Journal of Microbiological Methods 151 (2018) 16–19

Contents lists available at ScienceDirect

Journal of Microbiological Methods

journal homepage: www.elsevier.com/locate/jmicmeth

Enrichment of low-density symbiont DNA from minute insects

Corinne M. Stouthamer[a,b,*], Suzanne Kelly[b], Martha S. Hunter[b]

[a] Graduate Interdisciplinary Program in Entomology and Insect Science, PO Box 210036, The University of Arizona, Tucson, AZ 85721, USA
[b] Department of Entomology, 410 Forbes Building, The University of Arizona, Tucson, AZ 85721, USA

ARTICLE INFO

Keywords:
Cardinium
Symbiont
Microbial DNA enrichment protocol
Parasitic wasps
Parasitoid
Low titer infections
Wolbachia

ABSTRACT

Symbioses between bacteria and insects are often associated with changes in important biological traits that can significantly affect host fitness. To a large extent, studies of these interactions have been based on physiological changes or induced phenotypes in the host, and the genetic mechanisms by which symbionts interact with their hosts have only recently become better understood. Learning about symbionts has been challenging in part due to difficulties such as obtaining enough high quality genomic material for high throughput sequencing technology, especially for symbionts present in low titers, and in small or difficult to rear non-model hosts. Here we introduce a new method that substantially increases the yield of bacterial DNA in minute arthropod hosts, and requires less starting material relative to previous published methods.

1. Introduction

Intracellular, maternally inherited symbionts can influence their arthropod hosts' biology in various important ways. Obligate symbionts usually provide key nutrients missing in their hosts' diet (Moran et al., 2008), while facultative symbionts may provide conditional benefits such as defense against parasitoids (e.g Oliver et al., 2003; Xie et al., 2014), pathogens (e.g. Scarborough et al., 2005; Łukasik et al., 2013), or heat shock protection (Montllor et al., 2002; Russell and Moran, 2006). Facultative symbionts might also manipulate host reproduction by increasing the daughter production or fitness of infected females relative to their uninfected counterparts, as the symbionts are only passed on from mother to daughter, usually through the cytoplasm of the egg (Moran et al., 2008). Several different lineages of bacteria have evolved the ability to manipulate the daughter production of their hosts, including those in the genera *Wolbachia*, *Spiroplasma*, *Rickettsia*, and *Cardinium* (reviewed in Engelstädter and Hurst, 2009). Intracellular symbionts are found in the hemolymph and other insect tissues, and most are uncultivable outside of their hosts (Moran et al., 2008). Although there is considerable interest in how these bacteria interact with their hosts, classical microbiology techniques are limited for the study of these symbionts because they cannot generally be grown on plates. In this context, understanding the genomic capabilities of the symbiont can provide particular insight, but extraction of sufficient high quality bacterial DNA for sequencing may pose considerable technical challenges.

There are several obstacles to overcome when using whole hosts to sequence the genomes of their intracellular symbionts. First, there is often a large amount of contaminating host DNA, as eukaryotic genomes are much larger than bacterial genomes. Second is the issue of obtaining enough bacterial DNA for sequencing, particularly from hosts that are very small or hard to rear, because the absolute amount of bacterial DNA per host is at least somewhat proportional to host body size. Improved bioinformatics and less expensive short read technology now make it possible to obtain high quality draft genomes from samples of mixed host and symbiont DNA (e.g. Koutsovoulos et al., 2016; Brown et al., 2016), but long read, low throughput technologies, such as PacBio sequencing, still require a relatively large amount of high molecular weight DNA and are more efficient with greater concentrations of symbiont DNA. In fact, genome coverage for both low and high throughput technologies is improved when a higher proportion of the sample is the target DNA, in this case the symbiont DNA.

Sequencing technologies such as targeted capture methods can be used to enrich symbiont sequence, but reference genomes must be available to design the probes for symbiont DNA capture (Geniez et al., 2012; Jones and Good, 2016). Targeted capture is used quite effectively, for example, when only a portion of the genome is of interest (Sims et al., 2014), or for population genomics applications, when multiple individuals or closely related species will be sequenced (Christmas et al., 2017). The cost of development of the probe set can be prohibitive for one or a few genomes, and the need to have either a reference genome, or other fairly extensive genomic resources makes it less applicable than DNA enrichment and whole genome sequencing for exploration of genome function.

