RSC Advances

ELECTRONIC SUPPLEMENTARY INFORMATION

Cross-species biosynthesis of maytansine in *Maytenus serrata*

Parijat Kusari, Souvik Kusari, Dennis Eckelmann, Sebastian Zühlke,

Oliver Kayser and Michael Spiteller

TABLE OF CONTENTS

I. EXPERIMENTAL SECTION

II. SUPPLEMENTARY FIGURES

from *Putterlickia* plants,[1] the using T-Coffee, combined into one final alignment, based on their 16S rRNA regions.

Fig. S4. A phenogram-like rooted phylogenetic tree representing the correlation of our isolated endophytic bacterial communities of *M. serrata* with most closely related bacteria based on their 16S rRNA sequences. EC-MSR-R, EC-MSR-B, EC-MSR-L represent the endophytic bacterial community codes of *Maytenus* plants. Others represent the accession numbers of closely related bacterial species. The arrows provide information about the isolation source or natural host of the respective bacteria.

III. SUPPLEMENTARY TABLES

Table S1. List of primers used in this study

Table S2. Overview of the AHBA synthase and halogenase genes found in endophytic bacterial communities and plant tissues of *Maytenus serrata* plants

Table S3. Presence of detectable major precursors of maytansine in various tissues of the Cameroonian *M. serrata* plant (see also Fig. S2)

References

I. EXPERIMENTAL SECTION

Plant material

The *Maytenus serrata* plants were bioprospected from the Cameroonian rainforest at Bafoussam in September 2014 (leaves, stems/bark, and roots; voucher number 26298 HNC) and from the Ghanaian forest at Asakraka-Asikam (Eastern region) in January 2015 (stems/bark and roots; voucher number KNUST/HM/2015/M101).

Plant extraction

Dried plant material was chopped into pieces and grinded to dust using liquid nitrogen. A mixture of ethanol:trichlormethane (80:20 v:v) was used to suspend the material. Under chilled conditions, the samples were ultrasonicated for 15 min. After filtering using Whatman filter paper, the extract was kept and the residue again suspended in the solvent. The ultrasonication process was then repeated for another 15 min, followed by filtration. The solvent of the combined extracts was evaporated using a rotary evaporator (Laborota 4001, Heidolph, Schwabach, Germany) at 40°C (using a water bath) at 100 rpm. For HPLC-ESI-HRMS^n measurements, extracts were dissolved in 1 ml ethanol.

Isolation of endophytes from different *M. serrata* tissues

Isolation, maintenance and preservation of endophytic bacterial community, endophytic fungi and endophytic bacterial isolates were done according to the published methods.[1]

Genomic DNA extraction, PCR amplification and sequencing

The extraction of total genomic DNA (gDNA) of different tissues of *Maytenus serrata* plants (leaf, root and bark), bacterial endophytic communities and positive control *Actinosynnema* *pretiosum* subsp. *auranticum* (DSM 44131; obtained from Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures, DSMZ, Braunschweig, Germany; the activation of the bacterial strain was performed according to DSMZ guidelines) was done following previously established procedures.[1] Further, the PCR amplification, purification of PCR products and sequencing of 16S rRNA and AHBA synthase gene of *Maytenus* plants, positive control *Actinosynnema pretiosum* subsp. *auranticum* and bacterial endophytic communities was also done following previously established protocol.[1]

For the analysis of halogenase gene, two successive PCR amplifications were performed using primers described previously.[2,3] The PCR amplifications were performed in a 50 µL reaction mixture containing 45 µL Red Taq DNA Polymerase Master Mix (1.1x), 0.5 µL forward primer (100 µM), 0.5 µL reverse primer (100 µM), 3 µL template DNA and 1 µL of sterile double distilled water. The first PCR amplification with primers P1 (designed from *asm12* gene of *Actinosynnema pretiosum* subsp. *auranticum*) consisted of an initial denaturation at 95°C for 2 min, 30 cycles of denaturation, annealing and elongation at 95°C for 30 s, 63°C for 40 s and 72°C for 30 s. This was followed by a final elongation step at 72°C for 5 min. The genomic DNA of endophytic communities and plant tissues were subjected to PCR amplification. As a negative control, the template DNA was replaced by sterile double distilled water. As a positive control, the genomic DNA of *A. pretiosum* subsp. *auranticum* (DSM 44131) was used.

The second PCR amplification with primers P2 (designed from the conserved regions of the $FADH_2$-dependent halogenases) consisted of an initial denaturation at 95°C for 2 min, 30 cycles of denaturation, annealing and elongation at 95°C for 30 s, 62°C for 40 s and 72°C for 30 s. This was followed by a final elongation step at 72°C for 5 min. The genomic DNA of endophytic communities and plant tissues were subjected to PCR amplification. As a positive control, the PCR purified product (*asm12* gene; spanning around 1300 bp) of *A. pretiosum* subsp. *auranticum* was used. Simultaneously, the 1300 bp fragment of bark endophytic community (purified from first PCR amplification) was used as a template for this second PCR. As a negative control, the template DNA was replaced by sterile double distilled water.

The PCR amplified products spanning around 1300 bp (first PCR) and 550 bp (second PCR) was checked and confirmed by agarose gel electrophoresis. The PCR products were further extracted from the agarose gel and purified using GFX^{TM} PCR DNA and Gel Band Purification kit (GE Healthcare Life Sciences, Germany) strictly following manufacturer's instructions. The amplified products were then sequenced from both directions at Seqlab Sequence Laboratories (Göttingen, Germany).

Phylogenetic evaluation and detection of AHBA synthase and halogenase gene in different tissues of host plant *M. serrata* and the endophytic bacterial communities

The phylogenetic evaluation of the bacterial endophytic community and detection of purified PCR products of AHBA synthase gene sequences (approx. 755 bp) and halogenase gene sequences (approx. 1300 bp and 550 bp) were aligned and evaluated following previously established methods.[1] Briefly, for the phylogenetic evaluation, the sequences were aligned using the EMBOSS-Needle Pairwise Sequence Alignment creating an optimal global end-to-end alignment of the sequences based on Needleman-Wunsch algorithm. This was further subjected to multiple sequence alignment for deducing the phylogenetic relationships using the T-Coffee Multiple Alignment software. The alignments were used to construct a phenogram-like rooted phylogenetic tree diagram using Drawgram in PHYLIP 3.66 software with centered ancestral nodes. For the detection of the genes, the aligned sequences were matched against the public nucleotide database using the Basic Local Alignment Search Tool (BLASTn) of the US National Centre for Biotechnology Information (NCBI) for identification of the genes. The similarity of the community sequences with phylogenetically-related reference sequences were identified using the EMBL-European Nucleotide Archive (ENA). The coding sequences were further translated into protein sequences using BLASTx of UniProt Knowledgebase (UniProtKB). The maximum homology to the respective protein in each case was identified using the UniProtKB identifier. The sequences of all the positive PCR purified products have been deposited at EMBL-Bank.

Fermentation of endophytes

100 µL of the frozen stocks of the endophytic communities from different *M. serrata* tissues where incubated with 300 mL GYM *Streptomyces* broth (4 g glucose, 4 g yeast extract, 10 g malt extract, 1 L double-distilled water, adjusted to pH 7.2) for 24 h at 28 ± 2°C with shaking (150 rev min^{-1}) on a rotary shaker (INFORS HT Multitron 2, Einsbach, Germany). The fermentation culture was extracted 3 times using 100 mL ethyl acetate. The combined extracts were dried on a rotary evaporator (Laborota 4001, Heidolph, Schwabach, Germany) at 40°C (using a water bath) at 100 rpm and dissolved in ethanol for HPLC-HRMS measurements. Appropriate medium was used for the fermentation of the bacteria (nutrient broth) and the fungi (potato dextrose broth).

Preparative HPLC

Crude plant extracts were fractionated and purified by preparative HPLC on a Gilson (Middleton, USA) HPLC-system including a pump (322) and a UV/VIS-152 detector with a Nucleodur C18 Isis column (5 µm particle size, 250 x 16 mm, Macherey-Nagel, Düren, Germany) with a water (A) and methanol (B) gradient (flow rate 4 mL/min).

The gradient was set up as follows: 30% B isocratic for 2 min, linear gradient from 30% B to 100% B over 11 min, 100% B isocratic for 5 min, within 0.5 min the system returned back to initial conditions of 30% B and was equilibrated for 6.5 min.

HPLC-ESI-HRMS^n of plant extracts and endophytic fermentation extracts

HPLC-HRMS^n measurements of crude plant extracts, purified/enriched fractions, and endophytic fermentation extracts were carried out using our previously established procedures,[1] suitably modified, on a LTQ Orbitrap spectrometer (Thermo Scientific, USA) equipped with a HESI-II source. The spectrometer was operated in positive mode with a nominal mass resolving power of 60000 at *m/z* 400 with a scan rate of 1 Hz under following parameters: spray voltage 5 kV, capillary temperature 300°C, tube lens 100 V. He served as collision gas and N_2 was used as sheath gas (55 arbitrary units) and auxiliary gas (8 arbitrary units). N-butyl benzenesulfonamide was used as lock mass ($[M+H]^+$ *m/z* 214.08963). All MS^n measurements were performed with collision-induced dissociation with 35 eV. The spectrometer was equipped with an Agilent (Santa Clara, USA) 1200 HPLC system including pump, PDA detector, column oven (30°C) and auto-sampler. The separations were performed with a Luna C18 (2) column (50 x 3 mm, 3 µm particle size) from Phenomenex (Torrance, USA) with a H_2O (+ 0.1% HCOOH) (A) and CH_3CN (+ 0.1% HCOOH) (B) gradient (flow rate 300 µL min^{-1}). The gradient program was as follows: 70% A isocratic for 2 min, linear gradient to 80% B over 8.5 min, linear gradient to 100% B for 0.5 min, 100% B isocratic for 3 min, the system returned within 0.5 min to initial conditions of 70% A and was equilibrated for 6.5 min.

III. SUPPLEMENTARY FIGURES

Fig. S1. Proposed mass spectral fragmentation pathway of maytansine.

