Phylogeny, classification and metagenomic bioprospecting of microbial acetyl xylan esterases

Fiyinfoluwa A. Adesioye\textsuperscript{a}, Thulani P. Makhalanyane\textsuperscript{a}, Peter Biely\textsuperscript{b}, Don A. Cowan\textsuperscript{a}.

\textsuperscript{a}Centre for Microbial Ecology and Genomics, Genomics Research Institute, Natural Sciences 2 Building, University of Pretoria, Hatfield 0028, Pretoria, South Africa.

\textsuperscript{b}Institute of Chemistry, Slovak Academy of Sciences, Dúbravská cesta 9, 845 38, Bratislava, Slovak Republic.

Corresponding author: Don A. Cowan\textsuperscript{a}, don.cowan@up.ac.za

**Highlights**

- AcXEs occur in several CAZy families, and show promiscuous substrate specificities.
- A strong AcXE sequence to specificity link would be a valuable functional predictor.
- There is a clear need for novel AcXE enzymes with detailed substrate specificity data.
- Functional metagenomics would be the most effective means for identifying new AcXEs.
- A lack of suitable substrates limits high throughput metagenomic biomining of AcXEs.
ABSTRACT

Acetyl xylan esterases (AcXEs), also termed xylan deacetylasases, are broad specificity Carbohydrate-Active Enzymes (CAZymes) that hydrolyse ester bonds to liberate acetic acid from acetylated hemicellulose (typically polymeric xylan and xylooligosaccharides). They belong to eight families within the Carbohydrate Esterase (CE) class of the CAZy database. AcXE classification is largely based on sequence-dependent phylogenetic relationships, supported in some instances with substrate specificity data. However, some sequence-based predictions of AcXE-encoding gene identity have proved to be functionally incorrect. Such ambiguities can lead to mis-assignment of genes and enzymes during sequence data-mining, reinforcing the necessity for the experimental confirmation of the functional properties of putative AcXE-encoding gene products.

Although one-third of all characterized CEs within CAZy families 1-7 and 16 are AcXEs, there is a need to expand the sequence database in order to strengthen the link between AcXE gene sequence and specificity. Currently, most AcXEs are derived from a limited range of (mostly microbial) sources and have been identified via culture-based bioprospecting methods, restricting current knowledge of AcXEs to data from relatively few microbial species. More recently, the successful identification of AcXEs via genome and metagenome mining has emphasised the huge potential of culture-independent bioprospecting strategies. We note, however, that the functional metagenomics approach is still hampered by screening bottlenecks.

The most relevant recent reviews of AcXEs have focused primarily on the biochemical and functional properties of these enzymes. In this review, we focus on AcXE phylogeny, classification and the future of metagenomic bioprospecting for novel AcXEs.

Keywords: Acetyl xylan esterase; carbohydrate active enzymes; carbohydrate esterases; metagenomics
1.0 Introduction

Acetyl xylan esterases (AcXEs) (EC 3.1.1.72) are a class of hydrolytic enzymes primarily involved in the hydrolysis of acetyl xylan esters. They were first described by Biely, Puls [1, 2] and have since attracted steadily increasing attention from researchers (Figure 1A), with a number of recent reviews focusing principally on the chemistry and specificity of AcXE catalysis [3-6]. A number of AcXE-producing microorganisms have been isolated from various environments (further discussed below) and AcXE activity was recently reported for esterases from some plants; e.g., *Populus*-secreted pectin acetyl esterase [7] and *Arabidopsis thaliana* [8]. However, the classification of AcXEs remains unclear, with the CAZy (Carbohydrate Active Enzyme) database classification placing AcXEs within the very broad category of carbohydrate esterases (CEs). Such assignments introduce problems during sequence-based screening and genome annotation and it is therefore important that the linkage between sequence-based phylogenies and functional characteristics of AcXEs (particularly substrate specificity) are properly understood.

The importance of AcXEs in lignocellulose degradation [9] supports the need for continuous bioprospecting for novel AcXEs with improved functional properties [10] using both culture-dependent and genomic bioprospecting approaches. Although, several AcXEs have been identified, the discovery and characterization of new AcXEs of microbial origin with improved catalytic activities via functional metagenomics will result in the availability of a huge reservoir of xylan deacetylases for maximising xylan hydrolysis given different processing conditions [11]. A limited number of xylan deacetylating gene products, identified principally by functional screening of metagenomic libraries, have been reported [12]. While the convention is that novel genes are best identified by activity-based screening of metagenomic libraries [13], identification of putative AcXE genes by annotation of metagenomic sequence datasets
Figure 1A: Scopus estimate of AcXE–related publications since 1985; B. Search hits from various databases using ‘acetyl xylan esterases’, ‘acetylxylan esterases’ and ‘xylan deacetylases’ as key words.

* Some database search results include non-xylan deacetylases
is technically feasible, albeit confounded by the miss-annotation of some non-specific esterases as AcXEs. A careful analysis of conserved AcXE phylogenetic data should therefore assist in the discovery of novel, but distantly-related AcXEs.

We note, from a review of literature, that AcXEs are variously termed acetyl xylan esterases, acetylxylan esterases or xylan deacetylases. While all three terms are broadly appropriate, the use of multiple name variants impacts significantly on the search outputs from sequence databases, purely because each is perceived as a different input by search engines and therefore yields different search results (Figure 1B). Although the term *acetylxylan esterase* is the more correct chemical nomenclature, the term *acetyl xylan esterase* is in more common usage and is consistent with the widely accepted acronym AcXE, and we therefore argue strongly for the general acceptance and consistent usage of this term.

