House dust mites possess a polymorphic, single domain putative peptidoglycan D,L endopeptidase belonging to the NlpC/P60 Superfamily

Vivian H. Tang, Geoffrey A. Stewart, Barbara J. Chang *

Microbiology & Immunology, School of Pathology and Laboratory Medicine, The University of Western Australia, Crawley, WA, Australia

A B S T R A C T

A 14 kDa protein homologous to the γ-D-glutamyl-L-diamino acid endopeptidase members of the NlpC/P60 Superfamily has been described in *Dermatophagoides pteronyssinus* and *Dermatophagoides farinae* but it is not clear whether other species produce homologues. Bioinformatics revealed homologous genes in other Sarcopteformes mite species (*Psoroptes ovis* and *Blomia tropicalis*) but not in *Tetranychus urticae* and *Metaseiulus occidentalis*. The degrees of identity (similarity) between the *D. pteronyssinus* mature protein and those from *D. farinae*, *P. ovis* and *B. tropicalis* were 82% (96%), 77% (93%) and 61% (82%), respectively. Phylogenetic studies showed the mite proteins were monophyletic and shared a common ancestor with both actinomycetes and ascomycetes. The gene encoding the *D. pteronyssinus* protein was polymorphic and intronless in contrast to that reported for *D. farinae*. Homology studies suggest that the mite, ascomycete and actinomycete proteins are involved in the catalysis of stem peptide attached to peptidoglycan. The finding of a gene encoding a P60 family member in the *D. pteronyssinus* genome together with the presence of a bacterial promotor suggests an evolutionary link to one or more prokaryotic endosymbionts.

© 2015 The Authors. Published by Elsevier B.V. on behalf of the Federation of European Biochemical Societies. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Mites secrete a number of enzymes in their fecal pellets including proteases, carbohydrases and bacteriolytic enzymes, which reflect their digestive processes. With regard to the latter, we have characterised a 14 kDa protein [1] from the house dust mite (HDM) *Dermatophagoides pteronyssinus* which originally showed significant sequence homology with a variety of bacterial proteins such as the P60 proteins from *Listeria monocytogenes* [1,2] and the invasion proteins from mycobacterial species [1]. Such proteins were subsequently shown to be cysteine peptidases, with a catalytic triad similar to that seen in papain [3] (MEROPS Clan CO, Family C40) [4] and this activity is associated with the NlpC/P60 Pfam domain.

Proteins in this family cleave between the second and third residues of the peptidoglycan stem peptide to release the peptide L-Ala-γ-D-Glu and diaminopimelic acid (DAP) or Lys [5]. In bacteria, they are autolysins that may be cytoplasmic or secreted, and they play an important role in septum formation since gene deletion results in long, connected strings of bacteria [6–8]. However, they may also play a role in eukaryotic cell invasion and, in humans, the Listerial P60 protein induces both B and T cell responses, which help protect against infection [9]. Proteins possessing the NlpC/P60 domain may be single or multidomain proteins and, in the latter, other domains may play a role in targeting the enzyme to bacterial cell walls [10–13], or providing them with additional peptidoglycan cleaving ability [14]. The majority of proteins possessing the NlpC/P60 domain described in the literature are prokaryotic in origin, with only a relatively small number being described thus far in eukaryotes.

The mite 14 kDa protein is a single domain protein and the gene encoding it and a variant have now been designated *lytFM* and *lytFM1*, respectively, based on the terminology used to describe the autolysins from *Bacillus subtilis* [15,16]. Whilst the role of NlpC/P60 family members in bacterial species in peptidoglycan processing is more clearly delineated, that of the proteins in mites and other eukaryotes is yet to be established. As they do not

possess peptidoglycan, it is reasonable to assume that possession of NlpC/P60 proteins and other peptidoglycan-degrading proteins such as lysozyme and amidases are likely to enable eukaryotes to protect themselves against pathogenic bacteria, or to utilise bacteria as a food source or both.

Mites represent one the most ubiquitous groups of invertebrates, and they can be found in a diverse array of ecological niches. However, it is not clear whether all species of mites possess the NlpC/P60 protein described in the genus Dermatophagoides. In addition, it is possible that the gene encoding the mite NlpC/P60 protein with its relatively high sequence homology to bacterial proteins, entered into the mite genome from mite-associated microorganisms. In this regard, both Gram-positive and Gram-negative bacterial species have been demonstrated in mites by both conventional culture techniques and microbiome studies [17–19], that are capable of producing P60 family members, for example, *B. subtilis*, *Bacillus licheniformis* and *Bacillus cereus*. Whether *D. pteronyssinus* mites acquired the *lytFM* gene via horizontal gene transfer either as a single domain protein or one encoding a multi-domain protein, which then lost genetic material encoding non-NlpC/P60 domains, is unclear.

In this report, we have searched for *lytFM* homologues in other mite species using a bioinformatics approach. Similarly, we looked at flanking sequences of *lytFM* by PCR to determine whether there were remnants of a bacterial heritage, and used gDNA libraries to look for introns. Finally, we have modelled the mite P60 family members to gain information about the possible preferred peptidoglycan substrate. The data obtained indicate that the LytFM homologue is present in some Pyroglyphidae, Psoroptidae and *Echimyopodidae* families in the Acariformes but not yet in Parasitiformes mites.

2. Materials and methods

All PCR primers were obtained from Invitrogen (Mulgrave, Victoria, Australia). The *D. pteronyssinus* gDNA library was kindly provided by Dr. Wendy-Anne Smith (Telethon Kids Institute, Perth, Western Australia), respectively. PCR reagents used were HotStarTaq Master Mix (QIAGEN, Hilden, Germany), KAPA HiFi™ PCR Kit (KAPA Biosystems, Boston, USA), Taq buffer (10×, containing 15 mM MgCl$_2$) (Perkin Elmer, New Jersey, USA), the four 2′-deoxynucleotide 5′-triphosphates (Promega, Wisconsin, USA) and Taq DNA polymerase (Ampli Taq Gold) (Applied Biosystems, Victoria, Australia).

2.1. D. pteronyssinus gDNA library and Dermatophagoides farinae cDNA library

Briefly, the library was prepared using purified *D. pteronyssinus* genomic DNA, digested with *Sau* 3A I 7–20 kb fragments and packaged into *Xho* I. The library was propagated and the phage adsorbed to *Escherichia coli* ER1647 cells at a ratio of 4×10^5 pfu/ml host cells at 37 °C for 30 min. Plaques were allowed to reach semi-confluence (5–6 h at 37 °C) and the phage eluted with SM buffer. The gDNA library was quantified as described previously [20] before being used in PCR.

2.2. PCR primers

The annealing sites of all the primers used in PCR screening of the *D. pteronyssinus* gDNA library are shown in Fig. 1. The primers included those used previously to amplify *lytFM* and *lytFM1* from the *D. pteronyssinus* cDNA library [1], i.e., GSUTR1 (5′-CTAT TATGAAATTCTTCTTCACT-3′), GS2 (5′-CAAATTGGTGTTCCATATT CATGG-3′) and GSR3 (5′-TTACCAACATCGTGCAACATTAGC-3′), primers designed to anneal to the right and left arms of the λBlueS-TAR™ vector VP1 (5′-GCATACATTATACGAAGTTATGCG-3′) and VP2 (5′-GAAAAATAAACAAATAGGGGTTCCG-3′), and the 5′ and 3′ ends of *lytFM* GSP1 (5′-TGTACAGAATAAAGCTAAAGTGAAG-3′), GSP2 (5′-ATATGGAGTGATCATATTGCTAATG-3′), GSP2R (5′-TACATGCTCTTC TTTTATATACCTC-3′) and GSP1F (5′-GTTACATGTCTTATTTCGATTT CAC-3′) and FSPR (5′-TGAATATTTTTTGAACATCGTGCTG-3′) which
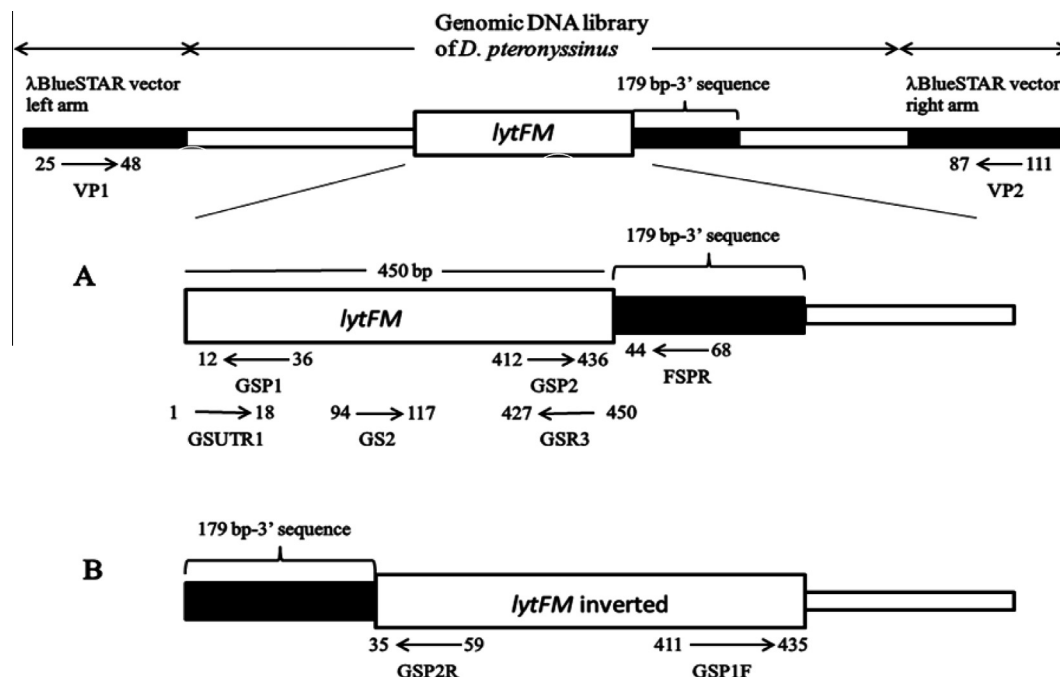
Fig. 1. Schematic diagram showing the annealing sites for all the primers used in the PCR screening of the λBlueSTAR™ gDNA library of *D. pteronyssinus*. The arrows and numbers above the primers indicate the directions of amplification and the nucleotide positions, respectively, spanned by the primers within the λBlueSTAR vector, *lytFM* and 179 bp-3′ sequence. Since the *Sau3A* I-digested genomic DNA fragments of *D. pteronyssinus* could be ligated to the *Xho* I-digested vector arms in either direction, the *lytFM* gene could exist in orientation 1 (A) or inverted (B).

