Viral origin of eukaryotic type IIA DNA topoisomerases

Julien Guglielmini[1], Morgan Gaia[2], Violette Da Cunha[3], Alexis Criscuolo[1], Mart Krupovic[4] and Patrick Forterre[3,4]

[1]Institut Pasteur, Université de Paris, Bioinformatics and Biostatistics Hub, F-75015, Paris, France.
[2]Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, 91000 Evry, France.
[3]Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), 91198 Gif-sur-Yvette, France.
[4]Archaeal Virology Unit, Department of Microbiology, Institut Pasteur, 75015 Paris, France.

Corresponding Author: patrick.forterre@pasteur.fr

Abstract

Type II DNA topoisomerases of the family A (Topo IIA) are present in all bacteria (DNA gyrase) and eukaryotes. In eukaryotes, they play a major role in transcription, DNA replication, chromosome segregation and modulation of chromosome architecture. The origin of eukaryotic Topo IIA remains mysterious since they are very divergent from their bacterial homologues and have no orthologues in Archaea. Interestingly, eukaryotic Topo IIA have close homologues in viruses of the phylum *Nucleocytoviricota*, an expansive assemblage of large and giant viruses formerly known as the nucleocytoplasmic large DNA viruses (NCLDV). Topo IIA are also encoded by some bacterioviruses of the class *Caudoviricetes* (tailed bacteriophages). To elucidate the origin of the eukaryotic Topo IIA, we performed in-depth phylogenetic analyses combining viral and cellular Topo IIA homologs. Topo IIA encoded by bacteria and eukaryotes form two monophyletic groups nested within Topo IIA encoded by *Caudoviricetes* and *Nucleocytoviricota*, respectively. Importantly, *Nucleocytoviricota* remained well separated from eukaryotes after removing both bacteria and *Caudoviricetes* from the dataset, indicating that the separation of *Nucleocytoviricota* and eukaryotes is probably not due to long branch attraction artefact. The topology of our tree suggests that the eukaryotic Topo IIA was probably acquired from an ancestral member of the *Nucleocytoviricota* of the class *Megaviricetes*, before the emergence of the last eukaryotic common ancestor (LECA). This result further highlights a key role of these viruses in eukaryogenesis and suggests that early proto-eukaryotes used a Topo IIB instead of a Topo IIA for solving their DNA topological problems.

Introduction

DNA topoisomerases are ubiquitous enzymes that are essential for solving topological problems inherent to the helical structure of DNA (Champoux 2001; Forterre and Gadelle 2009; Wang 2009; Vos et al. 2011; Forterre 2012; Pommier et al. 2016). Based on mechanistic properties, they have been classified into types I and II. Type I DNA topoisomerases (Topo I) produce transient single-strand breaks in double-stranded DNA (dsDNA) and catalyze the transfer of one DNA strand through this break. In contrast, type II DNA topoisomerases (Topo II) produce transient double-strand breaks and catalyze the transfer of a dsDNA segment (either from the same or different dsDNA molecule) through this break. Five different families of DNA topoisomerases have been defined based on amino-acid sequences and structural similarities: three families of Topo I (Topo IA, Topo IB and Topo IC)(Champoux 2001; Forterre 2006), and two families of Topo II (Topo IIA and Topo IIB) (Bergerat et al. 1997; Gadelle et al. 2003). All Topo II and some Topo I (IB and IC) can relax positive and negative superturns that otherwise would accumulate in front and behind the replication forks and transcription bubbles, respectively. In addition, Topo II can eliminate the catenanes that can accumulate at the end of the chromosome replication. In eukaryotes, Topo IIA are also intrinsic structural components of the chromosomal scaffold (Hizume et al. 2007) and play a major role in modulating chromosome architecture and long-range chromatin structure (Nitiss 2009; Nielsen et al. 2020).

DNA topoisomerases have been extensively studied because they are the targets of important antibiotics and antitumor drugs (Pommier 2013). These drugs interfere with the breakage-reunion mechanisms of the enzyme and transform the transient intermediate topoisomerase-DNA covalent complexes into stable poisons, interfering with replication and transcription. However, these enzymes are also very interesting (and intriguing) in terms of the history of life on our planet. Indeed, the distribution patterns of the different DNA topoisomerase families within the three domains of life, Archaea, Bacteria and Eukarya (eukaryotes), do not fit the usual distribution of informational proteins, such as ribosomal proteins or DNA-dependent RNA polymerases (Da Cunha et al. 2017), raising challenging questions and prompting unorthodox hypotheses (Forterre and Gadelle 2009). Whereas informational proteins from eukaryotes usually much more closely resemble their archaeal homologs than their bacterial ones, the universal eukaryotic Topo II (member of the Topo IIA family) has no obvious orthologue in Archaea. All archaea (except for certain Thermoplasmatales) contain an enzyme of the Topo IIB family, dubbed DNA topoisomerase VI (Topo VI), suggesting that the Last

Archaeal Common Ancestor (LACA) contained no Topo IIA but a Topo IIB for relaxation of positive superturns and chromosome decatenation (Forterre and Gadelle 2009). All bacteria encode a unique Topo IIA, DNA gyrase, which is a distant homologue of the eukaryotic enzyme. DNA gyrases are heterotetramers composed of two subunits (GyrA and GyrB) that are homologous to the C-terminal and N-terminal moieties of the homodimeric Topo IIA of eukaryotes, respectively (Fig. 1). Some Archaea encode a two-subunit DNA gyrase very similar to the bacterial enzyme and highly divergent from the eukaryotic Topo IIA. Phylogenetic analysis has indicated that these DNA gyrases were recruited from Bacteria by lateral gene transfer (Forterre et al. 2014). Similarly, some eukaryotes, such as *Archaeplastida*, encode a bacterial-like DNA gyrase (Topo IIA) present in chloroplasts and mitochondria that was most likely acquired from *Cyanobacteria* during the endosymbiotic event that led to the emergence of the chloroplasts (Wall et al. 2004). These eukaryotic Topo II are very similar to their bacterial counterparts and in phylogenetic analyses are nested within the clade of bacterial gyrases (Forterre et al. 2007). It seems unlikely that the very divergent eukaryotic Topo IIA originated through a similar endosymbiotic pathway. The origin of the Topo IIA in eukaryotes thus remains enigmatic.

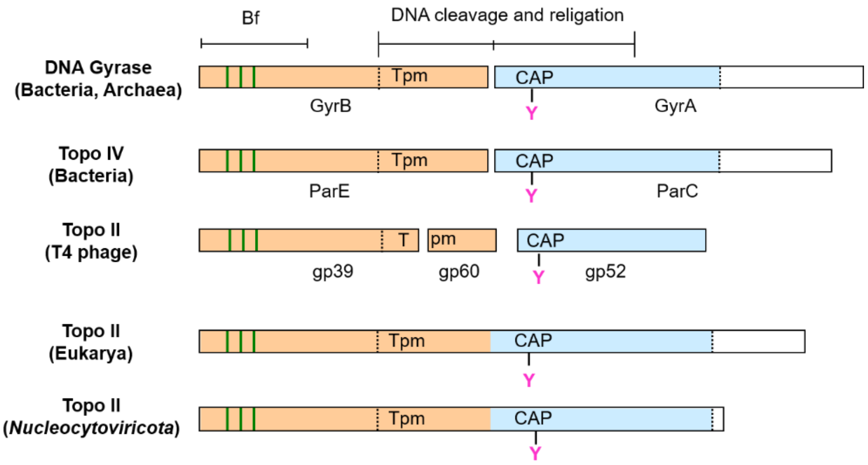

Fig. 1. Schematic representation of the domain composition of Type II DNA topoisomerases of the family A.
The different domains correspond to the Bergerat fold/ GHKL (Bf), TOPRIM (Tpm), the 5Y-CAP or winged helix (WHD) domain containing the catalytic Tyrosine.

A possible answer to this enigma could reside in the virosphere. The first viral Topo IIA was discovered in 1980 in the T4 bacteriovirus (Liu et al. 1979), the iconic virus of the class *Caudoviricetes* from the recently proposed family *Straboviridae*. Surprisingly, the T4 Topo

IIA, a heterotrimer (Fig. 1), is not specifically related to bacterial Topo IIA, but branched between DNA gyrases and eukaryal Topo IIA in phylogenetic trees (Gadelle et al. 2003). Later on, Topo IIA genes were discovered in several members of the phylum *Nucleocytoviricota* (Fig. 1), formerly known as nucleocytoplasmic large DNA viruses (NCLDV)(Lavrukhin et al. 2000; Gadelle et al. 2003; Coelho et al. 2015; Coelho et al. 2016; Erives 2017). Inhibition of this Topo IIA disrupts replication of the African swine fever virus (ASFV; family *Asfarviridae*) *in vitro* (Freitas et al. 2016), indicating that compounds active against the ASFV-Topo IIA, such as fluoroquinolones, are promising drugs against the highly contagious and fatal disease caused in pigs by ASFV.

The Topo IIA encoded by *Nucleocytoviricota* are very similar to the ubiquitous eukaryotic Topo IIA at the sequence level and in that they are homodimers devoid of gyrase activity (Fig. 1). In the traditional view that considers viruses as pickpockets of cellular proteins, this suggests that Topo IIA were acquired by *Nucleocytoviricota* from their eukaryotic hosts. However, in the framework of the "out of viruses" hypothesis for the origin of DNA (Forterre 2002), it is tempting to suggest that this gene transfer took place the other way around, and that eukaryotic Topo IIA was acquired from the *Nucleocytoviricota* (Forterre and Gadelle 2009). A preliminary phylogenetic analyses provided ambiguous results: some Topo IIA from *Nucleocytoviricota* branched between T4 and Eukarya, suggesting that Topo IIA was indeed transferred from viruses to cells, whereas other viral enzymes branched within eukaryotes in agreement with transfers from cells to viruses (Gadelle et al. 2003; Forterre et al. 2007).

