An unusual feruloyl esterase belonging to family VIII esterases and displaying a broad substrate range

Colin W Ohlhoff[e], Bronwyn M Kirby[a], Ana Casanueva[b], Robert J Huddy[c], Rolene Bauer[a], David LR Mutepfa[d], Don A Cowan[a,d] and Marla Tuffin[a]

[a]Institute for Microbial Biotechnology and Metagenomics, University of the Western Cape, Bellville, Cape Town, South Africa

[b]Technology Transfer Office, University of the Western Cape, Bellville, Cape Town, South Africa

[c]Centre for Bioprocess Engineering Research (CeBER), Department of Chemical Engineering, University of Cape Town, Rondebosch, Cape Town, 7701, South Africa

[d]Centre for Microbial Ecology and Genomics, University of Pretoria, Pretoria, South Africa

[e]Fruitique (Pty) Ltd, 27 Scheckter Rd Killarney Gardens, Cape Town, 7441, South Africa

*ituffin@uwc.ac.za, +27219599725, +27 21 9591506

Abstract

A thermophilic compost metagenomic library constructed in *E. coli* was functionally screened for novel esterases. Of the 110,592 fosmid clones screened, 25 clones demonstrated degradative activity on glyceryl tributyrate (a hit rate of 1:4,423). Four clones displayed ferulic acid esterase activity and were sequenced using 454 Titanium sequencing technology. EstG34, a 410 amino acid protein, was identified as having high sequence identity with a

number of bacterial β-lactamases. EstG34 has the S-X-X-K motif which is conserved in class C β-lactamases and Family VIII carboxylesterases. Purified recombinant Est34 had a molecular mass of 42 kDa and displayed hydrolytic activity towards a variety of *p*-nitrophenylesters, hydroxycinnamic acid esters and α-naphthol acetate. EstG34 represents the first Family VIII carboxylesterase and β-lactamase fold enzyme, able to hydrolyse ferulate and a number of other hydroxycinnamic acid esters. In addition, EstG34 is the first reported FAE to not adopt the α/β hydrolase conformation. The sequence similarity and wide substrate utilization capability of this esterase complicates its placement within current classification systems, but also draws attention to the enzyme's potential versatility in a range of industrial applications.

Introduction

Microbial communities are a rich source of metabolic and genomic diversity, and are critical components of functioning ecosystems. Both direct cultivation techniques and indirect molecular methods have been used to investigate and exploit this wealth of microbial diversity. It is now widely accepted that traditional methods involving the cultivation and isolation of microorganisms only provide limited access to the total genetic information within an environment [1]. Metagenomic methods entailing the isolation and cloning of microbial DNA directly from environmental samples circumvent the limitations of culturing and have been used to identify a variety of biotechnologically relevant microbial products [2, 3].

Composting is a self-heating, aerobic, solid-phase biological system which accelerates the natural process of biodegradation and mineralisation of organic matter. Microorganisms are essential to the composting process and compost microbial communities have been shown to produce a wide range of robust lignocellulolytic enzymes [4]. Carboxylic esterases are a diverse class of hydrolases which catalyze the cleavage and formation of ester bonds, including ester linkages in plant cell wall polysaccharides. They are commercially significant biocatalysts in biotechnological processes, and have a wide range of industrial and medicinal applications [5, 6]. Carboxylic esterases are ubiquitous in nature, having been identified in all domains of life (Bacteria, Archaea and Eukaryotes), as well as in some viruses [7]. Within the carboxylic ester hydrolase family, there are two recognized groups; the lipases (EC 3.1.1.1, triacyglycerol hydrolases) and esterases (EC 3.1.1.3, carboxylester hydrolases). The current sub-classification of esterases is based on biochemical and physiological properties of the enzymes, particularly substrate specificities, as well as primary and tertiary structures [8].

Here we present the use of metagenomics to identify a novel Group VIII esterase with broad feruloyl and aromatic ester substrate specificity.

Materials and Methods

Samples and Strains

E. coli strains EPI300 (Epicentre® Biotechnologies), GeneHogs (Invitrogen) and Rosetta pLysS (Novagen) were used for all cloning and expression studies. Compost samples were collected from a commercial compost production facility located in the Western Cape Province of South Africa during the summer season of 2009 (GPS position 34°S 2' 53.35'', 18°E 31' 45.71''). The compost source material consisted of an unspecified mix of wood chips and sawdust, with lesser amounts of plant debris and bovine manure. The average temperature of the compost at the point of sampling was 70 °C and the pH was 6.1. Elemental analysis (BemLab Laboratories, Strand, South Africa) indicated the composition of the compost was as follows: C 25.8%; N 1.1%; K 0.72%.

Metagenomic fosmid library construction

Metagenomic DNA was extracted using the chemical lysis method described by Zhou et al. [9] with slight modifications (addition of SDS (2% w/v) and PVPP (0.5% w/v) to the extraction buffer). Persistent humic- and phenolic compounds were removed by electrophoresis in low melting point (LMP) agarose. DNA samples were loaded onto a 0.7% LMP agarose gel in 1 X TAE buffer and electrophoresis was performed overnight at 1.5 V/cm. High-molecular weight (>35 kb) DNA fragments were excised from LMP agarose and treated with agarase (Fermentas) followed by isopropanol precipitation according to the manufacturer's instructions. DNA was quantitated using the Quanti-iT dsDNA BR assay kit with a Qubit fluorimeter as described by the manufacturer (Invitrogen). The metagenomic

DNA library was constructed using the CopyControl™ Fosmid Library production kit (Epicentre® Biotechnologies) according to the manufacturer's guidelines. Briefly, purified metagenomic DNA was end-repaired and ligated to the CopyControl pCC1Fos™ vector. Fosmid clones were packaged by MaxPlax Lambda phage and transfected into *E. coli* EPI300-T1® cells. Transformants were selected on Luria-Bertani (LB) agar containing chloramphenicol (12.5 μgmL^{-1}). To verify the size of the library fosmid DNA was extracted from 24 clones and digested to completion with *Hin*dIII and *Eco*RI (Fermentas). Insert size was estimated by agarose gel electrophoresis and visualization with an Alpha Imager 3400 imaging system.

Construction of 16S rRNA gene clone library and phylogenetic analysis

Transformants plated on LB agar were pooled and total fosmid DNA extracted using a Qiagen®Midi kit. The 16S rRNA gene was amplified from the purified fosmid DNA using the universal bacterial primers E9F and U1510R [10, 11]. Cycling conditions for the universal bacterial primers included denaturation (4 min at 94 °C), followed by 30 cycles of denaturation (30 s at 94 °C), annealing (30 s at 52 °C), and extension (105 s at 72 °C), followed by a final extension (10 min at 72 °C). PCR was carried out in 50 μl reaction volumes containing 2 mM MgCl$_2$, 1 U Dream *Taq* polymerase (Thermo Scientific), 1 X PCR buffer, 200 μM dNTPs, 0.5 μM of each primer and 100 ng fosmid DNA. PCR products were purified with a GFX^T PCR DNA purification kit (Illustra) and cloned into pGEM®-T Easy vector (Promega) according to the manufacturer's instructions. Putative recombinant clones were screened by colony PCR using universal M13 primers. Compost fosmid library (CFL) clones were grouped manually into ribotypes based on ARDRA restriction patterns generated by single digestions using *Alu*I and *Bsu*RI (isoschizomer of *Hae*III) (Fermentas). Partial 16S rRNA gene sequences were obtained for at least one representative clone from each ribotype. Chromatograms were edited with Chromas software version 7.0.5.2 [12] and sequences were

assembled in DNAMAN version 4.13 (LynnonBioSoft). Local alignments were obtained by performing a standard nucleotide-nucleotide BLAST search (BLASTn) [13] of the GenBank database. For phylogenetic analysis reference strains identified from the BLAST search were selected for comparison. Sequences were aligned using CLUSTAL_W [14] and checked manually for errors. Phylogenetic analyses were conducted using MEGA version 5.0 [15] and trees were constructed using the neighbour-joining [16] algorithm. The robustness of tree topology was evaluated by bootstrap analysis [17] based on 1,000 resamplings.

High-throughput screening of fosmid library clones

The library (approximately 10^5 clones) was inoculated into 384-well microtitre plates containing 50 µl LB broth supplemented with 12.5 µg mL^{-1} chloramphenicol using a QPix2 automated colony picker (Genetix). Primary screening was performed by gridding the library from the 384-well microtitre plates onto 22 x 22 cm Q-trays (Genetix) containing 500 ml basal medium (LB agar) supplemented with 12.5 µg mL^{-1} chloramphenicol (Sigma) and 0.01% (w/v) L-arabinose (Sigma). Clones were screened for general esterase activity on basal agar containing 1% (w/v) Gum Arabic (Sigma) and 0.1% (v/v) glycerol tributyrate (Sigma). A clear halo surrounding a colony indicated lipolytic and/or esterase activity [18]. Positive esterase clones were screened for ferulic acid esterase activity by plating onto basal agar supplemented with 0.4% (w/v) ethyl 4-hydroxy-3 methoxycinnamate (ethyl ferulate) (Sigma). A clear halo surrounding a colony indicated ferulic acid esterase activity [19].

Fosmid Pyrosequencing and ORF analysis

Equimolar concentrations of DNA from 36 fosmids were pooled and sequenced on a Roche 454 GS FLX sequencer by Inqaba Biotechnology, Pretoria, South Africa. Sequence reads were assembled with CLC Genomics Workbench software (www.clcbio.com) and Sequencher® sequencing software (www.genecodes.com) to produce contiguous sequences.

ORFs were predicted using Softberry's bacterial operon and gene prediction tool, FGENESB (http://linux1.softberry.com) [20]. Signal peptides were predicted with SignalP version 4.0 [21]. Homology and similarity searches of the translated sequences of the predicted ORFs were performed using the BLASTp program [13, 22, 23].

3   **DNA sequence analysis**

Multiple alignments of the predicted amino acid sequence of EstG34 were performed using MEGA 6.0 [24]. A phylogenetic tree of EstG34 with other family VIII esterases, class-C β-lactamases, ferulic acid esterases and DD-peptidases was constructed using the neighbour-joining method [16]. The robustness of tree topology was evaluated by bootstrap analysis [17] based on 1,000 resamplings.