### 2.0 Acetylated Xylan

Many plants have evolved mechanisms for protection of their tissues from physical and biochemical attack (e.g., from insects, microorganisms and enzymes) [14, 15]. Such mechanisms include acetylation and ‘feruloyation’ of plant cell polymers (xylan, mannan, pectin, peptidoglycan and chitin), thereby reducing the efficiency of polysaccharide hydrolases [4, 5]. Acetylation occurs mainly in hardwood (acetyl glucuronoxylan), typically at the C2 or C3 or both positions, depending on whether it is a mono- or di-O-acetylated xylopyranosyl unit (Figure 3). On non-reducing xylopyranosyl residues of oligosaccharides or xylopyranosides, acetyl groups can migrate to position C4 [5, 16]. The migration is accelerated by increased pH and temperature [17]. The migration between position 2 and 3 on internal xylopyranosyl residues in the polysaccharide is also anticipated, however it has not been proven and
**Figure 3:** Schematic representation of xylan deacetylation by acetyl xylan esterases (ACD/Chemsketch v.11.0 -Freeware).
demonstrated experimentally. Other substitutions, such as C5 acetylation of the L-arabinofuranosyl side chain residues of xyloglucan, have also been reported [18].

It has been established [3, 5, 19] that acetylation of xylans generally increases the recalcitrance of plant polymers to depolymerisation by interfering with access of endoxylanases to glycosidic linkages on the xylopyranose main chain and resulting in incomplete hydrolysis of polymeric xylan (hemicellulose). This is a significant factor in reducing the efficiency of saccharification of lignocellulosic biomass for biofuel production [3, 5, 19].

3.0 The Carbohydrate-Active enZyme (CAZy) database

3.1 Carbohydrate Esterases (CEs)

Carbohydrate esterases are a large group of carbohydrate-active enzymes that catalyse the removal of ester substituents from the glycan chains of polysaccharides. The catalysis of the de-O- or de-N-acylation of sugar substituents may be via a Ser-His-Asp catalytic triad, a Ser-His diad or a Zn$^{2+}$ catalytic pathway (or other mechanism) depending on the CE family [20]. With exception of CE15, which act on substrates in which the sugar acts as the acid, all other CE families hydrolyze esters in which sugars act as alcohols. Currently, there are 16 recognized CE families and a CE17 family has been proposed [4, 21]. CEs act on soluble substrates and de-esterify saccharide residues esterified with short carboxylic acids such as acetic, propionic, butyric and phenolic acids (ferulic or p-coumaric acid) including peptidoglycan (CE4, 9, 11), glucuronic acid (CE14 and 15), chitin (CE4, 14), rhizobial Nod factors (CE4), xylan (CE1-7 and 16), pectin (CE8, 12, 13) and rhamnogalacturonan (CE12) [20, 22]. These activities may facilitate the action of other CAZymes associated with plant cell wall degradation. Interestingly, CEs (CE2 and CE16) have also been implicated in transesterification reactions [23, 24].
3.2 **Acetyl xylan esterases (AcXEs)**

Acetyl xylan esterases (EC 3.1.1.72) belong to CE families 1 to 7 and 16 [4, 20]. Family CE12 includes two enzymes designated as AcXEs, but their ability to deacetylate xylan has not been confirmed. The classification into these separate enzyme families is primarily based on similarities in sequence, structure and activity on xylan and per-o-acetylated saccharides or methylated saccharides [25, 26]. While some acetyl esterases (AcEs) have been wrongly designated as AcXEs [27] due to the similarity of their sequences and inadequate information on their substrate specificities, clear distinctions between these enzyme classes have been made with the creation of the AcXE sub-class (EC 3.1.1.72) in 1999. In some cases, esterases exhibiting xylan deacetylation activity as a side reaction only were assigned as acetyl xylan esterases, such as the CE2 esterases (see below). General esterase activity assays, using α-naphtyl acetate [28] and 4-nitrophenyl acetate as substrates, have frequently been used to determine AcXE activity [19], but cannot distinguish AcEs from AcXEs. AcEs differ from AcXEs in that AcEs are generally not capable of deacetylating polymeric or oligomeric xylan. Although some AcEs may act non-specifically on short acetylated xylooligosaccharides [5], here we define AcXEs as enzymes that may deacetylate polymeric xylan; e.g., AcXE6A from *Fibrobacter succinogenes* S85 [29], and/or xylanase-cleaved XOS; e.g., AcXE from *Trichoderma reesei* [30], regardless of their activity on other substrates. The term ‘AcE’ is also generally used to describe enzymes which deacetylate substrates other than carbohydrates [31, 32]. Simply stated, all AcXEs are AcEs in that they are deacetylating esterases, but not all AcEs are AcXEs.

The closest to the natural substrate of AcXEs is partially depolymerized O-acetyl-4-O-methyl-D-glucurono-D-xylan (Figure 3), released from hardwood by steam explosion [33] or extracted from delignified hardwood pulp by DMSO [34]. Chemically acetylated alkali-extracted
glucuronoxylan is an alternative substrate used in functional assays to confirm AcXE activity [35]. While the substrate specificity of AcXEs has been associated with whether xylan was naturally or chemically acetylated [36], it remains unclear whether the acetyl groups in chemically acetylated xylan are located on xylopyranosyl residues only. Some AcXEs are produced as bimodular or multimodular enzymes [37, 38]. Frequently their catalytic domain is linked to a different catalytic module or to a carbohydrate binding module (CBM), emphasizing their role in plant cell wall degradation. The CBM may not be a xylan binding module but a cellulose binding module [39]. Certain AcXEs were reported to bind to other plant cell wall polymers such as cellulose and arabinoxylan while independently deacetylating xylan, thus increasing accessibility of other lignocellulases to decaying plant material [29, 40]. Other substrates on which AcXEs have been reported to be active include acetylated monosaccharides [23, 41], triacetin [30], α-napthyl acetate [42], tri-O-acetyl-D-galactal [43], p-nitrophenyl acetate [1], short chain alkyl acetates and 4-methylumbelliferyl acetate [44]. Certain AcXEs have also been shown to have a higher rate of deacetylation when in synergistic action with xylanases [43, 45].