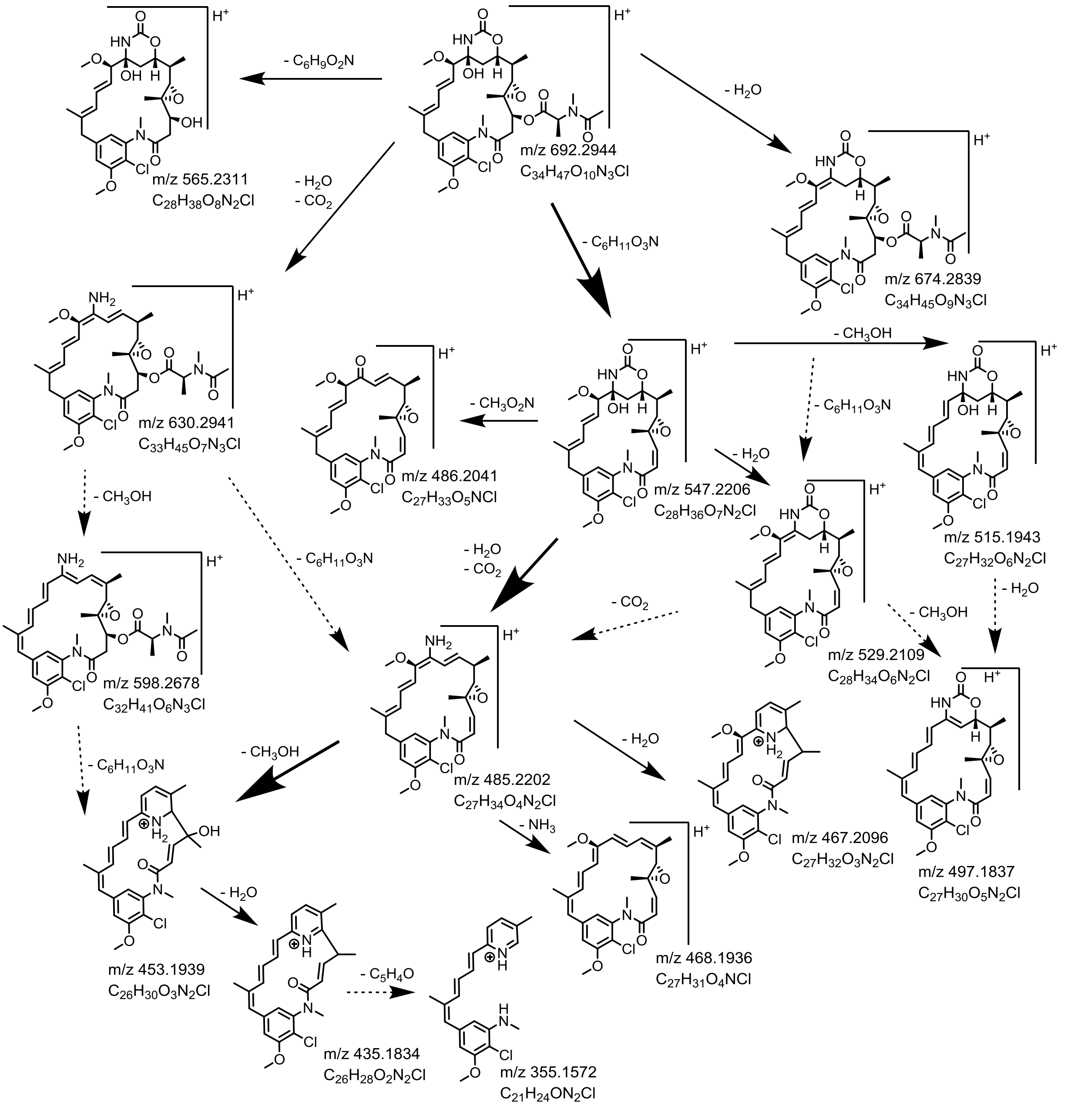
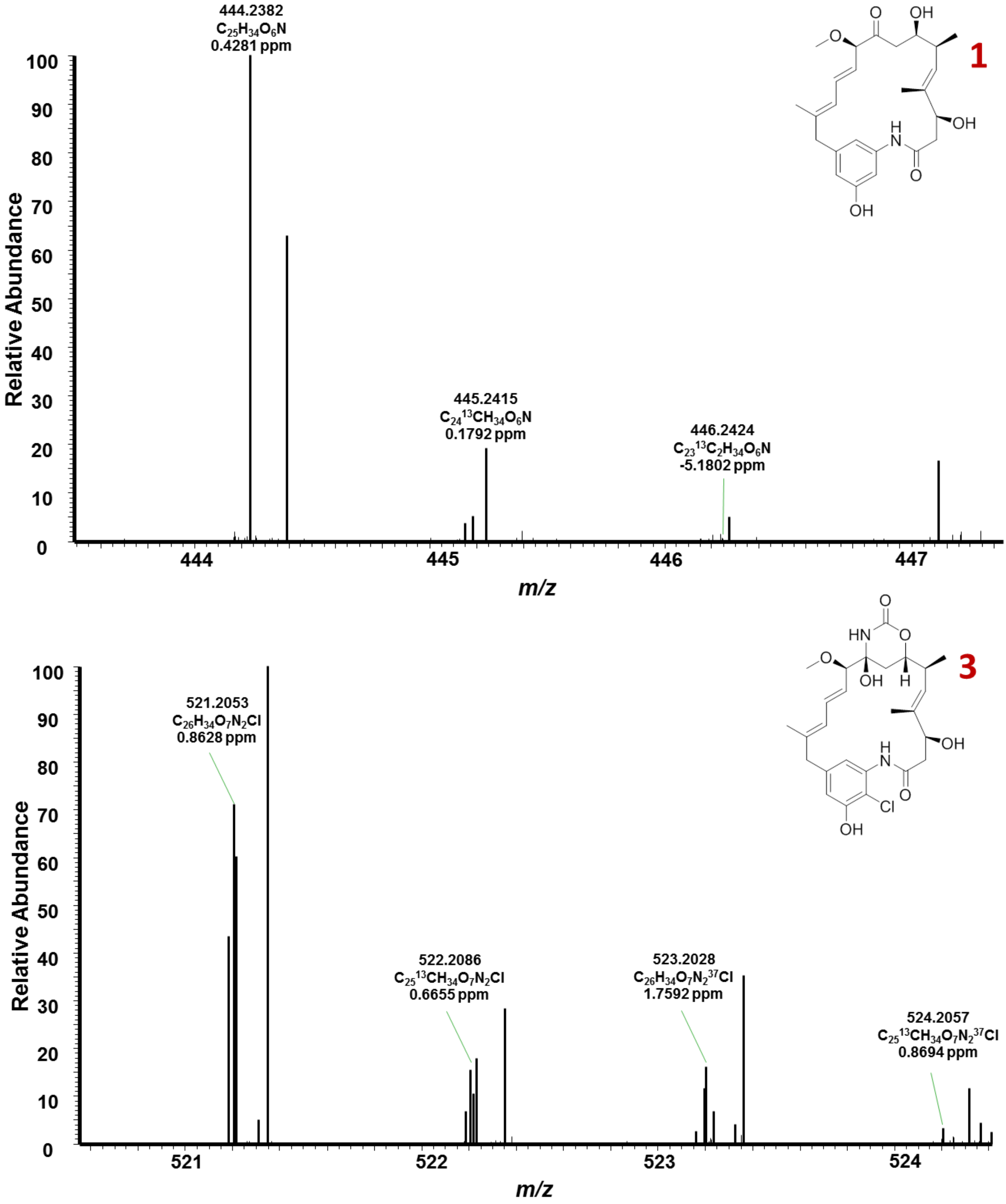
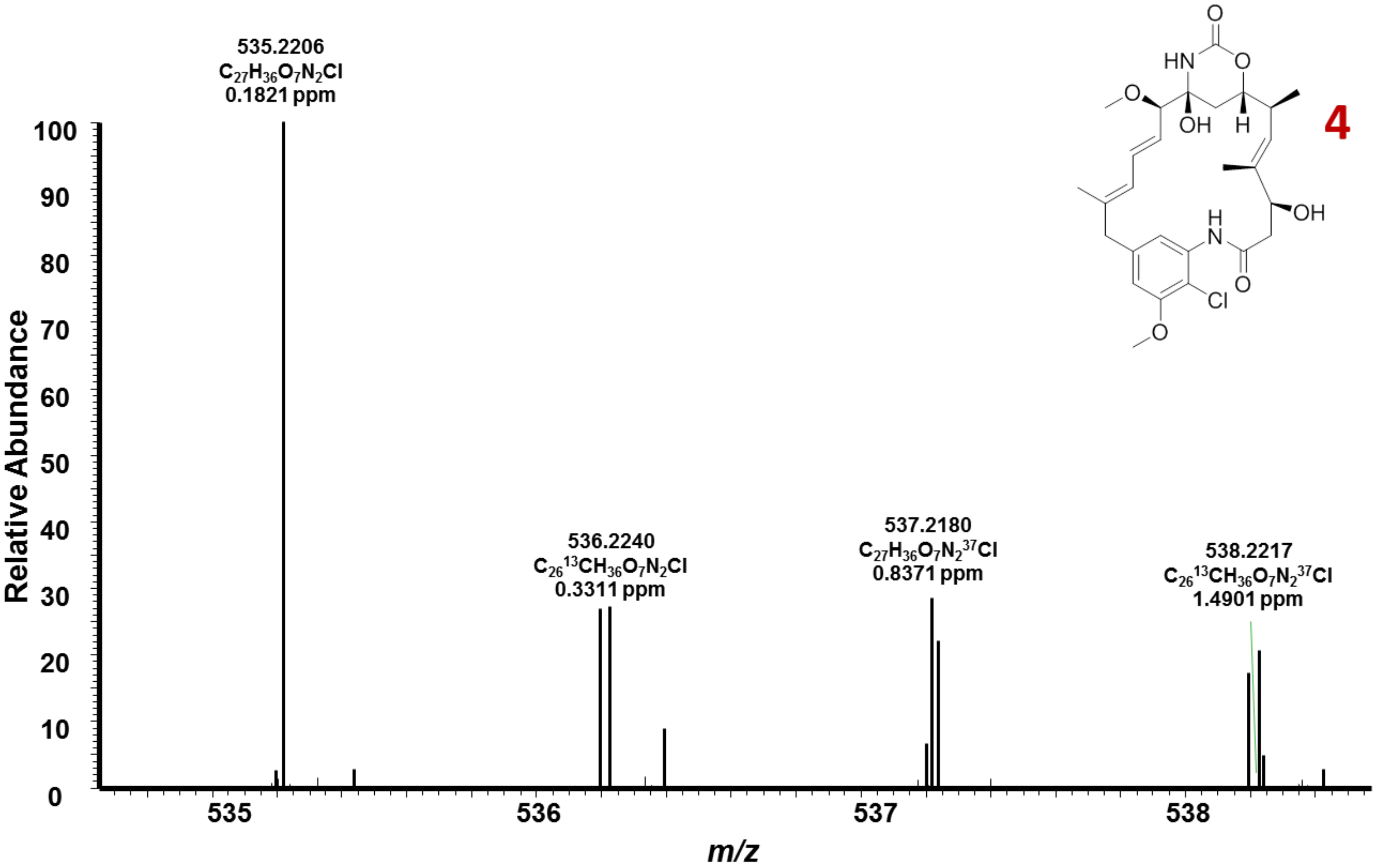
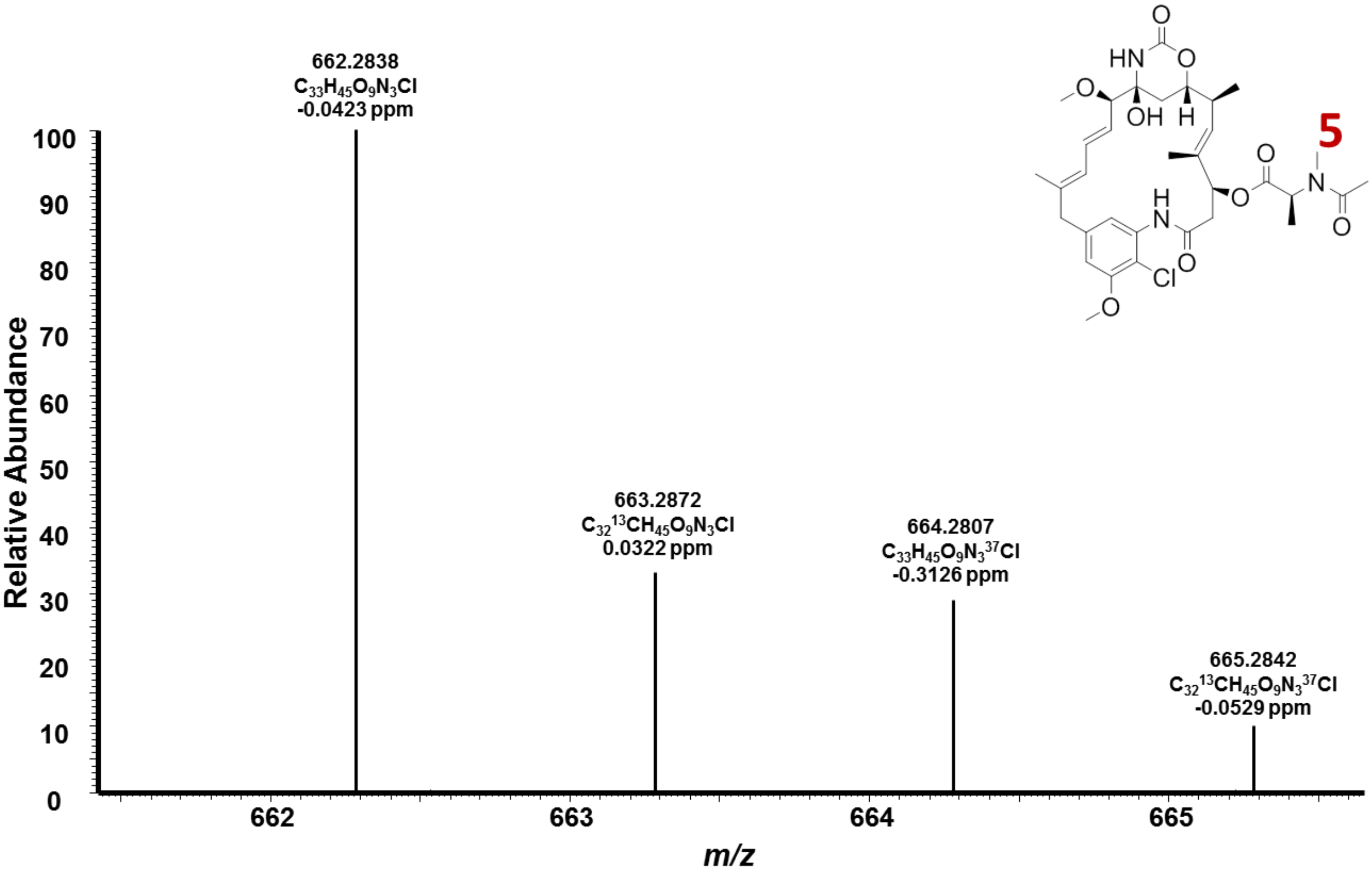


Fig. S2. Representative HPLC-ESI-HRMS full scan mass spectra of precursors of maytansine detected in the host plant (see also Table S3).




