environmental microbiology

Environmental Microbiology (2016) **18**(5), 1364–1378 doi:10.1111/1462-2920.12939

Species-specific diversity of novel bacterial lineages and differential abundance of predicted pathways for toxic compound degradation in scorpion gut microbiota

Luis M. Bolaños,[1]* Mónica Rosenblueth,[1] Santiago Castillo-Ramírez,[2] Gilles Figuier-Huttin[1] and Esperanza Martínez-Romero[1]

[1]*Programa de Ecología Genómica y*

[2]*Programa de Genómica Evolutiva, Centro de Ciencias Génomicas, Universidad Nacional Autónoma de México, Cuernavaca, Morelos 62210, México.*

Summary

Scorpions are considered 'living fossils' that have conserved ancestral anatomical features and have adapted to numerous habitats. However, their gut microbiota diversity has not been studied. Here, we characterized the gut microbiota of two scorpion species, *Vaejovis smithi* and *Centruroides limpidus*. Our results indicate that scorpion gut microbiota is species-specific and that food deprivation reduces bacterial diversity. 16S rRNA gene phylogenetic analysis revealed novel bacterial lineages showing a low level of sequence identity to any known bacteria. Furthermore, these novel bacterial lineages were each restricted to a different scorpion species. Additionally, our results of the predicted metagenomic profiles revealed a core set of pathways that were highly abundant in both species, and mostly related to amino acid, carbohydrate, vitamin and cofactor metabolism. Notably, the food-deprived *V. smithi* shotgun metagenome matched almost completely the metabolic features of the prediction. Finally, comparisons among predicted metagenomic profiles showed that toxic compound degradation pathways were more abundant in recently captured *C. limpidus* scorpions. This study gives a first insight into the scorpion gut microbiota and provides a reference for future studies on the gut microbiota from other arachnid species.

Received 26 February, 2015; accepted 31 May, 2015. *For correspondence. E-mail lbolanos@lcg.unam.mx; Tel. +527 773 175 581; Fax +527 773 175 581.

Introduction

Scorpions are among the oldest animals on Earth (Dunlop, 2010) and are considered to be living fossils. To date, over 2000 species of scorpions have been described, and these colonize a worldwide range of habitats (Prendini, 2011; Rein, 2012). Normally, scorpions feed on insects, small vertebrates or other scorpions, and are able to tolerate different stresses, such as food (Polis, 1990) and water deprivation (Hadley, 1970), and extreme temperatures (Hadley, 1974). Scorpions are efficient predators and have the ability to consume a large amount of food in one meal (Polis, 1990). In nature, two important aspects of scorpions feeding behaviour are the variability in prey types and the inconstancy of the time intervals between each feeding period.

In Mexico, there are more than 200 described species (González-Santillán, 2001; Francke, 2014). *Vaejovis smithi* (Pocock, 1902) and *Centruroides limpidus* (Karsch, 1879) scorpions are found in the state of Morelos, located in South-Central Mexico, and have different geographical distributions, with *V. smithi* occurring in more temperate and higher areas. As far as their venoms are concerned, that of *C. limpidus* is highly toxic with an LD_{50} of 1.72 mg kg^{-1} in mice (Padilla *et al.*, 2003), whereas *V. smithi* belongs to a non-neurotoxic genus (Rowe *et al.*, 2011).

Symbiotic bacteria are commonly found in arthropods and affect reproduction (Werren *et al.*, 2008; Sharon *et al.*, 2010), behaviour (Ridley *et al.*, 2012), longevity and immunity of their hosts (Guo *et al.*, 2014). Therefore, symbionts are recognized as key elements driving the evolution and adaptation of arthropods (Klepzig *et al.*, 2009). Clearly, arthropod gut microbiota are involved in many important and diverse physiological processes, such as host peritrophic membrane integrity maintenance (Narasimhan *et al.*, 2014), polysaccharide degradation (Brune and Ohkuma, 2011; Lee *et al.*, 2014), nitrogen fixation and recycling (Ohkuma, 2008; Morales-Jiménez *et al.*, 2013), toxin degradation (Kikuchi *et al.*, 2012), and protection against pathogens (Gonzalez-Ceron *et al.*, 2003; Dillon *et al.*, 2005; Cirimotich *et al.*, 2011). Remarkably, gut bacterial community composition in arthropods is

Table 1. Statistical analysis of 16S rRNA gene clone libraries.

	Vaejovis smithi			*Centruroides limpidus*		
	Recently captured	Laboratory-fed	Food-deprived	Recently captured	Laboratory-fed	Food-deprived
No. of individual libraries	7	6	5	6	5	5
Accession numbers	KM978218-KM978243	KM978244-KM978265	KM978266-KM978280	KM978281-KM978305	KM978306-KM978324	KM978325-KM978331
Statistics						
Mean of clones per library	47.1 ± 25.2	38.5 ± 7.7	39.4 ± 15.4	38.8 ± 17.7	35.4 ± 9.86	39.4 ± 6.7
No. of unique representative sequences	26	22	15	25	19	7
No. of OTUs[a]	18	15	11	20	16	4
Diversity index and measure						
Chao1	53	37	29	37.5	35.5	7
Faith's PD	1.40158	1.13899	0.89077	1.23759	1.34444	0.40619

a. OTUs were defined at 97% sequence identity.

influenced by the host diet (Huang *et al*., 2013), developmental stage (Vasanthakumar *et al*., 2008; Wong *et al*., 2011), geographical location (Corby-Harris *et al*., 2007; Adams *et al*., 2010) and species (Schloss *et al*., 2006; Tang *et al*., 2012; Wong *et al*., 2013).

To assess the gut microbiota of scorpions, we used a culture-independent approach to analyse the taxonomic diversity and metabolic capabilities from recently captured *V. smithi* and *C. limpidus* specimens. Furthermore, we compared these results to food-deprived and laboratory-fed scorpions to better understand the dynamics of the scorpion's gut microbiota and to analyse the effects of starvation and a homogeneous type of diet on it.

Results

Gut microbiota are species-specific and food deprivation reduces bacterial diversity

To study the gut bacterial diversity from the scorpion specimens, we created 16S rRNA libraries. First, a total of 34 libraries were constructed with 16S rRNA polymerase chain reaction (PCR) products from individual guts of both scorpion species under three different treatments: recently captured, laboratory-fed and food-deprived. Out of a total of 1365 clones categorized by amplified ribosomal DNA restriction analysis (ARDRA), 180 almost full-length 16S rRNA genes (~ 1400 bp.) were sequenced. We removed 16 chimeric, 10 low-quality sequences and identical sequences from each library. Identical sequences were considered for the clone count in further analysis. The total set of 16S rRNA was composed of 114 sequences, and the source and characteristics of these sequences are shown in Table 1. Using this dataset, we defined 55 operational taxonomic units (OTUs) at a 0.03 distance. We used the Ribosomal Database Project to classify these OTUs and found that 23.63%, 16.36% and 18.18% of the total OTUs corresponded to *Alpha*-, *Beta*- and *Gammaproteobacteria*, respectively. Bacilli and Clostridia (Firmicutes phylum) represented 12.72% and 7.27%, respectively, followed by Actinobacteria (12.72%), Mollicutes (3.63%) and Spirochaetes (1.81%). Two OTUs could not be classified: OTU4 with 90% identity to *Spiroplasma veloscicrescens* and OTU1 with 79% identity to *Spiroplasma lampyridicola*; we designated the latter as Scorpion Group 1 (SG1).

Recently captured specimens from *V. smithi* and *C. limpidus* had the most diverse microbiota with a total of 18 and 20 OTUs, respectively (Table 1). Pairwise comparisons between the Chao1 diversity indexes showed that microbiota from *V. smithi* laboratory-fed and food-deprived scorpions were significantly less diverse than that from the recently captured group (chi-square test, $P < 0.05$). In *C. limpidus*, bacteria from food-deprived scorpions were significantly less diverse than bacteria from laboratory-fed and the recently captured scorpions (chi-square test, $P < 0.05$), while the comparison of bacteria from laboratory-fed and recently captured scorpions was not statistically significant (chi-square test, $P = 0.842$). To further understand the bacterial diversity, we calculated the Faith's phylogenetic diversity (i.e. the sum of all the branches on a phylogenetic tree of a group of organisms) for the six groups (Table 1). Food-deprived scorpions of both species had lower measures relative to the other groups, with food-deprived *C. limpidus* being the lowest of the six groups (0.4062). The remaining four groups (laboratory-fed and recently captured of both species) had relatively similar measures with no statistical difference between them (chi-square test, *P*-value = 0.9985). Rarefaction curves showed great variance in the number of OTUs present among treatments independently of how deeply the library was sampled. However, in food-deprived individuals, the maximum numbers of phylotypes were four and two in *V. smithi* and *C. limpidus*, respectively (Fig. S1). Taken together, these results imply that, for both species, the recently captured

and laboratory-fed specimens had a more diverse microbiota. A possible explanation is that transient gut bacteria might have been eliminated under food deprivation.

