The structured diversity of specialized gut symbionts of the New World army ants

Piotr Łukasik, Justin A. Newton, Jon G. Sanders, Yi Hu, Corrie S. Moreau, Daniel Kronauer, Sean O'Donnell, Ryuichi Koga, Jacob A. Russell

Online Supplementary Material

Supplementary Table legends

Table S1. Details of the New World army ant colonies used in this project. Collection data is provided for all New World Army Ant colonies presented in Figure 1.

Table S2. Details of Old World army ant and other ant colonies for which data were generated. The list includes the Old World army ant colonies presented in Figure 1, as well as other colonies that workers typed at 16S rRNA or *rplB* gene originated from.

Table S3. Details of individual worker samples from the experimental New World army ant colonies. The list includes accession numbers for all sequence data generated during this project.

Table S4. PCR primers and reaction conditions used in this project.

Table S5. Fluorescent microscopy probes used in this project.

Table S6. Summary of 16S rRNA amplicon sequencing analysis: library classification, and read numbers at various stages of the analysis.

Table S7. 16S rRNA amplicon sequencing data: unique genotypes remaining after contaminant and quality filtering. For each genotype, we provided sequence, OTU-level classification, taxonomy assignment, and information on the number of times it occurred in each library.

Table S8. 16S rRNA amplicon sequencing data: 97% OTUs. For each OTU, we provided a representative sequence, taxonomy assignment, and information on the number of times it occurred in each library.

Table S9. The list of OTUs that differ significantly in relative abundance when comparing between *Eciton burchellii* and *Labidus praedator*, across locations that these species were collected from, or among replicate colonies of *E. burchellii* from the same location. Samples and data shown in Figure 2 were used. Prior to analyses, libraries were rarified to 1399 reads, and the means provided reflect the average number of reads in a given OTU after rarefaction.

Table S10. Results of tests for group differences in dispersion among samples of *Eciton burchellii* and *Labidus praedator*, across locations that these species were collected from, or among replicate colonies from the same locations. Tests were conducted using function betadisper() within the vegan package in R. Samples and data shown in Figure 2 were used. Each line represents the result of a separate analysis conducted for a subset of samples indicated in column 1.

Table S11. 454 pyrosequencing of *rplB* amplicon of Unclassified Firmicutes: distribution of 98% OTUs across experimental libraries

Table S12. Summary of genotype-level diversity data for Unclassified Firmicutes in individual ant workers, based on 454 pyrosequencing of *rplB* amplicon. For the four dominant OTUs, each of which accounted for over 99.5% of reads from one library, we analyzed polymorphism in alignment of reads classified to that OTU. For each position in the alignment, we provided the proportion of reads with each of five possible states (A, T, G, C, gap), and we manually inspected all positions with less than 99% consensus. Note that the alignment reference had gaps introduced every two bases, in order to facilitate alignment of reads with insertions.

Table S13. Estimated colony sizes of Dorylinae species, based on published studies

Supplementary Figure legends

Figure S1. The effects of removal of putative contaminants on microbial community composition across libraries. A) The relative abundance of the dominant OTUs, including those regarded as real ant-associated bacteria and those identified as contaminants, across all sequenced libraries. Analysis has mostly followed Mothur's Standard Operating Procedure – see Supplementary Material for details. We classified as contaminants those OTUs which were not at least 10x more abundant in one of the experimental libraries than in any of the blanks; B) The relative abundance of the dominant OTUs in experimental libraries (analysis as in panel A) after the OTUs regarded as contaminants have been removed. Data is not shown for blanks or for experimental libraries deemed to be heavily contaminated; C) The relative abundance of the dominant OTUs in experimental libraries after analysis that included the custom contaminant filtering step – see Supplementary Material for details. This analysis has been used for the current study; D) The relative abundance of the dominant unique genotypes assigned to the most abundant OTU classified as a contaminant (Pseudomonas). In all panels, the same set and order of libraries has been used.

Figure S2. Principal coordinates analyses (PCoA) plots for *Eciton burchellii* and *Labidus praedator* colonies sampled in replicate. The samples and data from Figure 2 were used. Legends from the third (rightmost) plot in a panel refer to all plots in that panel.

Figure S3. Genotype-level diversity of specialized symbionts of army ants, based on amplicon sequencing of the V4 region of the 16S rRNA gene. A) Relative abundance of 97% OTUs representing the two dominant ant symbionts, Unclassified Firmicutes (red) and Unclassified Entomoplasmatales (yellow), in all experimental libraries. B) Relative abundance of the dominant unique genotypes classified to OTU001 (Unclassified Firmicutes) within that OTU. Colours represent the 12 most abundant unique genotypes in the dataset, and black represents all other (rare) unique genotypes combined. Data are shown only shown for libraries where OTU001 is represented by at least 100 reads. C) Relative abundance of the dominant unique genotypes classified to OTU002 (Unclassified Entomoplasmatales) within that OTU. Colors represent 12 most abundant unique genotypes in the dataset, and black represents all other (rare) unique genotypes combined. Data shown only for libraries where OTU002 was represented by at least 100 reads.

