JBC Papers in Press. Published on July 9, 2015 as Manuscript M115.664847
The latest version is at http://www.jbc.org/cgi/doi/10.1074/jbc.M115.664847

Functional analyses of resurrected and contemporary enzymes illuminate an evolutionary path for the emergence of exolysis in polysaccharide lyase family two

Richard McLean[1], Joanne K. Hobbs[2], Michael D. Suits[2,3], Sami T. Tuomivaara[4], Darryl Jones[1], Alisdair B. Boraston[2], and D. Wade Abbott[1]

[1]Lethbridge Research Centre, Agriculture and Agri-Food Canada, 5403-1st Avenue South, Lethbridge, AB, T1J 4B1, Canada.
[2]Biochemistry and Microbiology, University of Victoria, Victoria, BC, Canada V8W 3P6.
[3]Chemistry and Biochemistry, Wilfred Laurier University, Waterloo, ON, Canada, N2L 3C5
[4]Complex Carbohydrate Research Center, University of Georgia, 315 Riverbend Road, Athens, Georgia 30602, USA.

***Running Title:** *Structure of a family 2 polysaccharide lyase from Vibrio vulnificus*

To whom all correspondence should be addressed: D. Wade Abbott, Lethbridge Research Centre, Agriculture and Agri-Food Canada, 5403-1st Avenue South, Lethbridge, AB, T1J 4B1, Canada,
Tel: (403)-317-3443; Fax: (403)-382-3156; Email: wade.abbott@agr.gc.ca.

Acknowledgements: This work was supported by Agriculture and Agri-Food Canada (DWA) and a Natural Sciences and Engineering Research Council of Canada Discovery Grant (ABB). MDS was a Michael Smith Foundation for Health Research (MSFHR) Scholar. STT is funded by The BioEnergy Science Center (BESC), a US Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science. We are grateful for the beamline staff at Stanford Synchrotron Lightsource beamline 12-2 for their support and technical assistance. VvPL2 coordinates haven been deposited in the Protein Data Bank (www.pdb.org) under the accession code 5A29.

Keywords: polysaccharide lyase; endolysis; exolysis; beta-elimination; ancestral sequence reconstruction

---

Background: The evolutionary history of family two polysaccharide lyases is unknown.
Results: Functional analysis highlights a key lysine-tryptophan transition involved in exolysis.
Conclusion: Subtle changes in amino acid structure can transform enzyme activity.
Significance: Combinatorial use of ancestral sequence reconstruction, gene resurrection, and structure-function analysis is valuable for elucidating the function and evolutionary history of polysaccharide lyases.

ABSTRACT

Family 2 polysaccharide lyases (PL2s) preferentially catalyze the β-elimination of homogalacturonan (HG) using transition metals as catalytic cofactors. PL2 is divided into two subfamilies that have been generally associated with secretion, Mg^{2+}-dependence, and endolysis (subfamily 1); and intracellular localization, Mn^{2+}-dependence, and exolysis (subfamily 2). When present within a genome, PL2s are typically found as tandem copies, which suggest that they provide complementary activities at different stages along a catabolic cascade. This relationship most likely evolved by gene duplication and functional divergence (i.e. neofunctionalization). Although the molecular basis of subfamily 1 endolytic activity is understood, the adaptations within the active site of subfamily 2 enzymes that contribute to exolysis have not been determined. In order to investigate this relationship, we have conducted a comparative enzymatic analysis of enzymes dispersed within the PL2 phylogenetic tree, and elucidated the structure of VvPL2 from *Vibrio vulnificus* YJ016, which represents a transitional member between subfamiles 1 and 2. In addition, we have used ancestral sequence reconstruction

(ASR) to functionally investigate the segregated evolutionary history of PL2 progenitor enzymes and illuminate the molecular evolution of exolysis. This study highlights that ASR in combination with the comparative analysis of contemporary and resurrected enzymes holds promise for elucidating the origins and activities of other carbohydrate active enzyme families and the biological significance of cryptic metabolic pathways, such as pectinolysis within the zoonotic marine pathogen *V. vulnificus*.

INTRODUCTION

Polysaccharide lyases (PLs) are a class of carbohydrate active enzymes (i.e. 'CAZymes') that have proven useful for investigating convergent enzyme evolution (1-3). PLs deploy a β-elimination mechanism to cleave glycosidic linkages within uronic acids such as homogalacturonan (HG), a homopolymer of galacturonic acid and a primary component of pectin within the cell wall of plants (4). This reaction generates products with a 4,5-unsaturation at the non-reducing end (Fig. 1A). To perform β-elimination, unrelated PL families are dependent on three convergent structural features: a Brønstead base (most commonly an ariginine) to depronotate the C5 carbon, a catalytic metal cofactor (most often Ca^{2+}) to acidify the departing C5 proton and stabilize the oxyanion intermediate, and a stabilizing arginine residue to interact with O2 and O3 of the modified GalA residue (1-3). Cleavage can occur indiscriminately at internal linkages throughout the polysaccharide (i.e. endolysis) or exclusively at the terminus of the substrate (i.e. exolysis; Fig. 1B).

The majority of PL family 2 members (PL2s) partition into one of two functionally distinct subfamilies. Intriguingly, many species contain two paralogous PL2 copies that appear to have arisen by gene duplication and functional divergence (i.e. neofunctionalization). Insights into the functional landscape of these two subfamilies of (PL2) have identified a correlation between cellular localization, mode of activity, and metal selectivity (1,5). Subfamily 1 (e.g. YePL2A) contains secreted endolytic members; whereas, subfamily 2 members (e.g. YePL2B) are intracellular, exolytic, and preferentially harness Mn^{2+} during catalysis (1,6). Interestingly, PaePL2 from *Paenibacillus* sp. Y412MC10, an outlier that is endolytic and preferentially utilizes Mg^{2+} (Fig. 1C; (5)), has provided a snapshot into the evolution of PL2s and the activity of the progenitor enzyme ((5); (Table 1)). A similar relationship has been described for the structurally unrelated PL22 cytoplasmic lyase family (Table 1; (2)). Preferential use of transition metals in PL2s and PL22s is mediated by histidines (PL2 coordination pockets display two histidines; PL22 coordination pockets display three histidines), which displace acidic residues found within Ca^{2+}-selective PLs (1,2). Nitrogen ligands provide more favorable coordination chemistries for transition metals (7).

The earliest diverging outgroup of PL2s is strictly cytoplasmic (2,5), which suggests that transition metals are a prerequisite for intracellular β-elimination. Ca^{2+} is an intracellular signalling molecule and it is present at limiting levels in the cytoplasm of bacteria (0.1-2 μM) to prevent signalling interference and modification of subcellular structures (8,9). In contrast, the periplasm is believed to be a more heterogeneous metallo-environment as extracellular ions are free to passively diffuse across the outer membrane (10). The β-helix PLs (PL1, PL3, and PL9), the largest group of PLs most commonly associated with phytopathogens and saprophytes, are secreted into the periplasm or extracellular environment. β-helix PL families active on HG preferentially coordinate Ca^{2+} (11) and appear to have evolved for colonization and modification of the plant cell wall. Ca^{2+} plays a crucial role in the maintenance of plant cell wall integrity and its levels are high in this environment (10 μM-10 mM; (12)).

Pectins are ubiquitous nutrients for environmental saprophytes, target substrates for macerating phytopathogens (e.g. soft rot), and components of dietary fibres that are digested by symbiotic microbes within the intestines of animals. Perhaps surprisingly, HG utilization and functional pectinases have also been reported for several human enteric pathogens, including *Yersinia* spp. (2,13,14) (Fig. 1C). Although the biological significance of pectinolysis within human pathogens is not clearly understood, several possible roles have been hypothesized, including environmental persistence, colonization of agricultural crops as vectors for transmission,

and utilization of pectic nutrients within the intestine of an infected animal host (15). In this light, the presence of a pectinolytic pathway, complete with transport machinery (KdgM-like porin and solute binding protein), polysaccharide lyases (PL2, PL9, and PL22), and a homologue of a unique HG-binding protein (CBM32; (16)), has been identified within the genome of *Vibrio vulnificus* (Fig. 2A). *V. vulnificus* is a marine-borne bacterium most commonly associated with gastroenteritis caused by the consumption of contaminated seafood, or septicemia resulting from wading in contaminated water with open wounds (17). Correspondingly, pectin represents a nutrient niche that is not consistent with its lifestyle (18). This pathway is not strictly conserved within Vibrionaceae, and whether it represents a historical remnant of a pectinolytic ancestor of *V. vulnificus* or if it evolved by horizontal gene transfer in response to its coastal water-zoonotic infectious lifecycle remains to be determined.

Further insights into the evolutionary history of PL2s after the gene duplication event will help illuminate the adaptations required for metal dependent activity and cellular specialization of pectin utilization, in addition to the biological significance of pectinolysis for various enteric pathogens. This study describes the structure and function of VvPL2, which is the first reported pectinase from *V. vulnificus*. Based upon its phylogenetic position within subfamily 2, VvPL2 represents a potential endolytic-exolytic transitional remnant. Additionally we perform ancestral sequence reconstruction (ASR) of the PL2 family and resurrect progenitor PL2s to compare their activities to contemporary enzymes from subfamilies 1 and 2. We propose that ASR is an underexploited approach within the CAZyme field that will assist in streamlining the characterization of unknown enzyme activities and illuminating the evolutionary basis of substrate recognition and modification in other PL and CAZyme families.

EXPERIMENTAL PROCEDURES

Biochemical characterization of VvPL2

Purification of Enzymes – Synthesized codon-optimized VvPL2, DdPL2, and PaPL2B genes were subcloned in pET28 (BioBasic Int., Mississauga ON), and YePL2A and YePL2B plasmids (1) were transformed into *E. coli* BL21 Star (DE3) cells and grown in LB broth containing 50 µg mL^{-1} kanamycin sulphate. Cells were grown at 37°C with agitation at 180 rpm until cell density reached an A_{600} ~ 0.8. Cultures were cooled to 16°C, agitation was reduced to 120 rpm and genes were induced with a final concentration of 200 µM IPTG. Overnight cultures were centrifuged at 7,000 g for 10 minutes. Cells were chemically lysed by resuspension in a solution of 8% (w/v) sucrose, 0.65% (v/v) deoxycholate, 0.65% (v/v) Triton X-100, 30 mM NaCl, 350 µg mL^{-1} lysozyme, 6 µg mL^{-1} DNase, 30 mM Tris pH 8.0. After lysis, lysate was centrifuged at 13,000 g for 45 minutes. The clarified supernatant was passed through a 0.45 µM filter and applied to a gravity flow nickel affinity chromatography column and eluted with 0.5 M NaCl, 20 mM Tris, pH (8.0) with a stepwise increase in imidazole concentration of 5, 10, 100, and 500 mM. Samples containing protein of interest were concentrated with an Amicon ultrafiltration cell (EMD Millipore) and passed through a HiPrep 16/60 Sephacryl S-200 HR size exclusion chromatography column (GE Healthcare Life Sciences) in 20 mM Tris-HCl, pH 8.0. Pure samples were pooled and concentrated.