535.2206
$C_{27}H_{36}O_7N_2Cl$
0.1821 ppm
536.2240
$C_{26}{}^{13}CH_{36}O_7N_2Cl$
0.3311 ppm
537.2180
$C_{27}H_{36}O_7N_2{}^{37}Cl$
0.8371 ppm
538.2217
$C_{26}{}^{13}CH_{36}O_7N_2{}^{37}Cl$
1.4901 ppm
Relative Abundance
100
90
80
70
60
50
40
30
20
10
0
535
536
537
538
m/z
O
HN
O
O
OH
H
OH
H
N
O
Cl
O
4


662.2838
$C_{33}H_{45}O_9N_3Cl$
-0.0423 ppm
663.2872
$C_{32}{}^{13}CH_{45}O_9N_3Cl$
0.0322 ppm
664.2807
$C_{33}H_{45}O_9N_3{}^{37}Cl$
-0.3126 ppm
665.2842
$C_{32}{}^{13}CH_{45}O_9N_3{}^{37}Cl$
-0.0529 ppm
Relative Abundance
100
90
80
70
60
50
40
30
20
10
0
662
663
664
665
m/z
O
HN
O
O
OH
H
O
O
N
O
H
N
O
Cl
O
5

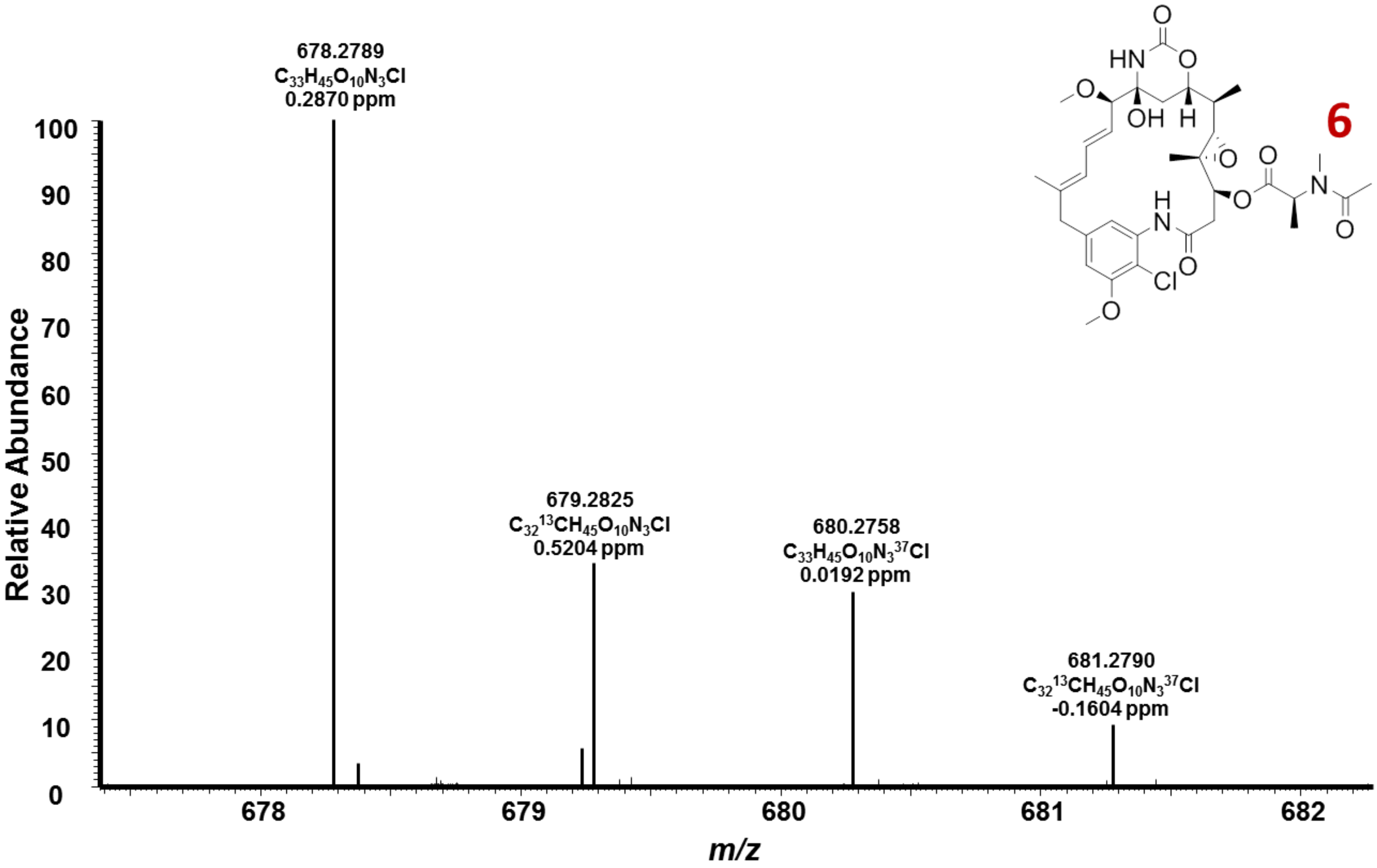
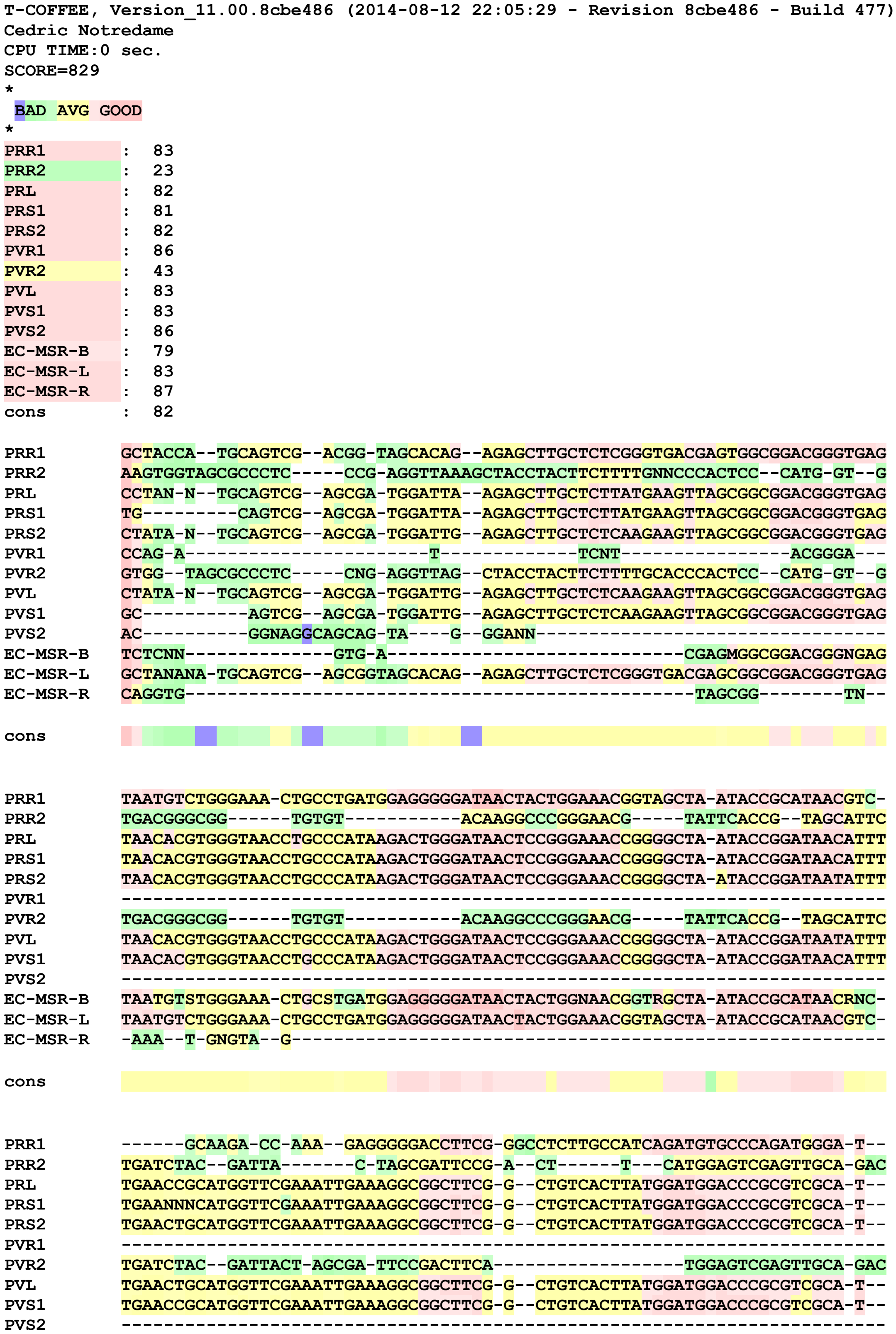
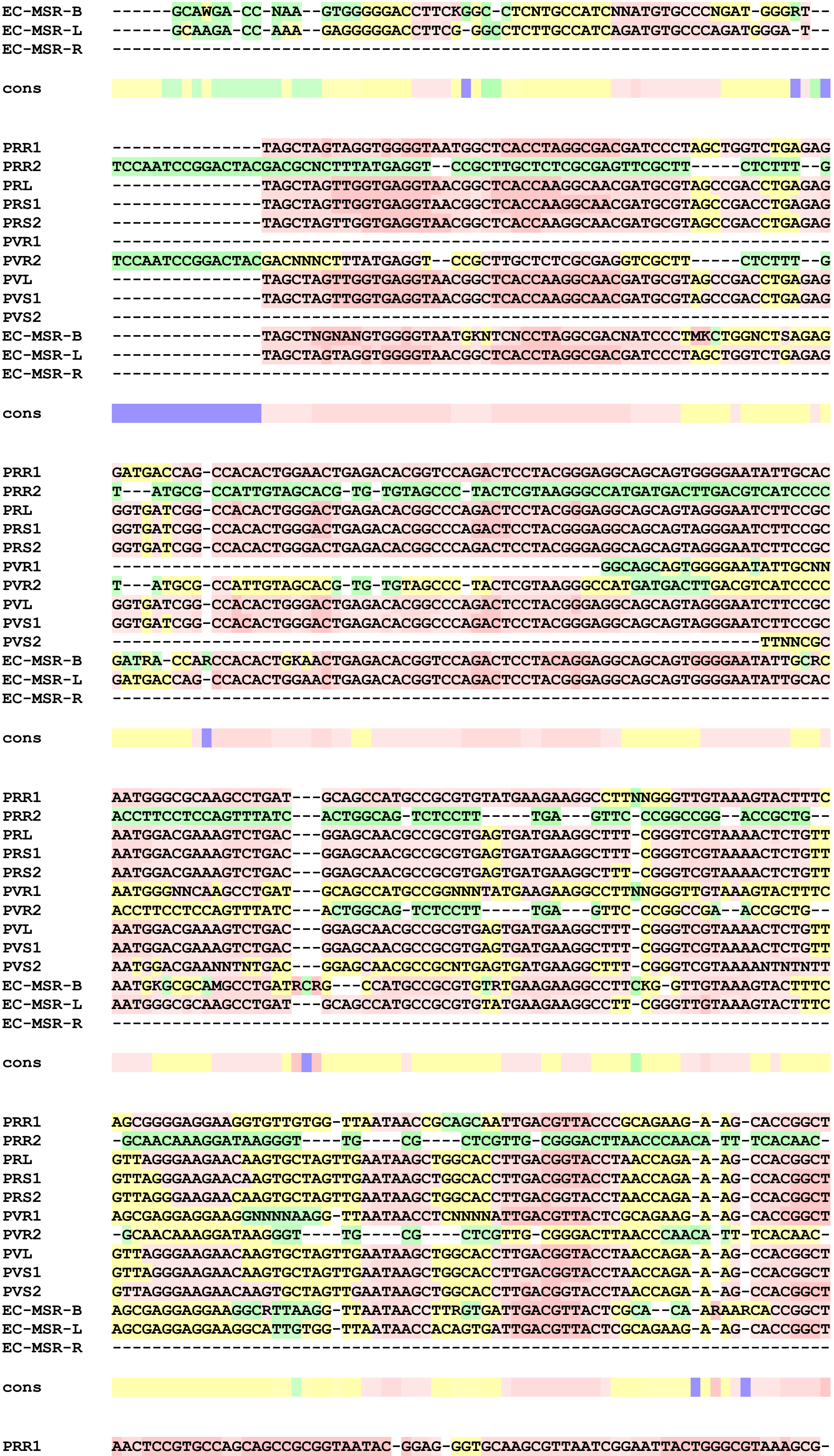

