their genetic origin (penton- or hexon-coding), with the only difference between the two trees concerning the HAdV-C. In fact, whereas the tree based on the penton gene sequences, this species shared common ancestry with HAdV-D and -E species, while the tree based on hexon gene sequences suggested common ancestry with the HAdV-D, -E but also with -B sequences. Beyond providing identical evolutionary relationships, the hexon gene dataset was more diverse, since sequences for most of the viral types were available in public nucleotide databases, while for the penton gene that was not the case. Considering this, the hexon gene dataset was chosen as the "best" target for the primer design. In addition, the topology of the phylogenetic tree obtained for the hexon gene was congruent with the one recently published by Kang et al. (72) with only one minor change in the tree topology regarding the HAdV-E and -D species. In our phylogenetic analysis, they were presented as sister taxa and shared an ancestor with the -B species, while in the analysis performed by Kang et al. (72), HAdV-B and -E species were sister taxa and shared an ancestor with -D. This difference was most probably due to the different datasets used for phylogenetic reconstruction

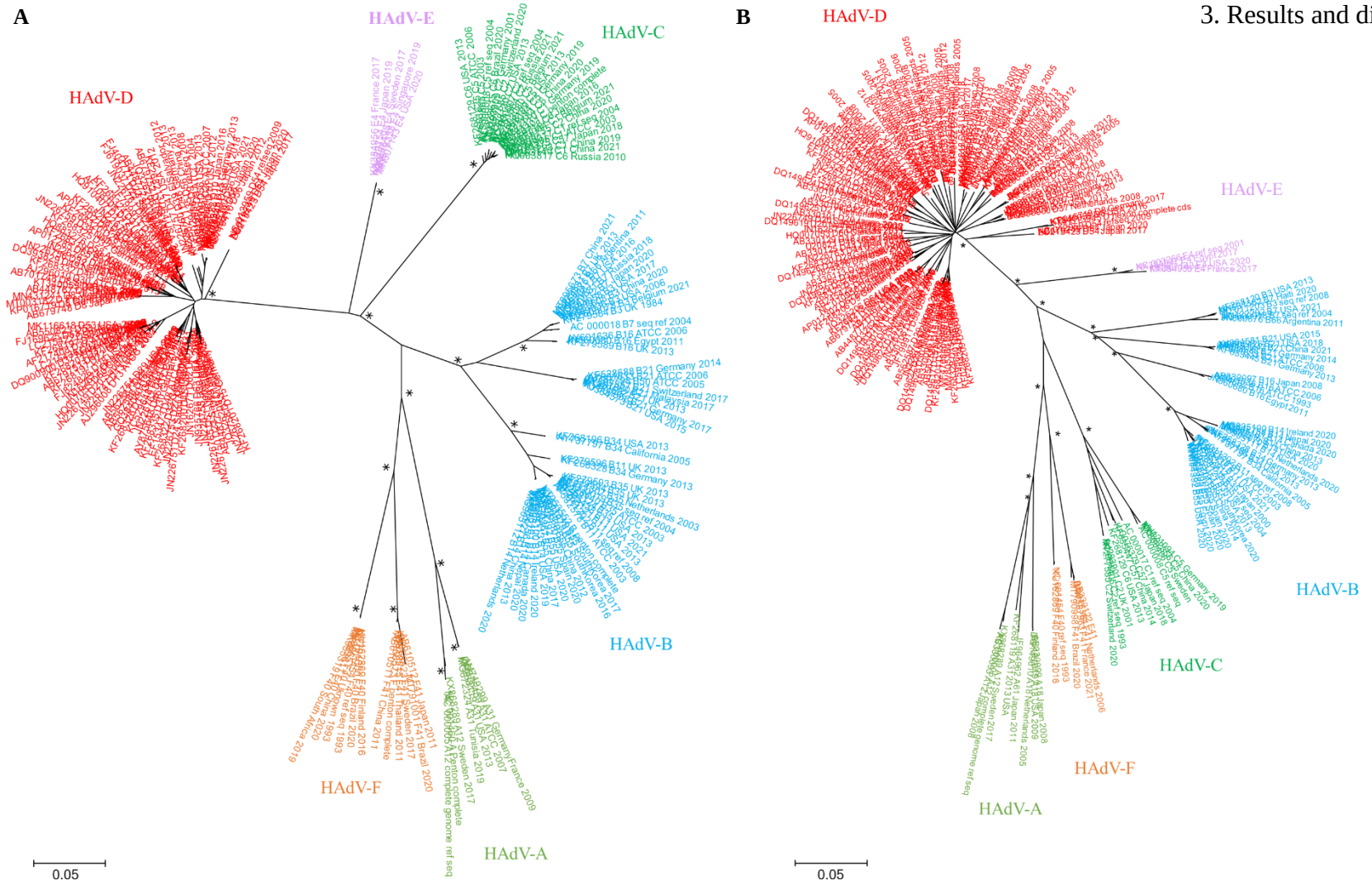

Figure 7. Phylogenetic tree by Neighbour-Joining of HAdV capsid nucleotide sequences. (A) Tree based on the penton gene sequences. **(B)** Tree based on the hexon gene sequences. The "*" indicate bootstrap values that support the suggested topology, considering 75% (of the total number of data resamplings) and above as relevant. Each HAdV species is indicated by a different colour: A – light green, B – blue, C – dark green, D – red, E – purple, and F – orange.

This study aimed for the genetic characterization of HAdV sequences obtained from water samples. In the literature, multiple sets of primers have already been described in order to amplify part of the hexon gene for HAdV screening (32,35,54,55,62,73–80). However, these primers usually produce amplicons with less than 500 bp, which did not suit our aim to obtain amplicons bigger than 500 bp to allow a more robust phylogenetic analysis, as the use of longer sequences usually improves the resolution of the phylogenetic analysis compared to when short-sequence datasets are used (81).

The great diversity of the hexon-coding sequences under analysis compromised the intended primer design when all sequences were considered together. Therefore, some HAdV species were combined according to the phylogenetic distance between the lineages observed in the tree (namely HAdV-A and -F, -B and -E). Species -D were not associated with any other viral lineage due to the high number of types (translated into a high genetic diversity) associated with this species. In other words, three groups were established (A+F, B+E and D) for the primer design. Even so, because of the genetic variability contained in each viral type, we verified that it would not be possible to design primers from the alignments of all viral types. Consequently, the primer design was directed toward the detection of the viral types most frequently found in Europe, and the remaining were excluded. Considering this, one pair of primers was designed for each group (A+F, B+E, D) to allow the differentiation of HAdV species through the size of the amplicon.

3.3. PCR performance assessment

Since we were studying the diversity of HAdV in water samples, where a low viral titer is expected (especially in the environmental ones), a nested-PCR approach was applied to viral detection. This method involves two amplification rounds, where the product of the 1st round is used as template in the 2nd round, where each amplification involves the use of different primers, increasing both the sensitivity and the specificity of the reaction (82). Several authors have reported the usefulness of this approach for HAdV detection in wastewater and environmental water samples (9,58,62,74,75,77–80,83).

In addition, the starting water samples contain DNA from multiple sources (including expected large amounts of non-viral DNA). Even though the used

amplification primers had been designed from alignments of HAdV sequences and were, a priori, virus-specific, the hypothesis of these primers hybridising with other targets, other than HAdV, could not be excluded. Also, the high number of degenerated positions in the designed primers could originate non-specific products. Considering this, a preliminary analysis was performed to determine if the designed primers enable the detection of HAdV DNA. For that, a comparison of the performance of conventional PCR protocol and touch-down PCR was made using the extracted DNA from a WWTP influent (data not shown). A touch-down PCR protocol involves, in an early phase of the thermal profile, the successive decrease of the annealing temperature at each amplification cycle (-1 ºC per cycle), whereas in a later phase, a generic amplification protocol is used (84). While with conventional PCR no amplification products were obtained, the successful amplification with the touch-down PCR approach came as no surprise (data not shown), since it was developed to increase specificity, sensitivity and reaction yield, and is particularly useful for the amplification of complex templates (84).

All things considered, three nested touch-down PCR protocols were established. An amplicon of 1108, 702 and 1339 bp was expected in case of successful amplification of HAdV-A and -F, -B and -E, and -D, respectively. Several cycling conditions and reagent concentrations were tested. For the optimization of all PCR assays, the extracted DNA from a WWTP influent was used (data not shown), as previous studies from the FS&M laboratory had disclosed the presence of HAdVs genomes in this sample. Although no amplicons were observed by electrophoresis after the 1st round of the nested-PCR protocol, probably due to the low viral titter presence in the water samples, we decided to proceed with the 2nd round where an amplicon of the expected size was obtained for each of the 2nd PCR assays (Figure 8A-C).

Additionally, the mentioned DNA extract was used to assess if a multiplex strategy might enhance the reaction's efficiency, allowing the detection of all HAdV species in the same reaction. Multiplex PCR techniques, which must be carefully optimized to maintain/increase the sensitivity of their singleplex equivalents (15), are advantageous for clinical diagnostics because they may quickly identify many viral types in a single reaction. Although this technique was previously successfully optimized to detect enteric viruses simultaneously, such as AdV, EV and HAV (77), in this study the use of multiplex PCR to distinguish HAdV species was not

accomplished. As can be observed in Figure 8D, some viral sequences from HAdV species that have amplified in the singleplex approach (Figure 8A-C) did not amplify in the multiplex approach. Hence, the presence of multiple primers, which may compete for the same matrix, resulted in a loss of sensitivity in the PCR assay.

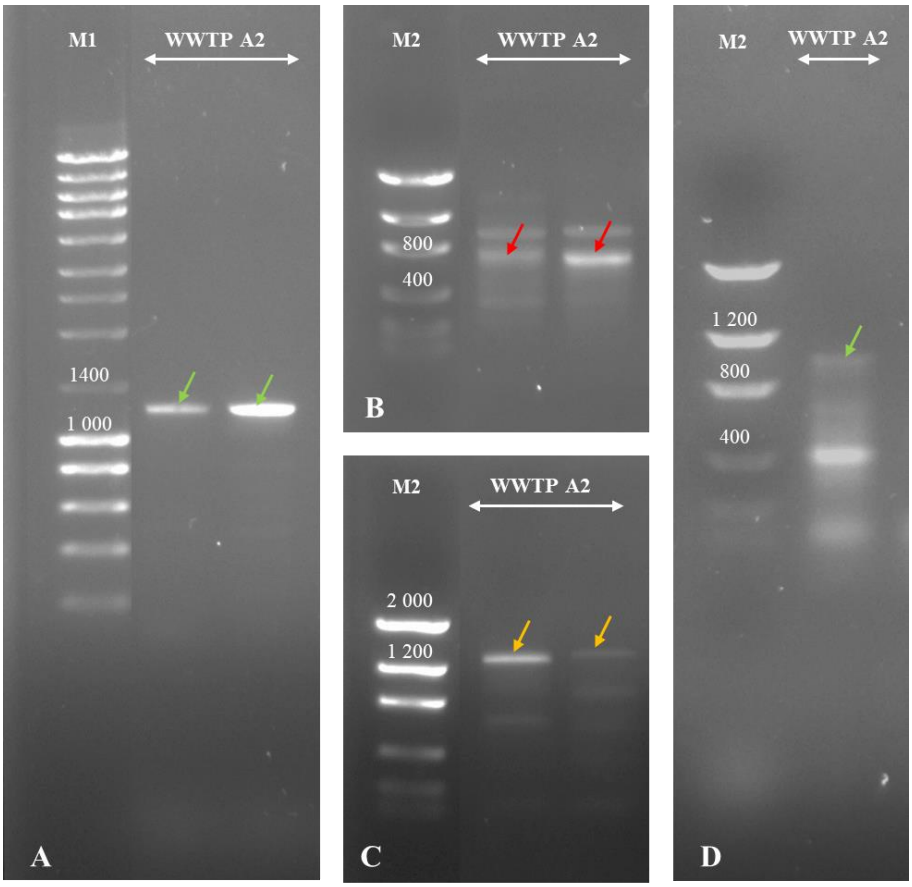

Figure 8. Electrophoretic analysis of the amplification products obtained from the 2nd round of the nested-PCRs using a singleplex and multiplex approach with WWTP A2 DNA extract. **(A)** Amplification products obtained by nested-PCR A+F. **(B)** Amplification products obtained by nested-PCR B+E. **(C)** Amplification products obtained by nested-PCR D. **(D)** Multiplex approach using all primers. The lane marked with "M1" corresponds to the NZYLadder III (NZYTech, Portugal) and "M2" corresponds to the Low Mass DNA ladder (Invitrogen, USA). The coloured arrows indicate the fragment with the expected size, according to the species (A or F – green, B or E – red, D – yellow). "WWTP" stands for "wastewater treatment plant" sample.

3.4. HAdV screening by touch-down nested-PCR assays

The DNA extracts obtained from the water samples were used as the template for the 1st round of the nested-PCR, and the obtained products for the 2nd round. An amplification was considered successful whenever a fragment of the expected size was observed on the agarose gel (Figure 9).

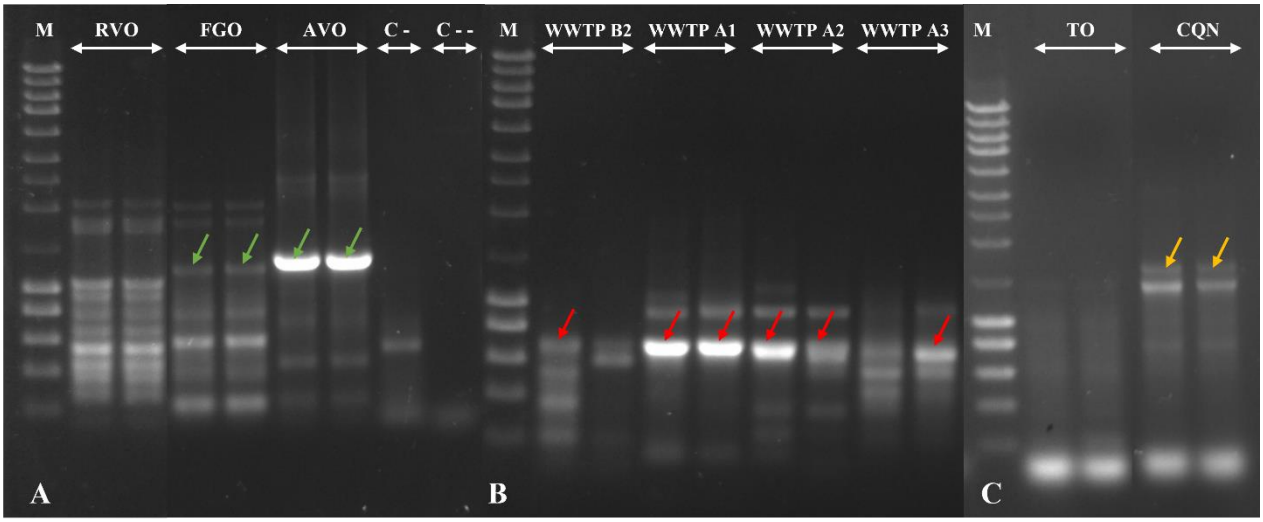

Figure 9. Electrophoretic analysis of the amplification products obtained in the 2nd PCR round for HAdV screening. **(A)** Amplification products obtained by nested-PCR A+F from environmental samples. **(B)** Amplification products obtained by nested-PCR B+E from WWTP samples. **(C)** Amplification products obtained by nested-PCR D from environmental samples. The lane marked with "M" corresponds to the NZYLadder III (NZYTech, Portugal). Lanes marked with "C-" and "C--" correspond to the negative controls. The coloured arrows indicate the fragment with the expected size, according to the species (A or F – green, B or E – red, D – yellow). "RVO" corresponds to the "Ribeira das Vinhas" sample, "FGO" to "Fogueteiro", "AVO" to "Alhos Vedros"; "WWTP" to "wastewater treatment plants samples", "TO" to "Trancão River, and "CQN" to "Ribeira do Carenque".

The amplification reactions were carried out using different Taq DNA polymerases according to each type of sample (environmental versus WWTP). While with the use of NZYTaq II DNA polymerase no amplifications were obtained for the environmental water samples, amplificons were, indeed, obtained with the Supreme NZYTaq II polymerase. This is a robust DNA polymerase created to have high processivity and PCR sensitivity, avoiding the extension of non-specifically annealed primers or primer dimers

and providing a highly specific DNA amplification by having a hot-start-like PCR capability.

Given that a combination of nested and touch-down approaches was used in this work, in an ideal scenario, a unique amplification product in each PCR protocol using the DNA extracts from all samples was expected. In this case, these unique amplification products would be directly used for molecular cloning and NGS steps. However, several non-specific amplicons were observed (as illustrated in Figure 9). Considering this, an additional purification step of the obtained DNA fragments from agarose gels, which separated those with the expected size from most other contaminants, was required (section 2.7) before the obtained amplicons could be either cloned or directly sequenced.

The developments in molecular methods have increased the detection and identification of viruses in different samples. However, the complexity of water samples remains a challenge associated with these matrices (85). Various components can contaminate VLP concentrates and, consequently, negatively interfere with the viral DNA extraction and/or with the performance of the PCR amplification protocol used. Frequently these inhibitors are not removed by the DNA extraction protocols used. Therefore, to reduce this inhibitory effect, the samples should be purified after VLP concentration (85). For example, filtration through a thin layer of fabric or a filter with a width of 0.45 or 0.20 μm is one of the possible options to remove unwanted elements that may interfere with the detection method. On the other hand, free nucleic acids can also be present, and nuclease treatment can be applied to eliminate the presence of non-target unprotected nucleic acids (85). However, in this study, collection, VLP concentration, and DNA extraction steps were only performed for seven environmental samples. The remaining samples used were DNA extracts from another study from the laboratory team (65). So, to avoid discrepancies due to the differential processing of the samples, the same protocols were used for the seven samples mentioned. Considering this, the hypothesis that some inhibitory factors might have interfered with our PCR amplification, and consequently led to a negative/less reliable result for HAdV DNA presence, cannot be neglected. In addition, the PCR amplification can introduce amplification biases since 1) some viral species can be more easily detected when compared to others and 2) the overrepresented nucleic acid molecules can be often enriched and detected by the PCR (52).

3.5. HAdV screening in the collected water samples

As summarized in Table 6, amplicons with the expected size were obtained in 33 out of the 72 PCR amplifications. Considering the 24 water samples in study, 14 (58.3 %), including eight WWTP and six environmental water samples, revealed the presence of HAdV-A and -F species. On the other hand, the presence of HAdV-B or -E genomic sequences were suggested by the observation of an amplicon of the expected size in 12 (50 %) of the samples under analysis, corresponding to three environmental and nine WWTP water samples. Finally, using the PCR protocol designed for HAdV-D, an amplicon with the expected molecular weight was observed in seven (29 %) samples (one environmental and six WWTP).

Table 6. HAdV screening results using three nested and touch-down PCR protocols for the amplification of A or F, B or E and D species of HAdV.

Sample type	Collection site	Expected species		
		A or F	**B or E**	**D**
Environmental	Alhos Vedros ditch	+ [a, c]	–	–
	Cascais ditch	–	–	–
	Corroios ditch	–	+ [a, c]	–
	Fogueteiro ditch	+ [a]	–	–
	Lizandro River	+ [a]	–	–
	Ribeira das Lajes	–	–	–
	Ribeira das Vinhas	–	+ [b, c]	–
	Ribeira de Caparide	+ [a]	+ [b, c]	–
	Ribeira do Carenque	–	–	+ [a, c]
	Sado River (Industrial Area of Setúbal)	+ [a]	–	–
	Sado River (Port of Setúbal)	–	–	–
	Sobreda ditch	–	–	–
	Tagus River (Algés)	–	–	–
	Tagus River (Barreiro)	–	–	–
	Trancão River	+ [a, c]	–	–
Wastewater (influent)	WWTP A1	+ [a, c]	+ [a, c]	+ [a, c]
	WWTP A2	+	+	+
	WWTP A3	+	+	+
	WWTP B1	+	+ [d]	–
	WWTP B2	+ [a, c]	+	–
	WWTP C	+ [a, c]	+ [d]	+ [a]
	WWTP D	–	+ [d]	–
	WWTP E	+ [a, c]	+ [d]	+ [a, c]
	WWTP F	+ [a, c]	+ [a, c]	+ [a, c]

[a] Purified amplicons used in the molecular cloning approach, [b] Purified amplicons not cloned due to the preliminary results of BLASTn (as mentioned in section 3.6), [c] Purified amplicons used in NGS analysis, [d] PCR products directly used in NGS analysis.

The results obtained for WWTP A1, A2, A3, B1 and B2 samples were the same regardless of their collection year and month (Table 6), which corroborates previous studies in which the presence of HAdV throughout the year remains unchanged (86). Based on this, only WWTP A1 and WWTP B2 samples were selected to proceed with the analysis through molecular cloning and NGS. Therefore, from those 33 amplicons described in Table 6, only 25 were used in subsequent analyses: 11, nine, and five amplicons corresponding to PCR A+F, PCR B+E and PCR D protocols, respectively.