Generation of Loop Swap Mutants − Two mutants were created by replacing residues 601–654 of YePL2A with 562–632 of YePL2B as described previously (Horton et al, 1989). Native YePL2A and YePL2B in pET28a were used as templates. 3'-regions were grafted to 5'-regions in a secondary PCR and full gene sequences were ligated into NheI and XhoI restriction enzyme cutsites in pET28a. Ligated transformants were sequenced. Enzymes were produced and purified as above.

Enzyme Assays−Optimal pH was determined by dialyzing samples of enzyme overnight into buffers: Bis-tris pH 6.6-7.2, tris pH 7.1-9.0, CAPSO pH 8.9-10.3, CAPS pH 9.7-11.1, and CABS pH 10.0-11.0. After equilibration, enzyme was incubated at 37°C in 1 mg mL^{-1} HG, 50 mM buffer and the reaction was monitored at 232 nm. Optimal temperature was determined by incubating samples of enzyme in water baths at temperatures ranging from 5-60°C for 15 minutes. Enzyme was then added to 1 mg mL^{-1} HG, 20 mM CAPSO pH 9.0 pre-equilibrated to temperature. Reactions were run for 3 minutes and monitored at 232 nm. Divalent metal cation preference was determined by dialyzing samples of enzyme into 2 mM EDTA in 20 mM CAPSO pH 9.0 to remove divalent metal cations from solution. Fractions were then dialyzed into deionized water and incubated with 1 mgmL^{-1} HG, 50mM CAPSO

pH 9.0 at 37°C to ensure activity had been ablated. Fractions were further dialyzed into solutions containing 1 mM $CaCl_2$, $MgCl_2$, $MnCl_2$, $CoCl_2$, $NiCl_2$, or $CuCl_2$, CAPSO pH 9.0. After equilibration, samples were incubated in 1 mg mL^{-1} HG, 50 mM CAPSO pH 9.0 and monitored at 232 nm.

Time course experiments were performed to determine product profiles. Enzymes were incubated in 1 mg mL^{-1} HG, 50 mM CAPSO pH 9.0, 1 mM $MnCl_2$ at 37°C. Reactions were stopped by heating the samples to 95°C for 10 minutes followed by flash freezing in liquid nitrogen. Samples were then resolved by thin layer chromatography with 1-butanol : distilled water : acetic acid (5 : 3 : 2 v/v/v) running buffer and visualized with 1% orcinol in a solution of ethanol sulfuric acid (70 : 3 v/v) followed by heating at 110°C for 10 minutes. Samples were compared to samples of GalA (Sigma: #48280), $GalA_2$ (Sigma: #D4288) and $GalA_3$ (Sigma: #T7407)

Enzyme Kinetics – PLs (100 nm – 1 μm) were incubated in increasing concentrations of HG and $GalA_3$ with 50mM CAPSO pH 9.0, 1mM $MnCl_2$. Samples were monitored in real time at 232nm and product formation was determined using the extinction coefficient 5,200 M^{-1} cm^{-1}. Data was analyzed and kinetic values determined using GraphPad Prism 6.

Crystallization and structure solution of VvPL2– crystals of VvPL2 developed via hanging drop vapour diffusion method at a protein concentration of 15 mg ml^{-1} by mixing 1.0 μl of the protein solution with an equal volume mother liquor consisting of 16% (w/v) polyethylene glycol 3,350, 0.14 M Na/K tartrate, and 0.1 M HEPES (pH 7.0) at 19°C. Crystals were cryoprotected by brief crystal immersion into a solution of the reservoir solution supplemented with 25% ethylene glycol, and subsequently frozen in a liquid nitrogen stream prior to diffraction experiments.

VvPL2 crystallized in space group $P6_5$ with one protein molecule in the asymmetric unit. Diffraction data for VvPL2 in complex with two molecules of tartrate was collected at the beamline SSRL 12-2 of the Stanford Synchrotron Radiation Lightsource. The dataset was processed with XDS and scaled with Scala (19). The correct phases were derived via molecular replacement with the program Phaser (20) using the *Y. enterocolitica* PL2 structure (YePL2A, PDBID 2V8J) as a search model (1). The structure of VvPL2 was rebuilt with the program Buccanneer and iteratively improved with cycles of manual building with Coot and positional refinement with Refmac (21-23). Data collection, processing, and refinement statistics were generated by Molprobity (24) are presented in Table 2. Ramachandran statistics were generated using Rampage (25). Coordinates for VvPL2 tartrate-bound structure are deposited in the Protein Data Bank (www.pdb.org) under the accession code: 5A29. Mapping of the degree of residue conservation was performed with the program Consurf (26) and figures were produced using Pymol (pymol.org).

Phylogenetic analysis of the PL2 Family−PL2 sequences were retrieved from the CAZy database (www.cazy.org) and curated to remove truncated or duplicated sequences. An initial amino acid sequence alignment was built using Gblocks (27) and a guide tree subsequently generated using PhyML (28). This guide tree was then utilized by (29) to align the full-length sequences. A maximum likelihood (ML) phylogeny was constructed using GARLI 2.0 (30) and the appropriate model of evolution (LG + I + G) as determined by ProtTest 3.4 (31). A Bayesian phylogeny was also generated using MrBayes 3.2.4 (32) and a mixed amino acid model with two parallel runs in order to ensure convergence. Both phylogenies were rooted on the branch between the gammaproteobacteria and outgroup sequences. Bootstrapping was performed in GARLI 2.0 using 1,024 replicates and a 10% burn-in. All trees were visualized using Geneious 6.1.8 (33).

Ancestral inference − Maximum likelihood ancestral inference was performed in PAML 4.3 (Yang, 2007) on the basis of nucleotide, codon and amino acid sequences using the ML phylogeny constructed for PL2. For nucleotide inference in BASEML, a nucleotide alignment exactly matching the PL2 amino acid alignment generated by PRANK was constructed using Geneious 6.1.8, and the appropriate model of nucleotide substitution (GTR) was implemented as determined by jModelTest 2 (34). For codon and amino acid inference in CODEML, the WAG amino acid rate file was employed. The sequences inferred by the three methods were compiled and a majority-rules approach was taken, with any remaining ambiguous sites resolved after consideration of the physicochemical properties of the inferred amino acids, their frequency among the contemporary sequences and the inference made by the codon method (which is considered to be the most accurate). Bayesian ancestral inference was performed in MrBayes 3.2.4 using a mixed amino acid model. Ancestral gaps were inferred using PRANK and incorporated into the ancestral sequences inferred by PAML and MrBayes.

Biochemistry of ancestral PL2s − Node 49, 52, 54, and 74 sequences were codon optimized, synthesized, and subcloned into a pET28 (Novagen, #69864) expression vector using NheI and XhoI directional restriction sites (Biobasic Inc., Markham, ON). The Node 52 gene was subsequently subcloned into pET32 (Novagen, #69015) to increase soluble yields. Gene products were purified by IMAC, eluted with a 0-500 mM imidazole gradient, and dialyzed into 20 mM Tris-HCl, pH 8.0. Digests were performed using 0.1 μM enzyme and 1 mg ml^{-1} HG at 37 ˚C. Node 52 and 54 were performed at pH 8.0 (4 mM Tris-HCl), and Node 74 at pH 9.4 (4 mM CAPS). Direct method metal recovery assays were performed by adding EDTA to a final concentration of 1 mM, and supplementing with 10 mM of $CaCl_2$, $MgCl_2$, or $MnCl_2$. Reactions were heat killed by boiling for 5 min and clarified by centrifugation at 13,000 g. Products were analyzed directly or following a 10-fold concentration by TLC (as above) or high performance anion exchange with pulsed amperometric detection (HPAEC-PAD). HPAEC-PAD analysis was performed with a Dionex ICS-3000 chromatography system (Thermo Scientific) equipped with an autosampler as well as a pulsed amperometric detector for total carbohydrates, and a UV-Vis detector for unsaturated galacturonides. Aqueous sample (typically 10 μl) was injected into an analytical (4 x 250 mm) CarboPac PA1 column (Thermo Scientific) and eluted at 0.4 mL min^{-1} flow-rate with a Na-acetate gradient (0 to 1 min: 250 mM, 1 to 17.5 min: 250 to 1,000 mM, 17.5 to 20 min: 1,000 mM, 20 to 21 min: 1,000 to 250 mM, 21 to 35 min: 250 mM) in a constant background of 100 mM NaOH.

Site directed mutagenesis of YePL2A and YePL2B − Nucleotide substitutions were generated via PCR mediated site-directed mutagenesis. Mutagenic primer sets were extended with KOD polymerase (Novagen #71086) using pETPL2A or pETPL2B (Abbott and Boraston, 2007), encoding the YePL2A and YePL2B proteins respectively, as template. The entire reaction mixture was then digested with DpnI (NEB #R0176), and 1/10th the reaction mixture transformed into DH5α competent cells. Plasmid was extracted (Omega #D6945) and constructs were verified by Sanger sequencing (McGill University and Génome Québec Innovation Centre).

RESULTS AND DISCUSSION

Biochemical characterization of PL2 subfamily 2 enzyme − To explore the full profile of PL2 subfamily 2 activities, sequence entries from its four major clades were analyzed. These representative enzymes include *Yersinia enterocolitica subsp. enterocolitica* 8081 (YePL2B; gene ID: YE1886), *Pectobacterium atrosepticum* SCRI1043 (PaPL2B; gene ID: ECA2402), *Dickeya dadantii* 3937 (DdPL2; gene ID: Dda3937_03361), and *V. vulnificus* YJ016 (VvPL2; gene ID: VVA1383; res. 18-556). In addition, the endolytic YePL2A (gene ID: YE4069) was purified to enable comparisons with a previously characterized subfamily 1 member ((1); Fig. 1C). As anticipated, YePL2B, PaPL2B, and DdPL2 exclusively released unsaturated digalacturonate ($\Delta GalA_2$ where $_{[n]}$ = the degree of polymerization, d.p.), which is consistent with an exolytic mode of activity (not shown). Unexpectedly, VvPL2 displayed an endolytic product profile (Fig. 2B) and pH optimum reminiscent of subfamily 1 enzymes (Fig. 2C); however, its temperature optimum reflected a similar distribution to YePL2B (Fig. 2D-F).

In order to compare specificities and activities, we have performed a comparative kinetic analysis between VvPL2, YePL2A, and YePL2B on both HG and the pectic fragment $GalA_3$ (Fig. 2G-2I, and Table 3). YePL2A is preferentially active on HG over $GalA_3$ (~11-fold); whereas its paralog YePL2B is preferentially active on $GalA_3$ over HG (~13-fold). This inverse relationship is consistent with their assigned roles in a degradative pathway. YePL2A is secreted and endolytic, which is tailored for upstream activity on polymeric HG, and YePL2B is intracellular and exolytic, and performs a downstream role in oligogalacturonide depolymerization (15). In comparison, VvPL2 displays a relatively high catalytic rate on both $GalA_3$ and HG, with preferential activity on $GalA_3$ (~5-fold). This plasticity may be explained by *V. vulfinicus* only containing one PL2 copy. In this context, VvPL2 appears to take on the roles of both YePL2A and YePL2B, and its position within the PL2 family tree suggests it represents a transitional member based upon both sequence relatedness and function (Fig. 1C).

