Microbial adaptation to venom is common in snakes and spiders.

E. Esmaeilishirazifard[1,2†], L. Usher[1,2†], C. Trim[3], H. Denise[4], V. Sangal[5], G.H. Tyson[6], A. Barlow[7], K.F. Redway[1], J.D. Taylor[1,2,8], M. Kremyda-Vlachou[1], T. D. Loftus[9], M.M.G. Lock[9], K. Wright[1], A. Dalby[1], L.A.S. Snyder[10], W. Wuster[11], S. Trim[9], and S.A. Moschos[1,2,5.*]

SUPPLEMENTARY FIGURES

Affiliations:

[1]Department of Biomedical Sciences, Faculty of Science and Technology, University of Westminster, London, U.K., E.U.

[2]Westminster Genomic Services, Faculty of Science and Technology, University of Westminster, London, U.K., E.U.

[3]School of Human and Life Sciences, Faculty of Social and Applied Sciences, Canterbury Christ Church University, Canterbury, Kent, U.K.

[4]EMBL-EBI European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK.

[5]Department of Applied Sciences, Faculty of Health and Life Sciences, Northumbria University, Newcastle, Tyne and Wear, UK.

[6]Food and Drug Administration, Center for Veterinary Medicine, Office of Research. Laurel, MD, USA.

[7]Institute for Biochemistry and Biology, University of Potsdam, Potsdam, Germany.

[8]School of Environment and Life Sciences, University of Salford, Salford, Greater Manchester, U.K.

[9]Venomtech Limited, Discovery Park, Sandwich, Kent, UK.

[10]School of Life Sciences, Pharmacy, and Chemistry, Kingston University, Kingston Upon Thames, London, UK.

[11]School of Biological Sciences, College of Natural Sciences, Bangor University, Bangor, Wales, U.K.

*Correspondence to: sterghios.moschos@northumbria.ac.uk.

†These authors contributed equally to this manuscript.