At the time of these analyses, only six Topo IIA from four families (*Asfarviridae*, *Mimiviridae, Iridoviridae, Phycodnaviridae*) within *Nucleocytoviricota* were known (Forterre et al. 2007). During the last decade, a great number of new *Nucleocytoviricota* genomes became available, including those of giant viruses from the families *Mimiviridae, Marseilleviridae* and *Pandoraviridae*, which encode Topo IIA (Abergel et al. 2015; Colson et al. 2017). Notably, it was shown that the Topo IIA encoded by *Marseilleviridae* branch as a sister clade to Eukarya (Erives 2017). We thus decided to update the Topo IIA phylogenetic classification, focusing on viral and eukaryotic Topo IIA. Our results strongly suggest that eukaryotic Topo IIA originated from a Topo IIA ancestor encoded by a virus closely related to modern *Megaviricetes*, a class of *Nucleocytoviricota* that includes many giant viruses, such as *Mimiviridae*. We have previously reported phylogenetic analyses suggesting that eukaryotic RNA polymerase II was probably recruited by eukaryotes from a virus related to *Imitervirales* in a tree including

Nucleocytoviricota and the three nuclear RNA polymerases present in all eukaryotes (Guglielmini et al. 2019). One can speculate that both RNA polymerase II and Topo IIA were possibly acquired simultaneously by a proto-eukaryote in the lineage leading to the last eukaryotic common ancestor (LECA), in agreement with the fact that these two enzymes interact functionally and physically in modern eukaryotes. Regardless, our results support the hypothesis that interactions between proto-eukaryotes and *Nucleocytoviricota* have played an important role in shaping the physiology of modern eukaryotic cells.

Material and methods

Data collection

For Bacteria, we downloaded the full proteomes of a set of 112 bacterial strains spanning 10 representative groups (Aquificae, Dictyoglomi, Elusimicrobia, FCB group, Nitrospirae, PVC group, Proteobacteria, Spirochaetes, Terrabacteria group, Thermotogae). We used BLASTP v2.9.0+ (Ramsay et al. 2000; Camacho et al. 2009) recursively to collect the homologs of *Escherichia coli* K12 GyrB and GyrA proteins (WP_000072067.1, NP_416734.1) in those 112 proteomes. Finally, we concatenated each related pair of GyrB and GyrA hits. We also added the sequence of *E. coli* Topo IV (Table S1) that branched at the base of DNA gyrase in previous analyses.

For *Caudoviricetes* (tailed phages), we downloaded all the 1,131,926 *Caudoviricetes* proteins, and next used BLASTP to search for the homologs of the three subunits of the T4 phage topoisomerase II (E-value lower than 1e-10). We kept only *Caudoviricetes* lineages for which we obtained a hit for the three subunits, and concatenated the corresponding sequences. Interestingly, beside the group of previously known Topo IIA closely related to T4 Topo IIA infecting Enterobacteriaceae, we detected several new related Topo IIAs sequences in distantly related members of the *Myoviridae* infecting Rhizobiaceae and Firmicutes, but also in three members of the *Ackermannviridae* infecting Enterobacter and Rhizobiaceae as well as three unclassified siphoviruses infecting Firmicutes (Table S1).

For *Nucleocytoviricota* (NCLDV), we searched for Topo IIA sequences in *Nucleocytoviricota* genomes that we previously used to determine the list of core genes of *Nucleocytoviricota* and the phylogeny based on the concatenation of 8 core genes (Guglielmini et al. 2019). Topo IIA turned out to be present in all *Imitervirales* (formerly *Megavirales* in (Guglielmini et al. 2019)) an order that includes *Mimiviridae*, the so-called 'extended Mimiviridae' and several more recently described large DNA viruses (Catovirus, Hokovirus, Indivirus, Klosneuvirus and Tupanvirus) (Guglielmini et al. 2019). Topo IIA is also present in some members of the *Pimascovirales*, in particular, in all members of *Marseilleviridae* and viruses of the Pitho-like group (Orpheovirus, Cedratvirus, Pithovirus). Finally, Topo IIA is present in all members of the order *Asfuvirales*, which includes *Asfarviridae* and related viruses (Kaumovirus, Faustovirus, Pacmanvirus).

It turned out to be more challenging to assemble the dataset of eukaryotic Topo IIA sequences, because we found many fragments of Topo IIA in eukaryotic proteomes. Thus, we produced position-specific scoring matrices (PSSM) for GyrA and GyrB proteins using alignments from the PFAM database (files provided as supplementary materials) and searched for coding sequences matching both profiles. We defined a list of 52 eukaryotic organisms, representative of the known eukaryotic diversity. When possible, we downloaded the corresponding proteomes and used PSSM as PSIBLAST queries to obtain Topo IIA homologs. For those organisms where no proteome was available, we looked for transcriptomic data and performed *de novo* assembly using Trimmomatic v0.36 (Bolger et al. 2014) for the read pre-processing step, SortMeRNA v2.1b (Kopylova et al. 2012) to filter out rRNA sequences, and Trinity v2.2.0 (Grabherr et al. 2011) for the assembly. We then used the GyrA and GyrB PSSMs as queries for a TBLASTN search against the assemblies and kept hits matching both profiles.

Phylogenetic analyses

All multiple amino-acid sequence alignments were performed using MAFFT v7.407 (Katoh and Standley 2013) and the E-INSi algorithm. Sites containing more than 50% of gaps were filtered out. Of note, for the tree with the largest taxonomic sampling, we used Divvier v1.0 (Ali et al. 2019) to reduce alignment errors with the MAFFT output.

All phylogenetic analyses were performed using IQ-TREE v1.6.7.2 (Nguyen et al. 2015). We selected the best-fit model using the IQ-TREE's model finder (Wong et al. 2017) according to the BIC criterion. For the tree with the largest taxonomic sampling we used a mixture model (selected according to the BIC criterion) and the PMSF implementation (Wang et al. 2018). We made the search for the best tree more thorough by using the "allnni" option as well as setting the "pers" parameter to 0.2 and the "nstop" parameter 500. We always used 10 independent runs (--runs option of IQ-TREE) and selected the best one. Confidence branch supports were assessed using the transfer bootstrap expectation (1000 replicates except for the tree including all sequences, where 100 replicates were used (Lemoine et al. 2018). We used iToL v4.4.2 (Letunic and Bork 2016) to generate the figures. All trees and alignments are available https://doi.org/10.5281/zenodo.5702416.

Results

Viral Topo IIA branch between bacterial and eukaryotic Topo IIA in a global phylogeny

We first built a tree spanning the whole diversity of Topo IIA by including sequences from Bacteria, *Caudoviricetes*, *Nucleocytoviricota* and eukaryotic Topo IIA (Fig. 2). We did not include archaeal DNA gyrases because they branch within bacterial DNA gyrases in previous phylogenetic analyses (Forterre et al. 2007; Raymann et al. 2014). Importantly, we did not detect orthologues of eukaryotic-like Topo IIA in the MAGs of different lineages of Asgard archaea, but only bacterial-like DNA gyrases.

The four groups of sequences (Bacteria, *Caudoviricetes, Nucleocytoviricota* and eukaryotes) were clearly separated in the tree (Fig. 2). The tree was arbitrarily rooted between *Nucleocytoviricota* and *Caudoviricetes* for convenience, dividing the tree in two clusters, one grouping eukaryotes and their viruses (*Nucleocytoviricota)* and the other grouping Bacteria and their viruses (*Caudoviricetes)*. Both Bacteria and eukaryotes were monophyletic. In contrast, it was not possible to obtain the monophyly of either *Caudoviricetes* or *Nucleocytoviricota*. Importantly, in contrast to our previous analysis (Forterre et al. 2007), *Nucleocytoviricota* and eukaryotic Topo IIAs were not intermixed.

