Adaptation during the transition from *Ophiocordyceps* entomopathogen to insect associate is accompanied by gene loss and intensified selection.

Chris M. Ward[1], Cristobal A. Onetto[1,2], Anthony R. Borneman[1,2]

[1] Australian Wine Research Institute, Australia

[2] School of Agriculture, Food and Wine, University of Adelaide, Australia

Abstract:

Fungal and bacterial symbiosis is an important adaptation that has occurred within many insect species, which usually results in the relaxation of selection across the symbiont genome. However, the evolutionary pressures and genomic consequences associated with this transition are not well understood. Pathogenic fungi of the genus *Ophiocordyceps* have undergone multiple, independent transitions from pathogen to associate, infecting soft-scale insects trans-generationally without killing them. To gain an understanding of the genomic adaptations underlying this transition, long-read sequencing was utilized to assemble the genomes of both *Parthenolecanium corni* and its *Ophiocordyceps* associate from a single insect. A highly contiguous haploid assembly was obtained for *Part. corni*, representing the first assembly from a single Coccoidea insect, in which 97% of its 227.8 Mb genome was contained within 24 contigs. Metagenomic-based binning produced a chromosome-level genome for *Part. corni*'s *Ophiocordyceps* associate. The associate genome contained 524 gene loss events compared to free-living pathogenic *Ophiocordyceps* relatives, with predicted roles in hyphal growth, cell wall integrity, metabolism, gene regulation and toxin production. Contrasting patterns of selection were observed between the nuclear and mitochondrial genomes specific to the associate lineage. Intensified selection was most frequently observed across nuclear orthologs, while selection on mitochondrial genes was found to be relaxed. Furthermore, scans for diversifying selection identified associate specific selection within three adjacent enzymes catalyzing acetoacetate's metabolism to acetyl-COA. This work provides insight into the adaptive landscape during the transition to an associate life history, along with a base for future research into the genomic mechanisms underpinning the evolution of *Ophiocordyceps*.

Introduction:

Microbes interact with higher organisms through diverse and specific mechanisms (Iturbe-Ormaetxe, et al. 2011; Fan, et al. 2015; Flórez, et al. 2017; Kobmoo, et al. 2018; Trivedi, et al. 2020) adapting through both antagonistic and mutualistic strategies. Mutualism between microbes and associated taxa exists whereby fitness benefits are conferred to both organisms by mutual and generally symbiotic, relationships. In insects, an association with specific prokaryotic and fungal species has been shown to provide mutual benefit to the host through increasing nutrient availability (Noda and Koizumi 2003; Vogel and Moran 2011) or detoxification of noxious compounds (Welte, et al. 2016). As the mutualistic relationships can insulate the associate from direct environmental influences, which can drastically change selection pressures that are applied to the symbiont genome and allow for previously constrained genes to escape selective pressures (Wernegreen 2011; Wertheim, et al. 2015). This is most dramatically observed when symbionts live within the body of the insect host or even within the insect cells (Wertheim, et al. 2015) and are inherited vertically, drastically decreasing the effective population size, while removing the need for sexual reproduction (Mira and Moran 2002). The loss of sexual recombination during the switch from a free-living to an asexually-inherited obligate lifestyle results in the accumulation of deleterious mutations via Muller's Rachet (Muller 1964). For example, facultative and obligate γ-proteobacteria endosymbionts of insects show clear signals for relaxed selection when compared to their free-living relatives (Wertheim, et al. 2015). In addition to selective relaxation, gene loss and genomic minimization are a common feature of the transition to an obligate lifestyle, with the loss of genes that encode now superfluous metabolic pathways (Osvatic, et al. 2023), in addition to those involved in DNA regulation or repair (Shigenobu, et al. 2000). Accumulation of deleterious mutations can also drive pseudogenization, whereby genes lose functionality while being retained in the genome (van der Burgt, et al. 2014; Esfeld, et al. 2018).

Fungal symbiosis has been described throughout insects, with many obligate fungal species providing benefits to their host (Dowd 1992; Noda and Koizumi 2003; Douglas 2015; Stefanini 2018). *Ophiocordyceps* are entomopathogenic and nematopathogenic fungi that have evolved highly complex life histories to target, control and kill their host (Kobmoo, et al. 2018). Many species are highly host-specific, with species complexes evolving via unique mechanisms to infect their host (Evans, et al. 2011; Beckerson, et al. 2023). However, in soft-scale insects (Hemiptera:Coccomorpha) *Ophiocordyceps* have undergone multiple independent transitions from a pathogenic to associate lifestyle, where they are inherited vertically from female insects directly to their eggs (Szklarzewicz, et al. 2021). *Ophiocordyceps* associates have been demonstrated to live within the lipid rich storage tissues of their host and upon female maturation and egg production, migrate to the ovum to infect the eggs as they are forming (Szklarzewicz, et al. 2021).

In this study long-read sequencing was used to produce a highly contiguous genome from a single *Parthenolecanium corni* (Coccomorpha:Coccidae) individual along with its *Ophiocordyceps* obligate. Genomic analysis of the *Part. corni* genome identified multiple structural rearrangements and a decrease in the frequency of the conserved Mariner repeat family. Phylogenetic reconstruction of the *Ophiocordyceps* obligate confirmed its closest ancestor as *Hirsutella rhossiliensis* and identified widespread gene loss and signals for intensified selection within the nuclear genome of the fungal associate, while in contrast, the mitochondrial genome exhibited signals for relaxed selection. This work provides the first genome assembly of an obligate *Ophiocordyceps* and suggests widespread intensified selection coupled with diversifying selection on enzymes in the acetoacetate metabolism as key players in the transition from pathogenic to associate lifestyles in this system.

Methods:

Sample collection, DNA extraction and sequencing

A single *Parthenolecanium* corni individual was collected from a grapevine in the Langhorne Creek wine region of South Australia. The *Part. corni* individual was inspected under a stereomicroscope and washed using 95% ethanol and brushed three times to remove potential environmental contaminants. DNA was then extracted via gentle homogenization of the insect tissue using a glass rod in lysis buffer followed by 24-hour proteinase-K digestion, RNAse A incubation, a modified phenol-chlorophorm extraction (Ward, et al. 2021) and glycogen/isopropanol precipitation at 4°C. DNA molecular weight and quality was determined using a TapeStation (Agilent) at the Australian Genome Research Foundation. A total of 222 ng of DNA was obtained and used to prepare a nanopore library using the SQK-LSK114 kit. The library was loaded into a FLO-PRO114M flow-cell and sequenced in a P2-Solo instrument. Basecalling was performed using Guppy v.6.4.6 with the SUP (r10.4.1_e8.2_400bps_sup) model. Duplex paired reads were identified using duplex-tools and re-basecalled using Guppy. After quality filtering (Q > 10) a total sequencing yield of 76.17 Gb (71.98 simplex and 4.19 Gb of duplex reads) was obtained.

k-mer analysis

Simplex reads were used to carry out k-mer analysis using Canu v2.2 (Koren, et al. 2017) to investigate 22-mer depth distribution and estimate ploidy and genome size. Genome size estimation was carried out by calculating the area under the curve for local maxima present in the 22-mer depth distribution.

Genome assembly and separation

Metagenome assembly was carried out using Meta-Flye v2.9.3 (Kolmogorov, et al. 2020) using the combined simplex and duplex reads under default settings. Combined simplex and duplex reads were then mapped to the metagenome using Minimap2 v2.26 (Li 2018) with the map-ont preset. Mapped reads and the metagenome were then passed to MetaBat2 v2.15 (Kang, et al. 2019) for metagenome binning. A combined approach was used to determine if binned subgenomes should be combined. First; binned sub-genome average read depth and variance were compared to the 22-mer depth distribution, second; contig GC content distribution was compared between bins, third; a custom database containing all insect, fungal, and bacterial proteins on NCBI was constructed and contigs were queried using Kaiju v1.10.0 (Menzel, et al. 2016) and the BUSCO v4.1.2 (Seppey, et al. 2019) Insecta and Fungi datasets. NgsReports v2.4 (Ward, To, et al. 2020) was then used to visualize and investigate BUSCO outputs to identify differences in completeness between organisms.