Figure S4. Strain-level diversity of Unclassified Firmicutes in four individual workers from *Eciton burchellii* colony PL028, based on 454 pyrosequencing of the *rplB* amplicon. Colors represent distinct 98% OTUs (symbiont strains).

Figure S5. Unconstrained ML phylogeny of Unclassified Firmicutes based on ribosomal protein L2 (*rplB*) gene. This tree is included as a counterpart to the phylogeny in Fig. 3B, constructed using the same data [573bp sequence of gene from 252 ant workers], yet with imposed constraints described in that figure's legend. Annotation details are the same as for Fig. 3B.

Figure S6. *rplB*-genotype level diversity within 16S rRNA genotypes of the two specialized symbionts of army ants. ML phylogenies based on 252-253bp sequence of 16S rRNA and 573-579bp sequence of rplB have been constructed for all samples for which unambiguous *rplB* sequence was obtained, and 16S rRNA amplicon data indicated the presence of a single dominant strain of a given symbiont. Black dots represent bootstrap support values of 95% and over.

Figure S7. The relationship between *rplB* genotypes of Unclassified Firmicutes and the microbial community composition in two army ant species. ML phylogenies based on a 573bp sequence of ribosomal protein L2 (*rplB*) gene (as in Fig. 3B) and the microbial community composition data based on 16S rRNA amplicon sequencing (as in Figs. 1 and 2) are constructed for all workers of a given species with both types of data generated. Bootstrap support values of 95% or more are represented by black dots on branches.

Methods: 16S rRNA amplicon sequencing and data analysis

The composition of microbial communities of ants was characterized using amplicon sequencing of the V4 region of 16S rRNA on Illumina MiSeq platform, conducted at Argonne National Laboratory, Argonne, IL. We submitted for processing the total of 210 DNA samples extracted from gasters of individual army ant workers, as well as 13 negative controls: 10 "extraction blanks" (samples "extracted" at the same time and using the same protocols, plastics and reagents as experimental samples, but with no starting material added to the initial lysis buffer), and three samples of molecular-grade water.

Libraries were prepared following the Earth Microbiome Project protocols (Caporaso et al. 2012)(EMP website) and sequenced across four Illumina MiSeq 151x151 lanes, which also included other types of samples. Data were analyzed using Mothur v 1.36.1 (Schloss et al. 2009). The analysis details – specific commands used, with some of our comments - are provided below. Briefly, reads were quality-filtered and trimmed and then assembled into contigs. The dataset was then reduced by grouping the >5M sequences into sets of identical sequences. Next, we have conducted the decontamination procedure outlined below. Subsequently, reads were aligned against a set of 16S rRNA references. The first time the analysis was done, we used against SILVA reference database, release 123. But because a small proportion of reads failed to align correctly, we then redid the analysis using as a reference a mafft alignment of representative sequences for all OTUs from the first iteration. In both cases, all sequences that failed to align correctly were removed. We then checked for the alignment for chimaeras using Uchime algorithm, removing all suspected chimaeras. Finally, all sequences were classified using the RDP training set (v. 9), and those classified as of chloroplast, mitochondrial, archaeal and eukaryotic origin, as well as those unclassified, were removed. The resulting dataset was used for clustering at 97% identity level using average neighbor algorithm.

An important and non-standard feature of our analysis pipeline was filtering out suspected contaminant genotypes from the dataset. It has been demonstrated that bacterial DNA present in reagents and the laboratory environment can considerably alter microbial community profiles, particularly in samples with relatively low template concentration (Salter et al. 2014). Artifacts resulting in sequences derived from one sample appearing in another can further affect the results of amplicon sequencing experiments (Kircher et al. 2011). Our preliminary analyses of the current dataset indicated that indeed the OTUs and unique genotypes found in blanks are also found in experimental samples, sometimes making up a large proportion of the total community. We decided that removing these putative contaminants is essential.

We used information about the distribution of unique sequences across experimental libraries, contained in the original count_table file generated by the unique.seqs command. Parameters were chosen with the intent of removing sequences most likely to have originated as contaminants from reagents or the laboratory, while retaining any sample-derived sequences that might have appeared in blanks as a result of processes such as barcode misassignment (Kircher et al. 2011).

#1: We first excluded from the dataset any unique sequence UNLESS its relative abundance in at least one experimental library was at least 10x of its maximum relative abundance in any of the blanks, reasoning that reagent-derived contaminants should make up the highest relative abundance in blank samples with no added biomass, while sample-derived sequences appearing in blanks should in general have a higher relative abundance in their sample of origin.