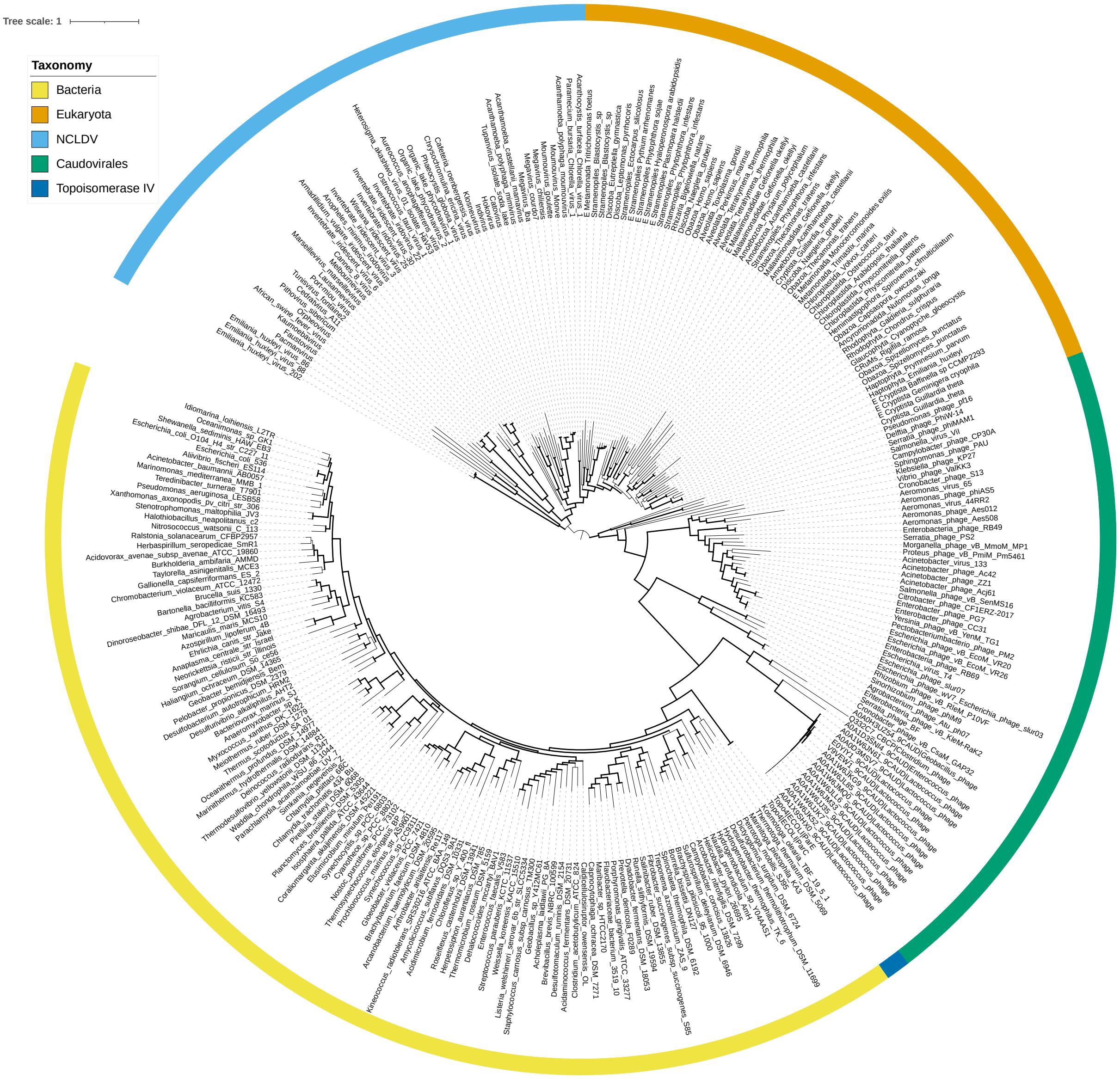

Fig. 2. Phylogenetic tree of the Topo IIA

Maximum-likelihood tree for 269 Topo IIA proteins for bacteria (113 sequences, including the two Topoisomerase IV proteins from *Escherichia coli*, ParC and ParE), eukaryotes (53 sequences), *Nucleocytoviricota* (47 sequences) and *Caudoviricetes* (56 sequences). The outer circle colors represent the group to which the sequences belong. The selected model was LG+F+R15. Thick branches have a branch support (TBE) greater than 70%.

Although DNA viruses encode many viral-specific DNA replication proteins, they can sometimes recruit cellular replisome components (Krupovič et al. 2010). We thus wondered if the grouping of *Caudoviricetes* and *Nucleocytoviricota* between Bacteria and eukaryotes was due to the long-branch attraction (LBA) artifact, with *Caudoviricetes* branching within Bacteria and *Nucleocytoviricota* within eukaryotes. This seemed unlikely considering the great divergence between viral Topo IIAs and their cellular counterparts. However, to test this hypothesis, we built several subtrees, both to remove groups with long branches and to enhance the signal by increasing the number of meaningfully aligned amino acids. After removing bacterial sequences, the most distant outgroup, we obtained a tree topology largely reproducing the relationships between *Caudoviricetes*, *Nucleocytoviricota* and eukaryotes observed in the global phylogeny (Fig. 3). More importantly, after removing both Bacteria and *Caudoviricetes, Nucleocytoviricota* remained well separated from eukaryotes (Fig. 4 a, b). This indicates that the separation of *Nucleocytoviricota* and eukaryotes in the tree was not due to an attraction of *Nucleocytoviricota* by *Caudoviricetes* and/or Bacteria. Similarly, *Caudoviricetes* remained well separated from bacteria after removing both *Nucleocytoviricota* and eukaryotes (Fig. 5).

The distribution and phylogeny of Topo IIA can provide information about their presence, or not, in the ancestors of each group of organisms. The ubiquity of DNA gyrase in bacteria leaves little doubt that this enzyme was present in the Last Bacterial Common Ancestor (LBCA). Similarly, the ubiquity of the single-subunit Topo IIA in Eukarya testifies to the presence of at least one Topo IIA in the LECA. However, we did not recover the monophyly of all major eukaryotic divisions in our phylogeny (Fig. 4a). Members of certain divisions were present in different parts of the tree, suggesting a complex history of Topo IIA during the diversification of eukaryotes, including gene duplication and gene loss. Several eukaryotes indeed contain more than one Topo IIA gene (Forterre et al. 2007). Some correspond to recent duplications (such as the Topo IIα and Topo IIβ in vertebrates), but others probably correspond to more ancient gene duplications or possibly gene transfers between eukaryotic lineages. With the root of the eukaryotic tree being still debated (Burki et al. 2019), it is difficult to propose a scenario for the evolution of Topo IIAs in eukaryotes. From our phylogenetic analyses, one cannot exclude that LECA already contained more than one Topo IIA.

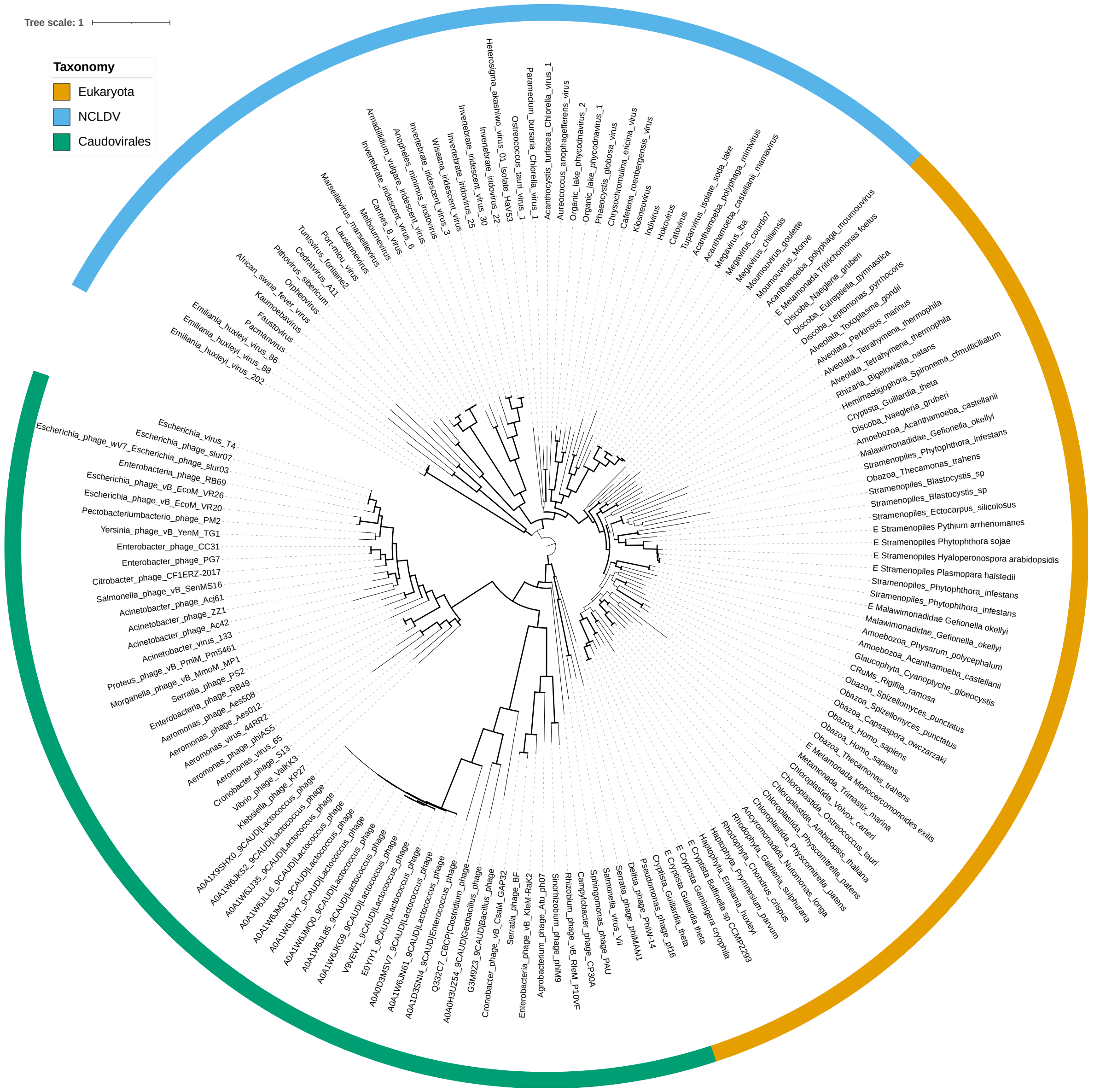
Tree scale: 1
Taxonomy
Eukaryota
NCLDV
Caudovirales
Emiliania_huxleyi_virus_202
Emiliania_huxleyi_virus_88
Emiliania_huxleyi_virus_86
Pacmanvirus
Faustovirus
Kaumoebavirus
African_swine_fever_virus
Orpheovirus
Pithovirus_sibericum
Cedratvirus_A11
Tunisvirus_fontaine2
Port-miou_virus
Lausannevirus
Marseillevirus_marseillevirus
Melbournevirus
Cannes_8_virus
Invertebrate_iridescent_virus_6
Armadillidium_vulgare_iridescent_virus
Anopheles_minimus_irodovirus
Invertebrate_iridescent_virus_3
Wiseana_iridescent_virus
Invertebrate_iridovirus_25
Invertebrate_iridescent_virus_30
Invertebrate_iridovirus_22
Heterosigma_akashiwo_virus_01_isolate_HaV53
Ostreococcus_tauri_virus_1
Paramecium_bursaria_Chlorella_virus_1
Acanthocystis_turfacea_Chlorella_virus_1
Aureococcus_anophagefferens_virus
Organic_lake_phycodnavirus_2
Organic_lake_phycodnavirus_1
Phaeocystis_globosa_virus
Chrysochromulina_ericina_virus
Cafeteria_roenbergensis_virus
Klosneuvirus
Indivirus
Hokovirus
Catovirus
Tupanvirus_isolate_soda_lake
Acanthamoeba_polyphaga_mimivirus
Acanthamoeba_castellanii_mamavirus
Megavirus_Iba
Megavirus_courdo7
Megavirus_chiliensis
Moumouvirus_goulette
Moumouvirus_Monve
Acanthamoeba_polyphaga_moumouvirus
E Metamonada Tritrichomonas foetus
Discoba_Naegleria_gruberi
Discoba_Eutreptiella_gymnastica
Discoba_Leptomonas_pyrrhocoris
Alveolata_Toxoplasma_gondii
Alveolata_Perkinsus_marinus
Alveolata_Tetrahymena_thermophila
Alveolata_Tetrahymena_thermophila
Rhizaria_Bigelowiella_natans
Hemimastigophora_Spironema_cfmulticiliatum
Cryptista_Guillardia_theta
Discoba_Naegleria_gruberi
Amoebozoa_Acanthamoeba_castellanii
Malawimonadidae_Gefionella_okellyi
Stramenopiles_Phytophthora_infestans
Obazoa_Thecamonas_trahens
Stramenopiles_Blastocystis_sp
Stramenopiles_Blastocystis_sp
Stramenopiles_Ectocarpus_silicolosus
E Stramenopiles Pythium arrhenomanes
E Stramenopiles Phytophthora sojae
E Stramenopiles Hyaloperonospora arabidopsidis
E Stramenopiles Plasmopara halstedii
Stramenopiles_Phytophthora_infestans
Stramenopiles_Phytophthora_infestans
E Malawimonadidae Gefionella okellyi
Malawimonadidae_Gefionella_okellyi
Amoebozoa_Physarum_polycephalum
Amoebozoa_Acanthamoeba_castellanii
Glaucophyta_Cyanoptyche_gloeocystis
CRuMs_Rigifila_ramosa
Obazoa_Spizellomyces_punctatus
Obazoa_Spizellomyces_punctatus
Obazoa_Capsaspora_owczarzaki
Obazoa_Homo_sapiens
Obazoa_Homo_sapiens
Obazoa_Thecamonas_trahens
E Metamonada Monocercomonoides exilis
Metamonada_Trimastix_marina
Chloroplastida_Volvox_carteri
Chloroplastida_Ostreococcus_tauri
Chloroplastida_Arabidopsis_thaliana
Chloroplastida_Physcomitrella_patens
Chloroplastida_Physcomitrella_patens
Ancyromonadida_Nutomonas_longa
Rhodophyta_Galdieria_sulphuraria
Rhodophyta_Chondrus_crispus
Haptophyta_Prymnesium_parvum
Haptophyta_Emiliania_huxleyi
Haptophyta Baffinella sp CCMP2293
E Cryptista Geminigera cryophila
E Cryptista Guillardia theta
Cryptista_Guillardia_theta
Pseudomonas_phage_pf16
Delftia_phage_PhiW-14
Serratia_phage_phiMAM1
Salmonella_virus_Vil
Sphingomonas_phage_PAU
Campylobacter_phage_CP30A
Rhizobium_phage_vB_RleM_P10VF
Sinorhizobium_phage_phiM9
Agrobacterium_phage_Atu_ph07
Enterobacteria_phage_vB_KleM-RaK2
Serratia_phage_BF
Cronobacter_phage_vB_CsaM_GAP32
G3M9Z3_9CAUD|Bacillus_phage
A0A0H3UZ54_9CAUD|Geobacillus_phage
Q332C7_CBCP|Clostridium_phage
A0A1D3SNI4_9CAUD|Enterococcus_phage
A0A1W6JN61_9CAUD|Lactococcus_phage
A0A0D3MSV7_9CAUD|Lactococcus_phage
E0YIY1_9CAUD|Lactococcus_phage
V9VEW1_9CAUD|Lactococcus_phage
A0A1W6JKG9_9CAUD|Lactococcus_phage
A0A1W6JL85_9CAUD|Lactococcus_phage
A0A1W6JMQ0_9CAUD|Lactococcus_phage
A0A1W6JJK7_9CAUD|Lactococcus_phage
A0A1W6JM33_9CAUD|Lactococcus_phage
A0A1W6JLL6_9CAUD|Lactococcus_phage
A0A1W6JJ35_9CAUD|Lactococcus_phage
A0A1W6JK52_9CAUD|Lactococcus_phage
A0A1X9SHX0_9CAUD|Lactococcus_phage
Klebsiella_phage_KP27
Vibrio_phage_ValKK3
Cronobacter_phage_S13
Aeromonas_virus_65
Aeromonas_phage_phiAS5
Aeromonas_virus_44RR2
Aeromonas_phage_Aes012
Aeromonas_phage_Aes508
Enterobacteria_phage_RB49
Serratia_phage_PS2
Morganella_phage_vB_MmoM_MP1
Proteus_phage_vB_PmiM_Pm5461
Acinetobacter_virus_133
Acinetobacter_phage_Ac42
Acinetobacter_phage_ZZ1
Acinetobacter_phage_Acj61
Salmonella_phage_vB_SenMS16
Citrobacter_phage_CF1ERZ-2017
Enterobacter_phage_PG7
Enterobacter_phage_CC31
Yersinia_phage_vB_YenM_TG1
Pectobacteriumbacterio_phage_PM2
Escherichia_phage_vB_EcoM_VR20
Escherichia_phage_vB_EcoM_VR26
Enterobacteria_phage_RB69
Escherichia_phage_wV7_Escherichia_phage_slur03
Escherichia_phage_slur07
Escherichia_virus_T4