Simplex and duplex fastq files were then mapped separately to the complete metagenome with Minimap2 v2.26 (Li 2018) using the map-ont preset without allowing for secondary mapping. Reads with a MAPQ>=20 were extracted from their respective fastq files and separated into subgenome fraction (ie. *Parthenolecanium* or *Ophiocordyceps* subgenome) using seqtk v75 (https://github.com/lh3/seqtk). The *Parthenolecanium* and *Ophiocordyceps* derived reads were then assembled separately using Verkko v2.0 (Rautiainen, et al. 2023) by providing duplex reads as the 'high accuracy' and simplex reads as the 'low accuracy' long-read arguments. The two Verkko assemblies were then combined along with Meta-Flye contigs that were not assigned to *Parthenolecanium* or *Ophiocordyceps.* All simplex and duplex reads were then mapped to the Meta-Flye/Verkko assembly and contig error correction was carried out using medaka (https://github.com/nanoporetech/medaka). Finally, corrected contigs were reassigned into *Parthenolecanium* or *Ophiocordyceps* derived genomes with SeqKit v2.7.0 (Shen, et al. 2016).

BUSCO v4.1.2 (Seppey, et al. 2019) was run at each stage of the assembly process using the Fungi and Insecta databases to determine single copy ortholog fraction being recovered and identify potential contamination.

Gene prediction and annotation of the Coccoidea and *Ophiocordyceps* genomes

Gene prediction was carried out on the *Parthenolecanium corni* assembly and publicly available Coccoidea genomes without gene models (Sup Table 1) using the braker2 pipeline (Brůna, et al. 2021) providing all Arthropod OrthoDB (Kriventseva, et al. 2019) orthogroups and Hemipteran proteins available on NCBI to generate hints for annotation.

The *Ophiocordyceps* genome was annotated using the Funannotate pipeline (10.5281/zenodo.2604804) by passing *Ophiocordyceps* proteins along with the SwissProt and UniProt databases to generate hints. Gene predictions were then functionally annotated for all genomes using the Funannotate (10.5281/zenodo.2604804) annotation function and Interproscan v5.59_91.0 (Jones, et al. 2014). GO terms were generated using the IPR domain to GO term table available through Gene Ontology (https://current.geneontology.org/ontology/external2go/interpro2go). Kegg terms were then generated using Kegg Mapper (Kanehisa and Sato 2020) and maps constructed using KEGG Reconstruct (Kanehisa 2017).

Analysis of the *Parthenolecanium* genome

Synteny analysis between the *Part. corni* genome and *Planococcus citri* (GCA_950023065.1) was carried out using MCScanX (Wang, et al. 2012) under default settings using contigs >1Mb in total length. Orthogroups were identified between *Part. corni* and publicly available Coccomopha genomes (Sup Table 1) protein annotations using OrthoFinder v2.2.7 (Emms and Kelly 2019) under default settings.

Repeat annotation was carried out by first modeling denovo repeats in the genome using RepeatModeler v2.0.5 (Flynn, et al. 2020) with the RepBase database 28.05 (Jurka, et al. 2005) and RepeatMasker databases (Chen 2004). Repeats were then annotated and masked in the genome using RepeatMasker v4.1.5 (Chen 2004).

Structural variation was characterized within the *Part. corni* genome by mapping the duplex reads to the *Part. corni* sub genome with Minimap2 v2.26 (Li 2018). Mapped reads were then called using Sniffles v2.2 (Sedlazeck, et al. 2018) to identify structural variants and IGV (Thorvaldsdóttir, et al. 2013) was used to manually inspect read mapping around these loci.

Analysis of the Ophiocordyceps genome

Synteny analysis between the *Ophiocordyceps* genome and *Cordyceps militaris* was carried out using MCscanX (Wang, et al. 2012) under default settings. Orthogroups were identified between *Ophiocordyceps* and Cordyceps genomes (Sup Table 2) protein annotations using OrthoFinder v2.2.7 (Emms and Kelly 2019) under default settings. A species tree was then constructed by combining individual single-copy ortholog gene trees using ASTRAL-III v5.7.1 (Zhang, et al. 2018). Gene loss analysis was carried out by parsing the OrthoFinder orthogroups output to identify genes present in the closest taxa, *H. rhossiliensis*, as identified using phylogenetic analysis. Two different levels of gene loss were then assessed by extracting genes using geaR v1.0 (Ward, Ludington, et al. 2020). First, genes present in all species within the same clade and second those that are present in all assessed *Ophiocordyceps* and *Cordyceps* genomes. GO and KEGG terms of lost genes were determined by first annotating the *H. rhossiliensis* genome using the same methodologies as the

Ophiocordyceps associate genome above. Annotations were then parsed for their GO and KEGG pathways. Enrichment analysis of GO terms was carried out using the R package clusterProfiler v3.18 (Yu, et al. 2012).

Selection testing

Single-copy orthologs between the *Ophiocordyceps* associate of *Parthenolecanium corni* (*OAc*), *Ophiocordyceps sinensis, Hirsutella rhossiliensis* and *Hirsutella minnesotensis* were identified using OrthoFinder v2.2.7 (Emms and Kelly 2019) and translation informed codon alignment was carried out with translatorX v1.1 (Abascal, et al. 2010). Gene alignments were then passed to HyPhy v2.5.59 (Pond, et al. 2005) for positive/diversifying selection testing using the BUSTED model (Murrell, et al. 2015). Two diversifying selection hypotheses were tested: first, only the *OAc* branch was tested for diversifying selection for each gene utilizing the species topology. Second, all other internal and terminal branchers excluding the *OAc* terminal branch were tested for diversifying selection. Genes were flagged as under diversifying selection if a statistically significant signal (pvalue ≤ 0.05) was observed in the *OAc* terminal branch without signal (pvalue > 0.05) detected across all other branches.

Intensification/relaxation of selection was calculated using RELAX model (Wertheim, et al. 2015) in HyPhy v2.5.59 (Pond, et al. 2005) utilizing the gene alignments of *OAc*, *Ophiocordyceps sine, Hirsutella rhossiliensis* and *Hirsutella minnesotensis* single copy orthologs. The *OAc* terminal branch was set to the foreground and evolutionary rate was determined against all other internal and terminal branches.

Duplicated orthogroups were identified from the nucleotide OrthoFinder output. Orthogroups that were found to be duplicated in only *OAc* were tested for diversifying positive selection using the ABSREL model (Smith, et al. 2015) in HyPhy v2.5.59 (Pond, et al. 2005). Each terminal branch was tested for selection and signal for positive diversifying selection was considered if significant signal (p value ≤ 0.05) in either one or both *OAc* paralogues.

Mitochondrial selection was investigated by using the mold codon model in HyPhy v2.5.59 (Pond, et al. 2005) with the ABSREL (Smith, et al. 2015) and RELAX (Wertheim, et al. 2015) models following the same methodology as nuclear genes.

Results:

Assembly of DNA sourced from a single scale individual recovers contiguous genomes for *Parthenolecanium corni* and a putative *Ophiocordyceps* associate

Sequencing the total genetic content within a single insect allows assembly of bacterial and/or fungal associates to be co-assembled (Ward and Baxter 2017). Due to the small size of soft-scale insects (<4 mm), recovery of sufficient high molecular weight DNA was challenging, however sufficient DNA was recovered for sequencing on a Nanopore P2 flow-cell, which produced 76.1 Gb of long read data (>q10, 71.9 Gb simplex; 4.2 Gb duplex). After removing shorter reads (<500 bp), 62 Gb (N_{50} of 8.05 kb) was retained for metagenomic assembly and which produced a 261.1 Mb assembly distributed across 1331 contigs and which was subsequently binned into 8 metagenome assembled genomes (MAGs) that varied in size from 212 Mb to 220 kb (Figure 1A, Sup Table 3). Combining MAGs based on taxonomic classification of the bins identified a single Insecta/Hemipteran and two Fungal genomes (Figure 1A) that were predicted to belong to the genera *Parthenolecanium* (Hemiptera), *Aureobasidium* (Dothideales) and *Ophiocordyceps* (Hypocreales), respectively (Figure 1A).

