TITLE PAGE

Full-length title: A metallo-beta-lactamase with both beta-lactamase and ribonuclease activity is linked with traduction in giant viruses

Short title (for the running head): Metallo-beta-lactamase fold in giant viruses

Author list: Philippe COLSON[1,2], Lucile PINAULT[2], Said AZZA[2], Nicholas ARMSTRONG[2], Eric CHABRIERE[1,2], Bernard LA SCOLA[1,2], Pierre PONTAROTTI[1,3], Didier RAOULT[1,2] *

Affiliations: [1] Aix-Marseille Univ., Institut de Recherche pour le Développement (IRD), Assistance Publique - Hôpitaux de Marseille (AP-HM), MEPHI, 27 boulevard Jean Moulin, 13005 Marseille, France; [2] IHU Méditerranée Infection, 19-21 boulevard Jean Moulin, 13005 Marseille, France; [3] CNRS, Marseille, France

*** Corresponding author:** Prof. Didier Raoult, IHU - Méditerranée Infection, 19-21 boulevard Jean Moulin, 13005 Marseille, France. Tel.: +33 413 732 401, Fax: +33 413 732 402; email: didier.raoult@gmail.com

Key words: Giant virus; mimivirus; tupanvirus; beta-lactamase; ribonuclease

Word counts: abstract, 269; text, 3,461

References: 41

Figures: 5; **Tables**: 0

ABSTRACT

Enzymatic proteins with a metallo-beta-lactamase (MBL) fold have been essentially studied in bacteria for their activity on beta-lactam antibiotics. However, the MBL fold is ancient and highly conserved, and these proteins are capable of cleaving a broad range of substrates. It has recently been shown that MBLs are present in a wide array of cellular organisms, including eukaryotes and archaea. We show here that Tupanvirus deep ocean, a giant virus, also encodes a protein with a MBL fold. Phylogeny showed its clustering with transfer ribonucleases (RNases) and the presence of orthologs in other giant viruses, mainly those harboring the largest sets of translation components. In addition, it suggests an ancient origin for these genes and a transfer between giant viruses and *Acanthamoeba* spp., a host of many giant viruses. Biologically, after its expression in *Escherichia coli*, the tupanvirus protein was found to hydrolyse nitrocefin, a chromogenic beta-lactam. We also observed an hydrolysis of penicillin G (10 µg/mL) and detected the metabolite of penicillin G hydrolysis, benzylpenilloic acid. This was inhibited by sulbactam, a beta-lactamase inhibitor. In addition, we tested the degradation of single-stranded DNA, double-stranded DNA, and RNAs, and observed a strong activity on RNAs from seven bacteria with G+C varying from 42% to 67%, and from *Acanthamoeba castellanii*, the tupanvirus host. This was not inhibited by sulbactam or ceftriaxone. RNase activity was estimated to be 0.45±0.15 mU/mg using a fluorescence-based assay. Our results still broaden the range of hosts of MBL fold proteins and demonstrate that such protein can have dual beta-lactamase/nuclease activities. We suggest that they should be annotated according to this finding to avoid further confusion.

TEXT

INTRODUCTION

The metallo-beta-lactamase (MBL) superfamily encompasses a large set of enzymes, including MBL and ribonuclease (RNase) Z enzymes[1]. These enzymes are pleitropic proteins that can hydrolyze a wide range of substrates, among which beta-lactams, and DNA or RNA[2,3]. Such capabilities rely on an ancient and highly conserved fold, which represents a stable scaffold that has evolved to perform a broad range of chemical reactions and on which various catalytic, regulatory and structural activities are based[2-4]. This wide array of activities is enabled by variations in the composition and size of loops located near the enzyme active site[3]. A well-known catalytic activity of MBLs consists in breaking beta-lactam rings, which was primarily identified in bacteria[5]. Nevertheless, this hydrolase activity is suspected to have evolved in response to the environmental beta-lactams from an ancestral protein whose function was not related to beta-lactams and which may have been devoid of such hydrolase capability[3]. Concurrently to their capability to interact with various substrates that likely emerged through adaptive evolution, members of the MBLs superfamily have been identified in a broad range of cellular organisms, including bacteria, but also eukaryotes and archaea with a beta-lactamase activity[2,6].

Giant viruses are *bona fide* microbes as their virions are visible under a light microscope and they display a complexity similar to that of small cellular microorganisms[7,8]. Since their discovery in 2003, their diversity has increased considerably, with nine families and more than 100 isolates cultured. Their classification alongside cellular microorganisms is still debated, but their characteristics clearly distinguish them from conventional viruses[9,10]. We have investigated whether genes encoding members of the MBLs superfamily may also be present in giant viruses. We found one in Tupanvirus deep ocean, a giant mimivirus

isolated from Brazilian Atlantic ocean sediments, and confirmed that its product harbored a biologically active MBL fold with both beta-lactamase and RNase activities.

RESULTS

While annotating the genome of Tupanvirus deep ocean, the second isolate of a new mimivirus genus, *Tupanvirus*[11], a gene (GenBank: AUL78925.1) that encodes a metallo-hydrolase-like MBL fold was identified (Conserved Protein Domain Family Accession no. cl23716)[12]. This gene has a homolog in the other tupanvirus isolate (Soda Lake) (AUL77644.1). Beyond, best BLASTp hits against cellular organisms included MBL fold harboring proteins from an unclassified deltaproteobacterium whose genome was assembled from a marine water metagenome (evalue, 5e-38; identity, 33.0; coverage, 83%), from an actinobacteria (*Nonnomuraea* spp.) (1e-36; 30.0; 86%), from *Microscilla marina* (6e-34; 28.5%; 89%) and from *Acanthamoeba castellanii* (4e-33; 29.8%; 81%) (***Fig. 1, Supplementary Fig. S1***). Significant BLASTp hits (evalues ranging from 1e-41 to 8e-6) against viruses were also obtained with genes from putative giant viruses whose genomes was assembled from metagenomes obtained from environmental samples[13-15] and from Cafeteria roenbergensis virus, a distant Mimivirus relative[16]. The 322 amino acid long tupanvirus protein exhibits the conserved MBL motif "HxHxDH" in amino acid positions 60-65. A search for domains using the NCBI conserved domain search (CD Search) tool[17] identified a MBL fold belonging to a ribonuclease Z (RNase_Z_T_toga, TIGR02650, interval= amino acids 24-273, E-value= 1.81e-14; RNaseZ_ZiPD-like_MBL-fold, cd07717, interval= amino acids 56-282, E-value= 1.63e-04), which is a transfer RNA (tRNA)-processing endonuclease. This Tupanvirus deep ocean protein was further analyzed by three-dimensional comparison against the Phyre2 web portal for protein modeling, prediction and analysis[18]. This analysis reported a best match with 100% confidence and 85% coverage (273 amino acid residues)

with the crystal structure of a long form ribonuclease Z (RNase Z) from yeast (template c5mtzA) (***Supplementary Fig. S2***). Proteome analysis conducted for Tupanvirus Soda Lake, and Tupanvirus deep ocean, as previously described[12], did not allow the detection of these proteins with a MBL fold in the virions. In addition, the dramatic RNA shutdown observed during the replication of this giant virus hindered the achievement of transcriptomic analyses. Interestingly, the genomes of 20 of the 21 (95%) giant viruses found to encode a MBL fold protein concurrently encode tRNAs, whereas this is only the case for 46 of the 122 (38%) giant viruses devoid of a MBL fold protein ($p<10^{-3}$; Yates-corrected chi-square test) (***Supplementary Fig. S3 and 4 and Table S2***). The presence of a MBL fold protein among Megavirales members was correlated with the size of the gene repertoire and the number of translation-associated components ($p<10^{-3}$; Anova test). Putative proteins with a MBL fold from giant viruses comprised two related phylogenetic clusters (***Fig. 1***). These clusters appeared deeply rooted in the phylogenetic tree, which suggests an ancient origin for these genes. In addition, one of the clusters of giant virus genes encoding MBL fold proteins appeared closely related to two genes from *Acanthamoeba castellanii*, an amoebal mimivirus hosts, suggesting a transfer from these giant viruses to *A. castellanii*.