9   **Expression and purification of recombinant EstG34**

The EstG34 gene was amplified using fosmid 6-C1 as the template and the following primers: 5'-GCT **CAT ATG** GAC GCC CAA TCT CAG TGG-3' and 5'-TAT **CTC GAG**GTC GAC GAT GGC CGA ATA-3' (nucleotides in bold denote *Nde*I and *Xho*I restriction enzyme sites, respectively). The EstG34 gene was cloned into the expression vector pET-21a (+) and the recombinant plasmid was transformed into *E. coli* Rosetta pLysS cells. Cells were cultured to an optical density of approximately 0.8 at 600 nm at 30 °C, after which 0.8 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to the culture for the induction of protein expression. Following 5 h of cultivation at 30 °C, cells were harvested by centrifugation (6,000 x *g* for 15 min at 4 °C) and resuspended in 1X PBS buffer (pH 7.4). Cells were disrupted by sonication and cellular debris was harvested by centrifuged (8,000 x *g* for 15 min at 4 °C). EstG34 was purified by HIS-tag nickel affinity chromatography using the HIS-Bind® resin and buffer kit (Novagen, USA) according to the

manufacturer's instructions. The eluate was dialysed overnight in a 6 ml Slide-A-Lyzer® cassette (Thermo Fisher Scientific) against 50 mM sodium phosphate; 50 mM sodium chloride (pH 7.9) buffer and stored at 4 °C. The protein concentration was estimated by the method of Bradford [25] using the Bio-Rad protein assay kit with bovine serum albumin as a standard. Protein purity was examined by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) under denaturing conditions, as described by Laemmli [26].

Esterase activity assay

Esterase activity assays were performed using a standard colorimetric method by measuring the release of *p*-nitrophenol from *p*-nitrophenyl (*p*-NP) esters. The release of *p*-nitrophenol was monitored continuously at 410 nm using a Cary 50 Bio spectrophotometer (Varian, CA, USA). All assays were prepared and analysed in triplicate. Enzyme activity was measured at 25 °C and one unit of activity was defined as the amount of enzyme releasing 1 μmol of *p*-nitrophenol per minute under the defined assay conditions. The effect of pH on esterase activity was studied by measuring activities on *p*-NP decanoate over a pH range of 4.5 – 11. The following range of buffer systems were used: 100 mM sodium acetate (pH 4.5 to 5.5), 100 mM morpholineethanesulfonic acid (MES) (pH 5.5 to 7), 100 mM TRIS-HCl (pH 7 to 9), and 100 mM *N*-cyclohexyl-3-aminopropanesulfonic acid (CAPS) (pH 9 to 11). The optimal temperature for enzyme activity was determined for a temperature range of 20 to 50 °C using the standard assay towards *p*-NP decanoate. To test thermal inactivation the recombinant enzyme was preincubated at a range of temperatures for 1 hr. Residual esterase activity was assayed on *p*-NP propanoate at 25 °C in 100 mM sodium phosphate pH 7.5, 100 mM sodium chloride.

Substrate specificity assays

p-nitrophenyl esters:

The hydrolytic activity of EstG34 against different fatty acid esters was investigated using the following *p*-nitrophenyl esters: acetate (C2); propanoate (C3); octanoate (C8); decanoate (C10); dodecanoate (C12) and hexadecanoate (C16).

Hydroxycinnamic acid methyl esters:

Activity towards the synthetic substrates methyl ferulate (Key Organics Ltd, Cornwall, UK), methyl sinapate, methyl caffeate and methyl *p*-coumarate (APIN chemicals, Abingdon, Oxon, UK) were based on the absorption difference of the free acid and the respective methyl ester. The reaction was initiated by the addition of enzyme solution to the assay buffer (100 mM sodium phosphate pH 7.5, 100 mM sodium chloride) containing 50 µM synthetic substrate and was performed at 25 °C for 1 hour in a SPECTROstar^{Nano} microplate reader (BMG Labtech). Absorbance was measured at 1 min intervals at 340 nm and activity was determined using calibration curves of the substrate/product. Relative activities were expressed as a percentage of the highest activity.

4-nitrophenyl ferulate:

Ferulic esterase activity of EstG34 was determined quantitatively using 4-nitrophenyl ferulate (4-NPF) as the substrate [27]. The substrate was synthesized according to the method of Hegde et al. [28]. The assay was carried out in 100 mM sodium phosphate (pH 7.5); 100 mM sodium chloride buffer containing 1 mM 4-NPF. The liberated free *p*-nitrophenol was measured at 410 nm. One unit of enzyme activity is defined was the amount of enzyme releasing 1 µmol of *p*-nitrophenol from 4-NPF in 1 min at 25 °C.

Feruloyl esterase activity was also determined using the substrates 4-nitrophenyl 5-O-trans-feruloyl-α-L-arabinofuranoside (NPh-5-Fe-Araf) and 4-nitrophenyl 2-O-trans-feruloyl-α-L-arabinofuranoside (NPh-2-Fe-Araf). These substrates were synthesised as detailed by

Mastihubova et al. [27]. The assay was carried out in 100 mM sodium phosphate buffer (pH 6.5) containing 12 µl DMSO and 3 µl Tween 20 per ml substrate (NPh-5-Fe-Araf or NPh-2-Fe-Araf at 2.5 mM) [29]. The reaction mixture was incubated to 30 $^{\circ}$C before addition of EstG34. Measurements of absorbance at 420 nm were taken over a period of 3 hours and 1 unit of activity was defined as an absorbance change of 1.0 min^{-1} (A_{420})

Naphthyl acetate:

Deacetylase activity was determined by measuring α-naphthol released from naphthyl acetate using the Fast Garnet liquid assay as described by Koseki et al. [30]. The assay was performed in 50 mM sodium phosphate buffer (pH 7.5) with 0.8 µmol α-naphthyl acetate (Fluka). After incubation at 37 $^{\circ}$C for 10 min, the reaction was terminated by the addition of 110 µL Fast Garnet GBC (6 mg mL^{-1} stock solution in 10% (w/v) SDS) and the absorbance measured at 560 nm. One unit (U) of acetyl xylan esterase (AXE) activity was defined as the amount of enzyme required to produce 1 µmol of product per min under the defined assay conditions.

Acetyl xylan hydrolysis:

To determine hydrolytic activity of EstG34 on acetyl xylan, an endoxylanase-free 1% solution of acetyl xylan in 0.1 M sodium phosphate buffer (pH 6.5) was prepared. The reaction mixtures with varying amounts of the enzyme (5 - 50 µl) were incubated overnight at 37 $^{\circ}$C in sealed, thin-walled glass test tubes. Precipitation in the reaction mixtures was visually determined [31].

Kinetic Measurements

Esterase activity kinetics were determined for four different substrates: *p*-nitrophenyl acetate, *p*-nitrophenyl propionate, *p*-nitrophenyl ferulate and α-naphthyl acetate. The reactions

Table 1: Kinetic parameters for EstG34 on selected synthetic substrates.

Substrate	K_M (mM)	Specific activity (U mg^{-1})	k_{cat}(s^{-1})	k_{cat}/K_M (mM^{-1}s^{-1})
p-nitrophenyl acetate	0.011 ± 0.001	26.25 ± 30	19.70 ± 30	1790 ± 204
p-nitrophenyl propanoate	0.023 ± 0.004	151 ± 11	113.25 ± 90	4913 ± 287
p-nitrophenyl ferulate	0.213 ±0.017	10.34 ±0.70	7.49 ± 0.81	35.13 ± 4.24
α-naphthyl acetate	0.84 ± 0.09	1337 ± 56	969 ± 48	1154 ± 74

contained different substrate concentrations (Table 1) and the parameters (K_M and k_{cat}) were calculated by Michaelis-Menten direct plots generated using Graphpad Prism® version 5.02 (La Jolla, CA).

Effects of inhibitors and detergents

The effect of various metal ions ($CaCl_2$, $CoCl_2$, $FeSO_4$, $MgSO_4$, $MnSO_4$, $NiSO_4$ and $ZnSO_4$), EDTA, organic solvents (ethanol, methanol, isopropanol, acetone, acetonitrile and DMSO) and chaotropic agents (CTAB, SDS, Tween 20, and Triton X-100) on EstG34 activity was determined. The enzyme was incubated in the presence of the inhibitor/detergent in 100 mM sodium phosphate (pH 7.5), 100 mM sodium chloride for 30 min at 25 °C, and assayed for enzyme activity. The activity of the enzyme without the addition of metal ions or no exposure to solvents or detergents was defined as 100% activity.

Nucleotide sequences

The 16S rRNA gene sequences amplified from the library have been submitted to the GenBank database under the Accession Nos. HQ694846.1–HQ694892.1.The nucleotide sequence for EstG34 has been deposited in the GenBank database under the accession number KJ937963.

Results

The compost metagenomic library was comprised of 150,000 clones with an average insert size of 30.67 kb. The estimated 4.65×10^9 bp of insert DNA represents a coverage equivalent to approximately 1,500 bacterial genomes [32]. Analysis of the bacterial diversity captured within the fosmid library identified 39 ribotypes, of which 9 ribotypes were represented by a

single clone. Phylotype richness estimators S_{ACE} and S_{Chao1} predicted 56 and 60 ribotypes, respectively (data not shown).