was designed to anneal to a 179 bp-3′ sequence downstream of *lytFM* (Fig. 1).

2.3. PCR screening of the D. pteronyssinus gDNA library

PCR to amplify *lytFM* from the gDNA library was performed with the HotStarTaq Master Mix kit under the conditions: 95 °C for 15 min, 94 °C for 1 min, 37 °C for 45 s, 68 °C for 1 min and 72 °C for 10 min for 35 cycles. PCR screening of the gDNA library to elucidate the sequences flanking the 5′ and 3′ ends of *lytFM* was performed using the HotStarTaq Master Mix kit at the annealing gradient temperatures of 29.6, 29.9, 31.2, 33.4, 36.6, 38.6, 39.8 and 40.1 °C (29–40 °C) or 40, 41.4, 43.6, 46.9, 51.4, 54.6, 56.8 and 58 °C (40–58 °C). For PCR performed using the KAPA HiFi™ PCR Kit, the conditions applied were 95 °C for 5 min, 98 °C for 20 s, 29–40 °C or 40–58 °C for 45 s, 72 °C for 2.5 min and 72 °C for 5 min for 30 cycles. For PCR performed using the conventional reagents, the samples prepared contained 1× Taq buffer (containing 1.5 mM MgCl₂), 0–5.5 mM supplementary MgCl₂, 0.1 μM forward and reverse primers, 200 μM of each of the four dNTPs, 2.5 units of AmpliTaq Gold and 100 ng of template to a total volume of 25 μl, and the PCR conditions applied were 95 °C for 15 min, 94 °C for 1 min, 37 °C for 45 s, 68 °C for 1 min and 72 °C for 10 min for 30 cycles. All amplifications were performed in a Bio-RAD iCycler IQ PCR Thermal Cycler (96 wells) (Bio-RAD, Hercules, California, USA).

2.4. Preparation of PCR products for DNA sequencing and analysis of sequencing data

DNA sequencing was performed by the Lotterywest State Biomedical Facility Genomics at the Royal Perth Hospital, Perth, Australia. The sequencing data were compared with sequences in the NCBI nucleotide database or the European Nucleotide Archive using BLASTN or translated into amino acid sequences using the ExPaSy translation tool (University of Geneva, Switzerland) and the sequences compared with those in the protein and translated nucleotide sequence databases of NCBI or EBI using BLASTP and TBLASTN, respectively.

2.5. Bioinformatic analyses

Various searches of the available databases were undertaken to determine whether the presence of the *lytFM* sequence had been previously reported in other mite species. In addition, appropriate nucleotide sequences were analysed for the presence of prokaryotic promoters using the Neural Network Promoter Prediction (NNPP) program at the Berkeley Drosophila Genome Project (BDGP) (www.fruitfly.org/seq_tools/promoter.html) and the program Softberry BPROM (Softberry, Inc., Mount Kisco, New York). The latter program predicts bacterial sigma 70 promoter sequences with about 80% accuracy and specificity (http://linux1.softberry.com/

berry.phtml). The likely presence of leader sequences was determined using the eukaryote non-TM data set and SignalP 4.1 [http://www.cbs.dtu.dk/services/SignalP/], and the theoretical isoelectric points (pI) determined using the ProtParam tool at ExPaSy (www.expasy.org). PHYRE 2.0 was used to model the structure of the mite protein [21].

Phylogenetic analysis of the mite proteins, and bacterial and eukaryotic homologues were performed using the Phylogeny.fr pipeline [22], with multiple sequence alignment performed using MUSCLE [23] after using the MaxAlign software (www.cbs.dtu.dk) and manual curation [24]. Sequences possessing the invariant Tyr and the Cys, His and a polar residue comprising the catalytic triad [5] were included. The approximate likelihood ratio test was used for branches instead of bootstrapping as described [25].

3. Results

3.1. The lytFM gene in D. pteronyssinus is polymorphic

A search was performed to ascertain whether *lytFM* or its variant, *lytFM1,* were curated in any of nucleotide databases. Three *D. pteronyssinus* cDNA clones (ESTs, GenBank: EX163010, EX163553 and EX162735) were identified. In addition, one of these (GenBank: EX163010) was found to contain a 179 bp 3′ flanking sequence (data not shown). A number of nucleotide polymorphisms were detected in all 5 sequences (Table 1), most of which were silent. However, polymorphisms in the 3′ region of the gene result in amino acid substitutions giving rise to another variant (LytFM2) in addition to the two originally described. The changes are predicted to alter the pI of the proteins; 8.57 for LytFM, and 8.92 for the remaining four sequences. A *D. pteronyssinus* gDNA library was then screened for the presence of *lytFM* variants with PCR conditions previously used for the successful amplification of the gene from the *D. pteronyssinus* cDNA library [1]. However, both primer pairs used, GSUTR1/GSR3 and GS2/GSR3 (Fig. 1) resulted in amplification of the *lytFM* isoform, but not the *lytFM1* or *lytFM2* isoforms (data not shown). In addition, the PCR product obtained was found to be intronless, as the sequence obtained was identical to that derived from the cDNA library (data not shown).

3.2. lytFM homologues in other mite species

A partial [26] and full-length sequence [19] of the *lytFM* homologue from *D. farinae* have recently been described. In this regard, an intron has been identified in the *lytFM* homologue of *D. farinae,* in contrast to the intronless gene reported here for *D. pteronyssinus.* The intron is between the sequence encoding the first 16 residues of the leader sequence and the sequence encoding the next four residues (Fig. 2) using the reported unplaced genomic scaffold (KN266412). A search of the available databases identified *lytFM* homologues in *P. ovis* (FR749374) and *B. tropicalis* (CB282085). An alignment of the deduced amino acid sequences for the four

Table 1
Polymorphisms found in the *lytFM* gene of *D. pteronyssinus.*

Sequence	[a]Nucleotide position (amino acid residue change)											
	12	69	72	105	138	165	396	414	415	417	418	420
LytFM	C	C	T	T	T	C	T(N)	A	T(W)	G	A(S)	T
LytFM1	C	C	T	T	T	C	G(K)	A	G(G)	A	G(G)	A
[b,c]EX163553	–	T	G	A	T	T	G(K)	C	T(W)	G	A(S)	T
[b,c]EX163010	T	C	T	A	A	C	G(K)	C	T(W)	G	A(S)	T
[b,c]EX162735	T	A	G	A	T	T	A(K)	A	T(W)	G	A(S)	T

[a] Nucleotide position based on LytFM sequence.
[b] Sequences obtained from the European Nucleotide Archive.
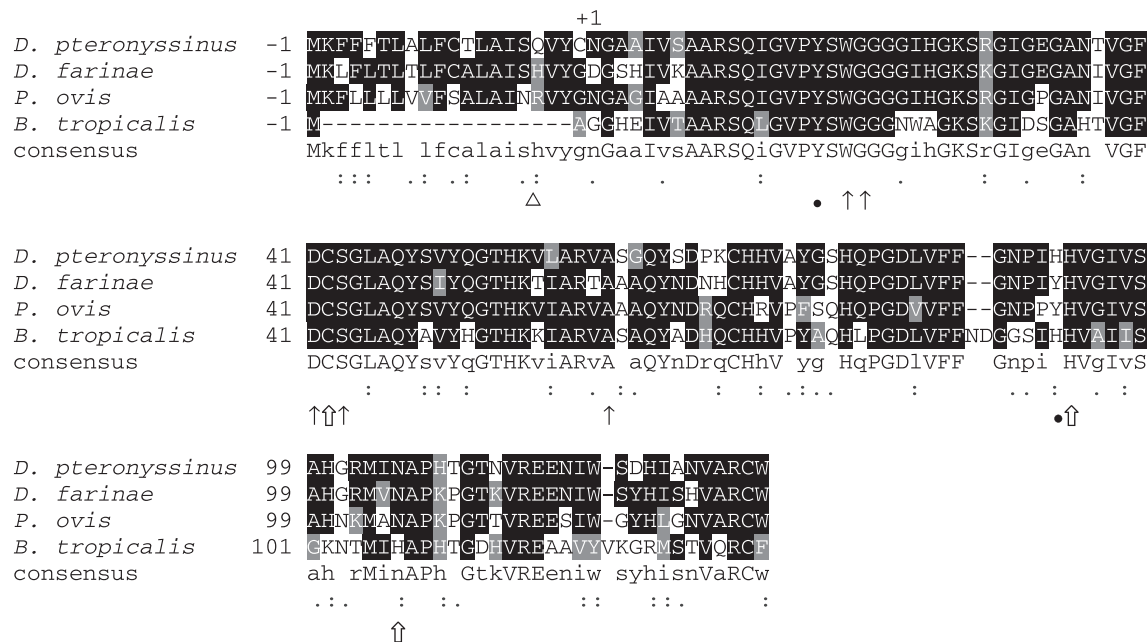[c] Protein with this polymorphism designated LytFM2.