Duplex and simplex reads that mapped to the *Ophiocordyceps* and *Part. corni* MAGs were then extracted and used to perform specific assemblies for each organism, producing two highly contiguous genomes that were contained within 6 and 85 contigs, respectively. Assessment of duplex and simplex 22-mer depth revealed two clear peaks in the approximate location of mapped contig read depth distributions (Figure 1B). The 21-mer distribution was then used to determine the estimated genome size of each of the sub-genomes. The difference between recovered and estimated genome size provides insight into over clustering or bin mis-assignment during the assembly and separation stages. Calculating the area under each local maxima estimated the genome size of *Ophiocordyceps* to be 33 Mb and 212 Mb for *Part. corni,* compared to the 27.7 Mb and 227.8 Mb recovered genomes (Figure 1C) suggesting the genomes for both *Part. corni* and its *Ophiocordyceps* associate were well recovered.
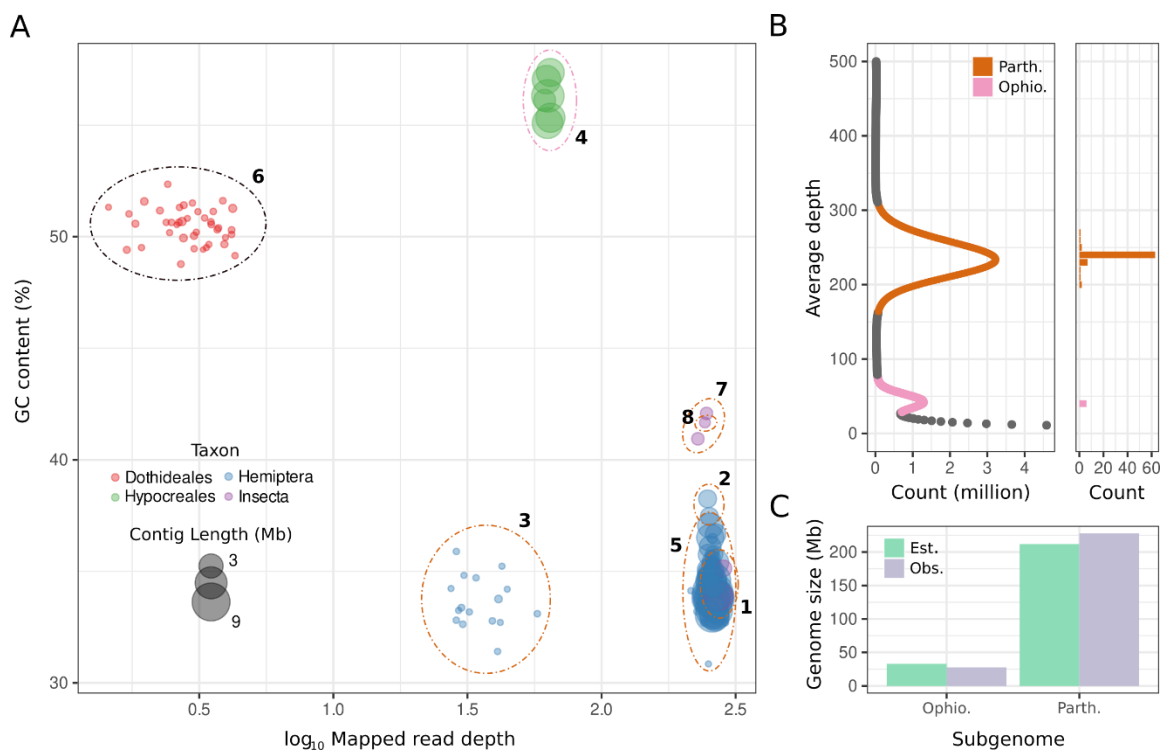

Figure 1: Metagenomic assembly of a single scale insect recovers two contiguous genomes. A) Metagenomic binning of contigs based on contig GC content and read depth. Contigs were then assigned to taxon using flexible protein matches and BUSCO gene content. Contigs are visualized based on contig length in megabases. B) Left: distribution of observed count vs read depth of 21-mers across the genome. Right: number of contigs within each read depth bin for each of the sub genomes. C) Genome size estimation through 21-mer frequency against the observed genome sizes of the *Ophiocordyceps* and *Parthenolecanium* sub-genomes post separation.

Structural rearrangements and decreased repetitive elements in the *Part. corni* genome

Analysis of completeness of the *Part. corni* genome revealed similar levels of BUSCO gene presence relative to genome assemblies of other Coccoid species (Sup Figure 1) with 85.6% single copy, 5.5% duplicated (91.1% complete) and 6.8% missing BUSCOs in the *Part. corni* genome. Gene annotation of the *Part. corni* genome identified 14,951 protein coding genes, which is well within the estimated

10,000 - 30,000 present in most insects (Adams, et al. 2000; Holt, et al. 2002; Consortium 2006; Consortium 2010; Ward, et al. 2021).

Clustering the *Part. corni* genes with other publicly available coccoid genomes identified 23,726 orthogroup clusters containing 86.3% of all genes within the search space, 951 of which represented single copy orthologs present across all genomes. Phylogenetic reconstruction of the *Part. corni* single copy orthologs against the background of mealybug and scale insect coccoid species identified the fellow scale insect *Ericerus pela* as the closest sequenced relative (Figure 2A). Scale insects were found to be in reciprocal monophyly with mealybugs (Figure 2A), agreeing with past literature (Vea and Grimaldi 2016). Synteny analysis of contigs greater than 1Mb in length (n = 24, 97% of total genome size) from the *Part. corni* genome against the chromosome level *Planococcus citri* (mealybug) genome identified almost no macro-synteny at the chromosome level (Figure 2B). To determine if this was due to misassembly in the *Part. Corni* genome, duplex ONT reads were mapped against the *Part. Corni* assembly. One hundred and sixty-nine putative structural variants (SVs) were identified within the *Part. Corni* genome and manual investigation of read alignments surrounding these loci confirmed 13 heterozygous InDels. No evidence for widespread mis-assembly was observed with called major SVs (e.g. inversions and translocations) being caused by mapping errors (Sup Figure 2).

Repeat modeling of the *Part. corni* genome identified 1,881 DNA, 38 LINE, 268 SINE and 1,364 LTR repetitive elements (Figure 2C), which represents far lower repeat content than the estimates performed on either its nearest neighbor, *E. pela* (Yang, et al. 2019) or the model Hemipteran, *Acyrthosiphon pisum* (Consortium 2010). For example, Mariner repeats are present throughout Hemiptera (Filée, et al. 2015; Bouallègue, et al. 2017; Yang, et al. 2019) and while 812 were observed in the *E. pela* repeat annotation (Yang, et al. 2019), *Part. corni* was found to only have 193.
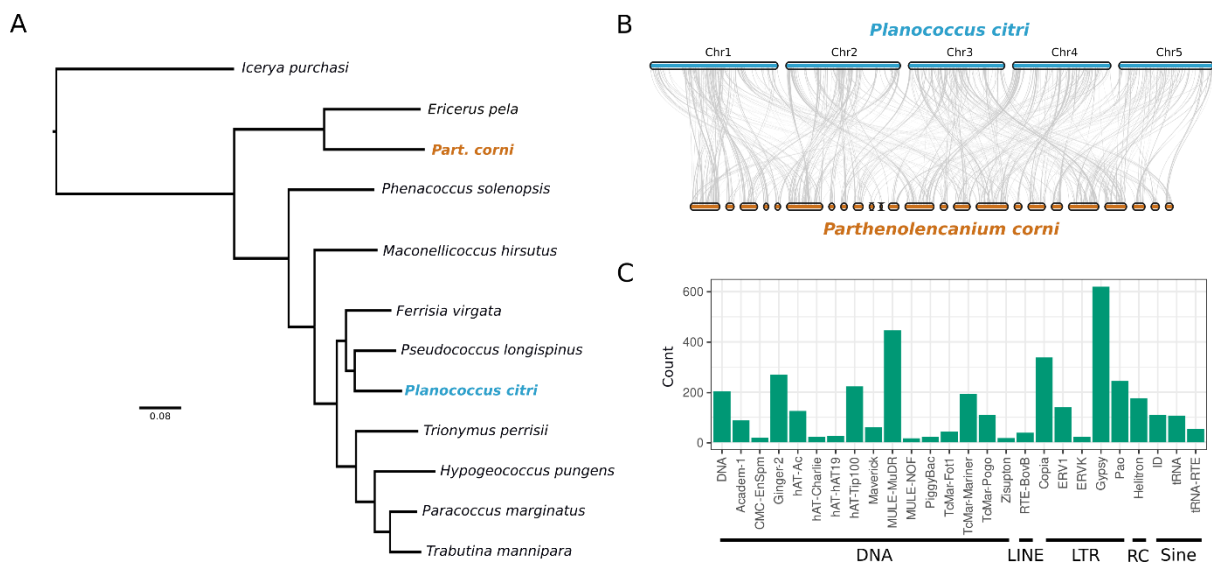