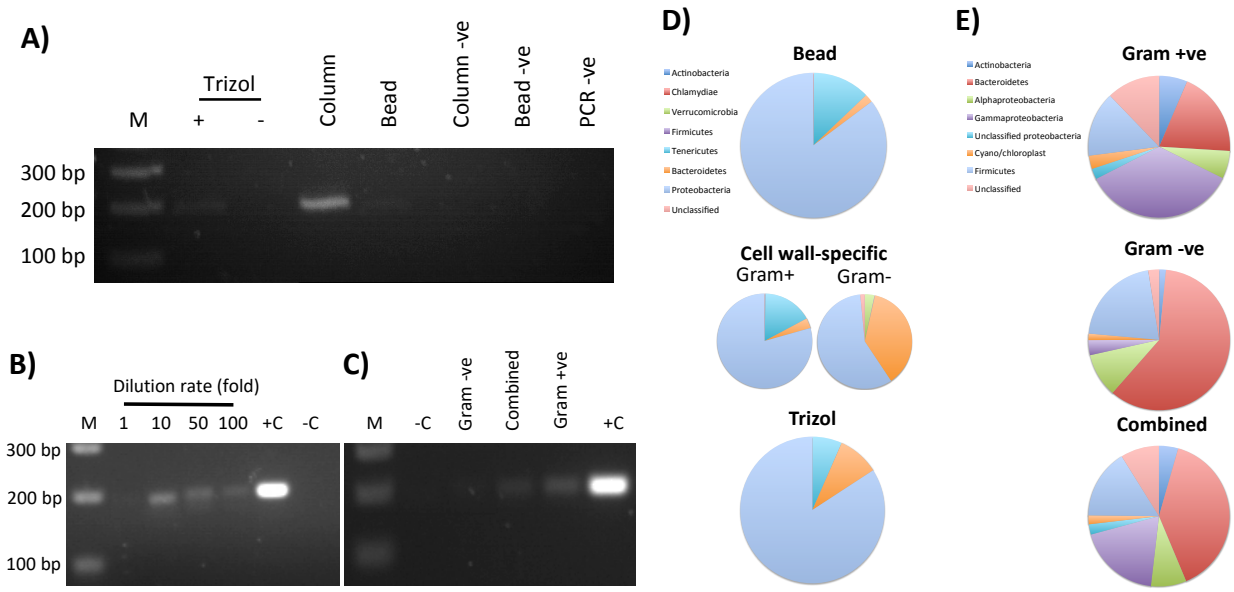

Figure S1: 16S Ribosomal RNA gene PCR output and phylogenetic differences on account of venom collection and extraction methods. The choice of extraction method (phenol-chloroform-based: Trizol; column based; magnetic bead based) impacts significantly on the recovery and amplification of bacterial DNA from lyophilized *B. atrox* venom (A). This bacterial DNA is not an artefact of lyophilisation process contamination as detection is maintained in aseptically collected, flash-frozen *Bitis arietans* venom, nor is it an artefact of diluent contamination by 18 MΩ water confirmed 16S free by PCR; however, >10x dilution of venom is necessary for PCR to progress (B). The yield of bacterial DNA is a function of upstream cell lysis methods selectivity for Gram +ve or Gram –ve cell walls (C). The cell lysis and extraction methodology also directly impact upon microbial diversity profiles as determined by 16S rRNA phylogenetics for either lyophilised (D) or aseptically collected, flash-frozen venoms (E), , with combined use of cell wall-specific extraction methods yielding more balanced profiles.. +C: positive control; -C: negative control; -ve: method specific negative controls.