The recombinant protein AUL78925.1 of Tupanvirus deep ocean (named TupBlac) was expressed in *Escherichia coli* and was then purified, as described previously[6]. Based on the phylogenetic analysis and as MBL folds can hydrolyse nucleic acids[2], both beta-lactamase and nuclease activities of this purified protein were thereafter tested. We first evaluated the beta-lactamase activity of a pure solution of TupBlac used at a concentration of 1 µg/mL by incubating it with nitrocefin, a chromogenic beta-lactam used to test the beta-lactamase activity[19]. A significant hydrolysis activity was observed (***Fig. 2***). A concentrate of protein extract (50 mg/mL) obtained from tupanvirus virions also degraded, albeit slightly, nitrocefin. Thereafter, we monitored by liquid chromatography-mass spectrometry the effect of TupBlac

on penicillin G (10 µg/mL) and observed a significant hydrolysis activity of this coumpound within 48h (***Fig. 3***). We also detected, in the presence of the tupanvirus protein, benzylpenilloic acid, the metabolite resulting from the enzymatic hydrolysis of penicillin G[20]. Finally, we confirmed that these observations were related to a beta-lactamase activity as both penicillin G degradation and benzylpenilloic acid appearance were inhibited by sulbactam, a beta-lactamase inhibitor (***Fig. 3***). We further tested if pre-treatment with sulbactam had an impact on the duration of the giant virus replication cycle and replication intensity. After replication on *A. castellanii* strain Neff in the presence of a high concentration (10 µg/mL) of sulbactam, the virions produced (10^6/mL) were inoculated on fresh amoebae at different concentrations. No differences were observed regarding viral growth in the absence or presence of pre-treatment with sulbactam as assessed using high content screening (***Supplementary Fig. S5***).

Finally, as some proteins with a MBL fold can hydrolyse DNA and RNA[2], we tested the capability of tupanvirus enzyme TupBLac to degrade synthetic single- and double-stranded DNAs and bacterial RNAs. We found no effect on both DNA types. In contrast, we observed a strong RNase activity (***Fig. 4***). Another set of experiments was conducted using *E. coli* RNA as a substrate with an assessment of RNA size distribution on a bioanalyzer (Agilent Technologies, Palo Alto, CA) after incubation with TupBlac. It showed a dramatic degradation of RNAs by the tupanvirus enzyme (***Fig. 5a***). In contrast with the beta-lactamase activity, this was not inhibited, neither by sulbactam (***Fig. 5a*** and ***Supplementary Fig. S6***), nor by ceftriaxone (***Fig. 5b***), a cephalosporin that inhibits human SNM1A and SNM1B, that are DNA repair nucleases with a MBL fold[21]. In addition, a RNase activity of the Tupanvirus protein was further observed on *A. castellanii* RNA, and not inhibited either by sulbactam or ceftriaxone (***Fig. 5c***). TupBLac also degraded RNA extracted from bacteria with genomes with different G+C contents ranging between 41.8% and 66.6% (***Fig. 5d***), suggesting an

absence of influence of the G+C richness on the RNase activity. Finally, TupBlac RNase activity was estimated to be 0.451±0.153 mU/mg using a fluorescence-based assay, without difference in the presence of sulbactam or ceftriaxone (0.520±0.003 and 0.551±0.024 mU/mg, respectively) (***Supplementary Fig. S7***).

DISCUSSION

Hence, we found herein by several bioinformatic approaches that a gene of Tupanvirus deep ocean, a recently discovered giant virus classified in family *Mimiviridae*[11,12], encodes for a protein with a MBL fold. We further observed that this protein exhibited dual beta-lactamase and RNase activities. This is the first evidence of the presence of a biologically-active protein with a MBL fold in a virus. In addition, this work parallels the one on a protein detected by functional screening of a metagenomic library from the deep-seep sediments[22], showing that the same enzyme has both beta-lactamase and RNase activities. Indeed, MBL fold proteins were previously biologically-tested for either activity, but not for both. It is noteworthy that the beta-lactamase activity of the MBL fold protein of Tupanvirus was inhibited by a beta-lactamase inhibitor but this was not the case for the RNase activity[23]. The phylogenetic study of this beta-lactamase shows the presence in several other giant viruses of phylogenetically-clustered counterparts, the origin of which seems very old. Interestingly, it also appears that there may have been a gene transfer between these giant viruses and *Acanthamoeba* sp., the amoebal host of many giant viruses. Such potential for horizontal transfer of these MBL fold proteins is well-recognized [3].

Beta-lactamases are *a priori* useless for giant viruses, which are grown in the presence of various antibiotics, including beta-lactams[24], but our findings enhance the recent reconsideration of the function of MBL fold proteins. Thus, the recent description of penicillin secretion by arthropods [25] and the demonstration of active beta-lactamase in

172 vertebrates including humans[19], as well as in archaea[6] and fungi [26] show that MBL fold
173 proteins have a dramatically broad distribution. In humans, 18 genes were annotated as beta-
174 lactamases, whose activity had not been biologically-tested until recently[19]. In addition, MBL
175 fold proteins were highlighted to digest DNA or RNA[2,19]. Thus, a class of enzymes, that were
176 named beta-lactamases because of their original discovery in bacteria resistant to beta-
177 lactamines, are in fact potentially versatile proteins. This differs from the drastically-
178 simplified paradigm consisting in enzymes with a beta-lactamase activity being secreted by
179 bacteria under the selective pressure of natural or prescribed antibiotics.
180 The RNase activity observed here for the Tupanvirus MBL fold protein could be
181 related to the host ribosomal shutdown observed in the presence of Tupanvirus deep ocean
182 with various protists, the mechanism of which has not been elucidated[12]. This activity could
183 allow these viruses to take over on their cellular hosts by degrading cellular messenger RNAs
184 and shutting down cellular gene expression. The giant virus mRNAs should be protected from
185 such a degradation, which may be explained by the encapsidation of RNA transcripts into
186 giant virions that was detected for some of these viruses[27]. Bioinformatic analyses suggested
187 that the tupanvirus MBL fold protein may belong to the RNase Z group that was proposed to
188 be one of the two main groups of the MBL superfamily with that encompassing MBLs[1].
189 RNase Z enzymes perform tRNA maturation by catalyzing the endoribonucleolytic removal
190 of the 3' extension of tRNA precursors that do not contain a chromosomally-encoded CCA
191 determinant[28-30]. The presence in giant viruses of RNases showing the greatest homology to
192 tRNases suggests a specific activity on tRNAs, which seems consistent with the presence of a
193 large set of translation components in these viruses, first and foremost Tupanvirus deep ocean
194 that is the current record holder of the number of translation components (including 70 tRNAs
195 targeting all 20 canonical amino acids). The presence of a putative tRNase in the virus that
196 currently has the most complete set of translation components of the whole virosphere is

likely not fortuitous. Furthermore, it was described for *Escherichia coli* that its RNase Z had endoribonucleasic activity on messager RNAs, being responsible for their decay in *in vitro* experiments[29]. This further argues that MBL fold proteins may contain a wide range of activities. PNGM-1, a MBL fold protein whose sequence was recently described from a deep-sea sediment metagenome by detection of its beta-lactamase activity [31], was also found to harbor dual beta-lactamase and RNase activities [22]. MBL fold proteins from giant viruses are clustered with this protein in the phylogenetic analysis. Interestingly, PNGM-1 was suspected to have evolved from a tRNase Z[22]. In conclusion, our data still broaden the range of biological hosts of MBL fold proteins and demonstrate that such proteins can display dual beta-lactamase and nuclease activities. Therefore, we reannotated the tupanvirus MBL fold protein as a beta-lactamase/nuclease.

