



HAL
open science

ExoMeg1: a new exonuclease from metagenomic library

Rita C. B. Silva-Portela, F C Carvalho, Carolina P Pereira, Nadja P de Souza-Pinto, Mauro Modesti, Robert P. Fuchs, Lucymara F. Agnez-Lima

► **To cite this version:**

Rita C. B. Silva-Portela, F C Carvalho, Carolina P Pereira, Nadja P de Souza-Pinto, Mauro Modesti, et al.. ExoMeg1: a new exonuclease from metagenomic library. Scientific Reports, 2016, 6, pp.19712. 10.1038/srep19712 . hal-01456273

HAL Id: hal-01456273

<https://amu.hal.science/hal-01456273>

Submitted on 4 Feb 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

SCIENTIFIC REPORTS



OPEN

ExoMeg1: a new exonuclease from metagenomic library

Rita C. B. Silva-Portela¹, Fabíola M. Carvalho¹, Carolina P. M. Pereira², Nadja C. de Souza-Pinto², Mauro Modesti^{3,5,6}, Robert P. Fuchs^{4,5,6,7} & Lucymara F. Agnez-Lima¹

Received: 08 July 2015

Accepted: 27 November 2015

Published: 27 January 2016

DNA repair mechanisms are responsible for maintaining the integrity of DNA and are essential to life. However, our knowledge of DNA repair mechanisms is based on model organisms such as *Escherichia coli*, and little is known about free living and uncultured microorganisms. In this study, a functional screening was applied in a metagenomic library with the goal of discovering new genes involved in the maintenance of genomic integrity. One clone was identified and the sequence analysis showed an open reading frame homolog to a hypothetical protein annotated as a member of the Exo_Endo_Phos superfamily. This novel enzyme shows 3'-5' exonuclease activity on single and double strand DNA substrates and it is divalent metal-dependent, EDTA-sensitive and salt resistant. The clone carrying the hypothetical ORF was able to complement strains deficient in recombination or base excision repair, suggesting that the new enzyme may be acting on the repair of single strand breaks with 3' blockers, which are substrates for these repair pathways. Because this is the first report of an enzyme obtained from a metagenomic approach showing exonuclease activity, it was named ExoMeg1. The metagenomic approach has proved to be a useful tool for identifying new genes of uncultured microorganisms.

In nature, microorganisms live under a variety of environmental stress conditions induced by physical or chemical agents that, in addition to byproducts of cellular metabolism, are able to generate different DNA lesions, which if not repaired may induce mutations, genetic recombination and inhibition or modification of DNA replication and/or transcription. Considering these deleterious effects of DNA damage, it is predictable that all forms of life possess specialized mechanisms to recognize and repair different DNA lesions, favoring direct or indirect approaches for their adaptation and survival^{1,2}. In fact, considerable conservation of repair functions among different species of prokaryotes and eukaryotes has been observed, which may or may not be accompanied by corresponding conservation of protein sequence^{1,3,4}. However, knowledge of DNA repair mechanisms in bacteria is based on cultured microorganisms, mainly using the model organism *Escherichia coli*^{5,6,7}, which represents less than 1% of the microbial diversity estimated to be on Earth because more than 99% of microbial species have not been cultured using traditional methods^{8,9}.

To circumvent the limitations of culture methods, the metagenomic approach, defined as collective microbial genome analysis¹⁰, has been developed and improved to complement culture-dependent traditional approaches. This methodology allows for the extraction of DNA from environmental samples and subsequent cloning using a cultivable host strain as well as direct shotgun sequencing by next generation sequencing strategies. Numerous studies have confirmed that access to genetic patrimony of metagenomes provides a rich source for finding new genes and for understanding biological processes and microbial ecology¹¹⁻¹⁴.

Functional screens of metagenomic libraries have allowed for the discovery of new genes that encode proteins with either known or new functions because these screens are independent of sequence similarity to known genes. The selection of active clones expressing a desired phenotype is followed by sequencing and characterization by biochemical specific assays. The major advantage of this approach is that, in general, only functional gene products are identified. However, successful heterologous expression of recombinant proteins is necessary,

¹Departamento de Biologia Celular e Genética, Centro de Biociências, Universidade Federal do Rio Grande do Norte, Natal, RN, 59072-970, Brasil. ²Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo (USP), São Paulo, Brasil. ³Homologous Recombination, NHEJ and Maintenance of Genomic Integrity CNRS, UMR7258, Marseille, F-13009, France. ⁴Team DNA Damage Tolerance CNRS, UMR7258, Marseille, F-13009, France. ⁵Inserm, U1068, CRCM, Marseille, F-13009, France. ⁶Institut Paoli-Calmettes, Marseille, F-13009, France. ⁷Aix-Marseille University, UM 105, F-13284, Marseille, France. Correspondence and requests for materials should be addressed to L.F.A.-L. (email: lfagnez@ufnet.br)

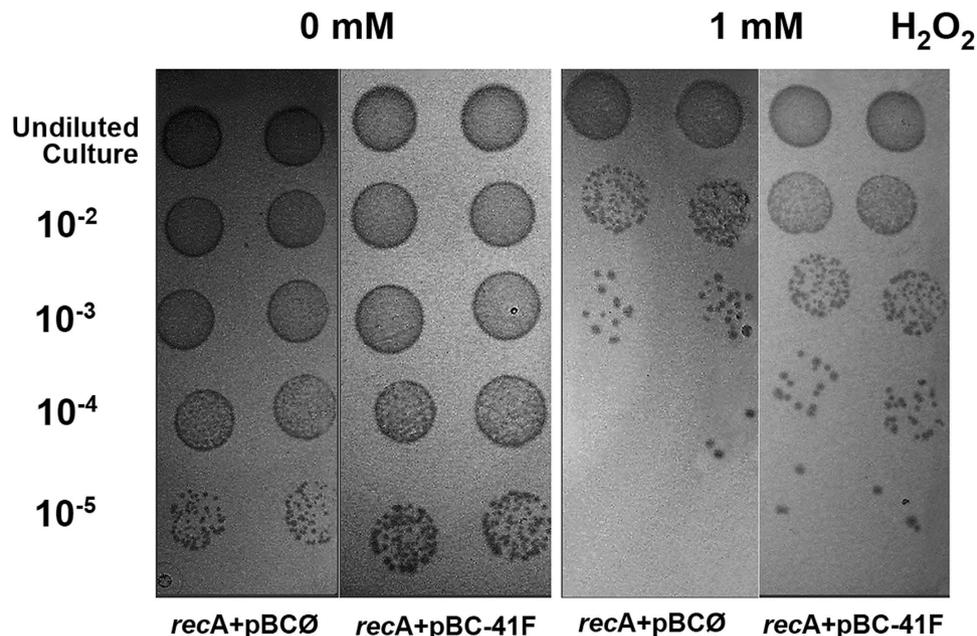


Figure 1. Complementation assay in the *E. coli recA*- strain. An increase in H_2O_2 resistance was observed in the DH10B strain due to the presence of an environmental clone (pBC-41F, containing the ExoMeg1 ORF). Undiluted cultures and serial dilutions (10^{-2} , 10^{-3} , 10^{-4} and 10^{-5}) from cultures (OD_{600} of 0.8) were plated on LB agar containing 1 mM H_2O_2 .

which is a primary limitation of this approach, in addition to the lack of innovative screening methods and novel activity assays^{9,15,16}.