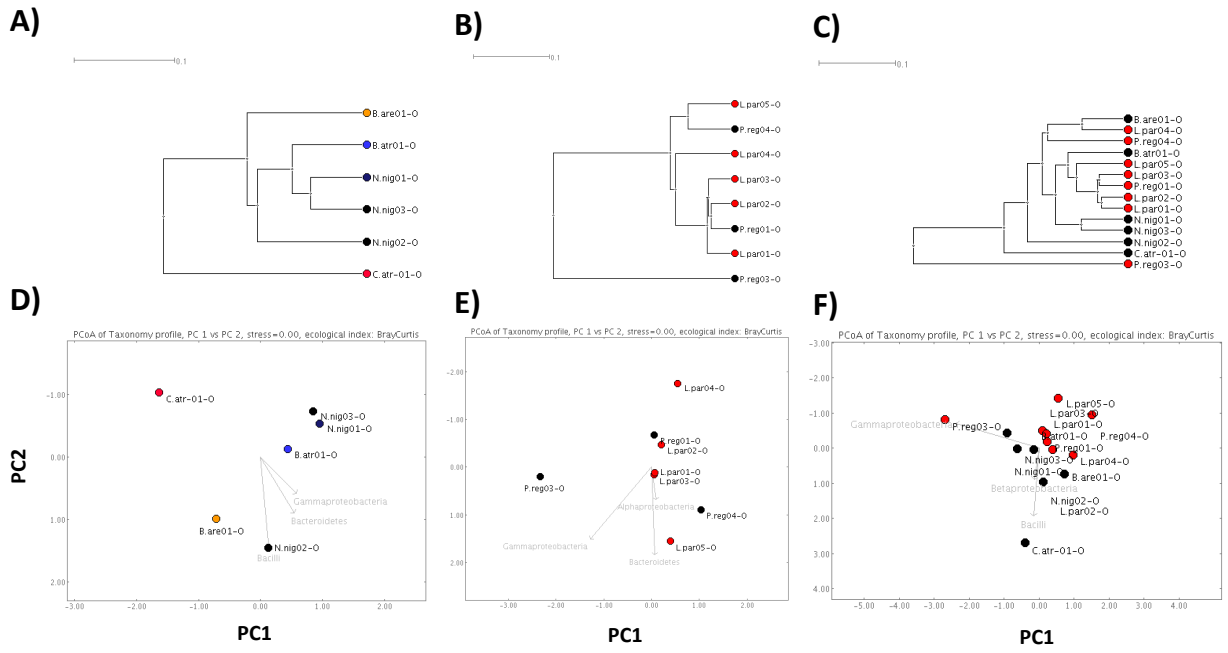

Figure S2: Comparison of the oral microbiomes of snakes and spiders suggests their oral microbiota is not host-species specific. UPGMA tree (A - C) and PCoA (D – F) analysis of the oral microbiome diversity of snakes (A, D; individuals identified by independently coloured dots), spiders (B, E; *L. parahybana*: red dots, *P. regalis*: black dots), or vertebrate vs invertebrate animals (C, F; black vs red dots), as determined by 16S rRNA phylogenetic analysis at class level indicate no host species-specific relationships. Dots represent single captivity individuals, labelled with short species name, enumerated for individual number and identified for the oral/fang (O) nature of the sample.

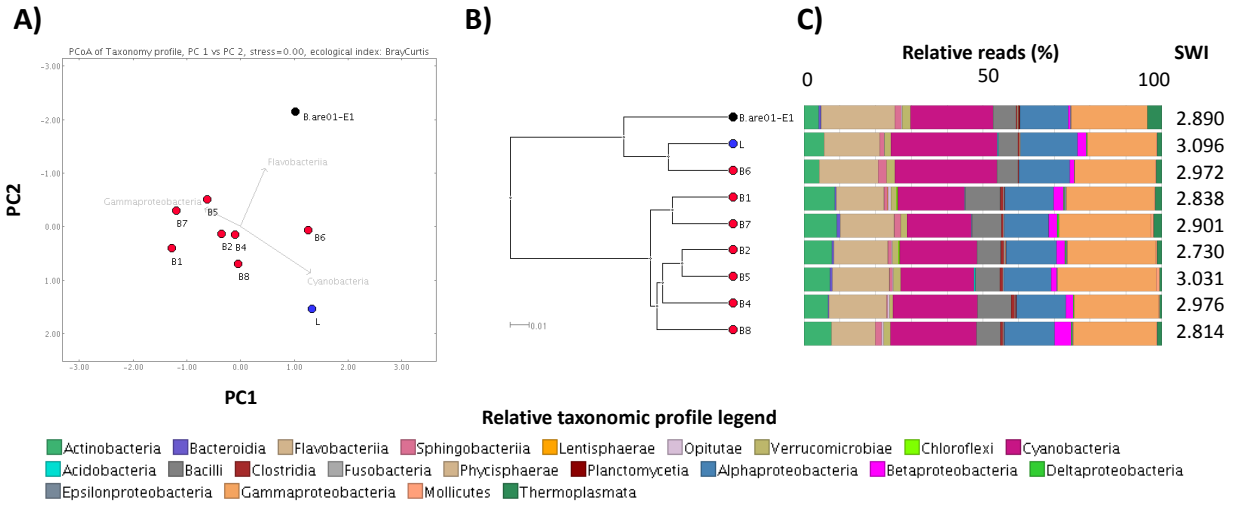

Figure S3: The origin of a *B. arietans* snake does not appear to influence the microbiome profile in the venom of each animal. *B. arietans* venom microbiome profiles do not present substantial differences on account of host geographical origin as determined by A) PCoA, B) UPGMA tree and C) class-level taxonomic profiling following 16S rRNA phylogenetic analysis. Dots in (A) and (B) represent individual animal data, are coloured and labelled by animal origin and number (red B1-8: wild; blue L: lyophilised captivity; black B. are01-E1: flash-frozen captivity). Relative taxonomic diversity profiles in (C) are aligned to the UPGMA tree sample labels, with the Shannon-Weiner Index (SWI) of each sample indicated.