6
678.2789
$C_{33}H_{45}O_{10}N_3Cl$
0.2870 ppm
679.2825
$C_{32}{}^{13}CH_{45}O_{10}N_3Cl$
0.5204 ppm
680.2758
$C_{33}H_{45}O_{10}N_3{}^{37}Cl$
0.0192 ppm
681.2790
$C_{32}{}^{13}CH_{45}O_{10}N_3{}^{37}Cl$
-0.1604 ppm
Relative Abundance
0
10
20
30
40
50
60
70
80
90
100
678
679
680
681
682
m/z
HN
O
OH
H
Cl
N
H
O

Fig. S3. The multiple sequence alignment of bacterial endophytic communities of the present study (EC-MSR-B, EC-MSR-L, EC-MSR-R), compared to ones reported earlier from *Putterlickia* plants,[1] the using T-Coffee, combined into one final alignment, based on their 16S rRNA regions.



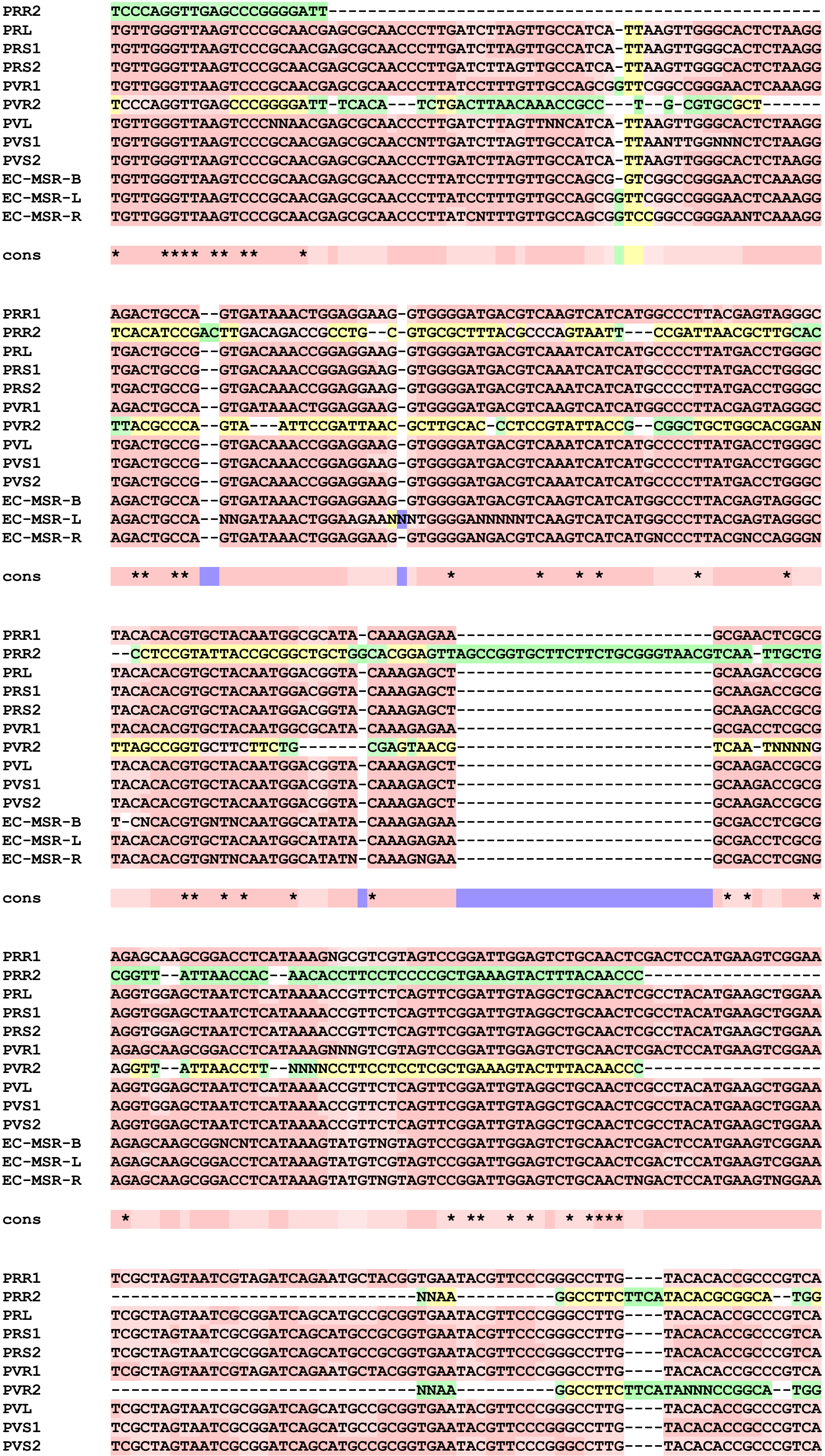
```
T-COFFEE, Version_11.00.8cbe486 (2014-08-12 22:05:29 - Revision 8cbe486 - Build 477)
Cedric Notredame
CPU TIME:0 sec.
SCORE=829
*
 BAD AVG GOOD
*
PRR1          :  83
PRR2          :  23
PRL           :  82
PRS1          :  81
PRS2          :  82
PVR1          :  86
PVR2          :  43
PVL           :  83
PVS1          :  83
PVS2          :  86
EC-MSR-B      :  79
EC-MSR-L      :  83
EC-MSR-R      :  87
cons          :  82

PRR1          GCTACCA--TGCAGTCG--ACGG-TAGCACAG--AGAGCTTGCTCTCGGGTGACGAGTGGCGGACGGGTGAG
PRR2          AAGTGGTAGCGCCCTC-----CCG-AGGTTAAAGCTACCTACTTCTTTTGNNCCCACTCC--CATG-GT--G
PRL           CCTAN-N--TGCAGTCG--AGCGA-TGGATTA--AGAGCTTGCTCTTATGAAGTTAGCGGCGGACGGGTGAG
PRS1          TG---------CAGTCG--AGCGA-TGGATTA--AGAGCTTGCTCTTATGAAGTTAGCGGCGGACGGGTGAG
PRS2          CTATA-N--TGCAGTCG--AGCGA-TGGATTG--AGAGCTTGCTCTCAAGAAGTTAGCGGCGGACGGGTGAG
PVR1          CCAG-A-----------------------T-------------TCNT----------------ACGGGA---
PVR2          GTGG--TAGCGCCCTC-----CNG-AGGTTAG--CTACCTACTTCTTTTGCACCCACTCC--CATG-GT--G
PVL           CTATA-N--TGCAGTCG--AGCGA-TGGATTG--AGAGCTTGCTCTCAAGAAGTTAGCGGCGGACGGGTGAG
PVS1          GC----------AGTCG--AGCGA-TGGATTG--AGAGCTTGCTCTCAAGAAGTTAGCGGCGGACGGGTGAG
PVS2          AC----------GGNAGGCAGCAG-TA----G--GGANN---------------------------------
EC-MSR-B      TCTCNN--------------GTG-A----------------------------CGAGMGGCGGACGGGNGAG
EC-MSR-L      GCTANANA-TGCAGTCG--AGCGGTAGCACAG--AGAGCTTGCTCTCGGGTGACGAGCGGCGGACGGGTGAG
EC-MSR-R      CAGGTG------------------------------------------------TAGCGG--------TN--

cons


PRR1          TAATGTCTGGGAAA-CTGCCTGATGGAGGGGGATAACTACTGGAAACGGTAGCTA-ATACCGCATAACGTC-
PRR2          TGACGGGCGG------TGTGT-----------ACAAGGCCCGGGAACG-----TATTCACCG--TAGCATTC
PRL           TAACACGTGGGTAACCTGCCCATAAGACTGGGATAACTCCGGGAAACCGGGGCTA-ATACCGGATAACATTT
PRS1          TAACACGTGGGTAACCTGCCCATAAGACTGGGATAACTCCGGGAAACCGGGGCTA-ATACCGGATAACATTT
PRS2          TAACACGTGGGTAACCTGCCCATAAGACTGGGATAACTCCGGGAAACCGGGGCTA-ATACCGGATAATATTT
PVR1          ------------------------------------------------------------------------
PVR2          TGACGGGCGG------TGTGT-----------ACAAGGCCCGGGAACG-----TATTCACCG--TAGCATTC
PVL           TAACACGTGGGTAACCTGCCCATAAGACTGGGATAACTCCGGGAAACCGGGGCTA-ATACCGGATAATATTT
PVS1          TAACACGTGGGTAACCTGCCCATAAGACTGGGATAACTCCGGGAAACCGGGGCTA-ATACCGGATAACATTT
PVS2          ------------------------------------------------------------------------
EC-MSR-B      TAATGTSTGGGAAA-CTGCSTGATGGAGGGGGATAACTACTGGNAACGGTRGCTA-ATACCGCATAACRNC-
EC-MSR-L      TAATGTCTGGGAAA-CTGCCTGATGGAGGGGGATAACTACTGGAAACGGTAGCTA-ATACCGCATAACGTC-
EC-MSR-R      -AAA--T-GNGTA--G--------------------------------------------------------

cons


PRR1          ------GCAAGA-CC-AAA--GAGGGGGACCTTCG-GGCCTCTTGCCATCAGATGTGCCCAGATGGGA-T--
PRR2          TGATCTAC--GATTA-------C-TAGCGATTCCG-A--CT------T---CATGGAGTCGAGTTGCA-GAC
PRL           TGAACCGCATGGTTCGAAATTGAAAGGCGGCTTCG-G--CTGTCACTTATGGATGGACCCGCGTCGCA-T--
PRS1          TGAANNNCATGGTTCGAAATTGAAAGGCGGCTTCG-G--CTGTCACTTATGGATGGACCCGCGTCGCA-T--
PRS2          TGAACTGCATGGTTCGAAATTGAAAGGCGGCTTCG-G--CTGTCACTTATGGATGGACCCGCGTCGCA-T--
PVR1          ------------------------------------------------------------------------
PVR2          TGATCTAC--GATTACT-AGCGA-TTCCGACTTCA------------------TGGAGTCGAGTTGCA-GAC
PVL           TGAACTGCATGGTTCGAAATTGAAAGGCGGCTTCG-G--CTGTCACTTATGGATGGACCCGCGTCGCA-T--
PVS1          TGAACCGCATGGTTCGAAATTGAAAGGCGGCTTCG-G--CTGTCACTTATGGATGGACCCGCGTCGCA-T--
PVS2          ------------------------------------------------------------------------
```