As previously mentioned (section 3.4), the obtained amplicons were purified from the agarose gel. A total of 21 out of 25 amplicons were purified to proceed with the cloning method (Table 6). The remaining four out of 25 PCR products were not purified due to their low amount revealed by the observation of a weak intensity in the agarose gels. This can indicate that probably the viral titer in these samples was very low or that the performance of the PCR protocol used was not the same for all the DNA extracts. Also, considering that with the purification step there is an expected loss in the extraction yield, these four PCR products were selected to be directly analysed by NGS (superscript "d" in Table 6). As a result, for NGS analysis, in addition to these four PCR products, other 16 out of 21, were also chosen to be purified and then sequenced.

The HAdV genome is very stable and occurs in higher frequency and abundance in water compared to other enteric viruses (1,11), so, the detection of HAdVs in some water samples was expected. In this study, a higher amplification rate was obtained for the WWTP influent samples when compared with the environmental samples used, corroborating previous studies where the presence of HAdV in untreated sewage was frequently disclosed (1,4,9,10,16,37,49,53,64,77,87–89).

The environmental sampling points were selected based on their proximity to possible faecal sources. However, the difference in the amplification efficiency regarding both types of water samples was expectable due to an expected significant decrease of the virus titer in the environmental samples, and the dilution effect. Therefore, the absence of HAdV DNA amplification in some environmental samples can be due to the absence of viral HAdV particles in those samples or due to the low viral titer which could directly influence the amplification rate, if the sensitivity of the PCRs protocols is low. Recently, Monteiro et al. (86) showed that rainfall incidence affects HAdV levels. Considering this,

and even though it is expected the presence of AdV throughout the whole year (13,86), all the environmental samples were collected during the autumn season, and the hypothesis that the rain could watered down the existing viruses, resulting in a decrease of HAdV titter, cannot be ignored. Furthermore, unlike the environmental samples where HAdV DNA was amplified (such as Alhos Vedros, Corroios and Ribeira do Carenque), corresponding to water samples collected from ditches/creeks, some of them were collected from river estuaries (Tejo and Sado basins), where the diluting impact of a large body of water under the influence of tides cannot be neglected.

3.6. Preliminary genetic analysis of AdV nucleotide sequences obtained by Sanger sequencing

A total of 21 gel-purified DNA fragments were cloned into the pGEM®-T Easy vector (section 2.8). Considering transformants with an HAdV-A and -F insert, an average of 98/290 white/blue colonies were obtained, while 310/954 and 161/433 were obtained for -B and -E, and -D inserts, respectively. Of these, ten white colonies were randomly selected to confirm the presence of the cloned fragment in the vector using crude DNA extracts obtained after cell boiling. A total of 116 recombinant plasmid DNA extracts were identified as carrying inserts of the expected sizes. These recombinant plasmid-DNA molecules were extracted using a column, and the purity ratios and DNA concentration were determined. Several contaminants from the extraction method can compromise the quality of sequencing results (90). For example, EDTA, often present in extraction buffers, is a commonly recognized inhibitor of the Sanger sequencing reaction, as well as salts and alcohols, which are used in the wash and precipitation of the DNA, can be present and negatively affect the Sanger sequencing quality (https://dnacore.missouri.edu/commoncontaminants.html accessed on 30/08/2022). As a result, considering their purity ratios, 60 out of the 116 plasmid molecules were chosen to proceed to Sanger sequencing. Table A2 of the appendix A section presents the results for the plasmid DNAs chosen. In sum, DNA concentration ranged between 10.31 – 541.82 ng/µL, 260/280 and 260/230 ratios varied between 1.75 – 2.17 and 0.62 – 2.35, respectively.

After Sanger sequencing, the quality of the obtained sequences was evaluated considering their respective chromatograms, and the obtained sequences were subsequently edited by removing the amplification primers and low-quality sections (at both ends). Then, the sequences were analysed using the BLASTn tool and later confirmed by phylogenetic reconstruction. Considering that the BLASTn tool finds regions of similarity between nucleotide sequences and determines the statistical significance of such findings, this aimed at a preliminary taxonomical identification of their origin by searching homologous sequences in the public genomic databases. Table 7 summarizes the obtained results.

Out of a total of 28 recombinant plasmids expected to have an HAdV-A or -F specific insert, 11 insert sequences (from WWTP C, Fogueteiro, Sado River – Industrial Area of Setúbal, Lizandro River, Ribeira de Caparide and Tranção River samples) showed either similarity with non-AdV sequences or their origin could not be suggested (as they were not highly similar to any sequence in the database). Later was verified that these amplicons were revealed as a smear on the agarose gel instead of a well-defined band. Unexpectedly, one sequence (from the WWTP B2 sample) was identified as HAdV-D. From the remaining sequences, 11 showed similarity with HAdV-A or -F and five were recognized only as HAdV.

Considering the aforementioned results, the PCR B+E amplification products that also presented a smear on the agarose gel were not cloned, namely, two out of five amplicons (Ribeira das Vinhas and Ribeira de Caparide samples; referred to as superscript "b" in Table 6), since the same result was expected (non-AdV sequences/no similarity). Therefore, only 11 recombinant plasmids expected to have an HAdV-B or -E specific insert were analysed. The preliminary taxonomic identification of the inserts showed similarity with HAdV-F and -D.

Finally, 21 recombinant plasmids expected to have a HAdV-D specific insert were sequenced. Of these, eight inserts (from the WWTP C sample) showed similarity with non-AdV sequences; and again, unexpectedly, six of them were identified as HAdV-C and only two as -D.

Table 7. Preliminary taxonomic identification by BLAST tool for the recombinant plasmids sequenced by the Sanger method.

Expected HAdV species	Sample	No. of recombinant plasmids analysed	BLASTn result	QC (%)	Id (%)
A or F	WWTP A1	7	HAdV-F41	100 %	98 – 100 %
	WWTP B2	4	HAdV-F41	100 %	98 – 100 %
			HAdV-A12	100 %	95 %
			HAdV-D	100 %	100 %
	WWTP C	3	No significant similarity found	n.a.	
	WWTP E	1	HAdV-A31	100 %	100 %
	WWTP F	2	HAdV	100 %	100 %
	Alhos Vedros	3	HAdV	100 %	100 %
	Fogueteiro	2	*Timema californicum*	4 – 7 %	95 %
	Lizandro River	2	No significant similarity found	n.a.	
	Ribeira de Caparide	1			
	Sado River (Industrial Area of Setúbal)	2			
	Trancão River	1	*Candidatus Nomurabacteria*	72 %	10 %
B or E	WWTP A1	6	HAdV-F41	100 %	100 %
	WWTP F	3	HAdV-F41	100 %	99 – 100 %
			HAdV-D32	100 %	100 %
	Corroios	2	HAdV-D	22 %	88 %
			HAdV	19 %	87 %
D	WWTP A1	3	HAdV-C5	100 %	100 %
	WWTP C	8	Uncultured bacterium plasmid	100 %	97 – 100 %
			Lathyrus sativus microsatellite	52 – 58 %	100 %
			Aeromonas veronii plasmid	100 %	99 %
			Shewanella putrefaciens strain	100 %	93 – 99 %
	WWTP E	2	HAdV-C2	100 %	100 %
	WWTP F	5	HAdV-D	100 %	93 – 96 %
			HAdV-C2	100 %	100 %
			HAdV-D25	100 %	96 %
	Ribeira do Carenque	3	Simian adenovirus 41	80 %	68 %
			Simian adenovirus	77 %	68 %
			MAdV 2	39 %	76 %

"QC" stands for "query cover", "Id" for "identity", "n.a." for "not applied", "HAdV" for "human mastadenovirus" and "MAdV" for "murine mastadenovirus".

These results suggest that the primers used for the amplification of a hexon fragment from the five HAdV species under analysis are not specific, since viral

sequences from heterologous HAdV species were also amplified/identified. Even though in the silico phase of primer design their specificity was tentatively ensured, the amplification was influenced by other parameters not evaluated in silico.

Considering this analysis, the presence of HAdV DNA was confirmed and, in addition, these results suggest that the HAdV-F41 type is apparently the most prevalent in the analysed samples. This might be a consequence of the PCR's easiness to amplify this type when compared to others, as previously mentioned.

However, some of the sequence homology search results were characterized by low values of query cover (QC) and identity (Id). The identity is a helpful indicator of the evolutionary distance, and identity values below 95 %, or greater than 95% but corresponding to low query covers, do not represent a meaningful result using MegaBLAST. Thus, all obtained sequences (from molecular cloning and NGS) whose origin could not be defined based only on BLASTn results were tentatively characterized by phylogenetic analysis.

3.7. Genetic characterization of AdV, as defined by molecular cloning followed by Sanger sequencing

After preliminary taxonomic identification of each one of the 60 nucleotide sequences obtained during this work using BLASTn, molecular characterization of the AdVs circulating in the LMA was also performed by phylogenetic reconstruction. The sequences of the inserts that showed similarity with non-AdV sequences or did not show any similarity with any sequence of the databases were excluded (n = 19), and only the remaining sequences (n = 41) were used. Hexon gene-specific datasets were constructed to create phylogenetic trees using an ML approach to confirm the previous results, and also to assign a HAdV type to the sequences that BLASTn only indicated the viral species.

Five sequences (out of 41), from Ribeira do Carenque and Corroios samples were analysed separately since it was not possible to assign a specific origin by BLASTn (query cover and identity values were not significant). We expected that these sequences were probably associated with a non-human AdV origin. Indeed, their unusual identity stood out from a preliminary phylogenetic analysis including all 41 sequences where the five sequences clustered together as a monophyletic out-group (data not shown). Therefore, a

phylogenetic tree (Figure B1 of appendix B section) was constructed with sequences from the hexon gene of non-human AdV. According to the topological stability of the phylogenetic tree mentioned (Figure B1 of appendix B section), these sequences were found to be closely related to Murine mastadenovirus (MAdV) 2. Furthermore, the five sequences clustered according to their geographical origins, suggesting geographic heterogeneity between them. Rodents are the most diverse family of mammals (91) and are distributed worldwide. The presence of AdV genomes in these mammals has been previously reported in other studies (91–94). In particular, two AdV types have already been isolated from the house mice (*Mus musculus*): MAdV-1 and MAdV-2, with the latter being responsible for the infection of cells from the gastrointestinal tract (92). In addition, some serological studies have suggested that AdV infections in rodents may be frequent (91). The identification of the MAdV viral genome was only supported by the phylogenetic tree obtained (in which was possible to observe a monophyletic group with relevant bootstrap/aLRT values) and not by BLASTn. However, its possible detection in Ribeira do Carenque and Corroios samples was not surprising, as these animals can also contaminate the water with their excreta.

Moreover, in the dataset constructed for the preliminary phylogenetic analysis with all 41 sequences (data not shown), three of them (from Alhos Vedros and WWTP F samples) were very short when compared to the others. For this reason, these sequences were analysed individually (Figure B2 and Figure B3 of appendix B section). This analysis showed that the sequence obtained from the Alhos Vedros sample clustered along with the HAdV-F41 viral references, while the remaining two (amplified from the WWTP F sample) clustered within the HAdV-D species.

A phylogenetic tree was constructed with the remaining 33 out of 41 sequences (Figure 10) to examine the distribution of these sequences within the HAdV species. As it can be observed, most sequences clustered together within the HAdV-F species and none of them clustered with -B or -E species.

This study aimed the evaluation of the molecular diversity of HAdV types in circulation in LMA. Thus, specific phylogenetic trees for each species (HAdV-A and -F, -C, and -D) were constructed to assign one type to each sequence (Figure 11, Figure 12, and Figure 13 for -A and -F, -C, and -D species, respectively).

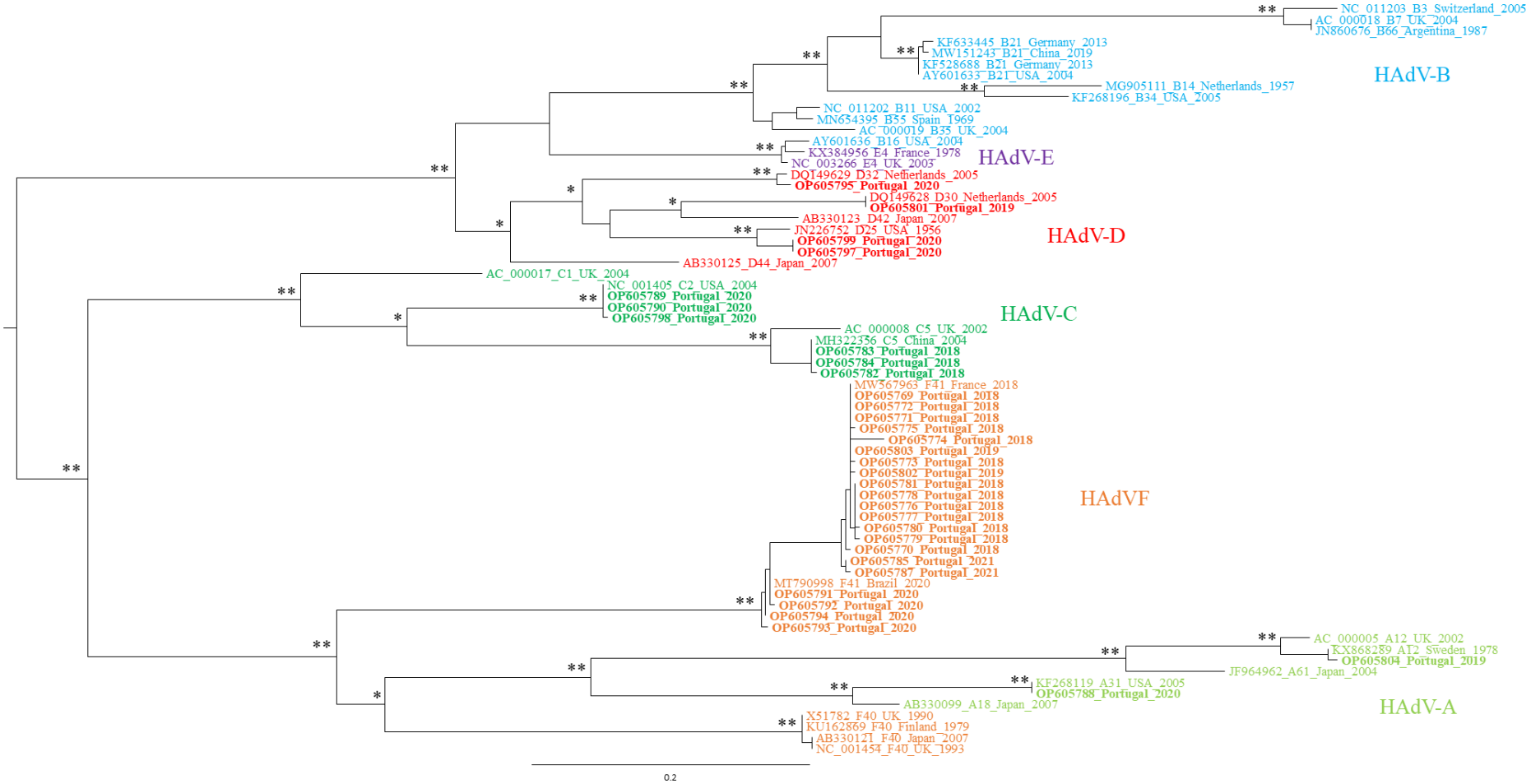

Figure 10. Phylogenetic tree by Maximum Likelihood of HAdV hexon capsid gene sequences, including 33 Portuguese sequences obtained by the Sanger method. Each sequence is identified by its accession number, HAdV type, country of origin and year. The number of "*" indicates the number of methods that support the demonstrated topology, considering 75% (of the total number of data resamplings) and above as relevant for aLRT and bootstrap values. Each HAdV species is indicated in a different colour: A – light green, B – blue, C – dark green, D – red, E – purple, and F – orange.

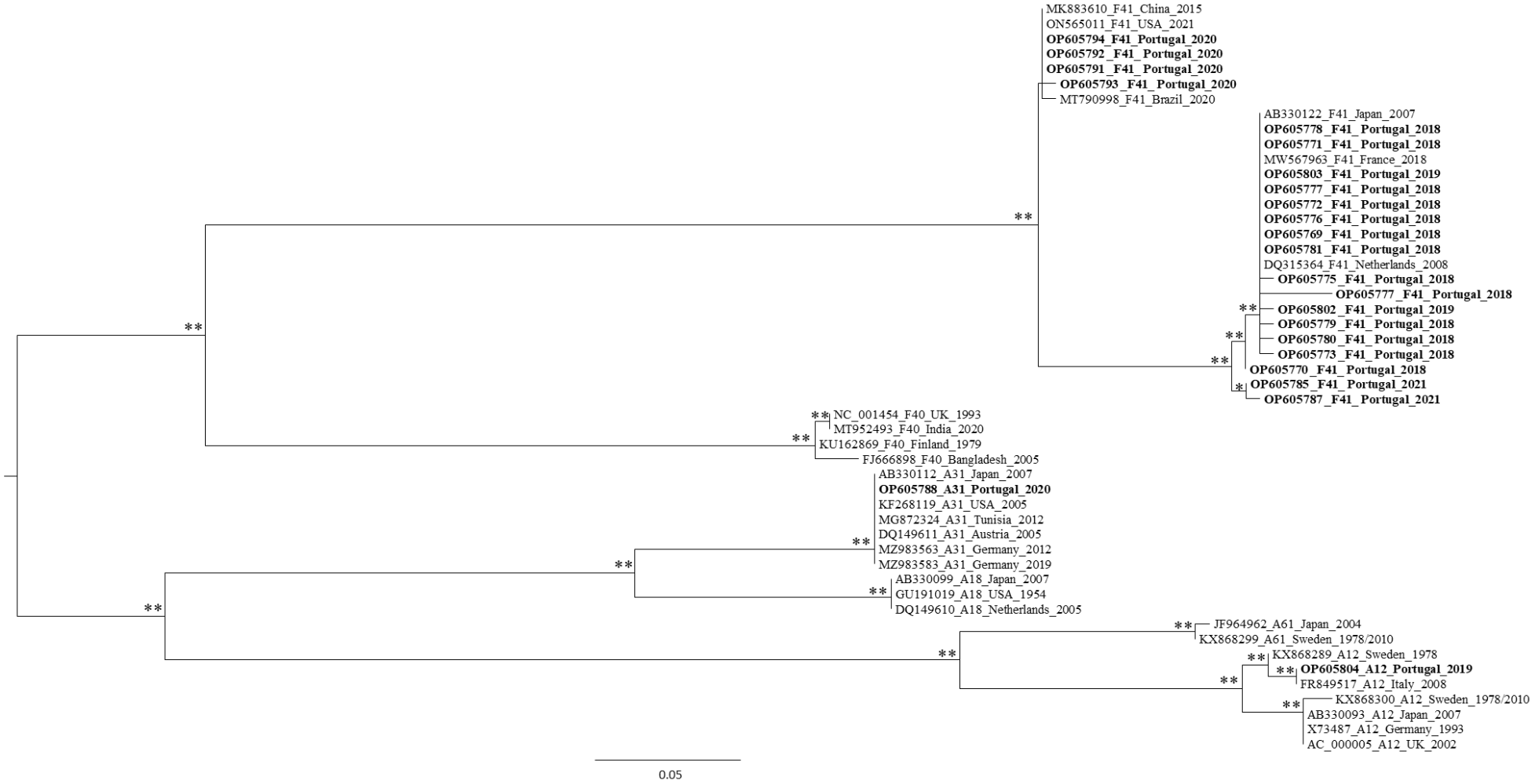

Figure 11. Phylogenetic tree by Maximum Likelihood of HAdV-A and -F hexon capsid gene sequences, including 23 Portuguese sequences obtained by the Sanger method. Each sequence is identified by its accession number, HAdV type, country of origin and year. The number of "*" indicates the number of methods that support the demonstrated topology, considering 75% (of the total number of data resamplings) and above as relevant for aLRT and bootstrap values.