3.3 AcXE phylogeny

The phylogenetic relationships between AcXEs are important in the classification of these enzymes into appropriate CE families. A maximum-likelihood phylogenetic tree (Figure 2), using only characterized AcXE sequences from the CAZy database, shows a clear correlation between sequence homology and predicted activity. This topology also provides strong evidence for inter- and intra-specific horizontal AcXE gene transfers, both within and between bacteria and fungi [46-48]. Non-AcXEs and AcXEs within the same family are shown to originate from the same ancestor (e.g., members of the CE4 family).
Figure 2: Unrooted maximum likelihood (1000 bootstrap) phylogenetic tree of characterised AcXEs within CE 1-7 and CE16 families, including some non-AcXEs, constructed with their protein sequence alignments (MAFFT align) using the protein phylogeny methods comparison tool of Mobyle 1.5.4 version of the Pasteur Institute programme [170].
Phylogenetic classification is also relevant for higher order classifications; e.g., of CAZyme classes into sub-families [48]. Consequently, certain CAZymes have been grouped into sub-families to serve as a platform through which CAZyme substrate specificities can be directly related to their structures [26], but no CE families have been sub-classified. So far, only the sub-classification of the CE16 family has been strongly supported [21, 49].

In the sequence homology-based annotation of AcXEs, it has been demonstrated [50, 51] that the use of conserved catalytic residues and block patterns in multiple sequence alignments for constructing phylogenetic trees is a more reliable guide to predicting enzyme specificity [52]. However, while phylogenetic relationships provide clear guidance to assigning AcXE sequences to most CE families, certain AcXEs belonging to families 2, 3 and 16 (all SGNH hydrolases) show substantial substrate promiscuity. Figure 2 shows up to 80% branch support for a cluster of CE2 and CE3 AcXEs together with non-AcXEs from the CE8 family. These observations highlight the risk to either functional annotation or prediction of enzyme substrate specificity based on phylogenetic relationships alone; and the importance of functional characterization data [3, 53]. Hence, the general consensus that there is still inadequate information on substrate specificity in relation to CAZyme sequence information for automated prediction of CAZyme substrates [26]. It is therefore clear that more phylogenetic data derived from characterized functional AcXEs are required in order to establish reliable predictive relationships between gene sequences and substrate specificities.

3.4 AcXE Classification – CAZy

AcXE phylogeny largely determines the classification of AcXEs. While biochemical properties and substrate specificities of AcXEs have been described in detail in previous reviews [3, 4], we have summarised the most important characteristics of these enzymes, including source organism, substrate preference, conserved sequences (Table 3) and the catalytic residues that
Table 3: Whole conserved sequence regions derived from multiple sequence alignments consisting of at least four characterised functional AcXE within each CE family and their encoded function depicted by listed substrate preference.

<table>
<thead>
<tr>
<th>CE family</th>
<th>AcXE source organism (GenBank accession number)</th>
<th>Substrate preference</th>
<th>Conserved sequences</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><em>Aspergillus oryzae</em> RIB40 (XP_001826329.1)</td>
<td>AX</td>
<td>LEQVTDFGDNPSNVKMIYVP; VAIIHNCYTQGA; GSPOYQQLA; GFFIVYFESP; CWDVSS-LTHNGGNSNFLANNMV; VFVTGTSQAMTAVAATYPNLFAAG; YAGVPAQCF; WNSTCAQGQ; MYPDYSGSRPKMQIYMHVNGSVTTLVPOQNYEE; KQWAGVFGY; PNLQGILAGGGVQNNQIY; DMKWFY</td>
<td>[58]</td>
</tr>
<tr>
<td>2</td>
<td><em>Neocallimastix patriciarum</em> (AAB69091.1)</td>
<td>α-Naphthyl acetate, AX</td>
<td>KIEFGDISITCAYG; ASKQLNA; SGFGI; PDLVVINLNGTND; GKKIDWHP; AELEVAEI</td>
<td>[45, 63]</td>
</tr>
<tr>
<td>3</td>
<td><em>Ruminococcus flavefaciens</em> 17 (CAB55348.1)</td>
<td>β-Naphthyl acetate, AX</td>
<td>IKIMPLGDSIT; DEGGRKYTL; VDLVPEG; YDDNHAGYSYTL; SFDILLQIGTNDVSNGH; YNELIKKVA; NVIYADIH; DGVHPNAGYKEMYG</td>
<td>[37]</td>
</tr>
<tr>
<td>4</td>
<td><em>Streptomyces lividans</em> 1326 (AAC06115.2)</td>
<td>AX</td>
<td>GYVGTLTFDDGP; ALRQRNLATMFNQQG; RAQVDAG; VANHSTYPMTIQ; SRTQQA1; GLCSGMISPQGTRAVAPDQSGGGGDGCG; FRPYGETNATLRSVEAKYGLTEVMDVSQDWNASTDAEVQASVRLNGQ; VILMHDWPANTLAIPIRQITLA</td>
<td>[80]</td>
</tr>
<tr>
<td>5</td>
<td><em>Chrysosporium lucknowense</em> (ADZ98863.1)</td>
<td>AX</td>
<td>SCPEVHVGARETTAPPYGTSGQLMNVMVQAYPGATSEAINYPACGQPEQSC; GGIYNTSANQGQTAVSVSATSVFQNPCDTELIGVSQGGQIMDNAYCGGA; SAILNAVATWFGNP; YRGVCQAGGFAARPFG; IKSYCAEDPYCCNQNDNAN; YGGQQLAFIKSKL</td>
<td>[62]</td>
</tr>
<tr>
<td>6</td>
<td><em>Orpinomyces</em> sp. PC-2 (AAC14690.1)</td>
<td>AX</td>
<td>PDPNFHIYLAQSNMEQQQG; GEWYGAPLPP; LGPDYFGRTL; AKKAQKAGVILKILHGETN; LNLKAEVPLAGEVV; LPVPIAHTVIDAE; DDLHF; YRILGERYA</td>
<td>[43]</td>
</tr>
<tr>
<td>7</td>
<td><em>Bacillus pumilus</em> (CAB76451.2)</td>
<td>short chain AcXOS</td>
<td>FDLSLEELKKY; DFSDFW; MLVRGQGG; DTTYYRGVYLDVAVRA; RIGVIGSQGGALIAAAAL; PPSSTVFAAYN; YFGHE</td>
<td>[88, 101]</td>
</tr>
<tr>
<td>8</td>
<td><em>Trichoderma reesei</em> (ABI34466.1)</td>
<td>Non-reducing end AcXOS</td>
<td>KLYTIFGDSY; TASGGLQW; WIGTND; AGGRFVIL; SYLWYDELHP</td>
<td>[4, 92]</td>
</tr>
</tbody>
</table>

*Pairwise alignment with *Myceliophthora thermophile* [94]; AX - acetylated xylan, AcXOS - Acetylated xylooligosaccharides; Bold letters are amino acids that are unique to the representative AcXE listed.
are unique to AcXEs within each family. These data clearly show the independence of AcXE classification on any individual characteristic, but strongly confirm the linkage between sequence homology and functional relationships with other CEs.