The study of genome stability and DNA repair mechanisms is extremely important for not only understanding molecular mechanisms and evolutionary implications but also because of medical relevance because several diseases are attributed to deficiencies in DNA repair mechanisms, as well as the resistance of microorganisms to host immune systems, which is an important aspect of the host-pathogen relationship^{1,2}. Despite this importance, few studies are available that use the metagenomic approach for the discovery of new genes involved in the maintenance of genome integrity^{13,17–19}. As such, no protein directly related to DNA repair activity has been successfully purified and biochemically characterized. In this study, a new exonuclease potentially related to DNA repair has been identified and experimentally characterized using a metagenomic approach with functional screening.

Results

In this study, 2,688 clones of one environmental DNA library was screened through a complementation assay using the DH10B *E. coli* strain as host, which is deficient in the DNA repair gene *recA*, rendering it sensitive to several mutagenic agents, including ultraviolet light (UV) and hydrogen peroxide (H_2O_2). The screening was conducted on H_2O_2 -LB agar plates, and one candidate clone, named pBC-41F, was pre-selected. Afterward, to avoid the isolation of a false-positive clone, the recombinant plasmid was isolated and retransformed into DH10B cells to perform a new activity confirmation assay in selective medium. Since it showed positive results in a complementation assay in the *recA*- strain (Fig. 1), it was selected for the functional characterization described in this work.

In silico analysis of the pBC-41F clone. The size of the environmental insert cloned into the pBC plasmid was 1.5 kb, and sequence analysis revealed the presence of an open reading frame (ORF) of 975 bp encoding a polypeptide of 324 amino acids with a predicted molecular weight and theoretical isoelectric point (pI) of approximately 37 kDa and 5.02, respectively. The search for homologous sequences revealed the absence of significant similarity to known proteins deposited in the non-redundant database (NCBI) but showed similarity to hypothetical protein BfaeM_06119 (ZP_09858401.1, presenting 17% identity, e-value $2e^{-62}$ and 99% coverage) of *Bacteroidesfaecis* MAJ27, containing a conserved domain related to the Exo_Endo_Phos (PF03372) superfamily (EEP).

The multiple sequence alignments of metagenomic ORF against different EEP family members showed the presence of six conserved catalytic residues, which suggests that the environmental protein belongs to this superfamily and shares the same catalytic mechanism with other EEP members (Fig. 2A). This ORF was named exonuclease 1 from the metagenome, ExoMeg1, due to its exonuclease activity described below. In the phylogenetic reconstruction, ExoMeg1 is present in a monophyletic group, composed of DNase I and MnuA subfamily members. The phylogeny obtained suggests that the divergence of ExoMeg1 is an ancestral event separate from the divergence of the DNase I and MnuA subfamilies. Thus, ExoMeg1 is in a separate branch, not grouping with any members of EEP subfamilies (Fig. 2B), suggesting that this enzyme belongs to a new sub family.

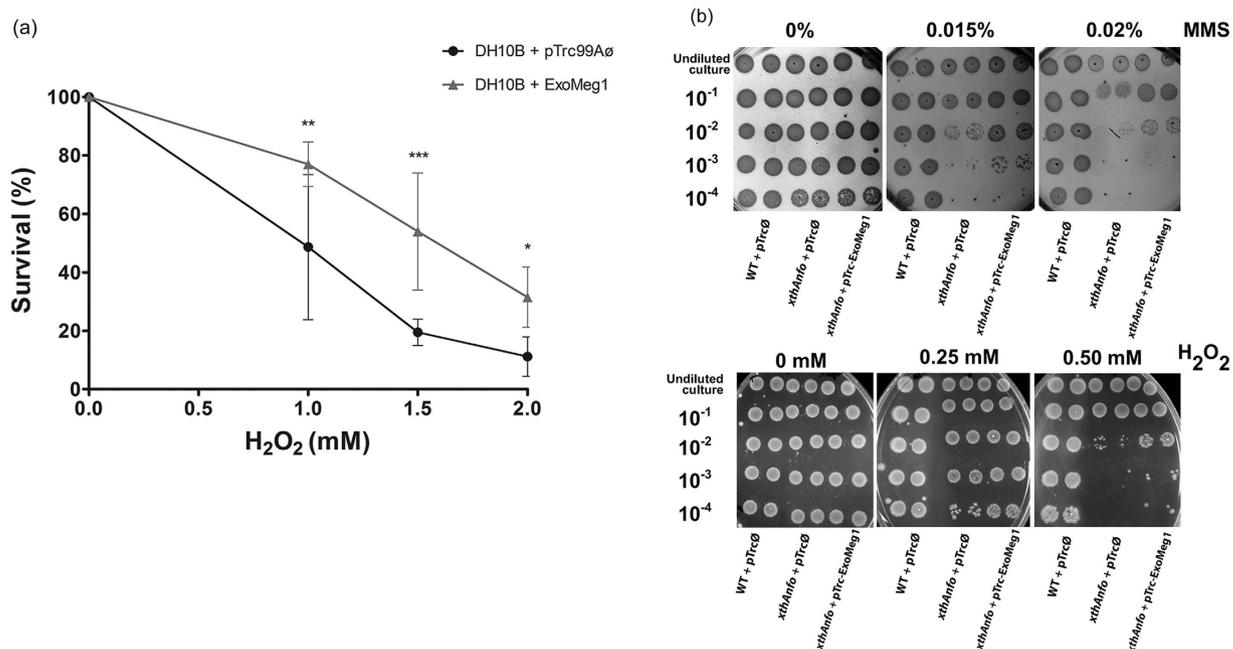


Figure 3. Complementation assay in the *E. coli* DNA repair deficient strains. (A) Survival curve of the *recA*- strain subjected to increasing concentrations of H_2O_2 (0–2 mM) for 30 min. *recA*- strain transformed with empty vector (●) and with pTrc-ExoMeg1 (▲). Values are the mean \pm SD of three independent experiments in duplicate. Statistical analysis by two-way ANOVA and Bonferroni post-test, * $P < 0.05$, ** $P < 0.01$ and *** $P < 0.001$. (B) Complementation assay in the *E. coli xthA-nfo*- strain. A survival increase in plates with MMS or H_2O_2 was observed in the presence of pTrc-ExoMeg1 in the *E. coli* double mutant strain. The cultures were induced by 0.1 mM IPTG. Undiluted cultures and serial dilutions (10^{-1} , 10^{-2} , 10^{-3} and 10^{-4}) from cultures (OD_{600} of 0.8) were plated on LB agar containing different concentrations of MMS or H_2O_2 . AB1157 was used as wild type (WT) strain proficient in DNA repair.