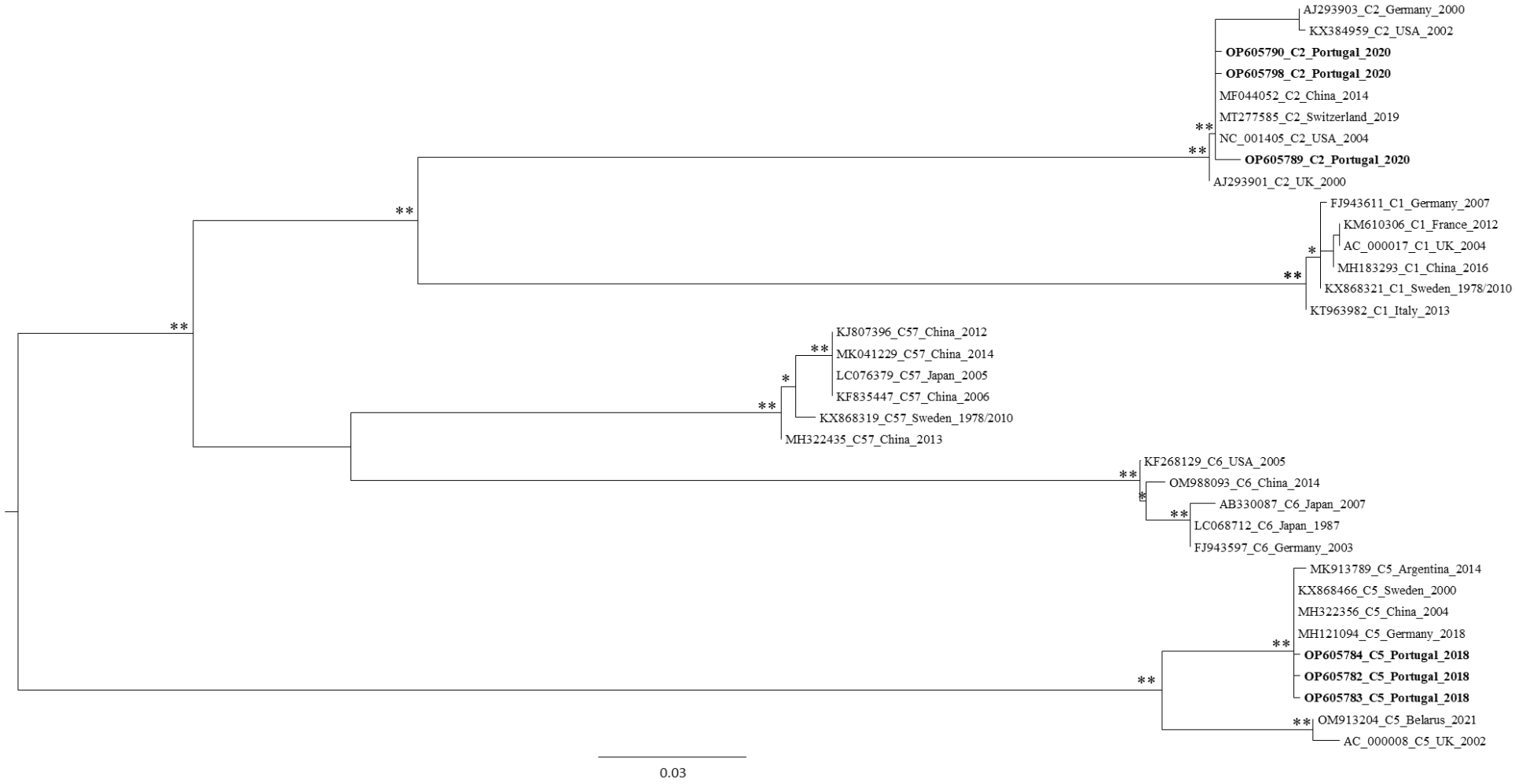

Figure 12. Phylogenetic tree by Maximum Likelihood of HAdV-C hexon capsid gene sequences, including six Portuguese sequences obtained by the Sanger method. Each sequence is identified by its accession number, HAdV type, country of origin and year. The number of "*" indicates the number of methods that support the demonstrated topology, considering 75% (of the total number of data resamplings) and above as relevant for aLRT and bootstrap values.

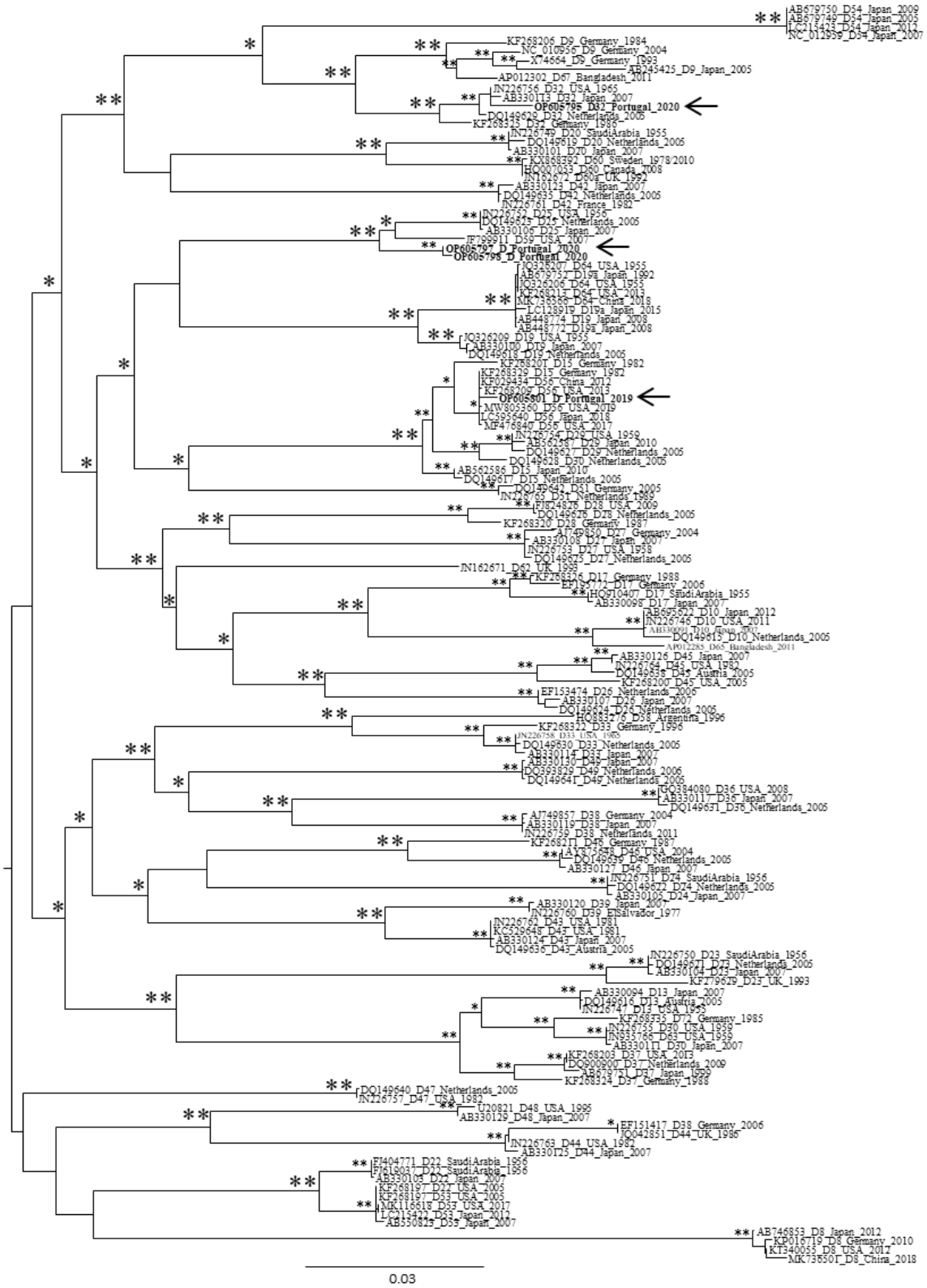

Figure 13. Phylogenetic tree by Maximum Likelihood of HAdV-D hexon capsid gene sequences, including four Portuguese sequences obtained by the Sanger method. Each sequence is identified by its accession number, HAdV type, country of origin and year. The number of "*" indicates the number of methods that support the demonstrated topology, considering 75% (of the total number of data resamplings) and above as relevant for aLRT and bootstrap values. The arrows indicate the Portuguese sequences.

The preliminary taxonomic identification by BLASTn suggested that 11 sequences that were assumed as corresponding to an HAdV-specific species, were similar to other species instead (Table 7). This was confirmed by phylogenetic reconstruction, e.g. one sequence from WWTP B2 amplified using PCR A+F primers clustered among HAdV-D; sequences from WWTP A1, WWTP F and Corroios samples amplified with PCR B+E clustered with HAdV-F41, -D32 or MAdV-2, and sequences from WWTP A1, WWTP E, WWTP F, and Ribeira do Carenque amplified by PCR D clustered with HAdV-C5, -C2 or MAdV-2 species. As previously mentioned, this corroborates that the primers were not absolutely specific for the species for which they were supposed to target, since heterologous HAdV species were also amplified. One possible explanation for this might be the high genomic similarity between the HAdV types, leading to the amplification of other sequences if the species for which the primer was designed was absent in the sample.

The assignment of one viral type to each sequence was performed for all sequences, revealing the genetic variability of the Portuguese sequences, except for some of the HAdV-D species. According to Amdiouni et al. (88) and Casas et al. (95), the types of this species present a high homology between them, making it impossible to distinguish them unambiguously. Figure 14 summarizes the distribution of AdV types assigned to the 41 Portuguese sequences obtained from molecular cloning followed by Sanger sequencing. When observing the distribution of each viral type, HAdV-F41 accounted for 54 % of the total sequences from all samples.

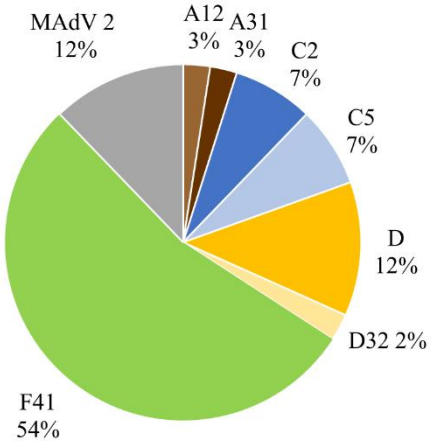

Figure 14. Graphical distribution of AdV types assigned to the 41 Portuguese sequences obtained from molecular cloning followed by Sanger sequencing. Each colour represents one viral type/species: A12 – light brown, A31 – dark brown, C2 – dark blue, C5 – light blue, D – yellow, D32 – light yellow, F41 – green, murine mastadenovirus (MAdV) 2 – grey.

Over time, several studies have reported the detection of HAdV in different water matrices, such as rivers, coastal water, and beaches, as well as in wastewater samples collected in several geographical locations around the world, and HAdV-F41 is the one most commonly detected (10–12,29,34,35,38,42,50,54–60,64). The obtained results highlight the high prevalence of the HAdV-F41 type, in accordance with the previous studies. HAdV-A12, -A31, -C2 and -C5 types were also detected in many studies worldwide (11,12,22,29,34,38,42,56,60) although some authors mentioned that these types were found with a lower prevalence than -F41, as demonstrated in this study. Although all HAdV types (enteric and non-enteric) can be excreted in high concentrations (34–36), the identification of some types was surprising given that not all of them were targeted in nested-PCR protocols, in particular HAdV-C species which were, a priori, excluded.

Figure 15 represents the distribution of the obtained AdV types according to the type of sample. The greatest diversity of HAdV types was observed in WWTP samples.

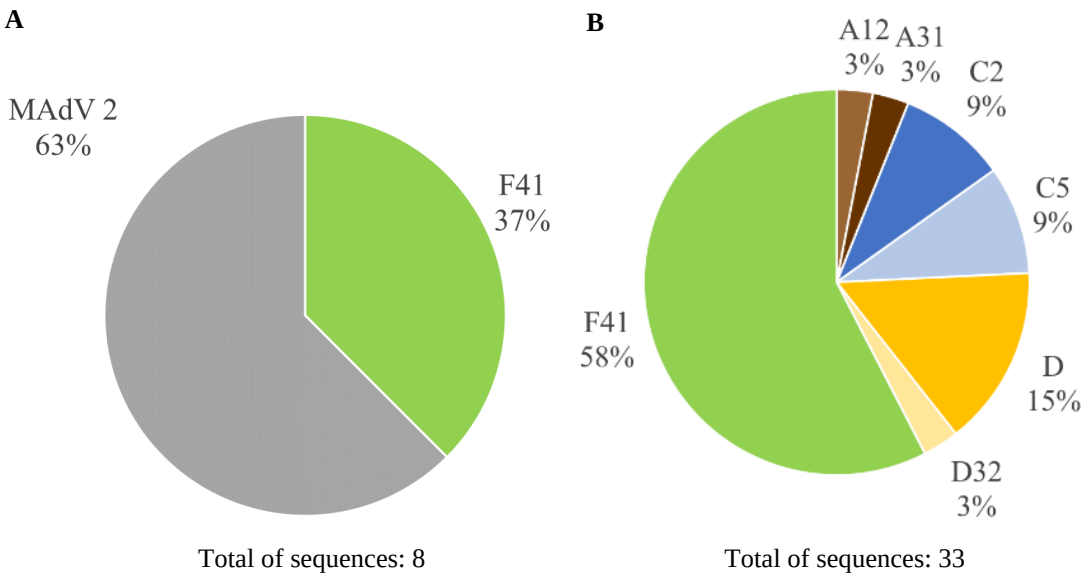

Figure 15. Graphical distribution of AdV types assigned to the Portuguese sequences obtained from recombinant plasmids and sequenced by the Sanger method, according to the water sample. **(A)** Detected types in environmental samples. **(B)** Detected types in WWTP samples. Each colour represents one viral type/species: A12 – light brown, A31 – dark brown, C2 – dark blue, C5 – light blue, D – yellow, D32 – light yellow, F41 – green, murine mastadenovirus (MAdV) 2 – grey. "WWTP" stands for "wastewater treatment plant".

HAdVs are directly excreted into sewage systems, so finding greater diversity of types when assessing raw sewage samples from WWTP was expectable. On the other hand, faecal contamination of environmental waters refers to point sources, such as the discharge of industrial/municipal/domestic wastewater as well as non-point sources including storm/urban/agricultural water runoffs (96). So, the lower diversity of detected genomes in the environment may suggest that in the case of sewage leakage, the majority of HAdV were already eliminated by WWTP treatments, and only remain the ones with a higher propensity to persist in the aquatic environment. HAdVs are more resistant to chemical and physical agents than other viruses and FIB, particularly -F species (13). For example, the predominance of HAdV-F41 in the environment might be explained by the greater ability to persist in natural environments when compared to other types, as suggested by Fong et al. (87). Hence, Rafie et al. (24) demonstrated that this virus possesses a capsid whose structure is not modified by stomach pH, and which virion surface has substantial changes when compared to non-enteric HAdVs. As a result, HAdV-F41 seems to be adapted to the specific conditions in the gastrointestinal tract, perhaps also resulting in higher resistance in the environment.

Traditional cloning and sequencing methods can be less sensitive than other methods and their utilisation may inaccurately reflect the circulation of the various HAdVs in the environment. The fact that only a small number of recombinant molecules were tested may provide an imprecise distribution of the circulating types in comparison to what may exist. Also, this approach might underestimate the prevalence of some less frequent types since the most prevalent can be favourably amplified by the primers. So, the hypothesis that the molecular cloning step may contribute to a biased result cannot be rejected. Thus, to corroborate our findings and confirm if some bias was introduced, an NGS analysis was also performed.

3.8. Genetic characterization of AdV, as defined by the analysis of AdV amplicons using an NGS (Illumina) approach

To ensure that an accurately genetic characterization of the diversity of HAdV types present in LMA waters was achieved, an NGS approach was also applied. Additionally, the utilisation of molecular cloning followed by Sanger sequencing and NGS approaches allow the evaluation of the limitations of both methods.

Seven, nine and four amplicons resulting from the PCR A+F, B+E and D protocols (Table 6), corresponding to 14 WWTP and six environmental samples, were used to construct a pooled DNA library (section 2.12.3). After the sequencing run, 127 contigs corresponding to the AdV hexon gene fragments with more than 350 bp were obtained. In this case, the same strategy (BLASTn search and phylogenetic reconstruction) was used.

First, the sequences for which it was not possible to achieve type classification were divided into datasets according to the species classification (HAdV-A, -F, -C and -D). The distribution of 11 of the analysed sequences within the HAdV species is represented in Figure 16, Figure 17, and Figure 18 for -A and -F, -C, and -D, respectively. Additionally, we verified that one of the contigs obtained did not even partially overlap with the others due to the random fragmentation of DNA molecules before library preparation. When this sequence was analysed separately it was found to cluster with the HAdV-D26 type (Figure B4 of appendix B section). Overall, in accordance with the results described for the molecular cloning approach (section 3.7), it was not possible to assign one specific HAdV-D type to some of the sequences.

Five additional trees were constructed to confirm the results with BLASTn identity values below 95 %. According to the topology of the obtained trees for the non-human AdV sequences, 34 are suggested to be related to MAdV-2 (Figure B5, Figure B6, Figure B7, and Figure B8 of appendix B section). The other three sequences clustered with HAdV-D42 (Figure B9 of appendix B section).

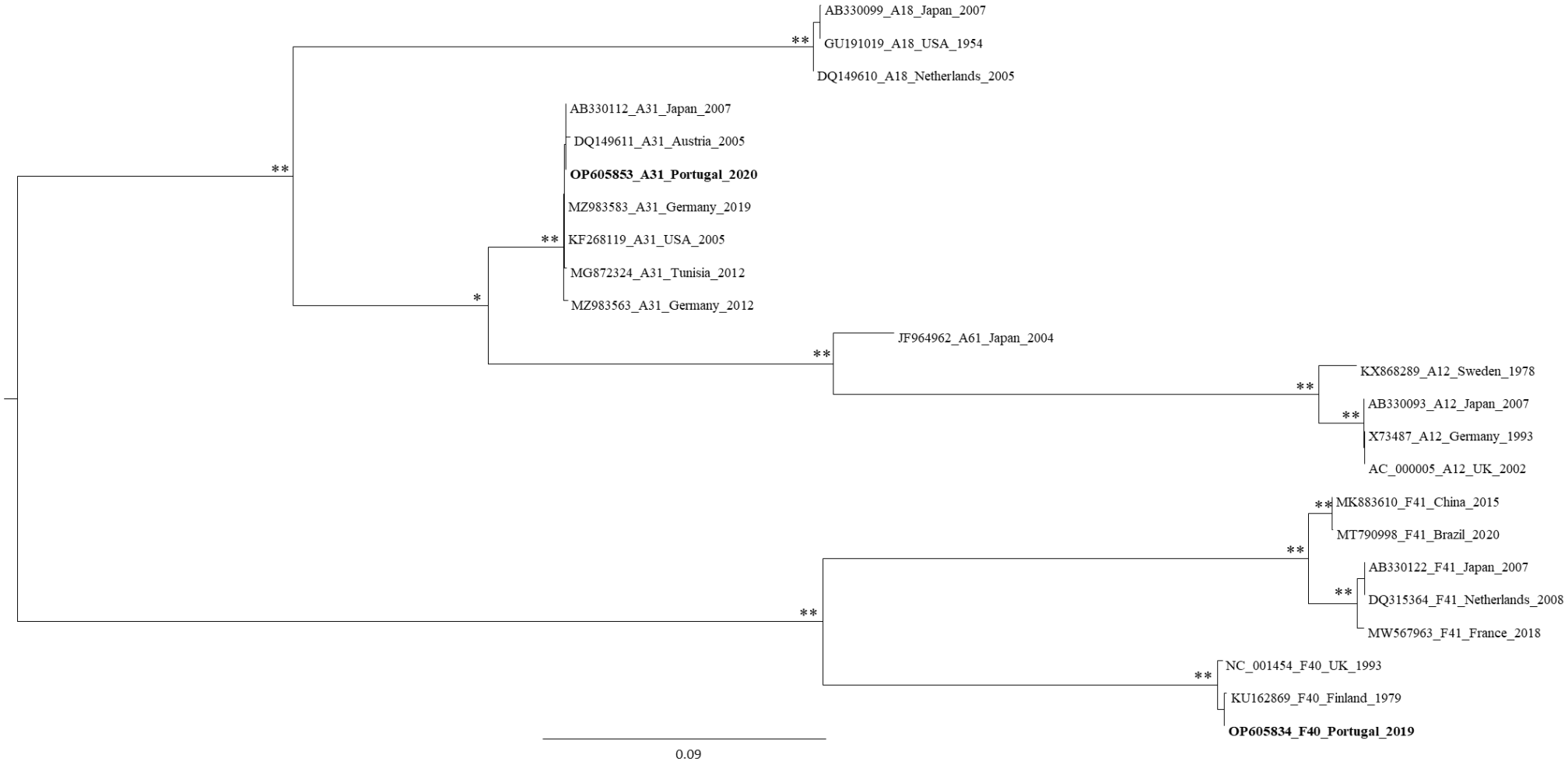

Figure 16. Phylogenetic tree by Maximum Likelihood of HAdV-A and -F hexon capsid gene sequences, including two Portuguese sequences obtained by NGS. Each sequence is identified by its accession number, HAdV type, country of origin and year. The number of "*" indicates the number of methods that support the demonstrated topology, considering 75% (of the total number of data resamplings) and above as relevant for aLRT and bootstrap values.

