**Support information**

The *C. hookeri* transcriptome was constructed from five different tissues derived from nine adults collected from Totara Park, Auckland, New Zealand (37º0.111 S, 174º55.039 E). The assembly was produced by combining three *de novo* assemblies: the pooled antennae, midgut and terminalia assembly, and the two previously assemblied leg, and head and prothorax assemblies. The head and prothorax assembly is described elsewhere (Dennis et al. 2015).

**Antennae, midgut and terminalia assembly**

Illumina sequencing reads were pooled together and quality assessed with FastQ (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). To reduce the effects from inherent sequencing biases resulting from library preparation, raw reads were trimmed by removing eight nucleotides (nt) from the 5’ end using PRINSEQ (v.0.20.3). To maximise quality, reads were then processed to remove ambiguities, trim adapters and low quality regions (Phred < 30) using cutadapt (v1.2.1). Remaining reads that were less than 50 bp in length, and those no longer paired were removed using PRINSEQ. To optimise the completeness of the transcriptome and the full-length transcripts generated from the pooled three libraries, the de novo assembly was obtained by combining assembled contigs derived from Trinity (v20140413) [27], SOAPdenovo-trans (v1.03), and Velvet/Oases (v0.2.08). The Kmer sizes 21, 25, 29, and sizes from 35 to 85 with stepwise increase of ten were used in SOAPdenovo-trans and Velvet/Oases, and scaffold gaps were filled by GapCloser (http://sourceforge.net/projects/soapdenovo2/files/GapCloser/). Each final transcript was given an ID consisting of species, assembly version, assembler, Kmer size if adjusted multiple times, and an unique seven digit ID, eg. Chv1VELVK21 0000001.

**Leg assembly**

Contigs were assembled using SOAPdenovo-trans (v1.01) with a Kmer of 31. The gapped sequences were filled gaps using GapCloser. Raw data were preprocessed by removing reads of poor quality including: 1) reads with adapter; 2) reads with a total number of >10 Ns; 3) reads with >50 base pairs of low quality (Phred quality score = 2, ASCII 66 ”B”, Illumina 1.5+ Phred+64). Each contig ID was replaced by the code ChvKITESOAP with a 7 digit ID, which vKITE stands for the assembly version produced for the project of ‘evolution in insects’ that has been supervised by 1K Insect Transcriptome Evolution (<http://www.1kite.org/>).

**Final assembly**

The final transcriptome of coding transcripts was created by combining a total of 19 assemblies described above using the EvidentialGene tr2aacds pipeline (http://arthropods.eugenes.org/EvidentialGene/). This pipeline employs CD-HIT-EST (v3.1.1) to cluster transcripts with identity of 90% or greater and then select the ‘best’ contig based on the open reading frame (ORF) and UTRs present from each cluster as the final coding transcript. The resulting transcript set was then quality evaluated with Core Eukaryotic Genes Mapping Approach (CEGMA: v2.4) software, which detects a core protein set of 248 highly conserved proteins that are found in a wide range of eukaryotes. The cleaned reads were also mapped back to both the raw and final transcriptomes using Bowtie2 (v2.1.0) to reproduce a coverage profile of the assemblies from short reads.

Please see our paper for references and more information