Phylogenetic analysis, as shown in Supplementary Figure S1, revealed that the majority of library clones were related to γ-*Proteobacteria* (63%), α-*Proteobacteria* (20%), *Bacteroidetes* (7%) and *Firmicutes* (4%). The library was dominated by clones (70%) that were most similar to uncultured bacteria from environmental or clinical sources, which is common to other soil metagenomes [33]. Previous studies have shown that compost is a rich source of microorganisms possessing lignocellulose hydrolyzing abilities [34, 35]. In this study, several of the ribotypes identified were related to known lignocellulolytic enzyme-producing strains, confirming that genomes encoding lignocellulolytic enzymes were well represented in the library (denoted by a black triangles (▲) in Supplementary Figure S1). Approximately 10^5 fosmid clones were functionally screened for general esterase activity on glyceryl tributyrate, and a total of 25 clones were confirmed positive for lipolytic activity (25:100,000), a hit rate of 0.025%. The reported hit rates for esterase clones from metagenomic studies appear to vary significantly, with values ranging from 0.87% from termite guts [18] to 0.001% from various environmental soils (1:60,000) [36]. The esterase positive clones were further analysed for ferulic acid esterase activity by screening on ethyl ferulate. Four of the 25 clones demonstrated the ability to degrade this substrate. One of these clones, pCC1fos6-C1, showed strong feruloyl esterase activity and contained a 1,233-bp ORF (EstG34) encoding a polypeptide of 410 amino acids. No signal peptide was identified. EstG34 displayed 77% primary sequence identity to a predicted β-lactamase from *Phenylobacterium zucineum* HLK1 (YP_002129844) (Figure 1). The closest characterised relative was EstM-N2 (49% amino acid identity), a metagenome derived cold-active esterase belonging to the family VIII esterase/lipase and class C β-lactamases [37]. Further sequence
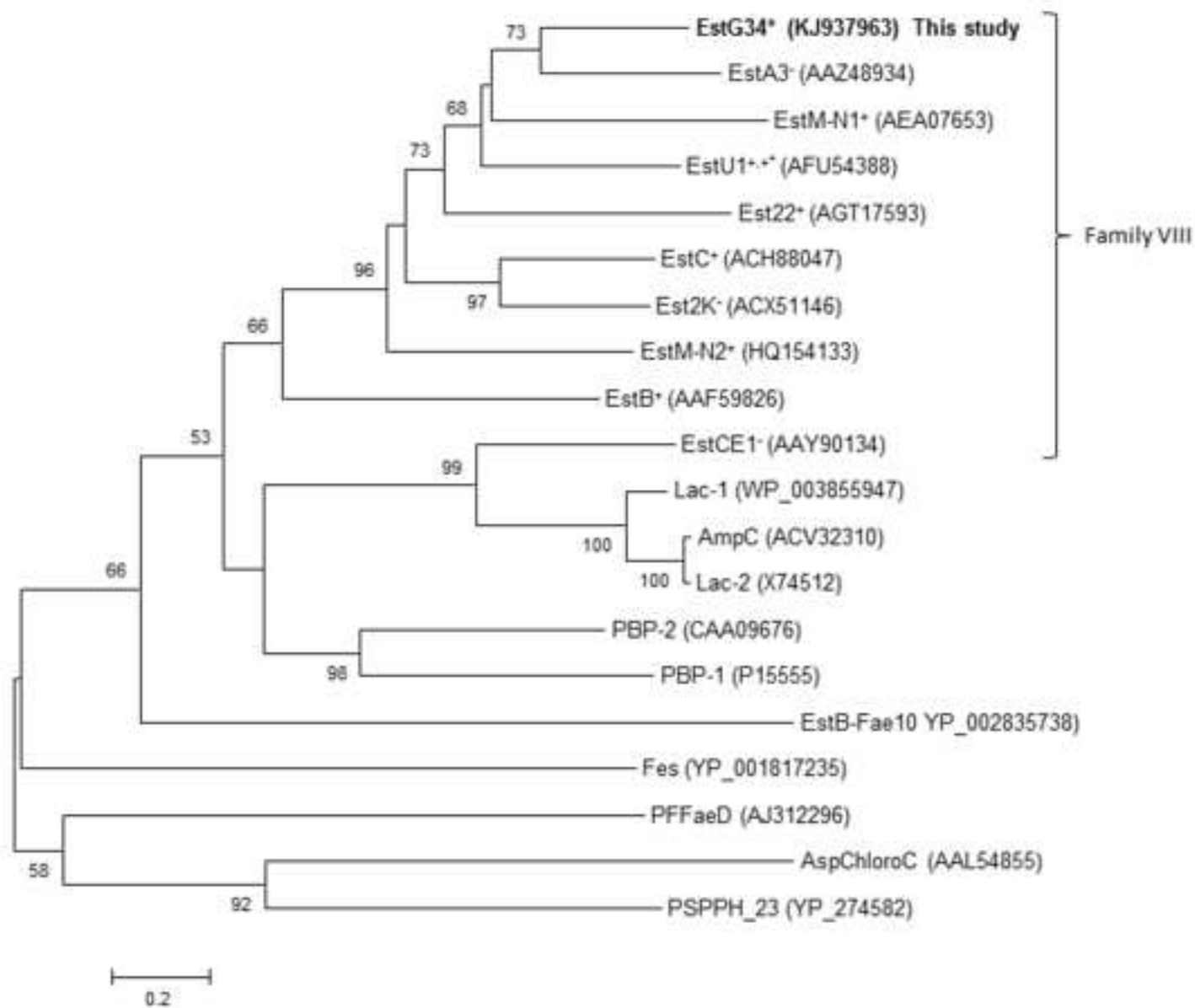
Fig 1 Evolutionary relationship of Est34 in relation to family VIII esterases, class C β- lactamases, type C and D feruloyl esterases [52], group 10 feruloyl esterases [53] and pencillin-binding proteins. The phylogenetic tree was generated using the neighbour-joining method (MEGA 6.0 software). All protein sequences were retrieved from GenBank, with the associated accession numbers indicated in parentheses. β-lactamase activity is reported for all family VIII esterases [+: hydrolysis of the β-lactam amide bond; +*: deacetylation of the β-lactam ring; +?: unknown if activity represents hydrolysis of the ester or amide bond; -: no β-lactamse activity]. The numbers at the nodes indicate the bootstrap percentages based on 1000 replicates. The bar represents 0.2 nucleotide substitutions per nucleotide position.

14

analysis searches using the Lipase Engineering database [38] and the Arpigny and Jaeger classification scheme [5] indicated that EstG34 was a family VIII carboxylesterase.

Unlike other microbial esterase families where the active site serine residue is typically located within the G-X-S-X-G pentapeptide motif, the serine residue of family VIII esterases is situated within the S-X-X-K motif and serves as the catalytic nucleophile [5]. The alignment (Figure 2) indicated that EstG34 contained the S-X-X-K consensus sequence (S-M-T-K[75]). This S-X-X-K motif is conserved within most of the β-lactamase superfamily proteins, which includes penicillin binding proteins (PBPs), DD-peptidases and other family VIII carboxylesterases [39, 40, 41]. However, two other highly conserved class C β-lactamase motifs (Y-A-N) and (K-T/S-G) [42] were not found in EstG34 (Figure 2).

EstG34 was expressed in *E. coli* and purified to near homogeneity by nickel-chelation chromatography (Figure 3). Using *p*-NP esters as a substrate, EstG34 was most active at a temperature of 41 ˚C and at pH 9. EstG34 showed a limited tolerance to low pH's, exhibiting only 50% of its maximum activity at pH 7.0, while being completely inactivated at pH 6.0 (Figure 4A). EstG34 can be characterized as being mildly alkaliphilic, with high activity at pH 9.0 and 70% of maximal activity at pH 10.0. The composting process involves different temperature stages, therefore despite having prepared the metagenomic library from material in the thermophilic stage it is not unexpected that the library contained DNA from mesophilic organisms. Furthermore, as the compost material contained bovine manure DNA from enteric organisms may be present in the library. This may account for the relatively low thermal stability of EstG34 (Figure 4B), which retained 90% of its activity after one hour at 40 °C and was completely deactivated after 30 minutes at 50 °C (Figure 4C). Divalent cations showed no significant influence on enzyme activity (Figure S2). Of the chaotropic agents tested, only 1 mM SDS showed significant inhibition of EstG34 activity (Figure S3), while the enzyme was stable in all the solvents tested (Figure S3).

S-X-X-K

```
 1. EstG34      F G S M D L E - - - R N K P M A D D A I F R I Y  S M T K  P I
 2. EstB        Q G L A D R E - - - A G R P M R E D T L F R L A  S V T K  P I
 3. EstC        Y G Y Q D V E - - - N Q I P V S E D T L F R I Y  S M T K  P V
 4. Est22       Q G M K D V E - - - R N Q P M T K D T I F R M A  S M T K  P I
 5. EstCE1      Y G V A S R Q - - - T G K P I T N Q T L F E I G  S L S K  T F
 6. EstA3       S G - - - - - - - - - - - V A D D A I F R I A    S M T K  P V
 7. Est2K       Y G Y Q D M E - - - N E V A M S E D S I F R I F  S M T K  P I
 8. EstU1       F G L R D P D - - - T K E P M T A E T I F R I Y  S M S K  P I
 9. EstM-N1     S G T S D E A - - - G K I P L A D D A I Y R I Y  S M T K  P I
10. EstM-N2     Q G H R D V E - - - R G T P M T E D S I L R I Y  S M S K  P I
11. AmpC        W G K A D I T - - - N N H P V T Q Q T L F E L G  S V S K  T F
12. Lac-1       F G K A D I A - - - A N K P V T P Q T L F E L G  S I S K  T F
13. Lac-2       W G K A D I A - - - N N H P V T Q Q T L F E L G  S V S K  T F
14. PBP-1       E G V A D R A - - - T G R A I T T T D R F R V G  S V T K  S F
15. PBP-2       A G I A D L R - - - T K K P M K T D F R F R I G  S V T K  T F
16. AspChloroC  M R L P V S W N - - - G R F L Y Q G T D G        N I A T  A T
17. pFFaeD      Y S Q C I Q G - - - - - - N C S P A S S T S S T  S S T K  T T
18. EstB-Fae10  N G T F D D L - - - - A T R - H G M L A V Y P T G  - V E Q  H F
19. Fes         E P P A D Y R - - - - - - - A L P N D A L R G R  L E P F  Q Y
20. PSPPH_23    F R L E L P T G T W A Q R Y L Q V G C G G L C G   N I N S  T V   82 bp
```

Y-A-N

```
 1. EstG34      P L L Y Q P G E R W M  Y S I  S T D I C G - - A L V E L I S
 2. EstB        A P L S A P G S G W Q  Y S -  L A L D V L G - - A V V E R A
 3. EstC        P L R Q Q P G T L W H  Y S V  S V D V Q G - - Y L V E V L S
 4. Est22       P L R Y E P G T R W E  Y S A  A T S V V G - - R L V E V I S
 5. EstCE1      Q P P H A V G S Y R V  Y S N  L G I G M L G - - M I T A K S
 6. EstA3       P L E F D P G T Q W N  Y S V  S T D V L G - - I L I E R I S
 7. Est2K       P L R Q Q P G T Q W H  Y S V  S V D V Q G - - Y L V E V L S
 8. EstU1       P L V Y Q P G T T W D  Y G H  S T D I L G - - R V V E V V S
 9. EstM-N1     P L A S Q P G S Q W R  Y S L  S T D V L A - - R V L E V A S
10. EstM-N2     P L E F S P G E R W N  Y S L  S T D V L G - - Y L V E V I S
11. AmpC        Q P Q W A P G A K R L  Y A N  S S I G L F G - - A L A V K P
12. Lac-1       Q P Q W K P G T T R L  Y A N  A S I G L F G - - A L A V K P
13. Lac-2       Q P Q W T P G A K R L  Y A N  S S I G L F G - - A L A V K P
14. PBP-1       G V T N A P G A A Y S  Y S N  T N F V V A G - - M L I E K L
15. PBP-2       P P D F A P G K G W S  Y S N  T G Y V L L G - - I L I E K V
16. pFFaeD      H W L G T M G N V V S  G S Y  Y G I Q P L A G D N A I F V A
17. AspChloro   A A Q L Y T A Q Q L R  R V A  T D E N D L S T A L T - - - -
18. EstB-Fae1   A S T L - - - - - - -  - - G  A G A N H A P T N - - - - - -
19. PSPPH_2     T L Y H G W M A R S N  R G A  D G K P I L F A S R L P L L H   225 bp
20. Fes         P N G R - - - - - - -  - -    - - - - - - - - - - - - - -
```