Considering that some EEP-family members present activity as AP endonuclease/3'-phosphodiesterases important for the repair of DNA lesions generated by H_2O_2 , such as abasic sites or 3' blocking damages, we compared ExoMeg1 with the DNA repair enzymes exonuclease III (Xth) from *E. coli* and Nexo from *N. meningitidis*, which are members of the Xth sub family. Alignment indicates the presence of conserved residues (N9, E47, I252, D195, N197, W217, S229, D253 and H314) potentially involved in AP endonuclease/3'-phosphodiesterase activities (Fig. 2C). However, the residue W212 found in exonuclease III of *E. coli* and some loops and α -helix regions important for AP-endonuclease activity are absent in ExoMeg1.

Complementation assay in an *E. coli* strain deficient in DNA repair. The recombinant plasmid containing the ExoMeg1 ORF (pTrc- ExoMeg1) was able to decrease the sensitivity of the *recA*- strain to H_2O_2 because it increased survival by approximately 20% when compared to the same mutant strain transformed with the empty vector (Fig. 3A). Additionally, due to the presence of domains peculiar to the EEP family (Fig. 2A), pTrc-ExoMeg1 construction was transformed into *E. coli* strain deficient in exonuclease III (*xthA*) and endonuclease IV (*nfo*). The results show that ExoMeg1 complemented the DNA repair deficiency (*xthA-nfo*-), allowing for better survival after methyl methanesulfonate (MMS) or H_2O_2 treatments (Fig. 3B). Together, these results suggest a repair activity for ExoMeg1.

Activity assays. The environmental protein was obtained by two different and consecutive purification processes using HisTrap FF and HiTrap Q HP columns, and, for the last step, a salt gradient was used, being the purified protein eluted in 20% buffer B as verified by SDS-PAGE (Fig. 4A).

To determine whether ExoMeg1 has AP endonuclease activity, we performed *in vitro* assays using 5'-Cy5-labeled substrates consisting of 21-mer oligonucleotides containing a single abasic site at position 10. Incubation of the purified protein with the AP-containing 21-mer oligonucleotide generated a 13-nt product, which could not be observed in the reaction using the control oligonucleotide without the abasic site. However, the AP endonuclease activity should result in an 11-mer product. Additionally, progressive shortening of the full length 21-mer substrate was observed, exhibiting a ladder pattern for the fragments in the gel, indicating that ExoMeg1 degrades the dsDNA and ssDNA substrates from the 3' end, suggesting an exonuclease activity with 3' to 5' polarity (Fig. 4B–G).

The effect of salt concentration and metal dependency on ExoMeg1 activity was also investigated. High salt concentration was able to affect ExoMeg1's 3'-5' exonuclease activity in both ssDNA and dsDNA substrates (Fig. 4D,E). However, the exonuclease activity on ssDNA is more resistant to high salt concentration because the ladder pattern is still observed at higher KCl concentrations. At lower KCl concentrations, it is possible to observe

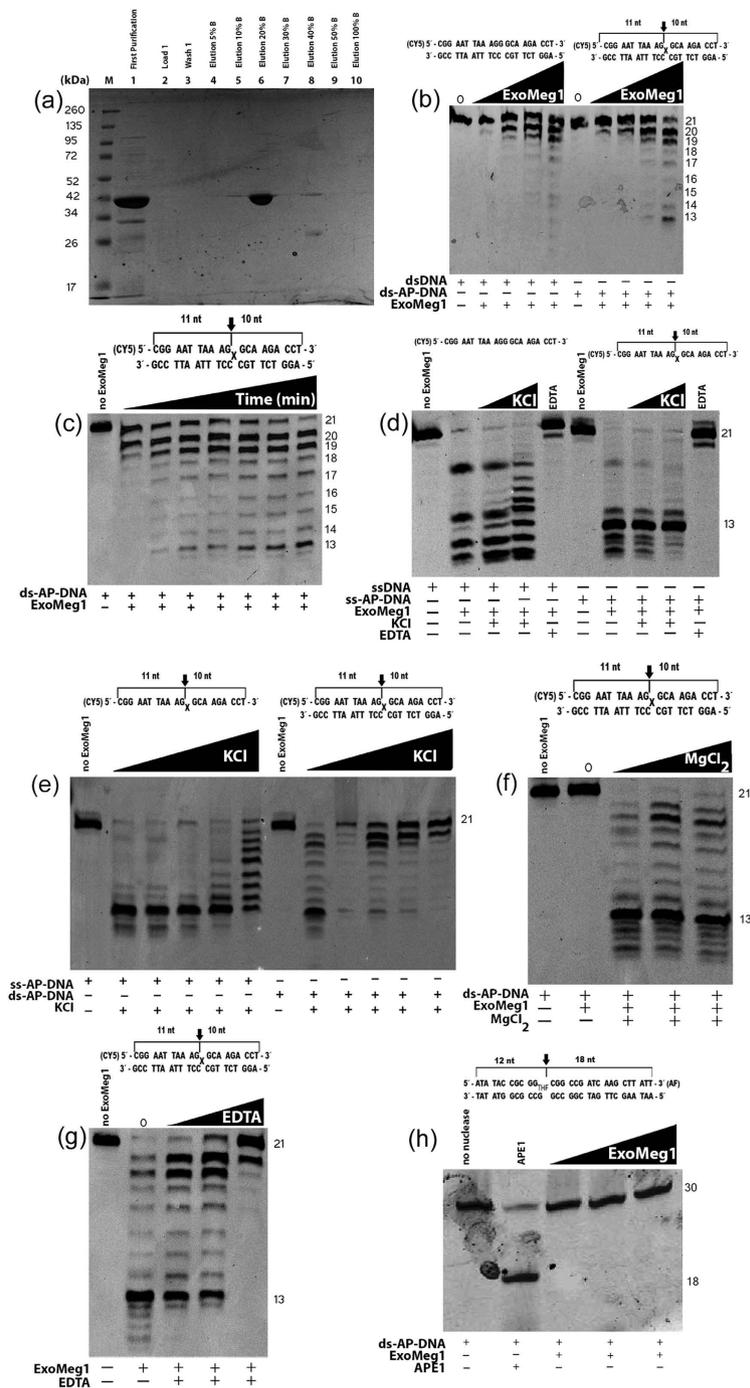


Figure 4. Purification and *in vitro* characterization of ExoMeg1. (A) SDS-PAGE analysis of metagenomic protein purification through a HiTrap Q HP. Aliquots (10 μ L) of each sample were electrophoresed on a 12% SDS-PAGE, which was stained with Coomassie blue. Lane M, molecular weight marker SpectraBR; Lane 1, purification by HisTrap column; Lanes 2–10, steps of purification by HiTrap Q HP. The protein was eluted with 20% buffer B (lane 6). (B) Identification of 3'-5' exonuclease activity on dsDNA or ds-AP-DNA with increasing concentrations of ExoMeg1. The AP endonuclease activity should result in an 11-mer product. However, ExoMeg1 has shown only the exonuclease activity, generating a 13-mer product. (C) Activity assay versus time of incubation at 37°C using ds-AP-DNA. (D) Identification of 3'-5' exonuclease activity on the single strand substrate and the effect of salt on the activity of ExoMeg1. Lanes 1 and 6, reactions in the absence of ExoMeg1 and KCl; Lanes 2 and 7, reactions in the presence of ExoMeg1 but absence of KCl; Lanes 3 and 8, reactions including ExoMeg1 and 50 mM KCl; Lanes 4 and 9, reactions including EXOMEG1 and 200 mM KCl; lanes 5 and 10, reactions containing 10 mM EDTA. (E) Analysis of salt effects on ExoMeg1 activity using increasing concentrations of KCl (50, 100, 250, 500 and 1000 mM). (F) Mg²⁺ dependence of ExoMeg1 activity. The reactions contain increasing concentrations of MgCl₂ (0, 5, 10 and 15 mM). The ExoMeg1 activity is totally inhibited by the absence of MgCl₂ (lane 2). (G) Effect of EDTA on ExoMeg1 activity. The reactions contain increasing concentrations of EDTA (0, 1, 5 and 10 mM). The progression of 3'-5' exonuclease activity is gradually inhibited by EDTA. (H) AP endonuclease assay using a 30-mer labeled at the 3' end with increasing concentrations of ExoMeg1. Commercial APE1 enzyme (NEB Biolabs) was used as the positive control (lane 2).