MATERIALS AND METHODS

Bioinformatics

Searches for Tupanvirus deep ocean protein AUL78925.1 homologs were performed using the BLAST tool[32]. Phylogeny reconstruction was performed after amino acid sequence alignment with the Muscle program[33] and the Maximum-Likelihood method using FastTree[34], and tree visualization used MEGA 6 software[35]. The amino acid sequences analyzed are Tupanvirus deep ocean protein AUL78925.1 and its homologs with the greatest BLASTp scores from the NCBI GenBank protein sequence database, our sequence database of giant virus genomes, and previously described draft genome sequences from 14 *Acanthamoeba* species[36] ; a set of previously described MBL fold proteins[19]; and a set of sequences from the UniProtKB database[1] previously used for phylogeny reconstructions. Three-dimensional comparisons for protein modeling, prediction and analysis were carried out against the Phyre2 web portal[18]. The set of translation components from each representative of the proposed

order Megavirales[37] was obtained through a BLASTp search[32] with their repertoire of predicted proteins against clusters of orthologous groups of proteins (COGs) involved in translation (category J)[38], using 10^{-4} and 50 amino acids as thresholds for e-values and sequence alignment lengths, respectively. The set of tRNAs from each virus was collected using the ARAGORN online tool (http://130.235.244.92/ARAGORN/)[39]. Hierarchical clustering was performed using the MultiExperiment Viewer software[40] based on the patterns of presence/absence of MBL fold protein, numbers of translation-associated components (number of tRNAs, aminoacyl tRNA-synthetases, other tRNA-associated proteins, other translation-associated proteins) and size of the gene repertoires for Megavirales members (***Supplementary Table S2***). For each item, the maximum value was determined, and values for each virus were considered relatively to these maximum values, being therefore comprised between 0 and 100%.

Cloning, expression and purification

The Tupanvirus deep ocean gene bioinformatically predicted to encode a beta-lactamase superfamily domain (AUL78925.1[12]) was designed to include a Strep-tag at the N-terminus and optimized for its expression by *Escherichia coli*. It was synthetized by GenScript (Piscataway, NJ, USA) and ligated between the NdeI and NotI restriction sites of a pET24a(+) plasmid. *E. coli* BL21(DE3)-pGro7/GroEL (Takara Shuzo Co., Kyoto, Japan) grown in ZYP-5052 media were used for the expression of the recombinant protein. When the culture reached an $O.D._{600\ nm}$= 0.6 at 37°C, the temperature was lowered to 20°C and L-arabinose (0.2% m/v) was added in order to induce the expression of chaperones. After 20 hours, cells were harvested by centrifugation (5,000 g, 30 min, 4°C) and the pellet was resuspended in washing buffer (50 mM Tris pH 8, 300 mM NaCl) and then stored at -80°C overnight. Frozen *E. coli* were thawed and incubated on ice for 1 hour after having added lysozyme, DNAse I and PMSF (phenylmethylsulfonyl fluoride) to final concentrations of 0.25 mg/mL, 10 µg/mL

and 0.1 mM, respectively. Partially lysed cells were then disrupted by 3 consecutive cycles of sonication (30 seconds, amplitude 45) performed on a Q700 sonicator system (QSonica). Cellular debris were discarded following a centrifugation step (10,000 g, 20 min, 4°C). The Tupanvirus protein was purified with an ÄKTA avant system (GE Healthcare, Bucks, UK) using Strep-tag affinity chromatography (wash buffer: 50 mM Tris pH 8, 300 mM NaCl, and elution buffer: 50 mM Tris pH 8, 300 mM NaCl, 2.5 mM desthiobiotin) on a 5 mL StrepTrap HP column (GE Healthcare). Fractions containing the protein of interest were pooled. Protein purity was assessed using 12.5% SDS-PAGE analysis (Coomassie staining). Protein expression was confirmed by performing MALDI-TOF MS analysis on gel bands previously obtained by SDS-PAGE. Protein concentrations were measured using a Nanodrop 2000c spectrophotometer (Thermo Scientific, Madison, WI, USA).

Spectrophotometry assay for the detection of beta-lactamase activity in Tupanvirus virions

Tupanvirus purified virions in solution were centrifuged at 5,000 RPM in order to collect 1g of humid matter. Virions were then suspended into 2 mL of a phosphate-buffered saline (PBS) solution at pH 7.4 prepared in water from a commercial salt mixture (bioMerieux, Marcy-l'Etoile, France). Virions were broken after five freeze-thaw cycles followed by 10 minutes of sonication (Q700 sonicator with a Cup Horn, QSonica, Newtown, Connecticut, USA). Integrity of virions was checked by scanning electron microscopy (TM 4000, Hitachi High-Technologie Corporation, Tokyo, Japan). Debris were discarded following a centrifugation step (15,000 g, 10 minutes). The clear supernatant was lyophilized and then reconstituted in 100 µL of PBS (corresponding to a final concentration of 50 mg/mL of soluble proteins). A pure solution of Tupanvirus protein was buffer-exchanged in PBS and the concentration was adjusted to 1 mg/ml. The degradation of nitrocefin (1 mM in PBS), a chromogenic cephalosporin substrate, was monitored as previously described after the

addition of virion protein extract or Tupanvirus protein to the solution[6].

Beta-lactam antibiotic degradation monitoring by liquid chromatography-mass spectrometry (LC-MS)

Penicillin G and sulbactam stock solutions at 10 mg/mL were freshly prepared in water from the corresponding high purity salts (Sigma Aldrich). A total of 30 µL of tupanvirus protein solution at 1 mg/mL was spiked with penicillin G and sulbactam at a final concentration of 10 µg/mL, before incubation at room temperature. Each time point corresponded to triplicate sample preparations. Negative controls consisted of PBS spiked with penicillin G and sulbactam. Then, 70 µL of acetonitrile were added to each sample, and tubes were vortexed 10 minutes at 16,000 g to precipitate the proteins. The clear supernatant was collected for analysis using an Acquity I-Class UPLC chromatography system connected to a Vion IMS Qtof ion mobility-quadrupole-time of flight mass spectrometer, as previously described [6].

Assessment of the effect of a beta-lactamase inhibitor on Tupanvirus growth

To evaluate the effect of a beta-lactamase inhibitor sulbactam on Tupanvirus growth, we tested Tupanvirus replication on *A. castellanii* pre-incubated with a high dose of sulbactam. Tests were performed in triplicate and amoebae cultivated in trypticase soy medium[12]. Four 1 mL culture wells containing 5.10^5 *A. castellanii* were incubated at 32°C, one of which contained 500 mg/L of sulbactam. After 24 hours, Tupanvirus was added at a multiplicity of infection (MOI) of 1 in the well with sulbactam. Two other wells were inoculated with Tupanvirus, including one in which 500 mg/L of sulbactam was added. The last well was used as control of amoeba survival. After 24h, amoebae were counted and Tupanvirus was titrated by qPCR as previously described[12]. In order to assess whether sulbactam could have affected newly formed virions, tupanviruses produced on amoebae incubated with sulbactam were inoculated on fresh amoebae at different concentrations. Their growth was monitored using high content screening microscopy every 8h for 48h[41]. Viral replication was compared

297 to that of tupanviruses produced on amoebae non-treated with sulbactam at the same MOIs.