We analysed the number of OTUs that are shared between the two species, regardless of the treatment. Of the 55 total OTUs, only eight were shared between both scorpion species. Six and seven OTUs were found in more than one treatment in *C. limpidus* and *V. smithi*, respectively. Evidently, this indicates that there is a great bacterial variation among species and among treatments. However, SG1 (related to the phylum Tenericutes) was found in 94.4% of the *V. smithi* scorpions regardless of the treatment. Clearly, the presence of this sequence is not affected by the treatment, and its absence in *C. limpidus* could suggest a strong association with *V. smithi*. Two other widely distributed OTUs were also related to the Mollicutes class. OTU2, here referred to as *Mycoplasma*-like from *V. smithi* (MVs), was present in 14% of the recently captured individuals, 16% of laboratory-fed specimens and 80% of food-deprived *V. smithi* scorpions. OTU3, here referred to as *Mycoplasma*-like from *C. limpidus* (MCl), was present in 50% of the recently captured specimens, 20% of laboratory-fed individuals, 80% of food-deprived *C. limpidus* scorpions and in one specimen of *V. smithi* that was recently captured. The MV and MCl sequences have an identity of 93.5% between them. The identity to the closest match in the National Center for Biotechnology Information non-redundant database was 89% and 88% to *Mycoplasma hyorhinis* GDL-1 for MVs and MCl, respectively. Seemingly, each species of scorpion has its own *Mycoplasma*-like lineage that persists across the treatments and even when scorpions are food-deprived (Fig. 1). Other frequently observed OTUs, present in more than one treatment group, are those related to *Streptococcus*, *Methylobacterium*, *Agrobacterium*, *Stenotrophomonas*, *Pseudomonas* and *Ochrobactrum*. Notably, all of them have been reported to colonize the intestinal tracts of many arthropods (Vasanthakumar *et al*., 2008; Geib *et al*., 2009; Shelomi *et al*., 2013).

We evaluated which of the two factors (scorpion species or diet treatment) has the major influence on the bacterial composition of the gut microbiota. Hierarchical clustering of the OTUs revealed that scorpion species is the main factor that divides the groups, followed by diet treatment (Fig. 1). Weighted and unweighted principal coordinate analysis (PCoA), based on UniFrac distances also show that samples are clustered preferentially by scorpion species (Fig. 2A). Principal coordinate analysis in which samples were tagged by treatment did not create a distinctive clustering pattern of recently captured, laboratory-fed or food-deprived scorpions (Fig. 2B). These results indicate that scorpion species explains most of the variation, and this is more influential than diet treatment in determining gut bacterial composition, although food deprivation reduces bacterial diversity.

Phylogenetic analysis reveals novel bacterial lineages

To gain further information on SG1 and the other previously unreported OTUs, we used seqmatch from the Ribosomal Database Project and BLAST searches against Greengenes database to classify the novel bacterial sequences that would correspond to SG1, *Mycoplasma*-like sequences and OTU4. The 16S rRNA of these novel lineages showed low levels of sequence identity to any other known bacteria in public databases and a distant relationship with the Mycoplasmatales and the Entomoplasmatales orders in Tenericutes phylum. Therefore, we conducted a phylogenetic analysis using type strain sequences of the 16S rRNA gene from the main taxa of the Tenericutes along with the novel sequences (Fig. 3A). SG1 is placed as a sister clade of the Pneumoniae group on a rather long branch that precludes a finer phylogenetic affiliation. The scorpion species-specific *Mycoplasma*-like MCl and MVs were placed as a new sister clade to the Hominis group, while OTU4 seems to have a common origin and diverged from them. MCl and MVs were well differentiated from one another and even more so from OTU4. Finally, in order to determine if SG1 represents a new bacterial phylum or just a basal member of the Tenericutes, we constructed a 16S rRNA gene phylogenetic tree of bacteria with available sequences from 54 bacterial phyla (including MCl, MVs, and SG1) and using archaeal sequences as an outgroup (Fig. 3B). Regardless of the high sequence divergence of the SG1, 16S rRNA phylogeny placed SG1 as a basal long branch in the Tenericutes phylum and not as an isolated branch representing a new phylum. The placement as a new basal long branch in the Tenericutes phylum tree and the sequence divergence supports the consideration of SG1 as a new bacterial lineage. Regarding the phylogenetic placement of the *Mycoplasma*-like sequences, they form a subclade within Tenericutes, representing a new lineage in this phylum. In summary, our phylogenetic analysis clearly demonstrates the presence of previously not described bacterial diversity within our dataset.

Conservation of functional capabilities and overrepresentation of toxic compound degradation pathways in recently captured C. limpidus *scorpions*

We carried out a functional analysis to test whether the microbial composition of each treatment group provided different metabolic capabilities to the scorpions. Notably, over the last years, there have been attempts to predict

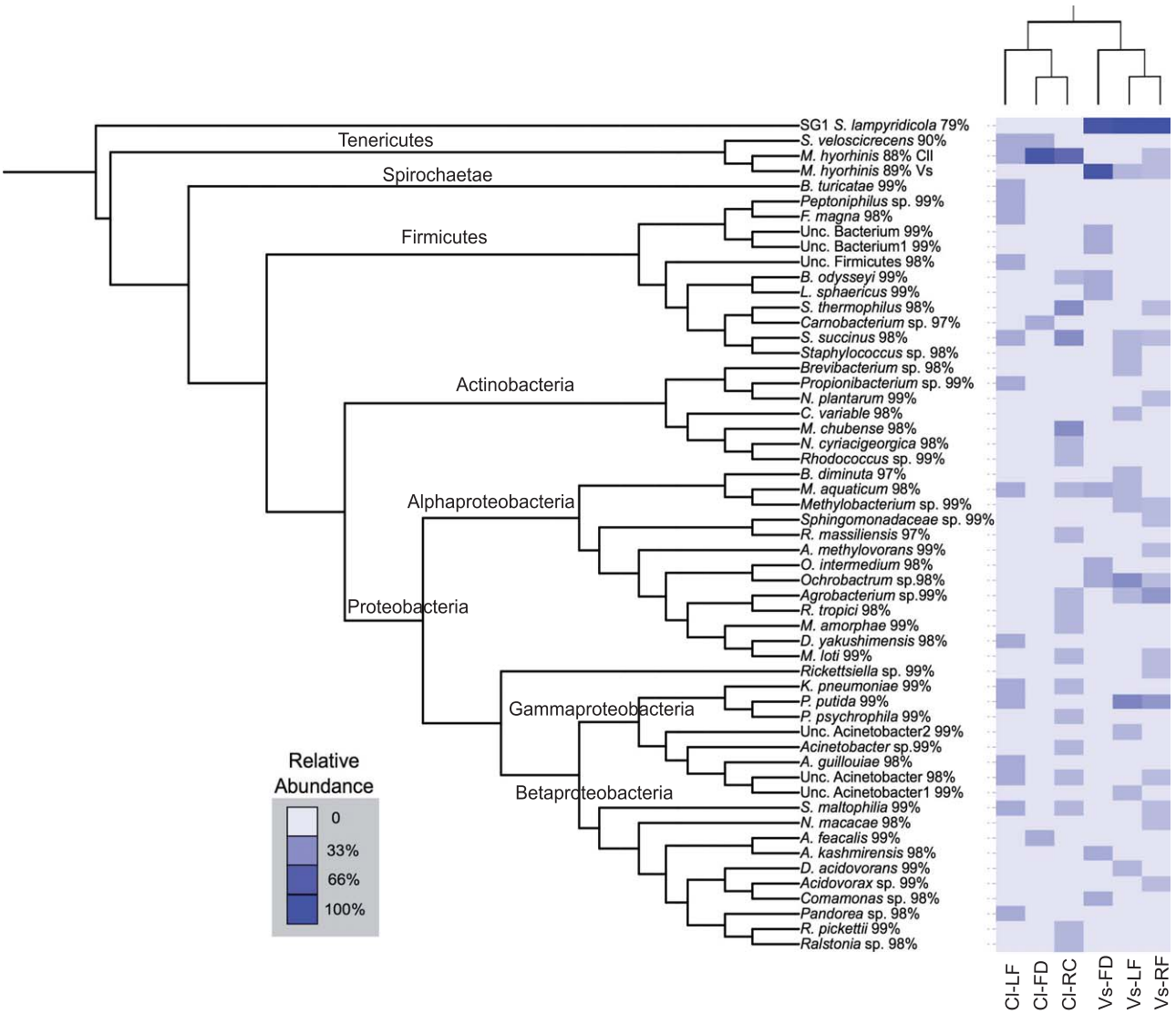

Fig. 1. Relative abundance heat map for the operational taxonomic units (OTUs) across treatment groups from both scorpion species. The range of colours indicates the OTUs' relative abundance for each group: from light blue (presence in 0% of specimens) to dark blue (presence in 100% of specimens). The 55 OTUs phylogenetic tree (RAxML 100 bootstrap replicates) is shown on the left hand side; the tip of the branches indicates the name of the closest sequence in the database and their identity. Euclidian distance clustering dendrogram is shown on top of the graph. Cl, *C. limpidus*; Vs, *V. smithi*; RC, recently captured; LF, laboratory-fed; FD, food-deprived.

the genomic diversity based on the ribosomal diversity (Zaneveld *et al*., 2010). For instance, Langille and colleagues recently developed PICRUST, a method to predict functional profiles from 16S rRNA sequences (Langille *et al*., 2013). We used this approach and extrapolated the presence of gene content and metabolic pathways from the genomes sequences of related taxa based on the 16S clone library analysis (see *Experimental procedures*). The PICRUST programme was used to calculate KEGG orthologue gene abundance by a hidden state prediction algorithm based on the phylogenetic position of the 16S rRNA gene and the available sequenced genomes in the clade. The whole dataset takes into account the abundance of clones of the treatment groups. It is important to highlight that sequences that did not fall in any OTU referenced in the Greengenes database with at least 97% identity were excluded from the prediction; therefore SG1, OTU4, MCl and MVs were not included. Comparison of the predicted gut microbiome of recently captured, food-deprived and laboratory-fed scorpions showed no significant differences at the gene metabolic categorization (Fig. S2A and B); this was true for both scorpion species. The more abundant functional gene categories across the different treatments in both scorpion species were membrane transport, amino acid metabolism and carbohydrate metabolism.