#2: We then removed from the dataset any rare unique genotypes, i.e. those that did not make up at least 0.001 of at least one experimental library; in addition to reducing analysis complexity, this step helped to avoid the retention of low-abundance contaminant-derived unique sequences in samples.

#3: Libraries retaining less than 25% of the starting number of reads after filtering steps #1 & #2 were deemed to be highly contaminated and were, hence, removed from the dataset. Indeed after contaminant removal (i.e. post steps #1 & #2) these libraries often shared microbial genotypes that were rare or absent in ant-associated samples with lower detectable contamination. Due to the likelihood that these represented further contaminants we felt it most prudent to remove entire libraries.

#4: Finally, we repeated step #2 to remove rare sequences, using maximum relative abundance values for the libraries remaining subject to initial filtering steps #s 1 & 2.

In our dataset, the decontamination procedure resulted in elimination of eight libraries that were deemed heavily contaminated (out of 210). From the remaining experimental libraries, it eliminated 12.7-74.9% (median 24.7%) of the starting number of reads, and 74.8-99.0% (median 91.8%) of the starting number of unique genotypes. These discarded sequences represent error-free "seed" contaminant genotypes (which are abundant in blanks), as well as those genotypes that are rare in the whole dataset (the majority of which derive from more abundant genotypes of any origin, but contain various sequencing errors). We feel that this simple approach is very effective and specific at removing contaminants as long as they differ from abundant host-associated bacteria by at least one base. It also removes a large proportion of sequences with multiple sequencing errors and especially chimaeras, many of which avoid detection by the current chimaera detection algorithms and negatively affect the precision of identification of operational taxonomic unit (Schloss et al. 2011). Thus, our contaminant filtering results in an increased OTU picking accuracy and dramatically reduced analysis times. At the same time, the filtering procedure may erroneously exclude some real microbes that are present in ants at low abundance, or happen to be identical to reagent-derived contaminants. We suspect that this procedure may also cause small shifts to the relative abundance of OTUs. However, we felt that for the current study, the goal of which is characterization of specialized bacteria abundant in many ants, this was less of a concern than the possibility of mistakenly identifying reagent-derived contaminants as real host-associated microbes.