K-T/S-G

```
 1. EstG34      D E V S T G S L G -  A G D  Y Y W G G A A S T I F L V D P K E
 2. EstB        D P A A A G T P Q H  A G T  L Q W G G V Y G H S W F V D R A L
 3. EstC        D P A R N N - G M S  E G S  Y W W W G I G G T W F W I D P V E
 4. Est22       D R A V A H T S D T  E G A  V S W G G A F G T L F W V D P E E
 5. EstCE1      K Q D Q G A A W Y N  K T G  S T G G - - - F S T Y A V F I P S
 6. EstA3       D S V A T L T P C S  T G D  F Y W G G M Y S T A F F V D P V E
 7. Est2K       D P A R N D - G M S  R G S  Y W W W G I G G T W F W I D P V E
 8. EstU1       E A G V T P V E G S  V G D  L S W G G A G G T V F W I D P K E
 9. EstM-N1     D L G Q A M S L T A  D G E  F G W E G A A G T Y F W V D P Q N
10. EstM-N2     D P V R N G H P G S  A G S  Y F W G G L A S T L F W V D P L E
11. AmpC        A P A V K A S W V H  K T G  S T G G - - - F G S Y V A F V P E
12. Lac-1       A P P V K A S W V H  K T G  S T G G - - - F G S Y V A F I P E
13. Lac-2       A P A V K A S W V H  K T G  S T G G - - - F G S Y V A F V P E
14. PBP-1       D L S C G I S V Y G  H T G  T V Q G - - - Y Y T Y A F A S K D
15. PBP-2       K L P S G V S I W G  H T G  G I L G - - - F T T L V G G K L G
16. pFFaeD      T E Y S G C D S E H  P V W  W I A F D G P H E P L A T D A G A
17. AspChloro   L S A T N S T Y S E  S A M  S F M I P P D E L N L D K L R L H
18. EstB-Fae1   P V E L V S V E G M  G H V  I P S G N E L D P - - - - - - - -
19. PSPPH_2     W H G W A D P N I S  P L N  T L A Y H E A V E A Q M G K T R T
20. Fes         V P D L A A A R A Q  L K L  L Y I S C G N K D G L I N F S Q R   438 bp
```

Fig. 2. Multiple sequence alignment of EstG34 and related Family VIII carboxylesterases, class C β-lactamases, penicillin-binding proteins, and class C and D feruloyl esterases. Family VIII esterases are represented by EstB (*Burkholderia gladioli*, AAF59826), and the metagenome-derived esterases EstC (ACH88047) Est22 (AGT17593), EstCE1 (AAY90134), EstA3 (AAZ48934), Est2K (ACX51146), EstU1 (AFU54388), AFU54388 EstM-N1 (AEA07653) and EstM-N2 (HQ154133). Class C β-lactamases are represented by Lac-1 (*Escherichia coli*, AAA23441), Lac-2 (*Enterobacter cloacae*, P05364) and AmpC (*Citrobacter freundii*, ACV32310). Penicillin-binding proteins are represented by PBP-1 (*Streptomyces* sp. R61, P15555) and PBP-2 (*Bacillus cereus*, CAA09676). Feruloyl esterases are represented by AspChloro (*Acinetobacter* sp. ADP1, AAL54855), EstB-Fae10 (*Corynebacterium aurimucosum*, YP_002835738), PSPPH (*Pseudomonas syringae*, YP-274582), pFFaeD (*Penicillium funiculosum*, AJ312296) and Fes (*Opituts terrae*, YP_001817235). The conserved S-X-X-K carboxylesterase family VIII motif, as well as the Y-A-N and K-T/S-G class C β-lactamase motifs are shown.

16
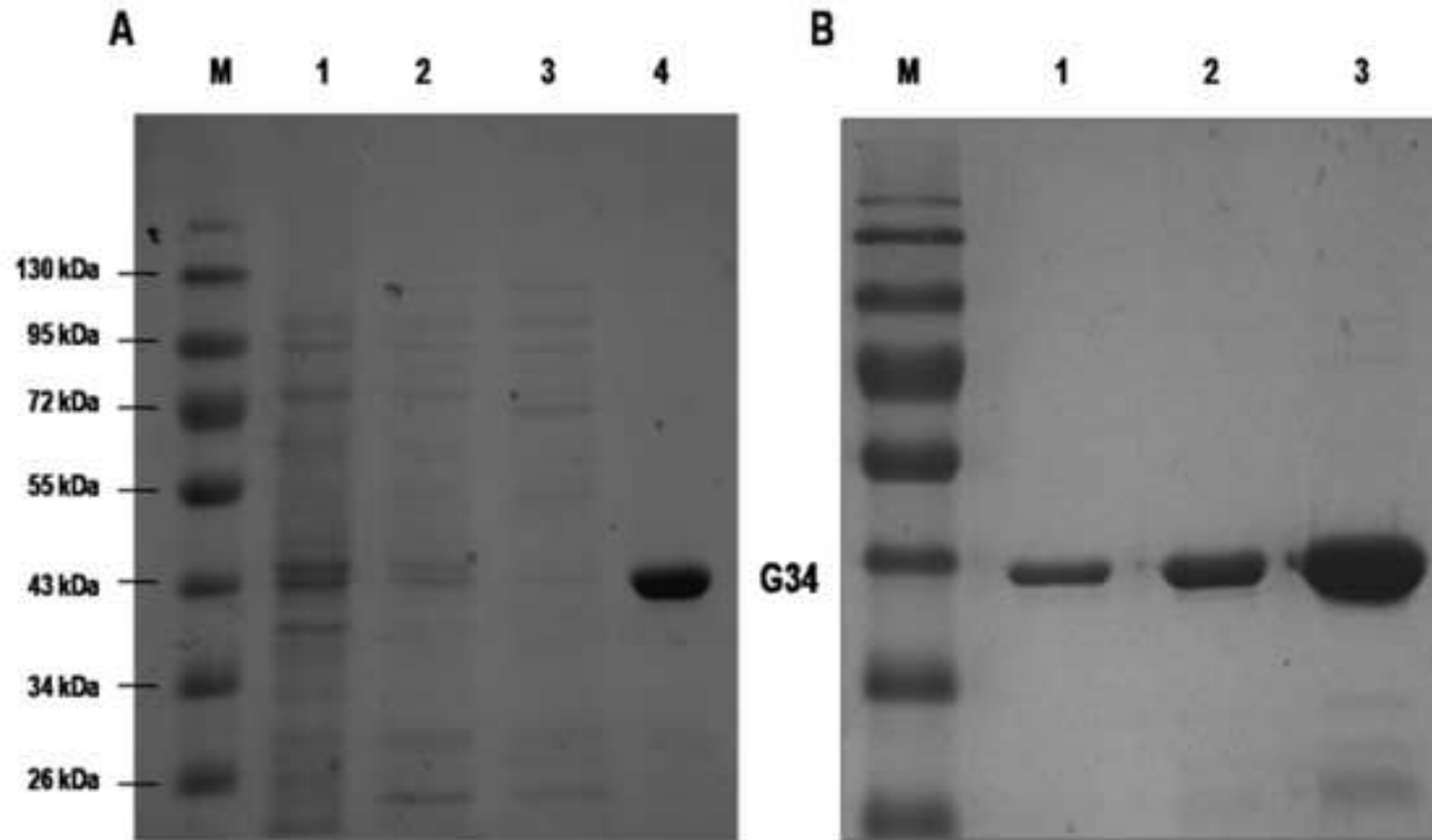
Fig 3 SDS-PAGE (12%) electropherogram showing different fractions collected for G34 purification by HIS-bind nickel affinity chromatography: lane M – molecular weight markers, lane 1 – flow through fraction, lane 2 – binding fraction, lane 3 – wash fraction and lane 4 – elute fraction. (B) Purified G34 recombinant protein following dialysis. Lanes 1 – 3 showing increased loading of the protein.
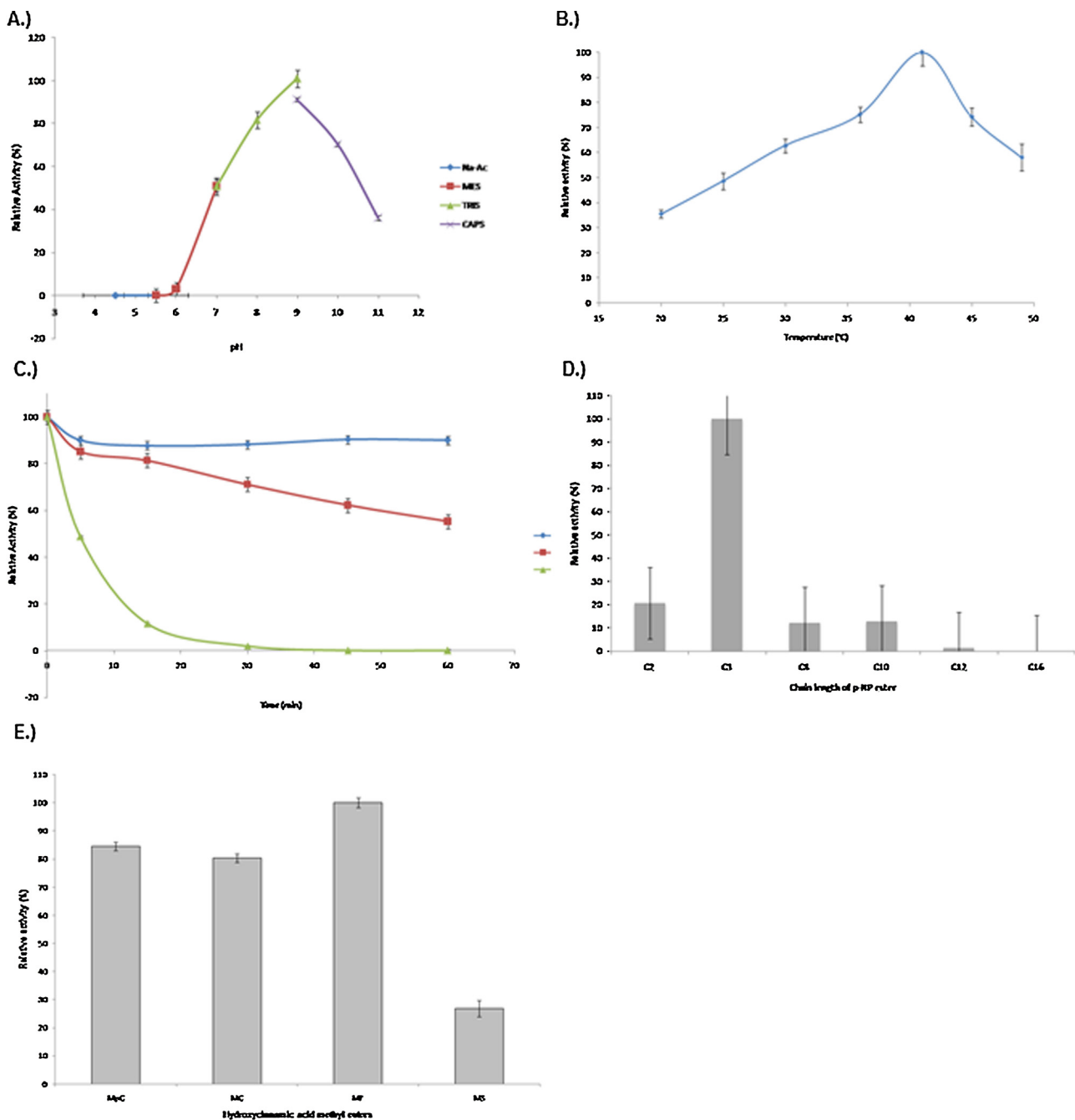
Fig. 4. (A) Effect of pH on activity of EstG34 toward *p*-NP decanoate. Assay buffers used were 100 mM sodium acetate (pH 4.5 to 5.5), 100 mM MES (morpholineethanesulfonic acid) (pH 5.5 to 7), 100 mM Tris-HCl (pH 7 to 9), and 100 mM CAPS (pH 9 to 11). (B) Effect of temperature on activity of EstG34 toward *p*-NP decanoate. (C) Thermal inactivation profile of EstG34 at 40 °C (◊), 45 °C (□) and 50 °C (Δ). Activity is expressed as a percentage of that at zero time in the standard assay (100 mM sodium phosphate, 100 mM NaCl (pH 7.5), 25 °C) toward *p*-NP propionate. (D) Activity of EstG34 at 25 °C towards*p*-NP esters of various chain lengths (C2, acetate; C3, propanonate; C8, octanoate; C10, decanoate; C12, dodecanoate; and C16, hexadecanoate) in 1 mL reactions containing 100 mM sodium phosphate (pH 7.5), 100 mM NaCl, 1% acetonitrile, and 0.5 mM *p*-NP ester substrate. (E) Ferulic acid esterase activity of G34 towards hydroxycinnamic acid methyl esters, represented as relative activity (% from the highest activity). Substrates are ordered (left to right) by increasing degree of substitution. MpC: methyl *p*-coumarate, MC: methyl caffeate, MF: methyl ferulate, MS: methyl sinapinate.