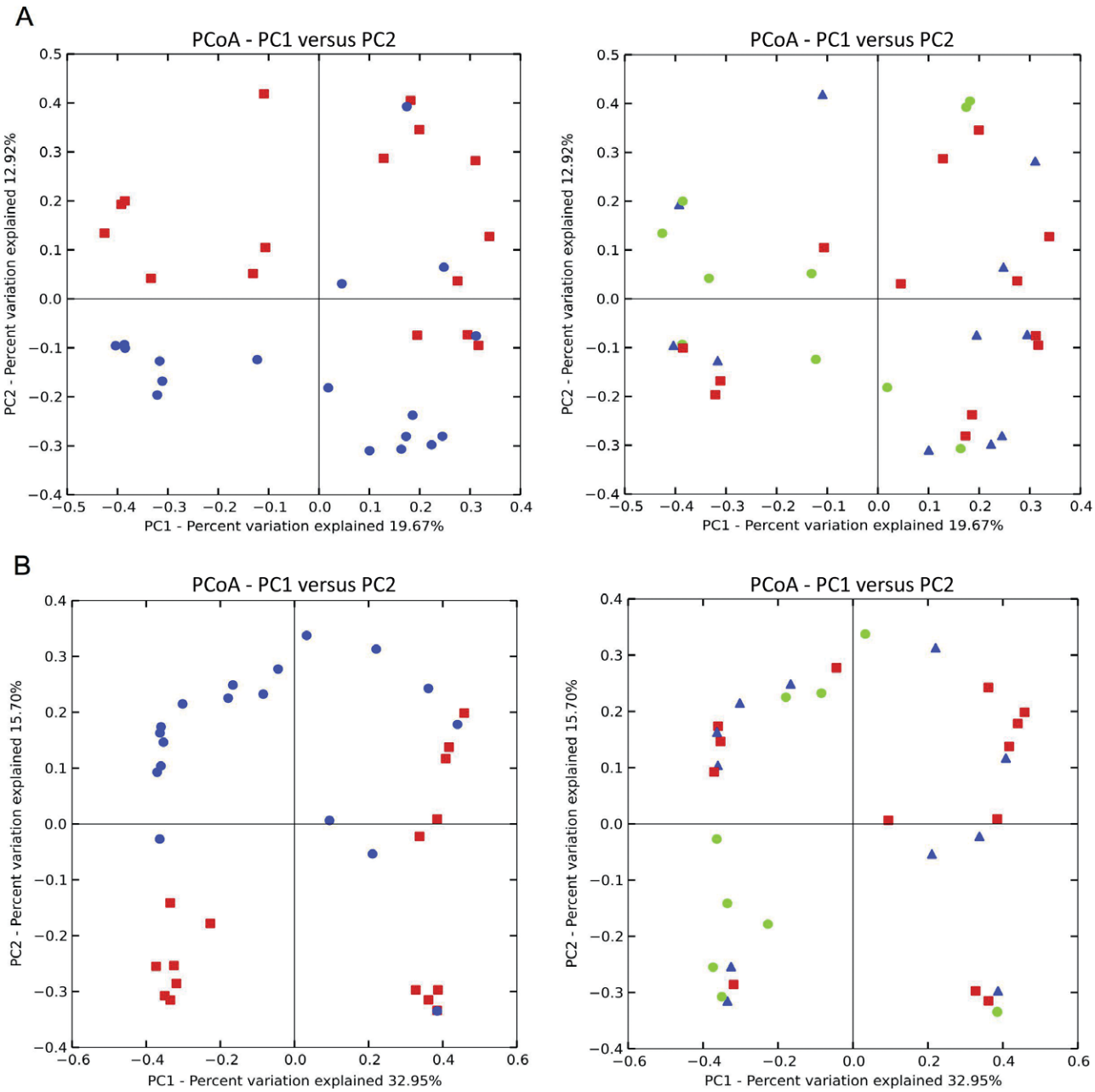

Fig. 2. Principal coordinate analysis of the 16S rRNA gene clone libraries. The ordination was constructed using weighted and unweighted UniFrac distances. Only the first two principal components are shown. Samples tagged for scorpion species (left panels): clone libraries from *C. limpidus* are shown in red squares and those from *V. smithi* in blue circles. Samples tagged for treatment group (right panels): recently captured group in red squares, laboratory-fed group in blue triangles and food-deprived group in green circles: (A) unweighted PCoA; (B) weighted PCoA.

The abundance of the orthologous gene families was used to reconstruct whole pathways using the MinPath approach (Ye and Doak, 2009) implemented in HUMAnN (Abubucker *et al.*, 2012). We define the core cluster of pathways from scorpion gut microbiota prediction profiles as follows: we considered a pathway to be a core pathway when its relative abundance was over 0.01 and was present in the six groups. Forty-one metabolic pathways were considered as core and mainly composed of pathways related to amino acid, carbohydrate, vitamin and cofactor metabolism (which together accounted approximately for 68% of core pathways). Among the remaining pathways, we found functions such as chloroalkane and chloroalkene degradation for detoxification, streptomycin biosynthesis, terpenoid backbone biosynthesis, and carbon fixation in prokaryotes (Fig. 4A).

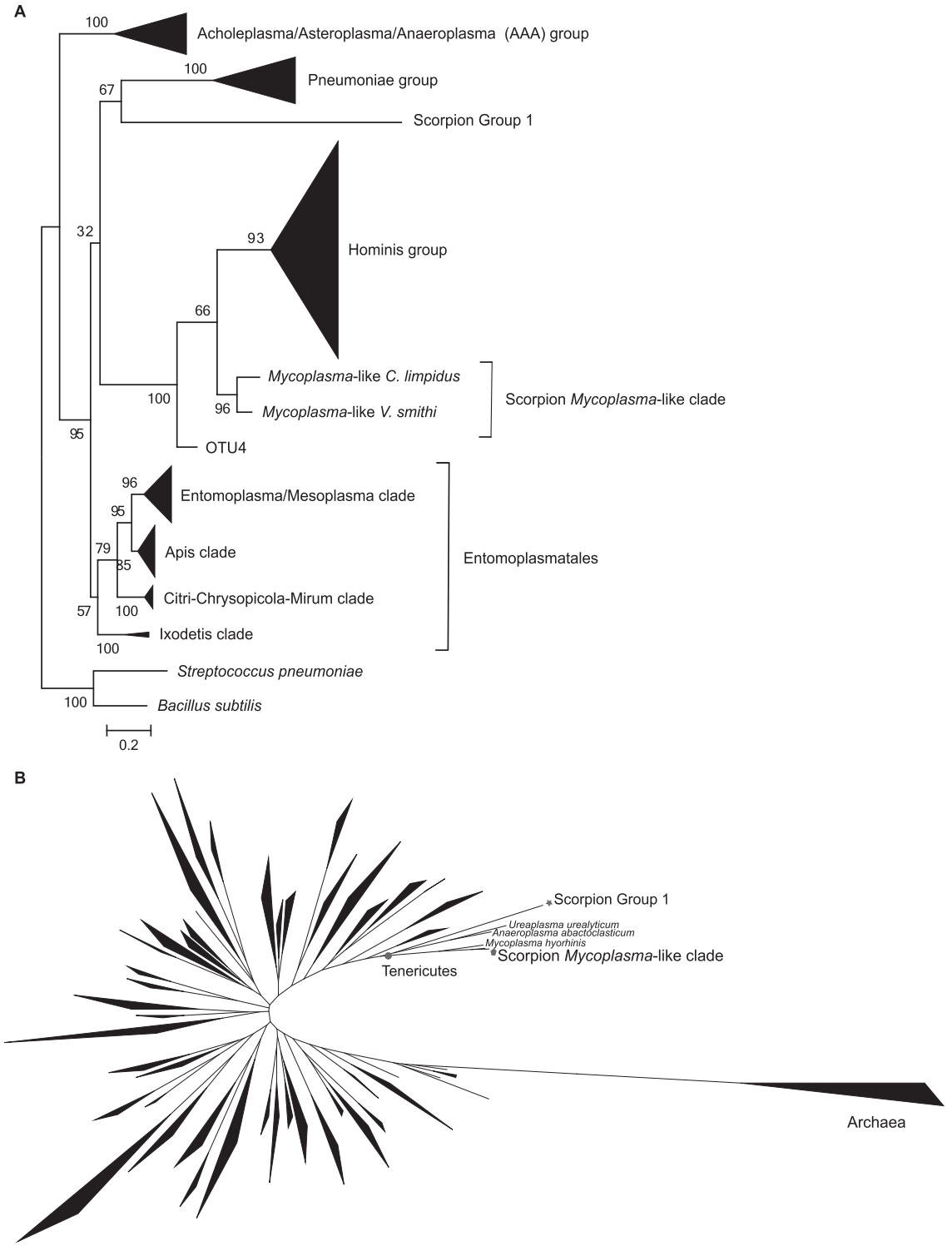

Fig. 3. Maximum likelihood phylogeny of the 16S rRNA gene sequences. Phylogenetic trees define the position of the novel lineages in the Tenericutes phylum and the bacteria domain.
A. Phylogenetic tree of the Tenericutes phylum constructed with 171 different species and the new lineages found in the scorpions. SG1 forms a long branch and shares a common ancestor with the pneumoniae group. MCl and MVs form a new sister clade to the Hominis group. *Streptococcus pneumoniae* and *Bacillus subtilis* were used as outgroups.
B. Phylogenetic tree of bacteria with three to five representative sequences for each phylum rooted with five archaeal sequences. The filled circle indicates the origin of the Tenericutes phylum. The star denotes the SG1 group, which forms a basal long branch of the Tenericutes phylum. MCl, MVs and OTU4 form the scorpion *Mycoplasma*-like clade (a pentagon) within the Tenericutes phylum.