Figure 2: Genome analysis of *Parthenolecanium corni*. A) Phylogenetic reconstruction of Coccoid single copy orthologs. Branch lengths are in coalescent units. B) Comparative synteny analysis of *Planococcus citri* chromosomes (blue) and contigs longer than 1Mb (n = 24) *Parthenolecanium corni* contigs (orange) which made up 97% of the total *Part. corni* genome. C) Repeat content analysis of the *Part*. corni genome. Repeat classes are annotated along the y-axis.

The *Ophiocordyceps* associate of *Part. corni* has undergone gene loss in conserved orthologs

Ophiocordyceps and *Hirsutella* are synonymous genera, with *Hirsutella* being used to refer to the anamorphic, asexual life stage of this fungus (Simmons, et al. 2015), rather than the two genera representing phylogenetically distinct taxa. Phylogenetic reconstruction of the *Ophiocordyceps* associate of *Part. corni* (referred to as OAc) and publicly available *Ophiocordyceps* and *Cordyceps* genomes identified *Hirsutella rhossiliensis* as the closest sequenced relative to OAc, with high posterior support (Figure 3A). *OAc, H. rhossiliensis, O. sinensis* and *H. minnesotensis* formed a monophyletic clade separate from all other *Ophiocordyceps* genomes (Figure 3A; Clade 1), suggesting that this lineage shared a most recent common ancestor that was separate from other *Ophiocordyceps* species.

Analysis of the completeness of each genome revealed similar levels of BUSCO gene presence across all of the sequenced *Ophiocordyceps* species (Sup Figure 3) with 96% single copy, 0.9% duplicated and 2.7% missing BUSCOs identified in the OAc genome. The proportion of genes assigned to orthogroups was also similar across the sequenced *Ophiocordyceps*, with no species having a significant number of species-specific genes that were not able to be grouped (Figure 3A). Of the 7357 predicted gene models in OAc, 7247 were successfully assigned to an orthogroup with at least one other *Ophiocordyceps* or *Cordyceps* gene model and the single copy orthogroups present in all species totaled 2319.

Gene gain and loss are essential adaptive strategies employed by many fungi to respond to novel environmental stressors (Osvatic, et al. 2023). To investigate gene loss in the OAc lineage, 524 orthologs were identified that were present in Clade 1 (Figure 3A) but absent in the OAc annotation (Figure 3A). Of these 524 genes, 159 were present in all other taxa throughout the phylogeny (Figure 3B), suggesting that they are likely to play important roles in Hypocreales biology. Genes present in all other taxa are therefore referred to as 'conserved' throughout the remainder of the text. OAc exhibited an increased rate of conserved gene loss relative to all other *Ophiocordyceps* species except *O. sinensis* (Sup Figure 4). Functional annotation of the *H. rhossiliensis* orthologs that were absent in OAc revealed roles in amino-acid and xenobiotic metabolism, genetic information processing, and environmental signal processing (Figure 3B). Furthermore, gene loss was observed across the fungal MAPK signaling pathway, with genes required for proper hyphal/cell wall growth absent from OAc genome, including *WSC1*, *RHO1*, *YPD1* and *SSN6*, which have all been implicated in hyphal generation and/or cell wall integrity (Verna, et al. 1997; Hwang, et al. 2003; Sohn, et al. 2003; Krysan, et al. 2005; Richthammer, et al. 2012).

Toxic compounds, such as enterotoxins and mycotoxins, are an important mechanism for pathogenic *Ophiocordyceps* to kill or negatively impact their infected host (Lacadena, et al. 2007; Kobmoo, et al. 2018). To investigate if loss of toxin producing proteins had occurred during the transition to an associate life history in OAc, the 'lost' gene set was queried for genes involved in toxin production/pathogenicity. This identified 10 orthogroups related to toxin production that were absent from the OAc annotation, but which ranged from non-conserved (present in few taxa) to highly conserved (present in all taxa other than OAc) throughout *Ophiocordyceps* (Figure 3C). Two highly paralogous conserved orthogroups that were both predicted to encode Heat Labile Enterotoxins (HLEs) were found to be absent from OAc (Figure 3C; HLE). Four orthogroups that were predicted to encode proteins containing either Aflatoxin (Figure 3C; AfT) or Mycotoxin (Figure 3C; MyT) domains were also absent from the OAc lineage, along with other toxin domain containing orthogroups including Diphtheria-like toxin-like (Figure 3C; OG0000073) and ribotoxin (Figure 3C; OG0008875).
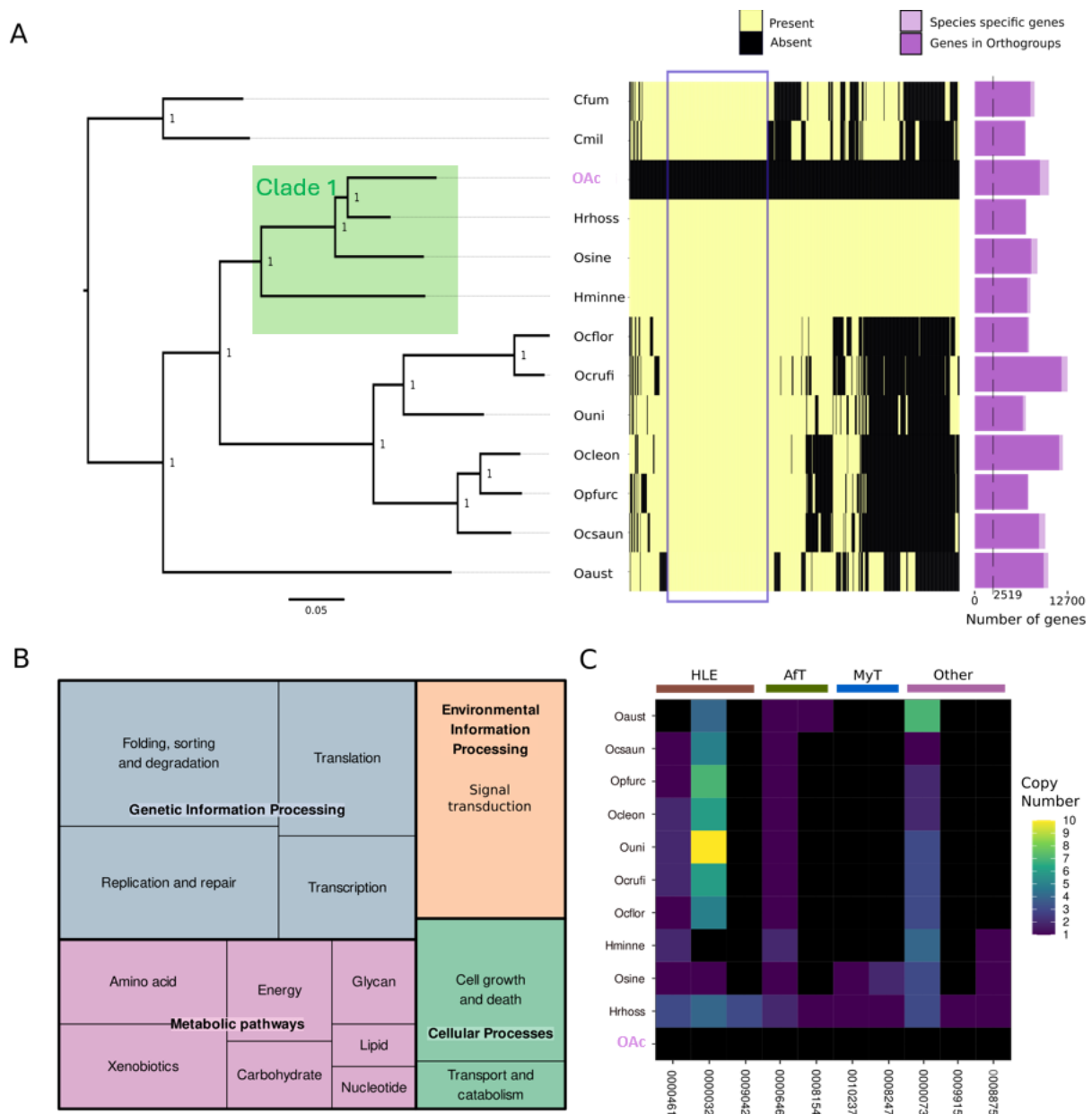
Figure 3: Widespread gene loss of conserved genes in OAc. A) Phylogenetic reconstruction of single copy orthologs of Cordyceps and Ophiocordyceps genomes (Sup Table 2). Gene loss in OAc for orthogroups (x-axis) that were present in the closest phylogenetic neighbors *H. rhossiliensis, O. sinensis* and *H. minnesotensis*. Cfum = *Cordyceps fumosorosea*; Cmil = *Cordyceps militaris*; Hrhoss = *Hirsutella rhossiliensis*; Osine = *Ophiocordyceps sinensis*; Hminne = *Hirsutella minnesotensis*; Ocflor = *Ophiocordyceps camponoti-floridani*; Ocleon = *Ophiocordyceps camponoti-leonardi*; Ocsaun = *Ophiocordyceps camponoti-saundersi*; Oaust = *Ophiocordyceps australis*; Opfur = *Ophiocordyceps polyrhachis-furcata*; Ocrufi = *Ophiocordyceps camponoti-rufipedis*; Ouni = *Ophiocordyceps unilateralis*. Purple box highlights orthogroups where all species contained at least one ortholog that were absent from OAc B) Kegg pathway breakdown of lost genes that are fixed with at least one copy in all entomopathogenic fungi in the tree. C) Loss of toxin related genes in OAc. Heat Labile Enterotoxin (HLE) (OG0000032, OG0000461, OG0009042), Aflatoxin (AFt) (OG0000646, OG0008154),