Fig. 3. Phylogenetic tree of the Topo IIA for the eukaryotes, *Nucleocytoviricota* and *Caudoviricetes*

Maximum-likelihood tree for 156 Topo IIA proteins for eukaryotes (53 sequences), *Nucleocytoviricota* (47 sequences) and *Caudoviricetes* (56 sequences). The outer circle colors represent the group to which the sequences belong. The selected model was LG+F+R10. Thick branches have a branch support (TBE) greater than 70%.

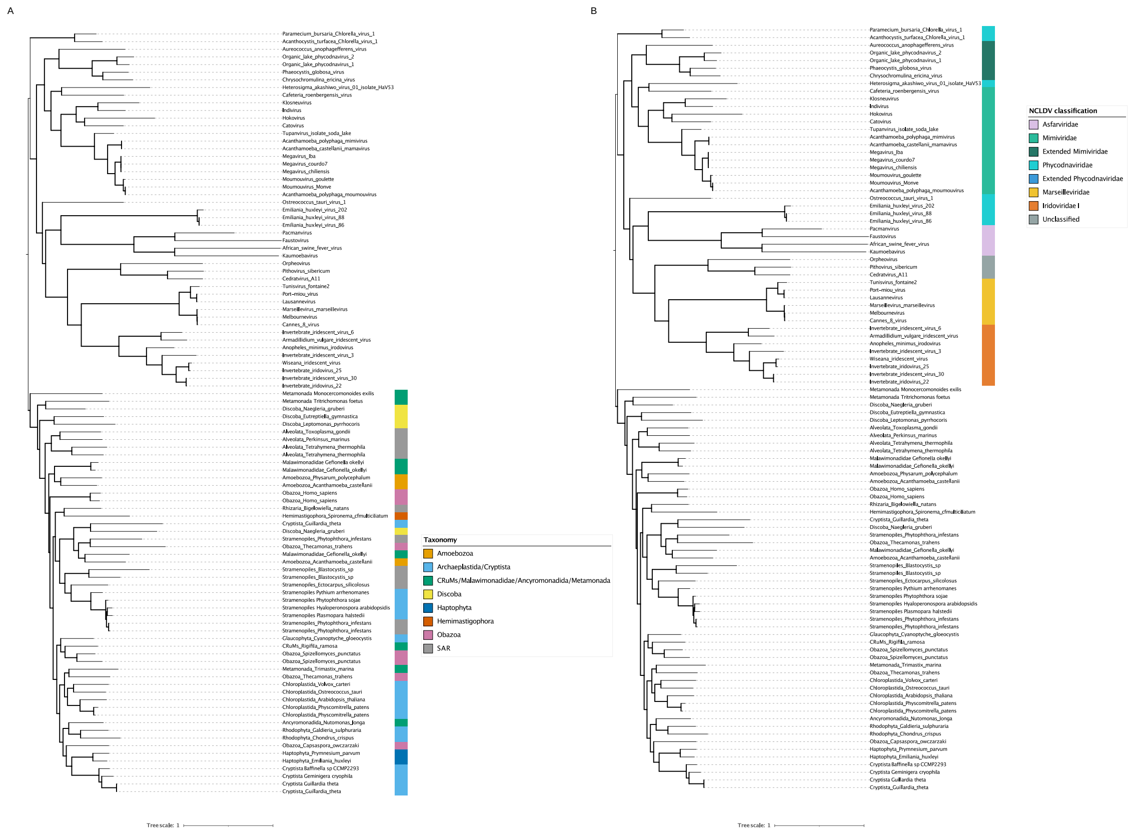

Fig. 4. Phylogenetic tree of the Topo IIA for the eukaryotes and *Nucleocytoviricota*
Maximum-likelihood tree for 100 Topo IIA proteins for eukaryotes (53 sequences) and *Nucleocytoviricota* (47 sequences). The selected model was LG+F+R6. Thick branches have a branch support (TBE) greater than 70%.
In panel A, the colored bar represents the eukaryotes classification. In panel B, the colored bar represents the *Nucleocytoviricota* classification.