the degradation of higher products generating smaller fragments, suggesting faster activity at lower salt concentrations (Fig. 4E, lane 6). Additionally, it was observed in this work that ExoMeg1 activity requires a divalent metal cofactor because the exonuclease activity was observed only in presence of magnesium (Fig. 4F). To confirm that ExoMeg1 is a metal-dependent enzyme, reactions were performed including increasing concentrations of EDTA and compared to the reaction without EDTA. From this assay, it was observed that EDTA inhibited the exonuclease activity, confirming ExoMeg1's dependence on a divalent metal (Fig. 4F,G).

The occurrence of an 13-nt product generated by the activity of ExoMeg1 raises the question of whether this product results from cleavage of the abasic site by a non-canonical mechanism, different from other AP endonucleases, such as exonuclease III, or whether such a product would be assigned to an activity pause site of ExoMeg1. To answer this question, we conducted an activity assay using a 3'-labeled oligonucleotide containing an abasic site analogue. The data showed that ExoMeg1 has no direct AP endonuclease activity or 5'-3' exonuclease activity because no cleavage product was observed (Fig. 4H). Therefore, ExoMeg1 is a 3'-5' exonuclease acting on ssDNA and dsDNA with ssDNA being its preferred substrate.

Discussion

The metagenomic approach has been considered to be an efficient method for obtaining novel and useful genes from uncultured microorganisms. In this work, functional screening allowed for the identification of a metagenomic clone that shows activity that is potentially related to DNA repair because it was able to complement the DNA repair deficiency in *recA*- and *xth-nfo*- strains. Sequence analysis revealed an ORF with similarity to a hypothetical protein containing a conserved domain related to the EEP superfamily. This superfamily contains various structural and functionally related enzymes, which share a common mechanism of phosphodiester bond cleavage but act on different substrates. EEP representative enzymes include magnesium dependent endonucleases (L1-EN, DNase I, APE1, APE2, Xth), exonucleases (Xth, REX1, REX2, APE1), phosphatases (I5PP), sphingomyelinases and others proteins with unknown functions^{20–25}.

ExoMeg1 has 3'-5' exonuclease activity, and an unexpected 13-nt product was observed when tested with an oligonucleotide containing a single abasic site. This 13-nt product can result from a pausing site of the exonuclease activity in the vicinity of the abasic site because direct AP endonuclease activity was not observed. Therefore, ExoMeg1 is a 3'-5' exonuclease that likely tends to pause near DNA damage similar to what has been observed for other 3'-5' exonuclease enzymes^{26,27}; however, it continues over the substrate to generate smaller products, as seen in Fig. 4.

In fact, the presence of most of the conserved residues important for AP endonuclease and 3' phosphodiesterase activities in Xth of *E. coli* and Nexo of *N. meningitidis* were found in the ExoMeg1 sequence, suggesting a DNA repair function. However, the absence of the equivalent residue to W212 found in Xth and the absence of structural elements, which produce direct contacts with nucleic acids, such as loops or α -helix that have been described to confer selectivity to the cleavage of abasic sites^{28–30}, provide possible reasons for the lack of classic AP endonuclease activity in the ExoMeg1 protein. Similarly, it was shown that Nexo from *N. meningitidis* is a Xth family member with significant 3' phosphodiesterase and exonuclease activity, which is fundamental in the repair of 3' blocking damages generated by oxidative stress, and has no AP endonuclease activity^{31,32}. The presence of these regions is crucial for AP endonuclease activity, as demonstrated by Cal *et al.*²⁸, which, with addition of some amino acids of Xth α -helix, converted a bovine pancreatic DNase I to a repair endonuclease with a high selectivity for abasic sites. This specific region is absent in ExoMeg1. This also corroborates our phylogenetic analysis in which ExoMeg1 is closest to the DNase1-like branch. These data emphasize that despite EEP members sharing the mechanism of phosphodiester bond cleavage, there are peculiarities that define the specificity of a given enzyme for different substrates²⁰.

Complementation assays showed that ExoMeg1 is able to complement the *recA*- and *xth-nfo*- strains, suggesting that this enzyme is involved in the repair of DNA lesions caused by H₂O₂ and MMS. Both agents can induce abasic sites and strand breaks with 3' blocking ends that are substrates for recombination or base excision repair^{33–35}. Therefore, the 3'-5' exonuclease activity of ExoMeg1 may be acting on the repair of single strand breaks with 3' blockers because AP endonuclease activity was not observed.

Some 3'-5' exonucleases are specialized to remove 3' blocking damages, which are very cytotoxic if not repaired. As an example, Nexo has 3'-5' exonuclease and 3' phosphodiesterase activities that are able to improve resistance to H₂O₂ and paraquat, but it has no detectable AP endonuclease activity, which is attributed to the presence of His167 instead of G or S, as found in other Xth subfamily members. Nexo is an essential enzyme specialized for 3'-end processing, mainly for removing 3' PO₄ blocking damages^{31,32} that are generated by some bi-functional glycosylases (as FPG) but can also arise directly from oxidative damage^{36–38}. In another example, the 3'-5' exonuclease activity of Rad1-Rad10 from *Saccharomyces cerevisiae* is required for removing 3' phosphoglycolate, a 3' blocking lesion generated by oxidative attacks on sugar in DNA³⁹. Mazouzi *et al.*⁴⁰ showed that Xth and APE1 efficiently removed the lesion (5'S)-8,5'-*cyclo*-2'-deoxyadenosine (S-cdA) when present at 3' termini of a recessed, nicked, or gapped DNA duplex. However, these enzymes are not able to remove the S-cdA adducts when present at 1 or more nt away from the 3' end.

Other functions has also been attributed to 3'-5' exonucleases, such as nucleotide incision repair (NIR) mediated by Nfo and in alternative pathways for repairing 8-oxoG, mismatches and 3' blocking groups by 3'-5' exonuclease function of Apn1^{41,42}. Likewise, human Ape2, aside from its inefficient AP endonuclease activity *in vitro*, has strong 3'-5' exonuclease activity that is able to remove mismatched nucleotides from the 3' prime end, suggesting a role as a proofreader in DNA repair synthesis similar to human APE1 proofreading errors by Pol- β ⁴³.