Mycotoxin (MyT) (OG0008247, OG0010237), and Other toxins (OG000073, OG0009915, OG0008875) related gene copy number are shown in species where the gene was observed.

Intensified and diversifying selection with potentially neofunctionalized paralogs shapes the transition to fungal symbiosis

Transition of life history from an entomopathogenic fungus to an associate likely requires complex changes in gene function and transcription. To understand the selection landscape of this transition, orthogroups were constructed using only species present in Clade 1 (Figure 3A) to maximize the number of nucleotide single-copy orthogroups for testing, identifying 2711 single-copy orthogroups for further analysis. Genome wide scans for relative changes in the intensity of selection in OAc against the background of non-associate *Ophiocordyceps* branches revealed 1120 (41.3%) orthologs that had k parameter values suggestive of intensified selection ($k \geq 1.1$), in addition to 912 (34.4%) with values consistent with relaxed selection ($k \leq 0.9$) (Figure 4A & B), while only 631 orthologs (23.3%) had a near-identical selective rate ($1.1 \geq k \geq 0.9$) across the entire phylogeny (Figure 4B). After testing for significant differences in evolutionary rate between the OAc and background non-associate *Ophiocordyceps* branches, 408 orthologs were identified that displayed a significant signal of either selection relaxation (n = 198) or intensification (n = 310) (Figure 4B), suggesting that intensified selective pressure, rather than relaxed selection, occurred more frequently in the OAc lineage since its separation from its most recent common ancestor (MCRA) with *H. rhossiliensis*.

To investigate the OAc genome for signatures of positive diversifying selection, ω (d_N/d_S) was calculated separately along the OAc terminal branch (Figure 4C; Foreground; ω_2) and all other branches in Clade 1 (Figure 4C; Background; ω_1) for each single-copy orthogroup. Comparing ω_1 and ω_2 values for each orthogroup revealed signals for elevated ω along the OAc terminal branch (ω_2) compared to the background (ω_1) (Figure 4C). Testing the foreground branch for diversifying positive selection in each single-copy ortholog identified 97 genes with significant signal of selection (p≤0.05) along the OAc lineage after its MCRA with *H. rhossiliensis* (Figure 4C). After filtering out orthologs with significant (p value ≤ 0.05) signal for selection across background branches, 82 genes were identified as displaying OAc specific diversifying. KEGG and GO functional analysis of these 82 genes under positive selection revealed functions involved in metabolism, cellular signalling and genetic information processing. With FAD binding (GO:0042254; p value = 0.019), ribosome biogenesis (GO:0042254; p value = 0.021), DNA-binding transcription factor activity (GO:0003700; p value = 0.025) and ATP-dependent chromatin remodeler activity (GO:0140658; p value = 0.027) GO terms enriched within the positive selected set compared to the entire OAc gene space.

In addition to the metabolism pathway selection above, significant (p≤0.05) signals for positive selection were detected specifically within acetoacetate metabolism (Sup Figure 5). The three genes (acetyl-CoA acetyltransferase, acetyl-Coenzyme A acyltransferase 1 and succinyl-CoA-3-oxaloacid CoA transferase), which are involved in catalyzing the reactions from acetyl −
Coenzyme A (CoA) ↔ acetoacetyl − CoA ↔ acetoacetate in the presence of CoA, showed clear signals for positive selection that were absent from the background branches.

Twenty-three gene duplications were identified that were specific to the OAc lineage in the Clade 1 orthogroup set. Tests for diversifying selection on duplicated paralogs revealed four OAc duplications with significant (p value ≤ 0.05) signal. Duplicated paralogs had increased omega values compared to all other lineages within their respective gene trees (Figure 4D), suggesting they may be undergoing neofunctionalization through directional selection. Gene function of putatively neofunctionalized genes included the Zinc Finger Protein 598 (OG0001049), TPR-like (OG0002189),

Fungal specific transcription factor (OG0002481) and DNA repair protein MutS (OG0002645) domain containing proteins.