**CE1:** The CE1 family consists of over 3700 CEs of archaeal, bacterial and eukaryotic (fungal and protozoan) origin including AcXEs, feruloyl esterases, carboxylesterases, cinnamoyl esterases, S-formylglutathione hydrolases, trehalose 6-O-mycollyltransferases, PHB depolymerases and diacylglycerol O-acyltransferases [20]. All members of this family possess an α-β-α sandwich three-dimensional structure and are grouped as serine-type esterases possessing the GXSXG conserved sequence with the Ser-His-Asp catalytic triad as well as lengthy, highly conserved sequence blocks. They are active on acetyl xylan as well as acetylated glucosides, galactosides, galactoglucomannan and cellulose [4, 54]. AcXEs belonging to this family of enzymes have been isolated largely from fungi such as *Schizophyllum commune* [44, 55], *Penicillium purpurogenum* (also *Talaromyces purpureogenus*) [56, 57], *Aspergillus ficuum* [58], *Aspergillus awamori* [59], *A. niger* [60], *Volvariella volvacea* [61] and *Chrysosporium lucknowense* (*Myceliophthora thermophile*) [62]. However, while CE1 AcXEs share sufficient homology with certain bacterial CEs to warrant classification into the same family, they do not possess strong phylogenetic relationships with any known AcXEs of bacterial origin. Evidently, classification of AcXEs into this family is not dependent on their possession of sequence homology unique to fungi or fungal AcXEs (which exist in other CE families).

**CE2:** This family consists of over 175 acetyl esterases (ACEs) and AcXEs of bacterial and fungal origins, only 6 of which have been characterized in detail [20]. There are currently no reported CE2 AcXE sequences from archaeal or protozoan sources. CE2 AcXEs, which were first described by Dalrymple, Cybinski [63] in the fungus *Neocallimastix patriciarum*, are
active on acetylated birchwood xylan and α-naphthyl acetate. The CE2 AcXEs from *Cellvibrio japonicus* and *Clostridium thermocellum* have been best characterised and are reported to be generally active on aryl esters and selectively active on acetylated xylan, showing higher activities on acetylated konjac glucomannan than on acetylated birchwood xylan and di-acetylated xylopyranosides [23, 64]. The CE2 AcXEs are serine-type esterases with a Ser-His catalytic diad and an α-β hydrolase protein fold [64]. They possess the GDS(L) conserved motif and exhibit deacetylase specificity for the C6 position of acetylated gluco- and manno-pyranosyl residues of hemicellulose substrates, and for the C4 and C3 positions on mono-acetylated xylopyranosyl residues [23]. 6-O-deacetylase activity has not been reported for any member of other CE families. Hence, it is unique to the CE2 family. The 4- and 6-O-deacetylation specificities of this family are unlike the regioselectivity for position 2 observed in AcXEs belonging to other CE families. AcXEs from this family catalyse transesterification reactions from vinyl acetate to the O-6 position of hexopyranosides and hexooligosaccharides [4, 23].

**CE3:** Members of the CE3 family belong to the SGNH hydrolase superfamily [21] and are included in the serine-type AcEs which possess the Ser-His-Asp (SHD) catalytic triad and α-β-α sandwich protein structure [39]. The CE3 family has over 190 members of archaeal, bacterial and eukaryotic (fungal) origin. This family also consists of the GDSL lipases and esterases which possess the GDSL motif rather than the GXSXG motif synonymous with lipases and esterases [65]. They are often multi-functional, including AcXE activity. However, studies on some of these enzymes to confirm their true substrate specificities are still required [4]. In the biochemical characterization of the catalytic module (CtCes3-1) of CtCes3 (*Clostridium thermocellum*) CE3 AcXE, it was observed that substrate specificity was “restricted to xylans”, typical of true AcXEs rather than AcEs [39]. Conversely, in a recent
study [66] only very low or no activity of CE3 AcXE (CtCE3 - ABN52033) on 2-O, 3-O, and 2,3-di-O-acetylated xylopyanosyl and 3-O-acetylated xylopyanosyl 2-O-substituted with methylglucuronic acid from *Eucalyptus* was detectable. Other characterized CE3 AcXEs have been obtained from animal rumen bacteria; such as *Ruminococcus flavefaciens* 17 [37], and the fungus *Neocallimastix patriciarum* [45, 63].

**CE4:** The CE4 family is the largest CE category with over 7600 sequenced enzyme members [20]. All enzymes in this CE family are referred to as ‘aspartate metalloenzymes’ [67]. CE4 family enzymes are characterised by their dependence, for catalytic activity (deacetylation), on coordination of either a Co$^{2+}$ or Zn$^{2+}$ cation [68], and by the use of the His-His-Asp metal binding triad[38]. CE4 AcXEs operate an acid (aspartic acid) - base (histidine) catalytic mechanism and are also known as ‘NodB homologs’ [69-71]. They possess the highly conserved catalytic NodB domain [72] characteristic of other rhizobial NodB enzymes in the CE4 family such as the chitin deacetylases, chitooligosaccharide deacetylases, rhizobial nod factor deacetylases as well as peptidoglycan N-acetylglucosamine deacetylases and N-acetylmuramic deacetylases which de-esterify N- or O-acetyl bonds of plant cell wall polymers or oligomers [73-75]. Family CE4 AcXEs possess an eight-stranded ($\beta$-$\alpha)_8$ barrel protein fold which has an irregular structure [38]. CE4 AcXEs prefer longer chain oligosaccharide substrates and exhibit positional deacetylation specificity [67, 73]. They are active on various acetylated xylan residues and are reported to show preference for methyl per-O-acetyl-\(\beta\)-D-xylopyranoside [25], but unlike most AcXEs in other families, do not show activity on para-nitrophenyl acetate (pNPA) or 4-methylumbelliferyl acetate [67, 73]. On the contrary, an enzyme recently reported as a CE4 AcXE from *Anoxybacillus flavithermus* showed activity on pNPA, but its activity on acetylated xylan was not tested to verify its identity as a ‘true’ AcXE [76]. On the polymeric substrate, CE4 AcXEs efficiently deacetylate only singly acetylated
xylopyranosyl residues. Doubly acetylated xylopyranosides do not serve as substrates because these esterases require a free vicinal OH-group for deacetylation of position 2 or position 3 [77-79]. Characterized CE4 AcXEs include enzymes from *Streptomyces lividans* [80] and *Clostridium thermocellum* [38].