298 **Nuclease activity assessment**

299 Nuclease activity was assessed using double-stranded DNA, (+) and (-) single-stranded
300 DNAs, and single-stranded RNAs as substrates. Single-stranded DNAs were synthetic
301 polynucleotides (***Supplementary Table S3***); double-stranded DNA was obtained by annealing
302 (+) and (-) single-stranded DNAs in a thermocycler at temperatures decreasing from 95°C to
303 25°C over 1h. RNAs used as substrate were from *Escherichia coli*, from different bacteria that
304 differ by the G+C content of their genomes (*Streptococcus parasanguinis* (41.8%), *Vibrio*
305 *parahaemolyticus* (45.4%), *Vitreoscilla massiliensis* (49.4%), *Aeromonas salmonicida*
306 (58.5%), *Aeromonas hydrophila* (61.5%) and *Pseudomonas aeruginosa* (66.6%)), and from
307 *Acanthamoeba castellanii*. RNAs were purified using RNeasy columns (Invitrogen, Carlsbad,
308 CA, USA). Enzymatic reactions were performed by incubating each polynucleotide (2 µg)
309 with 15 µg of the expressed Tupanvirus protein TupBlac in Tris-HCl buffer 50 mM, pH 8.0,
310 sodium chloride 0.3 M, using a final volume of 20 µL at 30°C for 2 h. After incubation, the
311 material was loaded onto denaturing polyacrylamide gel electrophoresis (dPAGE) at 12% or
312 analysed using the Agilent RNA 6000 Pico LabChip kit on an Agilent 2100 Bioanalyzer
313 (Agilent Technology, Palo Alto, CA, USA). Controls were carried out under the same
314 conditions. The action of TupBlac on RNAs was also assayed in the presence of ceftriaxone,
315 an inhibitor of human metallo β-lactamase fold DNA repair nucleases SNM1A and SNM1B[21].
316 To do this, enzymatic reactions were conducted at 30°C by incubating *E. coli* RNA (1 µg)
317 with TupBlac (15 µg) in the presence of ceftriaxone at 200 µM. At different times, reactions
318 were stopped by addition of proteinase K (10 µg) and incubated 1h at 37°C. For a quantitative
319 assessment of the RNase activity of the TupBlac enzyme, we used the RNaseAlert QC
320 System kit (Fisher Scientific, Illkirch, France) according to the manufacturer's protocol. This
321 assay uses as substrate a fluorescence-quenched oligonucleotide probe that emits a fluorescent

322 signal in the presence of RNase activity. RNase activities were assayed in the absence or
323 presence of sulbactam (10 µg/mL) or ceftriaxone (200 µM). Negative controls were made
324 with all the reagents used (RNase free water, enzyme buffer, sulbactam and ceftriaxone).
325 Fluorescence was monitored continuously at 37°C for 1h by a Synergy HT plate reader
326 (BioTek Instruments SAS, Colmar, France) with a 485/528 nm filter set. RNase activities of
327 TupBlac were estimated using supplied RNase A used as a standard (10 mU/mL). Two
328 independent experiments were conducted.
329
330
331 **Acknowledgments** This work was supported by the French Government under the
332 "Investments for the Future" program managed by the National Agency for Research (ANR),
333 Méditerranée-Infection 10-IAHU-03, and was also supported by Région Provence-Alpes-Côte
334 d'Azur and European funding FEDER PRIMMI (Fonds Européen de Développement
335 Régional - Plateformes de Recherche et d'Innovation Mutualisées Méditerranée Infection).
336 We are thankful to Rania Francis for her technical help.
337
338 **Competing interest** The authors declare no competing interests. Funding sources had no role
339 in the design and conduct of the study; collection, management, analysis, and interpretation of
340 the data; and preparation, review, or approval of the manuscript.
341
342 **Author contributions** Conceived and designed the study: DR, PC, PP, BLS. Designed and/or
343 performed experiments: DR, PC, LP, SA, NA, EC, BLS, PP. Analyzed and interpreted data:
344 DR, PC, EC, BLS, PP. Wrote the manuscript: PC and DR. All authors read and approved the
345 final manuscript.
346

REFERENCES

1. Alderson, R. G., Barker, D., Mitchell, J. B. One origin for metallo-beta-lactamase activity, or two? An investigation assessing a diverse set of reconstructed ancestral sequences based on a sample of phylogenetic trees. *J. Mol Evol.* **79**, 117 (2014).

2. Aravind, L. An evolutionary classification of the metallo-beta-lactamase fold proteins. *In Silico. Biol.* **1**, 69 (1999).

3. Bebrone, C. Metallo-beta-lactamases (classification, activity, genetic organization, structure, zinc coordination) and their superfamily. *Biochem. Pharmacol.* **74**, 1686 (2007).

4. Caetano-Anolles, G. and Caetano-Anolles, D. An evolutionarily structured universe of protein architecture. *Genome Res.* **13**, 1563 (2003).

5. Abraham, E. P. and Chain, E. An enzyme from bacteria able to destroy penicillin. Abraham, E. P. and Chain, E.8371940), p.837.

6. Diene, S. M., *et al.* Paradoxical ß-Lactamase activity of archaeal encoding enzymes. *BioRxiv* (2019). doi: https://doi.org/10.1101/667907

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

8. La Scola, B., *et al.* A giant virus in amoebae. *Science* **299**, 2033 (2003).

9. Raoult, D. and Forterre, P. Redefining viruses: lessons from Mimivirus. *Nat. Rev. Microbiol.* **6**, 315 (2008).

10. Raoult, D. TRUC or the need for a new microbial classification. *Intervirology.* **56**, 349 (2013).

11. Rodrigues, R. A. L., Mougari, S., Colson, P., La Scola, B., Abrahao, J. S. "Tupanvirus", a new genus in the family Mimiviridae. *Arch Virol* **164**, 325-331 (2018).

12. Abrahao, J., *et al.* Tailed giant Tupanvirus possesses the most complete translational apparatus of the known virosphere. *Nat. Commun.* **9**, 749 (2018).

13. Schulz, F., *et al.* Giant viruses with an expanded complement of translation system Components. *Science* **356**, 82-85 (2017).

14. Schulz, F., *et al.* Hidden diversity of soil giant viruses. *Nat. Commun.* **9**, 4881 (2018).

15. Backstrom, D., *et al.* Virus Genomes from deep sea sediments expand the ocean megavirome and support independent origins of viral gigantism. *MBio.* **10**, mBio-18 (2019).

16. Fischer, M. G., Allen, M. J., Wilson, W. H., Suttle, C. A. Giant virus with a remarkable complement of genes infects marine zooplankton. *Proc. Natl. Acad. Sci. U. S. A* **107**, 19508 (2010).

17. Marchler-Bauer, A., *et al.* CDD: conserved domains and protein three-dimensional structure. *Nucleic Acids Res.* **41**, D348-D352 (2013).

18. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845 (2015).

19. Diene, S. M., *et al.* Human metallo-ß-lactamase enzymes degrade penicillin. *Sci. Rep.* **9**, 1273 (2019).

20. Deshpande, A. D., Baheti, K. G., Chatterjee, N. R. Degradation of beta-lactam antibiotics. *Curr. Sci.* **87**, 1684-1695 (2004).