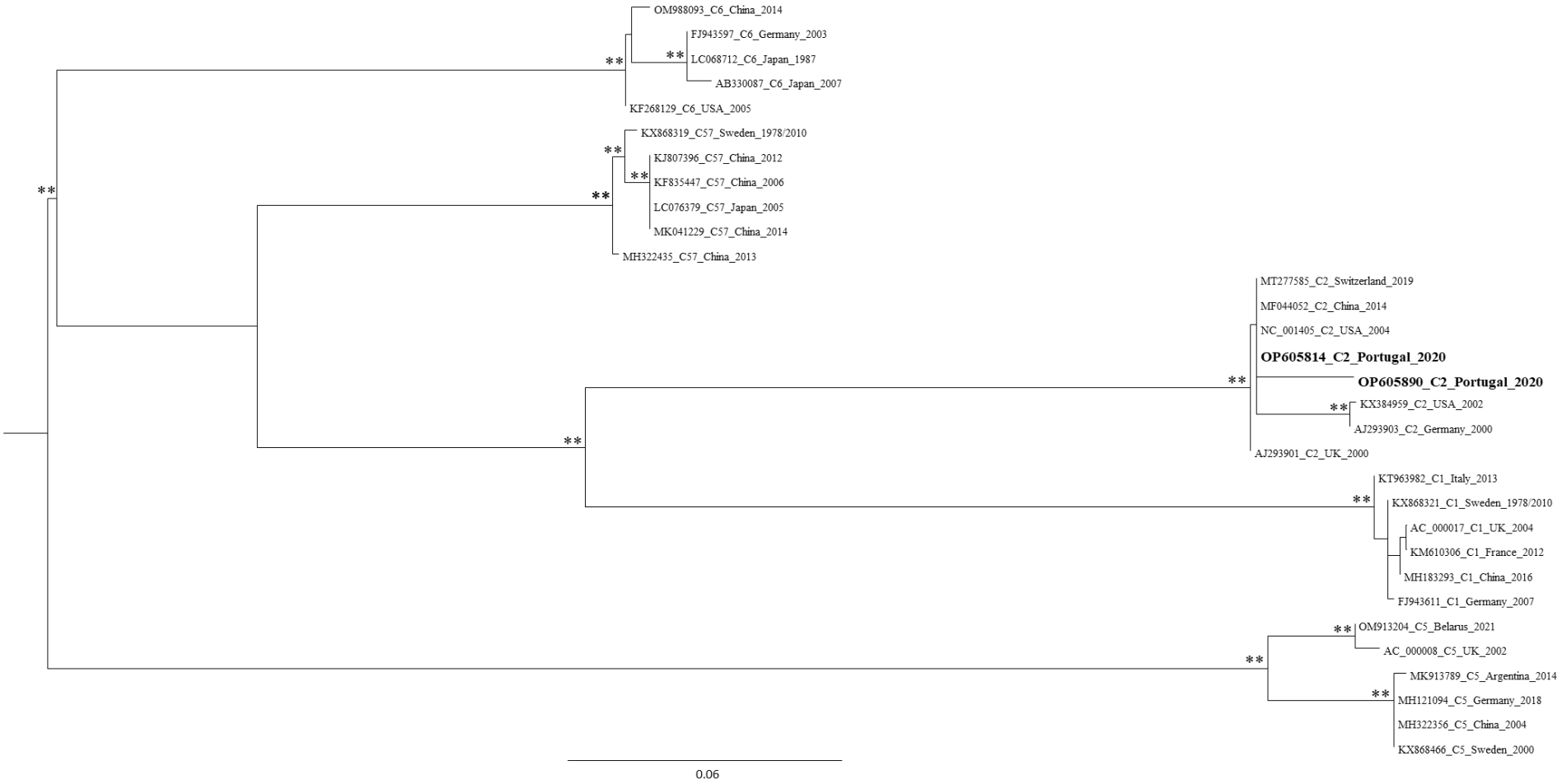

Figure 17. Phylogenetic tree by Maximum Likelihood of HAdV-C hexon capsid gene sequences, including two Portuguese sequences obtained by NGS. Each sequence is identified by its accession number, HAdV type, country of origin and year. The number of "*" indicates the number of methods that support the demonstrated topology, considering 75% (of the total number of data resamplings) and above as relevant for aLRT and bootstrap values.

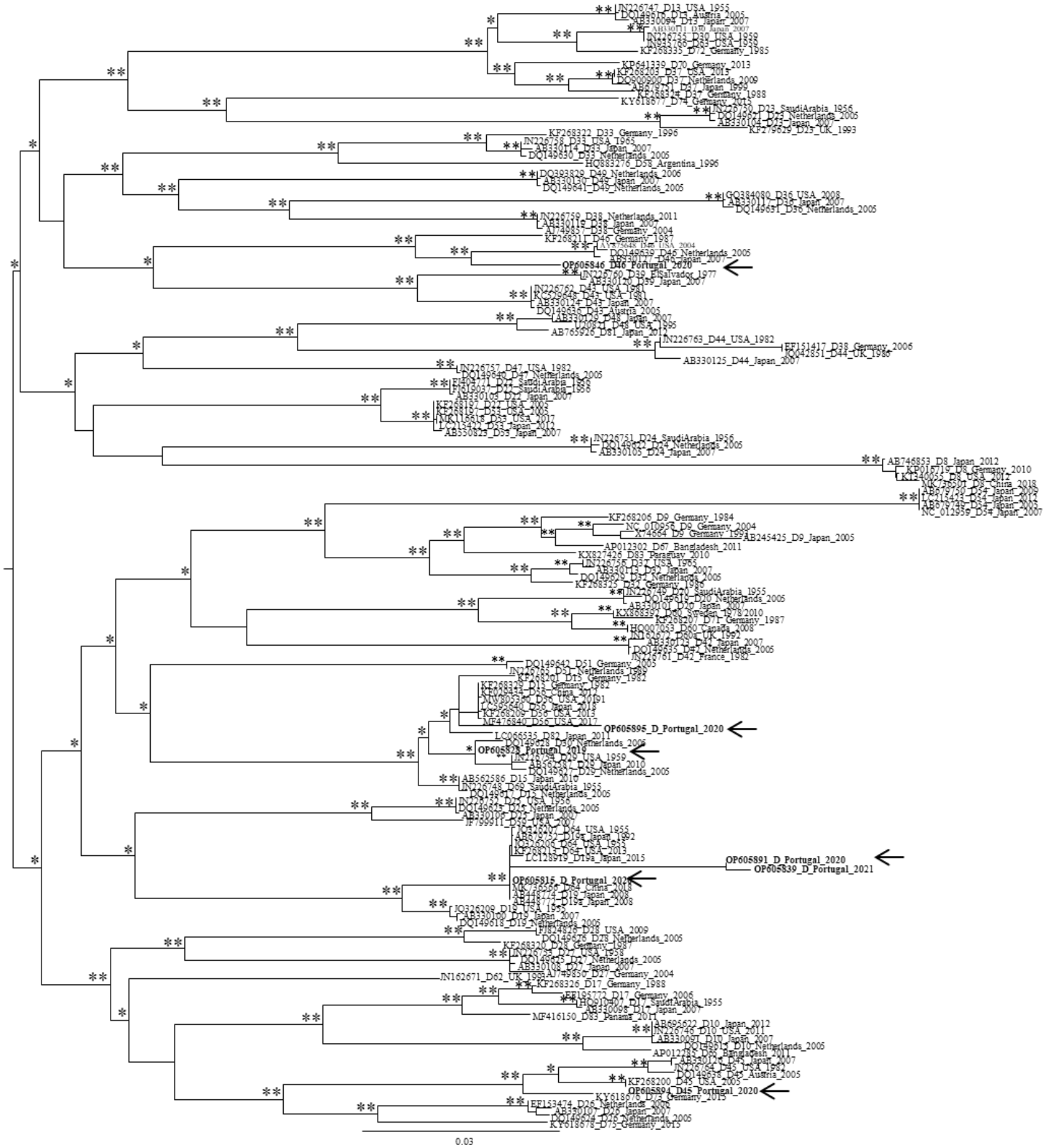

Figure 18. Phylogenetic tree by Maximum Likelihood of HAdV-D hexon capsid gene sequences, including seven Portuguese sequences obtained by NGS. Each sequence is identified by its accession number, HAdV type, country of origin and year. The number of "*" indicates the number of methods that support the demonstrated topology, considering 75% (of the total number of data resamplings) and above as relevant for aLRT and bootstrap values. The arrows indicate the Portuguese sequences.

Regarding all obtained contigs, this approach revealed the presence of AdV genomes in all samples analysed (14 WWTP and six environmental). In sum, from all contigs, 93 (73 %) were characterized as HAdV and 34 (27 %) were suggested as related to MAdV. Figure 19 represents all types assigned including the ones obtained by phylogeny.

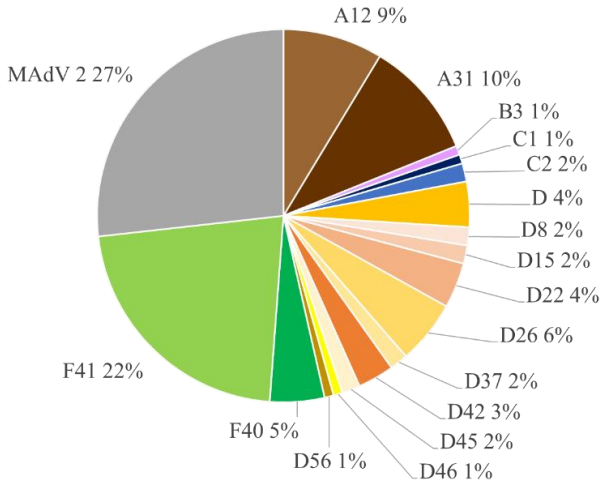

Figure 19. Graphical distribution of types assigned to the 127 Portuguese NGS sequences, including the ones also analysed by phylogeny. Each colour represents one viral type/species: A12 – light brown, A31 – dark brown, B3 – light purple, C1 – dark blue, C2 – light blue, D – a variation of yellow and orange, F40 – dark green, F41 – light green, murine mastadenovirus (MAdV) 2 – grey.

NGS provide a higher quantity of sequencing data, and the sequences produced can be used to characterise and identify pathogens more precisely (45) even in complex samples. Additionally, we verified that the first method, which involved a molecular cloning step before the sequencing step, only revealed the most prevalent types, while NGS allowed the identification of a wider diversity of types from HAdV species. As also suggested by Iaconelli et al. (34), these results show that a more comprehensive picture of the distribution of HAdV types in water can be predicted from water monitorization by NGS technologies, in particular using the Illumina MiSeq system, which is the most used for pathogen surveillance (45).

Regarding HAdV types, in accordance with previous studies and the obtained results by Sanger sequencing, NGS data also showed that -F41 was dominant (22 % of

all sequences in this case). In other words, regardless of the strategy used, this type was identified as the most frequent HAdV type. Furthermore, our findings support previous studies, since HAdV-A12, -A31, -B3, -C1, -C2, -D26, -D37, -D42, and -F40 types were also reported in European countries near Portugal (such as Italy and Spain) (10–12,22,29,32,34,56,57).

Figure 20 represents the distribution of the AdVs according to the type of sample analysed. NGS data also demonstrated that in WWTP samples a higher diversity of HAdV types was found, when compared with the environmental samples, corroborating the previous results.

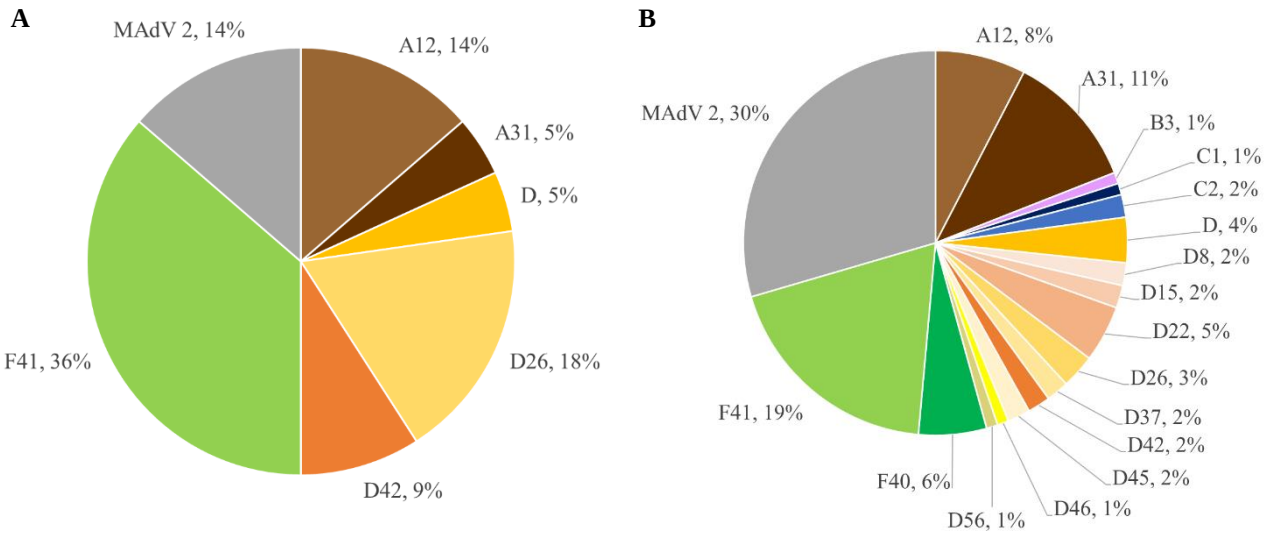

Figure 20. Graphical distribution of AdV types assigned to all Portuguese NGS sequences according to the water sample type. **(A)** Detected types in environmental samples. **(B)** Detected types in WWTP samples. Each colour represents one viral type/species: A12 – light brown, A31 – dark brown, B3 – light purple, C1 – dark blue, C2 – light blue, D – a variation of yellow and orange, F40 – dark green, F41 – light green, murine mastadenovirus (MAdV) 2 – grey.

Concerning the environmental samples, in this case, more than one HAdV type was found (Figure 20), suggesting that others (e.g. HAdV-A12, -A31, -D26, -D42) can also be resistant to environmental conditions. Osuolale et al. (51) demonstrated that the highest detection rates of HAdV were in the winter season, probably due to the reduction in efficacy that low temperatures have in WWTPs' treatment processes, which impacts the

quality of the treated effluent. As a result, this treated effluent can be leaked and contaminate the aquatic environment. Although our samples were not collected in the winter season, they were collected close to the beginning of winter when the temperatures were decreasing. This might explain the detection of some HAdV types in the environment, in particular in ditches/creeks.

3.9. Comparison of the genetic characterization of AdV by Sanger and Illumina approaches

Table 8 lists all the nucleotide sequences obtained in this work from the molecular cloning followed by Sanger sequencing and NGS methodologies, the corresponding HAdV species/type assigned and also the possible non-human AdV detected.

Table 8. List of all assigned types to each nucleotide sequence described in this work, respective BLAST and phylogeny result and corresponding accession number.

Expected species	Sample	NGS					Molecular cloning and Sanger sequencing				
		Accession number	BLAST result	Query Cover (%)	Identity (%)	Phylogeny result	Accession number	BLAST result	Query Cover (%)	Identity (%)	Phylogeny result
A or F	WWTP A1	ALO_1_28N_18	HAdV-D53	26	86	MAdV-2	OP605770	HAdV-F41	100	99	HAdV-F41
		ALO_1_7N_18	MAdV-2	62	71						
		OP605864	HAdV-F40	100	100	n.a	OP605771		100	100	
		OP605869	HAdV-F41	100	99		OP605772		100	100	
		OP605868		100	99		OP605773		100	99	
		OP605867	HAdV-D26	100	91	HAdV-D42	OP605774		100	97	
		OP605865	HAdV-A12	100	98	n.a	OP605775		100	99	
		OP605866	HAdV-A31	100	100		OP605769		100	99	
	WWTP B2	CHA_1_10N_19	MAdV-2	21	85	MAdV-2	OP605802	HAdV-F41	100	100	HAdV-F41
		OP605827	HAdV-D45	100	100	n.a					
		OP605831	HAdV-D37	100	100		OP605803		100	100	
		OP605828	HAdV-D	100	100	HAdV-D					
		OP605829	HAdV-F41	100	99	n.a	OP605804	HAdV-A12	100	99	HAdV-A12
		OP605830		100	99						
		OP605834	HAdV-F	100	100	HAdV-F40					
		OP605832	HAdV-A12	100	98	n.a	OP605801	HAdV-D	100	99	HAdV-D
		OP605833	HAdV-B3	100	99						
		OP605826	HAdV-F41	100	100						
	WWTP C	OP605842	HAdV-D22	98	96	n.a	Excluded due to the preliminary taxonomic results by BLASTn (non-similarity found)				
		OP605843	HAdV-A12	100	98						
		OP605844	HAdV-F40	98	99						
		OP605847	HAdV-F41	100	100						
		OP605841	HAdV-A31	100	100						
		OP605845	HAdV-C1	100	99						
		OP605846	HAdV-D	100	96	HAdV-D46					

A or F	WWTP E	BEN_1_11N_20	HAdV-D	13	88	MAdV-2	OP605788	HAdV-A31	100	99	HAdV-A31
		BEN_1_10N_20	HAdV-D81	13	87						
		OP605891	HAdV-D	100	99	HAdV-D					
		OP605886	HAdV-D8	100	98	n.a					
		OP605887	HAdV-D15	100	100						
		OP605885	HAdV-D22	100	99						
		OP605889	HAdV-F41	100	100						
		OP605888	HAdV-A31	100	100						
		OP605890	HAdV-C	100	99	HAdV-C2					
	WWTP F	CAN_1_24N_20	MAdV-2	100	71	MAdV-2	OP605791	Adenovirus	100	99	HAdV-F41
		CAN_1_21N_20	MAdV 2	52	82						
		OP605811	HAdV-A12	100	99	n.a					
		OP605808		100	99						
		OP605807	HAdV-A31	100	99						
		OP605810		99	99						
		OP605809	HAdV-F41	100	99		OP605792	Adenovirus	100	99	HAdV-F41
		OP605813	HAdV-D26	100	99						
		OP605812	HAdV-D42	100	90	HAdV-D42					
	Alhos Vedros	OP605880	HAdV-D26	100	99	n.a	OP605785	Adenovirus	100	100	HAdV-F41
		OP605882		100	98						
		OP605883	HAdV-A12	100	98		OP605786		100	100	
		OP605884	HAdV-F41	100	99						
		OP605879	HAdV-D42	99	91	HAdV-D42	OP605787		100	100	
		OP605881	HAdV-F41	100	99	n.a					
	Fogueteiro	Not included					Excluded due to the preliminary taxonomic results by BLASTn (non-AdV or non-similarity found)				
	Lizandro River										
	Ribeira de Caparide										
	Sado River – Industrial Area of Setúbal										
	Trancão River	OP605859	HAdV-A31	100	100	n.a					
		OP605860	HAdV-F41	100	99						
		OP605861	HAdV-D	99	91	HAdV-D42					
		OP605862	HAdV-D26	99	99	n.a					
		OP605863	HAdV-A12	99	98						

B or E	WWTP A1	ALO_2_16N_18	MAdV-2	20	86	MAdV-2	OP605776	HAdV-F41	100	99	HAdV-F41
							OP605777		100	99	
		OP605870	HAdV-F41	100	98	n.a	OP605778		100	99	
							OP605779		100	99	
		OP605871	HAdV-F41	100	97		OP605780		100	99	
							OP605781		100	99	
	WWTP B2	CHO_2_19N_18	HAdV-D81	15	87	MAdV-2	Selected to be directly analysed by NGS				
		OP605836	HAdV-A31	100	99	n.a					
		OP605835	HAdV-D56	100	100						
	WWTP C	QN_2_20N_20	HAdV	15	87	MAdV-2					
		QN_2_24N_20	HAdV-D	20	85						
		QN_2_26N_20	HAdV-D81	16	87						
		QN_2_27N_20	HAdV-D	15	85						
		QN_2_84N_20	HAdV-D	12	93						
		QN_2_10N_20	HAdV-D15	100	99	n.a					
		OP605849	HAdV-F41	100	99						
		OP605850		100	99						
		OP605851	HAdV-A31	100	99						
		OP605852	HAdV-D26	100	99						
		OP605853	HAdV-A	100	100	HAdV-A31					
	WWTP D	SN_2_9N_20	HAdV-D53	19	86	MAdV-2					
		SN_2_10N_20	MAdV-2	20	83						
		SN_2_25N_20	MAdV-2	8	91						
		SN_2_42N_20	HAdV-D81	21	88						
		OP605856	HAdV-F41	100	100	n.a					
	WWTP E	BEN_2_16N_20	MAdV-2	16	85	MAdV-2					
		BEN_2_19N_20		16	86						
		BEN_2_23N_20		13	86						
		BEN_2_31N_20		20	85						
		OP605892	HAdV-F41	100	99	n.a					
		OP605894	HAdV-D	100	99	HAdV-D45					
		OP605893	HAdV-A12	100	98	n.a					

B or E	WWTP F	CAN_2_28N_20	HAdV	16	88	MAdV-2	OP605793	HAdV-F41	100	99	HAdV-F41
		CAN_2_35N_20	MAdV-2	20	83						
		CAN_2_38N_20	HAdV-D81	13	87						
		CAN_2_39N_20	HAdV-D	14	88						
		OP605814	HAdV-C	100	100	HAdV-C2	OP605794	HAdV-F41	100	99	HAdV-F41
		OP605815	HAdV-D	100	100	HAdV-D					
		OP605818	HAdV-D37	100	99	n.a	OP605795	HAdV-D32	100	98	HAdV-D32
		OP605817	HAdV-A31	100	100						
		OP605820		100	100						
		OP605816	HAdV-D22	100	97						
		OP605819	HAdV-F41	100	99						
		OP605821	HAdV-F41	100	100						
	Ribeira das Vinhas	RVO_2_45N_21	MAdV-2	19	83	MAdV-2	Not cloned due to the preliminary taxonomic results by BLASTn for amplicons obtained from PCR A+F				
		RVO_2_79N_21		51	79						
		OP605855	HAdV-F41	100	99	n.a					
		OP605854	HAdV-F41	100	99						
	Ribeira de Caparide	OP605858	HAdV-A12	100	99						
		OP605857	HAdV-F41	100	99						
	Corroios	CRO_2_32N_21	MAdV-2	20	83	MAdV-2	CRO_2_3S_21	HAdV-D	22	87	MAdV 2
		OP605839	HAdV-D	96	94	HAdV-D					
		OP605840	HAdV-D	100	97	HAdV-D26	CRO_2_6S_21		18	86	
D	WWTP A1	ALO_3_54N_18	MAdV 2	37	81	MAdV-2	OP605783	HAdV-C5	100	99	HAdV-C5
		ALO_3_67N_18		100	68						
		OP605875	HAdV-F40	100	98	n.a	OP605784		100	99	
		OP605876	HAdV-F40	100	99						
		OP605877	HAdV-D8	100	98		OP605782		100	99	
		OP605872	HAdV-F41	100	99						
		OP605873	HAdV-A12	100	98						
		OP605874	HAdV-A31	100	99						
		OP605878		100	96						

D	WWTP E	BEN_3_7N_20	MAdV 2	45	79	MAdV-2	OP605790	HAdV-C2	100	99	HAdV-C2
		OP605806	HAdV-A12	100	99	n.a	OP605789		100	99	
		OP605896	HAdV-F41	100	99						
		OP605897		100	99						
		OP605805	HAdV-D22	100	97						
		OP605895	HAdV-D	100	99	HAdV-D					
	WWTP F	CAN_3_12N_20	MAdV-2	61	74	MAdV-2	OP605797	HAdV-D	100	96	HAdV-D
		CAN_3_11N_20		62	72		OP605798	HAdV-C2	100	99	HAdV-C2
		OP605824	HAdV-F40	100	98	n.a	OP605799	HAdV-D	100	95	HAdV-D
		OP605822	HAdV-F41	100	99		OP605800	HAdV-D	100	95	HAdV-D
		OP605825	HAdV-D26	100	99						
		OP605823	HAdV-D22	100	97		OP605796	HAdV-D25	100	95	HAdV-D
	Ribeira do Carenque	OP605837	HAdV-F41	100	99		CQN_3_2S_20	MAdV-2	80	69	MAdV-2
		OP605838	HAdV-F41	100	99		CQN_3_7S_20	MAdV-2	38	72	
							CQN_3_11S_20	MAdV-2	77	69	

“NGS” stands for “Next Generation Sequencing”, “n.a.” for “not applied”, “HAdV” for “human mastadenovirus” and “MAdV” for “murine mastadenovirus”.