This study introduces a protocol designed to enrich symbiont DNA in insect samples, particularly for hosts that are small and/or harbor

* Correspondence author at: Graduate Interdisciplinary Program in Entomology and Insect Science, PO Box 210036, The University of Arizona, Tucson, AZ 85721, USA.
E-mail address: cmstouthamer@email.arizona.edu (C.M. Stouthamer).

https://doi.org/10.1016/j.mimet.2018.05.013
Received 23 February 2018; Received in revised form 16 May 2018; Accepted 18 May 2018
Available online 21 May 2018

symbionts at relatively low densities. We tested our protocol in minute (~1 mm) *Encarsia* wasps, which harbor the facultative symbiont *Cardinium*.

Cardinium hertigii is a maternally inherited, intracellular symbiont of nematodes and arthropods, estimated to infect 9% of arthropods (Russell et al., 2012). Much like the very distantly related *Wolbachia*, *Cardinium* can manipulate its host's reproduction in several ways, including parthenogenesis (e.g. Zchori-Fein et al., 2004; Provencher et al., 2005), feminization (e.g. Weeks et al., 2001), and cytoplasmic incompatibility (e.g. Hunter et al., 2003). *Cardinium* infects several minute arthropods (≤1 mm long), including many species of mites, *Culicoides* biting midges, thrips, and parasitic *Encarsia* wasps (Lewis et al., 2014; Nguyen et al., 2017; Zchori-Fein and Perlman, 2004). *Encarsia* sp. are small (~1 mm, ~18 μg) and whole wasps harbor *Cardinium* at a low density; *Cardinium* genomes (~1 MB) are at roughly an equal ratio (1:1) with host genomes (200-400 MB) (Perlman et al., 2014). This makes 1) separating bacterial and host reads difficult, because genome coverage does not substantially differ between the two organisms, and 2) long read, low throughput technology impractical because the host genomic reads greatly outnumber those of the symbiont. Because 1000 adult *Encarsia* wasps are, in weight, equal to roughly 12 adult female *Drosophila melanogaster* (Katz and Young, 1975, Mann et al., 2017), most laboratories do not have the capabilities to raise enough wasps to follow previously published extraction protocols that start with 2000–5000 adult *Drosophila* (Iturbe-Ormaetxe et al., 2011), equivalent to approximately 167,000–417,000 *Encarsia* wasps.

The following protocol is based on the Penz et al. (2012) extraction protocol, which itself was modified from Braig et al. (1998). It starts with roughly the same inputs as the Penz et al. (2012) protocol, is no more labor intensive, and produces a higher yield of symbiont-enriched DNA of a quality and length appropriate for long and short read libraries. Although this protocol was developed for minute *Encarsia* wasps and their *Cardinium* endosymbionts, we anticipate that it can be used for hosts of any size, but will be particularly useful for other minute hosts with low-density symbionts.

2. Methods

2.1. DNA extraction method

Approximately 1000 wasps of each of *Encarsia hispida*, *Encarsia inaron* (Italy), and *Encarsia tabacivora* were used as the starting material. The wasps were homogenized using a tight fitting (0.025–0.076 mm) 1 ml Wheaton Dounce tissue grinder (catalog #357538) in 800 μl of Buffer A (35 mM Tris HCl, 250 mM sucrose, 250 mM EDTA, 25 mM KCl, 10 mM $MgCl_2$). The homogenate was then transferred to a 1.5 ml Eppendorf tube and the Dounce receptacle was rinsed with 400 μl of filtered Buffer A, which was also added to the Eppendorf tube. The 1.5 ml tube with homogenate was incubated for one hour at 4 °C, inverted every 10 min, and was then centrifuged at 600 x g at 4 °C for 10 min. Next, the supernatant was loaded into a sterile 5 ml Luer-Lok syringe (BD, #309646) attached to a 13 mm diameter filter cassette holder (Swinnex filter holder, Millipore,# × 0001300) with a 0.8 to 8 μm pore size glass fiber prefilter (Millipore, #AP2001300) on top of a strong protein-binding, mixed cellulose ester membrane (Millipore, #SMWP01300) with a 5 μm pore-size, and pushed slowly through. These steps remove most of the larger cellular fractions of the eukaryotic cells, while allowing the bacterial cells to pass through the filter. The resulting filtrate was then centrifuged to pellet the bacterial cells at 14,100 x g for 15 min at 4 °C. The supernatant was removed and the pellet was re-suspended in 150 μl lysis buffer (0.5% (w/v) SDS, (200 mM Tris, 25 mM EDTA, 250 mM NaCl, and 1.3 mg/ml Rnase A) and incubated for 30 min shaking at 250 RPM at 37 °C. One hundred fifty μl of buffered phenol and 150 μl of chloroform were added to the lysate, and the tube was inverted by hand for 10 min. Subsequently, the sample was centrifuged at 12,000 x g at room temperature for 10 min.