21. Lee, S. Y., *et al.* Cephalosporins inhibit human metallo beta-lactamase fold DNA repair nucleases SNM1A and SNM1B/apollo. *Chem. Commun. (Camb.).* **52**, 6727 (2016).

394 22. Lee, H. J., *et al.* Dual activity of PNGM-1, a metallo-beta-lactamase and tRNase Z,
395 pinpoints the evolutionary origin of subclass B3 metallo-beta-lactamases. *BioRxiv*
396 (2019). doi: https://doi.org/10.1101/575373.
397 23. Baier, F. and Tokuriki, N. Connectivity between catalytic landscapes of the metallo-
398 beta-lactamase superfamily. *J Mol. Biol.* **426**, 2442 (2014).
399 24. Khalil, J. Y., Andreani, J., La Scola, B. Updating strategies for isolating and discovering
400 giant viruses. *Curr. Opin. Microbiol.* **31**, 80 (2016).
401 25. Suring, W., *et al.* Evolutionary ecology of beta-lactam gene clusters in animals. *Mol*
402 *Ecol.* **26**, 3217 (2017).
403 26. Gao, M., Glenn, A. E., Blacutt, A. A., Gold, S. E. Fungal lactamases: their occurrence
404 and function. *Front Microbiol.* **8**, 1775 (2017).
405 27. Raoult, D., *et al.* The 1.2-megabase genome sequence of Mimivirus. *Science* **306**, 1344
406 (2004).
407 28. Ma, M., *et al.* The crystal structure of Trz1, the long form RNase Z from yeast. *Nucleic*
408 *Acids Res.* **45**, 6209 (2017).
409 29. Perwez, T. and Kushner, S. R. RNase Z in *Escherichia coli* plays a significant role in
410 mRNA decay. *Mol. Microbiol.* **60**, 723 (2006).
411 30. Schierling, K., Rosch, S., Rupprecht, R., Schiffer, S., Marchfelder, A. tRNA 3' end
412 maturation in archaea has eukaryotic features: the RNase Z from Haloferax volcanii. *J*
413 *Mol. Biol.* **316**, 895 (2002).
414 31. Park, K. S., *et al.* PNGM-1, a novel subclass B3 metallo-beta-lactamase from a deep-sea
415 sediment metagenome. *J. Glob. Antimicrob. Resist.* **14**, 302 (2018).
416 32. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., Lipman, D. J. Basic local alignment
417 search tool. *J. Mol. Biol.* **215**, 403 (1990).

418 33. Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and
419 space complexity. *BMC. Bioinformatics.* **5**, 113 (2004).
420 34. Price, M. N., Dehal, P. S., Arkin, A. P. FastTree 2--approximately maximum-likelihood
421 trees for large alignments. *PLoS. One.* **5**, e9490 (2010).
422 35. Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S. MEGA6: Molecular
423 Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725 (2013).
424 36. Chelkha, N., *et al.* A Phylogenomic Study of *Acanthamoeba polyphaga* draft genome
425 sequences suggests genetic exchanges with giant viruses. *Front Microbiol.* **9**, 2098
426 (2018).
427 37. Colson, P., *et al.* "Megavirales", a proposed new order for eukaryotic nucleocytoplasmic
428 large DNA viruses. *Arch. Virol.* **158**, 2517 (2013).
429 38. Tatusov, R. L., *et al.* The COG database: an updated version includes eukaryotes. *BMC.*
430 *Bioinformatics.* **4**, 41 (2003).
431 39. Laslett, D. and Canback, B. ARAGORN, a program to detect tRNA genes and tmRNA
432 genes in nucleotide sequences. *Nucleic Acids Res.* **32**, 11 (2004).
433 40. Saeed, A. I., *et al.* TM4: a free, open-source system for microarray data management
434 and analysis. *Biotechniques.* **34**, 374 (2003).
435 41. Francis, R., Ominami, Y., Bou Khalil, J. Y., La Scola, B. High-throughput isolation of
436 giant viruses using high-content screening. *Commun. Biol.* **2**, 216 (2019).
437
438

FIGURES

Figure 1 | Phylogeny reconstruction based on metallo-beta-lactamase (MBL) fold proteins

Phylogeny reconstruction was performed after amino acid sequence alignment with the Muscle program[33] with the Maximum-Likelihood method using FastTree[34], and tree was visualized with the MEGA 6 software[35]. The amino acid sequences analyzed are Tupanvirus deep ocean protein AUL78925.1 and its homologs with the greatest BLASTp scores from the NCBI GenBank protein sequence database (nr) (see ***Supplementary Table S1***), our sequence database of giant virus genomes, and previously described draft genome sequences from 14 *Acanthamoeba* species[36]; a set of previously described MBL fold proteins[19]; and a set of sequences from the UniProtKB database[1], previously used for phylogeny reconstructions. Extended tree is available in ***Supplementary Figure S1.***


Tetraselmis virus 1 lcl | KY322437.1 prot AUF82241.1 149 locus tagTetV 149 proteinalkyl/aryl-sulfatase protein
Pandoravirus salinus KC977571 AGO84666 gene ps 1328 protein ribonuclease Z protein
Pandoravirus macleodensis NC 037665.1 prot YP 009481559.1 875 locus tagDP161 gp080 db xrefGeneID:36842018 protein ribonuclease Z protein
Pandoravirus neocaledonia NC 037666.1 prot YP 009482545.1 935 locus tagDP160 gp0149 db xrefGeneID:36843255 protein ribonuclease Z protein
Pandoravirus quercus NC 037667.1 prot YP 009483520.1 829 locus tagDP159 gp0357 db xrefGeneID:36844392 protein ribonuclease Z protein
Pandoravirus dulcis KC977570 AGO82932 gene pd 1076 protein ribonuclease Z protein
Pandoravirus pampulha-1647 GMS
AWN09461.1 metallo-beta-lactamase PNGM-1 uncultured bacterium
MBLSF set | P54548 | RNZ BACSU Ribonuclease Z Bacillus subtilis (strain 168) OX224308 GNrnz PE1 SV1
MBLSF set | P0A8V0 | RBN ECOLI Ribonuclease BN Escherichia coli (strain K12) OX83333 GNrbn PE1 SV1
BLset | Human | AAH14624 | RNaseZ ElaC1
BLset | Human | CAG33542 | RNaseZ N-term MBL-fold ELAC2
BLset | Human | AAH20199 | Cleavage and Polyadenylation Endonuclease CPSF3L
BHGbk | AYV85464.1 hypothetical protein Satyrvirus18 1 partial Satyrvirus sp.
BHGbk | QBK87355.1 beta-lactamase superfamily domain protein Marseillevirus LCMAC201
BHGbk | AYV80377.1 tRNaseZ Harvfovirus sp.
BHGbk | AYV83813.1 tRNaseZ Hyperionvirus sp.
MBLSF set | Q9WZW8 THEMA Uncharacterized protein Thermotoga maritima (strain ATCC 43589 / MSB8 / DSM 3109 / JCM 10099) OX243274 GNTM 0864 PE1 SV1
AUL78925.1:1-322 beta-lactamase superfamily domain Tupanvirus deep ocean
AUL77644.1:1-320 hypothetical protein Tupanvirus soda lake
BHGbk | XP 004341230.1:18-268 nuclear RNase Z putative Acanthamoeba castellanii str. Neff
BHGbk | TNE48039.1:41-307 MBL fold metallo-hydrolase Deltaproteobacteria bacterium
BHGbk | MBK05979.1:37-296 hypothetical protein Deltaproteobacteria bacterium
Yasminevirus gene 793 | 317 aa | - | 1025767 | 1026720
BHGbk | XP 004335117.1:47-321 hypothetical protein ACA1 098010 Acanthamoeba castellanii str. Neff
BHGbk | EAY28652.1:10-300 hypothetical protein M23134 07750 Microscilla marina ATCC 23134
BHGbk | WP 045113612.1:18-308 MBL fold metallo-hydrolase Microscilla marina
Streptomyces spp./Streptosporangiaceae bacterium
Nonomuraea spp./Sinosporangium album
BHGbk | WP 081970712.1:39-313 hypothetical protein Dactylosporangium aurantiacum
BHGbk | QBK86134.1 hypothetical protein LCMAC101 07290 Marseillevirus LCMAC101
BHGbk | ARF10191.1 ribonuclease BN tRNA processing enzyme Hokovirus HKV1
BHGbk | AYV82246.1 hypothetical protein Homavirus20 4 Homavirus sp.
BHGbk | AYV83529.1 beta-lactamase superfamily domain Hyperionvirus sp.
BHGbk | AYV81277.1 beta-lactamase superfamily domain Harvfovirus sp.
BHGbk | AYV76769.1 beta-lactamase superfamily domain Terrestrivirus sp.
BHGbk | YP 003970092.1 putative metal-dependent hydrolase Cafeteria roenbergensis virus BV-PW1
Nucleases
Acanthamoeba spp.
B3 metallo-beta-lactamases
B2 metallo-beta-lactamases
B1 metallo-beta-lactamases
A-type flavoproteins
Glyoxalases II
0.5