Fig. 2. Deduced amino acid sequence alignment between LytFM and its *D. farinae*, *P. ovis* and *B. tropicalis* homologues. The DNA sequences of *lytFM* and its homologues from *D. farinae* (GenBank: KN266412), *P. ovis* (GenBank: FR749374) and *B. tropicalis* (GenBank: CB282085) were translated into amino acid sequences using the ExPaSy translation tool and the alignment was performed using MAFFT [34]. Conserved, consensus residues are highlighted in boxshades and chemically similar substitutions (:) and weakly similar substitutions (.) are marked as indicated. The predicted (SignalP 4.1) or demonstrated N-terminal residues are shown (+1) [1]. The open triangle indicates the exon–intron–exon boundary identified in the leader sequence of the *D. farinae* homologue [19]. The catalytic triad residues are indicated by open arrows and the residues comprising the S1 and S′ site are indicated by solid arrows and dots, respectively [5].

mite species *lytFM* homologues is shown in Fig. 2, where significant homology was detected across the mature proteins compared to LytFM for the mite species, *D. farinae* (82% identity, 96% similarity), *P. ovis* (77% identity, 93% similarity) and *B. tropicalis* (61% identity, 82% similarity). The percent GC contents of the nucleotide sequences were 38.9, 39.3, 38.5, 36.7, 41.9 and 58, for *lytFM*, *lytFM1*, *lytFM2*, *D. farinae*, *P. ovis* and *B. tropicalis* homologues, respectively. Two conservative polymorphisms were detected when the full-length [19] and partial [26] *D. farinae* sequences were compared, namely Lys for Arg at residue position 29 and Ile for Leu at position 45 (mature *D. pteronyssinus* numbering, Fig. 2). Signal peptide cleavage sites were predicted for *P. ovis*, *D. pteronyssinus* and *D. farinae* homologues (Fig. 2) but not for *B. tropicalis*. The cleavage positions predicted were between 20 and 21 for *P. ovis* and *D. farinae*, and between 16 and 17 for *D. pteronyssinus* using the eukaryotic data set. However, the cleavage position for *D. pteronyssinus* LytFM was previously determined to be between residues 20 and 21 (Fig. 2) using N-terminal sequencing (1), which coincided with that predicted using the Gram-negative data set. Using these data, the predicted pI's for the mature *D. farinae*, *P. ovis* and *B. tropicalis* LytFM homologues were 9.08, 9.55 and 8.86, respectively. A search of genomic data obtained for two other mite species, namely, the herbivorous mite *Tetranychus urticae* belonging to the Tetranychidae family and the western predatory mite *Metaseiulus occidentalis* (which feeds on spider mites) belonging to the Typhlodrominae subfamily was also performed. These studies showed that a LytFM homologue was absent in both.

3.3. Phylogenetic analysis of mite lytFM homologues

A BLAST search of the mite LytFM homologue against eukaryote and prokaryote sequences was performed. With regard to eukaryotes, a significant proportion of the available sequences were from the ascomycetes although homologues were detected in wasps, ants, trichomonads, protozoa and amoebae. Fig. 3 shows

the homology of the mite, eukaryotic and prokaryotic proteins. The sequences are similar although some differences were observed when the mite proteins were compared with both eukaryote and prokaryote sequences. For example, the position of the cysteine residues in mites is not mirrored in the other sequences, apart from the essential one involved in catalysis. Correspondingly, the other sequences can be differentiated from mites and other NlpC/P60 proteins based on the location of cysteine residues. For example, many of the ascomycetes possess cysteine residues at position 24 and 51 (mature *D. pteronyssinus* protein numbering) and in an insert between residues 88 and 89 in the *D. pteronyssinus* sequences that are not present in either the mite or actinomycete proteins. In this insert region, Cys may be present singly or as a Cys-Asp-Cys or a Cys-Ser-Cys motif. In contrast, the actinomycetes do not possess Cys other than the catalytic one characterising the NlpC/P60 domain. Fig. 4 shows the phylogenetic relationship between the homologues and it can be seen that the mite LytFM proteins form a monophyletic group. The mites and some ascomycetes and actinomycetes share a common ancestor.

3.4. The mite LytFM proteins are predicted to be peptidoglycan D,L endopeptidases

Members of the P60 family possess either γ-D-glutamyl-L-diamino acid or γ-D-glutamyl-L-lysyl endopeptidase activity [5]. Recently, the P60 protein from *B. cereus* (BcYkfC) has been crystallised with the bound reaction product, L-Ala-γ-D-Glu, thus establishing the residues involved in catalysis [5]. Fig. 2 shows that most of the residues comprising the substrate binding and catalytic site are present in the homologous mite proteins, indicating they are endopeptidases. However, some specific residues contributing to the S2 site of the binding cleft in the *B. cereus* protein are missing from the mite homologues; these being residues directly interacting with the free amino group of the L-Ala residue in the stem peptide. These include Asp_{256} (BcYkfC numbering), Glu_{83} and Tyr_{118},

```
D. pteronyssinus       1  -NCAAIVSAARSQIGVPYS---WGGGGIHGKSRGIGEGA--------NTVGFDCSGLAQY
D. farinae             1  -DCSHIVKAARSQIGVPYS---WGGGGIHGKSKGIGEGA--------NIVGFDCSGLAQY
P. ovis                1  -NCAGTAAAARSQIGVPYS---WGGGGIHGKSRGIGPGA--------NIVGFDCSGLAQY
B. tropicalis          1  AGCHEIVTAARSQLGVPYS---WGGGNWAGKSKGIDSGA--------HTVGFDCSGLAQY
B. bassiana            1  -DCPGIVKAANSQLGIDYV---WGGGGCKGPSKG----------------GYDCSGLTQY
M. guizhouense         1  -TCPGIVAAAEKMKCKPYV---WGGGNIHGPTNG----------------GFDCSGLTQY
C. maltaromaticum      1  -NVPALLAEAQKWIGTPYS---WGGGNTNGPSLGFGEGA--------NTVGFDCSSFVQY
T. hemipterigena       1  RTCPGIVSAAESQKGIDYV---YGGGGCKGPSKG----------------GYDCSGLTQF
P. omphalodes          1  -TCAAVLAAAKTQTGVPYS---WGGGNCNGKSLGIEQGA--------NTVGFDCSGLTQY
T. islandicus          1  ---ADIVNAAEKEKGIPYV---WGGGGCNGPSDG----------------GFDCSGLTQY
C. militaris           1  RDCPGIVKAADSQLGIDYV---YGGGGCKGPSKG----------------GYDCSGLTQY
O. maius               1  -TCSGIVAAAEKEKGIPYV---YGGGGCKGPSKG----------------GFDCSGLTQF
P. oxalicum            1  ---AAIVAAAEKEKGLPYV---FGGGCAGPSSG----------------GFDCSGLTQY
A. fumigatus           1  ----AIVAAAQKEKGLPYV---WGGGGCKGPSGG----------------GFDCSGLTQY
N. udagawae            1  ----AIVDAAEKEKGLPYV---WGGGGCNGPSGG----------------GFDCSGLTQF
B. bassiana            1  ATLHKVIARLKLQLGKPYV---WGGQTPE-------------------QGFDCSGLIYY
T. rubrum              1  --GQQILDKALTAAGVPYA---WGGGSCEGPTHDMPPWQ-------NGEIGYDCSGLIGW
M. gypseum             1  --GQQVLDKALTAAGMPYA---WGGGSCDGPTGDMPPWD-------NGEIGFDCSGLVSW
A. terreus             1  TVGQQILDTALTASGVPYA---WGGGDCAGPTGDQPPYD-------YGDVGYDCSGLVCW
T. vaginalis           1  --GEDIFNIARSKLGCPYV---YGAAGPN--------------------SFDCSGLTYY
H. capsulatum          1  --GQKVLDIAMKEKCTKYV---WGGGSCKGPTKG----------------GYDCSGLVAY
M. pharaonis           1  ----SLLRRAMTLLGTPYV---WGGSSPDS----------------GFDCSGLVGY
B. bassiana            1  -----------QLGKPYR---WGGTSPR------------------TGFDCSGLIYY
B. dermatitidis        1  --GQKILAVAMKEKGTRYV---WGGGSCKGPTKG----------------GYDCSGLVAY
Streptomyces sp.       1  -----VIDAARSQIGLPYA---WGGGSLNGPSGGSSPDV--------GVVGFDCSALARF
K. sinensis            1  ---SAVVDAALKWLGTPYS---WGGGGLEGPTFGIAHGA--------GTQGFDCSGLTRY
S. marina              1  ----AVISAARKWIGTPYA---WGGGTLNGPSQGTGIDA--------GVTGFDCSSIVRY
M. tuberculosis        1  --GRRIVAEAKAQRGVPYS---WGGGPSGKSRGICCSP--GGYDGRKTIGFDCSGLVQY
C. propinquum          1  QQAQAILSRARSQLGVLYA---WGGGNAQGPTAGIRDGGHADVNGDYGKVGFDCSGLVQY
T. elongata            1  ------------IGTPYS---WGGGDLDGPTLGIESGA--------NTVGFDCSGLARY
M. tuberculosis        1  ELCKRIIEGAEKWLGTRYV---WGGGNINGPTAG----------------GFDCSGLALY
G. effusa              1  --CPNIVAAAASQMGIRYS---WGGGDLHGPTNGIHDPV-GDQYGDYNHPGWDCSGLARY
M. chlorophenol        1  QASEVVIRRAMSQLGVPYS---WGGGNANGPSRGIDQGA--------NTVGFDCSGLMLY
M. tuberculosis        1  ARCALVARAALKWLGTPYS---WGGGNTFGPSHGVEQGA--------GTVGFDCSGLAMY
L. acidophilus         1  ----AIVKLAKKQVGKPYV---WGATGPY--------------------GFDCSGLTTY
Bacillus sp.           1  -SASSVISIAKQYLGVPYV---WGGTTPS--------------------GFDCSGFTSY
H. halophilus          1  RICNQIVDLAFKYDYAPYV---WGGTSTH--------------------GFDCSGYTQY
B. cereus BCYKFC       1  -AADDLINTGKXFLGLPYI---WAGTSGF--------------------GFDCSGFTHT
M. tuberculosis RipA   1  -ASEYVIRRCMSQLGVPYS---WGGGNAAGPSKGIDSGA--------GTVGFDCSGLVLY
C. solmsi              1  -VKSRIMDQYADWKGVRYR---LGGSTKT--------------------GIDCSGFVQR
C. capitata            1  -VLAALHDQMHTWQCAPYH---YGGTTLR--------------------GVDCSGFVWR
P. xylostella          1  -LRARILDQYQKWKCTQYQ---WGGTTRR--------------------CVDCSALMQH
D. fasciculatum        1  -ASAASKAGHYASCCCPYV---WGGTSCGCGGSG---------------GMDCSGLVYT
T. thermophile         1  -LRQKFIDQCKKYFGVPYHRKYWTEGESHHNAPL---------------YLDCCGLIRQ
consensus              1   g   iv a  igvpY   wggg   gps g              gfDCsgl  y
```