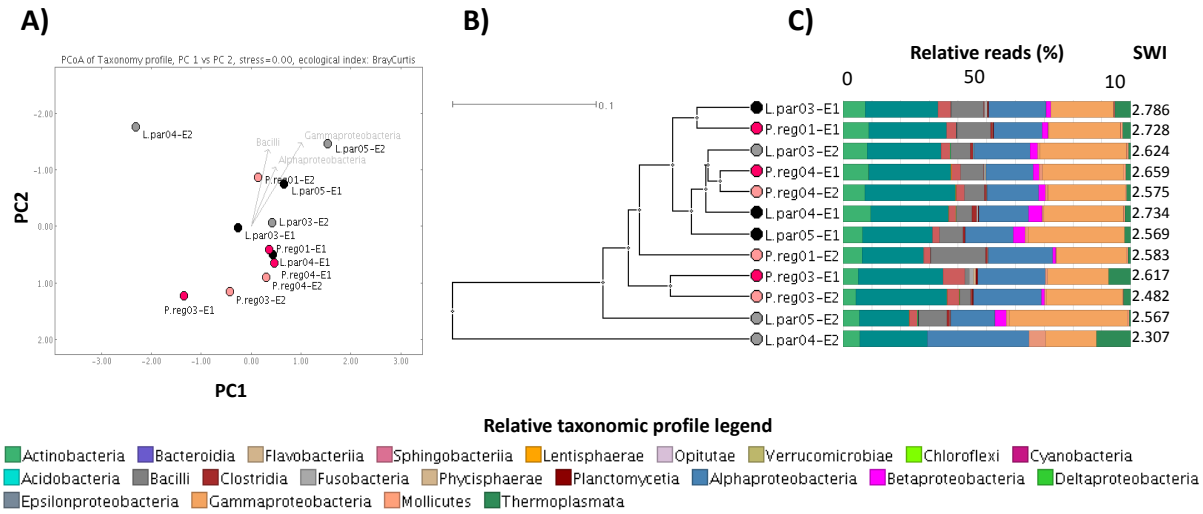

Figure S4: Spider venom microbiome profiles suggest closer relationships between consecutive envenomation samples within *P. regalis* individuals. Spider venom microbiomes were compared by A) PCoA, B) UPGMA tree and C) class-level taxonomic profiling following 16S rRNA phylogenetic analysis. Dots in (A) and (B) represent individual animal data, are colored/labelled by species and envenomation number (black and grey: *L. parahybana* envenomation 1 (E1) and 2 (E2) respectively; red and pink: *P. regalis* E1 and E2 respectively). Relative taxonomic diversity profiles in (C) are aligned to the UPGMA tree sample labels, with the Shannon-Weiner Index (SWI) of each sample indicated.