Only nucleotide sequences obtained during this work associated with HAdV were deposited in the GenBank/ENA/EMBL public genomic database.

The identification of viral genomes corresponding to HAdV-A12, -A31, -C2,-F40, and -F41 types was consistent in both approaches used in this work. However, in contrast with the study from Ogorzaly et al. (56) where Sanger sequencing results were highly consistent with NGS results for all analysed samples, our results differed between some samples. For two amplicons from PCR D of WWTP E and Ribeira do Carenque samples the results of both approaches were not the same (Table 8). This was probably because the two methods were based on PCR amplifications performed separately, where the use of a different aliquot, for example, could led to different results.

Moreover, the BLASTn identification performed did not detect AdV DNA in some sequences obtained from the molecular cloning approach, such as WWTP C, Lizandro River, Ribeira de Caparide, Sado River, Trancão River samples regarding PCR A+F, and WWTP C sample regarding PCR D. One possible explanation for this could be that non-viral DNA was associated with the purified products and was probably favourably cloned when compared to the viral DNA. Thus, in this case, the classic strategy was not able to confirm, on the number of clones analysed, the result suggested by the PCR amplification. However, given that there is a high probability of finding HAdV DNA at least in WWTP samples, this is quite an odd occurrence. Thus, in the NGS approach, some of these were included (Trancão River and WWTP C amplicons regarding PCR A+F; Ribeira das Vinhas e Ribeira de Caparide regarding PCR B+E). NGS results showed the presence of AdV DNA in all (Table 8). Therefore, probably the molecular cloning approach introduced some qualitative biases in our results. Indeed, this reflects the limitation of incorporating a molecular cloning strategy in the workflow, when only a limited number of molecules is studied. This highlights the advantages and the wide sequencing capacity of NGS even for complex water samples. The presence of HAdV DNA in Lizandro River, Ribeira de Caparide, Sado River, and WWTP C samples, as suggested by electrophoretic analysis (i.e., the amplicon obtained presented the expected size), cannot be neglected and these samples cannot be deemed as “false-positive”.

The results of BLASTn and phylogeny allowed us to unambiguously assign the corresponding type for each HAdV sequence, or at least, the species (in the case of some HAdV-D types). In contrast, this was not achievable for MAdV sequences. Although our

phylogenetic reconstructions suggest that the obtained sequences share an evolutionary relationship with MAdV-2, these sequences cannot be characterized with precision, since BLASTn query cover and identity values were below 90 % and 95 %, respectively. More comprehensive identification of these sequences would require, for example, the analysis of the total genome and not only the analysis of a region of the hexon gene.

LMA is one of the largest urban areas of Portugal, corresponding to 18 municipalities with a population of 2,870,770 in 2021 (https://censos.ine.pt/db_censos_2021.html accessed on 23/08/2022). This population is extremely diversified, including people from many ethnic origins, mostly from old Portuguese colonies, such as Angola, Cape Verde, Guinea-Bissau, and Mozambique, but also from Eastern Europe, India, Brazil, and China. Indeed, in 2019 the main immigrant nationalities in Portugal corresponded to Brazil (25.6 %), followed by Cape Verde (6.3 %) and the United Kingdom (5.8 %) (https://www.gee.gov.pt/estatisticas-de-imigrantes-em-portugal-por-nacionalidade accessed on 23/08/2022). Hence, identifying several HAdV types in LMA waters, including those described in other worldwide studies, is not surprising.

The community's viral genetic diversity can be identified objectively using wastewater for enteric virus surveillance (42). The data from both approaches applied in this study provided information about the molecular epidemiological distribution of HAdV in the Portuguese population. Taking our results into account, HAdV-A12, -A31, -B3, -B76, -C1, -C2, -C5, -D8, -D15, -D22, -D26, -D32, -D37, -D42, -D45, -D56, -F40, and -F41 types were in circulation in Portugal, especially in LMA, between 2018 and 2021 either in waste or environmental waters. Finding a high diversity of HAdV types is not, a priori, a concern since it does not mean they were causing disease. Moreover, these viruses can persist in the infected population, not inducing any kind of disease, and remaining silent. Of all types detected in this study, HAdV-B3, -C1, -C2, -C5, and -F41 were frequently reported in the literature as associated with human disease (22,38) but were detected in this study with low prevalence (1–2 % of NGS results and 9 % of molecular cloning approach results), except HAdV-F41 that was the most frequent (22 % of NGS results and 54 % of molecular cloning approach results). Although HAdVs are often efficiently eliminated in secondary treatment, with a considerable reduction in the titer of infectious viruses (42), wastewater can contaminate

surface water. Considering this, the relatively high detection of HAdV-F41, which is a common cause of severe gastroenteritis and diarrhoea-related death in young children across the world (24,40), can be a concern for Portuguese children. In a similar study, Lun et al. (42) demonstrated that HAdV-F41 was the most detected type in clinical samples. Although only water samples were analysed in this study, it is possible that this type could be associated with disease in some children from the LMA and, consequently, detected in the studied samples. HAdV detection in water samples highlights the effectiveness of using these samples for virus surveillance (42) but more studies must be performed to evaluate the infectivity of these viruses, and subsequently, to assess their risk to public health. Even though it is known that AdV can persist infectious for approximately one year (97), the detection of their genome does not always indicate a real risk to human health (10).

In addition, this study included water samples collected before and after the COVID-19 world pandemic, which led to restrictions on mobility between countries, and within the country, to control population flows (and, consequently, viral circulation). However, our results suggest that HAdV diversity remained almost the same between 2018 and 2021 in the LMA. For example, in WWTP A1 (collected in 2018) and WWTP C (collected in 2020), the genomes of the same types (HAdV-A12, -A31, -F40, -F41) were detected. In addition, HAdV-A12, -A31, -D26, -D42, and -F41 were detected in the Trancão River sample (collected in 2020) and in the Alhos Vedros sample (collected in 2021).

According to several studies that proposed the use of HAdV as an indication of viral contamination in water (4,12,13), this study demonstrated (through the detection of AdV genomes in wastewater and environmental samples), that HAdV can be useful to monitor faecal contamination. The presence of HAdV DNA in the environmental samples suggests that the streams and ditches analysed from LMA were, at some point, contaminated with faecal matter, probably due to the inefficiency of the conventional treatments applied by WWTPs, or due to direct contamination from sewage spills. This contamination can represent a public health problem. Moreover, Corroios and Alhos Vedros ditches were chosen due to the complaint of the nearest residents, regarding the smell after sewage discharges. Before the water collection, the Corroios ditch was cleaned, however, HAdV DNA was still detected (and possibly MAdV-2 DNA). This

underlines the extreme resistance of these viruses to environmental degradation and cleaning processes, especially the resistance of its genome, as revealed in other studies (1,11,33). On the other hand, it can also indicate that the contamination source was not eradicated.

4. Conclusions and perspectives

To our knowledge, only a restricted number of studies, which have included the Porto area (North of Portugal), surveyed the distribution of HAdV types in Portuguese waters (11,29). Additionally, a more recent study performed in the Tagus River basin only assessed the presence of HAdV and their quantification (86), but not their diversity. In contrast, this study focused on the molecular diversity of HAdV types circulating in LMA through the analysis of raw sewage (i.e., influent) and environmental water samples.

The obtained results concerning the molecular cloning and NGS approaches revealed the presence of viral genomes related to the HAdV-A, -B, -C, -D and -F species, and also possibly related to MAdV-2. HAdV-F41 was the most predominant type, corroborating the literature reports worldwide.

Despite its limitations, including the introduction of qualitative bias in the results by the molecular cloning step, this study allowed the characterization of different genomic types of HAdV circulating in LMA. Along with other studies, this work uncovers the occurrence of some HAdV types in wastewater. It also demonstrates the presence of faecal contamination in the environment, probably due to the inefficiency of the conventional treatments applied by WWTPs to remove some types or due to direct contamination from sewage. Furthermore, the assessment of the diversity of this virus (through the study of its genomes) provides important information about the patterns of its molecular epidemiological distribution in the general population.

There is still a lack of information regarding the diversity of these viruses in Portugal. Indeed, studies like the one here described are essential to further investigate the genetic diversity of HAdV circulating in other areas of the country. Finally, considering that molecular methods are the most used, but cannot provide information regarding the infectious status of the virus, it would be interesting to perform infectivity assays, to verify the effectiveness that the treatments carried out by the WWTP have in neutralizing these agents and to assess their risk to the public health.

5. Bibliography

1. Verani M, Federigi I, Donzelli G, Cioni L, Carducci A. Human adenoviruses as waterborne index pathogens and their use for Quantitative Microbial Risk Assessment. Sci Total Environ. 2019;651:1469–75.

2. Magana-Arachchi DN, Wanigatunge RP. Ubiquitous waterborne pathogens. In: Waterborne Pathogens. Elsevier. 2020;15–42.

3. Gonzales-Gustavson E, Rusiñol M, Medema G, Calvo M, Girones R. Quantitative risk assessment of norovirus and adenovirus for the use of reclaimed water to irrigate lettuce in Catalonia. Water Res. 2019;153:91–9.

4. La Rosa G, Pourshaban M, Iaconelli M, Muscillo M. Quantitative real-time PCR of enteric viruses in influent and effluent samples from wastewater treatment plants in Italy. Ann Ist Super SanItà. 2010;46(3):266–73.

5. U.S. Environmental Protection Agency. Bacteria: Indicators of Potential Pathogens, p. 25. In Agency. In: Volunteer Estuary Monitoring Manual, A Methods Manual. 2nd ed. 2006.

6. World Health Organization. Guidelines for drinking-water quality: fourth edition incorporating first addendum. 4th ed. Geneva: World Health Organization; 2017. 541 p.

7. Rusiñol M, Fernandez-Cassi X, Timoneda N, Carratalà A, Abril JF, Silvera C, et al. Evidence of viral dissemination and seasonality in a Mediterranean river catchment: Implications for water pollution management. J Environ Manage. 2015;159:58–67.

8. Calgua B, Fumian T, Rusiñol M, Rodriguez-Manzano J, Mbayed VA, Bofill-Mas S, et al. Detection and quantification of classic and emerging viruses by skimmed-milk flocculation and PCR in river water from two geographical areas. Water Res. 2013;47(8):2797–810.

9. Kokkinos PA, Ziros PG, Mpalasopoulou A, Galanis A, Vantarakis A. Molecular detection of multiple viral targets in untreated urban sewage from Greece. Virol J. 2011;8(1):195.

10. Iaconelli M, Muscillo M, Della Libera S, Fratini M, Meucci L, De Ceglia M, et al. One-year Surveillance of Human Enteric Viruses in Raw and Treated Wastewaters, Downstream River Waters, and Drinking Waters. Food Environ Virol. 2017;9(1):79–88.

11. Wyn-Jones, A.P., Carducci, A., Cook, N., D'Agostino, M., Divizia, M., Fleischer, J., Gantzer, C., Gawler, A., Girones, R., Höller, C., de Roda Husman, A.M., Kay, D., Kozyra, I., López-Pila, J., Muscillo, M., José Nascimento, M.S., Papageorgiou, G., Rutjes, S., Sellwood, J., Szewzyk, R., Wyer, M. Surveillance of adenoviruses and noroviruses in European recreational waters. Water Res. 2011;45:1025–1038. https://doi.org/10.1016/j.watres.2010.10.015

12. Bofill-Mas S, Albinana-Gimenez N, Clemente-Casares P, Hundesa A, Rodriguez-Manzano J, Allard A, et al. Quantification and Stability of Human Adenoviruses and Polyomavirus JCPyV in Wastewater Matrices. Appl Environ Microbiol. 2006;72(12):7894–6.

13. Hewitt J, Greening GE, Leonard M, Lewis GD. Evaluation of human adenovirus and human polyomavirus as indicators of human sewage contamination in the aquatic environment. Water Res. 2013;47(17):6750–61.

14. Bisseux M, Colombet J, Mirand A, Roque-Afonso AM, Abravanel F, Izopet J, et al. Monitoring human enteric viruses in wastewater and relevance to infections encountered in the clinical setting: a one-year experiment in central France, 2014 to 2015. Eurosurveillance. 2018;23(7).

15. Fong, T.T., Lipp, E.K. Enteric Viruses of Humans and Animals in Aquatic Environments: Health Risks, Detection, and Potential Water Quality Assessment Tools. Microbiol. Mol. Biol. Rev. 2005;69,357–371. https://doi.org/10.1128/MMBR.69.2.357-371.2005

16. Grøndahl-Rosado RC, Yarovitsyna E, Trettenes E, Myrmel M, Robertson LJ. A One Year Study on the Concentrations of Norovirus and Enteric Adenoviruses in Wastewater and A Surface Drinking Water Source in Norway. Food Environ Virol. 2014;6(4):232–45.

17. Okoh AI, Sibanda T, Gusha SS. Inadequately Treated Wastewater as a Source of Human Enteric Viruses in the Environment. Int J Environ Res Public Health. 2010;7(6):2620–37.

18. Benkő M, Aoki K, Arnberg N, Davison AJ, Echavarría M, Hess M, et al. ICTV Virus Taxonomy Profile: Adenoviridae 2022: This article is part of the ICTV Virus Taxonomy Profiles collection. J Gen Virol. 2022;103(3).

19. Besson S, Vragniau C, Vassal-Stermann E, Dagher MC, Fender P. The Adenovirus Dodecahedron: Beyond the Platonic Story. Viruses. 2020;12(7):718.

20. Maclachlan NJ, Dubovi EJ, Barthold SW, Swayne DE, Winton JR, editors. Fenner's veterinary virology. Fifth edition. Amsterdam: Elsevier/AP, Academic Press is an imprint of Elsevier. 2017;581p.

21. Mennechet FJD, Paris O, Ouoba AR, Salazar Arenas S, Sirima SB, Takoudjou Dzomo GR, et al. A review of 65 years of human adenovirus seroprevalence. Expert Rev Vaccines. 2019;18(6):597–613.

22. Allard A, Vantarakis A. Adenoviruses. In: Michigan State University, Rose JB, Jiménez Cisneros B, UNESCO - International Hydrological Programme, editors. Water and Sanitation for the 21st Century: Health and Microbiological Aspects of Excreta and Wastewater Management (Global Water Pathogen Project) [Internet]. Michigan State University. 2017. Available from: https://www.waterpathogens.org/book/adenoviruses

23. Gallardo J, Pérez-Illana M, Martín-González N, San Martín C. Adenovirus Structure: What Is New? Int J Mol Sci. 2021;22(10):5240.

24. Rafie K, Lenman A, Fuchs J, Rajan A, Arnberg N, Carlson LA. The structure of enteric human adenovirus 41—A leading cause of diarrhea in children. Sci Adv. 2021;7(2):eabe0974.

25. Pied N, Wodrich H. Imaging the adenovirus infection cycle. FEBS Lett. 2019;593(24):3419–48.

26. Crenshaw, B.J., Jones, L.B., Bell, C.R., Kumar, S., Matthews, Q.L. Perspective on Adenoviruses: Epidemiology, Pathogenicity, and Gene Therapy. Biomedicines. 2019;7,61. https://doi.org/10.3390/biomedicines7030061

27. Flint SJ, Racaniello VR, Rall GF, Skalka AM, Enquist LW. Principles of virology. 4th edition. Washington, DC: ASM Press. 2015.

28. Charman M, Herrmann C, Weitzman MD. Viral and cellular interactions during adenovirus DNA replication. FEBS Lett. 2019;593(24):3531–50.

29. Bofill-Mas S, Calgua B, Clemente-Casares P, La Rosa G, Iaconelli M, Muscillo M, et al. Quantification of Human Adenoviruses in European Recreational Waters. Food Environ Virol. 2010;2(2):101–9.

30. Purpari G, Macaluso G, Di Bella S, Gucciardi F, Mira F, Di Marco P, et al. Molecular characterization of human enteric viruses in food, water samples, and surface swabs in Sicily. Int J Infect Dis. 2019;80:66–72.

31. Rames E, Roiko A, Stratton H, Macdonald J. Technical aspects of using human adenovirus as a viral water quality indicator. Water Res. 2016;96:308–26.

32. Artieda J, Piñeiro L, González MC, Muñoz MJ, Basterrechea M, Iturzaeta A, et al. A swimming pool-related outbreak of pharyngoconjunctival fever in children due to adenovirus type 4, Gipuzkoa, Spain, 2008. Eurosurveillance. 2009;14(8).

33. Thompson SS, Jackson JL, Suva-Castillo M, Yanko WA, El Jack Z, Kuo J, et al. Detection of Infectious Human Adenoviruses in Tertiary-Treated and Ultraviolet-Disinfected Wastewater. Water Environ Res. 2003;75(2):163–70.

34. Iaconelli M, Valdazo-González B, Equestre M, Ciccaglione AR, Marcantonio C, Della Libera S, et al. Molecular characterization of human adenoviruses in urban wastewaters using next generation and Sanger sequencing. Water Res. 2017;121:240–7.

35. Sedji MI, Varbanov M, Meo M, Colin M, Mathieu L, Bertrand I. Quantification of human adenovirus and norovirus in river water in the north-east of France. Environ Sci Pollut Res. 2018;25(30):30497–507.

36. Mena KD, Gerba CP. Waterborne Adenovirus. In: Reviews of Environmental Contamination and Toxicology Volume 198. New York, NY: Springer New York; 2009;1–35.

37. Farkas K, Cooper DM, McDonald JE, Malham SK, de Rougemont A, Jones DL. Seasonal and spatial dynamics of enteric viruses in wastewater and in riverine and estuarine receiving waters. Sci Total Environ. 2018;634:1174–83.

38. Lynch JP, Kajon AE. Adenovirus: Epidemiology, Global Spread of Novel Serotypes, and Advances in Treatment and Prevention. Semin Respir Crit Care Med. 2016;37(4):586–602.

39. Radke JR, Cook JL. Human adenovirus infections: update and consideration of mechanisms of viral persistence. Curr Opin Infect Dis. 2018;31(3):251–6.

40. Rajan A, Palm E, Trulsson F, Mundigl S, Becker M, Persson BD, et al. Heparan Sulfate Is a Cellular Receptor for Enteric Human Adenoviruses. Viruses. 2021;13(2):298.

41. Heim A, editor. Diagnosis, Treatment and Prevention of Virus Infections. In: Encyclopedia of virology. 4th edition. Amsterdam Boston Heidelberg: Academic Press, Elsevier. 2021;197–205.

42. Lun JH, Crosbie ND, White PA. Genetic diversity and quantification of human mastadenoviruses in wastewater from Sydney and Melbourne, Australia. Sci Total Environ. 2019;675:305–12.

43. Jiang SC. Human adenoviruses in water: Occurrence and health implications: A critical review. Environ Sci Technol. 2006;40(23):7132–40.

44. Muscillo M, Pourshaban M, Iaconelli M, Fontana S, Di Grazia A, Manzara S, et al. Detection and Quantification of Human Adenoviruses in Surface Waters by Nested PCR, TaqMan Real-Time PCR and Cell Culture Assays. Water Air Soil Pollut. 2008;191(1–4):83–93.

45. Berry IM, Melendrez MC, Bishop-Lilly KA, Rutvisuttinunt W, Pollett S, Talundzic E, et al. Next Generation Sequencing and Bioinformatics Methodologies for Infectious Disease Research and Public Health: Approaches, Applications, and Considerations for Development of Laboratory Capacity. J Infect Dis. 2019;jiz286.