454

455 **Figure 2 | Effect on nitrocefin of expressed Tupanvirus protein (TupBlac).**

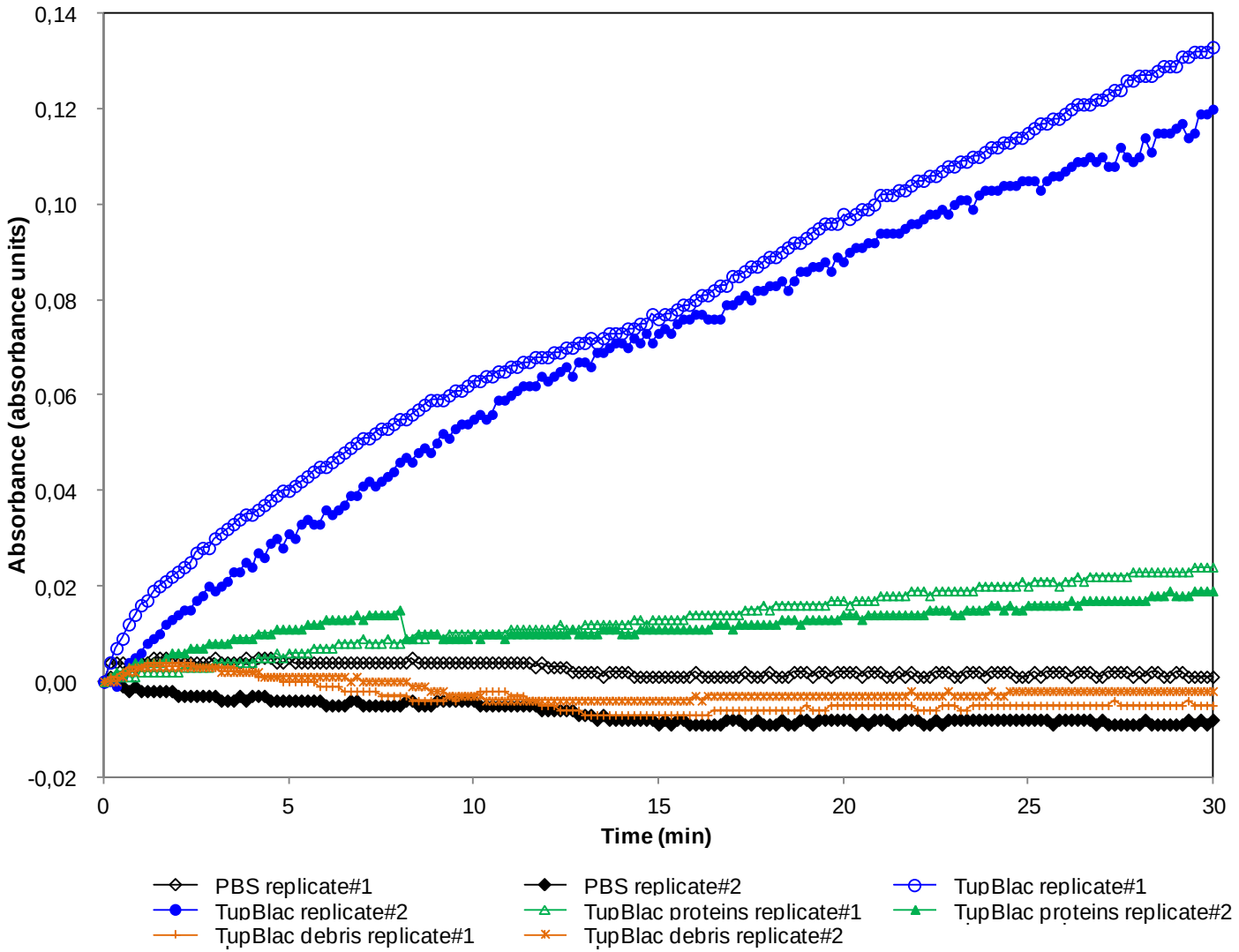
456 The effect on nitrocefin of the expressed Tupanvirus protein (TupBlac) was assessed by

457 monitoring the degradation of nitrocefin, a chromogenic cephalosporin substrate.

458 PBS, Phosphate-Buffered Saline; TupBlac, tupanvirus expressed protein

459

460


0,14
0,12
0,10
0,08
0,06
0,04
0,02
0,00
-0,02
Absorbance (absorbance units)
0
5
10
15
20
25
30
Time (min)
PBS replicate#1
PBS replicate#2
TupBlac replicate#1
TupBlac replicate#2
TupBlac proteins replicate#1
TupBlac proteins replicate#2
TupBlac debris replicate#1
TupBlac debris replicate#2

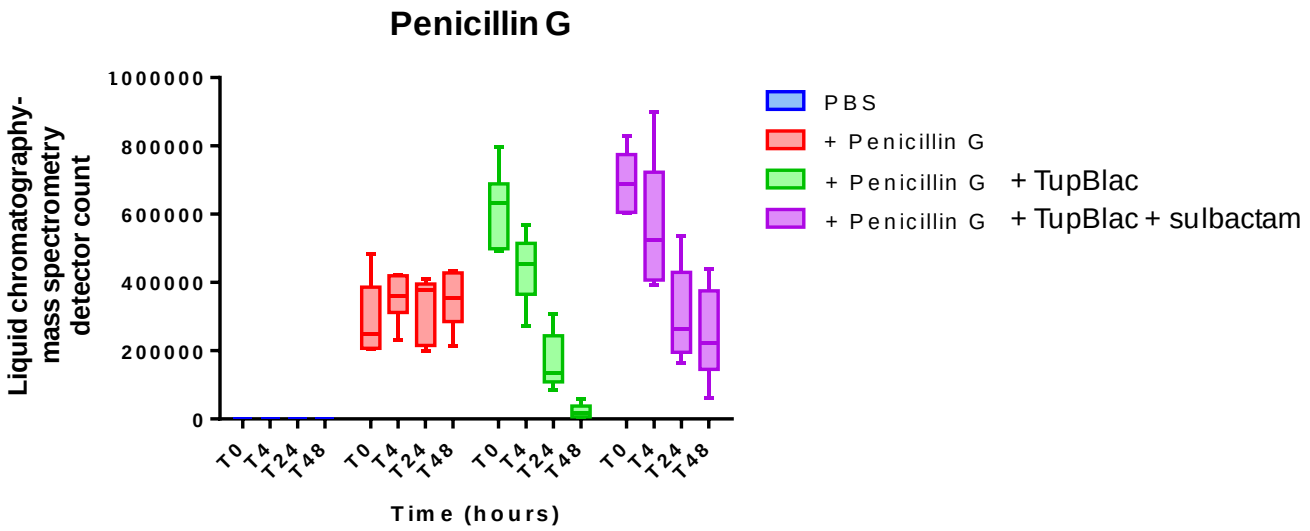
461

Figure 3 | Effect on penicillin G of expressed Tupanvirus protein (TupBlac).