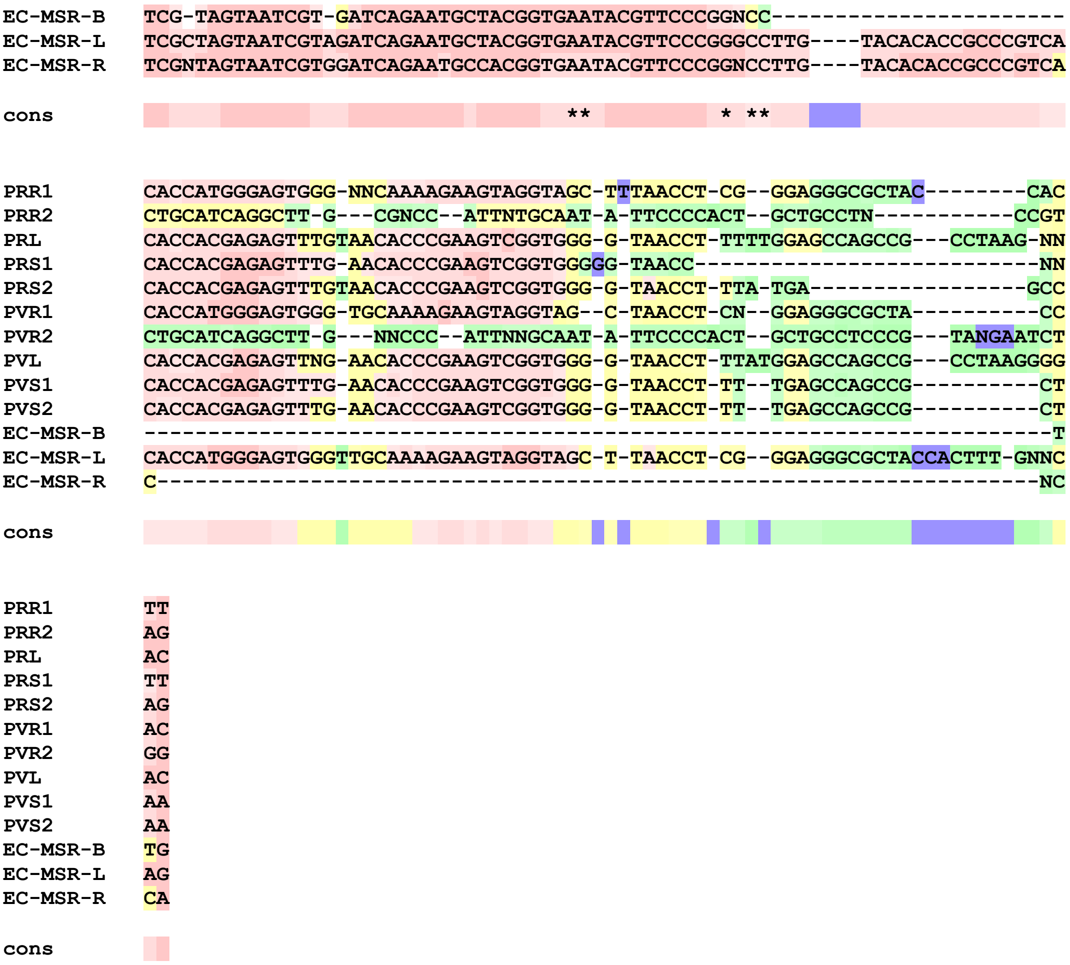
```
EC-MSR-B    ------GCAWGA-CC-NAA--GTGGGGGACCTTCKGGC-CTCNTGCCATCNNATGTGCCCNGAT-GGGRT--
EC-MSR-L    ------GCAAGA-CC-AAA--GAGGGGGACCTTCG-GGCCTCTTGCCATCAGATGTGCCCAGATGGGA-T--
EC-MSR-R    ------------------------------------------------------------------------

cons


PRR1        ---------------TAGCTAGTAGGTGGGGTAATGGCTCACCTAGGCGACGATCCCTAGCTGGTCTGAGAG
PRR2        TCCAATCCGGACTACGACGCNCTTTATGAGGT--CCGCTTGCTCTCGCGAGTTCGCTT-----CTCTTT--G
PRL         ---------------TAGCTAGTTGGTGAGGTAACGGCTCACCAAGGCAACGATGCGTAGCCGACCTGAGAG
PRS1        ---------------TAGCTAGTTGGTGAGGTAACGGCTCACCAAGGCAACGATGCGTAGCCGACCTGAGAG
PRS2        ---------------TAGCTAGTTGGTGAGGTAACGGCTCACCAAGGCAACGATGCGTAGCCGACCTGAGAG
PVR1        ------------------------------------------------------------------------
PVR2        TCCAATCCGGACTACGACNNNCTTTATGAGGT--CCGCTTGCTCTCGCGAGGTCGCTT-----CTCTTT--G
PVL         ---------------TAGCTAGTTGGTGAGGTAACGGCTCACCAAGGCAACGATGCGTAGCCGACCTGAGAG
PVS1        ---------------TAGCTAGTTGGTGAGGTAACGGCTCACCAAGGCAACGATGCGTAGCCGACCTGAGAG
PVS2        ------------------------------------------------------------------------
EC-MSR-B    ---------------TAGCTNGNANGTGGGGTAATGKNTCNCCTAGGCGACNATCCCTMKCTGGNCTSAGAG
EC-MSR-L    ---------------TAGCTAGTAGGTGGGGTAACGGCTCACCTAGGCGACGATCCCTAGCTGGTCTGAGAG
EC-MSR-R    ------------------------------------------------------------------------

cons


PRR1        GATGACCAG-CCACACTGGAACTGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCAC
PRR2        T---ATGCG-CCATTGTAGCACG-TG-TGTAGCCC-TACTCGTAAGGGCCATGATGACTTGACGTCATCCCC
PRL         GGTGATCGG-CCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTAGGGAATCTTCCGC
PRS1        GGTGATCGG-CCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTAGGGAATCTTCCGC
PRS2        GGTGATCGG-CCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTAGGGAATCTTCCGC
PVR1        -------------------------------------------------GGCAGCAGTGGGGAATATTGCNN
PVR2        T---ATGCG-CCATTGTAGCACG-TG-TGTAGCCC-TACTCGTAAGGGCCATGATGACTTGACGTCATCCCC
PVL         GGTGATCGG-CCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTAGGGAATCTTCCGC
PVS1        GGTGATCGG-CCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTAGGGAATCTTCCGC
PVS2        -----------------------------------------------------------------TTNNCGC
EC-MSR-B    GATRA-CCARCCACACTGKAACTGAGACACGGTCCAGACTCCTACAGGAGGCAGCAGTGGGGAATATTGCRC
EC-MSR-L    GATGACCAG-CCACACTGGAACTGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCAC
EC-MSR-R    ------------------------------------------------------------------------

cons


PRR1        AATGGGCGCAAGCCTGAT---GCAGCCATGCCGCGTGTATGAAGAAGGCCTTNNGGGTTGTAAAGTACTTTC
PRR2        ACCTTCCTCCAGTTTATC---ACTGGCAG-TCTCCTT-----TGA---GTTC-CCGGCCGG--ACCGCTG--
PRL         AATGGACGAAAGTCTGAC---GGAGCAACGCCGCGTGAGTGATGAAGGCTTT-CGGGTCGTAAAACTCTGTT
PRS1        AATGGACGAAAGTCTGAC---GGAGCAACGCCGCGTGAGTGATGAAGGCTTT-CGGGTCGTAAAACTCTGTT
PRS2        AATGGACGAAAGTCTGAC---GGAGCAACGCCGCGTGAGTGATGAAGGCTTT-CGGGTCGTAAAACTCTGTT
PVR1        AATGGGNNCAAGCCTGAT---GCAGCCATGCCGGNNNTATGAAGAAGGCCTTNNGGGTTGTAAAGTACTTTC
PVR2        ACCTTCCTCCAGTTTATC---ACTGGCAG-TCTCCTT-----TGA---GTTC-CCGGCCGA--ACCGCTG--
PVL         AATGGACGAAAGTCTGAC---GGAGCAACGCCGCGTGAGTGATGAAGGCTTT-CGGGTCGTAAAACTCTGTT
PVS1        AATGGACGAAAGTCTGAC---GGAGCAACGCCGCGTGAGTGATGAAGGCTTT-CGGGTCGTAAAACTCTGTT
PVS2        AATGGACGAANNTNTGAC---GGAGCAACGCCGCNTGAGTGATGAAGGCTTT-CGGGTCGTAAAANTNTNTT
EC-MSR-B    AATGKGCGCAMGCCTGATRCRG---CCATGCCGCGTGTRTGAAGAAGGCCTTCKG-GTTGTAAAGTACTTTC
EC-MSR-L    AATGGGCGCAAGCCTGAT---GCAGCCATGCCGCGTGTATGAAGAAGGCCTT-CGGGTTGTAAAGTACTTTC
EC-MSR-R    ------------------------------------------------------------------------

cons


PRR1        AGCGGGGAGGAAGGTGTTGTGG-TTAATAACCGCAGCAATTGACGTTACCCGCAGAAG-A-AG-CACCGGCT
PRR2        -GCAACAAAGGATAAGGGT----TG----CG----CTCGTTG-CGGGACTTAACCCAACA-TT-TCACAAC-
PRL         GTTAGGGAAGAACAAGTGCTAGTTGAATAAGCTGGCACCTTGACGGTACCTAACCAGA-A-AG-CCACGGCT
PRS1        GTTAGGGAAGAACAAGTGCTAGTTGAATAAGCTGGCACCTTGACGGTACCTAACCAGA-A-AG-CCACGGCT
PRS2        GTTAGGGAAGAACAAGTGCTAGTTGAATAAGCTGGCACCTTGACGGTACCTAACCAGA-A-AG-CCACGGCT
PVR1        AGCGAGGAGGAAGGNNNNAAGG-TTAATAACCTCNNNNATTGACGTTACTCGCAGAAG-A-AG-CACCGGCT
PVR2        -GCAACAAAGGATAAGGGT----TG----CG----CTCGTTG-CGGGACTTAACCCAACA-TT-TCACAAC-
PVL         GTTAGGGAAGAACAAGTGCTAGTTGAATAAGCTGGCACCTTGACGGTACCTAACCAGA-A-AG-CCACGGCT
PVS1        GTTAGGGAAGAACAAGTGCTAGTTGAATAAGCTGGCACCTTGACGGTACCTAACCAGA-A-AG-CCACGGCT
PVS2        GTTAGGGAAGAACAAGTGCTAGTTGAATAAGCTGGCACCTTGACGGTACCTAACCAGA-A-AG-CCACGGCT
EC-MSR-B    AGCGAGGAGGAAGGCRTTAAGG-TTAATAACCTTRGTGATTGACGTTACTCGCA--CA-ARAARCACCGGCT
EC-MSR-L    AGCGAGGAGGAAGGCATTGTGG-TTAATAACCACAGTGATTGACGTTACTCGCAGAAG-A-AG-CACCGGCT
EC-MSR-R    ------------------------------------------------------------------------

cons


PRR1        AACTCCGTGCCAGCAGCCGCGGTAATAC-GGAG-GGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCG-
```

```
PRR2        -ACGAGCTGACGACAGCCATG-CAGCAC-CTG---TCTC-AGAGTT--CCCGAA------GGCACCAATCC-
PRL         AACTACGTGCCAGCAGCCGCGGTAATAC-GTAG-GTGGCAAGCGTTATCCGGAATTATTGGGCGTAAAGCG-
PRS1        AACTACGTGCCAGCAGCCGCGGTAATAC-GTAG-GTGGCAAGCGTTATCCGGAATTATTGGGCGTAAAGCGC
PRS2        AACTACGTGCCAGCAGCCGCGGTAATAC-GTAG-GTGGCAAGCGTTATCCGGAATTATTGGGCGTAAAGCG-
PVR1        AANTCCGTGCCAGCAGCCGCGGTAATAC-GGAG-GGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCG-
PVR2        -ACGAGCTGACGACAGCCATG-CAGCAC-CTG---TCTC-AGAGTT--CCCGAA------GGCACCAATCC-
PVL         AACTACGTGCCAGCAGCCGCGGTAATAC-GTAG-GTGGCAAGCGTTATCCGGAATTATTGGGCGTAAAGCG-
PVS1        AACTACGTGCCAGCAGCCGCGGTAATAC-GTAG-GTGGCAAGCGTTATCCGGAATTATTGGGCGTAAAGCG-
PVS2        AACTACGTGCCAGCAGCCGCGGTAATAC-GTAG-GTGGCAAGCGTTATCCGGAATTATTGGGCGTAAAGCG-
EC-MSR-B    AACTCCGTGCCAGCAGCCGCGGTAATACRG-AGRGT-GCAAGCGTTAATCGSAATTACTGGGCGTAAAGCG-
EC-MSR-L    AACTCCGTGCCAGCAGCCGCGGTAATAC-GGAG-GGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCG-
EC-MSR-R    ------------------------------------------------------------------------

cons


PRR1        -------------------------------------------------CACGCAGGCGGTCTGTC-AAGTC
PRR2        -------------------------------------------------ATCTCTGGAAAGTTCTCTGGATG
PRL         -------------------------------------------------CGCGCAGGTGGTTTCTT-AAGTC
PRS1        GCGCAGGTGGTTTCTTAAGTCGTTATCCGGAATTATTGGGCGTAAAGCGCGCGCAGGTGGTTTCTT-AAGTC
PRS2        -------------------------------------------------CGCGCAGGTGGTTTCTT-AAGTC
PVR1        -------------------------------------------------CACGCAGGCGGTTTGTT-AAGTC
PVR2        -------------------------------------------------ATCTCTGGAAAGTTCTCTGGATG
PVL         -------------------------------------------------CGCGCAGGTGGTTTCTT-AAGTC
PVS1        -------------------------------------------------CGCGCAGGTGGTTTCTT-AAGTC
PVS2        -------------------------------------------------CGCGCAGGTGGTTTCTT-AAGTC
EC-MSR-B    -------------------------------------------------CACRCAGGCGGTCTGTC-AAGTC
EC-MSR-L    -------------------------------------------------CACGCAGGCGGTCTGTC-AAGTC
EC-MSR-R    ------------------------------------------------------------------------

cons


PRR1        GGATGTGAAATC-CCCGGGCTCAACCTGG-GAACTGCATTCGAAACTGGCAGGCTAGAGTCTTGTAGAGGGG
PRR2        TCAAGAGTAGG---TAAGGTTCT-TCGCG-TTGCATCGAATTAAACC-ACATGCTCCACCGCTTGTGCGG--
PRL         TGATGTGAAAGC-CCACGGCTCAACCGTG-GAGGGTCATTGGAAACTGGGAGACTTGAGTGCAGAAGAGGAA
PRS1        TGATGTGAAAGC-CCACGGCTCAACCGTG-GAGGGTCATTGGAAACTGGGAGACTTGAGTGCAGAAGAGGAA
PRS2        TGATGTGAAAGC-CCACGGCTCAACCGTG-GAGGGTCATTGGAAACTGGGAGACTTGAGTGCAGAAGAGGAA
PVR1        AGATGTGAAATC-CCCGGGCTCAACCTGG-GAACTGCATTTGAAACTGGCAAGCTAGAGTCTTGTAGAGGGG
PVR2        TCAAGAGTAGG---TAAGGTTCT-TCGCG-TTGCATCGAATTAAACC-ACATGCTCCACCGCTTGTGCGG--
PVL         TGATGTGAAAGC-CCACGGCTCAACCGTG-GAGGGTCATTGGAAACTGGGAGACTTGAGTGCAGAAGAGGAA
PVS1        TGATGTGAAAGC-CCACGGCTCAACCGTG-GAGGGTCATTGGAAACTGGGAGACTTGAGTGCAGAAGAGGAA