Methods: specific mother commands used in 16S rRNA amplicon data analysis

```
######## Specifying working directory:
set.dir(input=/Users/Piotr/Dropbox/Documents/Antz/00_AA_paper/MiSeq_20160509,
output=/Users/Piotr/Dropbox/Documents/Antz/00_AA_paper/MiSeq_20160509)


### I am extracting reads for army ant libraries from across five MiSeq runs.
### In all cases, data for these MiSeq runs have been processed the same way:
### make.contigs(file=RunA_file.txt, oligos=RunA_mapping.oligos, allfiles, processors=16)
###
screen.seqs(fasta=RunA.fasta,group=RunA.groups,maxlength=255,minlength=251,maxambig=0,maxhomop=
10,processors=16)

#################### Getting all reads together
# Extracting libraries from RunF (blanks as well as four Peruvian samples)
get.groups(group=/Users/Piotr/Dropbox/Documents/Antz/MiSeq_data/RunF/RunF.good.groups,
fasta=/Users/Piotr/Dropbox/Documents/Antz/MiSeq_data/RunF/RunF.good.fasta,  groups=water1-
water2-water3-Blank_P_AB-Blank_P_CDEF-Blank_P_GHI-Blank_P_JK-P_JSC093_W1-P_JSC120_W1-
P_JSC128_W1-P_JSC129_W1)
system(sed -i '' 's/water/BL_water/g' RunF.good.pick.groups)
system(sed -i '' 's/P_JSC/A_JSC/g' RunF.good.pick.groups)
# ....................from RunD (blanks; four Peru samples; original sample set that is being
redone)
get.groups(group=/Users/Piotr/Dropbox/Documents/Antz/MiSeq_data/RunD/RunD.good.groups,
fasta=/Users/Piotr/Dropbox/Documents/Antz/MiSeq_data/RunD/RunD.good.fasta, groups=A_BL20130514-
A_BL20130515-A_BL20131204-A_BL20140110-A_BL20140117-A_BL20140125-O_DSp8_F-L_HP060_W3-
L_HP116_W4-O_Dhe1_W1-O_Asp11_W1-L_CS491_W4-E_6_W5-L_PL039_W7-P_JSC098_W1-P_JSC152_W1-
P_JSC113_W1-P_JSC122_W1)
system(sed -i '' 's/A_BL/BL_Arg_/g' RunD.good.pick.groups)
system(sed -i '' 's/P_JSC/A_JSC/g' RunD.good.pick.groups)
# ....................from RunC
get.groups(group=/Users/Piotr/Dropbox/Documents/Antz/MiSeq_data/RunC/RunC.good.groups,
fasta=/Users/Piotr/Dropbox/Documents/Antz/MiSeq_data/RunC/RunC.good.fasta, groups=E_C3_W2-
O_Dco1_W1-L_PL039_W2-L_PL034_W3-E_CSM2407_W4-E_HP047_W1-A_PL150_W1-A_PL113_W1-L_PL034_W5-
```

```
L_CS491_W2-A_PL157_W1-E_HP047_W2-E_SOD016_W2-E_PL028_W8-L_CS491_W3-E_CSM2482_W5-A_Kaitlin_W1-
E_0537_W3-E_0537_W2-L_SOD010_W2-A_PL158_W1-L_CS386_W5-L_CS386_W3-L_PL039_W1-A_CS197_W1-
A_CS467_W4-A_CSM2375_W4-A_CS272_W1-A_Che1_W3-E_PL028_W5-E_PL028_W3-L_SOD001_W4-A_CSM1150_W1)
# .....................from RunB
get.groups(group=/Users/Piotr/Dropbox/Documents/Antz/MiSeq_data/RunB/RunB.good.groups,
fasta=/Users/Piotr/Dropbox/Documents/Antz/MiSeq_data/RunB/RunB.good.fasta, groups=E_6_W1-
E_6_W2-E_6_W3-E_6_W4-E_C1_W1-E_C1_W2-E_C1_W3-E_C1_W4-E_C1_W5-E_C3_W1-E_C3_W3-E_C3_W4-E_C3_W5-
E_CSM2407_W1-E_CSM2407_W2-E_CSM2407_W3-E_CSM2407_W5-E_CSM2482_W1-E_CSM2482_W2-E_CSM2482_W3-
E_CSM2482_W4-E_PL022_W1-E_PL022_W2-E_PL022_W3-E_PL022_W4-E_PL022_W5-E_PL028_W1-E_PL028_W4-
E_SOD013_W1-E_SOD013_W2-E_SOD013_W3-E_SOD013_W6-E_SOD015_W1-E_SOD015_W2-E_SOD015_W3-
E_SOD015_W4-E_SOD015_W5-E_SOD016_W1-E_SOD016_W3-E_SOD016_W4-E_SOD016_W5)
# ........and finally, from RunA
get.groups(group=/Users/Piotr/Dropbox/Documents/Antz/MiSeq_data/RunA/RunA.good.groups,
fasta=/Users/Piotr/Dropbox/Documents/Antz/MiSeq_data/RunA/RunA.good.fasta, groups=A_SOD014_W1-
A_HP103_W1-A_CvB286_W1-A_HP002_W1-O_Asp4_F-A_CvB94_W1-O_Asp13x1_F-O_Asp1x3_F-A_CvB177_W1-
A_SOD018_W1-O_Asp50_F-O_Dem2_F-O_Dsp6_F-A_SOD004_W1-A_PL042_W1-A_Ec8_W1-O_Q9_F-A_Nei13_W1-
A_PL041_W1-A_HP082_W1-O_Asp17_F-A_HP089_W1-A_PL027_W1-A_HP129_W1-A_RV4_W1-A_SOD007_W1-
A_PL038_W1-A_HP093_W1-A_HP009_W1-A_HP035_W1-A_HP132_W1-A_CvB192_W1-A_HP119_W1-A_HP024_W1-
L_PL037_W7-A_RV2_W1-A_PL025_W1-A_HP128_W1-A_PL036_W1-A_PL023_W1-A_HP127_W1-A_PL043_W1-
A_HP033_W1-A_SOD008_W1-A_HP006_W1-A_PL026_W1-A_HP111_W1-E_HP047_W5-L_SOD010_W3-L_PL040_W3-
A_CS324_W2-E_PL032_W4-L_SOD010_W5-E_0553_W2-E_HP092_W2-L_PL040_W2-E_0537_W5-E_HP047_W3-
A_Ec6_W1-E_PL032_W2-L_PL040_W4-E_PL032_W5-L_CS386_W4-L_HP003_W5-L_SOD001_W1-L_HP094_W1-
A_CSM2603_W1-A_CS470_W3-L_HP094_W2-L_SOD001_W3-E_HP047_W4-E_HP092_W1-L_SOD010_W1-L_PL040_W5-
E_0553_W4-E_HP092_W4-L_HP003_W4-L_HP060_W1-L_PL034_W2-L_PL037_W2-L_CS491_W1-L_SOD001_W5-
L_HP060_W5-L_HP060_W4-L_CS491_W5-L_PL040_W1-L_SOD010_W4-L_PL034_W1-L_PL037_W1-A_CSM2456_W1-
A_DK20_W1-O_Asp7_W1-L_HP094_W5-E_HP092_W3-E_0553_W1-E_PL032_W3-L_HP003_W2-L_CS386_W1-
L_HP094_W4-L_PL037_W8-A_CS281_W1-O_Dfu1_W1-L_PL037_W6-L_SOD001_W2-E_0553_W5-E_HP092_W5-
L_HP116_W2-E_0537_W1-A_SOD012_W1-L_HP116_W1-E_0537_W4-O_Dfi1_W1-L_HP003_W1-E_0553_W3-
L_HP116_W3-A_CSM2541_W3-L_PL034_W7-L_HP116_W5-A_CS201_W1-E_PL032_W1)


### Now, I am combining the files, and doing some basic processing
merge.files(input=RunA.good.pick.fasta-RunB.good.pick.fasta-RunC.good.pick.fasta-
RunD.good.pick.fasta-RunF.good.pick.fasta, output=Armies.fasta)
merge.files(input=RunA.good.pick.groups-RunB.good.pick.groups-RunC.good.pick.groups-
RunD.good.pick.groups-RunF.good.pick.groups, output=Armies.groups)
count.groups(group=Armies.groups)
# Grouping the >5M sequences into sets of identical sequences
unique.seqs(fasta=Armies.fasta)
# Making count tables
count.seqs(group=Armies.groups, name=current, processors=2)
# Summarizing the sequences
summary.seqs(fasta=Armies.unique.fasta, processors=2)
# Counting sequences in individual libraries
count.groups(group=current)


######## Removal of contaminants, rare unique genotypes, and heavily contaminated libraries
### I am using a custom Python script, which takes as input the original *.count_table file, as
well as a list of blanks. The script:
### 1) removes any unique genotype from experimental libraries UNLESS its relative abundance in
at least one experimental library is at least 10x of its maximum relative abundance in any of
the blanks;
### 2) removes any unique genotype that does not make up at least 0.001 of at least one
experimental library;
### 3) removes any library which after the filtering steps #1 & #2 retains less than 25% of the
starting number of contigs;
### 4) repeats step #2, using maximum relative abundance values for the libraries remaining
after step #3.
### The script outputs a *decontaminated.count_table file with the reduced set of contigs and
libraries

system(/Users/Piotr/Scripts/decontaminate/decontaminate3.py Armies.count_table blanks.txt 10
0.001 0.25)