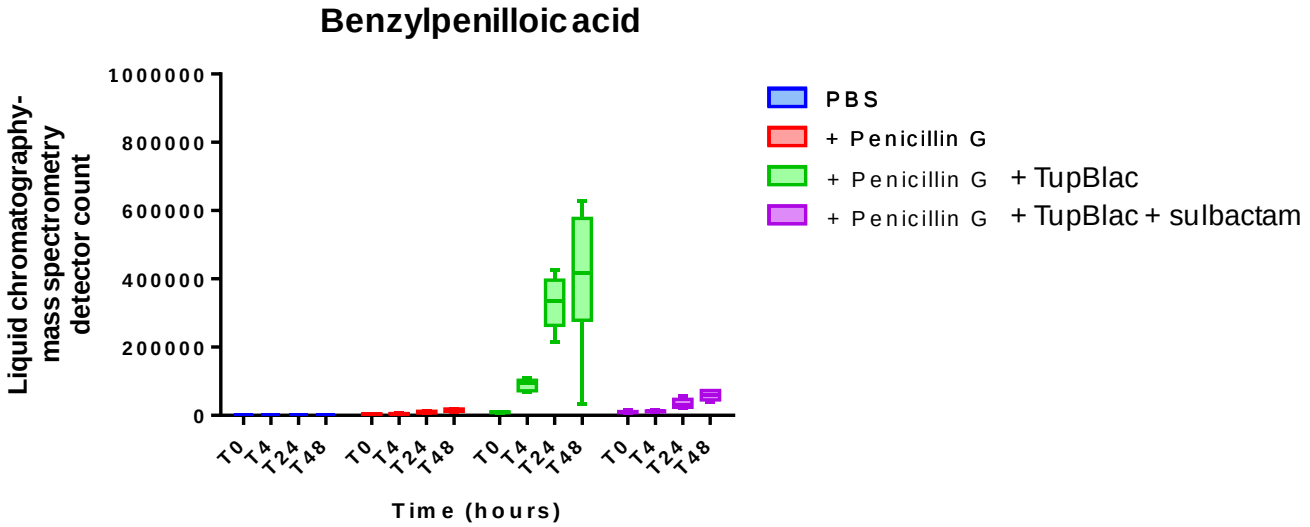
The effect on penicillin G of the expressed Tupanvirus protein (TupBlac) and its inhibition by sulbactam were assessed by monitoring by liquid chromatography-mass spectrometry (LC-MS) the degradation of penicillin G (a) and the appearance of benzylpenilloic acid, the metabolite resulting from the enzymatic hydrolysis of penicillin G (b), at times (T) T0 (0 hour), T4 (4 hours), T24 (24 hours), and T48 (48 hours).

PBS, Phosphate-Buffered Saline; TupBlac, tupanvirus expressed protein

a.


Penicillin G
Liquid chromatography-mass spectrometry detector count
1000000
800000
600000
400000
200000
0
T0 T4 T24 T48 T0 T4 T24 T48 T0 T4 T24 T48 T0 T4 T24 T48
Time (hours)
PBS
+ Penicillin G
+ Penicillin G + TupBlac
+ Penicillin G + TupBlac + sulbactam

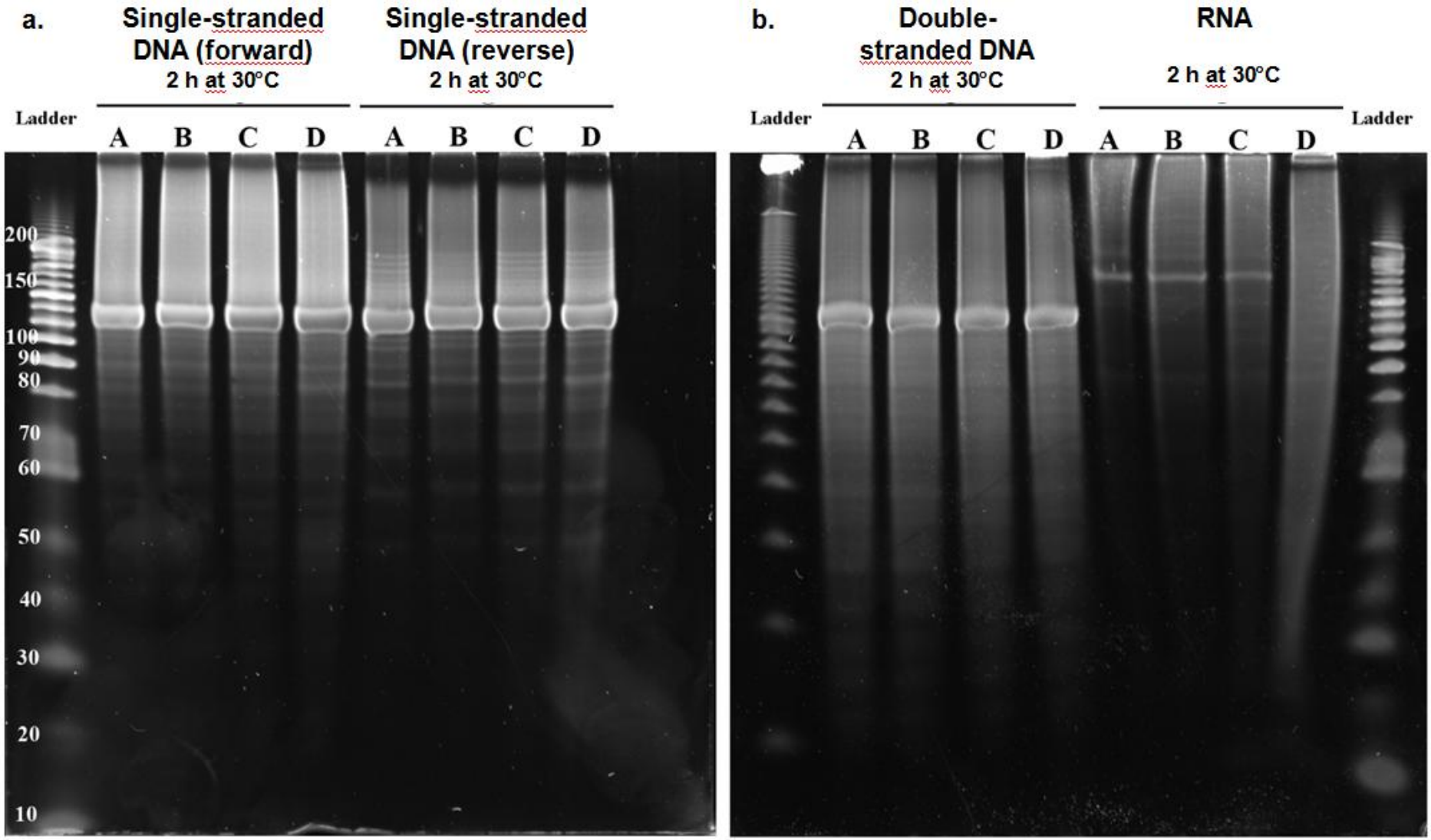
b.