**CE5:** The CE5 family consists of over 900 protein sequences of AcXEs and cutinases. Both are serine-type esterases with α-β-α sandwich structures and the Ser-His-Asp catalytic triad. CE5 AcXE sequences possess the GXSXG conserved motif as well as highly conserved blocks found within the cutinase domain [81] and belong to the cutinase superfamily [82]. These enzymes preferentially deacetylate xylopyranosyl residues of acetylated glucuronoxylan at the C2 position [4, 21] as well as at the C3 position of acetylated methyl xylopyranosides [67]. The first AcXE assigned to the CE5 family [81] was isolated from *Trichoderma reesei* and showed similarly as CE4 esterases. some requirement for a free unsubstituted hydroxyl group for deacetylation of adjacent positions [16]. Other characterized CE5 AcXEs have been isolated from *Chrysosporium lucknowense Cl* [62], *Penicillium aurantiogriseum* [83] and *P. purpurogenum* MYA 38 [84]. There are currently no records of characterised bacterial AcXEs in this CE family.

**CE6:** Over 160 enzymes designated as AcXEs belong to the CE6 family [20]. They are serine esterases with conserved regions similar to the GDS(L) family of esterases and belong to the SGNH hydrolase superfamily [85]. They possess an α-β-α sandwich formation and the Ser-His-Asp catalytic triad typical of esterases as well as a unique, conserved HQGE catalytic site motif positioned between the N-terminus and center of their sequences [86]. CE6 esterases possess a broad substrate specificity [4]. Unlike CE4 AcXEs, they do not require a free vicinal hydroxyl group for deacetylation activity. Characterized AcXEs belonging to this family
include Axe6A and Axe6B from the bacterium *Fibrobacter succinogenes* [29, 87], BnaA from *Neocallimastix patriciarum* [45, 63], AxeA from *Orpinomyces* sp. [43] and AxeA from a bovine metagenome [12], all of which are of rumen origin. These AcXEs have very low sequence homology to other CE6 enzymes and it is not certain if this is as a result of functionally-convergent or sequence-divergent evolution [29]. AxeA from *Orpinomyces* [43], which is the only commercially available CE6 esterase, deacetylates positions 2 and 3 on both mono- and di-O-acetylated xylopyranosyl residues of polymeric xylan [78].

**CE7:** The CE7 family is made up of over 450 AcXEs and cephalosporin C deacetylases, all of bacterial origin. All possess the α/β/α sandwich conformation common to other serine esterases, the Ser-His-Asp catalytic triad and the GXSXG conserved motif common to esterases, lipases and thioesterases [88]. They are intracellular enzymes which exhibit unique substrate and positional specificities, making the classification of some members of this family as AcXEs uncertain (Levisson *et al.* 2012). Most AcXEs in this family deacetylate 7-aminocephalosporanic acid and cephalosporin C but possess high sequence homology with non-AcXE cephalosporin deacetylases. Prior to the establishment of the CAZy database and the CE7 family, these enzymes were all classified as cephalosporin deacetylases [4]. AcXE from *Bacillus pumilus* [88] has been well characterized and is often considered as a model organism for this family. Some CE7 AcXEs possess a substrate preference for monoacetylated short chain XOS rather than polymeric xylan [88, 89], and act on acetates of simple sugars and various aryl alcohols [88, 90]. Other characterized CE7 AcXEs are derived from *Bacillus subtilis* [42] and the thermophiles *Thermotoga maritima* [90] and *Thermoanaerobacterium saccharolyticum* [89, 91]. It is also worth noting that all characterized AcXEs within this family are moderately (T_{opt} - 50°C) to highly (T_{opt} - 90°C) thermostable.
CE16: The CE16 family is the most recently established [92]. It consists of approximately 100 AcEs, only some of which exhibit AcXE activity. CE16 enzymes possess the GDS(L) catalytic motif and the Serine, Glycine, Asparagine and Histidine (SGNH) catalytic residues. They are mainly of bacterial and fungal origin and the most well-characterised member, a fungal AcE (Aes1) from *Hypocrea jecorina*, an anamorph of *Trichoderma reesei* [92, 93] is regarded as the model enzyme for this family [4]. Aes1 does not deacetylate polymeric xylan, but shows a preference for deacetylating positions 3 and 4 on non-reducing end xylopyranosyl residues in XOS. It behaves as an exo-acting deacetylase. It also deacetylates monoacetylated 4-nitrophenyl β-D-xylopyranoside residues and naphtyl acetate [92, 93] and can transacetylate saccharides in water or aqueous organic solvents [24, 67]. AcEs showing AcXE activity in this family have been referred to generally as AcXEs or specifically as ‘AcE-type AcXEs’ (Acetyl esterase-AcXEs) because of their inability to act on polymeric xylan [94]. However, this situation has recently changed as reported from two independent studies of CE16 members, one from *Podospora anserina* (PaCE16) [49] and one from *Aspergillus niger* (AnCE16) [66]. These enzymes showed some catalytic properties compatible with exo-deacetylase activity of Aes1 from *T. reesei*, but also exhibited some activity on polymeric xylan. Similarly, as typical AcXEs they did not attack the 3-O-acetyl group on methylglucuronic acid substituted xylopyranosyl residues. Other ‘AE-type AcXEs’ in this family include AcXEs from *Myceliophthora thermophila* and *Trichoderma harzanium* [94]. Analysis of the xylan-deacetylation activities of these enzymes on acetylated XOS and methylglucuronic acid substituted acetylated XOS showed low individual deacetylation efficiencies but synergistic catalysis when used with the CE1 AcXE from *M. thermophila* and the CE5 AcXE from *T. reesei* [94]. Little is known about the mechanistic properties of AcXEs in this family and no CE16 enzyme structural data is currently available.
3.5 Positional specificity of AcXEs

In addition to substrate specificity, AcXEs exhibit positional specificity (regioselectivity) where a specific carbon position is preferably deacetylated over others [95]. The positional specificities of AcXEs and other CEs have been investigated using techniques capable of monitoring the release of unique acetate groups from acetyl xylan: there include matrix assisted laser desorption/ionisation time of flight (MALDI-TOF) mass spectrometry [62], enzyme coupled assays [96], proton-NMR (¹H NMR) [66, 67, 78] and capillary electrophoresis with laser-induced fluorescence (CE-LIF) [62].