Fig. 3. Homology between the mite LytFM homologues and selected eukaryote and prokaryote NlpC/P60 proteins. The sequences of mature LytFM (*D. pteronyssinus*) and its homologues *D. farinae* (GenBank: KN266412), *P. ovis* (GenBank: FR749374) and *B. tropicalis* (GenBank: CB282085) were compared with eukaryote and prokaryote NlpC/P60 proteins and homology determined using MAFFT [34]. The alignments were then shaded using boxshade (http://www.ch.embnet.org/software/BOX_form.html). Accession numbers are those cited in Fig. 4. The Asp$_{256}$ residue in the *B. cereus* BCYKFC and its equivalent in the *P. oxalicum* protein that directly interact with the free amino group of the L-Ala residue in the stem peptide during catalytic cleavage [5] are highlighted in red.

with the latter two found within the SH3b domain (data not shown) rather than in the P60 domain [5]. All of the proteins shown in Fig. 3 share the catalytic residues involved in cleavage and only one, namely, the protein from *Penicillium oxalicum* has a corresponding residue to the Asp$_{256}$, although it does not possess a SH3 domain.

The availability of several three dimensional structures of NlpC/P60 Superfamily members has enabled the structure of the mite proteins to be modelled using PHYRE 2.0. For each mite protein, the highest-ranking model was obtained using the NlpC/P60 domain of *Mycobacterium tuberculosis* RipA and RipB proteins [27]. Each of the homologues was modelled with 97% coverage and 100% confidence and Fig. 5A and B show cartoon and surface models respectively, for the *D. pteronyssinus* homologue. Fig. 5C shows the *D. pteronyssinus* model superimposed on the *B. cereus* BcYkfC domain template, highlighting the location of a relatively large insertion between positions 22 and 38 in the mature

D. pteronyssinus LytFM sequence in the N-terminal region seen in some actinomycete and ascomycete proteins, but absent from some Firmicutes and from other eukaryotic sequences.

3.5. PCR analysis of a D. pteronyssinus gDNA library reveals the presence of a prokaryotic promoter, 3′ to the lytFM gene

A reverse primer, FSPR based on the 179 bp 3′ sequence of the *D. pteronyssinus* cDNA clone EX163010 was used in conjunction with a *lytFM* forward primer, GSP2 to amplify nucleotide sequences flanking the 3′ end of *lytFM* in the *D. pteronyssinus* gDNA library (Fig. 1). An amplicon (GenBank: KT595670) was obtained and shown to be completely homologous with the 179 bp 3′ sequence (data not shown) indicating that it is a consistent feature and that no intron splicing occurs in the region adjacent to the 3′ end of the gene (data not shown). Although the 179 bp-3′ sequence did not match any sequence in the nucleotide or protein databases, the

```
D. pteronyssinus        49  SVYQCT---HKVL----AR-VASGQYS-DP----KCHHV-----A-YGSHQPGDLVFFG-
D. farinae              49  STYQCT---HKTI----AR-TAAAQYN-DN----HCHHV-----A-YGSHQPGDLVFFG-
P. ovis                 49  SVYQCT---HKVI----AR-VAAAQYN-DR----QCHRV-----P-FSQHQPGDVVFFC-
B. tropicalis           50  AVYHCT---HKKI----AR-VASAQYA-DH----QCHHV-----P-YAQHLPGDLVFFND
B. bassiana             41  SVCKAQ---GKTI----PR-TIQQQYH-SP----MGKHL-----P-RAQAKPGDMLFWAT
M. guizhouense          41  SVYQAE---KKEI----PR-TAQAQYA-SK----LGKHI-----P-RAQAKAGDLLFWGK
C. maltaromaticum       49  VFGRL----GIGL----PR-TTYQQEY-------QGTYL-----E-LSQLQAGDLIFWGG
T. hemipterigena        42  SICKAQ---GKTV----PR-TAQEQYH-SS----MGKHI-----P-RAQAKAGDLLFWAT
P. omphalodes           49  AVCQATGK-SVVI----PR-TIRGQYA-------VGKRI-----P-LAQRQIGDLIYYAT
T. islandicus           39  AVCQAL---KKEI----PR-TAQTQYD-SS----LGKRL-----P-RSEAQEGDLLFWAE
C. militaris            42  SVCKAQ---GKTI----PR-TAQEQYH-SS----MGKHL-----P-RAQAKAGDMLFWAT
O. maius                41  AVCKAQ---GKTI----PR-TAQTQYH-ST----MGKRY-----P-RSQAKEGDLLFWAT
P. oxalicum             39  SLCKAL---NKKI----PR-DAQSQYDATS----QGKRI-----P-RAQAQEGDLLFWAN
A. fumigatus            38  AVCQAL---KKTI----PR-TAQTQYN-SN----MGKRI-----P-RAQAKEGDLLFWAT
N. udagawae             38  AVCKAL---KKTI----PR-TAQTQYH-SS----MGKRL-----P-RSQAKEGDLLFWAT
B. bassiana             38  AYNQLL---SQKL----PR-TANGMEQ-DR----RIKRI-----S-QRELRRGDLVFFSI
T. rubrum               49  AVCQVT---GRDLFSEGLR-VTRSMYCASEEKL-RYKKY-----P-FAERKPGDAVFFCG
M. gypseum              49  AVCQVT---GRDLFKEGLR-VTRTMYCASEEKL-RYKKY-----P-YAERQAGDAVFFCG
A. terreus              51  AVCQVT---GRDLFTEGLR-VTSSMYCADEATL-GYKKY-----P-YEERQAGDAVFFCG
T. vaginalis            35  CHQQV----GISI----PR-VASAQQA-------GGV--------ATNGMCGDICCFC-
H. capsulatum           40  AVCKAT---KRNLFKEGLR-VTYSMYCASSSRLGKFKKV-----P-FSKRRAGDAVFFGT
M. pharaonis            34  VFRSAL---GIEL----PR-VSREMAH-DD----NAELI-----NDRTALAAGDLVFFCR
B. bassiana             26  AYKDLV---NIRI----PR-TANEMYH-LR----DASSV-----D-RSELESGDLVFFRT
B. dermatitidis         40  AVCKVT---KRNLFREGLR-VTYSMYCASSKRLGKFKKV-----P-FSKRRAGDLVFFGS
Streptomyces sp.        45  AYYQCTNG-KITL----PR-SSKEQYR-AT----KGHTV-----P-VDQLQPGDLLFWGG
K. sinensis             47  SVFAAT---GRVI----PR-VASDQAR-------ALQRV-----P-TDQIQAGDLLFFHA
S. marina               46  AYYQATKG-KTTL----PR-TSAAQYQ-AT----AAHQI-----P-RDHLQPGDLLFYGS
M. tuberculosis         54  AVYRASKG-RITM----PR-VADEQVH-------RGKPV-----T-RNSMRPGDLIGFDH
C. propinquum           58  AFAGA----GISL----PH-YSGGQYR-------HGTHV-----S-TSNLQPGDLIFYGP
T. elongata             37  AIHQAT---GRVI----PR-VATDQAA-------ALTPV-----P-LDQIRAGDLLFFHN
M. tuberculosis         42  AVYHATGG-QIKL----PR-VADQQR-DP----RGKRV-----S-RSELQPGDLVFFTA
G. effusa               55  AVFVAT---GKVI----AR-VSQAQST-------GGMAV-----PNQAAAQPGDLVFFCG
M. chlorophenol         50  AFAGV----GIKL----DH-YSGSQYN-------AGRKI-----P-SAQMRRGDLIFYGP
M. tuberculosis         50  AWNKA----GVRL----DH-WIGTQWT-------SGPHI-----P-TGALRPGDLVFFAT
L. acidophilus          33  VYKNAA---NKTL----PR-TTYGQIT-IG----KSVSV-----S-TKSLKKGDLLFWGN
Bacillus sp.            36  VFRAV----GVSL----PR-VSSAQQN-------VGTQV-----S-TSSVQPGDLVFMC-
H. halophilus           37  VYDKA----GIDL----PR-TSRSQYG-------EGDYV-----S-RSNLQPGDLVFFG-
B. cereus BCYKFC        36  IYKSH----GITI----PR-DSGPQSR-------NGVAV-----D-KEHLQKGDLIFFAH
M. tuberculosis RipA    49  SFAGV----GIKL----PH-YSGSQYN-------LGRKI-----P-SSQMRRGDVIFYGP
C. solmsi               36  TFREQF---GLDL----PR-STYEQQE-------MGKSI-----S-RTKLRTGDLVLFRA
C. capitata             36  TLKDRF---NLPM----SRITTRELIH-------MGQRI-----E-PSRLQPGDLVFFRI
P. xylostella           36  LFSDVA---HLNL----PR-TIGEQIH-------RGVQV-----T-QYRLRPGDLVFFQT
D. fasciculatum         41  SYREAG---YTGI----QR-TIFEQIN------QGSSCGSCSPSNLGSCKVGALFFYCF
T. thermophile          44  VVYDLREDFGFTL----GR-WNQAYQF-DV----CPKEV-----K-FEEMKPGDLIFYSG
consensus               61  av a     gk l     pr t  qy         ghhv      p   l  Gdlvffg
```