# Getting accession numbers of sequences retained in count_table after contaminant removal,
extracting these sequences from fasta file
list.seqs(count=Armies.decontaminated.count_table)
get.seqs(accnos=current, fasta=Armies.unique.fasta)
count.groups(count=Armies.decontaminated.count_table)


##################### Removal of non-specific products and chimaeras, OTU picking
##### When first analyzing the results, I aligned unique genotypes against Silva database, and
then processed as below.
##### However, I noticed that in some cases nearly identical sequences got aligned differently
–
```

```
##### with gaps in different places. Subsequently, such sequences were classified to different
OTUs.
##### Thus, I have decided to create my own reference database, consisting of the dominant
##### sequence from each of 862 97% OTUs from that run. All these sequences were aligned using
mafft:
##### "/usr/local/bin/mafft"  --globalpair --maxiterate 16 --reorder "refs_unaligned.fasta" >
"refs_aligned.fasta"
##### I then redid the analysis (as below) using the new set of references.


################### Redoing the steps of the pipeline starting at 'align.seqs()'
# Aligning sequences against the custom reference, removing unaligned sequences, and then
deleting gaps.
align.seqs(fasta=Armies.unique.pick.fasta, reference=refs_aligned.fasta, processors=2)
summary.seqs(fasta=Armies.unique.pick.align)
# Summary.seqs() revealed that all sequences aligned correctly - no need to trim or filter
anything
# Chimera screen using uchime, and remove the identified chimeras from fasta and count files
chimera.uchime(fasta=Armies.unique.pick.align, reference=self,
count=Armies.decontaminated.count_table, dereplicate=f, processors=2, minh=1.5)
remove.seqs(accnos=current, fasta=Armies.unique.pick.align,
count=Armies.decontaminated.count_table)
# Classify sequences
classify.seqs(fasta=Armies.unique.pick.pick.align,
count=Armies.decontaminated.pick.count_table,
reference=/Applications/mothur/trainset14_032015.rdp/trainset14_032015.rdp.fasta,
taxonomy=/Applications/mothur/trainset14_032015.rdp/trainset14_032015.rdp.tax,cutoff=80,process
ors=2)
# Remove unclassified and non-bacterial sequences from the dataset
remove.lineage(fasta=Armies.unique.pick.pick.align,
count=Armies.decontaminated.pick.count_table,
taxonomy=Armies.unique.pick.pick.rdp.wang.taxonomy, taxon=Chloroplast-Mitochondria-unknown-
Archaea-Eukaryota)
# Rename final files
system(mv Armies.unique.pick.pick.pick.align Armies_final.fasta)
system(mv Armies.decontaminated.pick.pick.count_table Armies_final.count_table)
system(mv Armies.unique.pick.pick.rdp.wang.pick.taxonomy Armies_final.taxonomy)
count.groups(count=Armies_final.count_table)


#################### OTU picking, creating various summary tables
# Compute distance matrix
dist.seqs(fasta=Armies_final.fasta, processors=2, cutoff=0.20)
# OTU picking!
cluster(column=Armies_final.dist, count=Armies_final.count_table, cutoff=0.20, method=average)
# Classifying sequences
bin.seqs(list=current, fasta=Armies_final.fasta, label=0.03)
# Constructing OTU table
make.shared(list=current,count=Armies_final.count_table,label=0.03)
system(mv Armies_final.an.unique_list.shared Armies_final.an.unique_list.003.shared)
system(rm *.rabund)
# Assigning consensus taxonomic IDs for OTUs
classify.otu(list=current,count=Armies_final.count_table,taxonomy=Armies_final.taxonomy,
label=0.03)




#################################################
#################################################
### As a control, I decided to do OTU picking on all sequences from all libraries (experimental
and blanks) - the conventional way,
### so without any sort of contaminant removal.
### Starting files:
### Armies.count_table
### Armies.unique.fasta

# set environment
set.dir(input=/scratch/1/piotr/Armies, output=/scratch/1/piotr/Armies)
# remove singleton reads
split.abund(fasta=Armies.unique.fasta, count=Armies.count_table, cutoff=1)
# align sequences against a Silva reference, remove unaligned etc. sequences, and then delete
gaps
align.seqs(fasta=Armies.unique.abund.fasta, reference=silva.bacteria.fasta, processors=16)
screen.seqs(fasta=Armies.unique.abund.align, count=Armies.abund.count_table, start=13862,
end=23444, minlength=251)
filter.seqs(fasta=Armies.unique.abund.good.align, vertical=T, trump=.)
```

```
# Chimera screen using uchime, and remove the identified chimeras from fasta and count files
chimera.uchime(fasta=Armies.unique.abund.good.filter.fasta, reference=self,
count=Armies.abund.good.count_table, dereplicate=f, processors=16, minh=1.5)
remove.seqs(accnos=current, fasta=Armies.unique.abund.good.filter.fasta,
count=Armies.abund.good.count_table)
# Classify sequences
classify.seqs(fasta=Armies.unique.abund.good.filter.pick.fasta,
count=Armies.abund.good.pick.count_table, reference=trainset14_032015.rdp.fasta,
taxonomy=trainset14_032015.rdp.tax,cutoff=80,processors=2)
# Remove unclassified and non-bacterial sequences from the dataset
remove.lineage(fasta=Armies.unique.abund.good.filter.pick.fasta,
count=Armies.abund.good.pick.count_table,
taxonomy=Armies.unique.abund.good.filter.pick.rdp.wang.taxonomy, taxon=Chloroplast-
Mitochondria-unknown-Archaea-Eukaryota)
# Rename final files
system(mv Armies.unique.abund.good.filter.pick.pick.fasta Armies_final.fasta)
system(mv Armies.abund.good.pick.pick.count_table Armies_final.count_table)
system(mv Armies.unique.abund.good.filter.pick.rdp.wang.pick.taxonomy Armies_final.taxonomy)
count.groups(count=Armies_final.count_table)

# Compute distance matrix
dist.seqs(fasta=Armies_final.fasta, processors=64, cutoff=0.20)

# OTU picking
cluster.split(column=Armies_final.dist, count=Armies_final.count_table, cutoff=0.10,
method=average, processors=64)

# Constructing OTU table
make.shared(list=current,count=Armies_final.count_table,label=0.03)
# Classifying OTUs
classify.otu(list=current,count=Armies_final.count_table,taxonomy=Armies_final.taxonomy,
label=0.03)
# Selecting representative sequences for OTUs
get.oturep(column=Armies_final.dist, list=Armies_final.an.unique_list.list,
fasta=Armies_final.fasta, count=Armies_final.count_table)
# Classifying the sequences from the original Armies_final.fasta file
bin.seqs(list=Armies_final.an.unique_list.list, fasta=Armies_final.fasta, label=0.03)
```