The hydrolytic activity of EstG34 was determined with various p-NP fatty acid esters and other ester compounds. EstG34 demonstrated high catalytic activity against short-chain fatty acids (Figure 4D). Maximum activity was obtained with p-NP acetate (C3) with greatly decreased activity toward longer chain fatty acids (>C8). Kinetic analysis of EstG34 activity, however, suggests that p-NP propanoate is the preferred substrate. Although a 2-fold higher affinity for p-NP acetate was observed (Table 1), the k_{cat} for p-NP propanoate was 5.8-fold higher than that of p-NP acetate, and a higher k_{cat}/K_M value was observed for p-NP propanoate. The kinetics of EstG34 were also assessed on the acetylated substrate α-naphthyl acetate. EstG34 demonstrated the highest specific activity and turnover rate on this substrate, although catalysis was more efficient with p-NP acetate. The much higher K_M for α-naphthyl acetate suggests that the lower affinity for this substrate is due to steric hindrance by the bulky biphenolic structure of α-naphthyl. EstG34 showed no deacetylase activity on acetyl xylan as demonstrated by the absence of precipitation.

On hydroxycinnamic ester substrates, EstG34 demonstrated preference in the order of > methyl ferulate > methyl p-coumarate > methyl caffeate > methyl sinapinate (Figure 8). Of all the substrates tested methyl sinapinate contains the highest degree of substitution and EstG34 had approximately 80% less activity on this substrate compared to methyl ferulate. The lower activity on substrates with a higher degree of substitution can be attributed to increased steric congestion around the ester group. Feruloyl esterase activity was quantitated using the substrates p-NP ferulate, p-NP 5-O-trans-feruloyl-α-L-arabinofuranoside and p-NP 2-O-trans-feruloyl-α-L-arabinofuranoside. EstG34 had a specific activity of 10.3 U mg^{-1}, 0.65 U mg^{-1} and 0.11 U mg^{-1} on these substrates, respectively (Supplementary Figure S4). While there are relatively few characterised feruloyl esterases in the literature [43, 44, 45] EstG34 activity is comparable to two characterised fungal feruloyl esterases [46].

Discussion

Carboxylester hydrolases are ubiquitous enzymes and the current esterase classification scheme consists of eight families (families 1-VIII), based primarily on sequence similarity, and less so on biochemical properties [5]. EstG34, identified by the functional screening of a compost metagenomic library, grouped with the family VIII esterases. This is a poorly characterised esterase family, and has previously included class C β-lactamases, penicillin binding proteins, DD-peptidases and a range of carboxylesterases. All family VIII esterases demonstrate sequence identity to β-lactamases and of the top 100 BLASTp hits for EstG34 96 were annotated as β-lactamases. As a result, family VIII esterases are typically tested for their ability to hydrolyse a variety of β-lactam substrates (Table 2). The majority either lack the activity or show negligible activity, despite the high sequence identity to the β-lactamases, while others have been described as exhibiting "promiscuous β-lactamase activity" [18, 21, 41, 47, 48, 49]. Consequently, it has been suggested that these esterases have evolved from the class C β-lactamases, where some have maintained this remnant activity, while others have lost the capability due to steric interference resulting from structural evolution [18, 41, 50].

EstG34 displayed no detectable activity when tested against ampicillin (data not shown). Interestingly, detailed selectivity assays conducted for Est22 [49] and EstB [40] demonstrated that the enzymes selectively hydrolysed the ester bond of a number of cephalosporin-based substrates, while leaving the amide bond of the β-lactam ring intact. This clearly differentiates them from bona fide β-lactamases. EstU1, on the other hand, cleaves the amide bond of first generation β-lactam antibiotics, although it is more efficient at hydrolysing short chain esters [48]. Other family VIII esterases (EstM-N1, EstM-N2, EstC) have demonstrated activity on nitrocefin, although it is not clear from the data presented whether this constitutes deacetylation or amide bond hydrolysis [18, 37]. The differentiation of deacetylation or

Table 2: Comparison of the specific activities of known family VIII esterases.

Enzyme	MW (kDa)	Specific Activity (U mg⁻¹)			Additional Substrates	pH_{opt}	T_{opt} (°C)	β-lactamase activity	Origin	Reference
		C2	C3	C4						
EstG34	42	26.2	151.3	Nd	ρ-Nitrophenyl-octanoate ρ-Nitrophenyl-decanoate ρ-Nitrophenyl-ferulate ρ-Nitrophenyl-5-O-trans-feruloyl-α-L-arabinofuranoside ρ-Nitrophenyl-2-O-trans-feruloyl-α-L-arabinofuranoside α-naphthyl acetate methyl-ρ-coumarate methyl caffeate methyl ferulate methyl sinapinate	9	41	-	Compost metagenome	This study
EstB		+	+	+	σ-Nitrophenyl-acetate σ-Nitrophenyl-propanoateσ-Nitrophenyl-butyrate σ-Nitrophenyl-valerate ρ-Nitrophenyl-valerate ρ-Nitrophenyl-caproate naphthol AS-acetate naphthol AS-2-chloro-acetate 2-naphthol 2-chloro-propanoate	7	43	+* Cephalosporin C 7-ACA	*Burkholderia gladioli*	41
EstC	43.3	465	nd	776	ρ-Nitrophenyl-caproate ρ-Nitrophenyl-dodecanoate linalyl acetate	nd	37	+* Nitrocefin	Acidic leachate metagenome	18
Est22	45	nd	+	+	ρ-Nitrophenyl-valerate ρ-Nitrophenyl-octanoate ρ-Nitrophenyl-decanoate	8	30	+* Nitrocefin Cephalothin Cephalosporin C 7-ACA	Acidic leachate metagenome	50
EstCE1	44	nd	nd	31.1	ρ-Nitrophenyl-valerate ρ-Nitrophenyl-caproate triacetin	10	47	-	Drinking water metagenome	48

					tripropionin tributyrin vinyl acetate vinyl propanoate vinyl butyrate					
EstA3	42	nd	nd	513	ρ-Nitrophenyl-valerate ρ-Nitrophenyl-caproate ρ-Nitrophenyl-octanoate triacetin tributyrin vinyl acetate vinyl propanoate vinyl butyrate vinyl caproate vinyl parylate vinyl laurinate	9	50	-		
Est2K	44	11.4	nd	17.1	ρ-Nitrophenyl-octanoate ρ-Nitrophenyl-decanoate ρ-Nitrophenyl-dodecanoate	10	50	-	Compost metagenome	42
EstU1	44	6.24	nd	22.24	ρ-Nitrophenyl-caproate ρ-Nitrophenyl-octanoate ρ-Nitrophenyl-decanoate	8.5	45	+ Nitrocefin Cephaloridine Cephalothin Cefazolin	Soil metagenome	49
EstM-N1	45	nd	nd	37.8	ρ-Nitrophenyl-caproate ρ-Nitrophenyl-octanoate ρ-Nitrophenyl-decanoate	9	30	$+^?$ Nitrocefin	Arctic soil metagenome	38
EstM-N2	44	nd	nd	7.5	ρ-Nitrophenyl-caproate ρ-Nitrophenyl-octanoate ρ-Nitrophenyl-decanoate	9	20	$+^?$ Nitrocefin		