Fig. 3 (*continued*).

bacterial promotor prediction softwares indicated one between nucleotides 76 and 125 (NNPP score of 0.97, threshold 0.80) and another between nucleotides 97 and 129 (BPROM score 34, threshold 0.2) (data not shown).

PCR screening of the *D. pteronyssinus* gDNA library reveals homologies between the 5′ and 3′ flanking sequences of the *lytFM* gene with both eukaryotic and prokaryotic sequences, respectively

Attempts to identify sequences 5′ and 3′ to the *lytFM* gene were undertaken to determine possible ancestry. PCR was performed using a forward vector sequence primer, VP1 and a reverse *lytFM* gene sequence primer, GSP1 to amplify sequences upstream from the start of *lytFM* (Fig. 1). An amplicon was obtained that contained both vector and *lytFM* sequences (data not shown), as well as a sequence (GenBank: AKG95503) with homology to eukaryotic and prokaryotic hydroxysteroid dehydrogenases (Table 2). A similar screening approach using a *lytFM* forward primer, GS2 in conjunction with the reverse vector primer, VP2 (Fig. 1) amplified sequences downstream from the 179 bp-3′ sequence, some of which (GenBank: KT595671 and KT595672) showed high sequence homology with a hypothetical gene in *Escherichia coli* (Table 3).

4. Discussion

Homologues and variants of the *lytFM* gene have now been shown to be present in *D. pteronyssinus*, *D. farinae*, *P. ovis*, and *B. tropicalis* but not in *T. urticae* or *M. occidentalis*. With regard to *D. pteronyssinus*, five cDNA sequences have now been reported, and a comparative analysis indicates that the *lytFM* gene is polymorphic. Although most of the polymorphisms are silent, those in the 3′ region of the gene alter the pI of the LytFM1 and LytFM2 isoforms. In addition to identifying polymorphisms in *D. pteronyssinus*, a conservative polymorphism in the *D. farinae* homologue was also identified using the reported unplaced genomic scaffold (KN266412) [19]. These findings are consistent with those seen with other mite proteins, in particular the cysteine protease, trypsin and the MD-related group 2 protein [28,29].

This study also showed that there were no introns in the *D. pteronyssinus lytFM* gene which contrasts with that seen in *D. farinae* [19], where a single intron was observed in the leader sequence of the gene rather than in the region encoding the mature protein. In addition, the *D. pteronyssinus*, *D. farinae* and *P. ovis*

```
D. pteronyssinus       89  -----------NPIHHVGIVSAH--G-RMINAP-HTGT-NVREENIW-S-DHIANVARCW
D.farinae              89  -----------NPIYHVGIVSAH--G-RMVNAP-KPGT-KVREENIW-S-YHISHVARCW
P. ovis                89  -----------NPPYHVGIVSAH--N-KMANAP-KPGT-TVREESIW-G-YHLGNVARCW
B. tropicalis          91  G----------GSIHHVAIISGK--N-TMIHAP-HTGD-HVREAAVY-VKGRMSTVQRCF
B. bassianagi          82  GGDCK------NKVAHVGIFVKA--G-TMVNAA-HTGT-KVRQQAIW-T-----------
M. guizhouense         82  GGDCK------TGVVHVGIFTKP--G-WMVNAA-HTGV-PVREQKIW-TSY---------
C. maltaromaticum      87  R----------GSTHHVGIYMGG--N-QYIHAP-ETGD-VVKTSSVS-S-----------
T. hemipterigena       83  GGNCK------TGVEHVGIFVKA--G-TMINAA-HTGT-KVREQSIW-TS----------
P. omphalodes          90  GGDCSCSNGYCAALYHTAIWAGN--N-QMFEAQ-KTGT-TL------------------
T. islandicus          80  GGDCT------NSVAHVGIFIRD--G-LMINAA-HSGT-PVREQATW-TSSGGESICP--
C. militaris           83  NGDCK------NHVAHVGIFVKA--G-TMVNAA-HTGT-KVRQQAIW-T-----------
O. maius               82  NGDCT------SKVVHTGIYIKP--G-LMINAA-HTGT-PVREESIW-TSSGGEEICP--
P. oxalicum            81  NGNCN------TGVAHVGIFIRD--G-LMINAA-HTGT-PVREQSIW-T-----------
A. fumigatus           79  GGDCK------NKVVHVGIFIRD--G-WMINAA-KTGT-PVREQSIW-TSY---------
N. udagawae            79  GGDCA------NKVSHVGIFIRD--G-WMINAA-HTGV-PVREQSIW-TSYGGESICP--
B. bassiana            79  KSQ--------GSADHVGVYLGE--G-QFIEAP-RTGL-NIRISQLA-DNYWQDHYLGA-
T. rubrum              98  SCDCGNP----DTIHHVGLMMDS--GDRMWNAP-NDDVNQVQENNI-------------
M. gypseum             98  KCSCAQS----DTIHHVGLMMDS--GDRMWNAP-NDKINRVQEN--------------
A. terreus            100  ECSCADS----SSIHHVGLMMDS--GDRMWNAP-NDRVNAVQEN--------------
T. vaginalis           69  -----------NPAYHVGLCDGV--C-NYIHAP-QSGD-VVRYSNYW-NGVVVNFKR----
H. capsulatum          90  SCSCKN-----QNISHMGLMINS--GDRMIHAP-NPST-VVREGKV-------------
M. pharaonis           76  K----------GRVDHVGIYVGD--G-RFLHAP-SSGK-DVRVDTLL-SGYWGN------
B. bassiana            67  QGR--------GTADHVGVYVGN--G-KFIQSP-RSGR-DIQITSLS-EDYWQRHYVGA-
B. dermatitidis        90  NCSCKN-----QNISHMGLMINS--GDRMIHAP-NRRT-VVKEGRV-------------
Streptomyces sp.       88  TA---------ESIHHVALYIGG--G-KMIEAP-ESGK-KLRE---------------
K. sinensis            86  DGDPP------GFYHHVGIADGN--G-GMVHAP-RTGR-TVEVVPNVLSSGYFAEQL---
S. marina              89  SP---------SSIHHIAIYSGN--N-HQVEAP-QSGE-KIHEANL-------------
M. tuberculosis        95  G----------QGVTHIGIYIGD--G-RMVHAP-QTGD-VVKISSLA-S-----------
C. propinquum          96  -----------GGSQHVAIYSGN--G-MMIEAP-HSGS-HVRE---------------
T. elongata            76  SADPP------GVYHHVGIADGQ--G-GMIHAP-RTGA-FVEIVPSV-RDGYWMTEL---
M. tuberculosis        85  AGK--------TDAHHVGIYVGG--G-KMVNAP-HTGA-VVREAIG-WKEFSGG-----
G. effusa              95  E----------GSAHHVGIYMGG--G-KMVNAP-ESGK-KVSIADI-------------
M. chlorophenol        88  -----------NASQHEAMYLGD--G-MMIEAP-YTGS-VVR----------------
M. tuberculosis        88  DPSDP------GTIHHVGVYIGD--G-RMVEAP-YTGA-RVRISSI-------------
L. acidophilus         74  S--------------HVGIYIGN--G-KFVHAP-APGQ-NVKTQTLA-SFYPSAAK----
Bacillus sp.           73  -----------KPAYHVGIYIGG--G-QYIHAP-QTGD-VVK----------------
H. halophilus          74  -----------NPVWHVGIYIGN--N-KMISAE-NPSD-DI----------------
B. cereus BCYKFC       74  DQGK-------GSVHHVAXYIGD--G-NXIHSP-RAER-SVEIIPLN-TPGYIEEYAGAR
M. tuberculosis RipA   87  -----------NGSQHVTTYLGN--G-QMIEAP-DVGL-KVRVAPVR-TAGMTPYVVRY-
C. solmsi              75  G----------STGRHVGIYIGN--D-QFVHAS-TSSG--VTISSMN-EPYWKKRYNEAR
C. capitata            76  -----------HRELHVGFYDTD--R-HFLHAS-VHSG--VTRSSLD-AAYWRNVFLE--
P. xylostella          75  G----------PNRRHVGIYIGN--S-QFIHAS-SSQG--VTVSTLT-DAYWQAHYITA-
D. fasciculatum        86  SQ---------PCPDHVVMYLGG--G-KAAECP-EPGQ-DCHIITPY-SESYYACRTLC-
T. thermophile         88  TYYN-------EKMVHVEVFTGGETCEQTIGARWQKGFVQYHDSYKF-VSKNYYDIQFYY
consensus             121             iHvgiyvg   g  mi ap  tg   vr   i
```