The broad representation of Topo IIA in *Nucleocytoviricota* suggests that this enzyme was also present in the Last *Nucleocytoviricota* Common Ancestor (LNCA), and was subsequently lost in a few lineages. This hypothesis is supported by the congruence between the phylogenetic tree of *Nucleocytoviricota* Topo IIA (Fig. 4b) and the global phylogenetic classification of *Nucleocytoviricota* based on the concatenation of eight (core) genes present in most families of this phylum (Guglielmini et al. 2019). In the 8-core-genes phylogeny, *Nucleocytoviricota* were divided into two clusters that we named PAM (*Phycodnaviridae, Asfarviridae, Mimiviridae)* and MAPI (*Marseilleviridae*, *Ascoviridae*, Pitho-like viruses, *Iridoviridae)*, respectively. The PAM cluster included viruses corresponding to the recently proposed class *Megaviricetes* and *Pokkesviricetes*, whereas the MAPI cluster corresponded to the recently proposed order *Pimascovirales* (Koonin et al. 2020). In the Topo IIA NCLDV phylogenetic tree rooted with eukaryotes (Fig. 4b), we recovered the monophyly of all families of *Nucleocytoviricota*, as well as and the monophyly of *Pimascovirales*. Notably, the NCLDV Topo IIA tree was rooted deep within *Megaviricetes* when eukaryotes were used as the outgroup.

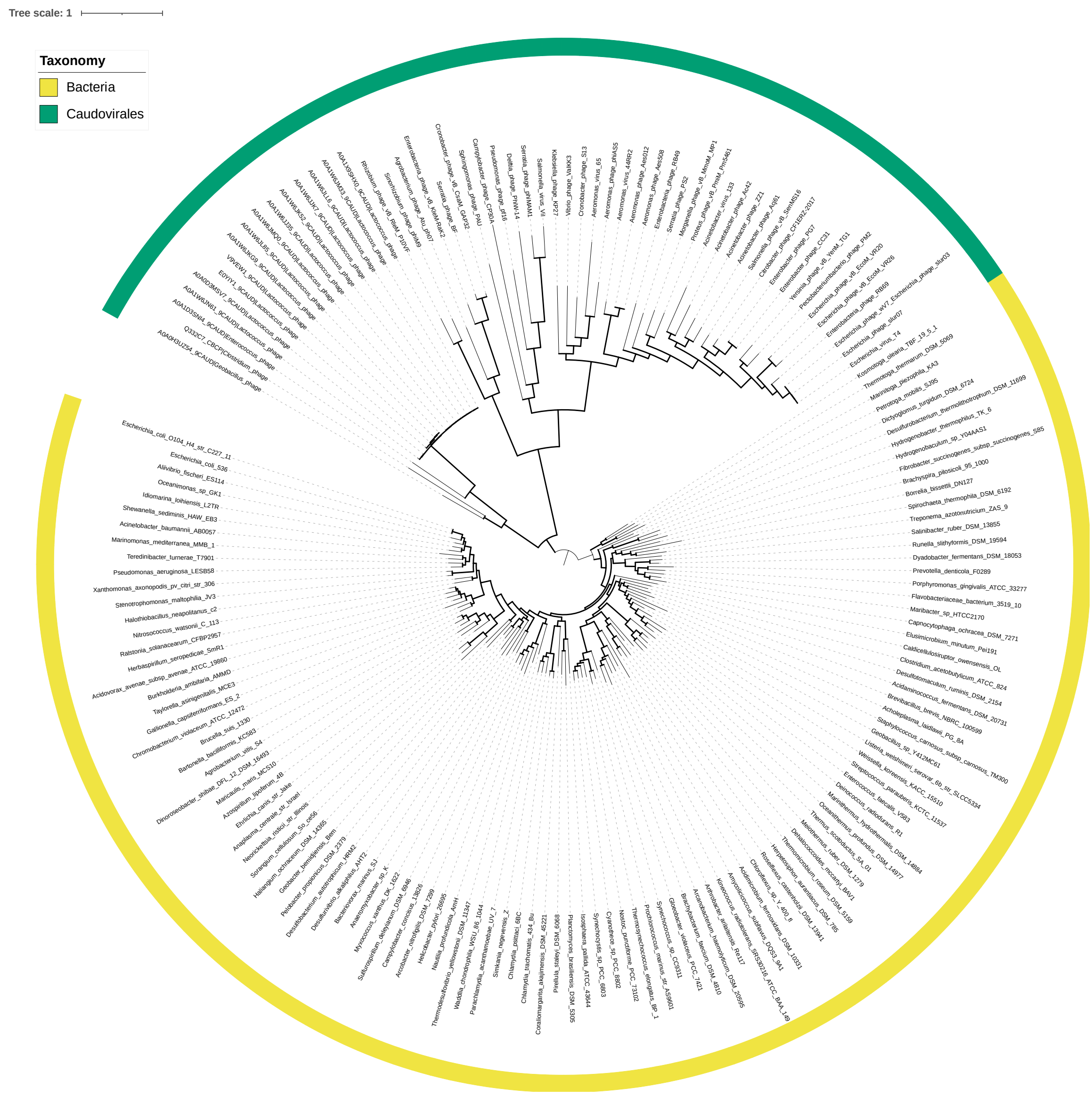
Tree scale: 1
Taxonomy
Bacteria
Caudovirales
A0A0H3UZ54_9CAUD|Geobacillus_phage
Q332C7_CBCP|Clostridium_phage
A0A1D3SNI4_9CAUD|Enterococcus_phage
A0A1W6JN61_9CAUD|Lactococcus_phage
A0A0D3MSV7_9CAUD|Lactococcus_phage
E0YIY1_9CAUD|Lactococcus_phage
V9VEW1_9CAUD|Lactococcus_phage
A0A1W6JKG9_9CAUD|Lactococcus_phage
A0A1W6JL85_9CAUD|Lactococcus_phage
A0A1W6JMQ0_9CAUD|Lactococcus_phage
A0A1W6JJ35_9CAUD|Lactococcus_phage
A0A1W6JK52_9CAUD|Lactococcus_phage
A0A1W6JJK7_9CAUD|Lactococcus_phage
A0A1W6JLL6_9CAUD|Lactococcus_phage
A0A1W6JM33_9CAUD|Lactococcus_phage
A0A1X9SHX0_9CAUD|Lactococcus_phage
Rhizobium_phage_vB_RleM_P10VF
Sinorhizobium_phage_phiM9
Agrobacterium_phage_Atu_ph07
Enterobacteria_phage_vB_KleM-RaK2
Serratia_phage_BF
Cronobacter_phage_vB_CsaM_GAP32
Sphingomonas_phage_PAU
Campylobacter_phage_CP30A
Pseudomonas_phage_pf16
Delftia_phage_PhiW-14
Serratia_phage_phiMAM1
Salmonella_virus_Vil
Klebsiella_phage_KP27
Vibrio_phage_VaIKK3
Cronobacter_phage_S13
Aeromonas_virus_65
Aeromonas_phage_phiAS5
Aeromonas_virus_44RR2
Aeromonas_phage_Aes012
Aeromonas_phage_Aes508
Enterobacteria_phage_RB49
Serratia_phage_PS2
Morganella_phage_vB_MmoM_MP1
Proteus_phage_vB_PmiM_Pm5461
Acinetobacter_virus_133
Acinetobacter_phage_Ac42
Acinetobacter_phage_ZZ1
Acinetobacter_phage_Acj61
Salmonella_phage_vB_SenMS16
Citrobacter_phage_CF1ERZ-2017
Enterobacter_phage_PG7
Enterobacter_phage_CC31
Yersinia_phage_vB_YenM_TG1
Pectobacteriumbacterio_phage_PM2
Escherichia_phage_vB_EcoM_VR20
Escherichia_phage_vB_EcoM_VR26
Enterobacteria_phage_RB69
Escherichia_phage_wV7_Escherichia_phage_slur03
Escherichia_phage_slur07
Escherichia_virus_T4
Kosmotoga_olearia_TBF_19_5_1
Thermotoga_thermarum_DSM_5069
Marinitoga_piezophila_KA3
Petrotoga_mobilis_SJ95
Dictyoglomus_turgidum_DSM_6724
Desulfurobacterium_thermolithotrophum_DSM_11699
Hydrogenobacter_thermophilus_TK_6
Hydrogenobaculum_sp_Y04AAS1
Fibrobacter_succinogenes_subsp_succinogenes_S85
Brachyspira_pilosicoli_95_1000
Borrelia_bissettii_DN127
Spirochaeta_thermophila_DSM_6192
Treponema_azotonutricium_ZAS_9
Salinibacter_ruber_DSM_13855
Runella_slithyformis_DSM_19594
Dyadobacter_fermentans_DSM_18053
Prevotella_denticola_F0289
Porphyromonas_gingivalis_ATCC_33277
Flavobacteriaceae_bacterium_3519_10
Maribacter_sp_HTCC2170
Capnocytophaga_ochracea_DSM_7271
Elusimicrobium_minutum_Pei191
Caldicellulosiruptor_owensensis_OL
Clostridium_acetobutylicum_ATCC_824
Desulfotomaculum_ruminis_DSM_2154
Acidaminococcus_fermentans_DSM_20731
Brevibacillus_brevis_NBRC_100599
Acholeplasma_laidlawii_PG_8A
Staphylococcus_carnosus_subsp_carnosus_TM300
Geobacillus_sp_Y412MC61
Listeria_welshimeri_serovar_6b_str_SLCC5334
Weissella_koreensis_KACC_15510
Streptococcus_parauberis_KCTC_11537
Enterococcus_faecalis_V583
Deinococcus_radiodurans_R1
Marinithermus_hydrothermalis_DSM_14884
Oceanithermus_profundus_DSM_14977
Thermus_scotoductus_SA_01
Meiothermus_ruber_DSM_1279
Dehalococcoides_mccartyi_BAV1
Thermomicrobium_roseum_DSM_5159
Herpetosiphon_aurantiacus_DSM_785
Roseiflexus_castenholzii_DSM_13941
Chloroflexus_sp_Y_400_fl
Acidimicrobium_ferrooxidans_DSM_10331
Amycolicicoccus_subflavus_DQS3_9A1
Kineococcus_radiotolerans_SRS30216_ATCC_BAA_149
Arthrobacter_arilaitensis_Re117
Arcanobacterium_haemolyticum_DSM_20595
Brachybacterium_faecium_DSM_4810
Gloeobacter_violaceus_PCC_7421
Synechococcus_sp_CC9311
Prochlorococcus_marinus_str_AS9601
Thermosynechococcus_elongatus_BP_1
Nostoc_punctiforme_PCC_73102
Cyanothece_sp_PCC_8802
Synechocystis_sp_PCC_6803
Isosphaera_pallida_ATCC_43644
Planctomyces_brasiliensis_DSM_5305
Pirellula_staleyi_DSM_6068
Coraliomargarita_akajimensis_DSM_45221
Chlamydia_trachomatis_434_Bu
Chlamydia_psittaci_6BC
Simkania_negevensis_Z
Parachlamydia_acanthamoebae_UV_7
Waddlia_chondrophila_WSU_86_1044
Thermodesulfovibrio_yellowstonii_DSM_11347
Nautilia_profundicola_AmH
Helicobacter_pylori_26695
Arcobacter_nitrofigilis_DSM_7299
Campylobacter_concisus_13826
Sulfurospirillum_deleyianum_DSM_6946
Myxococcus_xanthus_DK_1622
Anaeromyxobacter_sp_K
Bacteriovorax_marinus_SJ
Desulfurivibrio_alkaliphilus_AHT2
Desulfobacterium_autotrophicum_HRM2
Pelobacter_propionicus_DSM_2379
Geobacter_bemidjiensis_Bem
Haliangium_ochraceum_DSM_14365
Sorangium_cellulosum_So_ce56
Neorickettsia_risticii_str_Illinois
Anaplasma_centrale_str_Israel
Ehrlichia_canis_str_Jake
Azospirillum_lipoferum_4B
Maricaulis_maris_MCS10
Dinoroseobacter_shibae_DFL_12_DSM_16493
Agrobacterium_vitis_S4
Bartonella_bacilliformis_KC583
Brucella_suis_1330
Chromobacterium_violaceum_ATCC_12472
Gallionella_capsiferriformans_ES_2
Taylorella_asinigenitalis_MCE3
Burkholderia_ambifaria_AMMD
Acidovorax_avenae_subsp_avenae_ATCC_19860
Herbaspirillum_seropedicae_SmR1
Ralstonia_solanacearum_CFBP2957
Nitrosococcus_watsonii_C_113
Halothiobacillus_neapolitanus_c2
Stenotrophomonas_maltophilia_JV3
Xanthomonas_axonopodis_pv_citri_str_306
Pseudomonas_aeruginosa_LESB58
Teredinibacter_turnerae_T7901
Marinomonas_mediterranea_MMB_1
Acinetobacter_baumannii_AB0057
Shewanella_sediminis_HAW_EB3
Idiomarina_loihiensis_L2TR
Oceanimonas_sp_GK1
Aliivibrio_fischeri_ES114
Escherichia_coli_536
Escherichia_coli_O104_H4_str_C227_11