Table 1
Extractions of *Cardinium*-infected *E. inaron* with varying incubation times.[a]

Sample treatment	*Cardinium*: host cell ratio	Average fragment size	Total ng of DNA in sample
No incubation	34.64: 1	8595 bp	1753
1 h incubation	70.22: 1	8733 bp	1154.5
1.75 h incubation	110.81: 1	5958 bp	707.5

[a] The *Cardinium*: host ratios were derived with qPCR and calculated using the delta-delta Ct method. The average fragment size was estimated with the BioAnalyzer 2100. Qubit 3.0 was used to quantify the DNA.

The organic (lower) phase was then removed and 150 μl of chloroform and 100 μl of nuclease free water were added. The contents were then inverted by hand for 5 min and the sample was centrifuged at 12,000 x g at room temperature for 10 min. The supernatant was placed in a new tube, and, to precipitate DNA, 45 μl 5 M NaCl was added, then 1000 μl of 100% EtOH at room temperature. The mixture was left in the freezer overnight, then centrifuged at 12,000 x g for 15 min. The supernatant was decanted and the pellet was washed twice with 500 μl of 70% ethanol, dried, and suspended in TE buffer with 0.5 M EDTA.

2.2. DNA extraction with variation in incubation times

Earlier extractions with a version of this protocol omitted the initial 4 °C incubation step of the raw wasp homogenate in Buffer A. To test whether this step affects the symbiont to host ratio, DNA fragment size, or DNA yield, a separate extraction was performed where one aliquot of ~8000 homogenized wasps was split into three groups and extracted with varying initial incubation times at 4 °C in Buffer A: no incubation, 1 h incubation, 1.75 h incubation. Fragment analysis (Bioanalyzer 2100 with High Sensitivity DNA Kit (Agilent)), qPCR (Bio-Rad CFX Connect Real Time System. Maxima SYBR Green qPCR mix (ThermoFisher Scientific)), and DNA quantification (Qubit 3.0) were performed on the resulting DNA.

2.3. Relative quantification of cardinium: wasp genome copies with qPCR

The ratio of *Cardinium* to host genome copies was measured using qPCR and the delta-delta Ct method (Schmittgen and Livak, 2008). Quantitative PCR was performed using Maxima SYBR Green/ROX qPCR Master Mix (2×) (ThermoFisher Scientific) with primers targeting the single-copy EF-1alpha gene (EF_F: AGATGCACCACGAAGCC and EF_R: CCTTGGGTGGGTTGTTCTT) for all of the wasp species, and primers targeting the *Cardinium* gyrase B gene (gyrb737F: AAGTTATTGTAGC CGCTCAAG and gyrb911R: GCAGTACCACCAGCAGAG) (Perlman et al., 2014) for the *Cardinium* strains.

2.4. Short and long read sequencing sample extraction and analysis

Three sequencing technologies were chosen for this study: short reads were generated by either the Ilumina HiSeq platform (paired-end, 2 × 150 bp, insert size 500–1000 bp) or by the Ilumina MiSeq platform (paired-end, 2 × 300 bp, insert size 500–600 bp), and long reads were generated by Pacific Biosciences SMRT sequencing (hereafter "PacBio"). The non-enriched HiSeq samples were extracted from whole wasps using a DNeasy extraction kit (Qiagen). The samples destined for MiSeq sequencing were extracted using the enrichment protocol of the current study with a one-hour incubation step at 4 °C in Buffer A. The samples for PacBio sequencing were extracted using the protocol of the current study without the incubation step at 4 °C in Buffer A. For the MiSeq and HiSeq reads, the percentages of *Cardinium* reads were determined by mapping reads to reference *Cardinium* genomes (Stouthamer et al., unpublished) in Bowtie2 (version 2.3.2 default settings, paired reads) (Langmead and Salzberg, 2013). To determine the number of *Cardinium* reads in the PacBio samples, a custom BLAST