Biodiversity and Conservation

Differentiating the effects of organic management, pesticide reduction, and landscape diversification for arthropod conservation in viticulture

Marvin Kaczmarek*, Martin H. Entling, Christoph Hoffmann

Correspondence: Marvin Kaczmarek, Julius Kühn Institute, Federal Research Institute for Cultivated Plants, Institute for Plant Protection in Viticulture, Siebeldingen, Germany, Email: marvin.kaczmarek@julius-kuehn.de

Supplementary Methods

DNA metabarcoding and bioinformatics were conducted by AIM - Advanced Identification Methods GmbH. Following the protocol published in Hausmann et al. (2020), species identification of insect material of the Malaise trap samples was performed using DNA metabarcoding. Each sample was dried in a 60°C oven for at least eight hours and subsequently homogenized in a FastPrep96 machine (MP Biomedicals) using sterile steal beads to generate a homogeneous mixture of animal material. A subsample of each homogenate was transferred into sample vials prior to DNA extraction and processed using the DNeasy 96 Blood & Tissue Kit (Qiagen) following the manufacturer's instructions. A 2-step PCR was performed for amplification of the CO1-5P target region and preparation of the MiSeq libraries. First, using forward and reverse HTS primers, a 313 base pairs (bp) long mini-barcode region was amplified by PCR, equipped with complementary sites for the Illumina sequencing tails. Index primers with unique i5 and i7 inline tags and sequencing tails were used for amplification of indexed amplicons in a subsequent PCR reaction. Afterwards, equimolar amplicon pools of 100 ng/µL each were created and size checked using preparative gel electrophoresis. Using MagSi-NGSprep Plus beads (Steinbrenner Laborsysteme GmbH, Wiesenbach, Germany), the pooled DNA was purified. For a final check of the bp distribution and concentration of the amplicons, a bioanalyzer (High Sensitivity DNA Kit, Agilent Technologies) was used before the creation of the final library. Then, high-throughput sequencing was performed on an Illumina MiSeq using v2 (2*250 bp, 500 cycles, maximum of 20mio reads) chemistry (Illumina) aiming for 250k raw reads for each sample.

The bioinformatics processing of raw FASTQ files from Illumina was carried out with the VSEARCH suite 2.9.1 (Rognes et al. 2016) and Cutadapt 1.18 (Martin 2011). In each sample, forward and reverse reads were merged using the VSEARCH program fastq_mergepairs with a minimum overlap of 10 bp, yielding approximately 313 bp long sequences. Forward and reverse primers were removed using Cutadapt with the discard_untrimmed option to discard sequences for which primers were not reliably detected at ≥90% identity. Then, using the fastq_filter in VSEARCH, quality filtering was done, keeping sequences with zero expected errors (fastq_maxee 1). Sequences were dereplicated using derep_fulllength, first at the sample level and then concatenated into one FASTA file, which was subsequently dereplicated. Using the VSEARCH program uchime_denovo, chimeric sequences were filtered out from the FASTA file. The remaining sequences were clustered into OTUs at 97% identity with cluster_size, a greedy centroid-based clustering program. Afterwards, OTUs were blasted against a custom Animalia database which was downloaded from BOLD in Q4 2021, including taxonomy and BIN information, by means of Geneious 10.2.5 (Biomatters, Auckland, New Zealand), and following methods described in Morinière et al. (2016). The resulting CSV file, including the OTU ID, BOLD Process ID, BIN, Hit-%-ID value (percentage of overlap similarity (identical bp) of an OTU query sequence with its closest counterpart in the database), length of the top BLAST hit sequence and phylum, class, order, family, genus, and species information for each detected OTU, was exported from Geneious and combined with the OTU table generated by the bioinformatics pipeline.

Supplementary Results

The initial analysis showed that reads had high sequence quality scores. The number of paired-end sequences that came from the sequencer was 3,714,061. Across all samples, a median of 95.0% of reads merged (mean 94.5%, minimum 53.9%, maximum 99.1%). Adapters were detected in a median of 99.6% of the forward and 99.0% of reverse reverse-complement reads. After quality filtering, 3,070,422 sequences were kept. There were 2,035,286 unique sequences in all samples, of which 146,456 unique non-singleton sequences were kept after dereplication. Of a total of 10,915 OTUs, 4,325 OTUs were kept after de novo chimera detection, 3,245 of which found matches in the databases following OTU table cleaning. 2,150 OTUs had a BOLD Hit-%-ID equal to or higher than 97%, 1,809 of which were assigned to BINs. 61 BINs occurred doubled and were condensed to one entry.