PVS2        TGATGTGAAAGC-CCACGGCTCAACCGTG-GAGGGTCATTGGAAACTGGGAGACTTGAGTGCAGAAGAGGAA
EC-MSR-B    KGATRTGAAATCYCC-GGGCTCWACC-TGKGAACTGCRTTCGAAACTGGCAGGCTATAGTCTTGTAKAGGGG
EC-MSR-L    GGATGTGAAATC-CCCGGGCTCAACCTGG-GAACTGCATTCGAAACTGGCAGGCTAGAGTCTTGTAGAGGGG
EC-MSR-R    ------------------------------------------------------------------------

cons


PRR1        GGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGATC--------TGGAGGAATACCGGTGG-CGAAGGC
PRR2        ---------------GCCCCCGTCAATTCATTTGAGTTTTAACCTTGCGGCCGTACTCCCCAGG-CG--GTC
PRL         AGTGGAATTCCATGTGTAGCGGTGAAATGCGTAGAGATA--------TGGAGGAACACCAGTGG-CGAAGGC
PRS1        AGTGGAATTCCATGTGTAGCGGTGAAATGCGTAGAGATA--------TGGAGGAACACCAGTGG-CGAAGGC
PRS2        AGTGGAATTCCATGTGTAGCGGTGAAATGCGTAGAGATA--------TGGAGGAACACCAGTGG-CGAAGGC
PVR1        GGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGATC--------TGGAGGAATACCGGTGG-CGAAGGC
PVR2        ---------------GCCCCCGTCAATTCATTTGAGTTTTAACCTTGCGGCCGTACTCCCCAGG-CG--GTC
PVL         AGTGGAATTCCATGTGTAGCGGTGAAATGCGTAGAGATA--------TGGAGGAACACCAGTGG-CGAAGGC
PVS1        AGTGGAATTCCATGTGTAGCGGTGAAATGCGTAGAGATA--------TGGAGGAACACCAGTGG-CGAAGGC
PVS2        AGTGGAATTCCATGTGTAGCGGTGAAATGCGTAGAGATA--------TGGAGGAACACCAGTGG-CGAAGGC
EC-MSR-B    GGTAGAATTCCAGGTGTAKCGGTGAAATGCGTAGAGATC--------TSGAGGAGTAYCGGT-GKCGAAAGC
EC-MSR-L    GGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGATC--------TGGAGGAATACCGGTGG-CGAAGGC
EC-MSR-R    ----------------------------------ANATN--------TGGAGGAATACCGGTGG-NGAAGGC

cons                                          *  *           *  *     *    *  *    *


PRR1        GGCCCCCTGGACAAAGACTGACGCTCAGGTGCGAAAGCGTGGGGAGC--ANACAGGATTAGATACCCTGGTA
PRR2        GACTTAA----------------------CGCGTTAGCTCCGGAAGCCACGCCTCAAGGGCACAACCTCCAA
PRL         GACTTTCTGGTCTGTAACTGACACTGAGGCGCGAAAGCGTGGGGAGC--ANACAGGATTAGATACCCTGGTA
PRS1        GACTTTCTGGTCTGTAACTGACACTGAGGCGCGAAAGCGTGGGGAGC--AAACAGGATTAGATACCCTGGTA
PRS2        GACTTTCTGGTCTGTAACTGACACTGAGGCGCGAAAGCGTGGGGAGC--ANACAGGATTAGATACCCTGGTA
PVR1        GGCCCCCTGGACNAAGACTGACGCTCAGGTGCGAAAGCGTGGGGAGC--AAACAGGATTAGATACCCTGGTA
PVR2        GACTTAA----------------------CGCGTTAGCTCCGGAAGCCACNCCTCAAGGGNACAACCTCCAA
PVL         GACTTTCTGGTCTGTAACTGACACTGAGGCGCGAAAGCGTGGGGAGC--ANACAGGATTAGATACCCTGGTA
PVS1        GACTTTCTGGTCTGTAACTGACACTGAGGCGCGAAAGCGTGGGGAGC--ANACAGGATTAGATACCCTGGTA
PVS2        GACTTTCTGGTCTGTAACTGACACTGAGGCGCGAAAGCGTGGGGAGC--AAACAGGATTAGATACCCTGGTA
```

```
EC-MSR-B      GGCCCCCTGGACAAAGACTGACGCTCAGGTGCGAAAGCGTGCGGAGC--A-ACAGGATTAGATACCCTGGTA
EC-MSR-L      GGCCCCCTGGACAAAGACTGACGCTCAGGTGCGAAAGCGTGGGGAGC--AAACAGGATTAGATACCCTGGTA
EC-MSR-R      GGCCTCNTGGACGAAGACTGACGCTCAGGTGNNAAAGCGTGGGGAGC--AAACAGGATTAGATACCNTGGTA

cons          * *                           *    ***    * ***     *   *    * * * *   *


PRR1          GTCCACGCCGTANACGATGTCGACTTGGAGGTTGTGCCCTTGAGGCG-TGGCTTCCGGAGCTAACGCGTTAA
PRR2          GTCGACATCGTTTACGGCGTGGACTACC------AGGGTATCTAATCCTGTTTGCT---CCCCACGCTTT-C
PRL           GTCCACGCCGTANACGATGAGTGCTANGTGTTAGAGGGTTTCCGCCCTTNAGTGCTGNAGTTAACGCANTAA
PRS1          GTCCACGCCGTAAACGATGAGTGCTAAGTGTTAGAGGGTTTCCGCCCTTTAGTGCTGAAGTTAACGCATTAA
PRS2          GTCCACGCCGTANACGATGAGTGCTAAGTGTTAGAGGGTTTCCGCCCTTTAGTGCTGAAGTTAACGCATTAA
PVR1          GTCCACGCCGTAAACGATGTCGACTTGGAGGTTGTNCCCTTGAGGNG-TGGCTTCCGGAGCTAACGCGTTAA
PVR2          GTCGACATCGTTTACGGCGTGGACTACC------AGGGTATCTAATCCTGTTTGCT---CCCCACGCTTT-C
PVL           GTCCACGCCGTANACGATGAGTGCTAAGTGTTAGAGGGTTTCCGCCCTTTAGTGCTGAAGTTAACGCATTAA
PVS1          GTCCACGCCGTANACGATGAGTGCTAAGTGTTAGAGGGTTTCCGCCCTTTAGTGCTGAAGTTAACGCATTAA
PVS2          GTCCACGCCGTAAACGATGAGTGCTAAGTGTTAGAGGGTTTCCGCCCTTTAGTGCTGAAGTTAACGCATTAA
EC-MSR-B      GTCCACGCYGTAAACGATGTCSACTTGGAGGTTGTGCCCTTGWGGCG-TGTCTTCCGGAGCTAACGCGTTAA
EC-MSR-L      GTCCACGCCGTANACGATGTCGACTTGGAGGTTGTTCCCTTGAGGAG-TGGCTTCCGGAGCTAACGCGTTAA
EC-MSR-R      GTCCACGCNGTAAACGATGTCTATTTGGAGGTTGTNCCCNTGAGGCG-TGGCTTCCGGAGCTAACGCGTTAA

cons          *** **   **  ***  *     *               *       *   * *        ****  *


PRR1          GTCGACCGCCTGGGGAGTACGGCCGCAAGGTTAAAACTCANATGAATTGA--CGGGGGCCCGCACAAGCGGT
PRR2          G-----CACCTGAG---------CGTCAGTCTT-------TG----TCCAGGGGGCCGCCTTCGCCACCGGT
PRL           GCACTCCGCCGGGGGAGTACGGCCGCAAGGCTGAAACTCAAAGGAATTGA--CGGGGGCCCGCACAAGCGGT
PRS1          GCACTCCGCCTGGGGAGTACGGCCGCAAGGCTGAAACTCAAAGGAATTGA--CGGGGGCCCGCACAAGCGGT
PRS2          GCACTCCGCCTGGGGAGTACGGCCGCAAGGCTGANACTCANAGGAATTGA--CGGGGGCCCGCACAAGCGGT
PVR1          GTCGACCGCCTGGGGAGTACGGCCGCAAGGTTAAAACTCAAATGAATTGA--CGGGGGCCCGCACAAGCGGT
PVR2          G-----CACCTGAG---------CGTCAGTCTT-------NG----TCCAGGGGGCCGCCTTCGCCACCGGT
PVL           GCACTCCGCCTGGGGAGTACGGCCGCAAGGCTGAAACTCANAGGAATTGA--CGGGGGCCCGCACAAGCGGT
PVS1          GCACTCCGCCTGGGGAGTACGGCCGCAAGGCTGAAACTCANAGGAATTGA--CGGGGGCCCGCACAAGCGGT
PVS2          GCACTCCGCCTGGGGAGTACGGCCGCAAGGCTGAAACTCAAAGGAATTGA--CGGGGGCCCGCACAAGCGGT
EC-MSR-B      GTCGACCGCCTGGGGAGTACCGCCCCAMGGTT-AAACTCAAATGAATTGW--CGGGGGCCCCCACAAGCGGT
EC-MSR-L      GTCGACCGCCTGGGGAGTACGGCCGCAAGGTTAAAACTCAAATGAATTGA--CGGGGGCCCGCACAAGCGGT
EC-MSR-R      ATAGACCGCCTGGGGAGTACGGCCGCAAGGTTAAAACTCAAATGAATTGA--CGGGGGCCCGCACAAGCGGT

cons                * ** * *         *    *  *              *      **  ***  * * * ****


PRR1          GGAGCATGTGGTTTAATTCGATGCAACGCGAANAACCTTACCTACTCTTGACATCCAGAGAACTTTCCANAG
PRR2          ATTC------------CTCCA--GATCTCTACGCATTTCACCG---CT--ACA-------------------
PRL           GGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACATCCTCTGAAAACCCTAGAG
PRS1          GGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACATCCTCTGAAAACCCTAGAG
PRS2          GGAGCATGTGGTTTAATTCGAAACAACGCGAAGAACCTTACCAGGTCTTGACATCCTCTGAAAACCCTAGAG
PVR1          GGAGCATGTGGTTTAATTCGATGCAACGCGAAGAACCTTACCTACTCTTGACATCCAGAGAACTTTCCAGAG
PVR2          ATTC------------CTCCA--GATCTCTACGCATTTCACCG---CT--ACA-------------------
PVL           GGANCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACATCCTCTGAAAACCCTAGAG
PVS1          GGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACATCCTCTGANAACCCTAGAG
PVS2          GGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACATCCTCTGAAAACCCTAGAG
EC-MSR-B      GGAGCATGTGGTTTAATTCGATGCAACGCGAAGAACCTTACCTACTCTTGACATCCAGAGAACTTAGCAGAG
EC-MSR-L      GGAGCATGTGGTTTAATTCGATGCAACGCGAAGAACCTTACCTACTCTTGACATCCAGAGAACTTAGCAGAG
EC-MSR-R      GGAGCATGTGGTTTAATTNGATGCAACGCGAAGAACCTTACCTNGTCTTGACATCCACAGAAGTTTTCAGAG

cons                           *  *   * * * *   *  * ***    **  ***


PRR1          ATGNNTTGGTGCCTTC-GGGAACTCTGAGACAGGTGCTGCATGGCTGTCGTCAGCTCGT----GTTGTGAAA
PRR2          --------------CC-TGGAATTCTACC--CCCCTCTACAAGACTCTAGCCTGCCAGTTTCGAATGC-AGT
PRL           ATAGGGCTTCTCCTTC-GGGAGCAGAGTGACAGGTGGTGCATGGTTGTCGTCAGCTCGT----GTCGTGAGA
PRS1          ATAGGGCTTCTCCTTC-GGGAGCAGAGTGACAGGTGGTGCATGGTTGTCGTCAGCTCGT----GTCGTGAGA
PRS2          ATAGGNCTTCTCCTTCNGGGAGCANNGTGACAGGTGGTGCATGGTTGTCGTCAGCTCGT----GTCGTGAGA
PVR1          ATGGATTGGTGCCTTC-GGGAACTCTGAGACAGGTGCTGCATGGCTGTCGTCAGCTCGT----GTTGTGAAA
PVR2          --------------CC-TGGAATTCTACC--CCCCTCTACAAGACTCTAGCTTGCCAGTTTCAAATGC-AGT
PVL           ATAGGGCTTCTCCTTC-GGGAGCANAGTGACAGGTGGTGCATGGTTGTCGTCAGCTCGT----GTCGTGAGA
PVS1          ATAGGGCTTCTCCTTC-GGGANCAGAGTGACAGGTGGTGCATGGTTGTCGTCAGCTCGT----GTNGTGNNN
PVS2          ATAGGGCTTCTCCTTC-GGGAGCAGAGTGACAGGTGGTGCATGGTTGTCGTCAGCTCGT----GTCGTGAGA
EC-MSR-B      ATGCTTTGGTGCCTTC-GGGAACTCTGAGACAGGTGCTGCATGGCTGTCGTCAGCTCGT----GTTGTGAAA
EC-MSR-L      ATGCTTTGGTGCCTTC-GGGAACTCTGANACAGGTGCTGCATGGCTGTCGTCAGCTCGT----GTTGTGAAA
EC-MSR-R      ATGAGNNNNTNCCTTC-GGGAACTGTGAGACAGGTGNTGCATGGCTGTNGTCAGCTCGT----GTTGTGAAA

cons                         *  ***                * ** *  * * *   **  **       *


PRR1          TGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTATCCTTTGTTGCCAGCGGTCCGGCCGGGAACTCANAGG
```