Previously, PL2s have been reported to preferentially utilize transition metals over Ca^{2+} during catalysis (1,5,6). To further explore a

differential relationship between PL2 subfamilies and metal specificity, YePL2A, YePL2B, and VvPL2 were subjected here to a new preparative treatment that exchanges metal cofactors by performing an exhaustive dialysis against EDTA-buffered solutions, followed by exhaustive dialysis in divalent metal buffered solutions. This alternative method was developed to promote gradual exchange of cofactors and supplant the direct depletion-supplementation method that has been routinely used previously (1,6). The direct approach introduces highly concentrated metallo-microenvironments and can be deleterious to protein stability. For example, the characterization of YePL2A metal-dependence was not previously possible using the direct method (1). With the dialysis substitution method implemented here, YePL2A displayed very little precipitation during its preparation with all metals tested. Initial velocities for YePL2A, YePL2B, and VvPL2 in the presence of Co^{2+}, Mn^{2+}, Ni^{2+}, Mg^{2+}, Ca^{2+}, and Cu^{2+} were determined to compare related metal dependence of HG modification (Fig. 2J). For YePL2A and VvPL2 the highest catalytic rates were observed in the presence of Mg^{2+} (at 1 mg ml^{-1} HG), which agrees with what was reported for PaePL2 and supports a prominent role for Mg^{2+} as a metal cofactor in endolytic PL2s (1,6). In contrast, YePL2B displayed the highest activity when supplemented with Mn^{2+}.

To investigate the mechanistic contributions of various metals and potentially the biological significance of metal selectivity, full Michaelis-Menten kinetics were determined for YePL2A and YePL2B using the same spectrum of cofactors (Table 4, Fig. 2K & 2L). Mg^{2+} and Co^{2+} were confirmed to promote the highest turnover rate for YePL2A. Mn^{2+}; however, was associated with the lowest K_m, which translated into a four-fold higher catalytic specificity constant for Mn^{2+} than Mg^{2+} (Table 4). This suggests that Mn^{2+} is optimal under limiting concentrations of substrate; regardless, YePL2A demonstrates remarkable plasticity in cofactor selection. This property reflects the adaptation of secreted PL2s to the heterogeneous ionicity of the periplasm (10). The results for YePL2B indicate that the cytoplasmic enzyme has more selectivity for Mn^{2+} in both the rate of substrate turnover (k_{cat}) and catalytic efficiency (k_{cat}/K_m) (Table 4). Higher specificity (K_m) for and catalytic turnover (k_{cat}) with Mn^{2+} translates into a 8-fold increase over Mg^{2+} and ≈3 orders of magnitude increase over Ca^{2+} in catalytic rate. This finding highlights that within the confines of the cell, cofactor selection by PL2s for Mn^{2+} is more stringent. In the presence of Mn^{2+}, YePL2B appears to adopt a substrate inhibition profile when active on HG (Mn^{2+*}; Ki = 2.25 ± 1.07 mg mL^{-1}; Fig. 2L; Table 3). When fit to this model, the catalytic efficiency is lowered into the range of Mg^{2+}; however, this result should be interpreted with caution as the error values increase, which may compensate for this effect. Although this observation underpins the complexity of the metal-protein-HG interaction for YePL2B, such inhibitory effects are likely negligible in nature as HG concentrations would be limited inside the cell.

Structural Analysis of VvPL2, an endolytic member of subfamily 2 − In order to provide insight into the molecular determinants of subfamily 2 PL2 activities, we attempted to crystalize DdPL2, PaPL2B, YePL2B and VvPL2. Although we were able to produce high levels of each recombinant protein, they varied in solubility, stability, and crystallisability. Of the four proteins, only VvPL2 produced diffraction quality crystals which were used to solve the protein structure by molecular replacement to 1.90 Å using YePL2A as a homology model (PDBID: 2V8K; (1)). VvPL2 adopts an $(\alpha/\alpha)_7$ barrel fold with an extensive active site cleft that is characteristic of endolytic enzymes (Fig. 3A).

Superimposition of VvPL2 (residues 26 – 566) onto YePL2A using PDBeFold (http://www.ebi.ac.uk/pdbe) highlighted the close structural similarity between the two lyases (r.m.s.d. = 1.39 Å over 509 residues (35). Significantly, the Brønstead base (R191), catalytic pocket and stabilizing residue (R304) are structurally conserved, with R304 approximately 13.8 Å from the metal centre with reasonable geometry for interacting with 2-OH and 3-OH of the GalA in the +1 subsite (Fig. 3B). These three convergent features have been suggested to be critical for β-elimination in unrelated PL folds (1-3). Based on an inspection of B-factors and residual electron density, the metal coordinated in the VvPL2 structure appears to be Ni^{2+} or Mn^{2+} (Fig. 3C). While this may be influenced by the purification conditions, VvPL2 function was

modest in the presence of Ni^{2+} (Fig. 2J). Therefore, whereas the natural metalloenzyme complex is assumed to contain Mn^{2+}, we cannot negate with the available data that the metal-cofactor presented in the crystal structure has not been substituted with Ni^{2+}. Intriguingly, the uncleaved N-terminal histidine tag from VvPL2 introduced two artifactual interactions between the Nε2 of VvPL2 residues H4 and H6, and the bound metal assisted in stabilizing the coordination sphere with a perfect octahedral symmetry (Fig. 3C). Other interactions involve H129 (Nε2), E150 (Oε2), H192 (Nε2), and an ordered water molecule (HOH 383) that is activated by a 2.8 Å hydrogen bond with the Nδ1 imidazole nitrogen of H546.

Sequence comparison between the metal binding pockets of YePL2A and VvPL2 reveal the presence of a histidine residue in VvPL2 (H546) that replaces a glutamate in YePL2A (E515), which was presumed to contribute to Mn^{2+} selective chemistries for subfamily 2. Structural superimpositions of the metal binding pockets, however, reveals that H546 and E515 are spatially and functionally conserved (Fig. 3D). Both residues interact with an ordered water molecule, charging it for coordination of the metal cofactor. To investigate if the Nε of the imidazole group provided any definitive function, we performed both a single substitution (H530E) and insertion of the tripeptide sequence from YePL2A (F512/T513/E514) into the equivalent sequence space on YePL2B (Y528/I529/H530). These mutations did not reverse the metal selectivity in YePL2B as optimal digestion was still observed in the presence of Mn^{2+} after EDTA treatment; however, it did appear to be deleterious for the utilization of Ca^{2+} (not shown). This observation does not rule out a differential role for H546 in exolytic enzymes, but does suggest that the more stringent metal selectivity observed in the catalytic activity of YePL2B likely depends on other structural features within the metal binding site, such as residue geometry and distance, which would be supported by a greater network of interactions within the enzyme scaffold.

Based upon its activity (Table 3) and position within subfamily 2 (Fig. 1C), VvPL2 appears to represent a transitional sequence within the phylogeny of PL2. Therefore we examined the surface of VvPL2 to identify any structural features near the active site cleft that might help illuminate the structural transition between the endolytic and exolytic subfamilies. Subfamily 1 and subfamily 2 sequences were independently mapped onto the structure of VvPL2 using Consurf ((26); Fig. 3E & 3F). This program scales the conservation (magenta) and divergence (cyan) of residues to identify similar and distinct structural elements. Near its catalytic centre VvPL2 displays a high level of structural similarity with subfamily 1 sequences, which is consistent with its observed activity (Fig. 3E). Apart from the core catalytic residues, there is notably less conservation with PL2 subfamily 2 sequences (Fig. 3F). One such region includes a hallmark lysine residue (K300) that is poised near the exit of the active site cleft. This lysine is invariant in subfamily 1, and is replaced with a tryptophan through the majority of subfamily 2 sequences (Fig. 4). Intriguingly, the small outgroup of early diverging sequences in subfamily 2 that includes two *Marinomonas* spp. and *Acholeplasma brassicae* display a phenylalanine and glycine, respectively at this position (not shown).

Ancestral Sequence Reconstruction of Family 2 PLs − The biochemical properties of VvPL2 have revealed that subfamily boundaries assigned within the CAZy database do not provide enough sequence resolution to elucidate the evolutionary history of endolytic to exolytic transition in the PL2 family. Therefore, to trace lineage at the sequence level we constructed a robust maximum likelihood (ML) phylogeny of all available PL2 sequences and used this analysis to infer the sequences of ancestral PL2s positioned at a range of branch points (Fig. 4A). Almost all of the contemporary PL2 sequences available are from members of the gammaproteobacteria, with the exception of two sequences from *Paenibacillus sp.*, a member of the Firmicutes, and *Haloterrigena turkmenica*, an archaeon, which were used as an outgroup and to root the tree. The topology of the ML phylogeny shown in Fig. 4A is supported by high bootstrap percentages; furthermore, a Bayesian phylogeny was also constructed for comparison and found to display identical topology (not shown). The contemporary PL2 sequences form two major clades, subfamily 1 and subfamily 2, with subfamily 1 positioned

closest to the root. The ancestral nodes 49, 52, 54 and 74 were selected for ancestral inference and reconstruction given their positions at major branch points within the phylogeny (Fig. 4A). Node 49 represents the last common ancestor (LCA) of all PL2 sequences, including the outgroup sequences, whereas Node 74 represents the LCA of the subfamily 1 PL2s alone. Nodes 52 and 54 both represent ancestors of subfamily 2 post-divergence of VvPL2, with Node 54 being the ancestor of all subfamily 2 PL2s from plant pathogens and Node 52 being the ancestor of these same enzymes, plus the endolytic VvPL2 and PL2s from *Vibrio furnissi* and *Acholeplasma brassicae*. The positions of all four of these PL2 ancestors are supported by bootstrap percentages ≥98%. Ancestral inference was performed under the ML criterion and the average posterior probability for each of the four ancestors was >0.7 (this increases to >0.8 for Nodes 52, 54 and 74 when inference at ancestral gaps is not considered).

The four ancestral PL2s vary from their closest contemporary descendant by at least 15% (approx. 83 amino acids). As expected from its phylogenetic position, the closest contemporary descendant of Node 74 is a subfamily 1 PL2 from *Pectobacterium wasabiae* (84% sequence identity) and it possesses a lysine residue (K286) conserved within the endolytic subfamily 1. In contrast, Node 54 shares the greatest sequence identity with a subfamily 2 enzyme from *Yersinia pseudotuberculosis* (82%) and contains the conserved tryptophan residue (W286) at this same position in the active site cleft. Node 52 shares only 60% sequence identity with its closest contemporary descendant (VvPL2). Despite Node 54 sharing only 45% sequence identity with VvPL2, Nodes 52 and 54 share 65% sequence identity. Interestingly, Node 52 does not contain either the conserved lysine found in subfamily 1 or the conserved tryptophan found in subfamily 2, rather it contains an arginine residue (R283; Fig. 4B). The most divergent of the inferred ancestral PL2s is Node 49, sharing just 56% sequence identity with its closest contemporary descendant (a subfamily 1 PL2 from *P. wasabiae*). Based upon sequence alignments, Node 49 does not appear to contain a lysine or tryptophan residue at this position; however, structural modelling determined that this ancestor has a truncated loop and a lysine (K268) is spatially conserved (not shown). This observation suggests that a lysine at this position is correlated with endolytic activity, the LCA of PL2s was endolytic, which would be consistent with what was previously proposed for the early diverging PaePL2 (5).