46. Lee SH, Kim SJ. Detection of infectious enteroviruses and adenoviruses in tap water in urban areas in Korea. Water Res. 2002;36(1):248–56.

47. Dias J, Pinto RN, Vieira CB, de Abreu Corrêa A. Detection and quantification of human adenovirus (HAdV), JC polyomavirus (JCPyV) and hepatitis A virus (HAV) in recreational waters of Niterói, Rio de Janeiro, Brazil. Mar Pollut Bull. 2018;133:240–5.

48. Kundu A, McBride G, Wuertz S. Adenovirus-associated health risks for recreational activities in a multi-use coastal watershed based on site-specific quantitative microbial risk assessment. Water Res. 2013 Oct;47(16):6309–25.

49. Elmahdy EM, Shaheen MNF, Rizk NM, Saad-Hussein A. Quantitative Detection of Human Adenovirus and Human Rotavirus Group A in Wastewater and El-Rahawy Drainage Canal Influencing River Nile in the North of Giza, Egypt. Food Environ Virol. 2020;12(3):218–25.

50. Opere WM, John M, Ombori O. Molecular Detection of Human Enteric Adenoviruses in Water Samples Collected from Lake Victoria Waters Along Homa Bay Town, Homa Bay County, Kenya. Food Environ Virol. 2021;13(1):32–43.

51. Osuolale O, Okoh A. Incidence of human adenoviruses and Hepatitis A virus in the final effluent of selected wastewater treatment plants in Eastern Cape Province, South Africa. Virol J. 2015;12(1):98.

52. Quintão TSC, Silva FG, Pereira AL, Araújo WN, Oliveira PM, Souza MBLD, et al. Detection and molecular characterization of enteric adenovirus in treated wastewater in the Brazilian Federal District. SN Appl Sci. 2021;3(7):691.

53. Haramoto E, Katayama H, Oguma K, Ohgaki S. Quantitative analysis of human enteric adenoviruses in aquatic environments: Quantification of human adenoviruses in water. J Appl Microbiol. 2007;103(6):2153–9.

54. Xagoraraki I, Kuo DHW, Wong K, Wong M, Rose JB. Occurrence of Human Adenoviruses at Two Recreational Beaches of the Great Lakes. Appl Environ Microbiol. 2007;73(24):7874–81.

55. Ibrahim C, Hassen A, Pothier P, Mejri S, Hammami S. Molecular detection and genotypic characterization of enteric adenoviruses in a hospital wastewater. Environ Sci Pollut Res. 2018;25(11):10977–87.

56. Ogorzaly L, Walczak C, Galloux M, Etienne S, Gassilloud B, Cauchie HM. Human Adenovirus Diversity in Water Samples Using a Next-Generation Amplicon Sequencing Approach. Food Environ Virol. 2015;7(2):112–21.

57. La Rosa G, Sanseverino I, Della Libera S, Iaconelli M, Ferrero VEV, Barra Caracciolo A, et al. The impact of anthropogenic pressure on the virological quality of water from the Tiber River, Italy. Lett Appl Microbiol. 2017;65(4):298–305.

58. Shih YJ, Tao CW, Tsai HC, Huang WC, Huang TY, Chen JS, et al. First detection of enteric adenoviruses genotype 41 in recreation spring areas of Taiwan. Environ Sci Pollut Res. 2017;24(22):18392–9.

59. Nour I, Hanif A, Zakri AM, Al-Ashkar I, Alhetheel A, Eifan S. Human Adenovirus Molecular Characterization in Various Water Environments and Seasonal Impacts in Riyadh, Saudi Arabia. Int J Environ Res Public Health. 2021;18(9):4773.

60. Nagarajan V, Chen J, Hsu B, Hsu G, Wang J, Hussain B. Prevalence, Distribution, and Genotypes of Adenovirus and Norovirus in the Puzi River and Its Tributaries and the Surrounding Areas in Taiwan. GeoHealth. 2021;5(12).

61. Wolf S, Hewitt J, Greening GE. Viral Multiplex Quantitative PCR Assays for Tracking Sources of Fecal Contamination. Appl Environ Microbiol. 2010;76(5):1388–94.

62. Guerrero-Latorre L, Carratala A, Rodriguez-Manzano J, Calgua B, Hundesa A, Girones R. Occurrence of water-borne enteric viruses in two settlements based in Eastern Chad: analysis of hepatitis E virus, hepatitis A virus and human adenovirus in water sources. J Water Health. 2011;9(3):515–24.

63. Staggemeier R, Heck TMS, Demoliner M, Ritzel RGF, Röhnelt NMS, Girardi V, et al. Enteric viruses and adenovirus diversity in waters from 2016 Olympic venues. Sci Total Environ. 2017;586:304–12.

64. Thongprachum A, Fujimoto T, Takanashi S, Saito H, Okitsu S, Shimizu H, et al. Detection of nineteen enteric viruses in raw sewage in Japan. Infect Genet Evol. 2018;63:17–23.

65. Condez AC, Nunes M, Filipa-Silva A, Leonardo I, Parreira R. Human Polyomaviruses (HPyV) in Wastewater and Environmental Samples from the Lisbon Metropolitan Area: Detection and Genetic Characterization of Viral Structural Protein-Coding Sequences. Pathogens. 2021;10(10):1309.

66. Calgua B, Mengewein A, Grunert A, Bofill-Mas S, Clemente-Casares P, Hundesa A, et al. Development and application of a one-step low cost procedure to concentrate viruses from seawater samples. J Virol Methods. 2008;153(2):79–83.

67. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol. 2013;30(4):772–80.

68. Castresana J. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. Mol Biol Evol. 2000;17(4):540–52.

69. Yoon H, Leitner T. PrimerDesign-M: a multiple-alignment based multiple-primer design tool for walking across variable genomes. Bioinformatics. 2015;31(9):1472–4.

70. Padmanabhan S, Banerjee S, Mandi N. Screening of Bacterial Recombinants: Strategies and Preventing False Positives. In: Brown GG, editor. Molecular Cloning - Selected Applications in Medicine and Biology. InTech. 2011.

71. Trifinopoulos J, Nguyen LT, von Haeseler A, Minh BQ. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Res. 2016;44(W1):W232–5.

72. Kang J, Ismail AM, Dehghan S, Rajaiya J, Allard MW, Lim HC, et al. Genomics-based re-examination of the taxonomy and phylogeny of *human* and *simian*

Mastadenoviruses: an evolving whole genomes approach, revealing putative zoonosis, anthroponosis, and amphizoonosis. Cladistics. 2020;36(4):358–73.

73. Xu W, McDonough MC, Erdman DD. Species-Specific Identification of Human Adenoviruses by a Multiplex PCR Assay. J Clin Microbiol. 2000;38:7.

74. Allard A, Albinsson B, Wadell G. Rapid Typing of Human Adenoviruses by a General PCR Combined with Restriction Endonuclease Analysis. J Clin Microbiol. 2001;39(2):498–505.

75. Hernroth BE, Conden-Hansson AC, Rehnstam-Holm AS, Girones R, Allard AK. Environmental Factors Influencing Human Viral Pathogens and Their Potential Indicator Organisms in the Blue Mussel, *Mytilus edulis*: the First Scandinavian Report. Appl Environ Microbiol. 2002;68(9):4523–33.

76. Heim A, Ebnet C, Harste G, Pring-Åkerblom P. Rapid and quantitative detection of human adenovirus DNA by real-time PCR. J Med Virol. 2003;70(2):228–39.

77. Formiga-Cruz M, Hundesa A, Clemente-Casares P, Albiñana-Gimenez N, Allard A, Girones R. Nested multiplex PCR assay for detection of human enteric viruses in shellfish and sewage. J Virol Methods. 2005;125(2):111–8.

78. Lu X, Erdman DD. Molecular typing of human adenoviruses by PCR and sequencing of a partial region of the hexon gene. Arch Virol. 2006;151(8):1587–602.

79. Avellón A, Pérez P, Aguilar JC, Lejarazu R ortiz de, Echevarría JE. Rapid and sensitive diagnosis of human adenovirus infections by a generic polymerase chain reaction. J Virol Methods. 2001;92(2):113–20.

80. Puig M, Jofre J, Lucena F, Allard A, Wadell G, Girones R. Detection of adenoviruses and enteroviruses in polluted waters by nested PCR amplification. Appl Environ Microbiol. 1994;60(8):2963–70.

81. Heath TA, Hedtke SM, Hillis DM. Taxon sampling and the accuracy of phylogenetic analyses. J Syst Evol. 2008;46(3):239–57.

82. Green MR, Sambrook J. Nested polymerase chain reaction (PCR). Cold Spring Harb Protoc. 2019;2019(2):175–9.

84. Korbie DJ, Mattick JS. Touchdown PCR for increased specificity and sensitivity in PCR amplification. Nat Protoc. 2008;3(9):1452–6.

85. Hrdy J, Vasickova P. Virus detection methods for different kinds of food and water samples – The importance of molecular techniques. Food Control. 2022 Apr;134:108764.

86. Monteiro S, Ebdon J, Santos R, Taylor H. Elucidation of fecal inputs into the River Tagus catchment (Portugal) using source-specific mitochondrial DNA, HAdV, and phage markers. Sci Total Environ. 2021;783:147086.

87. Fong TT, Phanikumar MS, Xagoraraki I, Rose JB. Quantitative Detection of Human Adenoviruses in Wastewater and Combined Sewer Overflows Influencing a Michigan River. Appl Environ Microbiol. 2010;76(3):715–23.

88. Amdiouni H, Faouzi A, Fariat N, Hassar M, Soukri A, Nourlil J. Detection and molecular identification of human adenoviruses and enteroviruses in wastewater from Morocco: Molecular identification of HAdV and EV. Lett Appl Microbiol. 2012;54(4):359–66.

89. Van Heerden J, Ehlers MM, Van Zyl WB, Grabow WOK. Incidence of adenoviruses in raw and treated water. Water Res. 2003;37(15):3704–8.

90. Kieleczawa J. Fundamentals of Sequencing of Difficult Templates—An Overview. 2006;17(3):11.

91. Kumakamba C, N'Kawa F, Kingebeni PM, Losoma JA, Lukusa IN, Muyembe F, et al. Analysis of adenovirus DNA detected in rodent species from the Democratic Republic of the Congo indicates potentially novel adenovirus types. New Microbes New Infect. 2020;34:100640.

92. Hemmi S, Vidovszky MZ, Ruminska J, Ramelli S, Decurtins W, Greber UF, et al. Genomic and phylogenetic analyses of murine adenovirus 2. Virus Res. 2011;160(1–2):128–35.

93. Diffo J, Ndze VN, Ntumvi NF, Takuo JM, Mouiche MMM, Tamoufe U, et al. DNA of diverse adenoviruses detected in Cameroonian rodent and shrew species. Arch Virol. 2019;164(9):2359–66.

94. Zheng X yan, Qiu M, Ke X mei, Guan W jie, Li J ming, Huo S ting, et al. Detection of novel adenoviruses in fecal specimens from rodents and shrews in southern China. Virus Genes. 2016;52(3):417–21.

95. Casas I, Avellon A, Mosquera M, Jabado O, Echevarria JE, Campos RH, et al. Molecular Identification of Adenoviruses in Clinical Samples by Analyzing a Partial Hexon Genomic Region. J Clin Microbiol. 2005;43(12):6176–82.

96. Ahmed W, Goonetilleke A, Gardner T. Human and bovine adenoviruses for the detection of source-specific fecal pollution in coastal waters in Australia. Water Res. 2010;44(16):4662–73.

97. Maunula L, Söderberg K, Vahtera H, Vuorilehto VP, von Bonsdorff CH, Valtari M, et al. Presence of human noro- and adenoviruses in river and treated wastewater, a longitudinal study and method comparison. J Water Health. 2012;10(1):87–99.

6. Appendix

6.1. Appendix A – DNA concentration and purity ratios

Table A1. DNA concentration and purity ratios of DNA extracted from the VLP concentrates of environmental samples collected in 2021 in LMA, using Nanodrop® 1,000 (ThermoFisher Scientific, USA). The 260/280 ratio of 1.8 is generally accepted as "pure" for DNA and the ratio 260/230 of 2.0–2.2 is often considered "pure" for nucleic acid.

Sample	ng/µL	260/280	260/230
Alhos Vedros Ditch	112.8	2.03	1.40
Cascais Ditch	4.4	1.24	0.23
Corroios Ditch	18.4	1.34	0.46
Fogueteiro Ditch	26.7	1.60	0.71
Ribeira das Vinhas	3.5	1.02	0.20
Ribeira de Caparide	66.1	1.92	0.74
Sobreda Ditch	25.9	1.54	0.92

"ng" stands for "nanograms" and "µL" for "microliters".

Table A2. DNA concentration and purity ratios of plasmid DNA extracted from the saturated bacterial culture, using Nanodrop® 1,000 (ThermoFisher Scientific, USA). A total of 116 recombinant plasmids were obtained and quantified, but here it is only shown the obtained values for 60 of these plasmids, the ones chosen to be sequenced by the Sanger method.

Fragment	Sample	Colonie	ng/µL	260/280	260/230
A or F	Alhos Vedros Ditch	1	80.15	1.87	1.68
		4	129.33	1.88	1.85
		7	338.89	1.91	2.07
	Fogueteiro Ditch	3	170.35	1.87	2.04
		8	172.86	1.91	2.05
	Lizandro River	4	230.16	1.89	2.09
		6	292.40	1.88	2.10
		8	211.38	1.89	2.00
	Ribeira de Caparide	9	327.35	1.87	2.14
	Sado River (Industrial Area of Setúbal)	1	436.65	1.86	2.17
	Trancão River	2	191.41	1.90	2.06
	WWTP A1	1	171.01	1.90	1.97
		4	52.94	1.84	1.42
		5	94.68	1.89	1.74
		6	105.40	1.88	1.82
		7	72.36	1.88	1.63
		8	64.92	1.90	1.55
		11	86.49	1.89	1.72
	WWTP B2	2	318.61	1.87	2.11
		4	342.16	1.87	2.12
		5	387.47	1.87	2.13
		10	208.56	1.90	2.03
	WWTP C	3	400.20	1.87	2.16
		5	370.24	1.88	2.16
		7	311.82	1.87	2.13
	WWTP E	6	361.09	1.87	2.15
		9	398.33	1.88	2.15
	WWTP F	1	49.44	1.82	1.43
		7	530.29	1.87	2.12
B or E	Corroios Ditch	3	355.63	1.89	2.15
		6	359.76	1.88	2.14
	WWTP A1	1	257.14	1.89	2.10
		2	481.53	1.85	2.16
		4	541.82	1.84	2.16
		5	300.44	1.88	2.12
		7	527.45	1.84	2.15
		9	389.10	1.86	2.15

	WWTP F	2	376.62	1.87	2.28
		5	387.30	1.88	2.28
		7	279.44	1.89	2.29
D	Ribeira do Carenque	2	137.54	1.88	1.94
		7	47.13	1.88	1.44
		11	449.72	1.87	2.16
	WWTP A1	1	283.00	1.89	2.30
		6	222.37	1.89	2.34
		10	369.87	1.87	2.35
	WWTP C	3	262.44	1.92	2.11
		6	209.60	1.93	2.03
		9	309.03	1.92	2.13
		10	240.23	1.98	2.10
		11	254.17	1.92	2.00
		12	303.45	1.92	2.13
		13	286.50	1.92	2.08
		14	296.27	1.90	2.14
	WWTP E	3	407.52	1.87	2.17
		10	318.66	1.87	2.14
	WWTP F	2	279.50	1.87	2.11
		3	332.05	1.87	2.16
		4	83.53	1.95	1.73
		8	224.85	1.90	1.97
		10	186.05	1.92	2.00

“WWTP” stands for “wastewater treatment plant sample”, “ng” for “nanograms” and “µL” for “microliters”.

Table A3. DNA concentration obtained with Qubit® Fluorometer (Invitrogen, USA) of PCR purified fragments used for fragmentation and NGS analysis.

Fragment	Sample	Qubit® quantification (ng/µL)
A or F	Alhos Vedros Ditch	380
	Trancão River	73.4
	WWTP A1	332
	WWTP B2	204
	WWTP C	134
	WWTP E	402
	WWTP F	380
B or E	Ribeira das Vinhas	89.4
	Ribeira de Caparide	95.8
	Corroios Ditch	68.6
	WWTP A1	72.6
	WWTP B1	206
	WWTP C	208
	WWTP D	230
	WWTP E	208
	WWTP F	147
D	Ribeira do Carenque	122
	WWTP A1	318
	WWTP E	236
	WWTP F	199

"WWTP" stands for "wastewater treatment plant sample", "ng" for "nanograms" and "µL" for "microliters".

Table A4. DNA quantification and fragments' size of 11 randomly selected DNA fragmented samples using the Agilent 2100 Bioanalyzer equipment (Agilent Technologies, USA).

Fragment	Sample	Fragments' size (bp)	DNA quantification (ng/µL)	DNA mass (ng) in 50 µL of fragmented DNA
A or F	Trancão River	128	2.0	102.4
	WWTP A1	172	5.7	286.3
	WWTP C	148	5.8	291.5
	WWTP E	193	5.3	268.2
B or E	Ribeira de Caparide	n.d.	n.d.	n.d.
	WWTP B2	154	5.9	297.0
	WWTP C	131	6.2	311.5
	WWTP F	170	6.9	348.5
D	Ribeira do Carenque	192	2.9	148.0
	WWTP A1	185	2.9	147.5
	WWTP E	127	8.3	419.75

"n.d." stands for "non-determined" (this sample was contaminated with the ladder), "bp" for "base pairs", "ng" for "nanograms", "µL" for "microliters", and "WWTP" for "wastewater treatment plant sample".

Table A5. DNA libraries quantification using Agilent 2100 Bioanalyzer equipment (Agilent Technologies, USA), QuantStudio™ 5 (Thermo Fischer Scientific, USA), and KAPA Library Quantification template (Illumina®, USA).

Library no.	Sample name	Dilution	Size (bp)	Cq	The concentration of the undiluted library (nM)
1[a]	WWTP A1 (A+F)	10,000	295	10.54	61.24
				10.13	
		100,000		13.96	
				13.72	
2[a]	WWTP B2 (A+F)	10,000	260	11.62	38.42
				10.91	
		100,000		15.18	
				14.27	
3[a]	WWTP C (A+F)	10,000	270	11.45	43.78
				10.88	
		100,000		14.61	
				14.05	
4[a]	WWTP E (A+F)	10,000	280	10.57	73.48
				9.83	
		100,000		13.66	
				13.51	
5[a]	WWTP F (A+F)	10,000	280	11.01	60.28
				9.85	
		100,000		14.51	
				13.39	
6[a]	Trancão River (A+F)	10,000	280	10.54	48.34
		100,000		14.98	
				14.14	
7[a]	Alhos Vedros Ditch (A+F)	10,000	280	10.09	93.05
				9.19	
		100,000		14.05	
				12.81	
8[a]	WWTP A1 (B+E)	10,000	280	10.76	73.37
				9.35	
		100,000		14.38	
				13.08	
9[a]	WWTP F (B+E)	10,000	280	11.47	41.24
				10.22	
		100,000		14.73	
10	WWTP B1 (B+E)	10,000	280	10.28	52.56
		100,000		14.64	
11		10,000	280	14.08	5.86

	WWTP C (B+E)	100,000		17.45	
12	WWTP D (B+E)	10,000	320	11.43	25.68
		100,000		15.16	
13	WWTP E (B+E)	10,000	280	11.47	27.51
		100,000		15.33	
14	Ribeira das Vinhas (B+E)	10,000	278	15.60	1.82
		100,000		19.55	
15	Ribeira de Caparide (B+E)	10,000	275	11.67	28.13
		100,000		15.1025.76	
16	Corroios (B+E)	10,000	260	12.03	25.76
		100,000		15.21	
17	WWTP A1 (D)	10,000	274	13.43	11.33
		100,000		16.23	
18	WWTP E (D)	10,000	295	11.08	29.41
		100,000		15.46	
19	WWTP F (D)	10,000	284	11.02	36.25
		100,000		14.90	
20	Ribeira do Carenque (D)	10,000	330	13.51	8.26
		100,000		16.50	

[a] Libraries for which was possible to include duplicates of the dilutions.
“WWTP” stands for “wastewater treatment plant sample”, “bp” for “base pairs”, “Cq” for “quantification cycle” and “nM” for “nanomolar”.

Table A6. Pooling calculator results to create a 2 nM pool of 50 µL. According to the concentration of the undiluted library obtained, is shown each volume to use from the 30 µL library, the volume of 10 mM Tris-HCl to add to dilute each library, and the volume of each diluted library to use to create a 2 nM pool of 50 µL.