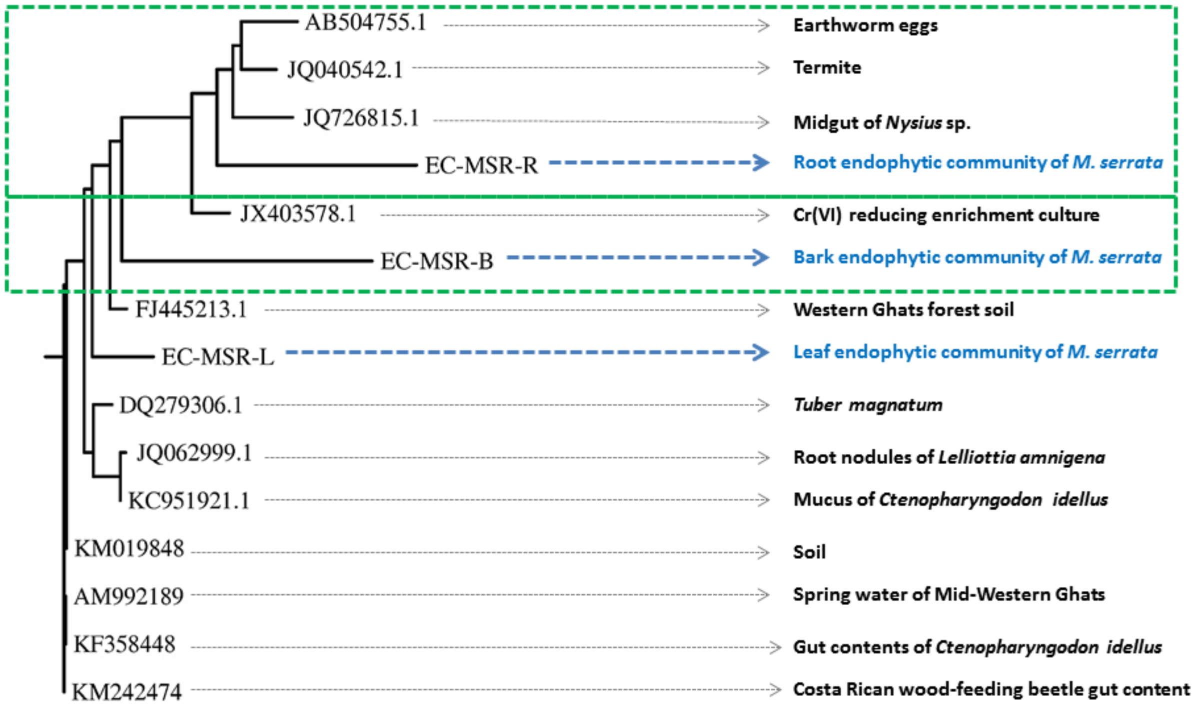


```
PRR2          TCCCAGGTTGAGCCCGGGGATT--------------------------------------------------
PRL           TGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGATCTTAGTTGCCATCA-TTAAGTTGGGCACTCTAAGG
PRS1          TGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGATCTTAGTTGCCATCA-TTAAGTTGGGCACTCTAAGG
PRS2          TGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGATCTTAGTTGCCATCA-TTAAGTTGGGCACTCTAAGG
PVR1          TGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTATCCTTTGTTGCCAGCGGTTCGGCCGGGAACTCAAAGG
PVR2          TCCCAGGTTGAGCCCGGGGATT-TCACA---TCTGACTTAACAAACCGCC---T--G-CGTGCGCT------
PVL           TGTTGGGTTAAGTCCCNNAACGAGCGCAACCCTTGATCTTAGTTNNCATCA-TTAAGTTGGGCACTCTAAGG
PVS1          TGTTGGGTTAAGTCCCGCAACGAGCGCAACCNTTGATCTTAGTTGCCATCA-TTAANTTGGNNNCTCTAAGG
PVS2          TGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGATCTTAGTTGCCATCA-TTAAGTTGGGCACTCTAAGG
EC-MSR-B      TGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTATCCTTTGTTGCCAGCG-GTCGGCCGGGAACTCAAAGG
EC-MSR-L      TGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTATCCTTTGTTGCCAGCGGTTCGGCCGGGAACTCAAAGG
EC-MSR-R      TGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTATCNTTTGTTGCCAGCGGTCCGGCCGGGAANTCAAAGG

cons          *    **** ** **    *


PRR1          AGACTGCCA--GTGATAAACTGGAGGAAG-GTGGGGATGACGTCAAGTCATCATGGCCCTTACGAGTAGGGC
PRR2          TCACATCCGACTTGACAGACCGCCTG--C-GTGCGCTTTACGCCCAGTAATT---CCGATTAACGCTTGCAC
PRL           TGACTGCCG--GTGACAAACCGGAGGAAG-GTGGGGATGACGTCAAATCATCATGCCCCTTATGACCTGGGC
PRS1          TGACTGCCG--GTGACAAACCGGAGGAAG-GTGGGGATGACGTCAAATCATCATGCCCCTTATGACCTGGGC
PRS2          TGACTGCCG--GTGACAAACCGGAGGAAG-GTGGGGATGACGTCAAATCATCATGCCCCTTATGACCTGGGC
PVR1          AGACTGCCA--GTGATAAACTGGAGGAAG-GTGGGGATGACGTCAAGTCATCATGGCCCTTACGAGTAGGGC
PVR2          TTACGCCCA--GTA---ATTCCGATTAAC-GCTTGCAC-CCTCCGTATTACCG--CGGCTGCTGGCACGGAN
PVL           TGACTGCCG--GTGACAAACCGGAGGAAG-GTGGGGATGACGTCAAATCATCATGCCCCTTATGACCTGGGC
PVS1          TGACTGCCG--GTGACAAACCGGAGGAAG-GTGGGGATGACGTCAAATCATCATGCCCCTTATGACCTGGGC
PVS2          TGACTGCCG--GTGACAAACCGGAGGAAG-GTGGGGATGACGTCAAATCATCATGCCCCTTATGACCTGGGC
EC-MSR-B      AGACTGCCA--GTGATAAACTGGAGGAAG-GTGGGGATGACGTCAAGTCATCATGGCCCTTACGAGTAGGGC
EC-MSR-L      AGACTGCCA--NNGATAAACTGGAAGAANNNTGGGGANNNNNTCAAGTCATCATGGCCCTTACGAGTAGGGC
EC-MSR-R      AGACTGCCA--GTGATAAACTGGAGGAAG-GTGGGGANGACGTCAAGTCATCATGNCCCTTACGNCCAGGGN

cons            **  **                          *        *   * *         *        *


PRR1          TACACACGTGCTACAATGGCGCATA-CAAAGAGAA--------------------------GCGAACTCGCG
PRR2          --CCTCCGTATTACCGCGGCTGCTGGCACGGAGTTAGCCGGTGCTTCTTCTGCGGGTAACGTCAA-TTGCTG
PRL           TACACACGTGCTACAATGGACGGTA-CAAAGAGCT--------------------------GCAAGACCGCG
PRS1          TACACACGTGCTACAATGGACGGTA-CAAAGAGCT--------------------------GCAAGACCGCG
PRS2          TACACACGTGCTACAATGGACGGTA-CAAAGAGCT--------------------------GCAAGACCGCG
PVR1          TACACACGTGCTACAATGGCGCATA-CAAAGAGAA--------------------------GCGACCTCGCG
PVR2          TTAGCCGGTGCTTCTTCTG-------CGAGTAACG--------------------------TCAA-TNNNNG
PVL           TACACACGTGCTACAATGGACGGTA-CAAAGAGCT--------------------------GCAAGACCGCG
PVS1          TACACACGTGCTACAATGGACGGTA-CAAAGAGCT--------------------------GCAAGACCGCG
PVS2          TACACACGTGCTACAATGGACGGTA-CAAAGAGCT--------------------------GCAAGACCGCG
EC-MSR-B      T-CNCACGTGNTNCAATGGCATATA-CAAAGAGAA--------------------------GCGACCTCGCG
EC-MSR-L      TACACACGTGCTACAATGGCATATA-CAAAGAGAA--------------------------GCGACCTCGCG
EC-MSR-R      TACACACGTGNTNCAATGGCATATN-CAAAGNGAA--------------------------GCGACCTCGNG

cons                 **  * *    *       *                                   * *      *


PRR1          AGAGCAAGCGGACCTCATAAAGNGCGTCGTAGTCCGGATTGGAGTCTGCAACTCGACTCCATGAAGTCGGAA
PRR2          CGGTT--ATTAACCAC--AACACCTTCCTCCCCGCTGAAAGTACTTTACAACCC------------------
PRL           AGGTGGAGCTAATCTCATAAAACCGTTCTCAGTTCGGATTGTAGGCTGCAACTCGCCTACATGAAGCTGGAA
PRS1          AGGTGGAGCTAATCTCATAAAACCGTTCTCAGTTCGGATTGTAGGCTGCAACTCGCCTACATGAAGCTGGAA
PRS2          AGGTGGAGCTAATCTCATAAAACCGTTCTCAGTTCGGATTGTAGGCTGCAACTCGCCTACATGAAGCTGGAA
PVR1          AGAGCAAGCGGACCTCATAAAGNNNGTCGTAGTCCGGATTGGAGTCTGCAACTCGACTCCATGAAGTCGGAA
PVR2          AGGTT--ATTAACCTT--NNNNCCTTCCTCCTCGCTGAAAGTACTTTACAACCC------------------
PVL           AGGTGGAGCTAATCTCATAAAACCGTTCTCAGTTCGGATTGTAGGCTGCAACTCGCCTACATGAAGCTGGAA
PVS1          AGGTGGAGCTAATCTCATAAAACCGTTCTCAGTTCGGATTGTAGGCTGCAACTCGCCTACATGAAGCTGGAA
PVS2          AGGTGGAGCTAATCTCATAAAACCGTTCTCAGTTCGGATTGTAGGCTGCAACTCGCCTACATGAAGCTGGAA
EC-MSR-B      AGAGCAAGCGGNCNTCATAAAGTATGTNGTAGTCCGGATTGGAGTCTGCAACTCGACTCCATGAAGTCGGAA
EC-MSR-L      AGAGCAAGCGGACCTCATAAAGTATGTCGTAGTCCGGATTGGAGTCTGCAACTCGACTCCATGAAGTCGGAA
EC-MSR-R      AGAGCAAGCGGACCTCATAAAGTATGTNGTAGTCCGGATTGGAGTCTGCAACTNGACTCCATGAAGTNGGAA

cons           *                                * **  * *   * ****


PRR1          TCGCTAGTAATCGTAGATCAGAATGCTACGGTGAATACGTTCCCGGGCCTTG----TACACACCGCCCGTCA
PRR2          -------------------------------NNAA----------GGCCTTCTTCATACACGCGGCA--TGG
PRL           TCGCTAGTAATCGCGGATCAGCATGCCGCGGTGAATACGTTCCCGGGCCTTG----TACACACCGCCCGTCA
PRS1          TCGCTAGTAATCGCGGATCAGCATGCCGCGGTGAATACGTTCCCGGGCCTTG----TACACACCGCCCGTCA
PRS2          TCGCTAGTAATCGCGGATCAGCATGCCGCGGTGAATACGTTCCCGGGCCTTG----TACACACCGCCCGTCA
PVR1          TCGCTAGTAATCGTAGATCAGAATGCTACGGTGAATACGTTCCCGGGCCTTG----TACACACCGCCCGTCA
PVR2          -------------------------------NNAA----------GGCCTTCTTCATANNNCCGGCA--TGG
PVL           TCGCTAGTAATCGCGGATCAGCATGCCGCGGTGAATACGTTCCCGGGCCTTG----TACACACCGCCCGTCA
PVS1          TCGCTAGTAATCGCGGATCAGCATGCCGCGGTGAATACGTTCCCGGGCCTTG----TACACACCGCCCGTCA
PVS2          TCGCTAGTAATCGCGGATCAGCATGCCGCGGTGAATACGTTCCCGGGCCTTG----TACACACCGCCCGTCA
```



```
EC-MSR-B    TCG-TAGTAATCGT-GATCAGAATGCTACGGTGAATACGTTCCCGGNCC-----------------------
EC-MSR-L    TCGCTAGTAATCGTAGATCAGAATGCTACGGTGAATACGTTCCCGGGCCTTG----TACACACCGCCCGTCA
EC-MSR-R    TCGNTAGTAATCGTGGATCAGAATGCCACGGTGAATACGTTCCCGGNCCTTG----TACACACCGCCCGTCA

cons                                         **          * **


PRR1        CACCATGGGAGTGGG-NNCAAAAGAAGTAGGTAGC-TTTAACCT-CG--GGAGGGCGCTAC--------CAC
PRR2        CTGCATCAGGCTT-G---CGNCC--ATTNTGCAAT-A-TTCCCCACT--GCTGCCTN-----------CCGT
PRL         CACCACGAGAGTTTGTAACACCCGAAGTCGGTGGG-G-TAACCT-TTTTGGAGCCAGCCG---CCTAAG-NN
PRS1        CACCACGAGAGTTTG-AACACCCGAAGTCGGTGGGGG-TAACC---------------------------NN
PRS2        CACCACGAGAGTTTGTAACACCCGAAGTCGGTGGG-G-TAACCT-TTA-TGA-----------------GCC
PVR1        CACCATGGGAGTGGG-TGCAAAAGAAGTAGGTAG--C-TAACCT-CN--GGAGGGCGCTA----------CC
PVR2        CTGCATCAGGCTT-G---NNCCC--ATTNNGCAAT-A-TTCCCCACT--GCTGCCTCCCG---TANGAATCT
PVL         CACCACGAGAGTTNG-AACACCCGAAGTCGGTGGG-G-TAACCT-TTATGGAGCCAGCCG---CCTAAGGGG
PVS1        CACCACGAGAGTTTG-AACACCCGAAGTCGGTGGG-G-TAACCT-TT--TGAGCCAGCCG----------CT
PVS2        CACCACGAGAGTTTG-AACACCCGAAGTCGGTGGG-G-TAACCT-TT--TGAGCCAGCCG----------CT
EC-MSR-B    -----------------------------------------------------------------------T
EC-MSR-L    CACCATGGGAGTGGGTTGCAAAAGAAGTAGGTAGC-T-TAACCT-CG--GGAGGGCGCTACCACTTT-GNNC
EC-MSR-R    C---------------------------------------------------------------------NC

cons


PRR1        TT
PRR2        AG
PRL         AC
PRS1        TT
PRS2        AG
PVR1        AC
PVR2        GG
PVL         AC
PVS1        AA
PVS2        AA
EC-MSR-B    TG
EC-MSR-L    AG
EC-MSR-R    CA

cons
```