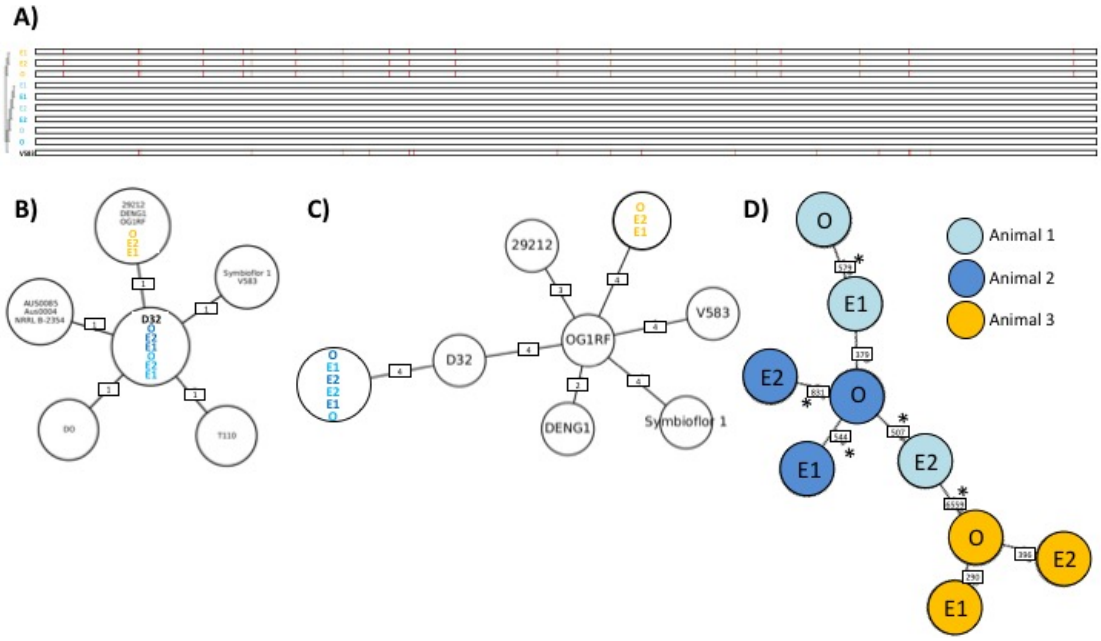

Fig. S5: MSA, MST and cgMLST define two novel *E. faecalis* sequence types isolated across *N. nigricollis* venom and oral cavities. Blinded MSA (A) of the *KatA* gene sequence across the nine *E. faecalis isolates* obtained from *N. nigricollis* oral swabs (O) and two consecutive envenomation samples (E1 and E2) from three independent animals (light blue (N.nig01), dark blue (N.nig02) and orange (N.nig03)) defines two alleles distinct to the V583 reference sequence (bottom lane). Base conservation is defined by similarity to the animal 1 and 2 *KatA* sequence using BoxShade v.3.3.1 on mobyle.pasteur.fr. Each pixel column represents a different nucleotide with orange and red columns indicating increasingly different nucleotides. Blinded MST analysis of these nine isolates against B) *E. faecalis* and *E. faecium* reference genomes (distance calculations based on *E. faecium* MLST), C) *E. faecalis* reference genomes with partial incidence locus data removed, and D) a custom cgMLST schema derived from *E. faecalis* OG1RF, D32 and DENG1 including loci with partial data between all study isolates (8101 targets). Allelic differences in excess of 5% of the cgMLST schema are highlighted by '*'.

Reference genomes: *E. faecalis*: V583, OG1RF, D32, DENG1, 29212, Symbioflor 1; *E. faecium*: T110; AUS0085, Aus0004, NRRL B-2354, DO.

A) >*pstS_E.faecalis_N.nig03_venom*

GTGACCATTGGAAATTCCGATGTGTTTGCCGAAGAGAAAGATGGCGTGGATGCCTCTAAACTAGTTGATCATCGGGTGGCCGTGGTTGGTATGGGACCAGTGG
TCAACAAAGAAGTCGGCGTGAAGAATTTAACAAAGCAACAATTGATTGATGTCTTTACTGGCAAAGTCAAAAACTGGAAAGAAGTGGGCGGCAAAGATCAAGA
AATCGTCGTAATAAACCGCGCAAACGGAAGTGGCACCCGAGCAACATTTGAAAAATGGGGCTTAGATGGAGCTAAACCAGTTCAATCACAAGAACAAGATTCT
TCGGGAACAGTTCGTAAAATTGTTGAACAAACACCAGGAGCAATCAGCTATTTAGCTTTTTCTTATATGGATGATTCCACCGTTGCTTTAAGCATTGATGGTG
TTGAACCAAAAGAAGAACACGTGAAAGACAATTCATGGAAAATCTGGTCTTATGAACATATGTATACAAAAGGGGAGCCTAATAAAGAAGTAAAAGCCTTCTT
AGACTATATGGTCACTGATGATGTTCAAAAAACAATTGTCAAAGACTTAGGTTATTTAGCCATCACAG

B) >*yqiL_E.faecalis_N.nig03_venom*

ATGTTACAACACAACTTTTAAAAAGACATTCCACTATTTCTGAAGAAATTGATCAAGTAATCTTTGGAAATGTTTTACAAGCTGGAAATGGGCAAAATCCCGC
ACGACAAATAGCAATAAACAGCGGTTTGTCTCATGAAATTCCCGCCATGACGGTTAATGAGGTCTGCGGATCAGGCATGAAGGCCGTTATTTTGGCGAAACAA
TTGATTCAATTAGGAGAAGCGGAAGTTTTAATTGCTGGTGGGATTGAGAATATGTCCCAAGCACCTAAATTACAACGATTTAATTACGAAACAGAAAGCTACG
ATGCGCCTTTTTCTAGTATGATGTATGATGGGTTAACGGATGCCTTTAGTGGTCAGGCAATGGGCTTAACTGCTGAAAATGTGGCCGAAAAGTATCATGTAAC
TAGAGAAGAGCAAGATCAATTTTC