In contrast to intensified selection across nuclear orthogroups (Figure 4A), mitochondrially-encoded genes exhibited frequent relaxed selection along the OAc lineage (Figure 4E). Six mitochondrial genes exhibited significant ($p \leq 0.05$) signal for relaxed selection (*NAD2*, *COX1*, *COX2*, *COX3* and *ATP8*) whereas only two (*NAD6* and *NAD3*) showed significant intensification. However, the majority (n = 7) of mitochondrial genes were found to lack significant signal for disparate selection intensities (Figure 4E).
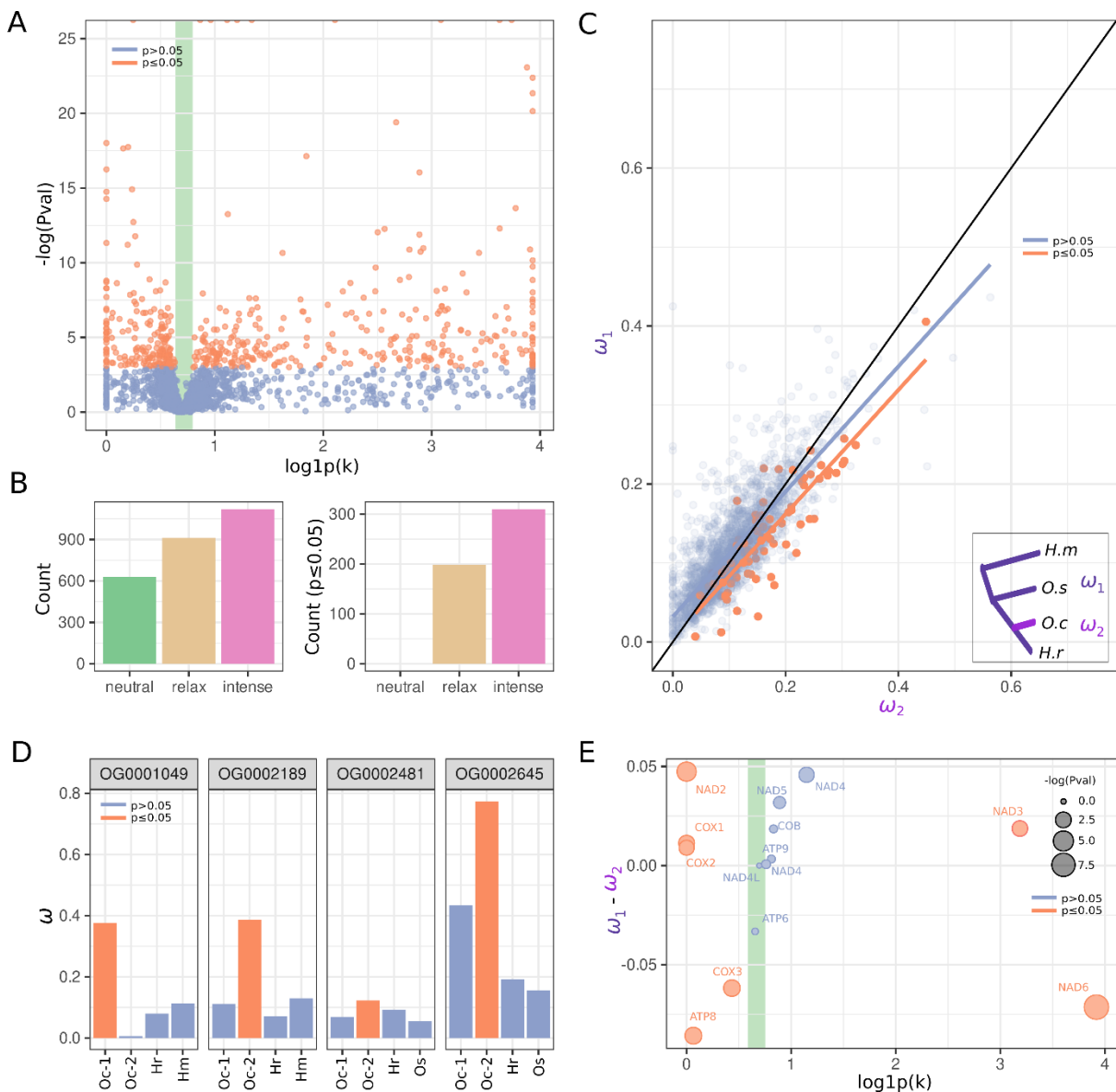

Figure 4: Widespread intensified and diversifying selection in the nuclear genome coupled with relaxed mitochondrial selection shapes the transition to symbiosis in OAc. A) Single copy ortholog (n = 2711) intensification parameter (k) and the significance of intensification ($k \geq 1.1$)/relaxation ($k \leq 0.9$) of selection observed in OAc against a background of *Ophiocordyceps sine, Hirsutella rhossiliensis* and *Hirsutella minnesotensis.* Green shading shows the expected value ($0.9 < k < 1.1$)

under the null hypothesis that both lineages are being selected at similar rates. B) Left: the total number of single-copy orthologs with $k \geq 1.1$ (intensified selection) and $k \leq 0.9$ (relaxed selection) along the OAc terminal branch. Right: the number of single-copy orthologs with significant signal for intensification ($k \geq 1.1$) or relaxation ($k \leq 0.9$) of selection. C) dN/dS ratio (ω) calculated along the OAc terminal branch (ω_2) and all other background branches (ω_1) of the species topology shown in the tree. Single-copy orthologs were tested for diversifying positive selection to a $p \leq 0.05$ cutoff. Solid colored lines represent the line of best fit calculated by glm for the significant and non-significant datapoints. Black line shows a linear relationship between ω_1 and ω_2. D) dN/dS ratio (ω) calculated across duplicated genes with positive selection. ω was calculated for all terminal branches of the tree shown in panel C and the significance level calculated for each terminal branch using a branch-site test for positive diversifying selection. E) Intensification parameter calculated across mitochondrial genes and the significance of the observed parameter. Green dashed line and shading shows the expected value ($0.9 < k < 1.1$) under the null hypothesis that both lineages are being selected at the near the same rate.

Discussion:

Long-read genome assembly of small insects is difficult due to the reduced amount of total DNA within individuals. To extract sufficient DNA researchers generally pool multiple individuals from the same clutch to carry out sequencing and assembly (Consortium 2010; You, et al. 2013; Yang, et al. 2019). Although this has benefits thorough the ability to carry out size selection, variant reads and those spanning recombination breakpoints can confound assembly algorithms resulting in fragmented assemblies (You, et al. 2013). In this study, genome assembly was performed from a single individual of *Parthenolecanium corni* producing contiguous assemblies for the insect and its *Ophiocordyceps* fungal associate.

Scale insects have a diverse set of reproductive life histories ranging from parthenogenic to sexual. *Parthenolecanium corni* is thought to have a parthenogenic life history, although males have been observed to exist within populations (Gimingham 1934; Santas 1985) and sexual reproduction is suggested to occur. In this study, almost no heterozygous genetic variation was observed, suggesting that the sequenced genome represents a highly-homozygous diploid, with only a small number of minor structural variants differentiating the two haplotypes. This may be due to the complexities of non-mendelian chromosome assortment during reproduction in *Part. corni* (Gavrilov and Kuznetsova 2004), as *Part. corni* exhibits a Comstockiella chromosome system (Gavrilov and Kuznetsova 2004; Gavrilov 2007), whereby paternally inherited chromosomes are condensed into a heterochromatic body in somatic cells and then eliminated (Hodson, et al. 2023). This, coupled with a suggested founder's effect (Ward, et al. 2023), may explain the limited genetic diversity present in the sequenced genome, rather than parthenogenesis. Chromosome number (1n) variability is high in scale insects (Coccidae), with various scale species estimated to have 5 – 18 chromosomes (Gavrilov 2007), compared to 4-32 chromosomes in mealybugs (Pseudococcidae) (Gavrilov 2007). High levels of structural rearrangements were observed between the chromosome level genome of the mealybug *Planococcus citri* and the *Part. corni* genome sequenced in this study, with low levels of macrosynteny between the two taxa. Taken together this suggests that the chromosome evolution of scale insects and mealybugs are exceptionally complex and more research is needed to more fully understand their evolutionary and population genetics.

Changes in gene function when transitioning life histories are well recorded, with strong selective pressures being imposed across the genome as the transitioning organism is exposed to novel

stressors (Schluter, et al. 1991; Kelly and Griffiths 2021). Symbionts are no exception with rapid diversification of key mechanisms such as metabolism and growth being observed (Brownlie, et al. 2007; Wernegreen 2015). After adaptation to their host, relaxed selection is generally observed in the genome as the symbiont becomes shielded from many of the selective pressures that affect their free-living sister taxa (Wernegreen 2011; Wertheim, et al. 2015). Changes in filamentous vegetative growth, production of fruiting bodies and metabolism efficiency have all been suggested to be key pathways required for the transition to an internal associate life history. In the *Ophiocordyceps* associate of *Parthenolecanium corni* (OAc), strong intensification of selection was observed across the genome. Although the transition to an associate life history occurred before the species diversification of the *Parthenolecanium* genus (Szklarzewicz, et al. 2021), OAc may still be adapting to an internal, non-pathogenic life history and may not yet be providing symbiotic benefit to its host. Transition from a pathogenic life history may require more complex evolutionary changes to occur, as the associate must first prevent host death while rapidly switching to vertical inheritance. Despite this, pathogenic fungi already have the ability for intracellular survival during the initial infection stages of their life cycle (Kobmoo, et al. 2018), potentially requiring fewer alterations in cell growth and response to abiotic and biotic stimuli to occur.