Fig. 5. Phylogenetic tree of the Topo IIA for the bacteria and *Caudoviricetes*. Maximum-likelihood tree for 167 Topo IIA proteins for bacteria (111 sequences) and *Caudoviricetes* (56 sequences). The outer circle colors represent the group to which the sequences belong. The selected model was LG+R11. Thick branches have a branch support (TBE) greater than 70%.

In contrast to the situation with Bacteria, eukaryotes and *Nucleocytoviricota*, Topo IIA are only present in a few subgroups of *Caudoviricetes*. Most Topo IIA are encoded by T4-like myoviruses (i.e., viruses with contractile tails recently reclassified into the family *Straboviridae*) with larger genomes, suggesting that Topo IIA was present in the last common ancestor of this phage group. Topo IIA encoded by *Ackermannviridae* (anther family of phages with contractile tails) branched within *Straboviridae* suggesting lateral gene transfer between these viral families (Fig. 5b). Three of the four Topo IIA encoded by viruses infecting Firmicutes have been tentatively assigned to the family *Siphoviridae* (phages with long non-contractile tails). They were grouped with Topo IIA of T4-like viruses, as a sister clade of bacterial homologs if the tree is rooted between *Nucleocytoviricota* and *Caudoviricetes***.**

Discussion

To discuss possible evolutionary scenarios, we arbitrarily rooted the Topo IIA phylogenetic tree (Fig. 2) at the three possible positions between the four clusters (Fig. 6 a,b,c). Rooting the tree between *Nucleocytoviricota* and eukaryotes (Fig. 6a) would suggest that *Nucleocytoviricota* and eukaryotic Topo IIA originated from a common viral or cellular ancestor. This scenario appears unlikely since it also implies that *Caudoviricetes* Topo IIA originated from *Nucleocytoviricota* Topo IIAs and *in fine* that bacterial DNA gyrases themselves originated from *Nucleocytoviricota* via *Caudoviricetes*. In that case, one should imagine that the LBCA originated after the diversification of *Nucleocytoviricota*. Since this diversification took place before LECA, at the time when ancestral *Nucleocytoviricota* infected proto-eukaryotic hosts, this scenario would suggest that proto-eukaryotes evolved before bacteria.

Rooting the tree between *Caudoviricetes* and *Nucleocytoviricota* (Fig. 6b) produced two clusters corresponding to Bacteria/*Caudoviricetes* and to *Nucleocytoviricota*/eukaryotes. This rooting suggests that bacterial DNA gyrase originated from Topo IIA of *Caudoviricetes*, whereas eukaryotic Topo IIA originated from those of *Nucleocytoviricota*. Considering the universal conservation of Topo IIA in Bacteria and eukaryotes, this scenario suggests that the transfer from viruses to cells took place before the emergence of the LBCA and LECA, respectively. Hence, both *Caudoviricetes* and *Nucleocytoviricota* should have originated and diversified before the emergence of the LBCA and LECA, infecting proto-bacterial and proto-eukaryotic hosts, respectively. Such ancient origin would explain the great divergence between the different versions of Topo IIA. The diversification of *Nucleocytoviricota* before LECA is indeed supported by the RNA polymerase phylogenetic tree including both viral and cellular enzymes (Guglielmini et al. 2019). Moreover it has been suggested that *Caudoviricetes,* which also infect archaeal hosts (Liu et al. 2021), have diverged even prior to the emergence of LUCA (Krupovic et al. 2020).

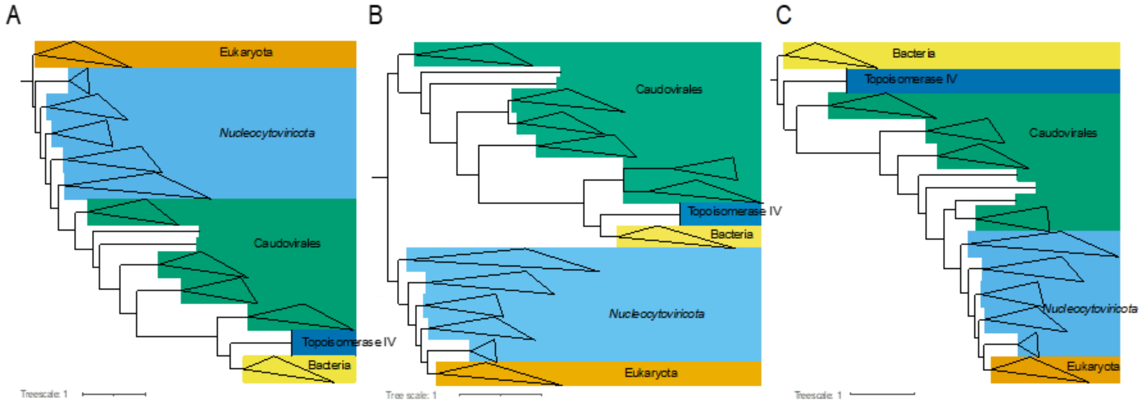

Fig. 6. Schematic representation of the different possible rooting for the Topo IIA phylogenetic tree.
The tree is the same as in Fig. 2. Monophyletic clades encompassing sequences from the same group have been collapsed and colored following Fig. 2. Panel A: rooting using the eukaryotes as an outgroup. Panel B: rooting using the bacteria and *Caudoviricetes* as an outgroup. Panel C: rooting using the bacteria as an outgroup.

Rooting the tree between Bacteria and *Caudoviricetes* (Fig. 6c) produced a tree in which Topo IIA of Bacteria and *Caudoviricetes* diverged from a common ancestor that predated the LBCA. In that case, the *Caudoviricetes* Topo IIA would have diverged from their bacterial counterparts before LBCA and continued during the diversification of Bacteria. The tree of Fig. 6c is consistent with the scenario in which eukaryotic viruses originated from a melting pot of bacterial viruses that infected the bacterium at the origin of mitochondria (Koonin, Krupovic, et al. 2015; Koonin, Dolja, et al. 2015) or another ancient bacterial endosymbiont present in a proto-eukaryotic ancestor of modern eukaryotes. In that case, the Topo IIA from a *Caudoviricetes* present in this putative bacterial endosymbiont would have been transferred to an ancestor of *Nucleocytoviricota*, potentially with other components of the DNA replication machinery shared between *Caudoviricetes* and *Nucleocytoviricota*, including NAD-dependent DNA ligase (Yutin and Koonin 2009). Notably, comparison of DNA replication machineries of all dsDNA viruses revealed a strong evolutionary and likely functional coupling between DNA topoisomerases and DNA ligases, with 96% of viruses encoding DNA topoisomerases also carrying a gene for a ligase (Kazlauskas et al. 2016). To explain the great divergence between the Topo IIA encoded by *Caudoviricetes* and those encoded by *Nucleocytoviricota* in terms of sequences and structure, this scenario entails that the rate of Topo IIA evolution increased dramatically following the transfer of the *Caudoviricetes* version into the lineage leading to the LNCA, with the fusion of the three Topo IIA subunits of *Caudoviricetes* Topo IIA into a single polypeptide.