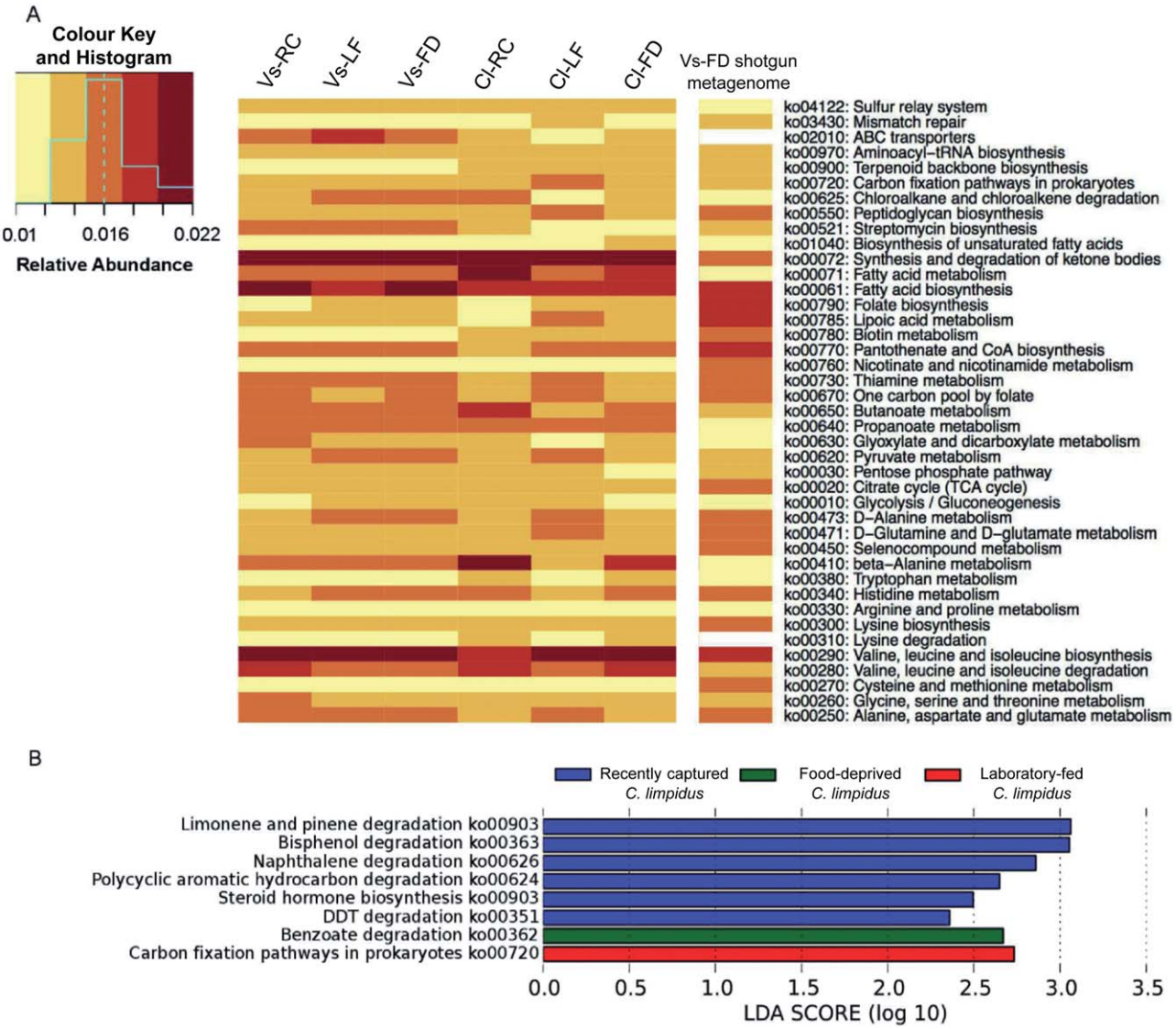

Fig. 4. Relative abundance heat map for the core pathways across treatments groups of both species. The range of colours indicates the average pathway relative abundance for each group: from light yellow (0.01) to dark red (0.022).
A. Set of pathways determined as the core cluster of metabolic pathways present in all the groups with a relative abundance greater than 0.01. The heat map is divided between predictions (left) and the shotgun metagenome from food-deprived scorpions (right). White colour (blank space) indicates that the pathway has a relative abundance below 0.01.
B. Linear discriminant analysis (LDA) of the pathways' relative abundance between diet-species groups. Degradation of toxic compounds such as bisphenol, limonene, pinene, naphthalene, dichlorodiphenyltrichloroethane and polycyclic aromatic hydrocarbons were more abundant in the recently captured *C. limpidus* scorpions. Only pathway categories meeting an LDA significant threshold of 2.0 are shown.

To determine which metabolic pathways are significantly more abundant under each treatment and species, we applied a non-parametric Wilcoxon sum-rank test, followed by a linear discriminant analysis (LDA) to assess the effect size of each of the differentially abundant pathway with the LEfSe programme (Segata *et al.*, 2011). Comparisons between the metabolic pathway profiles of the six groups revealed that some degradation pathways for toxic compounds (bisphenol, limonene, pinene, naphthalene, dichlorodiphenyltrichloroethane and polycyclic aromatic hydrocarbons) and steroid hormone biosynthesis were more abundant in the recently captured group of *C. limpidus* scorpions compared with the other groups (Fig. 4B). The carbon fixation pathways not only were among the most abundant in all groups (present in the core pathways), but also were overrepresented in the laboratory-fed *C. limpidus*.

We did not find significant differences in gut bacterial metabolic functions between groups as far as the functional classification at gene level is concerned. However,

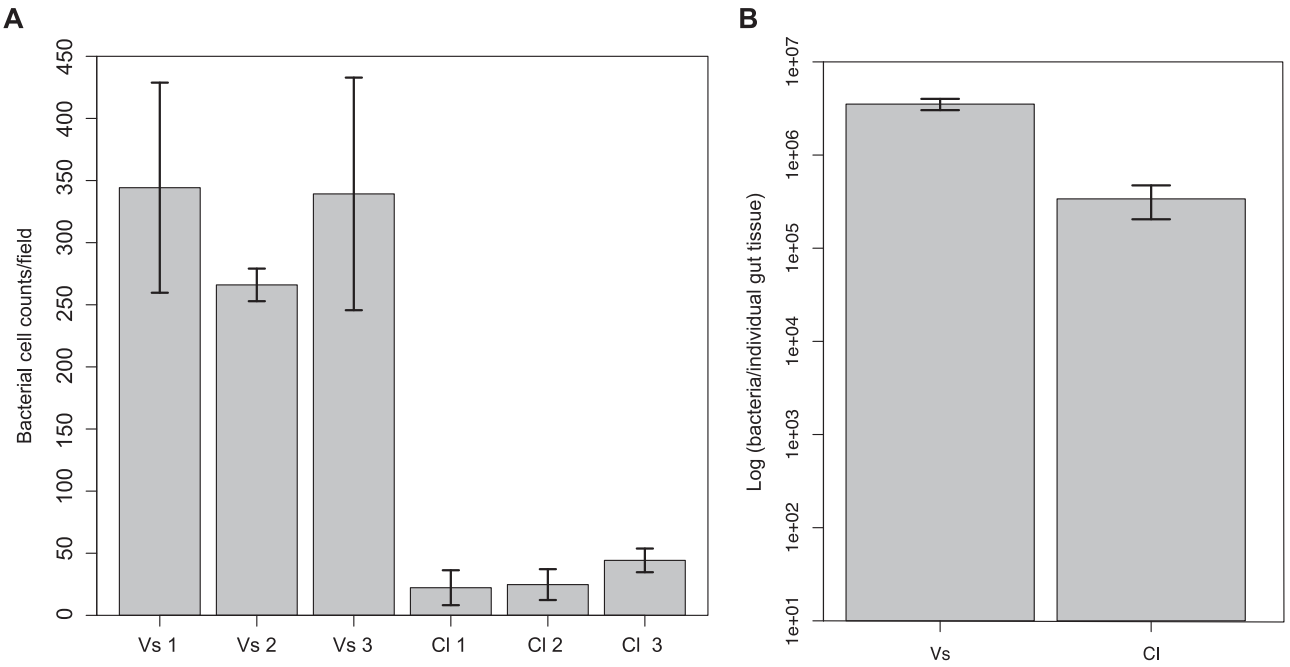

Fig. 5. Total bacteria quantification in food-deprived scorpions counted with confocal microscopy. (A) Total bacteria per field in each scorpion gut tissue. Three samples of each species were quantified. (B) Comparison between *V. smithi* and *C. limpidus* food-deprived groups. The average of the total counts (average of the fields) from the three samples is plotted. *C. limpidus* specimens have significantly reduced bacterial cell counts. Error bars represent standard deviations. Cl, *C. limpidus*; Vs, *V. smithi*.

at a pathway level, the presence of overrepresented pathways was biased to the *C. limpidus* species. The difference in the bacterial composition between scorpion species resulted in significant differential abundance of some pathways related to toxic compound degradation, hormone biosynthesis and carbon fixation. The defined core cluster of pathways was enriched in amino acid, carbohydrate, vitamin and cofactor metabolism, suggesting that these functions are conserved features in the gut microbiota regardless of species or treatment. Among the latter capabilities, the category of synthesis and degradation of ketone bodies is overrepresented in the core cluster. These results imply that although species and treatment influence the bacterial taxonomic composition, some gut microbiota functional capabilities are conserved. Nevertheless, species and treatment affect some traits, like enrichment of toxic compound degradation pathways in the recently captured *C. limpidus* scorpions.