Gene loss and genome shrinkage is a common occurrence during the transition to an associate or symbiotic lifestyle (Moran and Mira 2001; Fan, et al. 2015; Osvatic, et al. 2023). Gene loss was observed in the *Ophiocordyceps* associate across many conserved genes with important functions in metabolism, cell wall integrity and gene regulation. During the adaptation to a symbiotic lifestyle, the associate or symbiont coopts fundamental processes and nutrients from the host, with insect fungal associates being shown to aid in host sterol metabolism (Noda and Koizumi 2003; Body, et al. 2013; Oliver 2023). During the transition from pathogenic to an associate life histories it was expected that evolutionary changes would occur in genes involved in fundamental pathways such as toxicity, cell growth, and sexual reproduction. In line with these predictions, gene loss was observed across genes involved in hyphal growth and cell wall integrity as part of the fungal MAPK signaling pathway, suggesting that filamentous growth may not be necessary in this *Ophiocordyceps* associates. Therefore, the *Ophiocordyceps* associate of *Part. corni* may have reverted to a yeast-like life history as has been observed in other independent transitions in fungal insect associates (Suh, et al. 2001).

Pathogenicity and host control traits in *Ophiocordyceps* is suggested to be underpinned by complex interactions between toxin-related genes and the host organism (Kobmoo, et al. 2018; Will, Attardo, et al. 2023; Will, Beckerson, et al. 2023). Previous work has implicated Heat labile enterotoxins as compounds in *Ophiocordyceps* pathogenicity (Kobmoo, et al. 2018). Gene loss was also observed in multiple highly paralogous toxin genes including Heat liable enterotoxins which have been suggested to play an important role in *Ophiocordyceps* pathogenicity. Ribotoxins also play a key role in entomopathogenicity of Ophiocordyceptidae and other fungal taxa (Lacadena, et al. 2007; Foresman and Tartar 2023), which are thought to avoid exoskeleton drilling by favoring the production of ribotoxins early in the infection process (Foresman and Tartar 2023). However, only a single highly conserved ribotoxin gene was observed to be lost in the associate's genome suggesting that widespread loss of this gene family has not occurred. This suggests that key genetic pathways are only partially undermined through gene loss to reduce pathogenicity during the transition to an associate life history.

Diversifying selection provides an evolutionary mechanism for neofunctionalization and changes in the efficiency of preexisting genes for adaptive novelty (Brownlie, et al. 2007; Murrell, et al. 2012). The *Ophiocordyceps* associate of *Part. corni* has undergone diversifying selection in multiple

important mechanisms, such as genetic information processing, potentially changing the transcriptional landscape. In addition to changes in transcriptional control we observed diversifying selection across three sequential enzymes in the acetoacetate metabolism pathway suggesting that adaptive change has occurred across this pathway. Although this may be due to symbiotic production of extracellular acetoacetate for uptake by *Part. corni*, the presence of the *Ophiocordyceps* within the lipid bodies of *Part. corni* provides an attractive hypothesis (Szklarzewicz, et al. 2021). The *Ophiodordyceps* associate may be coopting nutrients from its host through metabolism of extracellular acetoacetate produced by *Part. corni*. This is supported by diversifying positive selection being observed in adjacent enzymes catalyzing the reversible reaction to acetyl-CoA. Furthermore, relaxed selection was observed within the mitochondrial genome providing support that the *Ophiocordyceps* associate may have lowered energy metabolism requirements than their pathogenic ancestors. Future work should investigate the validity of this hypothesis and determine the fitness of *Part. corni* without the associate as our results suggest that OAc may not be acting in symbiosis with its host.

This study investigated the evolutionary landscape during the transition from pathogenic to non-pathogenic associate life history in the *Ophiocordyceps* associate of *Parthenolecanium corni*. The results highlight metabolism and transcription as important mechanisms to reprogram during the transition. To achieve this, a long-read genome assembly was produced for *Parthenolecanium corni*, representing the first Coccoidea species assembled from a single individual, revealing divergent transposable element frequencies compared to other Hemipterans. Contiguous assembly of *Part. corni* allowed co-recovery of a chromosome level assembly for its *Ophiocordyceps* associate producing the first genome sequence for a non-pathogenic *Ophiocordyceps*. Investigation into the evolutionary history of the transition from pathogenic to associate life histories identified widespread gene loss of conserved orthogroups involved in toxicity, pathogenicity and transcriptional regulation. The selective landscape of the nuclear genome was intensified since the seperation of OAc from its closest pathogenic relative, contrasted with relaxed mitochondrial selection. Furthermore, we observed diversifying selection specific to OAc in genetic information processing and metabolism. Taken together our results suggest gene loss, transcriptional reprogramming and acetoacetate metabolism are important mechanisms for the transition from pathogenic to associate life histories in this *Ophiocordyceps*.

Data availability:

All associated read data can be found within NCBI BioProject PRJNA1083313. Gene annotations and their corresponding assemblies are deposited at 10.5281/zenodo.10780696.

Acknowledgements:

References:

Abascal F, Zardoya R, Telford MJ. 2010. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic acids research 38:W7-W13.
Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF. 2000. The genome sequence of Drosophila melanogaster. Science 287:2185-2195.

Beckerson WC, Krider C, Mohammad UA, de Bekker C. 2023. 28 minutes later: investigating the role of aflatrem-like compounds in Ophiocordyceps parasite manipulation of zombie ants. Animal Behaviour 203:225-240.

Body M, Kaiser W, Dubreuil G, Casas J, Giron D. 2013. Leaf-miners co-opt microorganisms to enhance their nutritional environment. Journal of Chemical Ecology 39:969-977.

Bouallègue M, Filée J, Kharrat I, Mezghani-Khemakhem M, Rouault J-D, Makni M, Capy P. 2017. Diversity and evolution of mariner-like elements in aphid genomes. BMC genomics 18:1-12.

Brownlie JC, Adamski M, Slatko B, McGraw EA. 2007. Diversifying selection and host adaptation in two endosymbiont genomes. BMC evolutionary biology 7:1-12.

Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. 2021. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. NAR genomics and bioinformatics 3:lqaa108.

Chen N. 2004. Using Repeat Masker to identify repetitive elements in genomic sequences. Current protocols in bioinformatics 5:4.10. 11-14.10. 14.

Consortium HGS. 2006. Insights into social insects from the genome of the honeybee Apis mellifera. Nature 443:931.

Consortium IAG. 2010. Genome sequence of the pea aphid Acyrthosiphon pisum. PLoS biology 8:e1000313.

Douglas AE. 2015. Multiorganismal insects: diversity and function of resident microorganisms. Annual review of entomology 60:17-34.

Dowd PF. 1992. Insect fungal symbionts: a promising source of detoxifying enzymes. Journal of industrial microbiology and biotechnology 9:149-161.

Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome biology 20:1-14.

Esfeld K, Berardi AE, Moser M, Bossolini E, Freitas L, Kuhlemeier C. 2018. Pseudogenization and resurrection of a speciation gene. Current Biology 28:3776-3786. e3777.

Evans HC, Elliot SL, Hughes DP. 2011. Hidden diversity behind the zombie-ant fungus Ophiocordyceps unilateralis: four new species described from carpenter ants in Minas Gerais, Brazil. PLoS One 6:e17024.

Fan H-W, Noda H, Xie H-Q, Suetsugu Y, Zhu Q-H, Zhang C-X. 2015. Genomic analysis of an ascomycete fungus from the rice planthopper reveals how it adapts to an endosymbiotic lifestyle. Genome Biology and Evolution 7:2623-2634.

Filée J, Rouault J-D, Harry M, Hua-Van A. 2015. Mariner transposons are sailing in the genome of the blood-sucking bug Rhodnius prolixus. BMC genomics 16:1-17.

Flórez LV, Scherlach K, Gaube P, Ross C, Sitte E, Hermes C, Rodrigues A, Hertweck C, Kaltenpoth M. 2017. Antibiotic-producing symbionts dynamically transition between plant pathogenicity and insect-defensive mutualism. Nature Communications 8:15172.

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. Proceedings of the National Academy of Sciences 117:9451-9457.

Foresman D, Tartar A. 2023. The transcriptome of the entomopathogenic fungus Culicinomyces clavisporus contains an ortholog of the insecticidal ribotoxin Hirsutellin. PeerJ 11:e16259.

Gavrilov I, Kuznetsova V editors. Proceedings of 10th international symposium on scale insect studies, Adana. 2004.

Gavrilov IA. 2007. A catalog of chromosome numbers and genetic systems of scale insects (Homoptera: Coccinea) of the world. Israel Journal of Entomology 37:1-45.