Salt concentration and specific metals may influence DNA cleavage by nucleases. Protein-DNA interactions are highly dependent on salt concentration, and the nuclease enzymes have conserved motifs that form active sites that coordinate catalytically essential divalent cations, such as magnesium, manganese, calcium or zinc as a cofactor^{44,45}. In this study, we observed that ExoMeg1 is divalent metal-dependent because it is inactive in the

Strain or plasmid	Genotype/Description	Reference
Strains		
DH10B	F ⁻ endA1 recA1 galE15 galK16 nupGrpsLΔlacX74 Φ80lacZΔM15 araD139 Δ(ara,leu)7697 mcrAΔ(mrr-hsdRMS-mcrBC) λ ⁻	51
AB1157	thr-1, araC14, leuB6(Am), Δ(gpt-proA)62, lacY1, tsx-33, qsr'-0, glnV44(AS), galK2(Oc), LAM-, Rac-0, hisG4(Oc), rfbC1, mgl-51, rpoS396(Am), rpsL31(strR), kdgK51, xylA5, mtl-1, argE3(Oc), thi-1	52
AB1157 <i>xthA</i> ⁻ <i>Info</i> ⁻	As AB1157 but info::Kan, Δ(<i>xthA</i> - <i>pncA</i>)DE	CNRS
BL21 (DE3) pLysS	F- ompT gal dcmlonhdsSB(rB- mB-) λ(DE3) pLysS(cmR)	Promega
Plasmids		
	Description	
pBC SK	Cloning vector, Cam ^r	Stratagene
pBC-41 F	pBC vector plus 1.5 kb environmental insert	This study
pCR-Blunt II-TOPO	Cloning vector, Kan ^r	Invitrogen
pCR-Blunt-EXOMEG1	pCR-Blunt vector plus 975 pb environmental ORF	This study
pHis ^{Parallel}	Expression vector, IPTG inducible, Amp ^r	53
His-EXOMEG1	pHis vector plus 975 pb environmental gene	This study
pTrc99A	Cloning vector, IPTG inducible, Amp ^r , it allows T7 polymerase-independent expression	54
pTrc-Exomeg1	pTrc99A plus 975 pb environmental gene	This study

Table 1. Strains and plasmids used in this study.

absence of any divalent metal but is active in presence of Mg²⁺. Furthermore, it is inactivated by EDTA, a chelating agent, similarly to that described for Dnase I and Xth^{45–47}. ExoMeg1 is a salt resistant enzyme because inhibition of the exonuclease activity was not completely achieved. The exonuclease activity was observed even with the highest tested concentration of 1 M of KCl. This concentration is approximately 10-fold higher than what was described for Xth and Nfo^{38,48}. This high resistance of ExoMeg1 to salt concentration may reflect environmental adaptation because the metagenomic library was obtained from DNA extracted from semiarid soil exposed to environmental stresses, including salinity, drought and UV radiation⁴⁹. These environmental stressors may lead to the generation of reactive oxygen species (ROS) and, consequently, oxidative damage of nucleic acids, including modified bases and single or double strand DNA breaks. In fact, it has been related that a high concentration of NaCl increases DNA breaks in both cell culture and *in vivo*⁵⁰. Thus, considering that well-known exonucleases of model organisms are sensitive to high salinity is very useful a new salt-resistant exonuclease, Exomeg1, able to decrease the sensitivity to oxidative stress and thus confer an important environmental adaptation.

In conclusion, in this study, a novel protein potentially related to DNA repair, initially annotated as a hypothetical protein of unknown function, has been characterized experimentally *in vivo* and *in vitro* and it is the first enzyme obtained from the metagenomic approach showing this function. In this first study on ExoMeg1, we described its 3'–5' exonuclease activity and its absence of AP endonuclease function. The next steps will be to test its 3' phosphodiesterase activity and identify the probable 3' blocking lesions that may be removed by this new enzyme. The results obtained in this work support the idea that metagenomes represent a rich source for finding new genes because it is expected that uncultivated fraction include many microorganisms distantly related to cultivated microorganisms. In fact, a vast portion of genes described in genome or metagenome projects are classified as coding hypothetical proteins. Hence, the discovery the functions of these new genes represents one of the biggest challenges of genome or metagenome projects because if there is no similarity to sequences already deposited in databanks, such as NCBI/nr or UniProt, in general, it is impossible to infer a function. Here, we were able to attribute function to a hypothetical ORF, contributing to the understanding of the diversity of the DNA repair process and generation of many new questions to be answered with new assays using various substrates and mutants of ExoMeg1.

Methods

Culture conditions. The *E. coli* strains and plasmids used in this study are listed in Table 1. The strains were routinely grown in Lysogeny Broth (LB) in g/L: Tryptone 10, Yeast Extract 5, NaCl 10; at 37 °C and, when necessary, were supplemented with 25 μg/mL chloramphenicol, 50 μg/mL kanamycin or 100 μg/mL ampicillin to maintain the plasmid vectors. All enzymes were purchased from New England Biolabs.

Library construction and functional screening. Soil samples (5–10 cm) were collected from João Câmara city, located in northeast Brazil (S5° 30' 51.81" O35° 54' 17.13"). This city is localized in a semiarid region, and it is included in the Caatinga biome, which is an exclusively Brazilian biome⁴⁹. The collection was performed using sterile tubes and spatulas, and environmental DNA (eDNA) was extracted using the commercial kit PowerMax TM Soil DNA Isolation Kit (MOBIO Laboratories) from 10 g of soil. The eDNA was partially digested with the enzymes EcoRV, SmaI and PvuII, and the 1.5–2 kb fragments were purified from agarose gel and linked to the cloning vector pBC previously EcoRV-linearized and CIAP-dephosphorylated. The library obtained was transformed into the DNA repair-deficient DH10B (*recA*-) *E. coli* strain and plated on chloramphenicol/X-gal-LB agar. For functional screening, the library was replicated on LB agar containing different concentrations of hydrogen peroxide (0–3 mM H₂O₂, Merck). The strain transformed with the empty vector was used as the negative

control. To avoid selecting a false positive clone, the recombinant plasmid was isolated and retransformed into DH10B to confirm that the acquired phenotype was due to the presence of the environmental insert.

ORF identification and sequence analysis. The positive clones were sequenced using a DYEnamic ET Dye Terminator Cycle Sequencing Kit for MegaBACE 500 (Amersham Biosciences) following the manufacturer's instructions. ORF predictions were performed using the program ORF Finder available online on the website (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>) provided by the National Center for Biotechnology Information (NCBI), accessed in April of 2013. The nucleotide sequence of the environmental nuclease characterized in this work has been deposited in the GenBank database under the Accession Number KJ436741.