Shotgun metagenomic sequencing of food-deprived scorpions and the dominance of Proteobacteria in the V. smithi *microbiome*

In order to infer and compare the gut microbiomes from both scorpion species in a food-deprived state and to validate the functional prediction, whole shotgun metagenomic sequencing was performed. Independent assemblies of data generated by 454 GS FLX Titanium and Illumina technologies together yielded 12 138 contigs for *V. smithi* (N50 = 1553) and 12 042 contigs for *C. limpidus* (N50 = 513). Both sets of contigs were restricted to have a minimal length of 250 bp. Gene predictions on the contigs estimated 18 009 open reading frames (ORFs) for *V. smithi* and 11 803 ORFs for *C. limpidus*. We obtained a similar number of contigs for both species, although N50 and ORF predictions indicated that *C. limpidus* dataset was underrepresented. Then, in order to focus exclusively on bacterial gene contents, we conducted BLASTP searches (of the predicted ORFs in MEGAN taxonomic classifier) and discarded all the ORFs that had a match with eukaryotic sequences. We kept 9048 predicted bacterial ORFs for *V. smithi* and only 28 for *C. limpidus* (Fig. S3A); therefore, there were much fewer bacterial reads in *C. limpidus* than in *V. smithi* metagenomes from the food-deprived samples. This might suggest that there were fewer bacteria in *C. limpidus* guts. To corroborate this, we quantified the bacterial cells from those samples via confocal microscopy. Notably, there were considerably less bacterial cells inside the guts from *C. limpidus* than in *V. smithi* food-deprived scorpions (chi-square test, *P*-value < 2.2e-16, the difference is almost one order of magnitude Fig. 5B). Of note, we found homogeneous bacterial load within the different individuals of each species (Fig. 5A). Thus, the low number of bacterial reads

detected in *C. limpidus* very likely is due to the considerably lower number of bacteria in *C. limpidus* compared with *V. smithi*.

Alpha- and *Gammaproteobacteria* accounted for 95.5% of the bacterial ORFs in the *V. smithi* dataset (Fig. S3B). In each of the classes, only one genus was the major contributor; for example, *Stenotrophomonas* accounted for 90.2% of the ORFs in the *Alphaproteobacteria*, whereas *Brevundimonas* explained 75.5% of the ORFs in the *Gammaproteobacteria*. A total of 4292 predicted bacterial ORFs from the *V. smithi* metagenome were associated with KEGG pathways, where carbohydrate metabolism and amino acid metabolism accounted for the 24.5% of the assignations (Fig. S3C).

To validate the functional predictions, we compared the *V. smithi* deprivation shotgun metagenome ORFs with the predicted functional profile of that group. We reconstructed whole metabolic pathways using the MinPath approach in HUMAnN based on the BLASTP results of the bacterial ORFs. First, we compared the abundance of the 'core' pathways determined from the predictions with the abundance of the same 41 pathways in the shotgun metagenome. From the shotgun metagenome set, only two categories – lysine degradation and ABC transporters – have lower abundance than 0.001 (Fig. 4A). This suggests that the core cluster of pathways determined from the predictions is a good proxy to determine the conserved metabolic features. As a second approach to validate the metagenome predictions, we conducted an LDA of pathway abundance between *V. smithi* food-deprived predicted metagenome and the shotgun sequencing. The analysis showed no significant differences in any of the pathway categories (the threshold on the logarithmic LDA score was 2.0).

Our analysis of the shotgun metagenomic sequencing of food-deprived scorpions reveals a lower bacterial abundance in *C. limpidus*, compared with *V. smithi*, and a high frequency of Proteobacteria in the *V. smithi* microbiome. Furthermore, the food-deprived *V. smithi* shotgun metagenome matched almost completely the metabolic features of the prediction, suggesting that predictions based on 16S rRNA sequences are a practical and reliable approach to determine the metagenomic profiles of the scorpion gut microbiota.

Discussion

Although scorpion species have evolved to survive in a wide variety of environmental conditions and have a clear medical importance, their gut microbiota had not been studied. Clearly, the majority of the studies of arthropod gut microbiota have targeted the insect class. Only few studies have started to elucidate the composition and functional properties of arachnid gut microbiota (Narasimhan *et al*., 2014). Thus, the main aim of this study was to characterize the gut microbiota of scorpions.

Notably, our 16S rRNA clone library analysis revealed great variation among individual samples; the number of OTUs within the libraries ranged from 1 to 8. This relatively low bacterial diversity has been previously observed in many other gut microbial communities of arthropods (Koch and Schmid-Hempel, 2011), with the exception of some termites (Warnecke *et al*., 2007). Of the total OTUs found in the 16S rRNA clone libraries survey, almost 60% belonged to just a few classes within the Proteobacteria (*Alpha*, *Beta*, *Gamma*), whereas the remaining OTUs belonged to the Firmicutes, Actinobacteria and Spirochaetes phyla – all these are known to be common colonizers of insect guts (Engel and Moran, 2013; Yun *et al*., 2014). The bacterial diversity, however, does not seem to be the same for the two scorpion species. We found that each scorpion species gut microbiota has a differential bacterial species composition. For instance, a novel bacterial lineage, namely SG1, was only found in the microbiota of *V. smithi* scorpions, present in approximately 94% of the specimens. Remarkably, this lineage is only 79% identical to the closest sequence (*Spiroplasma lampyridicola*) in the NCBI nr database. Furthermore, it could be the case that this lineage might not even be a member of the phylum Tenericutes, as SG1 sits on a long branch in the phylogenetic tree and has a basal position. Thresholds used to describe new bacterial lineages proposed that a 16S rRNA sequence with an identity equal or lower than 75% with another known type should be considered a different bacterial phylum and 95% a different genus (Yarza *et al*., 2014). However, it is known that some Tenericutes, such as *Spiroplasma* and *Mycoplasma*, dissatisfy these cut-offs (Bolaños *et al*., 2014). Although SG1 might constitute a new bacterial phylum, more gene markers – or even complete genome sequences – along with biochemical characterization are needed to firmly establish the taxonomic level of these novel bacteria. Additionally, we found another putative novel clade of bacteria belonging to the Mollicutes class. The proposed clade contains two different lineages, each one found in only one of the two different scorpion species. MVs and MCl are approximately 89% and 88% identical to *Mycoplasma hyorhinis* GDL-1, respectively, and are a sister clade to the Hominis group. Mollicutes bacteria are found in insects, and we are starting to elucidate the symbiotic interactions with non-insect arthropods. For instance, Candidatus *Hepatoplasma crinochetorum* colonizes the hepatopancreatic gland of *Porcellio scaber* isopod (Wang *et al*., 2004). Furthermore, '*Candidatus Hepatoplasma*' group seems to be common in many terrestrial isopods (Fraune and Zimmer, 2008). The finding of these novel bacterial lineages in just two scorpion species warrants further studies to unveil more of the bacterial

diversity likely to be found in the gut microbiota of other scorpion species. Additionally, a more comprehensive characterization of this novel bacterial diversity by next-generation sequencing and biochemical tests for cultivable isolates is needed to better understand how this unreported bacterial diversity is organized.

Food deprivation caused a significant reduction in bacterial diversity in the gut of both scorpion species. This was not completely unexpected, as it has been shown previously that, in locusts, starvation reduced gut bacterial diversity (Dillon *et al*., 2010). On the other hand, this treatment produced the enrichment of some taxa. For instance, we noted a proportional increase of *Mycoplasma*-like sequences in both species. We think this effect may be due to a reduction in the number of transient bacteria and the persistence of some taxa (in the case of the *Mycoplasma*-like bacteria) under stressful conditions. The fact that these novel *Mycoplasma*-like organisms were still found under the starvation condition could indicate a stable relationship with scorpions.

Functional bacterial predictions show a considerable number of core pathways; mainly related to amino acid, carbohydrate, vitamin and cofactor metabolism. However, we found a differential abundance of categories in the different treatment groups. Pathways for toxic compound degradation were more abundant in recently captured *C. limpidus* scorpions, suggesting that this could be an important function of the microbiota in this species' natural environment. It is known that some compounds, such as terpenoids, are toxic to some insects (Tripathi *et al*., 2003). Plants produce terpenoids, such as limonene and pinene. These are involved in repelling insect pests and pathogens; in addition, they are attractors to insects for herbivore control, pollination and seed dispersal (Rodríguez *et al*., 2011; Pontin *et al*., 2015). Other substances, such as benzoate and naphthalene, and some polycyclic aromatic hydrocarbon are also present in plants and have similar properties (Dudareva *et al*., 2004; Piskorski *et al*., 2011; Wöll *et al*., 2013). So, we expect that many insects that feed on plants may harbour them, and as scorpions feed on these insects they need bacteria that are able to degrade these compounds. In line with our findings, it has been shown that gut microbiota help degrade these compounds (Adams *et al*., 2011). The abuse and overuse of insecticides spread in the environment may lead also to the selection of bacteria capable of degrading these toxic compounds in *C. limpidus* scorpions.

Shotgun metagenomic sequencing of samples from food-deprived specimens showed a lower bacterial abundance and diversity in *C. limpidus* compared with *V. smithi*. Bacterial contigs from *C. limpidus* were scarcely found and were not included in the analysis to determine the metabolic pathway abundance comparison. Microscopy cell counts showed a very low abundance of bacteria in food-deprived *C. limpidus*, which would explain the scarcity of reads in its metagenome. We observed that *C. limpidus* scorpions had a larger mortality rate after capture and in the food-deprived treatment compared with *V. smithi* (data not shown). Additionally, *C. limpidus* scorpions are less active after food deprivation; thus, they seem less tolerant to the deprivation treatment, which seemingly affects negatively their gut microbiota. The decrease of bacterial loads in *C. limpidus* could be a strategy to conserve energy or nutrients and decrease metabolic rates for long food deprivation periods.

Shotgun metagenome of *V. smithi* matched almost completely the core pathways abundance and showed no significant difference of pathways abundance compared with the metagenome prediction.

On the whole, our study provides an initial characterization of the scorpion gut microbiota that sets a point of reference for further studies on the gut microbiota of scorpions and other arachnid species.