C)

```
yqiL_E.faecalis_N.nig02_E1      AAAGCTACGATGCGCCTT-TTTCTAGTATGATGTATGATGGGTTAACGGATGCCTTTAGT
yqiL_E.faecalis_allele-8        AAAGCTACGATGCGCCTT-TTTCTAGTATGATGTATGATGGGTTAACGGATGCCTTTAGT
yqiL_E.faecalis_N.nig02_E2      AAAGCTACGATGCGCCTTTTTTCTAGTATGATGTATGATGGGTTAACGGATGCCTTTAGT
                                ****************** ****************************************
yqiL_E.faecalis_N.nig02_E1      GGTCAGGCAATG-GCTTAACTGCTGAAAATGTGGCCGAAAAGTATCATGTAACTAGAGAA
yqiL_E.faecalis_allele-8        GGTCAGGCAATGGGCTTAACTGCTGAAAATGTGGCCGAAAAGTATCATGTAACTAGAGAA
yqiL_E.faecalis_N.nig02_E2      GGTCAGGCAATGGGCTTAACTGCTGAAAATGTGGCCGAAAAGTATCATGTAACTAGAGAA
                                ************ **********************************************
```

Figure S6: Novel *pstS* and *yqiL* allele sequences obtained from *N. nigricollis* venom-derived *E. faecalis*. The sequences of the novel *pstS* (A) and *yqiL* (B) alleles found in a novel *E. faecalis* sequence type obtained from *N. nigricollis* venom (animal 3). Clustal omega alignments of the *yqiL* sequences (C) from *E. faecalis* isolates derived from animal 2 venom against *E. faecalis* *yqiL* allele 8 found in the orally-derived isolate. The alignment is focused to positions 301-319 of the 436 nt allele and single base pair indels are highlighted in red.

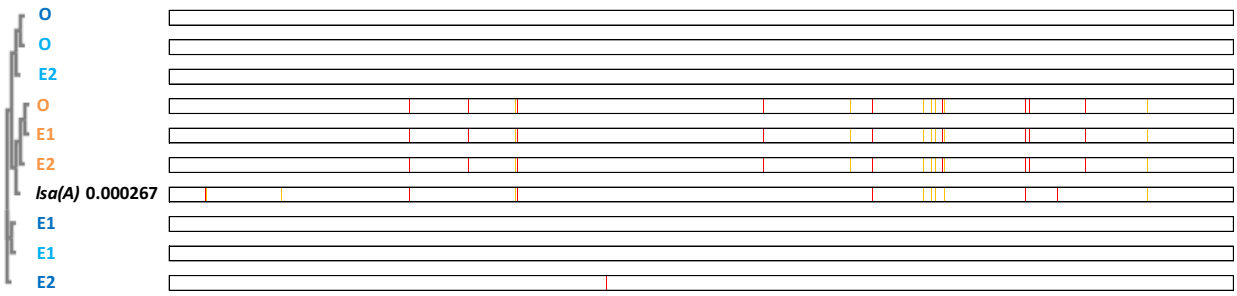

Figure S7: Multiple sequence alignment of the *lsa(A)* gene reinforces the clustering of the *N. nigricollis*-derived *E. faecalis* isolates. Blinded MSA of the *lsa(A)* antibiotic resistance gene sequence across the nine *E. faecalis isolates* obtained from *N. nigricollis* oral swabs (O) and two consecutive envenomation samples (E1 and E2) from three independent animals (light blue (N.nig01), dark blue (N.nig02) and orange (N.nig03)) defines two alleles distinct to the TX0263 reference strain gene sequence (accession no. AY737526.1). Base conservation is defined by similarity to the animal 1 and 2 *lsa(A)* sequence using BoxShade v.3.3.1 on mobyle.pasteur.fr. Each pixel column represents a different nucleotide with orange and red columns indicating increasingly different nucleotides.