Library no.	Sample name (expected species)	Library volume (µL)	10 mM Tris-HCl to add (µL)	Pooling volume (µL)
1	WWTP A1 (A+F)	2	56.8	2.4
2	WWTP B2 (A+F)	2	34.9	2.4
3	WWTP C (A+F)	2	41.8	2.5
4	WWTP E (A+F)	2	71.5	2.5
5	WWTP F (A+F)	2	58.3	2.5
6	Trancão River (A+F)	2	46.3	2.5
7	Alhos Vedros Ditch (A+F)	2	91.1	2.5
8	WWTP A1 (B+E)	2	71.4	2.5
9	WWTP F (B+E)	2	39.2	2.5
10	WWTP B1 (B+E)	2	50.6	2.5
11	WWTP C (B+E)	2	3.9	2.5
12	WWTP D (B+E)	2	23.7	2.5
13	WWTP E (B+E)	2	25.5	2.5
14	Ribeira das Vinhas (B+E)	undiluted	0	2.7
15	Ribeira de Caparide (B+E)	2	26.1	2.5
16	Corroios (B+E)	2	23.8	2.5
17	WWTP A1 (D)	2	9.3	2.5
18	WWTP E (D)	2	27.4	2.5
19	WWTP F (D)	2	34.3	2.5
20	Ribeira do Carenque (D)	2	6.3	2.5

“WWTP” stands for “wastewater treatment plant sample”, and “µL” for “microliters”.

Table A7. Number of reads that passed the internal quality filtering Illumina procedure for each sample sequenced by NGS and corresponding contigs obtained after OmicsBox Bioinformatics Software edition.

Sample	Number of PF reads	Number of contigs after OmicsBox edition	Number of contigs after filtering (hexon gene, >350 bp)
WWTP A1 (A+F)	167,956	326	10
WWTP B2 (A+F)	156,004	407	8
WWTP C (A+F)	125,045	239	7
WWTP E (A+F)	160,415	293	9
WWTP F (A+F)	156,809	322	9
Trancão River (A+F)	121,012	503	5
Alhos Vedros Ditch (A+F)	82,485	170	6
WWTP A1 (B+E)	134,370	386	3
WWTP F (B+E)	159,324	1555	12
WWTP B1 (B+E)	30,757	463	3
WWTP C (B+E)	114,557	423	11
WWTP D (B+E)	145,625	542	5
WWTP E (B+E)	171,896	685	7
Ribeira das Vinhas (B+E)	146,018	2584	4
Ribeira de Caparide (B+E)	189,575	1596	2
Corroios (B+E)	99,011	311	3
WWTP A1 (D)	200,250	628	9
WWTP E (D)	173,961	151	6
WWTP F (D)	227,571	443	6
Ribeira do Carenque (D)	558	15	2

"WWTP" stands for "wastewater treatment plant sample", "PF" for "pass-filter".

6.2. Appendix B – Phylogenetic analysis

Table B1. The appropriate evolution model for each dataset using the Model Selection feature of IQ TREE Web Server for phylogenetic tree construction. From the three available criteria, the Akaike Information Criterion was chosen.

Sequencing method	Dataset	Corresponding phylogenetic tree	Evolution model
Sanger	HAdV hexon and 33 Sanger sequences	Figure 11	TIM2+F+I+G4
	Non-human AdV and five Sanger sequences	Figure B1	GTR+F+I+G4
	HAdV-A and -F species and 23 Sanger sequences	Figure 12	TIM2e+I+G4
	HAdV-A and -F species and one Sanger sequence	Figure B2	GTR+F+I+G4
	HAdV-C species and six Sanger sequences	Figure 13	TIM2+F+I
	HAdV-D species and four Sanger sequences	Figure B3	GTR+F+I+G4
	HAdV-D species and two Sanger sequences	Figure 14	GTR+F+I+G4
NGS	HAdV-A and -F species and two NGS sequences	Figure 16	GTR+F+I+G4
	HAdV-C species and two NGS sequences	Figure 17	TIM2+F+I
	HAdV-D species and seven NGS sequences	Figure 18	GRT+F+I+G4
	HAdV-D species and one NGS sequence	Figure B4	GTR+F+I+G4
	Non-human AdV and 30 NGS sequences	Figure B5	SYM+I+G4
	Non-human AdV and one NGS sequence	Figure B6	GTR+F+I+G4
	Non-human AdV and two NGS sequences	Figure B7	GTR+F+I+G4
	Non-human AdV and one NGS sequence	Figure B8	GTR+F+I+G4
	HAdV-D species and three NGS sequences	Figure B9	GTR+F+I+G4

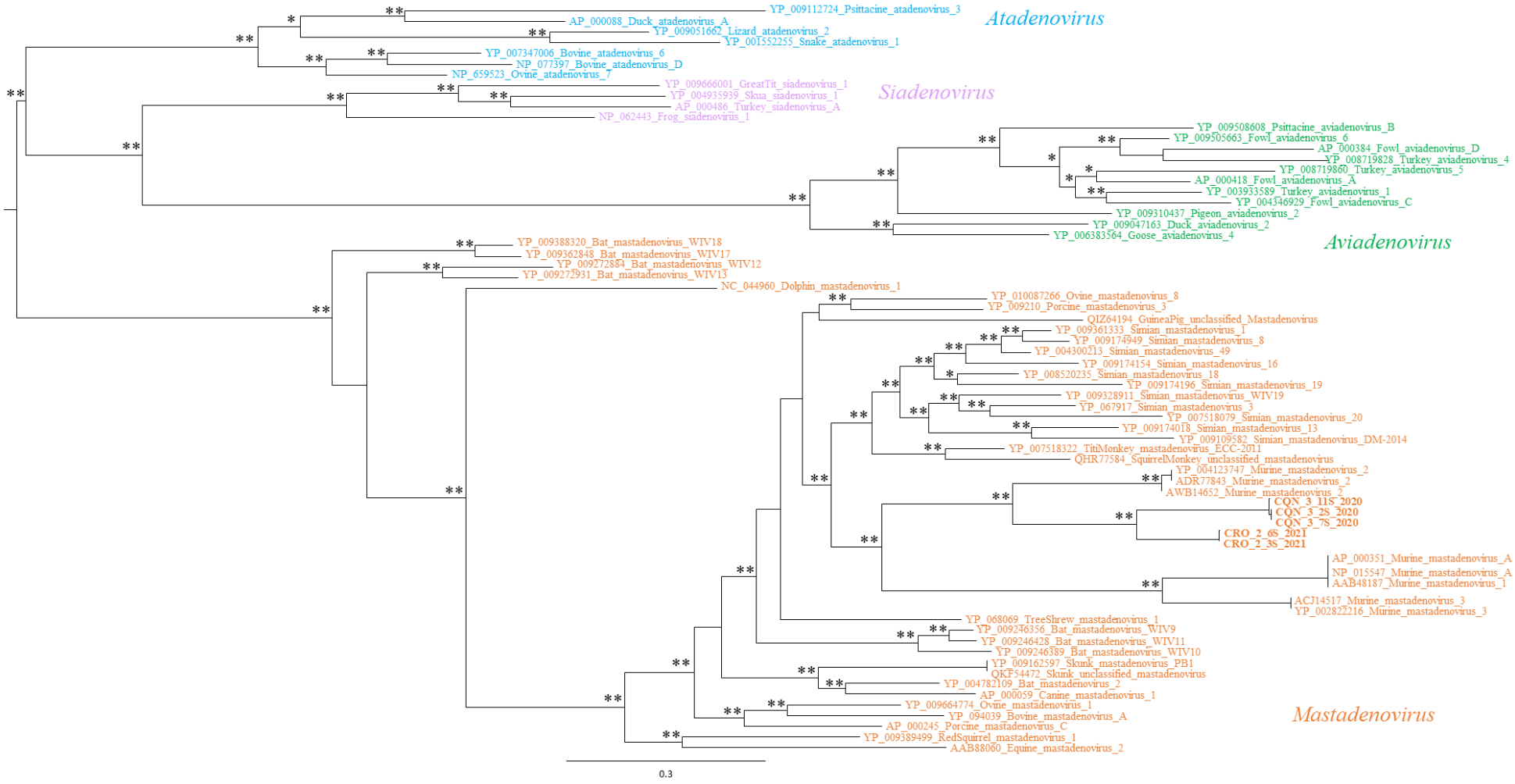

Figure B1. Phylogenetic tree by Maximum Likelihood of non-human AdV hexon capsid gene sequences, including five Portuguese sequences obtained by the Sanger method. Each sequence is identified by its accession number and AdV species, except the ones obtained in this study. The number of "*" indicates the number of methods that support the demonstrated topology, considering 75% (of the total number of data resamplings) and above as relevant for aLRT and bootstrap values. Each AdV genus is indicated in a different colour: *Atadenovirus* – blue, *Siadenovirus* – purple, *Aviadenovirus* – green, and *Mastadenovirus* – orange.

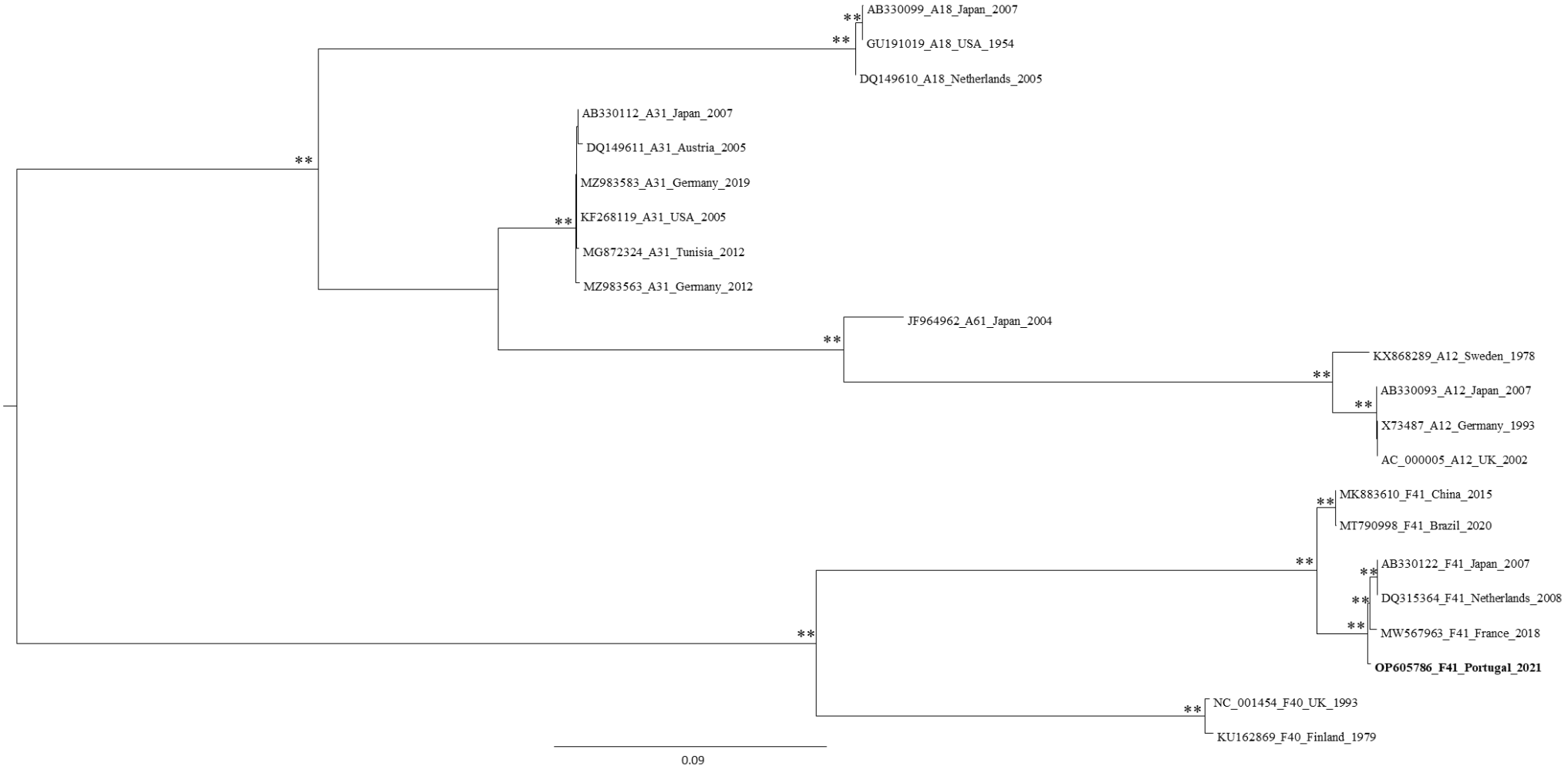

Figure B2. Phylogenetic tree by Maximum Likelihood of HAdV-A and -F hexon capsid gene sequences, including one Portuguese sequence obtained by the Sanger method. Each sequence is identified by its accession number, HAdV type, country of origin and year. The number of "*" indicates the number of methods that support the demonstrated topology, considering 75% (of the total number of data resamplings) and above as relevant for aLRT and bootstrap values

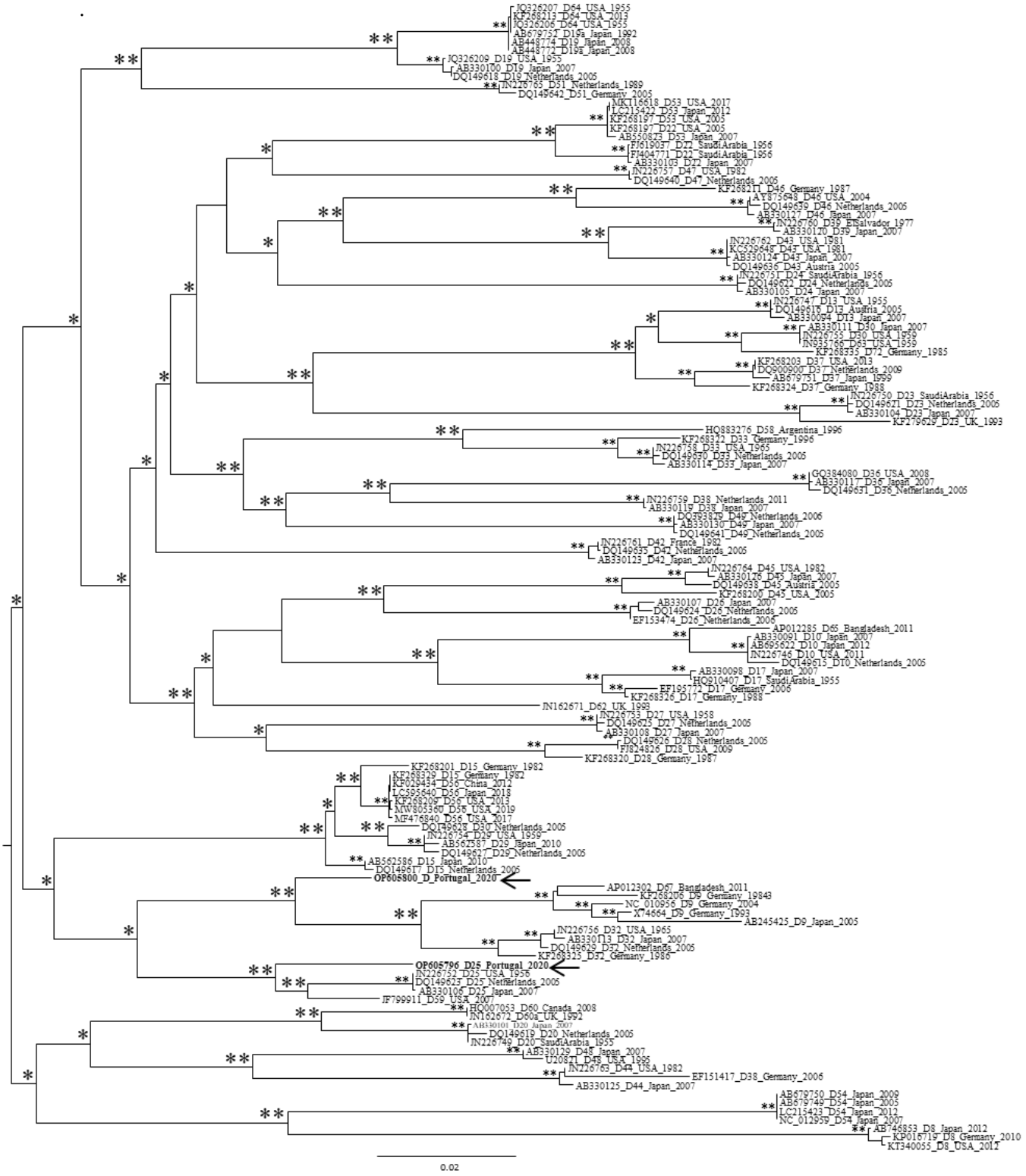

Figure B3. Phylogenetic tree by Maximum Likelihood of HAdV-D hexon capsid gene sequences, including two Portuguese sequences obtained by the Sanger method. Each sequence is identified by its accession number, HAdV type, country of origin and year. The number of "*" indicates the number of methods that support the demonstrated topology, considering 75% (of the total number of data resamplings) and above as relevant for aLRT and bootstrap values. The arrows indicate the Portuguese sequences.

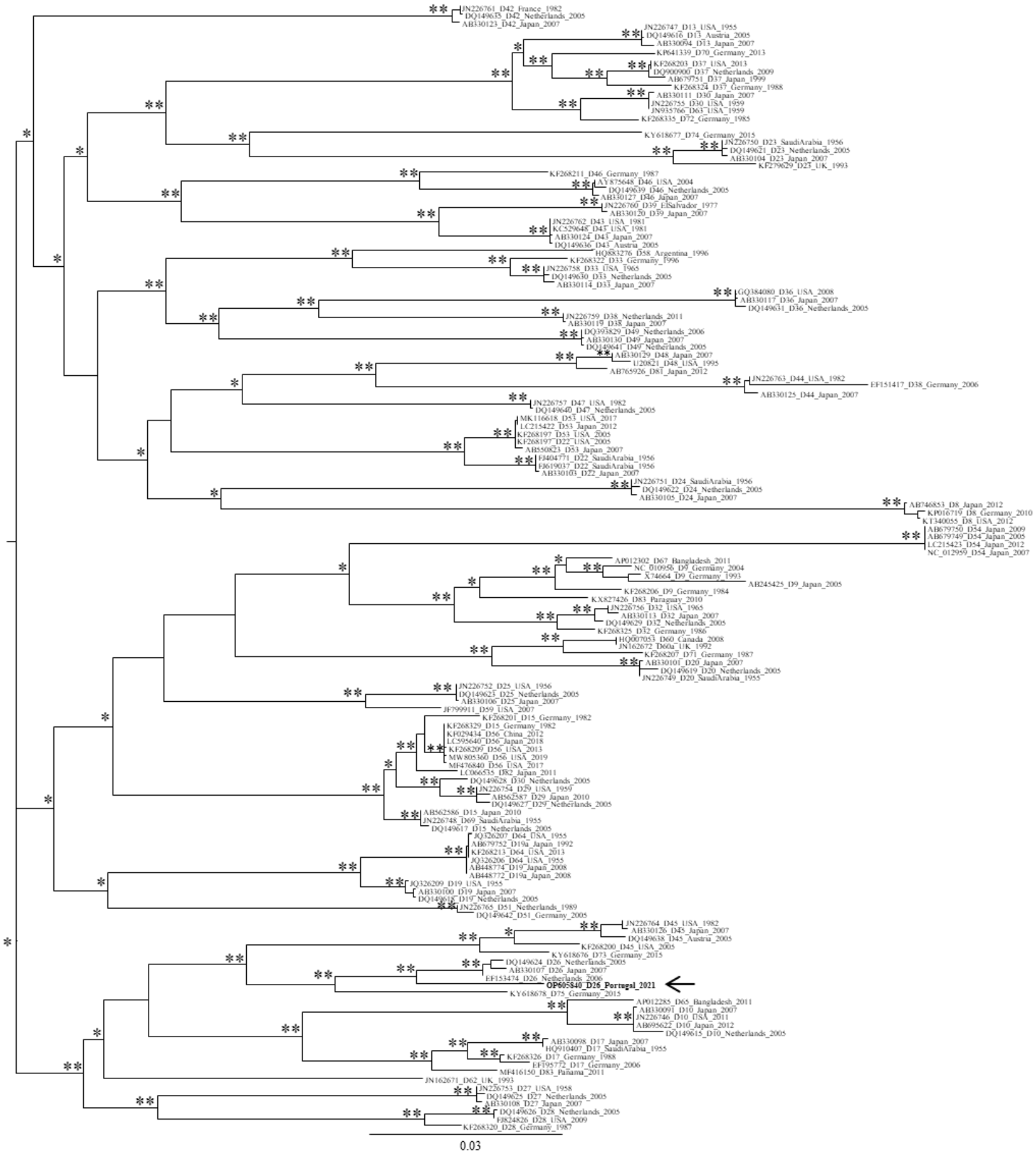

Figure B4. Phylogenetic tree by Maximum Likelihood of HAdV-D hexon capsid gene sequences, including one Portuguese sequence obtained by NGS. Each sequence is identified by its accession number, HAdV type, country of origin and year. The number of "*" indicates the number of methods that support the demonstrated topology, considering 75% (of the total number of data resamplings) and above as relevant for aLRT and bootstrap values. The arrow indicates the Portuguese sequence.