Fig. 3 (*continued*)

LytFM proteins are secreted, given the presence of a leader sequences. The predicted cleavage site for *D. pteronyssinus,* however, did not correspond with that observed experimentally [1] using the eukaryotic data set but did so with the Gram-negative data set. In contrast, the *B. tropicalis* protein did not appear to possess a leader sequence but confirmatory sequence data are required before it can be concluded that the latter is a cytoplasmic protein cleaving murein peptides intracellularly.

The *lytFM* gene is present in Sarcopteformes mites occupying different ecological niches. In this regard, both *Dermatophagoides* spp are widely distributed in temperate climates whereas *B. tropicalis* is associated with the tropics. In contrast, *P. ovis* is a parasitic mite infesting the skin of animals such as rabbits and sheep. Interestingly, a NlpC/P60 homologue is absent from the genome of the hematophagous tick *Ixodes scapularis* [4], and is yet to be described in the human skin-dwelling mite *Sarcoptes scabei.*

Phylogenetically, the greatest similarity was found to occur between the two *Dermatophagoides* spp and *P. ovis*, rather than with *B. tropicalis* and this is consistent with the relationships observed using homologues of the mite cysteine protease allergens (e.g. Der p 1) and the MD-related allergens (e.g. Der p 2) (data not shown). However, the degree of similarity between the *D. pteronyssinus* LytFM protein and the *B. tropicalis* homologue (61% identity) is higher than that seen with the *B. tropicalis* homologues of Der p 1 (35%) and Der p 2 (42%). The mite proteins showed some homology with homologues from other eukaryotes, and share a common ancestor with the actinomycetes and ascomycetes, with the Firmicutes, trichomonads, wasp, fruit fly, moths and protozoa being more distantly related.

Several of the eukaryote proteins including the mite LytFM homologues are single domain proteins as revealed by a Pfam search (data not shown), although three proteins (*Aspergillus fumigatus* GI 846917091 and GI 70997053, and *Neosartorya udagawae*, GI 849274716) were shown to possess a SH3_3 domain and one protein was a multidomain protein (Chemotaxis protein LafT, *Beauveria bassiana* GI 701777570). In addition, disulphide bond arrangement in the mite homologues is clearly different from that likely to occur in the ascomycetes. Here, there were distinct differences with cysteines being absent from the actinomycetes and a number of arrangements for the ascomycetes particularly in a C terminal insertion absent from the mite sequences.

The availability of domain and enzymatic substrate data [5,27,30] enables deductions to be made regarding the possible peptidoglycan substrates susceptible to the mite enzymes as well
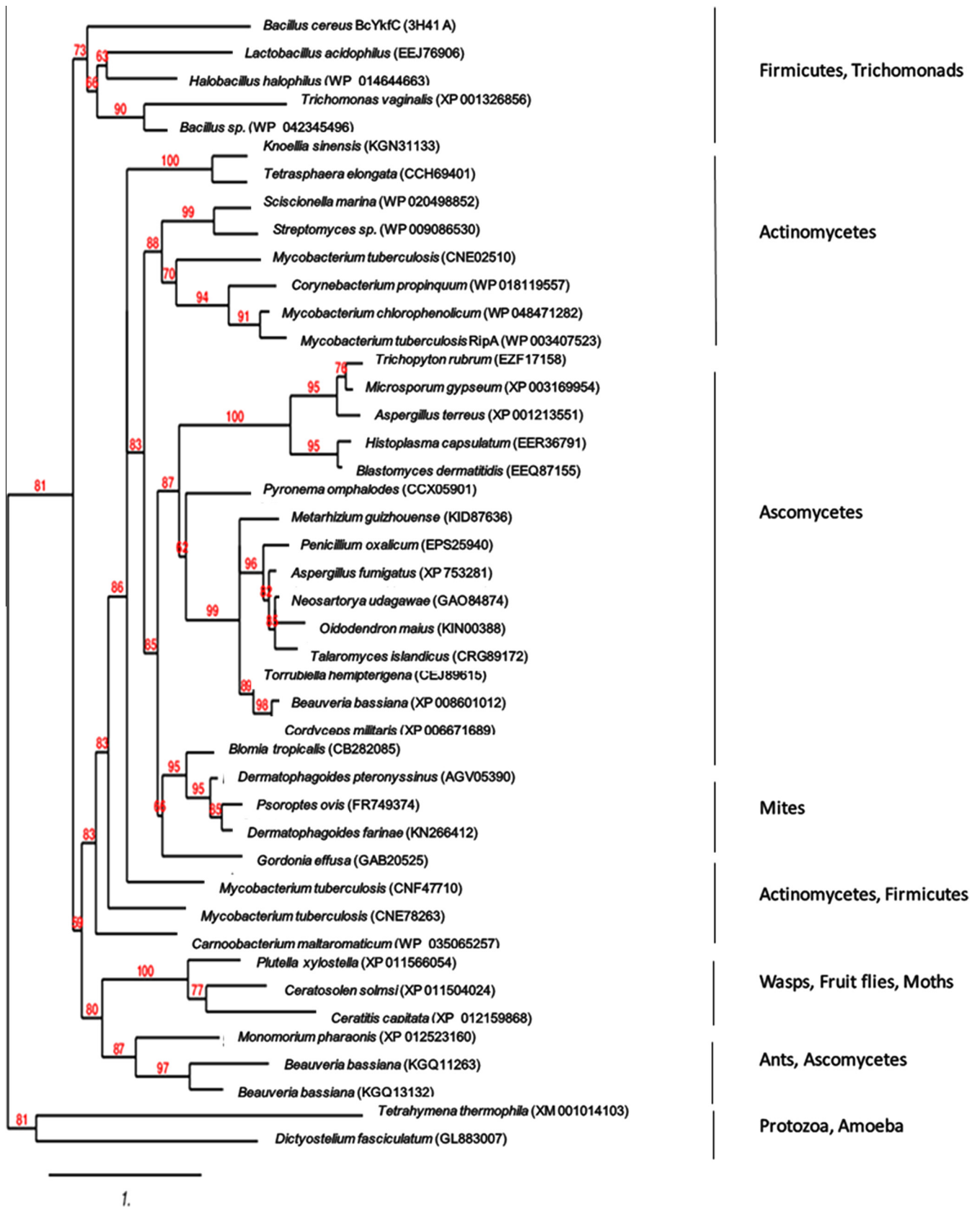
Fig. 4. Phylogenetic relationships of mite LytFM homologues and selected eukaryote and prokaryote NlpC/P60 proteins. The phylogenetic tree was created using Phylogeny.fr pipeline. The approximate likelihood ratio test was used and the confidence values shown on the branches.
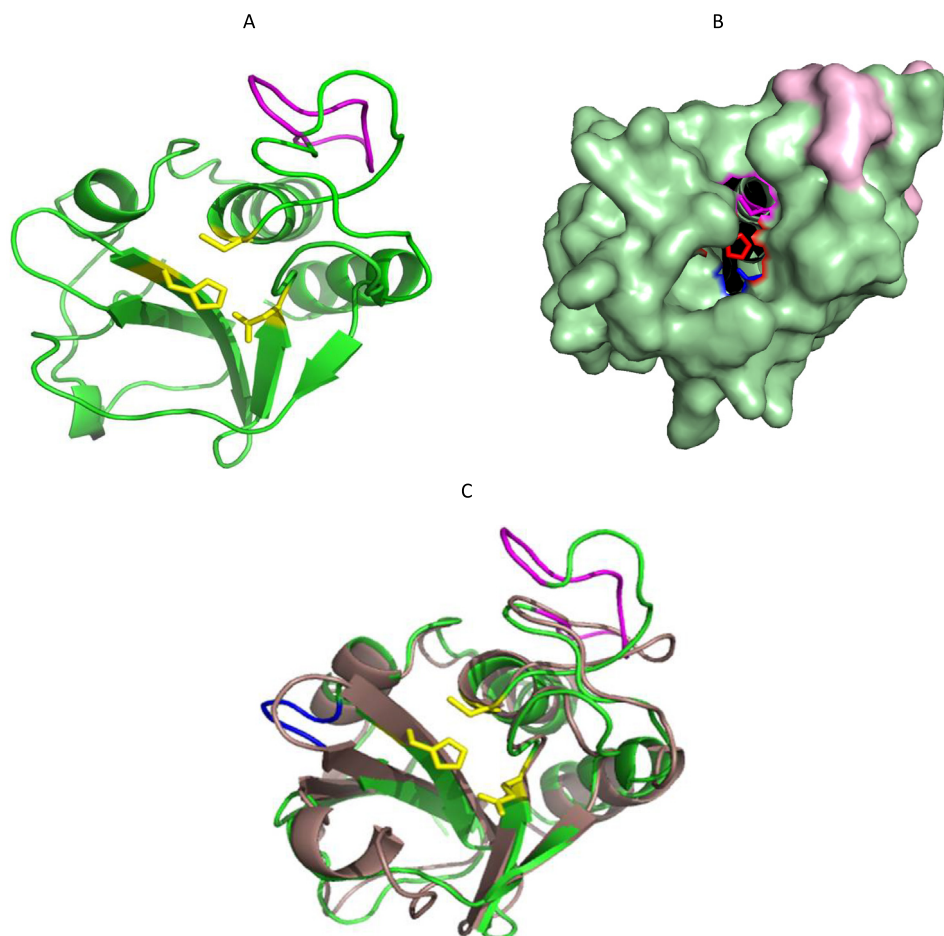
Fig. 5. Model of the mite LytFM protein. (A) A model of *D. pteronyssinus* LytFM generated by PHYRE2, using *M. tuberculosis* RipA (PDB: 2XIV) and *B. cereus* BcYkfC (PDB: 3H41). The 13-residue insertion compared to the *B. cereus* BcYkfC protein and catalytic residues are highlighted in magenta and yellow, respectively. (B) A surface representation of the LytFM showing the putative substrate-binding groove with the Cys_{42}, His_{93} and Asn_{105} catalytic residues revealed and shown in magenta, red and blue, respectively. The 13-residue insertion found in LytFM and the *M. tuberculosis* protein is highlighted in pink. (C) The LytFM model (green) is superimposed on the *B. cereus* BcYkfC template (light brown) to highlight the relative location of the 13-residue insertion (magenta) in the former. A 4-residue deletion in the mite LytFM homologues and *M. tuberculosis* protein relative to the *B. cereus* BcYkfC is shown in blue. The catalytic residues are highlighted in yellow.