The predicted ORF was used as the query in DELTA-Blast of the BLAST package (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) and compared with protein sequences in the non-redundant (nr) GenBank and CDD databases. The molecular weight and theoretical isoelectric point (pI) were predicted using the Compute pI/Mw tool from ExPASy (http://www.expasy.org/tools/pi_tool.html)⁵⁵, and the prediction of subcellular localization was computed by the PSLpred method (<http://www.imtech.res.in/raghava/pslpred/>)⁵⁶. The T-coffee program⁵⁷ was used to align the deduced protein with other subfamilies of the Exo_Endo_Phos (EEP) superfamily, which was obtained by the CDD hierarchy in addition to the data obtained from DELTA-Blast. To reduce the sequences sampled, two EEP members from each subfamily were selected based on better score alignment with the metagenomic protein. Phylogenetic reconstruction was performed by the MEGA5 program⁵⁸ using the maximum likelihood method and WAG model. Non-uniformity of evolutionary rates among sites was modeled using a discrete Gamma distribution (+G). All positions with less than 95% site coverage and redundant EEP members were eliminated. Bootstrap values were based on 1000 replicates. Myo-inositol-1(or 4)-monophosphatase sequences belonging to the IMPase-like superfamily were used as the outgroup.

Subcloning of the metagenomic ORF. ORF-specific forward (5' CCA TGG AAG TGC GCA TAG CGA CAT) and reverse (5' CTC GAG TCA AGT TAG TTC CAC AAT CAC CCT) primers were designed that contain restriction sites for the *NcoI* and *XhoI* enzymes added to their 5' regions, respectively. The PCR product was initially cloned into the pCR-Blunt vector (Invitrogen) and sequenced. Then, the ORF of interest was excised from the cloning vector with the enzymes *NcoI* and *XhoI* and subcloned into linearized pTrc99A and pHis parallel vectors with the enzymes *NcoI/SalI* and *NcoI/XhoI*, respectively.

Complementation assay. Wild type- or DNA repair-deficient strains were transformed with the appropriate vector. The overnight cultures were diluted 50X in fresh LB medium supplemented with an adequate amount of the appropriate antibiotic and grown at 37 °C under constant agitation at 180 rpm to achieve an optical density (OD) at 600 nm of 0.4 when IPTG was added to the culture. The construction of the pBC and pTrc99A vectors was induced by 1 mM and 0.1 mM IPTG, respectively. To perform a survival curve, we incubated the cultures at 37 °C with shaking for an additional 40 min to obtain cells that were still in the logarithmic phase, and then challenged them with H₂O₂ (0–2 mM) for 30 min. The cultures were diluted with LB broth and plated onto LB agar plates, and survival was assessed by counting colonies after 24 h. To perform the drop assay, aiming to verify if ExoMeg1 was able to affect the double mutant *xthA/nfo* after reaching OD₆₀₀ of 0.8, the induced cultures were diluted in LB liquid medium (10⁻¹, 10⁻², 10⁻³ and 10⁻⁴) and 10 µl of each dilution, as well as undiluted culture (10⁰) and were dripped in LB agar containing the appropriate antibiotic with different concentrations of H₂O₂ and MMS agents. The wild type and mutant strains transformed with empty vector were used as controls. The plates were incubated at 37 °C for 24 h.

GraphPad Prism version 5.00 for Windows (GraphPad Software, San Diego California USA) was used to perform the statistical analyses of survival curve. Two-way analysis of variance (ANOVA) followed by the Bonferroni post-test were performed. Difference was considered significant when P < 0.05.

Expression of the recombinant protein. To obtain the recombinant protein, *E. coli* BL21 carrying pHis-ExoMeg1 was grown in two liters of LB medium, containing 100 µg/mL ampicillin and 34 µg/mL chloramphenicol at 37 °C. The cells were cultured to an OD₆₀₀ of 0.45, and expression of the environmental gene was induced by adding IPTG to a final concentration of 0.1 mM and culturing for 20 h at 16 °C. The cell extract was collected by centrifugation at 15,000 g for 15 min, resuspended in PBS (pH 7.2). Lysis was accomplished by the addition of lysis buffer (0.5 M NaCl, 10% glycerol, 20 mM Hepes pH 8.0, 5 mM imidazole and 2 mM β-mercaptoethanol), 1 mM of the protease inhibitor PMSF, 1 mg/mL lysozyme, 0.5% of 10% Triton X-100 and 150 U Benzonase nuclease (Novagen, 10 KU, 25 U/µL). After incubation on ice for 1 h, the sample was sonicated until viscosity was diminished. The soluble cell extract obtained by ultracentrifugation for 45 min at 4 °C was subjected to chromatography on a HisTrap HP column (GE Healthcare). The protein was eluted at 300 mM imidazole and was dialyzed against buffer A (0.05 M KCl, 20 mM Hepes pH 8.0, 1 mM EDTA pH 8.0, 1 mM Dithiothreitol and 10% Glycerol) using membrane SnakeSkin dialysis tubing 10,000 MWCO (Thermo Scientific) over a period of 18 h at 4 °C under agitation. The dialysate was loaded onto a Hitrap Q HP column (GE Healthcare), which was loaded with a salt gradient using Buffer A with the addition of 5, 10, 20, 30, 40, 50 and 100% buffer B (1 M KCl, 20 mM Hepes pH 8.0, 1 mM EDTA pH 8.0, 1 mM Dithiothreitol and 10% glycerol). At the end, all steps of the purification process, including the induced and uninduced samples, were applied to a 12% SDS-PAGE. The fraction containing the eluted proteins was frozen with liquid nitrogen and kept in a freezer at -80 °C.

Substrates and activity assays. The AP endonuclease activity of environmentally purified protein was verified on single or double-stranded substrates with or without an abasic site at position 10 of the oligonucleotide 21-mer (ssDNA, ss-AP-DNA, dsDNA and ds-AP-DNA). This oligonucleotide was 5'-fluorescently labeled with

Control (1)	5' Cy5-CGG AAT TAA AGG GCA AGA CCT-3'
AP (1)	5' Cy5-CGG AAT TAA AG _X GCA AGA CCT-3'
Complementary (1)	5' AGG TCT TGC CCT TTA ATT CCG-3'
AP (2)	5'-ATA TAC CGC GG _{THF} CCG CCG ATC AAG CTT ATT -AF3'
Complementary (2)	5'- AAT AAG CTT GAT CCG CCG GCC GCG GTA TAT- 3'

Table 2. Labelled Oligonucleotides: -X: AP- site; THF: tetrahydrofuran.

Cy5. The 30-mer oligonucleotide, containing the abasic site analogue tetrahydrofuran at position 19 (Table 2), was 3'-labeled (ds-AP-DNA) with dCTP-Alexa Fluor and Terminal Transferase (Promega).