Fig. S4. A phenogram-like rooted phylogenetic tree representing the correlation of our isolated endophytic bacterial communities of *M. serrata* with most closely related bacteria based on their 16S rRNA sequences. EC-MSR-R, EC-MSR-B, EC-MSR-L represent the endophytic bacterial community codes of *Maytenus* plants. Others represent the accession numbers of closely related bacterial species. The arrows provide information about the isolation source or natural host of the respective bacteria.



III. SUPPLEMENTARY TABLES

Table S1. List of primers used in this study

Primer Pair	**Sequence**
P1 (forward)	5’-AAAGGATCCTCATATGCTCGACGCGATCGTGAT-3’
P1 (reverse)	5’-AAAGAATTCACCGGCCGACCGTGCCCG-3’
P2 (forward)	5’-TTCCCSCGSTACCASATCGGSGAG-3’
P2 (reverse)	5’- GSGGGATSWMCCAGWACCASCC-3’
P3 (forward)	5’-AGAGGATCCTTCGAGCRSGAGTTCGC-3’
P3 (reverse)	5’-GCAGGATCCGGAMCATSGCCATGTAG-3’
27F (16S rRNA forward)	5’- AGAGTTTGATCMTGGCTCAG -3’
1492R (16S rRNA reverse)	5’- TACGGYTACCTTGTTACGACTT -3’

Table S2. Overview of the AHBA synthase and halogenase genes found in endophytic bacterial communities and plant tissues of *Maytenus serrata* plants

Name of plant tissue or endophytic bacterial community	**AHBA synthase genes (bp)**	**Halogenase genes (bp)**	**Most closely related gene (EMBL accession number)**	**Most closely related translated protein (UniProt identifier)**	**EMBL- Bank accession numbers of submitted sequences (present study)**
Maytenus serrata plant root tissue (P-MSR-R)	765	N.A.	LN681221.1	F4NAR6	LN874065
Maytenus serrata plant leaf tissue (P-MSR-L)	770	N.A.	AAC13997.1	Q44131	LN874066
Maytenus serrata plant bark tissue (P-MSR-B)	731	N.A.	LN831790.1	F4NAR6	LN874067
Maytenus serrata root endophytic community (EC-MSR-R)	N.A.	N.A.	N.A.	N.A.	N.A.
Maytenus serrata leaf endophytic community (EC-MSR-L)	N.A.	N.A.	N.A.	N.A.	N.A.
Maytenus serrata bark endophytic community (EC-MSR-B)	N.A.	486	CP001630.1	C6W874	LN874068

Table S3. Presence of detectable major precursors of maytansine in various tissues of the Cameroonian *M. serrata* plant (see also Fig. S2)

Compound (see also Fig. S2)	**Bark**	**Leaf**	**Root**
1	✓	✗	✗
2	✗	✗	✗
3	✓	✗	✗
4	✓	✓	✓
5	✓	✗	✓
6	✓	✗	✓
Maytansine	✓	✓	✓

References

1. S. Kusari, M. Lamshöft, P. Kusari, S. Gottfried, S. Zühlke, K. Louven, *et al.*, *J. Nat. Prod.*, 2014, **77**, 2577.
2. Y. Wu, Q. Kang, Y. Shen, W. Su, L. Bai, *Mol. Biosyst.*, 2011, **7**, 2459.
3. X-G. Li, X-M. Tang, J. Xiao, G-H. Ma, L. Xu, S-J. Xie, *et al.*, *Mar. Drugs.*, 2013, **11**, 3875.