A recent evaluation of the positional specificity of eight fungal AcXEs from seven CE families (CE1-6 and CE16) on mono- and di-O acetylated XOS and on 3-O acetylated XOS 2-O substituted with methylglucuronic acid [66], based on the rate of deacetylation at particular positions, and other recent studies [67, 78] show that CE4 AcXEs deacetylate only singly acetylated xylopyranosyl residues, while CE1, CE5 and CE6 AcXEs are capable of also deacetylating doubly acetylated xylopyranosyl residues. None of the typical AcXEs recognized the 3-O-acetyl groups on methylglucuronic acid-substituted xylopyranosyl residues as a substrate. 3-O-acetyl groups were also resistant to CE16 esterase, but became a substrate when located on the non-reducing end xylopyranosyl residue. An example of such a substrate is the acetylated aldotetrauronic acid (Ac³MeGlcA³Xyl³) generated by GH10 xylanase [3, 94]. *Trichoderma reesei* CE16 (TrCE16) deacetylated in this aldouronic acid both position 3 and 4 (Puchart et al., unpublished data). In a comparative study of *Pa*CE16 and *An*CE16, their positional specificities were observed to differ from TrCE16 in two ways (Puchart et al., unpublished data). Firstly, they deacetylated the aldouronic acid only when the acetyl group migrated to position 4. This position was originally involved in glycosidic linkage. Secondly, they showed ability to deacetylate polymeric substrate at positions similar to AcXEs of CE.
families 1, 4, 5 and 6. Based on these studies, it is viewed that the positional specificity of AcXEs varies considerably with the enzyme and the assigned CE family, implying that the positional preference of any AcXE within a CE family cannot be confidently assumed. However, it has also been suggested that AcXE positional specificity may vary with natural and artificial substrates [4, 66].

4.0 Bioprospecting for microbial AcXEs

Bioprospecting of microbial cultures and genomes for lignocellulose-degrading enzymes has increased dramatically over the past decade [97] and a substantial number of novel AcXEs have been identified, though not all have been functionally characterised. A range of environments containing a diversity of lignocellulose-degrading organisms, such as forest soils, termite guts, animal rumens and compost have yielded a range of AcXE-expressing organisms, including the fungal species Aspergillus [58, 59], Anoxybacillus [76], Trichoderma [98], Orpinomyces [43], Penicillium [57], Schizophyllum [98], Fusarium [99], Coprinopsis [100], Rhodotorula [41] and Neocallimastix [63] and bacteria such as Bacillus [101], Butyrivibrio [102], Caldicellulosiruptor [103], Caldocellum [104], Clostridium [72], Fibrobacter [87], Geobacillus [21], Pseudomonas [105], Streptomyces, Thermoanaerobacterium [106] and Thermobifida [107]. Notably, genomic mining for AcXEs has, until recently, been limited to cultured species. With the rapidly increasing number of available fungal and bacterial genomes, sequence-based genomic mining for AcXEs would be a useful tool in AcXE bioprospecting. However, the use of culture-independent metagenomic screening methods to access novel AcXE genes is an obvious objective [13, 97, 108] and is the focus of this review.
4.1 Mining metagenomes for novel AcXEs

The techniques of metagenomics can be used to either identify putative enzyme genes in large environmental metagenome data sets using sequence-based homology searches or by identification of functional expression of genes without prior sequence knowledge: both have potential for the discovery of a wide range of novel biomolecules [109, 110].

Sequence homology-based screening involves nucleotide sequence-based screening of metagenomic datasets for genes of interest, where the selection of ‘hits’ is based on the identification of consensus sequences or motifs that are unique to specific gene product classes or functions. For identification of AcXE-encoding genes from metagenomic sequence data, it is generally recommended to conduct searches based on conserved AcXE domains [26, 38, 86, 111]. However, the prediction of AcXE genes based on CE family sequence homologies from the CAZy database is complex and unreliable without confirmatory functional studies, since individual CE families include non-AcXE members and members of different CAZy classes may be grouped into a similar Protein Family (PFAM) class. For confirmatory studies, ‘synthetic metagenomics’ [112] can be employed to synthesize predicted genes or the genes can be mined from the actual metagenomic libraries via PCR analysis using specific oligonucleotide probes [113]. Sequence homology-based metagenomic screening has been employed to identify CAZymes [50, 51, 114], but only one of such experiments have resulted in the functional characterization of AcXEs [51]. However, considering their sequence similarities with other non-AcXE CEs, sequence homology-predicted AcXEs may not provide accurate information on functional properties. A recent study [51] showed that only one of two predicted and synthesized AcXE-encoding genes was expressed in the functional screen.

Activity-based screening, also referred to as functional metagenomics, involves functional detection, typically by enzyme assay, of genes expressing from metagenomic DNA fragments
cloned into a suitable expression vector system [109, 115]. AcXEs have been identified using functional screening of metagenomic libraries from a bovine metagenome [12]. The screening technologies specifically used in identifying AcXEs are summarized in Table 2. Functional metagenomic mining can be targeted to highly specific functional properties, while genome or sequence-based metagenome mining only accesses enzymes by class, without explicit specificity of properties. Few reported metagenomic screening studies have specifically targeted AcXEs, with most focusing on other CAZy classes or failing to classify identified CEs into families [116-119]. The limited focus on AcXEs is due, in part, to a lack of high throughput (HTP) metagenomic screening methods for direct detection of AcXE activity. In consequence, the general approach is to use a non-specific HTP screen for lipases/esterases, followed by secondary screening for esterases and third-level assays for AcXEs using acetylated xylan [12]. The polymeric substrate would enable the monitoring of AcXE production on the basis of precipitation of soluble polysaccharide in the screening medium due to deacetylation [98]. Acidification of the screening medium would be a less reliable alternative. The fact that acetylated xylan is not currently commercially available also represents a bottleneck in the AcXE screening technologies. Several researchers [29, 36, 90, 101, 120] report the in-house (and rather time-consuming) synthesis of this substrate using the method of [35], while others extract acetylated xylan from natural substrates using DMSO [34, 66, 78, 121] or steam [33, 58, 62].