Standard reactions (20 μ l) containing buffer 1 (5 mM Tris-HCl, 5 mM MgCl₂, 0.5 mM dithiothreitol pH 7.0), 1 μ M substrate, single (ssDNA) or double-stranded (dsDNA) DNA, and increasing concentrations of purified protein (0, 10, 20, 50 and 100 μ M) were maintained at 37 °C for 1 hr. The reactions were terminated with the "STOP" solution (98% formamide and 0.5 M EDTA). The samples were heated at 95 °C for 3 min and immediately chilled on ice before applying to a PAGE-Urea gel (20% acrylamide and 8 M urea). The gels were run for 1 hr at 15 W and scanned with an FLA-5000 image analyzer (FUJIFILM) to detect the labeled DNA. Subsequently, all assays were performed using 100 μ M of purified protein in each reaction, and the control reactions were run without any added protein. For time courses, the reactions were maintained at 37 °C and an aliquot of 10 μ L of reaction was taken after 5, 10, 20, 30, 40, 50 and 60 min. To investigate the dependence of the divalent metal cofactor for protein activity, the reactions were performed in buffer 2 (5 mM Tris-HCl, 0.5 mM dithiothreitol pH 7.0) using increasing concentrations of MgCl₂ (0, 5, 10 and 15 mM). To test whether nuclease activity can be inactivated by EDTA, we performed standard reactions with increasing concentrations of EDTA (0, 1, 5 and 10 mM). In addition, to examine whether the activity of the metagenomic nuclease may be affected by salt concentration, we performed standard reactions by adding increasing concentrations of KCl (0–1 M).

References

- Eisen, J. A. & Hanawalt, P. C. A phylogenomic study of DNA repair genes, proteins, and processes. *Mutat Res* **435**(3), 171–213 (1999).
- Hoeijmakers, J. H. Genome maintenance mechanisms for preventing cancer. *Nature* **411**(6835), 366–74(2001).
- Aravind, L., Walker, D. R. & Koonin, E. V. Conserved domains in DNA repair proteins and evolution of repair systems. *Nucleic Acids Res* **27**(5), 1223–42 (1999).
- Grogan, D. W. The question of DNA repair in hyperthermophilic archaea. *Trends Microbiol* **8**(4), 180–5 (2000).
- Jiricny, J. Postreplicative mismatch repair. *Cold Spring Harb Perspect Biol* **5**(4), a012633, 1–23 (2013).
- Krokan, H. E. & Bjoras, M. Base excision repair. *Cold Spring Harb Perspect Biol* **5**(4), a012583, 1–22 (2013).
- Kisker, C., Kuper, J. & Van Houten, B., Prokaryotic nucleotide excision repair. *Cold Spring Harb Perspect Biol* **5**(3), a012591, 1–18 (2013).
- Amann, R. L., Ludwig, W. & Schleifer, K. H., Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol Rev* **59**(1), 143–69 (1995).
- Ferrer, M., Beloqui, A., Timmis, K. N. & Golyshin, P. N., Metagenomics for mining new genetic resources of microbial communities. *J Mol Microbiol Biotechnol* **16**(1–2), 109–23 (2009).
- Handelsman, J. *et al.* Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol* **5**(10), R245–9 (1998).
- Rondon, M. R. *et al.* Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl Environ Microbiol* **66**(6), 2541–7 (2000).
- Lammler, K. *et al.* Identification of novel enzymes with different hydrolytic activities by metagenome expression cloning. *J Biotechnol* **127**(4), 575–92 (2007).
- Simon, C., Herath, J., Rockstroh, S. & Daniel, R. Rapid identification of genes encoding DNA polymerases by function-based screening of metagenomic libraries derived from glacial ice. *Appl Environ Microbiol* **75**(9), 2964–8 (2009).
- Jeon, J. H. *et al.* Identification of a new subfamily of salt-tolerant esterases from a metagenomic library of tidal flat sediment. *Appl Microbiol Biotechnol* **93**(2), 623–31 (2011).
- Li, G., Wang, K. & Liu, Y. H., Molecular cloning and characterization of a novel pyrethroid-hydrolyzing esterase originating from the Metagenome. *Microb Cell Fact* **738**, 1–10 (2008).
- Jiang, C. *et al.*, Biochemical characterization of two novel beta-glucosidase genes by metagenome expression cloning. *Bioresour Technol* **102**(3), 3272–8 (2011).
- Mori, T., Suenaga, H. & Miyazaki, K., A metagenomic approach to the identification of UDP-glucose 4-epimerase as a menadione resistance protein. *Biosci Biotechnol Biochem* **72**(6), 1611–4 (2008).
- Yang, H., To, K. H., Aguila, S. J. & Miller, J. H., Metagenomic DNA fragments that affect *Escherichia coli* mutational pathways. *Mol Microbiol* **61**(4), 960–77 (2006).
- Ogata, H. *et al.* Two new subfamilies of DNA mismatch repair proteins (MutS) specifically abundant in the marine environment. *ISME J* **5**(7), 1143–51 (2011).
- Dlagic, M. Functionally unrelated signalling proteins contain a fold similar to Mg²⁺-dependent endonucleases. *Trends Biochem Sci* **25**(6), 272–3 (2000).
- Tamura, H. *et al.* Mutation in aspartic acid residues modifies catalytic and haemolytic activities of *Bacillus cereus* sphingomyelinase. *Biochem J* **309** (Pt 3), 757–64 (1995).
- Whisstock, J. C. *et al.* The inositol polyphosphate 5-phosphatases and the apurinic/apyrimidinic base excision repair endonucleases share a common mechanism for catalysis. *J Biol Chem* **275**(47), 37055–61 (2000).
- Chen, J., Chiang, Y. C. & Denis, C. L. CCR4, a 3'-5' poly(A) RNA and ssDNA exonuclease, is the catalytic component of the cytoplasmic deadenylation. *EMBO J* **21**(6), 1414–26 (2002).
- Rogers, K., Gao, G. & Simpson, L. Uridylate-specific 3' 5'-exoribonucleases involved in uridylate-deletion RNA editing in trypanosomatid mitochondria. *J Biol Chem* **282**(40), 29073–80 (2007).
- Wang, H. *et al.* Crystal structure of the human CNOT6L nuclease domain reveals strict poly(A) substrate specificity. *EMBO J* **29**(15), 2566–76 (2010).
- Feddersen, A. *et al.* *Saccharomyces cerevisiae* Ngl3p is an active 3' 5'-exonuclease with a specificity towards poly-A RNA reminiscent of cellular deadenylases. *Nucleic Acids Res* **40**(2), 837–46 (2012).