Characterization of resurrected ancestral PL2s − From their inferred sequences it appears that Node 74 is endolytic and Node 54 is exolytic but, as we have observed with VvPL2, sequence information alone cannot fully predict function. Furthermore, Node 52 contains a divergent amino acid in place of the highly conserved lysine or tryptophan residue associated with endolytic and exolytic activity, respectively. Therefore, we resurrected and characterized the enzymes from Nodes 49, 52, 54 and 74 *in vitro* using gene synthesis and enzyme product profiling. Nodes 52, 54 and 74 were produced as soluble protein in *E. coli* and found to be active on HG (Fig. 5A); however, Node 49 did not produce and could not be studied further. In agreement with its phylogenetic position as the LCA of subfamily 1 and the presence of K286, Node 74 displayed characteristic endolytic activity on HG, with detected products ranging in size from ΔGal_2-ΔGal_4. Similarly, as an ancestor of subfamily 2, Node 54 displayed an exolytic profile and almost exclusively generated ΔGal_2. Node 52 also generated an exolytic-like digestion profile of HG despite possessing an arginine at the K268 position. This residue may represent a key transition in the evolution of subfamily 2 sequences as despite having related charge potentials, arginine has more steric bulk than lysine. Intriguingly, both residues can display identical adenine bases in their first and third codon positions (lysine = AAA/AAG; arginine = AGA/AGG), which suggests substitutions can arise by in-frame substitutions. Analysis of the nucleotide sequence of YePL2A reveals that K265 is encoded by a tri-adenine codon, and the second position of W300 in YePL2B contains a guanine. Therefore it seems plausible that the AAA-lysine encoding position may have evolved first to an AGA-arginine, and subsequently to a TGG-tryptophan (Fig. 5B). Alternatively, it could have proceeded through an AAA>AAG silent mutation, and then an AGG-arginine intermediate. In either case, this pathway would suggest that exolysis

arose in part though increases in the steric bulk of this positional residue (lysine > arginine > tryptophan) and neutralization of its charge (positive > neutral) (Fig. 5B).

Node 74 purified in high yields and was therefore used to further probe ancestral functions and relationships between PL2 subfamilies. This enzyme displays a similar pH profile (not shown) to YePL2A and maximal enzyme recovery with Mg^{2+} (Fig. 5C). These data indicate that ASR can accurately determine the functional relatedness of PL2s back to the subfamily divergence in their lineage, and suggests that ASR will have utility for helping to define the phylogenetic relationships in other CAZyme families.

Evolution of exolytic and endolytic activities within the PL2 family − Despite numerous attempts (e.g. YePL2B, DdPL2, PaPL2B) we have been unable to solve the structure of an exolytic PL2, and currently the molecular basis of exolytic activity in this family remains to be determined. Previously, an endolytic-exolytic switch was proposed to be the result of a loop insertion near the catalytic centre (residues: 200-218 of YePL2A and 188-212 of YePL2B; (1)). Loop insertions have commonly been observed to be responsible for exolytic-endolytic transformations within CAZymes, including polygalacturonases (1) and family 11 PLs (36). In YePL2B, and by extension other exolytic PL2s, the catalytic cleft would need to be remodeled to accommodate the reducing end of HG with subsites +1 and +2 for the exclusive release of $\Delta GalA_2$ ((37) Fig. 5D). Therefore, we attempted to define the structural role of the loop in YePL2B by performing loop-swapping experiments between YePL2A and YePL2B to generate the hybrid enzymes YePL2A-B (containing the B-loop) and YePL2B-A (containing the A-loop). Swapping the loops between the two enzymes lowered the rate of HG digestion but did not alter their respective product profiles (Table 3 and not shown), which suggests that the predicted YePL2B loop is not the molecular determinant of exolysis. Intriguingly, the YePL2B loop shifted the pH optimum of the YePL2A towards YePL2B (not shown) and reduced the affinity for HG but not $GalA_3$ (Table 3), which indicates that the loop may contain residues that contribute to formation of distal subsites for accommodating polymerized HG.

In order to identify more subtle features that contribute to the structural basis of exolytic activity, we next performed a thorough examination of the primary structure alignments of the node enzymes and a homology model of YePL2B (not shown). Consistent with what was revealed through the ASR analysis, there is a surface exposed tryptophan conserved within all contemporary exolytic enzymes (YePL2B: W300) and Node 54 (W286), which underpins that it may have a functional role. To test this possibility, we performed substitutive mutagenesis on this tryptophan (W300K and W300A). The product profile of ΔYePL2B-W300K and ΔYePL2B-W300A contained several populations, which suggests that the mutants had become endolytic (Fig. 5E). This effect was enhanced in the presence of EDTA (Fig. 5F). It appears that W300 functions to stabilize the exolytic cleft, perhaps by occluding access to the active site and restricting interactions with polymerized HG to the reducing end (Fig. 5D). Additionally, the role of EDTA in generating this phenotype suggests that the modified cleft structure of YePL2B may be fortified by interactions with the catalytic metal. In the absence of a structure from an exolytic PL2, these results shed new light on how subtle transitions in primary structure can transform enzyme activity within closely related enzyme families.

Biological significance and evolution of HG utilization pathways within human enteric pathogens − PL2s are disproportionately found in human enteric pathogens and there are often paralogous copies within a genome that partition into subfamilies 1 and 2 (Fig.1C & 4A; (5)). The presence of two distinct PL2 activities that display differential cellular localization highlights that they are contributing to upstream endolytic (secreted) and downstream exolytic (cytoplasmic) stages of HG depolymerization (Fig. 6; (18)). Several examples of species with a single copy of either an exolytic or endolytic PL2 entry do exist (Fig. 4A; (5)); however, in these cases alternative pathways for HG saccharification have evolved (15,18). The pectinolyic pathway from *V. vulnificus* is one such example (Fig. 6). *V. vulnificus* is predicted to deploy an extracellular PL9, a periplasmic HG binding protein (endoVvCBM32) and endoVvPL2, and an

intracellular oligogalacturonte lyase (exoVvPL22). HG transport is facilitated through a KdgM-like anionic porin (38,39) and intracellular transport is predicted to be facilitated by a solute-binding protein and an ABC-transporter that is distally located in the genome but under similar regulation (18). This pathway differs from what has been biochemically defined for *Y. enterocolitica* (Fig. 6; (15)). *Y. enterocolitica* deploys an extracellular pectin methylesterase (YeCE8; (40)); periplasmic HG binding protein (endoYeCBM32; (16)), endolytic PL2 (endoYePL2A; (1)), and exolytic polygalacturonase (exoYeGH28; (14)); and two intracellular depolymerases, which cleave $\Delta GalA_2$ (exoYePL2B; (1)) and ΔGalA (exoYePL22; (2)), respectively from oligogalacturonide substrates. The signature architectures of these pathways may reveal subtle variances in the structure of pectic nutrients and symbioses (e.g. marine vs. terrestrial) or differential roles in environmental persistence and colonization of competitive ecosystems such as the gastrointestinal tract of animals. Further investigation into biochemical function and evolution of pectin utilization pathways containing PL2s will be foundational for understanding the roles of these enzymes in the lifecycle, and potentially in the pathogenesis, of human enteric pathogens.

CONCLUSION

Assigning ancestry and biological function to sequence-based CAZyme families has been complicated by the realization that many families display great diversity in substrate specificity or mode of activity. In this light, recent efforts to partition CAZyme families into subfamilies (41) and develop *in silico* tools to predict function based upon structural signatures (42-44) have helped to facilitate the sequence-to-function based characterization of protein-carbohydrate interactions and carbohydrate modifying enzymes. We have demonstrated here, however, that a higher level of resolution may be required to define the functional boundaries and evolution of activities within some CAZyme subfamilies. Within PL2s, the progenitor enzyme appears to have been endolytic and to preferentially harness Mg^{2+} for β-elimination (PaePL2 and Node 74); however, it is clear that contemporary endolytic PL2s (subfamily 1) display plasticity in metal selectivity, which can be explained by the heterogeneous metallo-environment of the periplasm and extracellular environment of the niches that these bacteria colonize. In contrast, intracellular PL2s (subfamily 2) display the highest rate of substrate turnover in the presence of Mn^{2+} and are exolytic.

Insights into the structure of VvPL2, which represents a 'transitional' enzyme that exhibits some properties of both subfamilies, and the biochemical characterization of resurrected enzymes from the lineages of both subfamilies has revealed that the molecular basis of an endolysis to exolysis transition is not loop-dependent, but rather relies on subtle changes to functional groups at the opening to the active site cleft. For example it appears that a lysine to tryptophan transition is in part responsible for the emergence of exolysis, and this mutation may have evolved through a lysine (AAA) > arginine (AGA) > tryptophan (TGG) transition. Future research aimed at illuminating the evolution of CAZyme subfamily substrate specificity and mode of activity will be central to defining general properties in the evolution of pectin recognition and modification, and the colonization of intriguing nutrient niches by microbes, such as pectinolysis by foodborne pathogens.

Conflict of Interest Statement: The authors have no competing interests to report.

Author Contributions: RM performed and processed enzyme kinetics and metal supplementation assays, performed YePL2A/B swapping mutagenesis, crystallized VvPL2, and assisted in figure preparation; JKH performed ancestral sequence reconstruction, provided comparative analysis of contemporary and resurrected enzyme sequences, and assisted in manuscript writing and figure preparation; MDS solved the structure of VvPL2, refined and analyzed the model, and assisted in figure preparation; STT performed the product profiling of PLs and assisted in figure preparation; DJ performed site directed mutagenesis, and digestions with mutant and ancestral enzymes; ABB assisted in structural analysis and project conceptualization; DWA conceived and coordinated the study, wrote the paper, and prepared figures. All authors analyzed results and approved of the final version of the manuscript.