The trees of Fig. 6b and 6c both can be also interpreted in the framework of the "*out of virus hypothesis*" for the origin of DNA topoisomerases (Forterre 2002; Forterre and Gadelle 2009). Conceivably, the different versions of Topo IIA originated in an ancient viral world. The scenario illustrated in Fig. 6b explicitly posits that proto-bacteria acquired their Topo IIA from an ancient *Caudoviricetes*, whereas in Fig. 6c, the bacterial and *Caudoviricetes* Topo IIA evolved from a common ancestor, which may or may not have been a virus. Regardless, in both scenarios, the eukaryotic Topo IIA has been acquired from *Nucleocytoviricota*. We have previously proposed a similar scenario to explain the ubiquitous distribution of Topo VI in Archaea and the restricted distribution of Topo VIII (both members of the Topo IIB family) in some archaea and bacterial mobile elements (Gadelle et al. 2014). In that scenario, the restricted distribution of Topo IIA to a few subgroups of *Caudoviricetes* seems surprising, but it resembles the restricted distribution of a recently described new version of RNA polymerase in a subgroup of these viruses (Weinheimer and Aylward 2020).

Importantly, if we exclude the unlikely conjecture in which the Topo IIA phylogenetic tree is rooted between *Nucleocytoviricota* and eukaryotes (Fig. 6a), the branching of all eukaryotes within *Nucleocytoviricota* in all other configurations suggests that a Topo IIA was introduced into eukaryotes from a member of this viral phylum. If the node at the base of the eukaryotic monophyletic clade corresponds to the position of LECA, as expected from the ubiquity of this enzyme in eukaryotes, the transfer of Topo IIA should have occurred before the emergence of LECA, i.e., from a member of *Nucleocytoviricota* to a proto-eukaryote. Alternatively, Topo IIA could have been introduced from *Nucleocytoviricota* to a particular eukaryotic lineage and later transferred from this lineage to all other lineages by horizontal gene transfer. This last scenario seems unlikely considering that Topo IIAs are present in all contemporary lineages of eukaryotes, without exception, and the enzyme is essential for several key functions conserved in all eukaryotic lineages. The fact that eukaryotes emerge in our analysis within the PAM group is consistent with the possibility that divergence of *Nucleocytoviricota* into several major families has predated the emergence of LECA (Guglielmini et al. 2019).

Our result raises an interesting question: which Topo II did proto-eukaryotes use before they captured the viral Topo IIA? A likely answer is that they relied on Topo IIB, since this enzyme is ubiquitous in Archaea, but also present in many eukaryotes. A Topo IIB-like protein with a very divergent V-B subunit is present in all eukaryotes and is part of the complex responsible for initiation of meiotic recombination (Vrielynck et al. 2016; Robert et al. 2017) whereas several eukaryotic lineages, e.g., Viridiplantae, contain a *bona fide* archaeal-like Topo IIB (Forterre et al. 2007; Malik et al. 2007; Forterre and Gadelle 2009).

In comparing the *Nucleocytoviricota* core genes' phylogeny with the phylogeny of the three eukaryotic nuclear RNA polymerases and those of *Nucleocytoviricota*, we have previously shown that two of the eukaryotic RNA polymerases, Pol I and Pol II, were probably introduced into the proto-eukaryotic lineage from *Nucleocytoviricota* (Guglielmini et al. 2019). This possibility was strongly supported in the case of the RNA polymerase II. It is worth noting that, like the position of Topo IIA in the present study, the RNA Pol II branched within *Megaviricetes* in the RNA polymerase tree. One can speculate that these two proteins (that play a major role in the eukaryotic transcription machinery) were recruited together from the same virus. This would make sense from the viewpoint of cell physiology, since the two enzymes interact both functionally and structurally. Indeed, it has been shown that Topo IIA is a structural component of the holo-Pol II complex and is essential for efficient RNA synthesis of nucleosomal DNA by this complex (Mondal and Parvin 2001). Topo IIA is required to produce long RNA Pol II transcripts (>3 kb) in *Saccharomyces cerevisiae* (Joshi et al. 2012) and enhances the recruitment of RNA Pol II to promoters in budding yeast (Sperling et al. 2011). It is possible that both Topo IIA and RNA Pol II were domesticated by a proto-eukaryote following the integration of a *Nucleocytoviricota* encoding these genes into the host chromosome. Integration of entire or large portions of the genomes of some *Nucleocytoviricota* into the chromosome of modern eukaryotes has been well documented (Delaroque and Boland 2008; Cock et al. 2010; Filée 2014; Moniruzzaman et al. 2020).

The viral origin of eukaryotic Topo IIA, in addition to those of RNA Pol II and possibly RNA Pol I, strengthens the idea that giant viruses of the phylum *Nucleocytoviricota* (especially members of the PAM group) played a major role in shaping the identity of modern eukaryotes (Forterre and Gaïa 2016). It is likely that other important proteins involved in eukaryotic physiology originated from *Nucleocytoviricota*. This has been proposed for eukaryotic histones, since the four histones from Medusavirus and Marseilleviruses branch at the base of the eukaryotic clades of their respective homologues (Erives 2017; Yoshikawa et al. 2019) and for enzymes involved in mRNA capping (Bell 2020). However, in those cases, robust phylogenetic analyses remain to be carried out since the published papers are based on limited sampling of the eukaryotic and *Nucleocytoviricota* diversity. The viral origin of some of the major players in eukaryotic cell biology was probably not limited to nuclear components since we have recently reported that the eukaryotic cytoplasmic actin might have been recruited by proto-eukaryotes from an actin-like protein (viractin) encoded by some *Imitervirales*, an order of *Megaviricetes* (Da Cunha et al. 2020).

The eukaryotic molecular fabric appears to be a melting pot of proteins that originated in *Nucleocytoviricota* (mainly *Megaviricetes*), those that emerged *de novo* in the eukaryotic stem branch, proteins inherited from the bacterial ancestor of mitochondria and chloroplasts, and proteins that had ancestors in Archaea (in two domains scenarios) or in the common ancestor of Archaea and eukaryotes (in three domains scenario). Sorting out the viral component of our eukaryotic ancestors is now a major task in understanding eukaryogenesis.

References

Abergel C, Legendre M, Claverie J-M. 2015. The rapidly expanding universe of giant viruses: Mimivirus, Pandoravirus, Pithovirus and Mollivirus. *FEMS Microbiol. Rev.* 39:779–796.

Ali RH, Bogusz M, Whelan S. 2019. Identifying Clusters of High Confidence Homologies in Multiple Sequence Alignments. *Mol. Biol. Evol.* 36:2340–2351.

Bell PJL. 2020. Evidence supporting a viral origin of the eukaryotic nucleus. *Virus Res.* [Internet] 289:198168. Available from: https://doi.org/10.1016/j.virusres.2020.198168

Bergerat A, de Massy B, Gadelle D, Varoutas P-C, Nicolas A, Forterre P. 1997. An atypical topoisomerase II from Archaea with implications for meiotic recombination. *Nature* 386:414–417.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120.

Burki F, Roger AJ, Brown MW, Simpson AGB. 2019. The New Tree of Eukaryotes. *Trends*

Ecol. Evol. [Internet] xx:43–55. Available from: http://dx.doi.org/10.1016/j.tree.2019.08.008

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: Architecture and applications. *BMC Bioinformatics* 10:1–9.

Champoux JJ. 2001. DNA Topoisomerases: Structure, Function, and Mechanism. *Annu. Rev. Biochem.* [Internet] 70:369–413. Available from: https://doi.org/10.1146/annurev.biochem.70.1.369

Cock JM, Sterck L, Rouzé P, Scornet D, Allen AE, Amoutzias G, Anthouard V, Artiguenave F, Aury J-M, Badger JH, et al. 2010. The Ectocarpus genome and the independent evolution of multicellularity in brown algae. *Nature* 465:617–621.

Coelho J, Ferreira F, Martins C, Leitão A. 2016. Functional characterization and inhibition of the type II DNA topoisomerase coded by African swine fever virus. *Virology* 493:209–216.

Coelho J, Martins C, Ferreira F, Leitão A. 2015. African swine fever virus ORF P1192R codes for a functional type II DNA topoisomerase. *Virology* 474:82–93.

Colson P, La Scola B, Levasseur A, Caetano-Anollés G, Raoult D. 2017. Mimivirus: leading the way in the discovery of giant viruses of amoebae. *Nat. Rev. Microbiol.* 15:243–254.

Da Cunha V, Gaia M, Gadelle D, Nasir A, Forterre P. 2017. Lokiarchaea are close relatives of Euryarchaeota, not bridging the gap between prokaryotes and eukaryotes. *PLoS Genet.* 13.

Da Cunha V, Gaia M, Ogata H, Jaillon O, Delmont TO, Forterre P. 2020. Giant viruses encode novel types of actins possibly related to the origin of eukaryotic actin: the viractins. *bioRxiv*:1–15.

Delaroque N, Boland W. 2008. The genome of the brown alga Ectocarpus siliculosus contains a series of viral DNA pieces, suggesting an ancient association with large dsDNA viruses. *BMC Evol. Biol.* 8:110.

Erives AJ. 2017. Phylogenetic analysis of the core histone doublet and DNA topo II genes of Marseilleviridae: Evidence of proto-eukaryotic provenance. *Epigenetics and Chromatin* [Internet] 10:1–11. Available from: https://doi.org/10.1186/s13072-017-0162-0

Filée J. 2014. Multiple occurrences of giant virus core genes acquired by eukaryotic genomes: the visible part of the iceberg? *Virology* 466–467:53–59.

Forterre P. 2002. The origin of DNA genomes and DNA replication proteins. *Curr. Opin. Microbiol.* 5:525–532.

Forterre P. 2006. DNA topoisomerase V: a new fold of mysterious origin. *Trends Biotechnol.* 24:245–247.

Forterre P. 2012. Introduction and Historical Perspective. In: Pommier Y, editor. DNA Topoisomerases and Cancer. New York, NY: Springer New York. p. 1–52. Available from: https://doi.org/10.1007/978-1-4614-0323-4_1

Forterre P, Gadelle D. 2009. Phylogenomics of DNA topoisomerases: Their origin and putative roles in the emergence of modern organisms. *Nucleic Acids Res.* 37:679–692.