Methods: Analysis of *rplB* amplicon sequencing data

We conducted 454 pyrosequencing of the *rplB* amplicon for Unclassified Firmicutes for four *Eciton burchellii parvispinum* workers from colony PL028, each of which hosted a distinct dominant symbiont genotype (as determined by direct Sanger sequencing of the amplicon). Product for pyrosequencing was generated using the same *rplB* primers as for Sanger sequencing, complete with linkers and barcodes, and sequenced in a multiplexed lane by Research and Testing Laboratory (Lubbock, TX). A total of 12922 reads were obtained. These data were analysed using mothur v. 1.33 (Schloss *et al.* 2009). Reads were extracted from sff files, quality-filtered and trimmed to the length of 360bp. Then, they were then aligned against the reference alignment of Sanger sequences of the dominant strains in these four samples, which had gaps introduced every two nucleotides in order to facilitate aligning reads with insertions. "Gotoh" was selected as the alignment method, and the reward for a match was set to +2. The alignment, manually inspected and filtered by removing all nucleotide positions without sequence data, was then used for calculation of an uncorrected pairwise distance matrix (with gaps ignored), and then for clustering using "average neighbour" method at 98% sequence similarity, the level at which reference Sanger sequences of the two most similar strains could be distinguished. We aligned 95% consensus sequences for all OTUs, and alignments of reads classified to dominant OTUs were manually inspected at all positions with less than 99% consensus.