Experimental procedures

Scorpions, feeding treatments and dissection

C. limpidus and *V. smithi* specimens (16 and 18, respectively) were collected in the urban area of Cuernavaca, Morelos. The specimens of each species were split into three groups: recently captured scorpions, food-deprived and laboratory-fed. The groups are as follows: six recently captured individuals, five food-deprived and five laboratory-fed individuals for *C. limpidus*; and seven recently captured scorpions, five food-deprived and six laboratory-fed individuals for *V. smithi*. Recently captured scorpions were dissected 1 or 2 days after being sampled; food-deprived scorpions were kept individually isolated in the laboratory without any food for 30 days, and the laboratory-fed group was kept in the laboratory with one *Tenebrio molitor* larvae every 7 days for 30 days. Scorpions were anaesthetized by placing them in closed containers with chloroform, and surface was disinfected by rinsing it with three cycles of 70% ethanol–sterile water rinses. Gut dissections were performed under sterile conditions, during which the hepatopancreatic gland was removed.

DNA extraction

Each gut tissue (lacking the hepatopancreatic gland) was placed in 100 μl of sterile PBS. The tissue was macerated with sterile polypropylene micro pestle inside a 1.5 ml tube. The tubes were centrifuged at low speed to pellet the macerated of gut tissue. DNA from the pellets was extracted using the UltraClean Microbial DNA Isolation Kit (MoBio Laboratories, Solana Beach, CA, USA) according to the manufacturer's directions. DNA was observed in a 0.8% agarose gel (90 V, 35 min).

PCR amplification and 16S rRNA gene clone libraries

DNA extracted from each tissue was used as template for PCR amplification of community 16S rRNA using the uni-

versal primers fD1 and rD1 (Weisburg *et al*., 1991). Final concentrations for 20 μl PCR reactions were as follows: 1 μl DNA (25 ng μl^{-1}), 0.2 nM of each primer, 0.2 mM dNTPs, 5 U of *Taq* DNA Polymerase (Invitrogen, Carlsbad, CA, USA), 1X *Taq* polymerase buffer and 1.5 mM MgCl. The reaction conditions were 94°C for 3 min, 30 cycles at 94°C for 50 s, 56°C for 1 min, 72°C for 2 min and a final extension at 72°C for 10 min. Near full-length (approximately 1400 bases) 16S rRNA gene sequences were amplified in triplicate. PCR products were cloned into pCR 4-TOPO (Invitrogen) cloning vector and transformed in *Escherichia coli* DH5α competent cells. Transformed cells were plated on Luria–Bertani agar with carbenicillin 100 μg ml^{-1} and kanamycin 30 μg ml^{-1}, and incubated overnight at 37°C. Inserts from transformed colonies were amplified with M13R and M13F primers. Correct size amplicons were subjected to an ARDRA using restriction enzymes *Rsa1* (10 U) and *HindIII* (10 U) (New England Biolabs, Ipswich, Massachusetts, USA). Finally, representative isolates from each ARDRA pattern were sequenced.

16S rRNA sequence analysis

Sequences were checked for chimeras using Bellerophon (Huber *et al*., 2004) and manually inspected for quality. After chimera filtering, the remaining sequences were aligned using CLUSTALW (Thompson *et al*., 1997) with the default settings. Sequence alignment was used to construct a distance matrix with the DNADIST programme from the PHYLIP package (Felsenstein, 1989). The distance matrix was the input for the MOTHUR programme (Schloss *et al*., 2009), with which sequences were clustered at a distance of 0.03 in order to define the OTUs. Diversity indexes and rarefaction curves of each group of scorpions were evaluated. Additionally, we used SPLITSTREE4 programme (Huson and Bryant, 2006) to calculate the phylogenetic diversity (Faith, 1992).

BLASTN (Altschul *et al*., 1990) and the Ribosomal Database Project (Maidak *et al*., 2001) were used to classify the 16S rRNA gene sequences that were deposited in the NCBI GenBank database under accession numbers KM978218–KM978331. An un-rooted tree generated by PHYML (Guindon *et al*., 2010) from OTUs and a chart for their relative abundance in the six scorpion groups were used as input to generate a heat map in the web page of Interactive Tree of Life, iTol (Letunic and Bork, 2007). A clustering cladogram – using Euclidian distances – of the scorpion groups based on the absence/presence of bacterial OTUs was constructed with HeatMap R package (R Core Team, 2014). Clustering cladogram was merged to the iTol phylogenetic heat map. Weighted and unweighted PCoA were conducted with the UNIFRAC online server (Lozupone and Knight, 2005) using as input a rooted tree file generated by the neighbour programme, which constructs neighbour-joining trees, of the PHYLIP package (Felsenstein, 1989) and an environment file that links each sequence to a clone library.

Phylogenetic analysis

Representative sequences of the denominated SG1, *Mycoplasma* amplified from *V. smithi* (MVs), *Mycoplasma* amplified from *C. limpidus* (MCl) and OTU4 were used to construct a Tenericutes phylogeny. Additionally, a bacterial domain phylogeny was assembled with the above representative sequences except OTU4. The set of sequences to generate the Tenericutes phylogenetic tree was created from the Ribosomal Database Project and was composed of 171 species. *Bacillus subtilis* and *Streptococcus pneumoniae* 16S rRNA sequences were used as outgroup. The sequences were < 1200 bp, type strains and good quality. The set of sequences to generate the bacterial domain was created in the Greengenes database (DeSantis *et al*., 2006). The set consisted of 194 bacterial and archeal species, three to five representative sequences from each bacterial phylum and five archeal sequences that were used to root the tree. For both sets, sequences were aligned using CLUSTALW. Tenericutes phylogenetic tree was constructed with PHYML (Guindon *et al*., 2010) with NNI + SPR searches and 100 bootstrap replicates. Bacterial domain phylogenetic tree was constructed on RAXML (Stamatakis *et al*., 2008) using maximum likelihood and the GTR + γ model of evolution with 100 bootstrap replicates. Tenericutes tree was visualized and edited with the MEGA6 (Tamura *et al*., 2013). The bacterial domain tree was visualized in the iTol web page.

Metagenome predictions from 16S rRNA surveys and functional analysis

Metagenomic predictions of each of the three treatment groups of both species were made taking as input the clone abundance of previously predicted OTUs. The dataset that consisted of the total six sets was inputted to QIIME (Caporaso *et al*., 2010) using a 'closed-reference' OTU picking protocol. Sequence search was done against the Greengenes reference collection at 97% identity. The biom-formatted OTU table obtained from QIIME was used as input to PICRUST (Langille *et al*., 2013), which estimates the gene families content of bacteria for which no genome sequence is available, using their sequenced relatives as a reference. The OTU table was normalized and used to create the final metagenomic functional predictions. We used QIIME to plot the functional categories collapsing the PICRUST predictions to the hierarchical KEGG levels 2 and 3. Downstream pathway coverage prediction from the KEGG orthologue datasets of the predictions and the shotgun metagenomic sequencing was done with HUMANN version 0.99 (Abubucker *et al*., 2012) and visualized with GraPhlAn (Asnicar *et al*., 2015). HUMANN predicted pathways were subjected to a differential abundance analysis with LEfSe (Segata *et al*., 2011). For the factorial Kruskal–Wallis test among classes and the pairwise Wilcoxon test between subclasses, we used a significance level (alpha) of 0.05. The threshold on the logarithmic LDA score to discriminate significantly abundant pathway categories was 2.0.

Shotgun metagenomic sequencing

Total DNA from 10 pooled guts of food-deprived *V. smithi* and *C. limpidus* was extracted, sheared and sized to produce DNA whole-genome shotgun library according to the manufacturer's protocol (GS FLX Titanium General Library Preparation Kit Roche Applied Science, USA). DNA sequencing

was performed on a 454 GS FLX Titanium platform according to manufacturer's instructions (Roche 454 Life Sciences, USA) by a sequencing provider (Macrogen, Korea). One-eighth plate corresponded to a sample's 3 kb mate pair. The *V. smithi* and *C. limpidus* samples yielded a total of 87 373 and 141 214 reads, respectively. The total pooled DNA extracted from three *V. smithi* and three *C. limpidus* guts were sequenced by means of an Illumina HiSeq 1000 platform (Illumina, USA) 72×72 bp paired-end reads in multiplex, one sixth of lane per sample, at the sequencing unit of the National Autonomous University of Mexico. The *V. smithi* sample yielded a total of 15 511 990 reads and the *C. limpidus* sample 12 860 650 reads.

Filtering, assembly, annotation and analysis of shotgun metagenomic sequences

Filtering analysis was performed with PRINSEQ (Schmieder and Edwards, 2011) for 454 and Illumina reads. After filtering the 454 GS FLX datasets, we kept 81 279 and 134 812 reads from *V. smithi* and *C. limpidus*, respectively, and for the Illumina data 12 931 972 reads from *V. smithi* and 10 351 398 from *C. limpidus*.

Metagenomic *de novo* assembly was generated independently for each dataset – Illumina and 454 reads – with IDBA-UD (Peng *et al.*, 2012) with a minimum contig length of 250 bp. Validation of contigs was achieved by mapping the reads against the contigs with Bowtie (Langmead *et al.*, 2009) for Illumina paired-end reads and NEWBLER 2.5 for 454 reads. Gene calling of contigs was performed with FragGeneScan 1.18 (Rho *et al.*, 2010). Taxonomic assignment of the putative coding genes was performed via BLASTP searches against the NCBI protein nr database with an e-value of 1×10^{-1}. Annotation results for BLASTP against the NCBI nr database were loaded into the MEtaGenome ANalyzer 4 (MEGAN 4) software (Huson *et al.*, 2011). We selected only the contigs that contained predicted coding genes, which align with bacterial sequences. A function-based classification was determined using KEGG and COG identifiers, according to the parameters for the lowest common ancestor algorithm (maximum number of match per coding sequence: 10; min support: 5; min score: 50; and top per cent: 10). Bacterial BLASTP results were used as HUMAnN input with a contig-reads alignment to predict full pathway abundance in the dataset and compare it with the metagenome predictions. Shotgun bacterial metagenomic sequence data have been deposited at the NCBI under Bioproject PRJNA266890.