REFERENCES

1. Abbott, D. W., and Boraston, A. B. (2007) A family 2 pectate lyase displays a rare fold and transition metal-assisted beta-elimination. *The Journal of biological chemistry* **282**, 35328-35336
2. Abbott, D. W., Gilbert, H. J., and Boraston, A. B. (2010) The active site of oligogalacturonate lyase provides unique insights into cytoplasmic oligogalacturonate beta-elimination. *The Journal of biological chemistry* **285**, 39029-39038
3. Charnock, S. J., Brown, I. E., Turkenburg, J. P., Black, G. W., and Davies, G. J. (2002) Convergent evolution sheds light on the anti-beta -elimination mechanism common to family 1 and 10 polysaccharide lyases. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 12067-12072
4. Caffall, K. H., and Mohnen, D. (2009) The structure, function, and biosynthesis of plant cell wall pectic polysaccharides. *Carbohydrate research* **344**, 1879-1900
5. Abbott, D. W., Thomas, D., Pluvinage, B., and Boraston, A. B. (2013) An ancestral member of the polysaccharide lyase family 2 displays endolytic activity and magnesium dependence. *Applied biochemistry and biotechnology* **171**, 1911-1923
6. Shevchik, V. E., Condemine, G., Robert-Baudouy, J., and Hugouvieux-Cotte-Pattat, N. (1999) The exopolygalacturonate lyase PelW and the oligogalacturonate lyase Ogl, two cytoplasmic enzymes of pectin catabolism in Erwinia chrysanthemi 3937. *Journal of bacteriology* **181**, 3912-3919
7. Harding, M. M. (2001) Geometry of metal-ligand interactions in proteins. *Acta Crystallogr D Biol Crystallogr* **57**, 401-411
8. Gangola, P., and Rosen, B. P. (1987) Maintenance of intracellular calcium in Escherichia coli. *The Journal of biological chemistry* **262**, 12570-12574
9. Yanyi, C., Shenghui, X., Yubin, Z., and Jie, Y. J. (2010) Calciomics: prediction and analysis of EF-hand calcium binding proteins by protein engineering. *Science China. Chemistry* **53**, 52-60
10. Ferguson, A. D., and Deisenhofer, J. (2004) Metal import through microbial membranes. *Cell* **116**, 15-24
11. Garron, M. L., and Cygler, M. (2010) Structural and mechanistic classification of uronic acid-containing polysaccharide lyases. *Glycobiology* **20**, 1547-1573
12. Hepler, P. K., and Winship, L. J. (2010) Calcium at the cell wall-cytoplast interface. *Journal of integrative plant biology* **52**, 147-160
13. Boraston, A. B., and Abbott, D. W. (2012) Structure of a pectin methylesterase from Yersinia enterocolitica. *Acta crystallographica. Section F, Structural biology and crystallization communications* **68**, 129-133
14. Abbott, D. W., and Boraston, A. B. (2007) The structural basis for exopolygalacturonase activity in a family 28 glycoside hydrolase. *Journal of molecular biology* **368**, 1215-1222

15. Abbott, D. W., and Boraston, A. B. (2008) Structural biology of pectin degradation by Enterobacteriaceae. *Microbiol Mol Biol Rev* **72**, 301-316
16. Abbott, D. W., Hrynuik, S., and Boraston, A. B. (2007) Identification and characterization of a novel periplasmic polygalacturonic acid binding protein from Yersinia enterolitica. *Journal of molecular biology* **367**, 1023-1033
17. Oliver, J. D. (2013) Vibrio vulnificus: death on the half shell. A personal journey with the pathogen and its ecology. *Microbial ecology* **65**, 793-799
18. Rodionov, D. A., Gelfand, M. S., and Hugouvieux-Cotte-Pattat, N. (2004) Comparative genomics of the KdgR regulon in Erwinia chrysanthemi 3937 and other gamma-proteobacteria. *Microbiology* **150**, 3571-3590
19. Kabsch, W. (2010) Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125-132
20. McCoy, A. J. (2007) Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr D Biol Crystallogr* **63**, 32-41
21. Cowtan, K. (2012) Completion of autobuilt protein models using a database of protein fragments. *Acta Crystallogr D Biol Crystallogr* **68**, 328-335
22. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501
23. Vagin, A. A., Steiner, R. A., Lebedev, A. A., Potterton, L., McNicholas, S., Long, F., and Murshudov, G. N. (2004) REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr D Biol Crystallogr* **60**, 2184-2195
24. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21
25. Lovell, S. C., Davis, I. W., Arendall, W. B., 3rd, de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S., and Richardson, D. C. (2003) Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins* **50**, 437-450
26. Ashkenazy, H., Erez, E., Martz, E., Pupko, T., and Ben-Tal, N. (2010) ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res* **38**, W529-533
27. Talavera, G., and Castresana, J. (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic biology* **56**, 564-577
28. Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology* **59**, 307-321
29. Loytynoja, A. (2014) Phylogeny-aware alignment with PRANK. *Methods in molecular biology* **1079**, 155-170
30. D.J., Z. (2006) *Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion.* Ph.D., The University of Texas at Austin
31. Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2011) ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* **27**, 1164-1165
32. Ronquist, F., and Huelsenbeck, J. P. (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572-1574
33. Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., and Drummond, A. (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647-1649
34. Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature methods* **9**, 772

35. Krissinel, E., and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* **60**, 2256-2268
36. Ochiai, A., Itoh, T., Mikami, B., Hashimoto, W., and Murata, K. (2009) Structural determinants responsible for substrate recognition and mode of action in family 11 polysaccharide lyases. *J Biol Chem* **284**, 10181-10189
37. Davies, G. J., Wilson, K. S., and Henrissat, B. (1997) Nomenclature for sugar-binding subsites in glycosyl hydrolases. *Biochem J* **321 (Pt 2)**, 557-559
38. Hutter, C. A., Lehner, R., Wirth, C., Condemine, G., Peneff, C., and Schirmer, T. (2014) Structure of the oligogalacturonate-specific KdgM porin. *Acta crystallographica. Section D, Biological crystallography* **70**, 1770-1778
39. Blot, N., Berrier, C., Hugouvieux-Cotte-Pattat, N., Ghazi, A., and Condemine, G. (2002) The oligogalacturonate-specific porin KdgM of Erwinia chrysanthemi belongs to a new porin family. *The Journal of biological chemistry* **277**, 7936-7944
40. Boraston, A. B., and Abbott, D. W. (2012) Structure of a pectin methylesterase from Yersinia enterocolitica. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **68**, 129-133
41. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic acids research* **42**, D490-495
42. Abbott, D. W., and van Bueren, A. L. (2014) Using structure to inform carbohydrate binding module function. *Current opinion in structural biology* **28**, 32-40
43. Terrapon, N., Lombard, V., Gilbert, H. J., and Henrissat, B. (2014) Automatic prediction of polysaccharide utilization loci in Bacteroidetes species. *Bioinformatics*
44. Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., and Xu, Y. (2012) dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic acids research* **40**, W445-451

FIGURE LEGENDS

FIGURE 1. Function and phylogeny of PL2s. *A*, β-elimination reaction coordinate resulting in a 4,5-unsaturated product. The dashed semi-circles indicate subsites in the positive (towards the reducing end) and negative (towards the non-reducing end) of the scissile bond. *B*, Schematic of exolytic and endolytic modes of activity on HG. Exolytic enzymes digest HG strictly from the terminus of the polysaccharide and release a single defined product. Endolytic enzymes cleave internal glycosidic linkages to generate a mixed product profile. *C*, Representative unrooted tree highlighting the phylogeny of subfamily 1 and 2 members discussed in this study. Boundaries identified by CAZy (www.cazy.org) are indicated with dashed circles. The associated modes of activity are represented with endolytic and exolytic models. The presumed transitional sequence space between these activities is shown with a black triangle. Abbreviations: *Ye (Yersinia enterocoliticus), Dd (Dickeya dadantii), Pa (Pectobacterium atrosepticum)* and *Pae (Paenibacillus sp.).*

FIGURE 2. Characterization of HG modification by VvPL2. *A*, The *V. vulnificus* HG utilization locus is displayed as a schematic with representative gene sizes shown to scale. Genes that have been classified to assigned CAZy families (CBM = CBM32) and genes predicted in be involved in transport (KdgM, SBP) are labeled. Gene IDs are displayed below. *B*, Product profile of HG digestion by VvPL2 over time. A saturated marker of GalA-$GalA_3$ is shown on the left and unsaturated products (d.p. 2-5) are displayed on the right of the figure. *C*, pH profile of VvPL2. The temperature optima of *D*, VvPL2, *E*, YePL2A, and *F*, YePL2B. Representative kinetic plots for YePL2A *G*, YePL2B *H*, and VvPL2 *I*, on $GalA_3$. Product formation represents the detection of ΔGalA by UV absorbance at 232 nm. *J*, Metal exchange assays for YePL2A, YePL2B, and VvPL2 using the dialysis method and quantified using initial velocities with 1 mg mL^{-1} HG and equivalent enzyme concentrations. The two x-axes scales represent product values for YePL2B and VvPL2 (left) and YePL2A (right). Michaelis-Menten plots for YePL2A *K*, and YePL2B *L*, in the presence of various catalytic metals using the dialysis method.

FIGURE 3. Three-dimensional structure of VvPL2. *A*, Cartoon model of VvPL2 color ramped blue (N-terminus) to red (C-terminus), and with its catalytic metal modeled as a Mn^{2+} shown as a purple sphere. *B*, Superimposition of GalA from the +1 site of YePL2A complex (2v8k) within the active centre of VvPL2. The backbone of VvPL2 is shown as a grey cartoon with the metal binding residues displayed as grey sticks, ordered waters as red spheres, Mn^{2+} as a purple sphere, and the stabilizing residue (R304) and Brønstead base (R191) as cyan sticks. The distances between the 2-OH and 3-OH of GalA and R304; C5 and R191; and the Mn^{2+} ion and the uronate group oxygens are labeled and shown as red dashes. *C*, The metal binding pocket of VvPL2 with N-terminal histag. The map of the active centre residues coordinating the transition metal are presented as maximum likelihood/σ_A weighted $2F_{obs}$-F_{calc} densities, contoured at 1.0 σ and carved at 1.5 Å. The coordinated Mn^{2+} and ordered water are displayed as silver and red spheres respectively. The presence of two tartrate molecules within the N-terminal histag complex are rendered as yellow ball and stick models. *D*, Alignment of the YePL2A (2v8j; gray) and VvPL2 (green) metal coordination pocket. Residues are modeled as sticks, Mn^{2+} as purple spheres, and waters as red spheres. Residues are labeled using VvPL2/YePL2A numbering. Bond distances are indicated with yellow dashed lines. Consurf mapping displaying the conserved and divergent surface features of VvPL2 with subfamily 1 *E*, and subfamily 2 *F*, members. The highly conserved residues (K152, R191, R304, and metal pocket) and location of the lysine-tryptophan site (K300W) are labeled.

FIGURE 4. Ancestral sequence reconstruction of the PL2 family. *A*, Phylogeny of the PL2 family with calculated ancestor sequences nodes are highlighted (•) and labeled. The node sequences (49, 52, 54, and 74) that were targeted for gene synthesis and biochemical characterization are indicated with a black triangle. Previously characterized enzyme activities are boxed, activities reported in this study are noted with a single asterisk (see supplemental Table 1 for references), and family members with solved three-dimensional structures are designated with a double asterisk (1). *B*, Primary structure alignment of YePL2A, YePL2B, VvPL2, Node 52, Node 54, and Node 74. Residues involved in catalysis and metal coordination are indicated with a black and white triangle, respectively. The putative lysine to tryptophan switch is highlighted with a black circle.