Forterre P, Gaïa M. 2016. Giant viruses and the origin of modern eukaryotes. *Curr. Opin. Microbiol.* 31:44–49.

Forterre P, Gribaldo S, Gadelle D, Serre M-C. 2007. Origin and evolution of DNA topoisomerases. *Biochimie* 89:427–446.

Forterre P, Krupovic M, Raymann K, Soler N. 2014. Plasmids from Euryarchaeota. *Microbiol. Spectr.* 2:PLAS-0027-2014.

Freitas FB, Frouco G, Martins C, Leitão A, Ferreira F. 2016. In vitro inhibition of African swine fever virus-topoisomerase II disrupts viral replication. *Antiviral Res.* 134:34–41.

Gadelle D, Filée J, Buhler C, Forterre P. 2003. Phylogenomics of type II DNA topoisomerases. *BioEssays* 25:232–242.

Gadelle D, Krupovic M, Raymann K, Mayer C, Forterre P. 2014. DNA topoisomerase VIII: A novel subfamily of type IIB topoisomerases encoded by free or integrated plasmids in Archaea and Bacteria. *Nucleic Acids Res.* 42:8578–8591.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29:644–652.

Guglielmini J, Woo AC, Krupovic M, Forterre P, Gaia M, Gaïa M. 2019. Diversification of giant and large eukaryotic dsDNA viruses predated the origin of modern eukaryotes. *Proc Natl Acad Sci U S A*:455816.

Hizume K, Araki S, Yoshikawa K, Takeyasu K. 2007. Topoisomerase II, scaffold component, promotes chromatin compaction in vitro in a linker-histone H1-dependent manner. *Nucleic Acids Res.* 35:2787–2799.

Joshi RS, Piña B, Roca J. 2012. Topoisomerase II is required for the production of long Pol II gene transcripts in yeast. *Nucleic Acids Res.* 40:7907–7915.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* 30:772–780.

Kazlauskas D, Krupovic M, Venclovas C. 2016. The logic of DNA replication in double-stranded DNA viruses: insights from global analysis of viral genomes. *Nucleic Acids Res.* 44:4551–4564.

Koonin E V., Dolja V V., Krupovic M. 2015. Origins and evolution of viruses of eukaryotes: The ultimate modularity. *Virology* 479–480:2–25.

Koonin E V., Dolja V V., Krupovic M, Varsani A, Wolf YI, Yutin N, Zerbini FM, Kuhn JH. 2020. Global Organization and Proposed Megataxonomy of the Virus World. *Microbiol. Mol. Biol. Rev.* 84:1–33.

Koonin E V., Krupovic M, Yutin N. 2015. Evolution of double-stranded DNA viruses of eukaryotes: From bacteriophages to transposons to giant viruses. *Ann. N. Y. Acad. Sci.* 1341:10–24.

Kopylova E, Noé L, Touzet H. 2012. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* 28:3211–3217.

Krupovic M, Dolja V V., Koonin E V. 2020. The LUCA and its complex virome. *Nat. Rev. Microbiol.* [Internet]. Available from: http://dx.doi.org/10.1038/s41579-020-0408-x

Krupovič M, Gribaldo S, Bamford DH, Forterre P. 2010. The evolutionary history of archaeal MCM helicases: A case study of vertical evolution combined with Hitchhiking of mobile genetic elements. *Mol. Biol. Evol.* 27:2716–2732.

Lavrukhin O V, Fortune JM, Wood TG, Burbank DE, Van Etten JL, Osheroff N, Lloyd RS. 2000. Topoisomerase II from Chlorella virus PBCV-1. Characterization of the smallest known type II topoisomerase. *J. Biol. Chem.* 275:6915–6921.

Lemoine F, Domelevo Entfellner J-B, Wilkinson E, Correia D, Dávila Felipe M, De Oliveira T, Gascuel O. 2018. Renewing Felsenstein's phylogenetic bootstrap in the era of big data. *Nature* [Internet]. Available from: http://www.nature.com/articles/s41586-018-0043-0

Letunic I, Bork P. 2016. Interactive Tree Of Life (iTOL) v3: An online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* 44:242–245.

Liu LF, Liu CC, Alberts BM. 1979. T4 DNA topoisomerase: a new ATP-dependent enzyme essential for initiation of T4 bacteriophage DNA replication. *Nature* 281:456–461.

Liu Y, Demina TA, Roux S, Aiewsakun P, Kazlauskas D, Simmonds P, Prangishvili D, Oksanen HM, Krupovic M. 2021. Diversity, taxonomy and evolution of archaeal viruses of the class Caudoviricetes. *bioRxiv* [Internet]:2021.05.26.445827. Available from: http://biorxiv.org/content/early/2021/05/26/2021.05.26.445827.abstract

Malik S-B, Ramesh MA, Hulstrand AM, Logsdon JMJ. 2007. Protist homologs of the meiotic Spo11 gene and topoisomerase VI reveal an evolutionary history of gene duplication and lineage-specific loss. *Mol. Biol. Evol.* 24:2827–2841.

Mondal N, Parvin JD. 2001. DNA topoisomerase IIalpha is required for RNA polymerase II transcription on chromatin templates. *Nature* 413:435–438.

Moniruzzaman M, Weinheimer AR, Martinez-Gutierrez CA, Aylward FO. 2020. Widespread endogenization of giant viruses shapes genomes of green algae. *Nature* [Internet] 588:141–145. Available from: http://dx.doi.org/10.1038/s41586-020-2924-2

Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32:268–274.

Nielsen CF, Zhang T, Barisic M, Kalitsis P, Hudson DF. 2020. Topoisomerase IIα is essential for maintenance of mitotic chromosome structure. *Proc. Natl. Acad. Sci. U. S. A.* 117:12131–12142.

Nitiss JL. 2009. DNA topoisomerase II and its growing repertoire of biological functions. *Nat. Rev. Cancer* 9:327–337.

Pommier Y. 2013. Drugging topoisomerases: lessons and challenges. *ACS Chem. Biol.* 8:82–95.

Pommier Y, Sun Y, Huang SN, Nitiss JL. 2016. Roles of eukaryotic topoisomerases in transcription, replication and genomic stability. *Nat. Rev. Mol. Cell Biol.* [Internet] 17:703–721. Available from: https://doi.org/10.1038/nrm.2016.111

Ramsay L, Macaulay M, Degli Ivanissevich S, MacLean K, Cardle L, Fuller J, Edwards KJ, Tuvesson S, Morgante M, Massari A, et al. 2000. A simple sequence repeat-based linkage map of Barley. *Genetics* 156:1997–2005.

Raymann K, Forterre P, Brochier-Armanet C, Gribaldo S. 2014. Global phylogenomic analysis disentangles the complex evolutionary history of DNA replication in archaea. *Genome Biol. Evol.* 6:192–212.

Robert T, de Massy B, Grelon M. 2017. [TopoVIL : a molecular scissor essential for reproduction]. *Med. Sci. (Paris).* 33:512–518.

Sperling AS, Jeong KS, Kitada T, Grunstein M. 2011. Topoisomerase II binds nucleosome-free DNA and acts redundantly with topoisomerase I to enhance recruitment of RNA Pol II in budding yeast. *Proc. Natl. Acad. Sci. U. S. A.* 108:12693–12698.

Vos SM, Tretter EM, Schmidt BH, Berger JM. 2011. All tangled up: How cells direct, manage and exploit topoisomerase function. *Nat. Rev. Mol. Cell Biol.* [Internet] 12:827–841. Available from: https://doi.org/10.1038/nrm3228

Vrielynck N, Chambon A, Vezon D, Pereira L, Chelysheva L, De Muyt A, Mézard C, Mayer C, Grelon M. 2016. A DNA topoisomerase VI-like complex initiates meiotic recombination. *Science* 351:939–943.

Wall MK, Mitchenall LA, Maxwell A. 2004. Arabidopsis thaliana DNA gyrase is targeted to chloroplasts and mitochondria. *Proc. Natl. Acad. Sci. U. S. A.* 101:7821–7826.

Wang HC, Minh BQ, Susko E, Roger AJ. 2018. Modeling Site Heterogeneity with Posterior Mean Site Frequency Profiles Accelerates Accurate Phylogenomic Estimation. *Syst. Biol.* 67:216–235.

Wang JC. 2009. Untangling the Double Helix: DNA Entanglement and the Action of the DNA Topoisomerases. (Cold Spring Harbor Laboratory Press, editor.).

Weinheimer AR, Aylward FO. 2020. A distinct lineage of Caudovirales that encodes a deeply branching multi-subunit RNA polymerase. *Nat. Commun.* [Internet] 11:1–9. Available from: http://dx.doi.org/10.1038/s41467-020-18281-3

Wong TKF, Jermiin LS, Minh BQ, Kalyaanamoorthy S, von Haeseler A. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* [Internet] 14:587–589. Available from: http://dx.doi.org/10.1038/nmeth.4285

Yoshikawa G, Blanc-Mathieu R, Song C, Kayama Y, Mochizuki T, Murata K, Ogata H, Takemura M. 2019. Medusavirus , a Novel Large DNA Virus Discovered from Hot Spring Water. *J. Virol.* 93:1–25.

Yutin N, Koonin E V. 2009. Evolution of DNA ligases of Nucleo-Cytoplasmic Large DNA viruses of eukaryotes: A case of hidden complexity. *Biol. Direct* 4:1–14.

Funding

This work is supported by a European Research Council (ERC) grant from the European Union’s Seventh Framework Program (FP/2007-2013)/ Project EVOMOBIL-ERC Grant Agreement no. 340440.

Acknowledgments:

Author contributions:

Conceptualization: JG, MG, VDC, PF

Methodology: JG, AC, PF

Investigation: JG, MG, PF

Supervision: AC, MK, PF

Writing – original draft: JG, PF

Writing – review & editing: JG, MG, VDC, AC, MK, PF

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: All data are available in the main text or upon request.