Methods: Fluorescence *in situ* hybridization – visualization of bacteria in digestive tracts of ants

A. Specimen fixing

In case of many insects, specimens directly preserved in ethanol or in acetone can be used for FISH. This does not generally work for ant guts – tissues which had not been fixed prior to preservation lose structure, become brittle, and are generally unsuitable for FISH.

Prior to fixation, you may want to carefully dissect the guts in PBS or fixative. However, it is sufficient (and much more convenient under field conditions) to do a partial dissection - open up the gaster by pulling sclerites apart. This should be done in PBS, or the gaster should be dipped in PBS as soon as possible. Make sure that no large gas bubbles block access of the fixative to the tissue. Such partially dissected guts can be fixed as effectively and following the same protocol as guts that have been fully dissected - and dissection can be completed later, prior to hybridization.

Also, if you are going to do hybridization soon after fixation, there is no need to replace water with ethanol in the sample. In such case, skip steps 5 and 6 below, as well as the rehydration step in the hybridization section. If interval between fixation and hybridization is longer than an hour, keep samples refrigerated but DO NOT FREEZE them, as forming ice crystals may damage tissue.

1. Dissect out the gut, completely or partially (as above), in **PBS (-)**, preferably on ice.
2. Transfer the gut to a tube containing 10 volume or more of ice-cold **PBS-4% formaldehyde** (prepare by mixing 9 volumes of PBS and 1 volume of commercially available 40% formaldehyde solution) and fix the tissues at 4°C overnight, or at room temperature for two hours.
3. After fixation, wash the tissue with 10 or more volumes of PBS at room temperature for 10 minutes, preferably with gentle agitation.
4. Repeat step #3 two or preferably three times - wash the tissue with PBS 3-4 times in total
5. Wash the tissue with 75% ethanol at room temperature for 10 min, repeat for the total of four washes
6. Then, transfer tissues into absolute ethanol and keep at -20 to -80°C, until needed.

B. Hydrogen peroxide treatment

Tissues of many insects display strong autofluorescence in several channels, and particularly at shorter wavelengths. This can interfere with detection of specific probe signal. This was a major problem for ant guts. One way of reducing autofluorescence of insect tissues relative to the specific signal is hydrogen peroxide treatment, originally proposed by Koga et al. 2009 (*Appl. Entomol. Zoolog.*, 44: 281-291). This protocol has a more positive effect in some organisms than in others; also, the optimal treatment length depends on specimen size and other characteristics. Seven-day treatment did reduce autofluorescence in tissues of ants, but it remained high.

1. Place the fixed ethanol-preserved specimen in **6% hydrogen peroxide solution in 80% ethanol**. Prepare the solution it by mixing 1 volume of commercially available 30% aqueous H_2O_2 with 4 volumes of absolute ethanol.
2. Bleach the specimen at room temperature until the tissues (including any fragments of cuticle) become whitish. It usually takes several days. For the first few days exchange the peroxide solution daily, later every 2-3 days.
3. Wash specimen with absolute ethanol at room temperature for 10 minutes, three times
4. Store specimen in absolute ethanol at -20°C until needed.

C1. Whole-mount FISH

Whole-mount FISH is a quick and easy way of localizing bacteria within complete dissected guts. After hybridization, specimens can be imaged using epifluorescence microscope or confocal microscope.

1. Rehydrate fixed ethanol-preserved specimen: wash three times with **PBSTx** (PBS buffer with 3 ml of Triton X-100 added to each liter), 10 minutes each time
2. If the specimen (ant gut) has not been fully dissected before, do it now. Remove all fragments of cuticle and as much of fat body as possible. Ideally, you will end up with intact digestive tract – and nothing else.
3. Wash specimen three times with **hybridization buffer** (see recipe at the end of this document)
4. Add **hybridization solution** (consisting of the buffer + DAPI + probes – see the recipe at the end of this document), keep specimen overnight at room temperature
5. Wash specimen three times with **PBSTx**, 10 minutes each time
6. Suspend specimen in a medium containing antifade reagent (for example **DABCO-glycerol**: 90ml glycerol, 10ml PBS, 1.25g DABCO (1,4-Diazabicyclo[2.2.2]octane)) on a microscope slide, cover with cover glass, seal with nail polish or similar
7. Image!