FIGURE 5. Product profiling and biochemical characterization of ancestral PL2 sequences. *A*, HAEPC-PAD analysis of HG digestion profiles generated by Node 74, Node 52, and Node 54 ancestral enzymes. PAD detection (left axis) of oligogalacturonide elution is displayed as a black trace. The presence of ΔGalA oligosaccharides with a d.p. >2 are indicated with black triangles. *B*, Model for nucleotide progression of lysine-arginine-tryptophan stepwise mutation. *C*, The metal cofactor analysis of Node 74 relative activity with the inset showing the product generation in the presence of Mn^{2+} and Mg^{2+}. *D*, Putative subsite structure of PL2 endolytic (top) and exolytic (bottom) active sites displaying the molecular basis of $\Delta GalA_2$ generation by introduction of W00. R = reducing end and N = non-reducing end. Brackets indicate terminal residues within the active site that can be extended. HPLC-PAD analysis of YePL2B digestions of HG with W300 substitution with lysine and alanine in the absence *E*, and presence *F*, of EDTA. The large peak after 14 minutes represents the primary $\Delta GalA_2$ products and the black arrows indicate the appearance of $\Delta GalA_3$ and $\Delta GalA_4$ products.

FIGURE 6. Alternative pathways for HG utilization by human enteric pathogens that colonize distinct ecosystems. Key differences between the pathways from *Y. enterocolitica* and *V. vulnificus* include: *Extracellular*: secreted pectin methylesterase (YeCE8) in *Y. enterocolitica* and pectate lyase (VvPL9) in *V. vulnificus*; *Periplasmic*: *Y. enterocolitica* possesses two depolymerases (endoYePL2A and exoYeGH28) and *V. vulnificus* only a single enzyme (endoVvPL2); *Intracellular Transport:* appears to be mediated through unrelated systems; and *Cytoplasmic*: *Y. enterocolitica* possesses two exolytic enzymes (exoYePL2B and exoYePL22) whereas *V. vulnificus* only one (exoVvPL22). The key conserved features include the KdgM-anionic porin and the HG binding protein endoCBM32 that is proposed to retain polymerized substrates within the periplasm (1). In Enterobacteriaceae this cluster exists as a KdgM-endoPL2-CBM32 operon, which is not architecturally conserved in *V. vulnificus* (Fig. 2A; (18)). Enzyme activities that have been biochemically characterized are underlined. Proteins that belong to annotated subfamilies are indicated with [#] and those families that have not been assigned with [x].

Table 1: Catalytic properties of characterized PL2s and PL22s

Enzyme	SF	pH$^{\Delta}$	Activity	Location	Metal*	Reference
DdPL2B (PelW)	2	ND	*Exo*	Cytoplasm	Co^{2+}, Mn^{2+}, Ni^{2+}	(6), †
PaePL2	N/A	7.4	*Endo*	Cytoplasm	Mg^{2+}	(5)
PaPL2B	2	ND	*Exo*	Cytoplasm	ND	†
VvPL2	2	9.3	*Endo*	Secreted	Mg^{2+}	†
YePL2A	1	9.6	*Endo*	Secreted	Mg^{2+}	(1),†
YePL2B	2	8.6	*Exo*	Cytoplasm	Mn^{2+}	(1),†
DdPL22 (Ogl)	1	ND	*Exo*	Cytoplasm	Mn^{2+}	(6)
YePL22	1	7.6	*Exo*	Cytoplasm	Mn^{2+}	(2)

SF = subfamily. ND = not determined.

$^{\Delta}$pH refers to experimentally determined pH optimum

*Represents the preferential metal that provides maximal activity in recovery assays.

†This study. These assays were done with the exhaustive dialysis technique as opposed to the depletion - supplementation approach.

Table 2: **X-ray data collection, processing and VvPL2 model refinement statistics.**

Data collection statistics	**Tartrate-bound**
Wavelength	0.97949
Beamline	SSRL 12-2
Space group	$P6_5$
Resolution	46-1.90 (1.95-1.90)
Cell dimension	141.1, 141.1, 72.4
α, β, γ (Å)	90.0, 90.0, 120.0
R_{merge}	0.096 (0.419)
Completeness (%)	99.8 (100)
$<I/\sigma I>$	29.1 (4.6)
Redundancy	4.8 (5.0)
Total reflections	311612
Unique reflections	64421
Refinement statistics	
R_{work} (%), R_{free} (%)	15.6, 19.5
RMSD	
Bond lengths (Å)	0.010
Bond angles (^{o})	1.909
Average *B*-factors (Å^2)	
Protein molecule	26.4
Transition metal	21.0
Solvent atoms	40.3
Number of atoms	
Protein atoms	4538
Transition metal	1
Solvent atoms	596
Ramachandran statistics*	
Most favored (%)	97.9 (550)
Additional allowed (%)	2.1 (12)
Disallowed (%)	0.0 (0)

*Ramachandran statistics were calculated by Rampage (25).

Table 3: Kinetics of VvPL2, YePL2A, YePL2B, and YePL2A-B on $GalA_3$ and HG

$GalA_3$	k_{cat} (min^{-1})	K_m (mM)	k_{cat}/K_m (min^{-1} / mM)
VvPL2	$1.28 \times 10^3 \pm 2.21 \times 10^1$	$1.37 \times 10^{-1} \pm 8.40 \times 10^{-3}$	$9.33 \times 10^3 \pm 5.94 \times 10^2$
YePL2A	$3.94 \times 10^2 \pm 1.4 \times 10^1$	$6.89 \times 10^{-1} \pm 9.00 \times 10^{-2}$	$5.71 \times 10^2 \pm 7.76 \times 10^1$
YePL2B	$1.58 \times 10^3 \pm 2.20 \times 10^1$	$2.61 \times 10^{-1} \pm 9.00 \times 10^{-2}$	$6.06 \times 10^3 \pm 2.66 \times 10^2$
YePL2A-B	$2.43 \times 10^2 \pm 2.17 \times 10^1$	$6.80 \times 10^{-1} \pm 1.40 \times 10^{-1}$	$3.56 \times 10^2 \pm 7.99 \times 10^1$
HG	**k_{cat} (min^{-1})**	**K_m (mg mL^{-1})**	**k_{cat}/K_m (min^{-1} / mg mL^{-1})**
VvPL2	$2.89 \times 10^2 \pm 1.20 \times 10^1$	$1.70 \times 10^{-1} \pm 2.50 \times 10^{-2}$	$1.70 \times 10^3 \pm 2.66 \times 10^2$
YePL2A	$3.83 \times 10^3 \pm 3.80 \times 10^2$	$6.00 \times 10^{-1} \pm 1.5 \times 10^{-1}$	$6.41 \times 10^3 \pm 1.80 \times 10^3$
YePL2B	$4.32 \times 10^2 \pm 5.80 \times 10^1$	$9.00 \times 10^{-1} \pm 2.70 \times 10^{-1}$	$4.81 \times 10^2 \pm 1.60 \times 10^2$
YePL2A-B	$8.90 \times 10^2 \pm 9.51 \times 10^1$	$7.20 \times 10^{-1} \pm 1.90 \times 10^{-1}$	$1.24 \times 10^3 \pm 3.56 \times 10^2$

Errors (±) represent standard deviations from three independent reactions.

Table 4: Kinetics of YePL2A and YePL2B on HG in the presence of varying metal cofactors

YePL2A	k_{cat} (min^{-1})	K_m (mg mL^{-1})	k_{cat}/K_m (min^{-1} / mg mL^{-1})
Mg^{2+}	$3.00 \times 10^3 \pm 1.20 \times 10^2$	$4.70 \times 10^{-1} \pm 5.60 \times 10^{-2}$	$6.39 \times 10^3 \pm 8.00 \times 10^2$
Mn^{2+}	$1.89 \times 10^3 \pm 4.10 \times 10^1$	$7.80 \times 10^{-2} \pm 8.90 \times 10^{-3}$	$2.43 \times 10^4 \pm 2.80 \times 10^3$
Ca^{2+}	$2.04 \times 10^3 \pm 6.40 \times 10^1$	$5.90 \times 10^{-1} \pm 5.00 \times 10^{-2}$	$3.44 \times 10^3 \pm 3.10 \times 10^2$
Co^{2+}	$3.15 \times 10^3 \pm 8.50 \times 10^1$	$6.80 \times 10^{-1} \pm 4.70 \times 10^{-2}$	$4.60 \times 10^3 \pm 3.44 \times 10^2$
YePL2B	**k_{cat} (min^{-1})**	**K_m (mg mL^{-1})**	**k_{cat}/K_m (min^{-1} / mg mL^{-1})**
Mg^{2+}	$4.58 \times 10^2 \pm 2.10 \times 10^1$	$1.84 \times 10^{-1} \pm 4.20 \times 10^{-2}$	$2.49 \times 10^3 \pm 5.70 \times 10^2$
Mn^{2+}	$2.44 \times 10^3 \pm 1.30 \times 10^2$	$1.26 \times 10^{-1} \pm 3.80 \times 10^{-2}$	$1.93 \times 10^4 \pm 5.90 \times 10^3$
Mn^{2+*}	$6.90 \times 10^2 \pm 1.47 \times 10^2$	$3.35 \times 10^{-1} \pm 1.33 \times 10^{-2}$	$2.06 \times 10^3 \pm 9.28 \times 10^2$
Ca^{2+}	$5.68 \times 10^2 \pm 4.40 \times 10^1$	$8.61 \times 10^{-1} \pm 1.70 \times 10^{-1}$	$6.60 \times 10^2 \pm 1.40 \times 10^2$
Co^{2+}	$5.24 \times 10^1 \pm 2.30 \times 10^0$	$1.55 \times 10^{-1} \pm 3.60 \times 10^{-2}$	$3.39 \times 10^2 \pm 8.00 \times 10^1$

*Values were calculated to account for substrate inhibition effects. Mn^{2+*}= substrate inhibition kinetics; K_i = 2.25 ± 1.07 mg mL^{-1}. Errors (±) represent standard deviations from three independent reactions.