+ : hydrolysis of the β-lactam amide bond

$+^*$:- deacetylation of the β-lactam

$+^?$: unknown if activity represents hydrolysis of the ester or amide bond

-    : no β-lactamase activity

7-ACA: 7-aminocephalosporanic acid

lactamase activity is a crucial determinant and a key differentiation between family VIII carboxylesterases from the β-lactamases. It is clear from the phylogenetic tree (Figure 1) that the true β-lactamases cluster separately from the family VIII carboxylesterases. However, the resolution of primary sequence comparisons does not sufficiently demonstrate the functional evolution (clustering based on deacetylation versus amide bond hydrolysis, Figure 1), which is expected to have resulted from a few critical amino acid changes in the catalytic site, as opposed to global structural modifications. Crystallisation and mutation studies of EstU and other family VIII esterases which have no activity on β-lactams should elucidate the active site changes which determine the differences in substrate specificity. Irrespective of β-lactam hydrolysis capacity, esters are the preferred substrates for all family VIII carboxylesterases (Table 2) and the β-lactams are poor substrates. Due to the clear functional distinction between class C β-lactamases and the carboxylesterases in family VIII, and for classification clarification purposes, we propose that this family be referred to as carboxylesterases and not β-lactamase fold enzymes.

EstG34 hydrolysed a wide variety of model esters, including hydroxycinnamic acids, ferulated *p*-nitrophenyl, and α-naphthyl acetate, showing considerable preference for the acetylated ester compounds. To the best of our knowledge, EstG34 represents the first family VIII carboxylesterase with feruloyl esterase activities. Specificity towards hydroxycinnamic acids has previously been used to sub-classify ferulic acid esterases (FAEs) as type A, B, C or D [51]. Interestingly, despite the fact that EstG34 demonstrated hydrolytic properties indicative of type C and D feruloyl esterases, it showed no significant sequence identity to type C and D FAEs (Figure 2). However, the A-D classification system is restricted to sequence similarity and substrate specificity on four model substrates only, and few fungal FAEs are included [51, 52]. Considering that FAEs belong to highly divergent protein

superfamilies, with each family having a different evolutionary pathway to the genesis of FAE specificity, a more robust classification scheme has been developed. The novel descriptor-based classification system is a more reliable scheme which groups functionally related FAEs that have common properties [52]. This was developed using FAEs belonging to all protein superfamilies and representing fungal, bacterial and plant origins, and classifies FAEs into 12 clusters (FEF1-12), which can be further sub-grouped based on the constellation and distance between the catalytic triad residues. Furthermore, its robustness is applicable even to poorly characterised enzyme families. EstG34 was predicted to group in FEF10, which does not include any of the type C or D FAEs. Furthermore, EstG34 does not show sequence similarity to FEF10 representatives (Figure 2). All FEF1-10 representatives contain the catalytic triad (Ser, His, Asp), as well as the consensus "nucleophilic elbow" (GXSXG, the universally conserved pentapeptide in which the catalytic serine residue is located) [5], although the positioning of these in the sequence varies greatly. This complicates EstG34 classification using the descriptor-based system since these conserved motifs, which EstG34 lacks, are a prerequisite for the classification using the FAE descriptor-based system [52]. Furthermore, feruloyl estersases are characteristically α/β-fold hydrolases, another characteristic that EstG34 lacks. EstG34 is the first FAE to be described which lacks the characteristics traditionally used to classify feruloyl esterases.

This presents a fascinating question regarding the general classification of feruloyl esterases. Both in this study and others, phylogenetic clustering of carboxylesterases does not correlate with substrate specificity [52]. EstG34 also demonstrates that novel bacterial FAEs adopting different structural folds and conserved motifs are yet to be discovered. While a preference for short to medium chain length p-NP esters remains the benchmark for classifying an enzyme as a 'true carboxylesterase', the results presented here suggest that a refinement of the classification of feruloyl esterases, and particularly the family VIII esterases, is required.

Specifically, the FAE classification now needs to consider the Family VIII type esterases which harbour the (Ser, Lys, Tyr) catalytic triad where the serine is located in the conserved (SXXK) motif. Given that many carboxylesterases show broad substrate ranges, and that a limited number of key amino acid substitutions in the active site can dramatically affect substrate specificity, we suggest that classification of these enzymes by 'substrate preference' may be misleading. Certainly, characterisation using a limited range of synthetic nitrophenyl-ester substrates offers little information on either the *in vivo* substrate or the classification of the enzyme. EstG34, with high sequence identity to class C β-lactamases and falling within the current classification scheme of both family VIII esterases and feruloyl esterases, emphasises the current classification dilemma. As more EstG34 homologues are discovered, and more in depth substrate characterisations of family VIII esterases is conducted, its classification will become clearer.

There now exists a number of reports of CAZy enzymes possessing more than one catalytic domain: for example, an enzyme from *Prevotella ruminicola* 23 contains two carbohydrate esterase domains which separately confer acetyl- and ferulic esterase activities [53]. Another example, *Clostridium thermocellum* XynY [54], contains xylanase and feruloyl esterase domains (clustering with Type D FAEs) [45]. However, based on preliminary modelling analyses, the multifunctional capability of EstG34 appears to be catalysed by a single catalytic domain, with no obvious carbohydrate binding domains (data not shown).

EstG34 displayed the ability to hydrolyse a wide variety of ester bonds found in natural lignocellulosic polysaccharides. Ferulated polysaccharides are essential in directing cell wall cross-linkages and serve as a defensive mechanism against invasion by plant pathogens [54]. The deconstruction of hemicellulose requires the cooperative effect of a number of different enzymes and microorganisms have developed cell associated multi-protein complexes, such as cellulosomes [56] and xylosomes [57], containing cellulases, xylanases and carbohydrate-

binding modules to achieve this. *O*-acetyl and methyl esterification products are the most frequently occurring substitutions found in various plant cell wall polysaccharides [58], where short xylo-oligosaccharides could be de-acetylated by non-specific acetyl esterases [31]. As the activity of endoxylanases can be partially or completely hindered by the presence of acetyl groups [59], EstG34 could serve to improve the accessibility of endoxylanases to the xylan backbone of plant polysaccharides.

Acknowledgements

Reference

[1] R.I Amann, W. Ludwig, K.H. Schleifer Microbiol. Rev. 59 (1995) 143-169.

[2] C.J. Duan, L. Xian, G.C. Zhao, Y. Feng, H. Pang, X.L. Bai, J.L. Tang, Q.S. Ma, J.X. Feng J. Appl. Microbiol. 107 (2009) 245–256.

[3] N. Ilmberger, D. Meske, J. Juergensen, M. Schulte. P. Barthen, U. Rabausch, A. Angelov, M. Mientus, W. Liebl, R.A. Schmitz, W.R. Streit Appl. Environ. Microbiol. 95 (2012) 135-146.

[4] M. Allgaier, A. Reddy, J.I. Park, N. Ivanova, P. D'haeseleer, S. Lowry, R. Sapra, T.C. Hazen, B.A. Simmons, J.S. Van der Gheynst, P. Hugenholtz, PLoS ONE 5 (2010) e8812.

[5] J.L. Arpigny, K.-E. Jaeger, Biochem. J. 343 (1999) 177-183.

[6] R. Gupta, N. Gupta, P. Rathi, Appl. Microbiol. Biotechnol. 64 (2004) 763-781.

[7] M. Levisson, J. van der Oost, S.W.M. Kengen, Extremophiles 13 (2009) 567-581.

[8] U.T. Bornsheuer, FEMS Microbiol. Rev. 26 (2002) 73-81.

[9] J. Zhou, M.A. Bruns, J.M. Tiedje, Appl. Environ. Microbiol. 62 (1996) 316-322.

[10] A.-L. Reysenbach, N.R. Pace, in F.T. Robb, A.R. Place (Eds.), *Archaea: A Laboratory Manual – Thermophiles*, Cold Spring Harbour Laboratory Press, New York, 1995, pp 101-107.

[11] M.C. Hansen, T. Tolker-Neilson, M. Givskow, S. Molin, FEMS Microbiol. Ecol. 26 (1998) 141-149.

[12] T.A. Hall, Nucleic acid Symposium Series 41 (1999) 95-98.

 [13] S.F. Altschul, T.L. Madden, A.A. Schaffer, J.H. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Nucleic Acids Res. 25 (1997) 3389-3402.

[14] J.D. Thompson, D.G. Higgins, T.J. Gibson, Nucleic Acids Res. 22 (1994) 4673-4680.

[15] S. Kumar, K. Tamura, M. Nei, Brief. Bioinform. 5 (2004) 150-163.

[16] N. Saito, M. Nei, Mol. Biol. Evol. 4 (1987) 406-425.

[17] J. Felsenstein, Evolution 39 (1985) 783–791.

[18] K. Rashamuse, V. Magomani, T. Ronneburg, D. Brady, Appl. Microbiol. Biotechnol. 83 (2009) 491-500.

[19] J. Donaghy, P.F. Kelly, A.M. McKay, Appl. Microbiol. Biotechnol. 50 (1998) 257–260.

[20] V. Solovyev, A. Salamov, R.W. Li (Ed.), Automatic Annotation of Microbial Genomes and Metagenomic Sequences, Nova Science Publishers, 2011, pp. 61-78.

[21] T.N. Petersen, S. Brunak, G. von Heijne, H. Nielsen, Nature Methods 8 (2011) 785-786.

[22] S.F. Altschul, J. Theor. Biol. 138 (1989) 297-309.