Quantification of bacterial loads in food-deprived scorpions

A cultivation-independent method was used to quantify the bacterial loads in the gut tissue of food-deprived scorpions from both species. Filtered samples enriched with bacteria cells and stained with DAPI (4′,6-diamidino-2-phenylindole) were screened and counted under a confocal microscope.

Whole gut tissues from three *V. smithi* and three *C. limpidus* scorpions in food-deprived treatment were dissected and homogenized in PBS (as previously described). The homogenized samples containing scorpion and bacterial cells were centrifuged and then re-suspended in 500 μl of a fixative solution with 4% formaldehyde in PBS for 2 h. Samples were re-suspended in 500 μl of PBS. Fixated samples were passed through a 5 μm filter, enriching the samples with bacterial cells. Of each sample, 15 μl was stained with 2.5 μl of DAPI (300 nM). As a positive control for bacterial DAPI staining, we used a swab from *E. coli* DH5α, and the negative control was the PBS used during the bacterial enrichment steps. We chose randomly four fields from each slide, using a 60× objective mounted onto an FV100 confocal microscope (Olympus, Japan) in the National Laboratory of Advanced Microscopy, UNAM Campus Morelos facility. Total bacterial cells were counted manually in each field using IMAGEJ image analysis programme (Abràmoff *et al.*, 2004). Auto-fluorescence scorpion gut debris (non-uniform particles) was selectively discarded with the support of bright-field images. The average of bacterial cells per field was used to estimate the total bacterial count per sample with the next formula: Bacteria/sample = (bacteria average/field volume) × 60 μl. Bacterial cells per field values for each individual and total sample bacterial counts comparison between food-deprived *V. smithi* and *C. limpidus* were analysed (chi-square goodness-of-fit test) and graphed with R package.

Acknowledgements

Financial support was from CONACyT Basic Science (Grant Number 154453). We are thankful to the Programa de Doctorado en Ciencias Biomédicas from UNAM. L. M. Bolaños had a CONACyT PhD scholarship. We thank Soledad Ochoa-Mendez for technical assistance and M. Dunn for critically reading the manuscript. Authors have no conflict of interest to declare.

References

Abràmoff, M.D., Magalhães, P.J., and Ram, S.J. (2004) Image processing with ImageJ. *Biophotonics Int* **11:** 36–43.

Abubucker, S., Segata, N., Goll, J., Schubert, A.M., Izard, J., Cantarel, B.L., and Huttenhower, C. (2012) Metabolic reconstruction for metagenomic data and its application to the human microbiome. *PLoS Comput Biol* **8:** e1002358.

Adams, A.S., Adams, S.M., Currie, C.R., Gillette, N.E., and Raffa, K.F. (2010) Geographic variation in bacterial communities associated with the red turpentine beetle (Coleoptera: Curculionidae). *Environ Entomol* **39:** 406–414.

Adams, A.S., Boone, C.K., Bohlmann, J., and Raffa, K.F. (2011) Responses of bark beetle-associated bacteria to host monoterpenes and their relationship to insect life histories. *J Chem Ecol* **37:** 808–817.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol* **215:** 403–410.

Asnicar, F., Weingart, G., Tickle, T.L., Huttenhower, C., and Segata, N. (2015) Compact graphical representation of phylogenetic data and metadata with GraPhlAn. *PeerJ* **3:** e1029.

Bolaños, L.M., Servín-Garcidueñas, L.E., and Martínez-Romero, E. (2014) Arthropod–*Spiroplasma* relationship in the genomic era. *FEMS Microbiol Ecol* **91:** 1–8.

Brune, A., and Ohkuma, M. (2011) Role of the termite gut microbiota in symbiotic digestion. In *Biology of Termites: A Modern Synthesis*. Bignell, D.E., Roisin, Y., and Lo, N. (eds). Dordrecht, the Netherlands: Springer, pp. 439–476.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., and Knight, R. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7:** 335–336.

Cirimotich, C.M., Dong, Y., Clayton, A.M., Sandiford, S.L., Souza-Neto, J.A., Mulenga, M., and Dimopoulos, G. (2011) Natural microbe-mediated refractoriness to Plasmodium infection in *Anopheles gambiae*. *Science* **332:** 855–858.

Corby-Harris, V., Pontaroli, A.C., Shimkets, L.J., Bennetzen, J.L., Habel, K.E., and Promislow, D.E. (2007) Geographical distribution and diversity of bacteria associated with natural populations of *Drosophila melanogaster*. *Appl Environ Microbiol* **73:** 3470–3479.

DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., and Andersen, G.L. (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* **72:** 5069–5072.

Dillon, R.J., Vennard, C.T., Buckling, A., and Charnley, A.K. (2005) Diversity of locust gut bacteria protects against pathogen invasion. *Ecol Lett* **8:** 1291–1298.

Dillon, R.J., Webster, G., Weightman, A.J., and Charnley, A.K. (2010) Diversity of gut microbiota increases with aging and starvation in the desert locust. *Antonie Van Leeuwenhoek* **97:** 69–77.

Dudareva, N., Pichersky, E., and Gershenzon, J. (2004) Biochemistry of plant volatiles. *Plant Physiol* **135:** 1893–1902.

Dunlop, J.A. (2010) Geological history and phylogeny of Chelicerata. *Arthropod Struct Dev* **39:** 124–142.

Engel, P., and Moran, N.A. (2013) The gut microbiota of insects–diversity in structure and function. *FEMS Microbiol Rev* **37:** 699–735.

Faith, D.P. (1992) Conservation evaluation and phylogenetic diversity. *Biol Conserv* **61:** 1–10.

Felsenstein, J. (1989) PHYLIP-phylogeny interference package (version 3.2). *Cladistics* **5:** 164–166.

Francke, O.F. (2014) Biodiversidad de Arthropoda (Chelicerata: Arachnida ex Acari) en México. *Revista mexicana de biodiversidad* **85:** 408–418.

Fraune, S., and Zimmer, M. (2008) Host-specificity of environmentally transmitted *Mycoplasma*-like isopod symbionts. *Environ Microbiol* **10:** 2497–2504.

Geib, S.M., Jimenez-Gasco, M.D.M., Carlson, J.E., Tien, M., and Hoover, K. (2009) Effect of host tree species on cellulase activity and bacterial community composition in the gut of larval Asian longhorned beetle. *Environ Entomol* **38:** 686–699.

Gonzalez-Ceron, L., Santillan, F., Rodriguez, M.H., Mendez, D., and Hernandez-Avila, J.E. (2003) Bacteria in midguts of field-collected *Anopheles albimanus* block *Plasmodium vivax* sporogonic development. *J Med Entomol* **40:** 371–374.

González-Santillán, E. (2001) Catálogo de escorpiones de la Colección Nacional de Arácnidos (CNAN). Dissertation, Tesis. DF, Mexico: Facultad de Ciencias, Universidad Nacional Autónoma de México.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59:** 307–321.

Guo, L., Karpac, J., Tran, S.L., and Jasper, H. (2014) PGRP-SC2 promotes gut immune homeostasis to limit commensal dysbiosis and extend lifespan. *Cell* **156:** 109–122.

Hadley, N.F. (1970) Water relations of the desert scorpion, *Hadrurus arizonensis*. *J Exp Biol* **53:** 547–558.

Hadley, N.F. (1974) Adaptational biology of desert scorpions. *J Arachnol* **2:** 11–23.

Huang, X.F., Bakker, M.G., Judd, T.M., Reardon, K.F., and Vivanco, J.M. (2013) Variations in diversity and richness of gut bacterial communities of termites (*Reticulitermes flavipes*) fed with grassy and woody plant substrates. *Microb Ecol* **65:** 531–536.

Huber, T., Faulkner, G., and Hugenholtz, P. (2004) Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. *Bioinformatics* **20:** 2317–2319.

Huson, D.H., and Bryant, D. (2006) Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* **23:** 254–267.

Huson, D.H., Mitra, S., Ruscheweyh, H.J., Weber, N., and Schuster, S.C. (2011) Integrative analysis of environmental sequences using MEGAN4. *Genome Res* **21:** 1552–1560.

Karsch, F. (1879) Scorpionologische Beiträge. I. *Mitt Münch Entomol Ver* **3:** 6–22.

Kikuchi, Y., Hayatsu, M., Hosokawa, T., Nagayama, A., Tago, K., and Fukatsu, T. (2012) Symbiont-mediated insecticide resistance. *PNAS* **109:** 8618–8622.

Klepzig, K.D., Adams, A.S., Handelsman, J., and Raffa, K.F. (2009) Symbioses: a key driver of insect physiological processes, ecological interactions, evolutionary diversification, and impacts on humans. *Environ Entomol* **38:** 67–77.

Koch, H., and Schmid-Hempel, P. (2011) Bacterial communities in central European bumblebees: low diversity and high specificity. *Microb Ecol* **62:** 121–133.

Langille, M.G., Zaneveld, J., Caporaso, J.G., McDonald, D., Knights, D., Reyes, J.A., and Huttenhower, C. (2013) Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotechnol* **31:** 814–821.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10:** R25.

Lee, F.J., Rusch, D.B., Stewart, F.J., Mattila, H.R., and Newton, I.L. (2014) Saccharide breakdown and fermentation by the honeybee gut microbiome. *Environ Microbiol* **17:** 796–815.