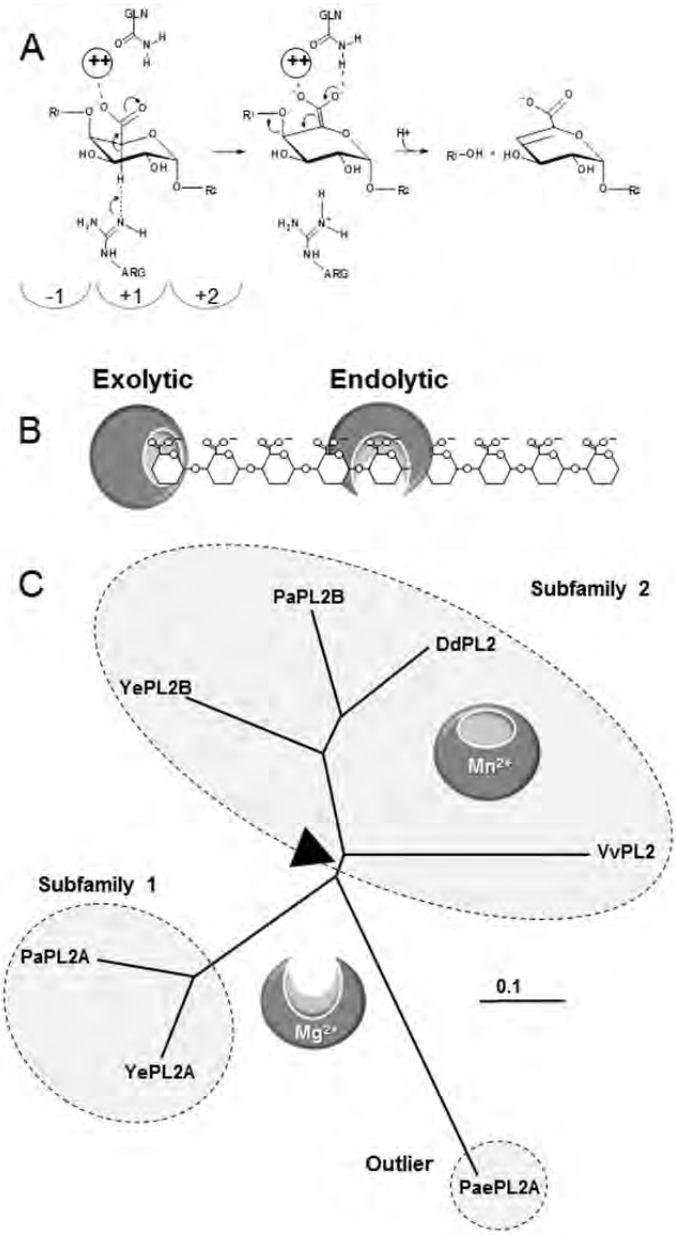


Figure 1: McLean et al (2015)

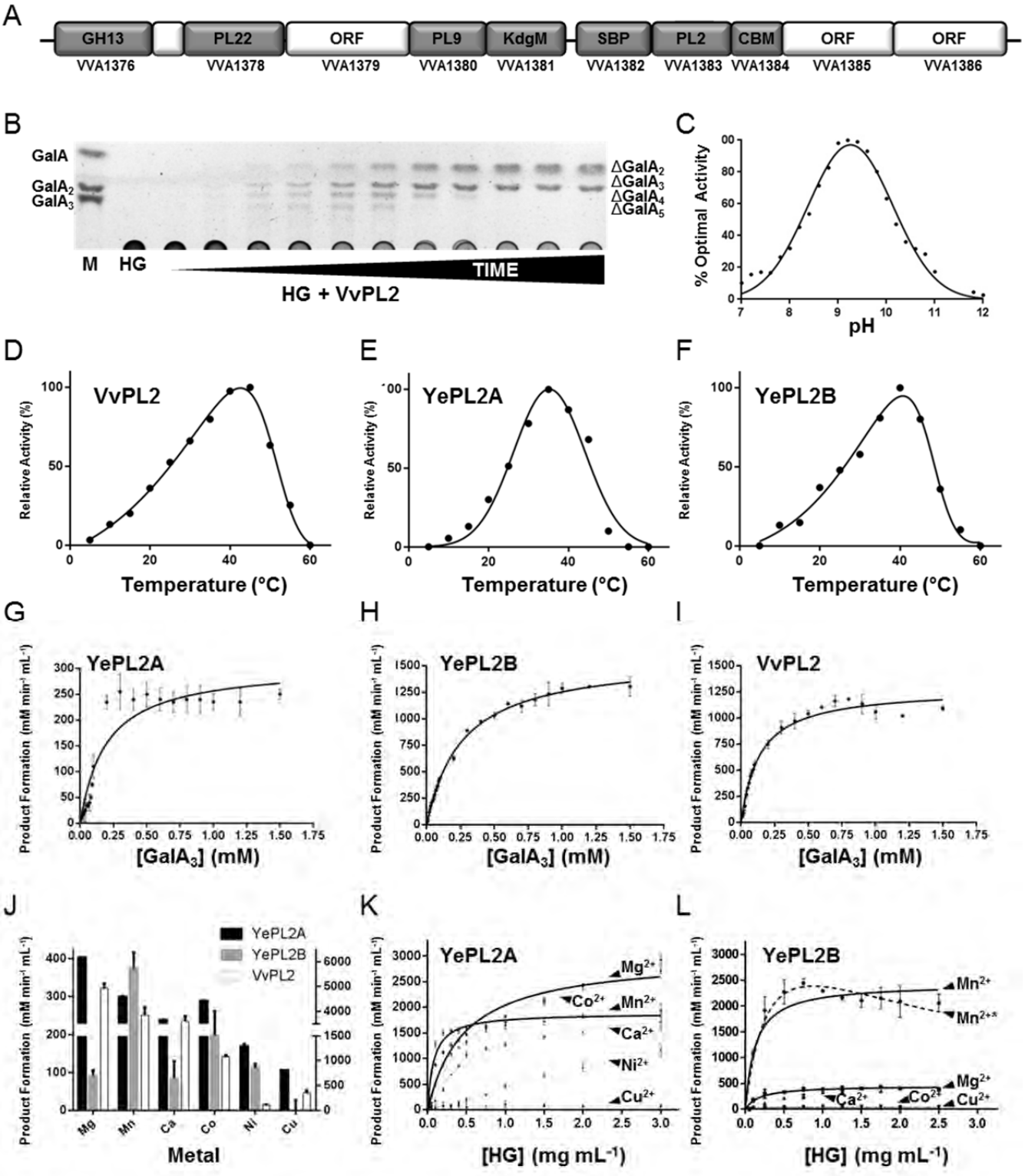


Figure 2: McLean et al (2015)

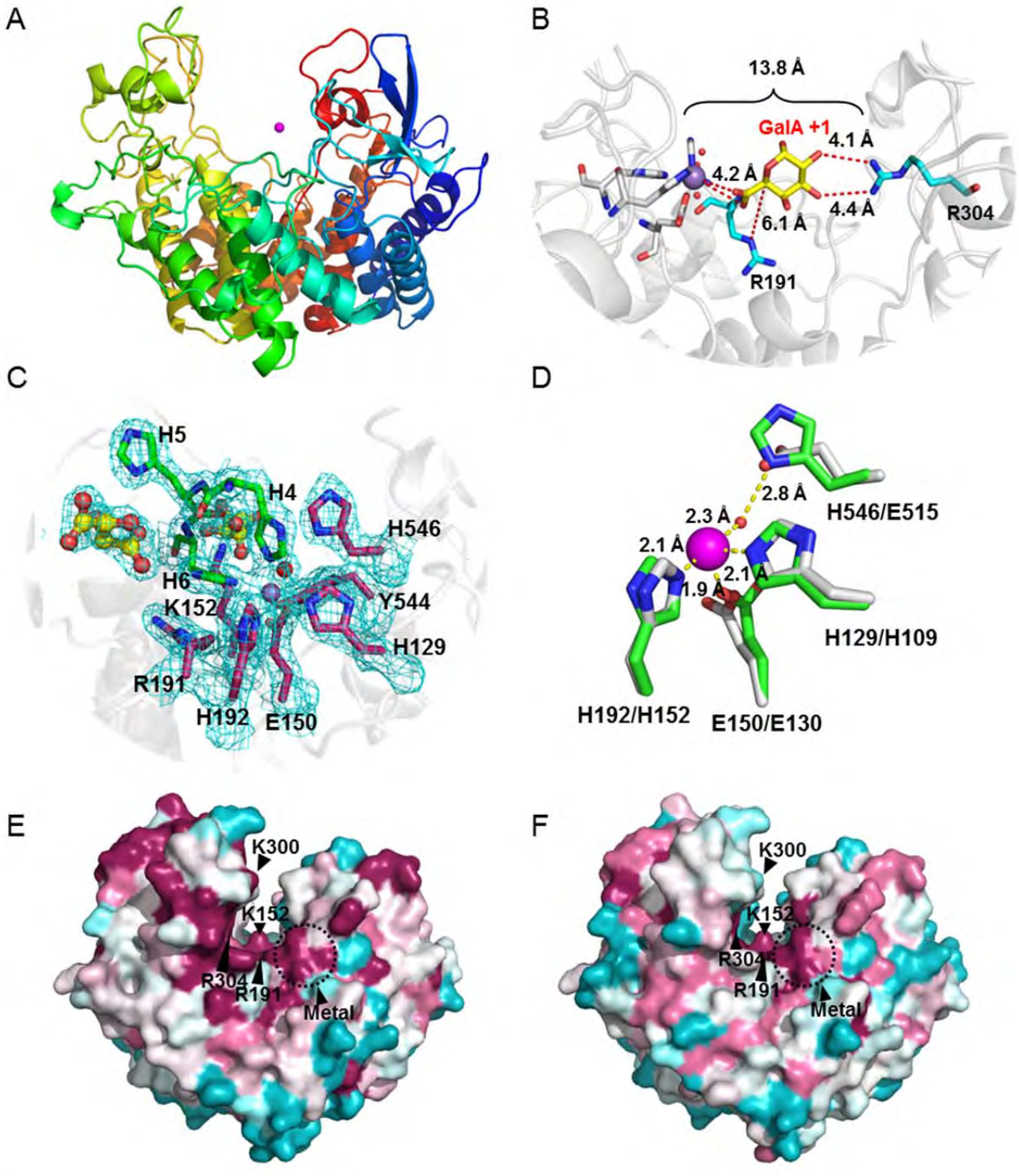


Figure 3: McLean et al (2015)