C2. FISH of resin sections - embedding

This is a powerful technique for visualizing bacteria within tissues at high magnification, in sections as thin as 1 μm. Embedding according to our protocol requires **Technovit 8100 Embedding Kit** and some related products: **Technovit 3040 Mounting Medium**, **HistoForm S** and **HistoBlocks** (https://www.emsdiasum.com/microscopy/products/embedding/technovit.aspx). Alternatively, **Technovit 7100 Embedding Kit** can be used, and the protocol modified according to manufacturer's instructions. Also, access to a quality microtome is needed, as well as coated microscope slides (hydrophilic coating preferred).

~~~EMBEDDING~~~

1. If your ethanol-preserved specimen is fully dissected, go directly to step #2. If the specimen (ant gut) has not been fully dissected, do it now:
   - Rehydrate tissue: wash three times with **PBSTx**, 10 minutes each time
   - Dissect: Remove all fragments of cuticle and as much of fat body as possible. Ideally, you will end up with intact digestive tract – and nothing else.
   - Hold your dissected gut straightened up on a glass Petri dish and add acetone. Tissue will harden rapidly in the position it was held at
2. Wash specimen three times with acetone, 10 minutes each time. The last wash can be extended to an hour or more.
3. Replace acetone with **resin monomer solution** (prepared by mixing 100ml of GMA monomer with 0.6g of Hardener I from Technovit 8100 kit; <u>keep the solution on ice or in the fridge</u> for up to a month. Do three washes with resin monomer solution, at 4°C, for at least 10 minutes each time. After the last replacement, keep the specimen in monomer solution overnight to ensure optimal tissue penetration with the monomer.
4. Cool down the embedding mold (**HistoForm S**) by placing it in the fridge for 1 hour

5. Add Hardener II from Technovit 8100 Kit to monomer solution (1 to 30 ratio; plan to use 2ml of solution with hardener per specimen) and mix well; the mixture turns blue. In the tube with specimen, replace monomer solution with 1ml of solution + Hardener II. Wash at 4°C for 5 minutes, with gentle agitation.

6. <u>On ice</u>, pour approx.1 ml of monomer solution + Hardener II to each well in the mold. Place the specimen in a mold, arrange it, then cover up the mold with PE film to reduce oxygen access. Leave mold at 4°C overnight.

7. Remove film and any non-polymerized resin. Then, attach embedded specimen to **HistoBlocks** using **Technovit 3040 Mounting Medium**: pour monomer powder over the resin block and into Histoblock, then slowly add liquid Hardener until the mixture reaches muddy consistence. Spread it well over embedded specimen and leave for 30 minutes.

8. Remove embedded specimen from the mold. You will need a tool such as groove joint plier for that. Store resin blocks with embedded specimens at -20°C.

9. Section the embedded specimen. First, remove excess resin around the specimen, using power drill and scalpel, or other tools. Thicker sections (3-4 μm) are easier to make, thin sections (1 μm) may offer superior images, but 2 μm sections offer a good balance between ease of sectioning and image quality. *Hint: gently blowing humid air on the microtome blade when sectioning helps to prevent sections from rolling*

10. Using a tool such as paintbrush or forceps, gently transfer sections onto droplets of water on coated microscope slides (we used slides with hydrophilic MAS coating made by Matsunami Glass Ind. Ltd, Japan, but the availability of this specific product might be limited outside Japan). Drain excess water, place slide on a plate heated to 37°C to speed up vaporation of remaining water, wait until slides are completely dry, store them at 4°C.

~~~HYBRIDIZATION~~~

11. Add 100 μl **hybridization solution** (buffer + DAPI + probes) for 24 x 24 mm^2 area, cover with cover glass, hybridize in humidified chamber (closed box filled with wet tissue) overnight at room temperature

12. Wash specimen three times with excess of **PBS**, 10 minutes each time

13. Drain residual buffer but do not dry the specimen to prevent bubble formation. Add antifade reagent such as **ProLong Gold Antifade Reagent** by Life Technologies, cover with a cover slip

14. Image!

D. Buffers
1. **Hybridization buffer** – mix:
   - 1M Tris-HCl (pH 8.0) – 2.0 ml   final concentration: 0.02M
   - 5M NaCl – 18.0 ml   final concentration: 0.9M
   - 20% SDS – 50 μl   final concentration: 0.01%
   - formamide – 30 ml   final concentration: 30%
   - distilled water – up to 100 ml
   -

   If possible, remove dusts and large suspended particles by filtering through a 0.45 μm filter

2. **Hybridization solution** – mix:
    - Hybridization buffer, as above – 1000 µl
    - Probes, each at 10 µM concentration – 10 µl (final concentration: approx. 100 nM)
    - DAPI, 1% solution – 1 µl

For whole-mount FISH, prepare approx. 300 µl of solution per sample. For sections, prepare 100 µl of solution per slide.