It is widely acknowledged that metagenomic screening approaches result in low hit rates [122, 123], the reasons for which (low gene frequency, promoter incompatibility, rare codon usage, nascent protein folding limitations etc.) have been extensively reviewed [113, 124, 125]. Attempts to address the various limitations of functional metagenomics include selective enrichment of the uncultured microbial communities prior to construction of metagenomic
Table 2: Acetyl xylan esterases identified via metagenomic screening

<table>
<thead>
<tr>
<th>Predicted AcXE/CE Family</th>
<th>Source</th>
<th>library/cloning vector</th>
<th>Host</th>
<th>Library/dataset size</th>
<th>No. of clones screened</th>
<th>Hit rate/No. of hits</th>
<th>Screening substrate</th>
<th>Assay substrate</th>
<th>Xylan deacetylation</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>CE12</td>
<td>Compost</td>
<td>PDonR221/pET57&amp;60 DEST</td>
<td><em>E. coli</em></td>
<td>NS</td>
<td>NS</td>
<td>2</td>
<td>NA</td>
<td>pNP-acetate</td>
<td>non-specific</td>
<td>[51]</td>
</tr>
<tr>
<td>AcXE1 (geneB); CE1, CE4</td>
<td>Tamar Wallaby</td>
<td>Fosmid (pCC1Fos)</td>
<td><em>E. coli</em></td>
<td>2.3Mbp</td>
<td>2.3Mbp</td>
<td>9</td>
<td>NA</td>
<td>Not examined</td>
<td>Not tested</td>
<td>[50]</td>
</tr>
<tr>
<td>AcXEs (NS)</td>
<td>Termite gut</td>
<td>Fosmids (pCC1Fos); pBK-CMV</td>
<td><em>E. coli</em></td>
<td>71Gbp</td>
<td>71Gbp</td>
<td>4-34</td>
<td>NA</td>
<td>Not examined</td>
<td>Not tested</td>
<td>[114]</td>
</tr>
<tr>
<td>AcXEs (CE1, CE6)</td>
<td>Bovine* Rumen</td>
<td>λ Phage; pBKr</td>
<td><em>E. coli</em></td>
<td>200,000 clones</td>
<td>14,000 clones</td>
<td>1:24</td>
<td>α-naphthyl acetate</td>
<td>Acetylated birchwood xylan, glucose pentaacetate, tri-O-acetyl-D-galactal, xylose tetraacetate.</td>
<td>Specific</td>
<td>[12]</td>
</tr>
</tbody>
</table>

NA: Not applicable; NS: Not specified; *function-based metagenomic screen
libraries [126] and the use of heterologous hosts other than *Escherichia coli*; e.g., *Streptomyces lividans*, *Pichia pastoris*, *Saccharomyces cerevisae*, *Pseudomonas putida*, *Rhizobium leguminosarum* and *Bacillus subtilis* [59, 109, 127-131] to improve functional expression of heterologous genes. Recently, metagenomic libraries from thermophilic habitats were successfully screened for novel esterases using the bacterium *Thermus thermophilus* as host [132].

The consequence of limitations in functional metagenomic studies, however, is a relatively slow rate of discovery of novel AcXEs and this is evident in CAZy statistics. Only five (Table 2) of all functionally-characterized AcXEs catalogued in the CAZy database (Table 1) were discovered via metagenomic analysis and all derived from one metagenomic screening study [12].

### 4.2 AcXEs from extremophiles

Extremophiles, organisms which typically inhabit environments which are considered to offer one or more ‘extreme’ characteristics (high or low temperatures, pH, salinity, pressures, high radiation; severe desiccation; etc.) [133, 134], are popular targets for bioprospecting on the basis that their gene products are very likely to be adapted to function optimally under the specific extreme conditions [135-137]. A number of AcXEs have been identified from thermophiles and hyperthermophiles, including *Caldicellulosiruptor owensensis* [103], *Thermomonospora fusca* [138], *Thermoanaerobacterium* sp. [106], *Thermobifida fusca* [107, 120], *Thermotoga maritima* [90], *Talaromyces emersonii* [139] and *Geobacillus stearothermophilus* [21]. AcXE1 from *T. maritima*, with optimal activity at 90°C, is the most thermostable AcXE reported [90]. AcXEs have also been identified from alkaliphiles such as
Table 1: Characteristics of CE families to which AcXEs belong