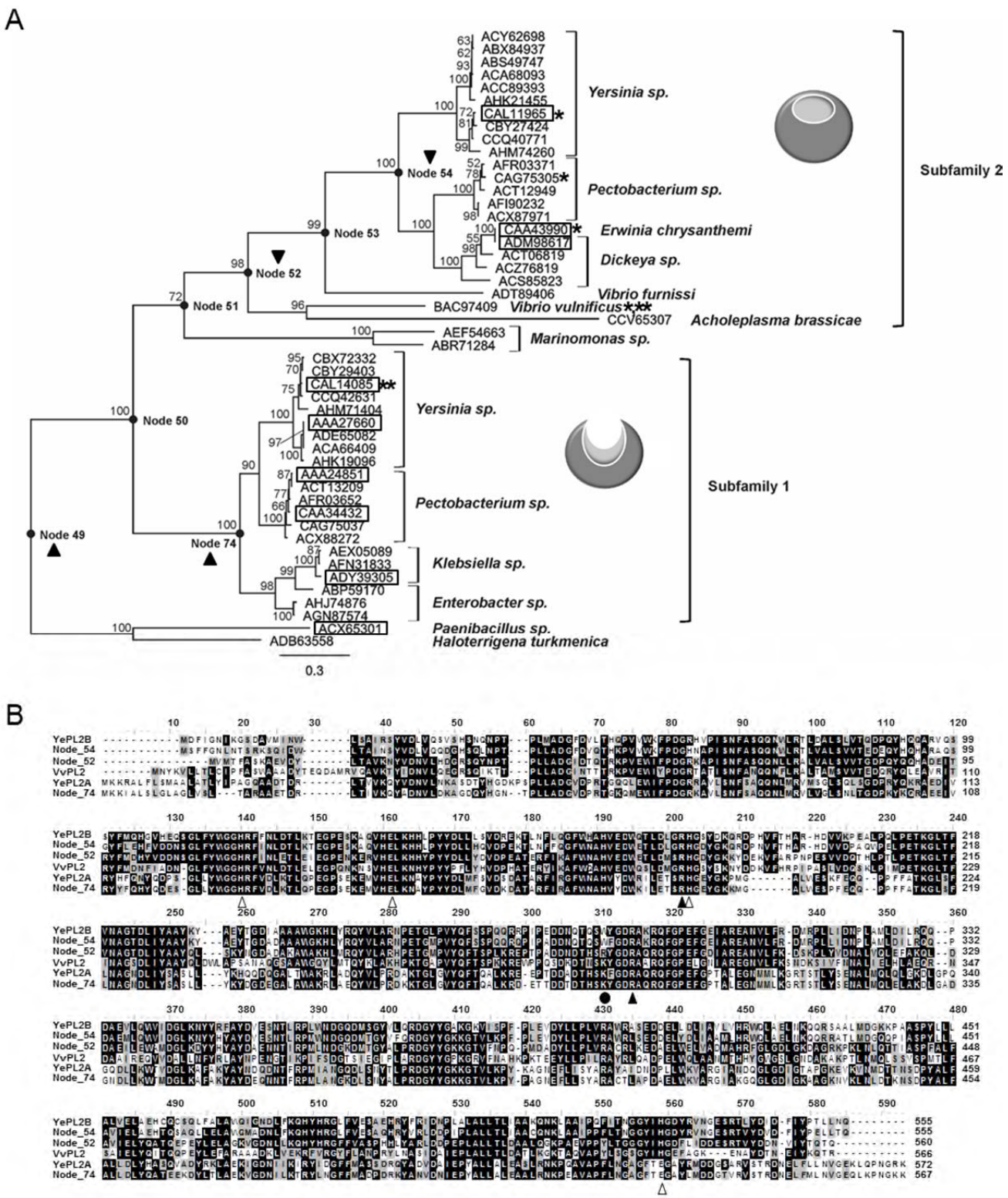


Figure 4: McLean et al (2015)

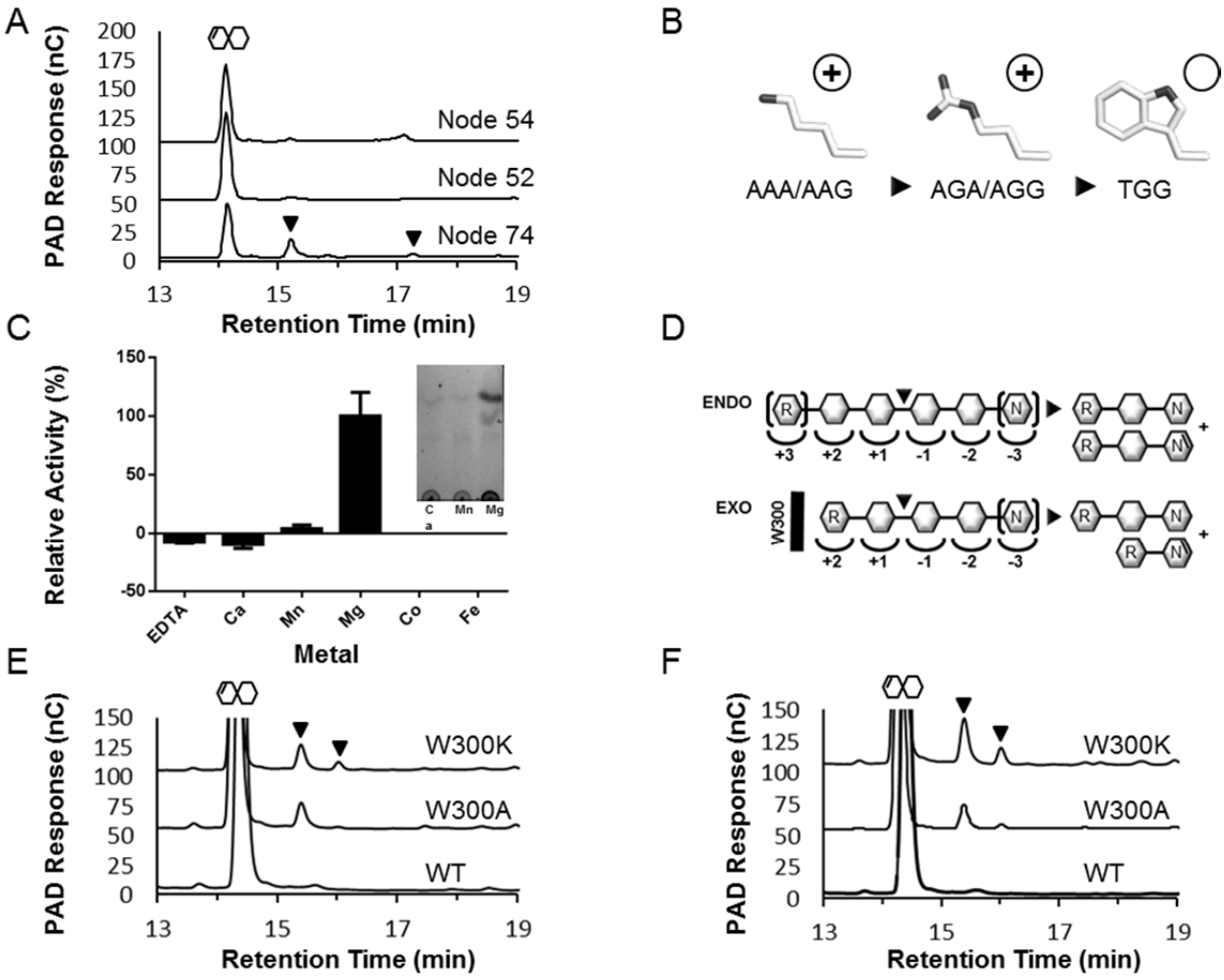


Figure 5: McLean et al (2015)

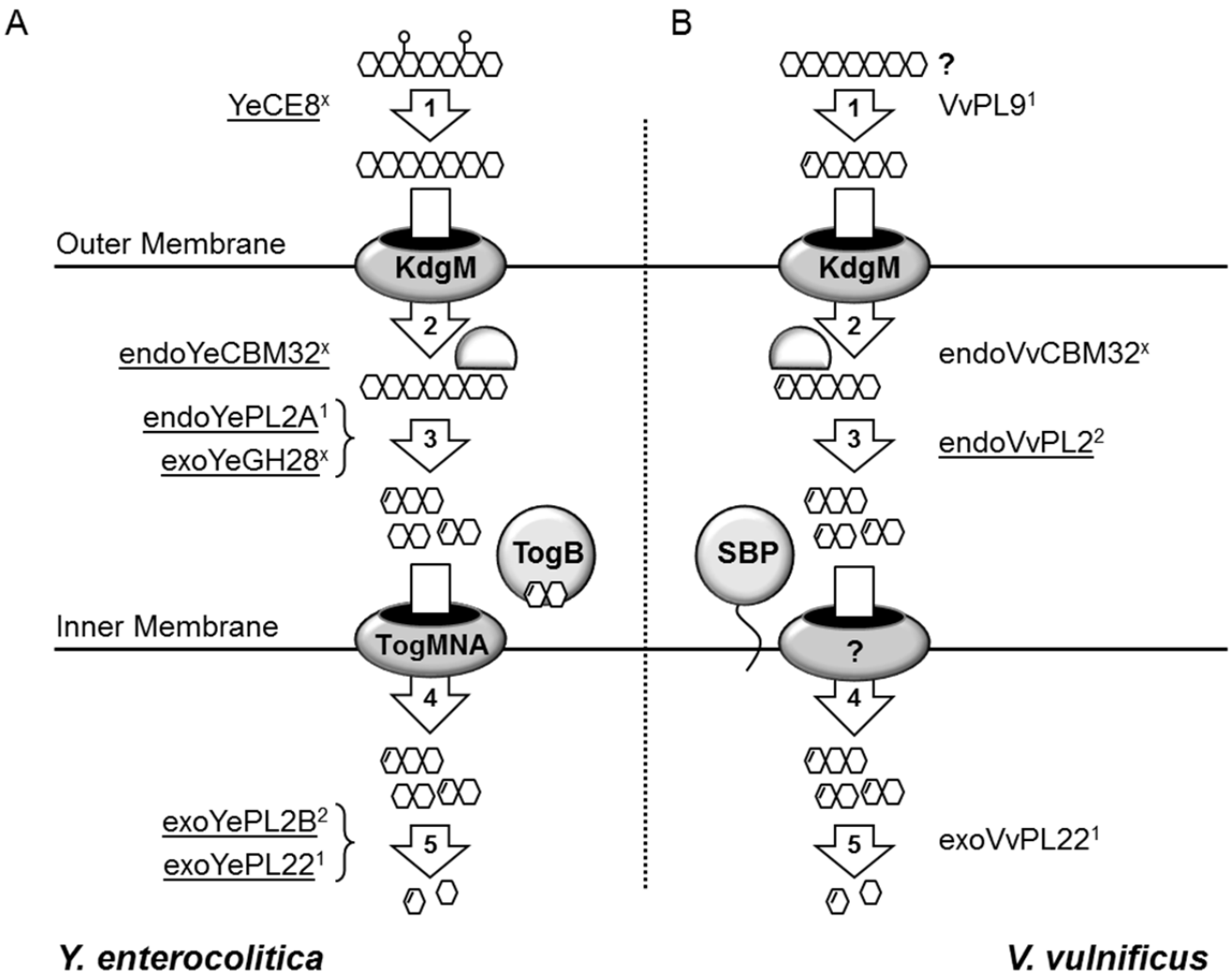
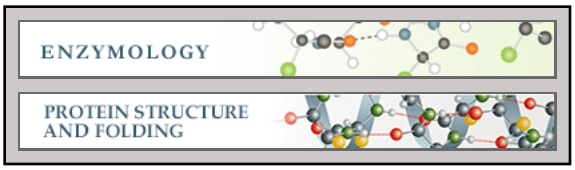


Figure 6: McLean et al (2015)

Enzymology:

Functional analyses of resurrected and contemporary enzymes illuminate an evolutionary path for the emergence of exolysis in polysaccharide lyase family two



Richard McLean, Joanne K. Hobbs, Michael D. Suits, Sami T. Tuomivaara, Darryl Jones, Alisdair B. Boraston and D. Wade Abbott

J. Biol. Chem. published online July 9, 2015