Letunic, I., and Bork, P. (2007) Interactive Tree of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23:** 127–128.

Lozupone, C., and Knight, R. (2005) UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* **71:** 8228–8235.

Maidak, B.L., Cole, J.R., Lilburn, T.G., Parker, C.T., Jr, Saxman, P.R., Farris, R.J., and Tiedje, J.M. (2001) The

RDP-II (Ribosomal Database Project). *Nucleic Acids Res* **29:** 173–174.

Morales-Jiménez, J., de León, A.V.P., García-Domínguez, A., Martínez-Romero, E., Zúñiga, G., and Hernández-Rodríguez, C. (2013) Nitrogen-fixing and uricolytic bacteria associated with the gut of *Dendroctonus rhizophagus* and *Dendroctonus valens* (Curculionidae: Scolytinae). *Microb Ecol* **66:** 200–210.

Narasimhan, S., Rajeevan, N., Liu, L., Zhao, Y.O., Heisig, J., Pan, J., and Fikrig, E. (2014) Gut microbiota of the tick vector *Ixodes scapularis* modulate colonization of the lyme disease spirochete. *Cell Host Microbe* **15:** 58–71.

Ohkuma, M. (2008) Symbioses of flagellates and prokaryotes in the gut of lower termites. *Trends Microbiol* **16:** 345–352.

Padilla, A., Govezensky, T., Possani, L.D., and Larralde, C. (2003) Experimental envenoming of mice with venom from the scorpion *Centruroides limpidus limpidus*: differences in mortality and symptoms with and without antibody therapy relating to differences in age, sex and strain of mouse. *Toxicon* **41:** 959–965.

Peng, Y., Leung, H.C., Yiu, S.M., and Chin, F.Y. (2012) IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28:** 1420–1428.

Piskorski, R., Ineichen, S., and Dorn, S. (2011) Ability of the oriental fruit moth *Grapholita molesta* (Lepidoptera: Tortricidae) to detoxify juglone, the main secondary metabolite of the non-host plant walnut. *J Chem Ecol* **37:** 1110–1116.

Pocock, R.I. (1902) Arachnida, Scorpiones, Pedipalpi and Solifugae. In *Biologia Centrali-Americana*. Godman, F.D., and Salvin, O. (eds). Londres: Taylor & Francis, p. 71.

Polis, G.A. (ed.) (1990) *The Biology of Scorpions*. Standford, CA, USA: Stanford University Press.

Pontin, M., Bottini, R., Burba, J.L., and Piccoli, P. (2015) *Allium sativum* produces terpenes with fungistatic properties in response to infection with *Sclerotium cepivorum*. *Phytochemistry* **S0031-9422:** 50–53. doi:10.1016/j.phytochem.2015.02.003.

Prendini, L. (2011) Order Scorpiones CL Koch, 1850. In Zhang, Z.-Q. (ed.) Animal biodiversity: an outline of higher-level classification and survey of taxonomic richness. *Zootaxa* **3148:** 115–117.

R Core Team (2014) R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing [WWW document]. URL http://www.R-project.org/.

Rein, J.O. (2012) The scorpion files. Trondheim, Norway: Norwegian University of Science and Technology [WWW document]. URL http://www.ntnu.no/ub/scorpion-files/.

Rho, M., Tang, H., and Ye, Y. (2010) FragGeneScan: predicting genes in short and error-prone reads. *Nucleic Acids Res* **38:** e191.

Ridley, E.V., Wong, A.C., Westmiller, S., and Douglas, A.E. (2012) Impact of the resident microbiota on the nutritional phenotype of *Drosophila melanogaster*. *PLoS ONE* **7:** e36765.

Rodríguez, A., San Andrés, V., Cervera, M., Redondo, A., Alquézar, B., Shimada, T., *et al*. (2011) The monoterpene limonene in orange peels attracts pests and microorganisms. *Plant Signal Behav* **6:** 1820–1823.

Rowe, A.H., Xiao, Y., Joseph Scales, J., Klaus, D., Linse, K.D., Rowe, M.P., *et al*. (2011) Isolation and characterization of CvIV4: a pain inducing α- scorpion toxin. *PLoS ONE* **6:** e23520.

Schloss, P.D., Delalibera, I., Jr, Handelsman, J.O., and Raffa, K.F. (2006) Bacteria associated with the guts of two wood-boring beetles: *Anoplophora glabripennis* and *Saperda vestita* (Cerambycidae). *Environ Entomol* **35:** 625–629.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., and Weber, C.F. (2009) Introducing Mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75:** 7537–7541.

Schmieder, R., and Edwards, R. (2011) Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27:** 863–864.

Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W.S., and Huttenhower, C. (2011) Metagenomic biomarker discovery and explanation. *Genome Biol* **12:** R60.

Sharon, G., Segal, D., Ringo, J.M., Hefetz, A., Zilber-Rosenberg, I., and Rosenberg, E. (2010) Commensal bacteria play a role in mating preference of *Drosophila melanogaster*. *PNAS* **107:** 20051–20056.

Shelomi, M., Lo, W.S., Kimsey, L.S., and Kuo, C.H. (2013) Analysis of the gut microbiota of walking sticks (Phasmatodea). *BMC Res Notes* **6:** 368.

Stamatakis, A., Hoover, P., and Rougemont, J. (2008) A rapid bootstrap algorithm for the RAxML web servers. *Syst Biol* **57:** 758–771.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* **30:** 2725–2729.

Tang, X., Freitak, D., Vogel, H., Ping, L., Shao, Y., Cordero, E.A., and Boland, W. (2012) Complexity and variability of gut commensal microbiota in polyphagous lepidopteran larvae. *PLoS ONE* **7:** e36978.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25:** 4876–4882.

Tripathi, A.K., Prajapati, V., Khanuja, S.P.S., and Kumar, S. (2003) Effect of d-limonene on three stored-product beetles. *J Econ Entomol* **96:** 990–995.

Vasanthakumar, A., Handelsman, J., Schloss, P.D., Bauer, L.S., and Raffa, K.F. (2008) Gut microbiota of an invasive subcortical beetle, *Agrilus planipennis* Fairmaire, across various life stages. *Environ Entomol* **37:** 1344–1353.

Wang, Y., Stingl, U., Anton-Erxleben, F., Geisler, S., Brune, A., and Zimmer, M. (2004) 'Candidatus hepatoplasma crinochetorum,' a new, stalk-forming lineage of Mollicutes colonizing the midgut glands of a terrestrial isopod. *Appl Environ Microbiol* **70:** 6166–6172.

Warnecke, F., Luginbühl, P., Ivanova, N., Ghassemian, M., Richardson, T.H., Stege, J.T., and Leadbetter, J.R. (2007) Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* **450:** 560–565.

Weisburg, W.G., Barns, S.M., Pelletier, D.A., and Lane, D.J. (1991) 16S ribosomal DNA amplification for phylogenetic study. *J Bacteriol* **173:** 697–703.

Werren, J.H., Baldo, L., and Clark, M.E. (2008) Wolbachia: master manipulators of invertebrate biology. *Nat Rev Microbiol* **6:** 741–751.

Wong, A.C., Chaston, J.M., and Douglas, A.E. (2013) The inconstant gut microbiota of *Drosophila* species revealed by 16S rRNA gene analysis. *ISME J* **7:** 1922–1932.

Wong, C.N.A., Ng, P., and Douglas, A.E. (2011) Low-diversity bacterial community in the gut of the fruitfly *Drosophila melanogaster*. *Environ Microbiol* **13:** 1889–1900.

Wöll, S., Kim, S.H., Greten, H.J., and Efferth, T. (2013) Animal plant warfare and secondary metabolite evolution. *Nat Prod Bioprospect* **3:** 1–7.

Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F.O., Ludwig, W., Schleifer, K.H., and Rosselló-Móra, R. (2014) Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat Rev Microbiol* **12:** 635–645.

Ye, Y., and Doak, T.G. (2009) A parsimony approach to biological pathway reconstruction/inference for genomes and metagenomes. *PLoS Comput Biol* **5:** e1000465.

Yun, J.H., Roh, S.W., Whon, T.W., Jung, M.J., Kim, M.S., Park, D.S., and Bae, J.W. (2014) Insect gut bacterial diversity determined by environmental habitat, diet, developmental stage, and phylogeny of host. *Appl Environ Microbiol* **80:** 5254–5264.

Zaneveld, J.R., Lozupone, C., Gordon, J.I., and Knight, R. (2010) Ribosomal RNA diversity predicts genome diversity in gut bacteria and their relatives. *Nucleic Acids Res* **38:** 3869–3879.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Fig. S1. Rarefaction curves of the 16S rRNA clone libraries from the treatment groups of each scorpion species. Rarefaction analysis was performed under an OTU threshold of 97% identity of the representative clone sequences for each individual library.

Fig. S2. KEGG gene categorization of the predicted profiles. The gene metabolic functions of microbiome predictions show a high consistency across samples, suggesting that they are not affected by diet treatment or scorpion species. No significant differences between the groups could be identified at different categorization levels. (A) Relative abundance of genes clustered in KEGG level 2 categories. (B) Relative abundance of the 20 most abundant gene clusters in KEGG level 3 categories.

Fig. S3. Shotgun metagenomic description of the assembled contigs for food-deprived scorpions. (A) Distribution of bacterial and non-bacterial contigs in each dataset. (B) Taxonomic distribution of the bacterial contigs in *V. smithi* bacterial fraction. *Gamma-* and *Alphaproteobacteria*, accounting for ≈ 95%, dominate the dataset. (C) Open reading frames from *V. smithi* shotgun metagenome assigned to KEGG metabolic categories. Carbohydrate and amino acid metabolism comprised the main fraction (24.5%) as in the predicted metagenomes.