Table 2

BLASTX search results indicating homology of the deduced amino acid sequence of the PCR[a] product amplified from the *D. pteronyssinus* gDNA library with eukaryotic and prokaryotic 3 hydroxysteroid dehydrogenases.

GenBank accession number	Amino acid sequences bearing homology	Query coverage (%)	*E* value	Maximum identity (%)
KFM77175	Short-chain dehydrogenase reductase family 42E member from *Stegodyphus mimosarum* (spider)	97	3e−22	49
XP_002434360	Putative 3 hydroxysteroid dehydrogenase from *Ixodes scapularis* (tick)	96	1e−17	42
XP_003743458	Predicted 3 BETA-hydroxysteroid dehydrogenase/delta 5 → 4-isomerase type 7-like hydroxysteroid dehydrogenase from *Metaseiulus occidentalis* (mite)	96	4e−15	39
XP_003743458	3 beta-hydroxysteroid dehydrogenase/delta 5 → 4-isomerase type 1 from *Zootermopsis nevadensis* (dampwood termite)	96	1e−14	42
XP_008216462	3 beta-hydroxysteroid dehydrogenase/isomerise type 4 from *Nasonia vitripennis* (wasp)	96	6e−14	41
XP_009051883	Hypothetical protein LOTGIDRAFT 114779 from *Lottia gigantea* (sea snail)	96	8e−13	39
XP_011181291	Predicted 3 beta hydroxysteroid dehydrogenase/delta 5 → 4 isomerase type 3 from Bactrocera cucurbitae (melon fly)	96	1e−12	44
WP_012470322	3 beta-hydroxysteroid dehydrogenase from *Geobacter lovleyi*	87	2e−9	41

[a] The PCR product was amplified using the primer set VP1/GSP1 and found to align with the 5′ end of *lytFM* within the λBlueSTAR™ *D. pteronyssinus* gDNA library.

as the actinomycetes and other eukaryotes. In this regard, structural data show that the *B. cereus* endopeptidase, BcYkfC, cleaves the bond between γ-D-Glu and DAP in the stem peptide but only when the substrate has a free N-terminal Ala residue, and is, therefore, devoid of peptidoglycan [5]. The bacillus protein comprises both the P60 domain as well as two SH3b domains, one of which contributes to the S2 site, as well as sterically restricting access

to larger substrates containing peptidoglycan [5]. The first two residues constituting the catalytic triad (Cys and His) are invariant throughout the sequences including the mite proteins but the third residue varies slightly between members, and may be His, Glu, Asn, Gln. The Asp interacting with the γ-D-Glu residue of the stem peptide in the *B. cereus* P60 domain is invariant in the mite, ascomycete and actinomycete homologues whereas the Asp_{256} residue in

Table 3
BLASTN search results showing homology of DNA sequences of PCR products[a] amplified from the *D. pteronyssinus* gDNA library with Gram-negative bacterial genes.

GenBank accession number	DNA sequences bearing homology	Gene showing sequence homology	Function of gene product	Query coverage (%)	E value	Maximum identity (%)
AP012030	*Escherichia coli* DH1 (ME8569) DNA, complete genome	htrL	Hypothetical protein	79	6e−166	96
AE005674	*Shigella flexneri* 2a strain 301, complete genome	htrL	Protein involved in lipopolysaccharide biosynthesis	79	2e−160	95
CP001383	*Shigella flexneri* 2002017, complete genome	htrL	Protein involved in lipopolysaccharide biosynthesis	79	2e−160	95
FN554766	*Escherichia coli* 042, complete genome	rlpA	Lipoprotein A	98	2e−165	97
		mrdB	Rod shape-determining protein RodA			
FN649414	*Escherichia coli* ETEC H10401, complete genome	ETEC_0661	Rare lipoprotein A	79	2e−160	95
		ETEC_0662	Rod shape-determining protein RodA			

[a] The PCR product was amplified using the primer set GS2/VP2 and found to align with the 3′ end of *lytFM* within the λBlueSTAR™ *D. pteronyssinus* gDNA library.

the P60 domain of the bacillus protein along with Glu_{83} and Tyr_{118} from the SH3b domain that interact with the free amino group of the L-Ala residue are missing. It is also missing from the two SH3 domain containing ascomycete homologues. In addition, the Tyr involved in catalysis is also invariant [5].

In contrast to the bacillus protein, enzymatic studies show that although the mycobacterial RipA and RipB endopeptidases have the same cleavage specificity, they can accommodate peptidoglycan attached to stem peptide as they lack the equivalent SH3b domain and possess two cavities, one at each end of the substrate binding groove; the first allows for interaction with the carbohydrate components of peptidoglycan and the second allows for interaction with the residues beyond the D-Glu of the stem peptide [27]. Similarly, *Lactobacillus casei* Lc-P75 protein cleaves the murein tetrapeptide of lysine type peptidoglycans [30]. Given the sequence homology with the mycobacterial proteins, it is likely that the mite and other eukaryote proteins will cleave peptide substrates attached to peptidoglycan.

In searching for bacterial ancestry, 5′ and 3′ sequences were analysed. With regard to 3′ sequences, the 179 bp flanking sequence downstream of the 3′ end of *lytFM2* was found to contain several putative prokaryotic promoter sequences although promoter activity was not assessed. However, these sequences might suggest an early capture of bacterial sequences during mite evolution, and further study of the 3′ region identified a hypothetical protein from *E. coli*. The weak homology of this sequence with the corresponding genes in *E. coli* K-12 from which ER1647 (the host in which the *D. pteronyssinus* gDNA library was propagated) was derived by a series of mutations (data not shown) [31–33] indicates that they were unlikely to be chromosomal contaminants but generated by bacterial sequences in the gDNA library. If confirmed, these may have arisen from endosymbionts incorporated during preparation of the library. With regard to the 5′ flanking region, a sequence with high homology to a putative arachnid 3 hydroxysteroid dehydrogenase was identified, suggesting it was mite-derived.