[23] A. Marchler-Bauer, J.B. Anderson, P.F. Cherukuri, C. De Weese-Scott, L.Y. Geer, M. Gwadz, S. He , D.I. Hurwitz, J.D. Jackson, Z. Ke, C.J. Lanczycki, C.A. Liebert, C. Liu, F. Lu, G.H. Marchler, M. Mullokandov, B.A. Shoemaker, V. Simonvan, J.S. Song, P.A. Thiessen, R.A. Yamashita, J.J. Yin, D. Zhang, S. Bryant, Nucleic Acids Res. 33 (2005): doi: 10.1093/nar/gki069

[24] K. Tamura, D. Peterson, N. Peterson, G. Stecher, M. Nei, S. Kumar, Mol. Biol. Evol. 28 (2011) 2731-2739.

[25] M.M. Bradford, Anal Biochem 72 (1976) 248-254.

[26] U.K. Laemmli, Nature 15 (1970) 680-685.

[27] V. Mastihuba, L. Kremnicky, M. Mastihubová, J.L. Willett, G.L. Côté, Anal. Biochem. 309 (2002) 96-101.

[29] S. Hegde, P. Srinivas, G. Muralikrisha, Anal. Biochem. 387 (2009) 128-129.

[30] P. Biely, M. Mastihubová, W.H. van Zyl, B.A. Prior, Anal. Biochem. 311(2002) 68-75.

[31] T. Koseki, K. Takahashi, S. Fushinobu, H. Iefuji, K. Iwano, K. Hashizume, H. Matsuzawa, Biochim. Biophys. Acta. 1722 (2005) 200-208.

[32] K. Poutanen, M. Sundberg, H. Korte, J. Puls, Appl. Microbiol. Biotechnol. 33 (1990) 506–510.

[33] E.M. Gabor, W.B.L. Alkema, D.B. Janssen, Environ. Microbiol. 6 (2004) 879-886.

[34] S.J. Joseph, P. Hugenholtz, P. Sangwan, C.A. Osborne, P.H. Janssen, Appl. Environ. Microbiol. 69 (2003) 7210-7215.

[35] M.C. Vargas-Garcia, F. Suárez-Estrella, M.J. Lopez, J. Morena, Int. Biodeterior. Biodegradation. 59 (2007) 322-328.

[36] C.M. Wang, C.L. Shyu, S.P. Ho, S.H. Chiou, Lett. Appl. Microbiol. 47 (2008) 46-53.

[37] Y.-J. Kim, G.-S. Choi, S.-B. Kim, G.-S. Yoon, Y.-S. Kim, Y.-W, Protein Expr. Purif. 45 (2006) 315-323.

[38] E.Y. Yu, M.-A. Kwon, M. Lee, J.Y. Oh, J.-E. Choi, J.Y. Lee, B.-K. Song, D.-H. Hahm, J.K. Song, Appl. Environ. Microbiol. 90 (2011) 573-581.

[39] J. Pleiss, M. Fischer, M. Peiker, C. Thiele, R.D. Schmid, J. Mol. Catal. B: Enzym 10 (2000) 491-508.

[40] J.A. Kelly, O. Dideberg, P. Charlier, J.P. Wery, M. Libert, P.C. Moews, J.R. Knox, C. Duez, C. Fraipont, B. Joris, J. Dusart, J.M. Frere, J.M. Ghuysen, Science 231 (1986) 1429–1431.

[41] E.I. Petersen, G. Valinger, B. Solkner, G. Stubenrauch, H. Schwab, J. Biotechnol. 89 (2001) 11–25.

[42] Y.H. Kim, E.J. Kwon, S.K. Kim, Y.S. Jeong, J. Kim, H.D. Yun, H. Kim, Biochem. Biophys. Res. Commun. 393 (2010) 45-49.

[43] B. Joris, J.M. Ghuysens, G. Dive, A. Renard, O. Dideberg, P. Charlier, J.M. Frere, J.A. Kelly, J.C. Boyington, P.C. Moews, J. Biochem. 250 (1988) 313 – 324.

[44] M. Estaban-Torres, I. Reverón, J.M. Mancheño, B. de las Rivas, R. Muñoz, Appl. Environ. Microbiol. 79 (2013) 5130-5136.

[45] D.W.S. Wong, V.J. Chan, H. Liao, M.J. Zidwick, J. Ind. Microbiol. Biotechnol. 40 (2013) 287-295.

[46] K. Rashamuse, T. Ronneburg, W. Sanyika, K. Mathiba, E. Mmutlane, D. Brady Appl. Microbiol. Biotechnol. 98 (2014) 727-737.

[47] S. Hegde, G. Muralikrishna, World J. Microb. Biot. 25 (2009) 1963-1969.

[48] C. Elend, C. Schmeisser, C. Leggewie, P. Babiak, J.D. Carballeira, H.L. Steele, J.-L. Reymond, K.-E. Jaeger, W. Streit, Appl. Environ. Microbiol. 72 (2006) 3637-3645.

[49] J.H. Jeon, S.-J. Kim, Y.S. Lee, S.-S. Cha, J.H. Lee, S.-H. Yoon, B.-S. Koo, C.-M. Lee, S.H. Choi, S.H. Lee, S.G. Kang, J.-H. Lee, Appl. Environ. Microbiol. 77 (2011) 7830-7836.

[50] N. Mokoena, K. Mathiba, T. Tsekoa, P. Steenkamp, K. Rashamuse, Biochem. Biophys. Res. Comm. 437 (2013) 342-348.

[51] U.G. Wagner, E.I. Petersen, H. Schwab, C. Kratky, Protein Sci. 11 (2002) 467-478.

[52] V.F. Crepin, C.B. Faulds, I.F. Connerton, Appl. Microbiol. Biotechnol. 63 (2004) 647-652.

[53] D.B. Udatha, I. Kouskoumvekaki, L. Olsson, G. Panagiotou, Biotechnol. Adv. 29 (2011) 94-110.

[54] M.A. Kabel , C.J. Yeoman , Y. Han , D. Dodd , C.A. Abbas , J.A. de Bont , M. Morrison, I.K. Cann , R.I. Mackie, Appl. Environ. Microbiol. 77 (2011) 5671-5681.

[55] Blum

[56] N.D. Mohne, M. Bar-Peled, C. Somerville. In: M.E. Himmerl (Ed.) Biomass recalcitrance, Blackwell, Oxford, 2008, pp 266 – 277.

[57] R. Lamed, E.A. Bayer, Adv. Appl. Microbiol. 33 (1988) 1-46.

[58] L.L. Lin, J.A. Thomson, FEMS Microbiol. Lett. 84 (1991) 197–204.

[59] P.M.-A. Pawar, S. Koutaniemi, M. Tenkanen, E.J. Mellerowicz, Front. Plant. Sci. 4 (2013) 1-8.

[60] P. Biely, C.R. Mackenzie, J. Puls, H. Schneider, Nature Biotechnology 4 (1986) 731–733.