27. Machwe, A., Ganunis, R., Bohr, V. A. & Orren, D. K., Selective blockage of the 3'→5' exonuclease activity of WRN protein by certain oxidative modifications and bulky lesions in DNA. *Nucleic Acids Res* **28**(14), 2762–70 (2000).
28. Cal, S., Tan, K. L., McGregor, A. & Connolly, B. A., Conversion of bovine pancreatic DNase I to a repair endonuclease with a high selectivity for abasic sites. *EMBO J* **17**(23), 7128–38 (1998).
29. Gorman, M. A. *et al.*, The crystal structure of the human DNA repair endonuclease HAP1 suggests the recognition of extra-helical deoxyribose at DNA abasic sites. *EMBO J* **16**(21), 6548–58 (1997).
30. Kaneda, K., Sekiguchi, J. & Shida, T. Role of the tryptophan residue in the vicinity of the catalytic center of exonuclease III family AP endonucleases: AP site recognition mechanism. *Nucleic Acids Res* **34**(5), 1552–63 (2006).
31. Carpenter, E. P. *et al.* AP endonuclease paralogues with distinct activities in DNA repair and bacterial pathogenesis. *EMBO J* **26**(5), 1363–72 (2007).
32. Silhan, J. *et al.* Specialization of an Exonuclease III family enzyme in the repair of 3' DNA lesions during base excision repair in the human pathogen *Neisseria meningitidis*. *Nucleic Acids Res* **40**(5), 2065–75 (2012).
33. Takemoto, T. *et al.* 3'-blocking damage of DNA as a mutagenic lesion caused by hydrogen peroxide in *Escherichia coli*. *J Radiat Res* **39**(2), 137–44 (1998).
34. Grzesiuk, E. The role of mutation frequency decline and SOS repair systems in methyl methanesulfonate mutagenesis. *Acta Biochim Pol* **45**(2), 523–33 (1998).
35. Otterlei, M. *et al.*, Repair of chromosomal abasic sites *in vivo* involves at least three different repair pathways. *EMBO J* **19**(20), 5542–51 (2000).
36. Agnez, L. F., Costa de Oliveira, R. L., Di Mascio, P. & Menck, C. F., Involvement of *Escherichia coli* exonuclease III and endonuclease IV in the repair of singlet oxygen-induced DNA damage. *Carcinogenesis* **17**(5), 1183–5 (1996).
37. Levin, J. D., Johnson, A. W. & Demple, B. Homogeneous *Escherichia coli* endonuclease IV. Characterization of an enzyme that recognizes oxidative damage in DNA. *J Biol Chem* **263**(17), 8066–71 (1988).
38. Demple, B., Johnson, A. & Fung, D., Exonuclease III and endonuclease IV remove 3' blocks from DNA synthesis primers in H₂O₂-damaged *Escherichia coli*. *Proc Natl Acad Sci USA* **83**(20), 7731–5 (1986).
39. Guzder, S. N. *et al.* Requirement of yeast Rad1–Rad10 nuclease for the removal of 3'-blocked termini from DNA strand breaks induced by reactive oxygen species. *Genes & development* **18**(18), 2283–91 (2004).
40. Mazouzi, A. *et al.* Insight into mechanisms of 3'-5' exonuclease activity and removal of bulky 8,5'-cyclopurine adducts by apurinic/apyrimidinic endonucleases. *Proc Natl Acad Sci USA* **110**(33), E3071–E3080 (2013).
41. Ishchenko, A. A., Yang, X., Ramotar, D. & Saparbaev, M. The 3'→5' exonuclease of Apn1 provides an alternative pathway to repair 7,8-dihydro-8-oxodeoxyguanosine in *Saccharomyces cerevisiae*. *Mol Cell Biol* **25**(15), 6380–90 (2005).
42. Golan, G. *et al.* Coupling of the nucleotide incision and 3'→5' exonuclease activities in *Escherichia coli* endonuclease IV: Structural and genetic evidences. *Mutat Res* **685**(1–2), 70–9 (2010).
43. Burkovics, P., Szukacsov, V., Unk, I. & Haracska, L. Human Ape2 protein has a 3'-5' exonuclease activity that acts preferentially on mismatched base pairs. *Nucleic Acids Res* **34**(9), 2508–15 (2006).
44. Nishino, T. & Morikawa, K., Structure and function of nucleases in DNA repair: shape, grip and blade of the DNA scissors. *Oncogene* **21**(58), 9022–32 (2002).
45. Butcher, S., Hainaut, P. & Milner, J. Increased salt concentration reversibly destabilizes p53 quaternary structure and sequence-specific DNA binding. *Biochem J* **298 Pt 3**, 513–6 (1994).
46. Lehtinen, D. A. *et al.* The TREX1 double-stranded DNA degradation activity is defective in dominant mutations associated with autoimmune disease. *J Biol Chem* **283**(46), 31649–56 (2008).
47. Erzberger, J. P. & Wilson, D. M., 3rd The role of Mg²⁺ and specific amino acid residues in the catalytic reaction of the major human abasic endonuclease: new insights from EDTA-resistant incision of acyclic abasic site analogs and site-directed mutagenesis. *J Mol Biol* **290**(2), 447–57 (1999).
48. Kerins, S. M., Collins, R. & McCarthy, T. V. Characterization of an endonuclease IV 3'-5' exonuclease activity. *J Biol Chem* **278**(5), 3048–54 (2003).
49. Pacchioni, R. G. *et al.* Taxonomic and functional profiles of soil samples from Atlantic forest and Caatinga biomes in northeastern Brazil. *MicrobiologyOpen* **3**(3), 299–315 (2014).
50. Dmitrieva, N. I., Cai, Q. & Burg, M. B. Cells adapted to high NaCl have many DNA breaks and impaired DNA repair both in cell culture and *in vivo*. *Proc Natl Acad Sci USA* **101**(8), 2317–22 (2004).
51. Grant, S. G., Jessee, J., Bloom, F. R. & Hanahan, D. Differential plasmid rescue from transgenic mouse DNAs into *Escherichia coli* methylation-restriction mutants. *Proc Natl Acad Sci USA* **87**(12), 4645–9 (1990).
52. Bachmann, B. J. Pedigrees of Some Mutant Strains of *Escherichia-Coli* K-12. *Bacteriological Reviews* **36**(4), 525–57 (1972).
53. Sheffield, P., Garrard, S. & Derewenda, Z. Overcoming expression and purification problems of RhoGDI using a family of “parallel” expression vectors. *Protein Expr Purif* **15**(1), 34–9 (1999).
54. Amann, E., Ochs, B. & Abel, K. J. Tightly regulated tac promoter vectors useful for the expression of unfused and fused proteins in *Escherichia coli*. *Gene* **69**(2), 301–15 (1988).
55. Gasteiger, E. *et al.* ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* **31**(13), 3784–8 (2003).
56. Bhasin, M., Garg, A. & Raghava, G. P. PSLpred: prediction of subcellular localization of bacterial proteins. *Bioinformatics* **21**(10), 2522–4 (2005).
57. Notredame, C., Higgins, D. G. & Heringa, J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* **302**(1), 205–17 (2000).
58. Tamura, K. *et al.* MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**(10), 2731–9 (2011).

Acknowledgements

This work was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), FAPESP (Grants 08/57721-3 and 10/51906-1) to NCS-P, and CNRS-France. The work was also supported by CAPES/COFECUB International Cooperation Program. We are also grateful for the technical assistance of Rita Lobo Napolitano and Gaëlle Philippin (CNRS, France) and Delanne C.S de Sena and Larissa A. Medeiros (UFRN, Brazil).

Author Contributions

R.C.B. Silva-Portela and L.F. Agnez-Lima wrote the main manuscript.

R.C.B. Silva-Portela, L.F. Agnez-Lima, N.C. de Souza-Pinto, M. Modesti and R.P. Fuchs analyzed the data. R.C.B. Silva-Portela, C.P.M. Pereira e F.M. Carvalho performed the experiments. R.C.B. Silva-Portela, C.P.M. Pereira and M. Modesti designed and prepared the samples. All authors discussed and reviewed the manuscript.

Additional Information

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Silva-Portela, R. C. B. *et al.* ExoMeg1: a new exonuclease from metagenomic library. *Sci. Rep.* **6**, 19712; doi: 10.1038/srep19712 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>