In summary, the data presented reveal that the mite LytFM proteins are single domain proteins, that are likely to cleave stem peptide bound to N-acetylmuramic acid in peptidoglycan. They are found in several families of Sarcopteformes mites but are absent from the herbivorous Trombidiformes mite *T. urticae* and the Parasitiformes predatory mite *M. occidentalis*. They are present amongst dust mite species and are likely to act in concert with other enzymes to completely degrade peptidoglycan for either defence or nutrition. Their presence raises the possibility that inhalation of mite fecal material containing peptidoglycan breakdown products may be pro-inflammatory via activation of NOD (nucleotide-binding oligomerisation domain-containing protein)-like receptors. Finally, the finding of a gene encoding a P60 family member in the *D. pteronyssinus* genome together with the presence of a bacterial promotor suggests an evolutionary link to one or more prokaryotic endosymbionts.

References

[1] Mathaba, L.T., Pope, C.H., Lenzo, J., Hartofillis, M., Peake, H., et al. (2002) Isolation and characterisation of a 13.8-kDa bacteriolytic enzyme from house dust mite extracts: homology with prokaryotic proteins suggests that the enzyme could be bacterially derived. FEMS Immunol. Med. Microbiol. 33, 77–88.

[2] Bubert, A., Kuhn, M., Goebel, W. and Kohler, S. (1992) Structural and functional properties of the p60 proteins from different Listeria species. J. Bacteriol. 174, 8166–8171.

[3] Aramini, J.M., Rossi, P., Huang, Y.J., Zhao, L., Jiang, M., et al. (2008) Solution NMR structure of the NlpC/P60 domain of lipoprotein Spr from *Escherichia coli*: structural evidence for a novel cysteine peptidase catalytic triad. Biochemistry 47, 9715–9717.

[4] Rawlings, N.D., Waller, M., Barrett, A.J. and Bateman, A. (2014) MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. Nucl. Acids Res. 42, D503–D509.

[5] Xu, Q., Abdubek, P., Astakhova, T., Axelrod, H.L., Bakolitsa, C., et al. (2010) Structure of the gamma-D-glutamyl-L-diamino acid endopeptidase YkfC from *Bacillus cereus* in complex with L-Ala-gamma-D-Glu: insights into substrate recognition by NlpC/P60 cysteine peptidases. Acta Crystallogr., Sect. F: Struct. Biol. Cryst. Commun. 66, 1354–1364.

[6] Garcia, P., Gonzalez, M.P., Garcia, E., Lopez, R. and Garcia, J.L. (1999) LytB, a novel pneumococcal murein hydrolase essential for cell separation. Mol. Microbiol. 31, 1275–1281.

[7] Takahashi, J., Komatsuzawa, H., Yamada, S., Nishida, T., Labischinski, H., et al. (2002) Molecular characterization of an atl null mutant of *Staphylococcus aureus*. Microbiol. Immunol. 46, 601–612.

[8] Vollmer, W., Joris, B., Charlier, P. and Foster, S. (2008) Bacterial peptidoglycan (murein) hydrolases. FEMS Microbiol. Rev. 32, 259–286.

[9] Kolb-Maurer, A., Pilgrim, S., Kampgen, E., McLellan, A.D., Brocker, E.B., et al. (2001) Antibodies against listerial protein 60 act as an opsonin for phagocytosis of *Listeria monocytogenes* by human dendritic cells. Infect. Immun. 69, 3100–3109.

[10] Marino, M., Banerjee, M., Jonquieres, R., Cossart, P. and Ghosh, P. (2002) GW domains of the *Listeria monocytogenes* invasion protein InlB are SH3-like and mediate binding to host ligands. EMBO J. 21, 5623–5634.

[11] Korndorfer, I.P., Danzer, J., Schmelcher, M., Zimmer, M., Skerra, A., et al. (2006) The crystal structure of the bacteriophage PSA endolysin reveals a unique fold responsible for specific recognition of Listeria cell walls. J. Mol. Biol. 364, 678–689.

[12] Lu, J.Z., Fujiwara, T., Komatsuzawa, H., Sugai, M. and Sakon, J. (2006) Cell wall-targeting domain of glycylglycine endopeptidase distinguishes among peptidoglycan cross-bridges. J. Biol. Chem. 281, 549–558.

[13] Xu, Q., Sudek, S., McMullan, D., Miller, M.D., Geierstanger, B., et al. (2009) Structural basis of murein peptide specificity of a gamma-D-glutamyl-L-diamino acid endopeptidase. Structure 17, 303–313.

[14] Anantharaman, V. and Aravind, L. (2003) Evolutionary history, structural features and biochemical diversity of the NlpC/P60 superfamily of enzymes. Genome Biol. 4, R11.1–R11.12.

[15] Margot, P., Pagni, M. and Karamata, D. (1999) *Bacillus subtilis* 168 gene lytF encodes a gamma-D-glutamate-meso-diaminopimelate muropeptidase expressed by the alternative vegetative sigma factor, sigmaD. Microbiology 145 (Pt 1), 57–65.

[16] Smith, T.J., Blackman, S.A. and Foster, S.J. (2000) Autolysins of *Bacillus subtilis*: multiple enzymes with multiple functions. Microbiology 146 (Pt 2), 249–262.

[17] Oh, H., Ishii, A., Tongu, Y. and Itano, K. (1986) Microorganisms associated with the house-dust mite Dermatophagoides. Jpn. J. Sanit. Zool. 37, 229–235.

[18] Hubert, J., Kopecky, J., Perotti, M.A., Nesvorna, M., Braig, H.R., et al. (2012) Detection and identification of species-specific bacteria associated with synanthropic mites. Microb. Ecol. 63, 919–928.

[19] Chan, T.F., Ji, K.M., Yim, A.K., Liu, X.Y., Zhou, J.W., et al. (2015) The draft genome, transcriptome, and microbiome of *Dermatophagoides farinae* reveal a broad spectrum of dust mite allergens. J. Allergy Clin. Immunol. 135, 539–548.

[20] Sambrook, J. and Russell, D.W. (2001) Molecular Cloning: A Laboratory Manual, third ed, Cold Spring Harbour Laboratory Press, Cold Spring Harbour, New York. pp. 6.11–6.12.

[21] Kelley, L.A. and Sternberg, M.J.E. (2009) Protein structure prediction on the Web: a case study using the Phyre server. Nat. Protoc. 4, 363–371.

[22] Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., et al. (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucl. Acids Res. 36, W465–W469.

[23] Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucl. Acids Res. 32, 1792–1797.

[24] Castresana, J. (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17, 540–552.

[25] Anisimova, M. and Gascuel, O. (2006) Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. Syst. Biol. 55, 539–552.

[26] Erban, T.C., Di, Presa A., Kopecky, J., Poltronieri, P. and Hubert, J. (2013) PCR detection of the 14.5 antibacterial NlpC/P60-like *Dermatophagoides pteronyssinus* protein in *Dermatophagoides farinae* (Acari: Pyroglyphidae). J. Med. Entomol. 50 (4), 931–933.

[27] Both, D., Schneider, G. and Schnell, R. (2011) Peptidoglycan remodeling in *Mycobacterium tuberculosis*: comparison of structures and catalytic activities of RipA and RipB. J. Mol. Biol. 413, 247–260.

[28] Smith, W.A., Hales, B.J., Jarnicki, A.G. and Thomas, W.R. (2001) Allergens of wild house dust mites: environmental Der p 1 and Der p 2 sequence polymorphisms. J. Allergy Clin. Immunol. 107, 985–992.

[29] Smith, W.A. and Thomas, W.R. (1996) Sequence polymorphisms of the Der p 3 house dust mite allergen. Clin. Exp. Allergy 26, 571–579.

[30] Regulski, K., Courtin, P., Meyrand, M., Claes, I.J., Lebeer, S., Vanderleyden, J., Hols, P., Guillot, A. and Chapot-Chartier, M.P. (2012) Analysis of the peptidoglycan hydrolase complement of *Lactobacillus casei* and characterization of the major gamma-D-glutamyl-L-lysyl-endopeptidase. PLoS One 7 (2), e32301.

[31] Imamura, N. and Nakayama, H. (1982) *ThiK* and *thiL* loci of *Escherichia coli*. J. Bacteriol. 151, 708–717.

[32] Raleigh, E.A., Trimarchi, R. and Revel, H. (1989) Genetic and physical mapping of the *mcrA* (*rglA*) and *mcrB* (*rglB*) loci of *Escherichia coli* K-12. Genetics 122, 279–296.

[33] Woodcock, D.M., Crowther, P.J., Doherty, J., Jefferson, S., DeCruz, E., et al. (1989) Quantitative evaluation of *Escherichia coli* host strains for tolerance to cytosine methylation in plasmid and phage recombinants. Nucl. Acids Res. 17, 3469–3478.

[34] Kuraku, S., Zmasek, C.M., Nishimura, O. and Katoh, K. (2013) ALeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. Nucl. Acids Res. 41, W22–W28.