Benzylpenilloic acid
Liquid chromatography-mass spectrometry detector count
1000000
800000
600000
400000
200000
0
T0 T4 T24 T48 T0 T4 T24 T48 T0 T4 T24 T48 T0 T4 T24 T48
Time (hours)
PBS
+ Penicillin G
+ Penicillin G + TupBlac
+ Penicillin G + TupBlac + sulbactam

471

Figure 4 | Nuclease activity on various types of nucleic acids of expressed Tupanvirus protein (TupBlac) as assessed by dPAGE.

Denaturant polyacrylamide gel electrophoresis (12% dPAGE) of nuclease activity on synthetic (+) and (-) single-stranded DNAs (130 nucleotide-long) (a), synthetic double-stranded DNA (b) (see also ***Supplementary Table S3*****),** or *Escherichia coli* RNA (b). No treatment (A); buffer (B); succinate dehydrogenase enzyme produced and purified by the same process and collected in the same fractions as Tupanvirus beta-lactamase TupBlac, used as negative control (C); Tupanvirus beta-lactamase TupBlac (D).


a.
Single-stranded DNA (forward)
2 h at 30°C
Single-stranded DNA (reverse)
2 h at 30°C
Ladder
A B C D A B C D
200
150
100
90
80
70
60
50
40
30
20
10
b.
Double-stranded DNA
2 h at 30°C
RNA
2 h at 30°C
Ladder
A B C D A B C D
Ladder

482

Figure 5 | Digital gel images of RNase activity of expressed Tupanvirus protein TupBlac on *E. coli* RNA.

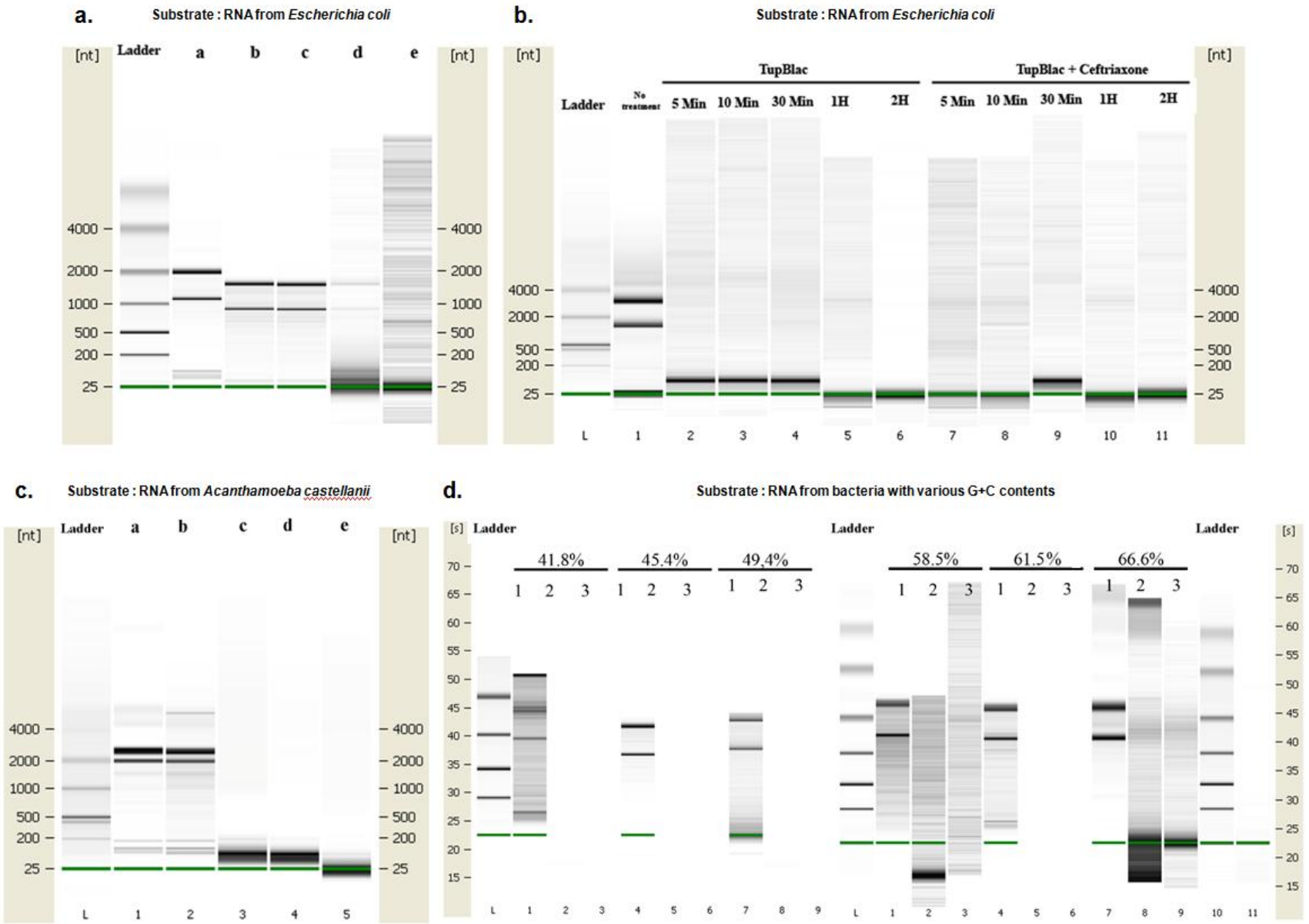
RNA samples (1 µg) were incubated with 15 µg of TupBlac at 30°C in the absence or presence of 10 µg/mL of sulbactam or 200 µM of ceftriaxone. Nuclease activity was visualized as digital gel images performed using the Agilent Bioanalyzer 2100 with the RNA 6000 Pico LabChip (Agilent Technologies, Palo Alto, CA).

a: RNA used as substrate was from *Escherichia coli;* no treatment (a); buffer (b); sulbactam (c); TupBlac in the absence (d) or presence (e) of sulbactam.

b: RNA used as substrate was from *Escherichia coli;* reactions were stopped at different times (5 min, 10 min, 30 min, 1 h and 2 h) by the addition of proteinase K (10 µg) and incubation for 1h at 37°C. The first lane corresponds to no treatment; lanes 2 to 6 to RNA treatment with TupBlac in the absence of ceftriaxone; lanes 7 to 11 to RNA treatment with TupBlac in the presence of ceftriaxone.

c: nuclease activity on RNAs originating from *Acanthamoeba castellanii*; no treatment (a); buffer (b); TupBlac in the absence (c, d) or presence (e) of sulbactam.

d: nuclease activity on RNAs originating from bacteria that differ by the G+C-content of their genome, as indicated at the top of the digital gel image. For each RNA, three samples were analyzed: no treatment (1); treatment with TupBlac in the absence of sulbactam (2); and treatment with TupBlac in the presence of sulbactam (3).


a.
Substrate : RNA from Escherichia coli
[nt]
Ladder
a
b
c
d
e
4000
2000
1000
500
200
25
b.
Substrate : RNA from Escherichia coli
TupBlac
TupBlac + Ceftriaxone
No treatment
5 Min
10 Min
30 Min
1H
2H
L
1
2
3
4
5
6
7
8
9
10
11
c.
Substrate : RNA from Acanthamoeba castellanii
d.
Substrate : RNA from bacteria with various G+C contents
[s]
41.8%
45.4%
49,4%
58.5%
61.5%
66.6%