Appendix A. Supplementary data
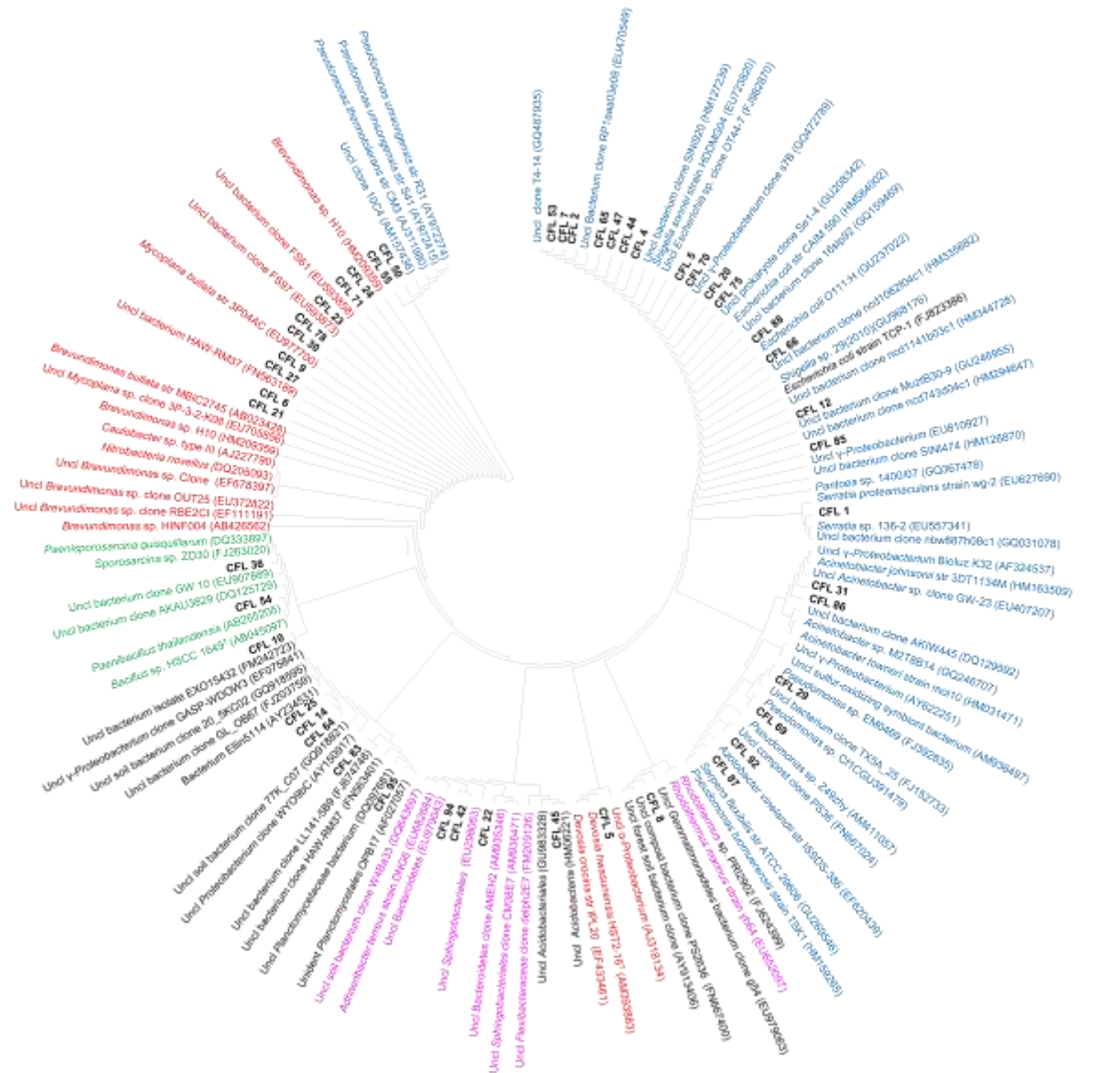
Supplementary Fig S1 Unrooted 16S rRNA gene phylogenetic tree of the compost fosmid library (CFL) clones generated with general eubacterial primers. The tree was based on 615 bp of common sequence and was obtained using the neighbour-joining method. Strains that produce lignocellulolytic enzymes are indicated with by a black triangle ( ). GenBank accession numbers are given in parentheses.
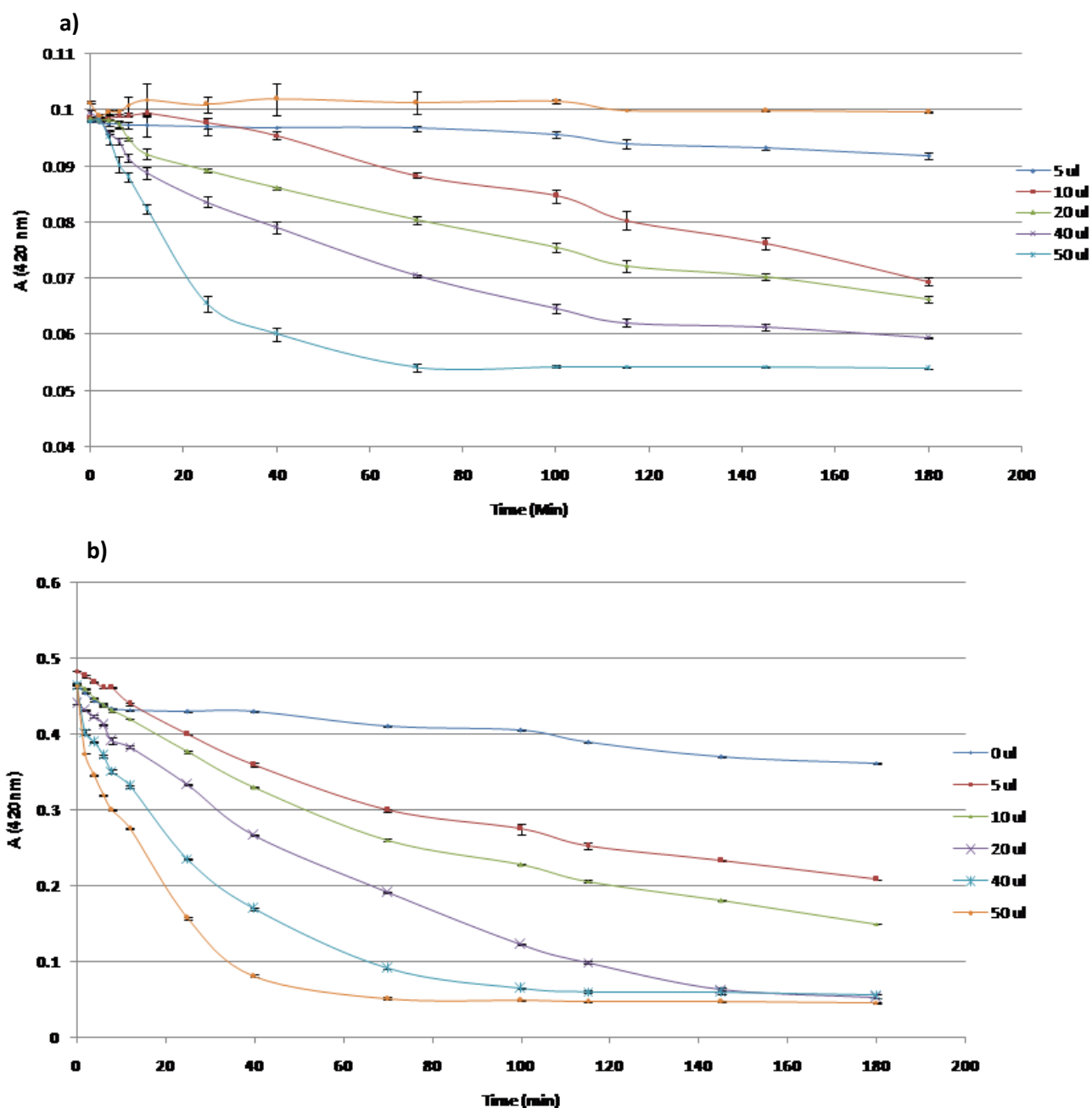
▲
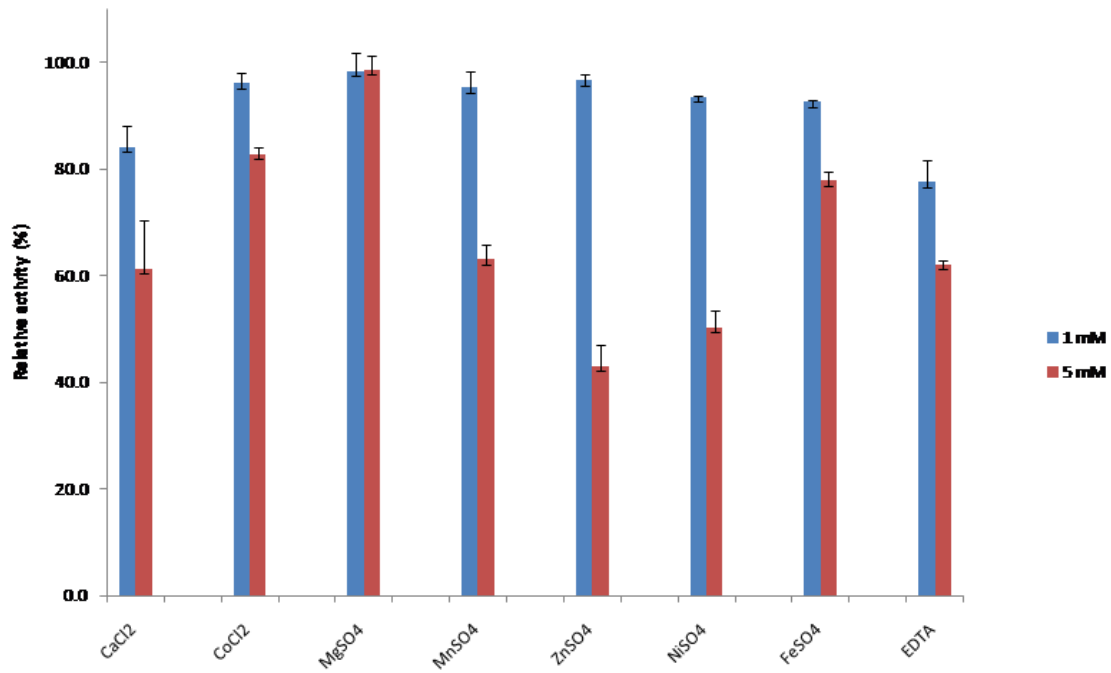
Supplementary Fig S2 Effect of divalent cations and EDTA on the activity of G34. Relative activities are provided as a percentage of the activity in the absence of cation.
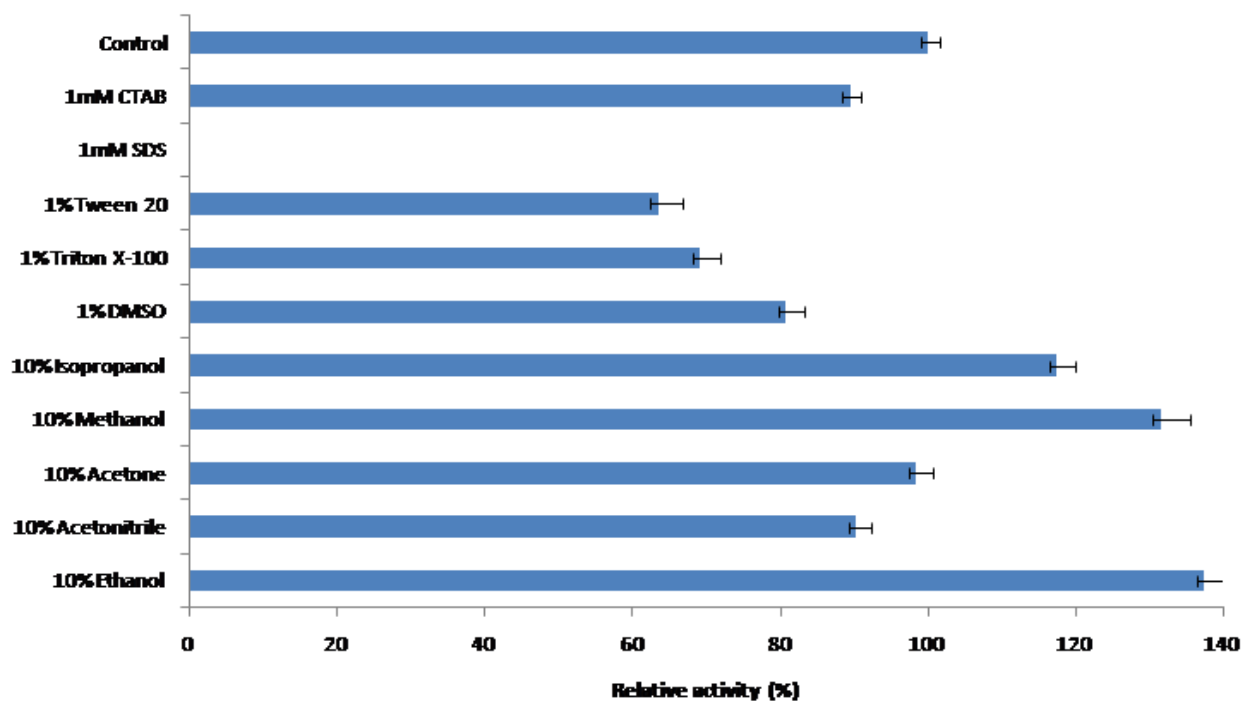
Supplementary Fig S3 Effect of various denaturants, inhibitors and organic solvents on the activity of G34.

Supplementary Fig S4 Activity of G34 at 25°C towards a) 4-nitrophenyl 5-O-trans-feruloyl-α-L-arabinofuranoside (NPh-5-Fe-Araf) and b) 4-nitrophenyl 2-O-trans-feruloyl-α-L-arabinofuranoside (NPh-2-Fe-Araf) in 100 mM sodium phosphate buffer (pH 6.5) Measurements of absorbance at 420 nm were taken over a period of 3 hours and 1 unit of activity was defined as an absorbance change of 1.0 min^{-1} (A_{420}).