<table>
<thead>
<tr>
<th>CE Family</th>
<th>Member Enzymes</th>
<th>3D Structure</th>
<th>Enzyme source</th>
<th>Other Eukaryotes</th>
<th>Unclassified</th>
<th>Characterised</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>Archaea</strong></td>
<td><strong>Bacteria</strong></td>
<td><strong>Fungi</strong></td>
<td><strong>Eukaryotes</strong></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>AcXEs, cinnamoyl esterases, feruloyl esterases, carboxylesterases, S-formylglutathione hydrolase, diacylglycerol O-acyltransferases, trehalose 6-O-mycolytransferases</td>
<td>α/β/α sandwich</td>
<td>3</td>
<td>3561</td>
<td>99</td>
<td>3</td>
<td>54</td>
</tr>
<tr>
<td>2</td>
<td>6-O deacetylases with some activity on xylan</td>
<td>-</td>
<td>-</td>
<td>171</td>
<td>8</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>AcXEs</td>
<td>α/β/α sandwich</td>
<td>8</td>
<td>131</td>
<td>54</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>AcXEs, chitin deacetylases, chitooligosaccharide deacetylases, peptidoglycan GlcNAc deacetylases, peptidoglycan N-acetylmuramic acid deacetylases</td>
<td>(β/α)γ barrel</td>
<td>16</td>
<td>7329</td>
<td>237</td>
<td>13</td>
<td>16</td>
</tr>
<tr>
<td>5</td>
<td>AcXEs and Cutinases</td>
<td>α/β/α sandwich</td>
<td>-</td>
<td>795</td>
<td>149</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>6</td>
<td>AcXEs</td>
<td>α/β/α sandwich</td>
<td>1</td>
<td>135</td>
<td>3</td>
<td>22</td>
<td>4</td>
</tr>
<tr>
<td>7</td>
<td>AcXEs and cephalosporin C deacetylases</td>
<td>α/β/α sandwich</td>
<td>-</td>
<td>442</td>
<td>-</td>
<td>-</td>
<td>5</td>
</tr>
<tr>
<td>16</td>
<td>Acetyl esterase (multiple substrates)</td>
<td>-</td>
<td>1</td>
<td>29</td>
<td>61</td>
<td>-</td>
<td>2</td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td></td>
<td></td>
<td><strong>29</strong></td>
<td><strong>12593</strong></td>
<td><strong>611</strong></td>
<td><strong>40</strong></td>
<td><strong>90</strong></td>
</tr>
</tbody>
</table>

|                | **Culled from** www.cazy.org 2015                                                   |

**Structurally characterized**

**Functionally characterized**
*Acremonium alcalophilum* [140]. To our knowledge, no characterized ‘extremozymes’ capable of deacetylating xylans have been reported from metagenomic screening projects.

There is clearly considerable scope to expand the bioprospecting of other extreme environments with the capacity for lignocellulosic degradation for AcXEs. For example, the seasonal or intermittent plant productivity of hot desert soils [141] offers an alternative, and largely unexplored, catalogue of thermophilic and/or halophilic and/or alkaliphilic lignocellulose-degrading organisms [132, 142-145]. Cold-active AcXEs might also become a focus for future biomining projects, given the recent development of strategies for bioconversion of lignocellulosic substrates using psychrophilic CAZymes [146, 147].

### 5.0 Applications of AcXEs

The aims of most enzyme bioprospecting projects are largely application-based. The most prominent potential application of AcXEs is specifically in the saccharification of acetylated hemicellulose extracted from plant biomass under non-alkaline conditions [148-150]. The potential role of AcXEs in enhanced lignocellulosic biomass degradation processes has been extensively reviewed [3, 5, 6, 9, 19, 151-155]. While alkaline and certain non-alkaline pre-treatment methods in current use remove most acetyl groups from lignocellulosic biomass, they each have their drawbacks based on factors such as type of lignocellulosic feedstock, temperature requirements, intensity of pre-treatment, detoxification requirements and economic viability [154-156]. No single pre-treatment method can be considered to be ideal for deacetylation of xylans. A recent analysis of the possible role of AcXEs in the bioconversion of lignocellulosic substrates without pre-treatment steps suggests an increased future focus on this enzyme class [103].
AcXEs are also potentially usable in other, less prominent, processes such as in the paper and pulp industry for treating complex carbohydrate waste to enhance its accessibility to xylanases during bleaching of sulfite pulp prepared under mild acidic conditions [157]. Deacetylated xylan is a substrate for the chemical production of hydrogels, which have potential as drug delivery agents [158]. Deacetylated xylooligosaccharides (XOS) are used as food additives and in the synthesis of stereoisomers of monosaccharide sugars [159]. XOS are classified as prebiotics and recommended as high roughage-diets for enhanced gastrointestinal activity in humans and animals [107, 159]. Some AcXEs were found to catalyze perhydrolysis to produce peroxycarboxylic acids, that is, transacylation to peroxide instead of water [160-162]. A CE7 AcXE gene has recently been modified by site-directed mutagenesis to create a perhydrolytic enzyme with improved activity for the production of perhydrolytic acids applicable in laundry care and disinfectant formulations [163]. AcXEs have also been employed in deacetylation of other acetylated polysaccharides [40] as well as in the post-synthetic (in planta) deacetylation of xylan to enhance lignocellulose saccharification [7, 8, 164, 165]. In synthetic carbohydrate chemistry, AcXEs can potentially be harnessed in catalyzing regioselective transacetylation reactions [23, 24, 166].

6.0 Future prospects and conclusion

It is argued that the need for improvements in the efficiency of industrial scale biodegradation of plant biomass necessitates continuous bioprospecting for novel lignocellulose-hydrolysing enzymes [10]. One of the areas where significant advances are possible is considered to be the exploitation of synergies derived from use of accessory enzymes (such as AcXEs) in conjunction with endo-acting cellulases and hemicellulases [2, 29, 45, 107, 167-169]. Options for identification of novel (and possibly superior) variants of such enzymes include the use of
new screening methodologies, and the targeting of under-explored biological communities. Microbial communities within unique (and extremophilic) ecological niches, such as haloalkaline lacustrine habitats, acid-mine drainage, hot and cold deserts soils, etc., are valid targets for metagenomic bioprospecting for novel AcXEs.

The current level of research interest in AcXEs, shown by their proportion of characterized CEs (>30%: Table 1) and publications surveys (Figure 1A), is moderately high. Despite this fact, only four functional metagenomic surveys have targeted AcXE-encoding genes (Table 2), suggesting that the potential for identifying novel genetic, functional and structural AcXE variants remains high. However, we argue that the effective and efficient exploitation of this potential is dependent on the future development of HTP functional screening assays, for example using acetylated xylan or XOS substrates, which are specific to this CAZy group.

**Acknowledgements**

This research was funded by the South African Department of Science and Technology Biocatalysis Initiative and National Research Foundation (DAC, TPM) and the University of Pretoria’s Genomics Research Institute (DAC) and Research Development Program (TPM). FAA was supported by funds from the Organisation for Women in Science in the Developing World (OWSD).

**Conflict of interest**

The authors declare no financial or commercial conflict of interest.
REFERENCES


oligosaccharides generated by a GH10 endoxylanase. BBA - General Subjects, 1830(2013) 5075-86.