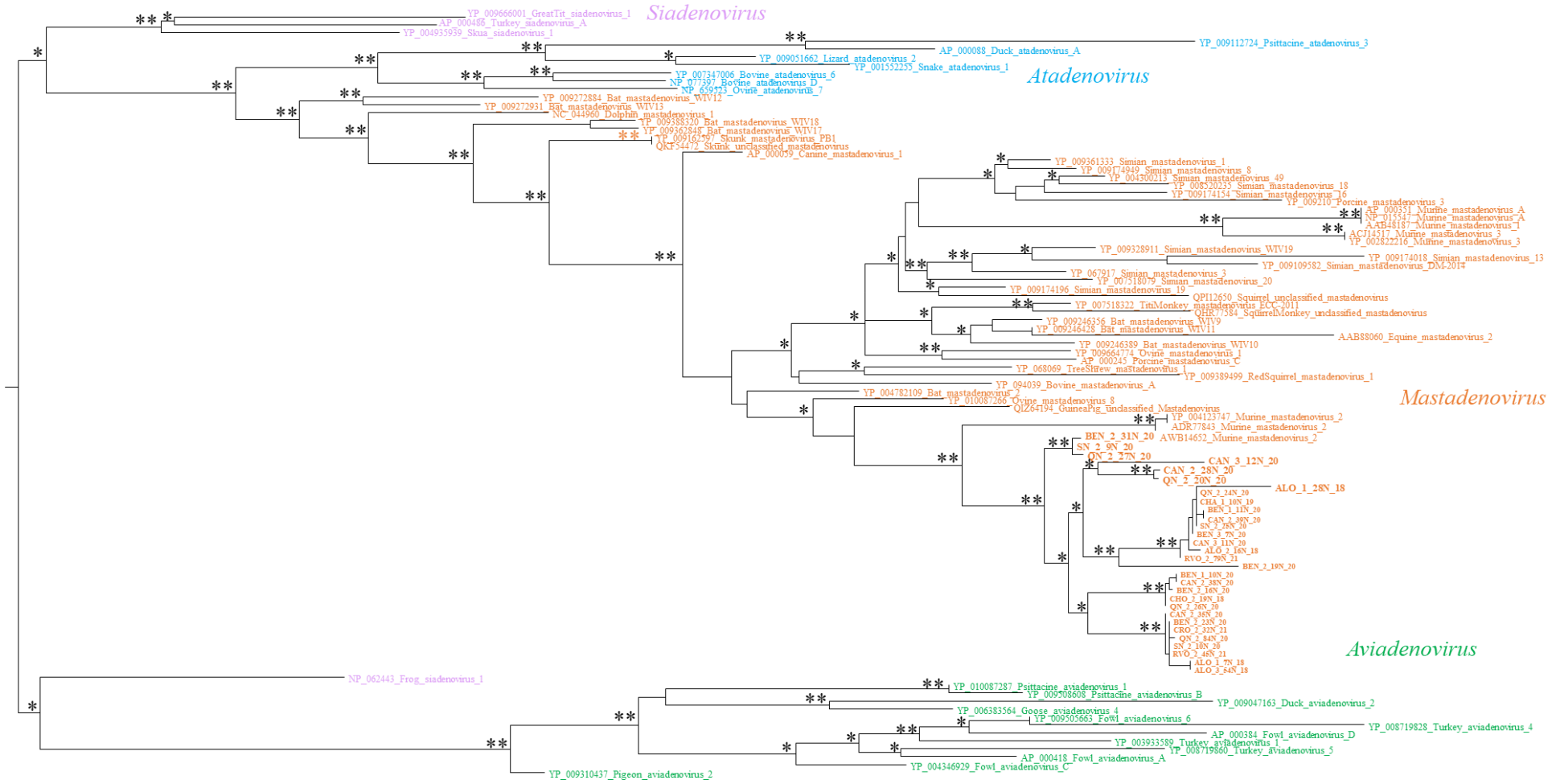

Figure B5. Phylogenetic tree by Maximum Likelihood of non-human AdV hexon capsid gene sequences, including 30 Portuguese sequences obtained by NGS. Each sequence is identified by its accession number and AdV species, except the ones obtained in this study. The number of "*" indicates the number of methods that support the demonstrated topology, considering 75% (of the total number of data resamplings) and above as relevant for aLRT and bootstrap values. Each AdV genus is indicated in a different colour: *Atadenovirus* – blue, *Siadenovirus* – purple, *Aviadenovirus* – green, and *Mastadenovirus* – orange.

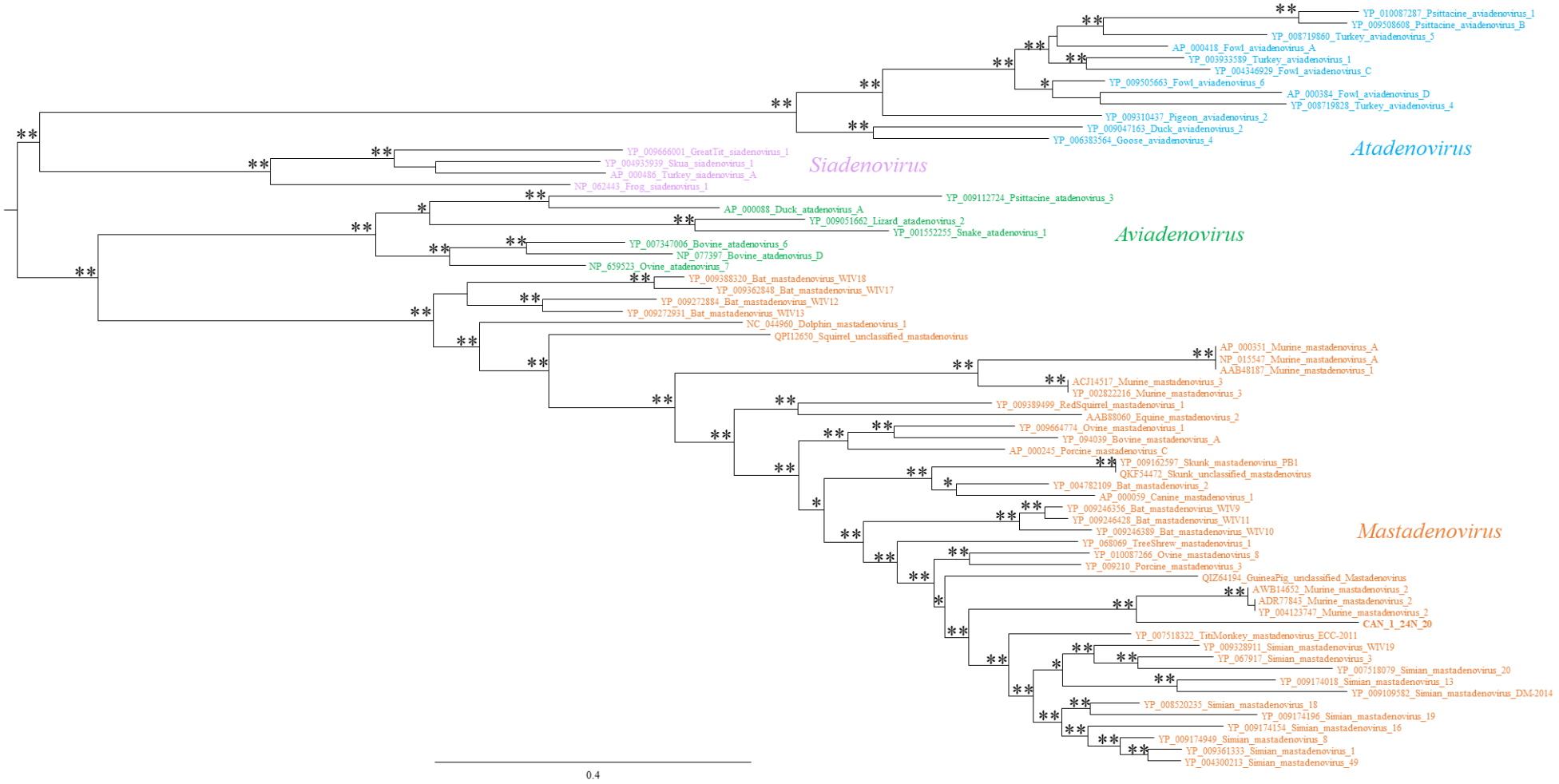

Figure B6. Phylogenetic tree by Maximum Likelihood of non-human AdV hexon capsid gene sequences, including one Portuguese sequence obtained by NGS. Each sequence is identified by its accession number and AdV species, except the one obtained in this study. The number of "*" indicates the number of methods that support the demonstrated topology, considering 75% (of the total number of data resamplings) and above as relevant for aLRT and bootstrap values. Each AdV genus is indicated in a different colour: *Atadenovirus* – blue, *Siadenovirus* – purple, *Aviadenovirus* – green, and *Mastadenovirus* – orange.

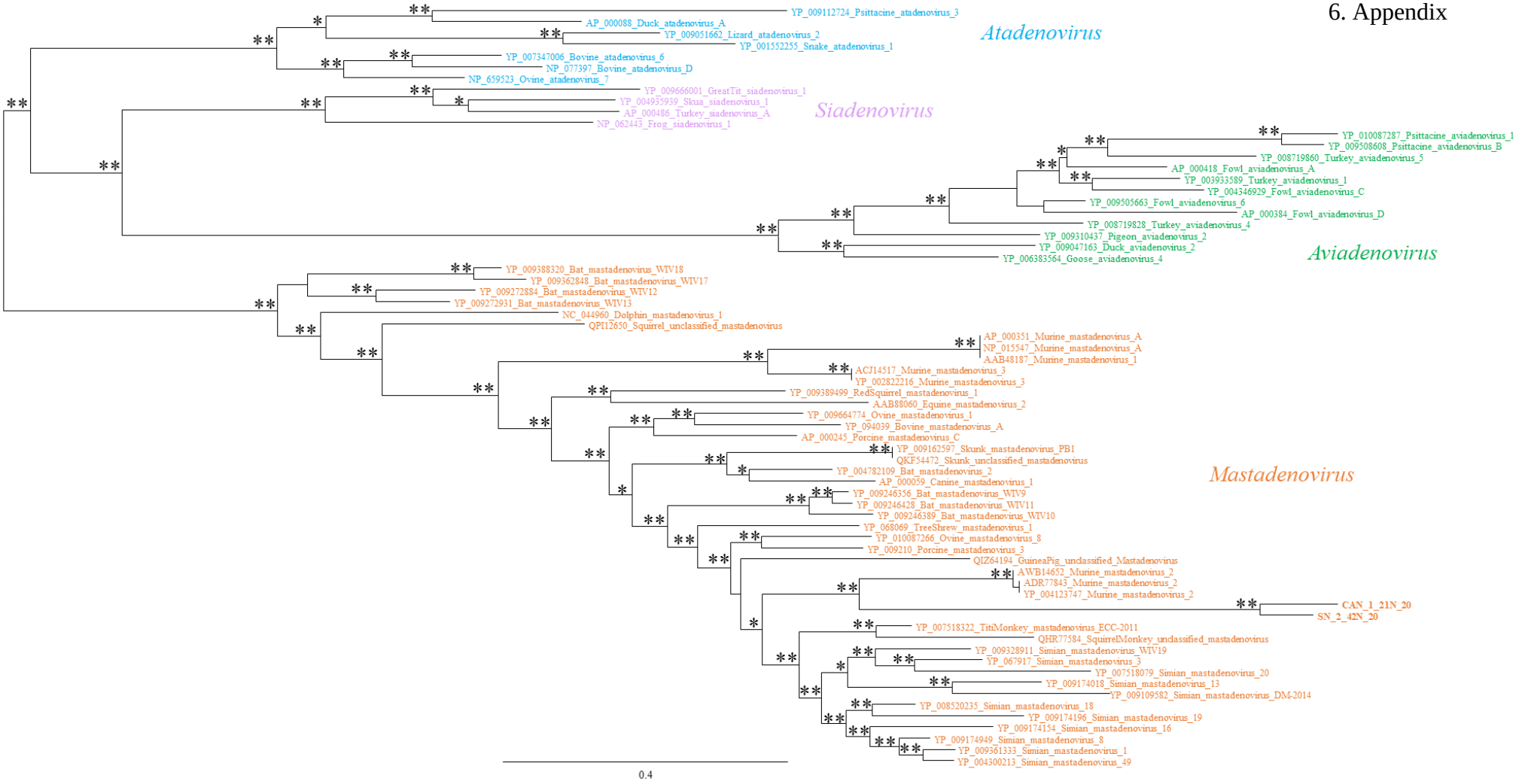

Figure B7. Phylogenetic tree by Maximum Likelihood of non-human AdV hexon capsid gene sequences, including two Portuguese sequences obtained by NGS. Each sequence is identified by its accession number and AdV species, except the ones obtained in this study. The number of "*" indicates the number of methods that support the demonstrated topology, considering 75% (of the total number of data resamplings) and above as relevant for aLRT and bootstrap values. Each AdV genus is indicated in a different colour: *Atadenovirus* – blue, *Siadenovirus* – purple, *Aviadenovirus* – green, and *Mastadenovirus* – orange.

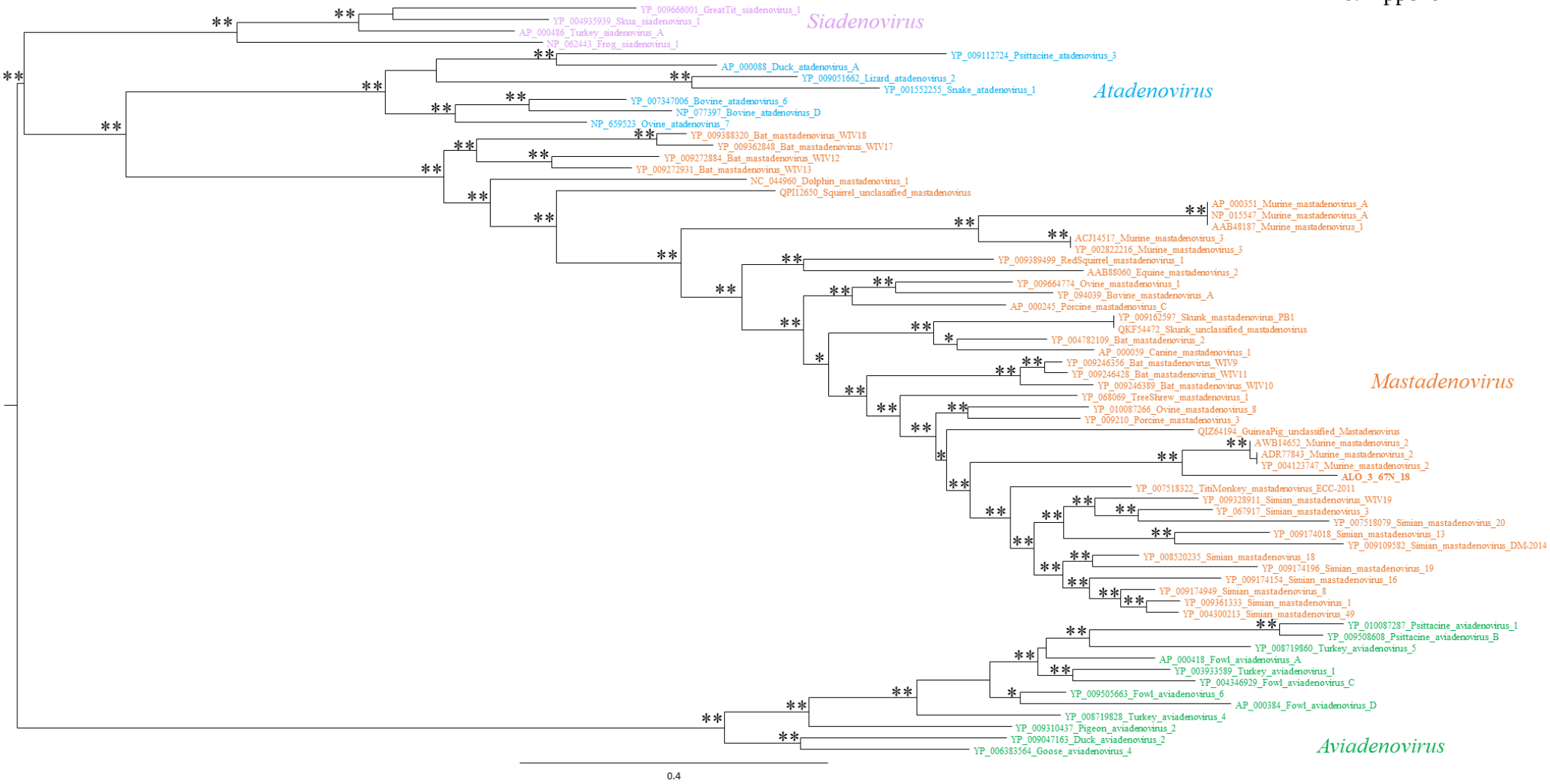

Figure B8. Phylogenetic tree by Maximum Likelihood of non-human AdV hexon capsid gene sequences, including one Portuguese sequence obtained by NGS. Each sequence is identified by its accession number and AdV species, except the one obtained in this study. The number of "*" indicates the number of methods that support the demonstrated topology, considering 75% (of the total number of data resamplings) and above as relevant for aLRT and bootstrap values. Each AdV genus is indicated in a different colour: *Atadenovirus* – blue, *Siadenovirus* – purple, *Aviadenovirus* – green, and *Mastadenovirus* – orange.

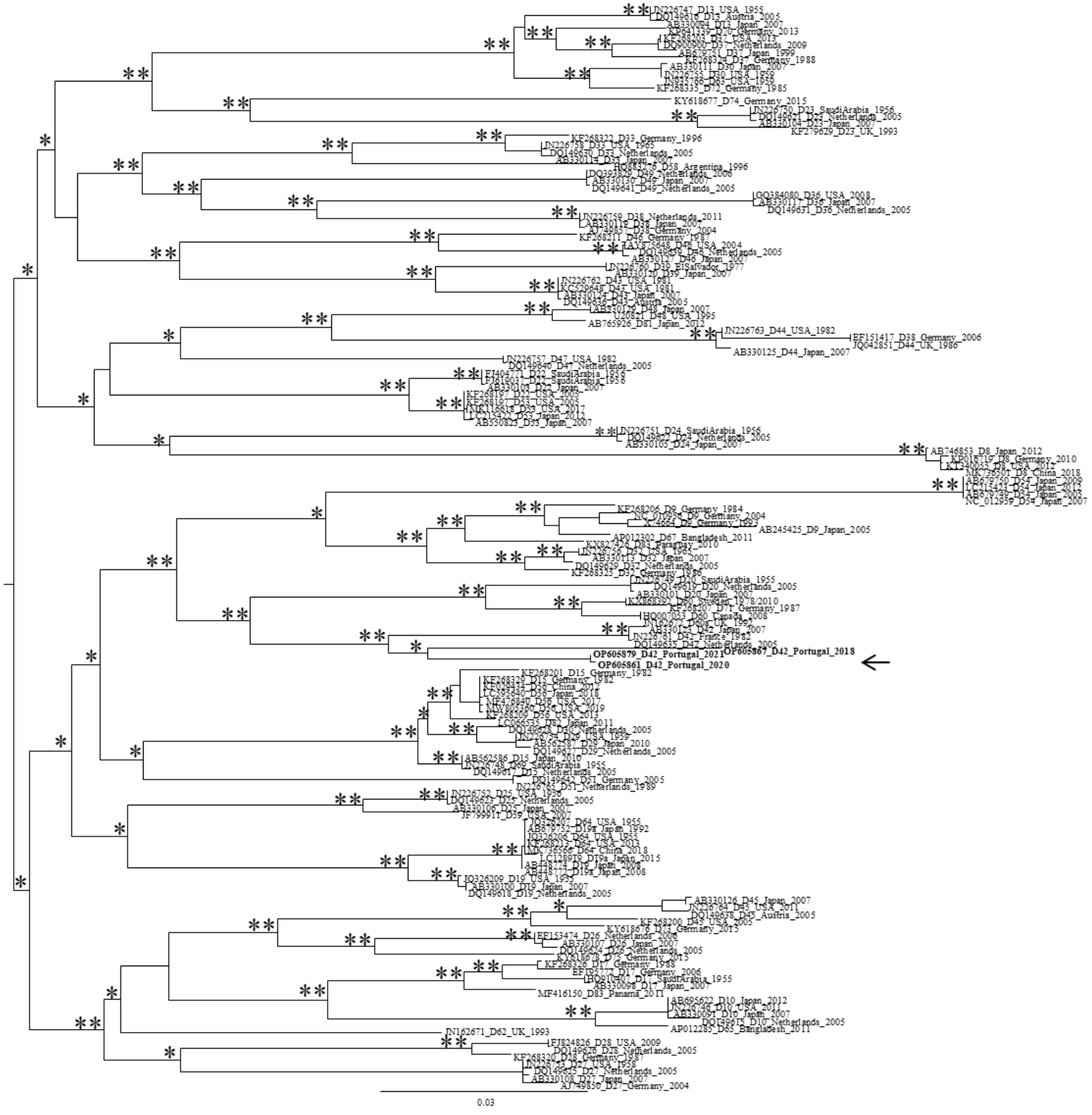

Figure B9. Phylogenetic tree by Maximum Likelihood of HAdV-D hexon capsid gene sequences, including three Portuguese sequences obtained by NGS. Each sequence is identified by its accession number, HAdV type, country of origin and year. The number of "*" indicates the number of methods that support the demonstrated topology, considering 75% (of the total number of data resamplings) and above as relevant for aLRT and bootstrap values. The arrow indicates the Portuguese sequences.

2022	Joana Filipa Nunes Cavadas	Evaluation of Human *mastadenovirus* molecular diversity in wastewater and environmental water from the Lisbon Metropolitan Area