Gimingham C. 1934. The Male Lecanium corni Bouché.

Hodson CN, Toon A, Cook LG, Ross L. 2023. Are asymmetric inheritance systems an evolutionary trap? Transitions in the mechanism of paternal genome loss in the scale insect family Eriococcidae. Genetics:iyad090.

Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, Nusskern DR, Wincker P, Clark AG, Ribeiro JMC, Wides R. 2002. The genome sequence of the malaria mosquito Anopheles gambiae. Science 298:129-149.

Hwang CS, Oh JH, Huh WK, Yim HS, Kang SO. 2003. Ssn6, an important factor of morphological conversion and virulence in Candida albicans. Molecular microbiology 47:1029-1043.

Iturbe-Ormaetxe I, Walker T, O'Neill SL. 2011. Wolbachia and the biological control of mosquito-borne disease. EMBO reports 12:508-518.

Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30:1236-1240.

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. 2005. Repbase Update, a database of eukaryotic repetitive elements. Cytogenetic and genome research 110:462-467.

Kanehisa M. 2017. Enzyme annotation and metabolic reconstruction using KEGG. Protein function prediction: methods and protocols:135-145.

Kanehisa M, Sato Y. 2020. KEGG Mapper for inferring cellular functions from protein sequences. Protein Science 29:28-35.

Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, Wang Z. 2019. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ 7:e7359.

Kelly MW, Griffiths JS. 2021. Selection experiments in the sea: What can experimental evolution tell us about how marine life will respond to climate change? The Biological Bulletin 241:30-42.

Kobmoo N, Wichadakul D, Arnamnart N, Rodríguez De La Vega RC, Luangsa-ard JJ, Giraud T. 2018. A genome scan of diversifying selection in Ophiocordyceps zombie-ant fungi suggests a role for enterotoxins in co-evolution and host specificity. Molecular ecology 27:3582-3598.

Kolmogorov M, Bickhart DM, Behsaz B, Gurevich A, Rayko M, Shin SB, Kuhn K, Yuan J, Polevikov E, Smith TP. 2020. metaFlye: scalable long-read metagenome assembly using repeat graphs. Nature Methods 17:1103-1110.

Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome research 27:722-736.

Kriventseva EV, Kuznetsov D, Tegenfeldt F, Manni M, Dias R, Simão FA, Zdobnov EM. 2019. OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. Nucleic acids research 47:D807-D811.

Krysan DJ, Ting EL, Abeijon C, Kroos L, Fuller RS. 2005. Yapsins are a family of aspartyl proteases required for cell wall integrity in Saccharomyces cerevisiae. Eukaryotic cell 4:1364-1374.

Lacadena J, Álvarez-García E, Carreras-Sangrà N, Herrero-Galán E, Alegre-Cebollada J, García-Ortega L, Oñaderra M, Gavilanes JG, Martínez del Pozo Á. 2007. Fungal ribotoxins: molecular dissection of a family of natural killers. FEMS microbiology reviews 31:212-237.

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34:3094-3100.

Menzel P, Ng KL, Krogh A. 2016. Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nature Communications 7:11257.

Mira A, Moran NA. 2002. Estimating population size and transmission bottlenecks in maternally transmitted endosymbiotic bacteria. Microbial Ecology 44:137-143.

Moran NA, Mira A. 2001. The process of genome shrinkage in the obligate symbiont Buchnera aphidicola. Genome biology 2:1-12.

Muller HJ. 1964. The relation of recombination to mutational advance. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis 1:2-9.

Murrell B, Weaver S, Smith MD, Wertheim JO, Murrell S, Aylward A, Eren K, Pollner T, Martin DP, Smith DM. 2015. Gene-wide identification of episodic selection. Molecular biology and evolution 32:1365-1371.

Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. 2012. Detecting individual sites subject to episodic diversifying selection. PLoS genetics 8:e1002764.

Noda H, Koizumi Y. 2003. Sterol biosynthesis by symbiotes: cytochrome P450 sterol C-22 desaturase genes from yeastlike symbiotes of rice planthoppers and anobiid beetles. Insect biochemistry and molecular biology 33:649-658.

Oliver KM. 2023. Flies co-opt bacterial toxins for use in defense against parasitoids. Proceedings of the National Academy of Sciences 120:e2304493120.

Osvatic JT, Yuen B, Kunert M, Wilkins L, Hausmann B, Girguis P, Lundin K, Taylor J, Jospin G, Petersen JM. 2023. Gene loss and symbiont switching during adaptation to the deep sea in a globally distributed symbiosis. The ISME Journal 17:453-466.

Pond SLK, Frost SD, Muse SV. 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics 21:676-679.

Rautiainen M, Nurk S, Walenz BP, Logsdon GA, Porubsky D, Rhie A, Eichler EE, Phillippy AM, Koren S. 2023. Telomere-to-telomere assembly of diploid chromosomes with Verkko. Nature Biotechnology:1-9.

Richthammer C, Enseleit M, Sanchez-Leon E, März S, Heilig Y, Riquelme M, Seiler S. 2012. RHO1 and RHO2 share partially overlapping functions in the regulation of cell wall integrity and hyphal polarity in Neurospora crassa. Molecular microbiology 85:716-733.

Santas L. 1985. Parthenolecanium corni (Bouche), an orchard scale pest producing honeydew foraged by bees in Greece. Entomologia Hellenica 3:53-58.

Schluter D, Price TD, Rowe L. 1991. Conflicting selection pressures and life history trade-offs. Proceedings of the Royal Society of London. Series B: Biological Sciences 246:11-17.

Sedlazeck FJ, Rescheneder P, Smolka M, Fang H, Nattestad M, Von Haeseler A, Schatz MC. 2018. Accurate detection of complex structural variations using single-molecule sequencing. Nature Methods 15:461-468.

Seppey M, Manni M, Zdobnov EM. 2019. BUSCO: assessing genome assembly and annotation completeness. Gene prediction: methods and protocols:227-245.

Shen W, Le S, Li Y, Hu F. 2016. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. PLoS One 11:e0163962.

Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H. 2000. Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. APS. Nature 407:81-86.

Simmons DR, Kepler RM, Renner SA, Groden E. 2015. Phylogeny of Hirsutella species (Ophiocordycipitaceae) from the USA: remedying the paucity of Hirsutella sequence data. IMA fungus 6:345-356.

Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Kosakovsky Pond SL. 2015. Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. Molecular biology and evolution 32:1342-1353.

Sohn K, Urban C, Brunner H, Rupp S. 2003. EFG1 is a major regulator of cell wall dynamics in Candida albicans as revealed by DNA microarrays. Molecular microbiology 47:89-102.

Stefanini I. 2018. Yeast-insect associations: It takes guts. Yeast 35:315-330.

Suh S-O, Noda H, Blackwell M. 2001. Insect symbiosis: derivation of yeast-like endosymbionts within an entomopathogenic filamentous lineage. Molecular biology and evolution 18:995-1000.

Szklarzewicz T, Michalik K, Grzywacz B, Kalandyk-Kołodziejczyk M, Michalik A. 2021. Fungal associates of soft scale insects (Coccomorpha: Coccidae). Cells 10:1922.

Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Briefings in bioinformatics 14:178-192.

Trivedi P, Leach JE, Tringe SG, Sa T, Singh BK. 2020. Plant–microbiome interactions: from community assembly to plant health. Nature reviews microbiology 18:607-621.

van der Burgt A, Karimi Jashni M, Bahkali AH, de Wit PJ. 2014. Pseudogenization in pathogenic fungi with different host plants and lifestyles might reflect their evolutionary past. Molecular plant pathology 15:133-144.

Vea IM, Grimaldi DA. 2016. Putting scales into evolutionary time: the divergence of major scale insect lineages (Hemiptera) predates the radiation of modern angiosperm hosts. Scientific Reports 6:23487.

Verna J, Lodder A, Lee K, Vagts A, Ballester R. 1997. A family of genes required for maintenance of cell wall integrity and for the stress response in Saccharomyces cerevisiae. Proceedings of the National Academy of Sciences 94:13804-13809.

Vogel KJ, Moran NA. 2011. Sources of variation in dietary requirements in an obligate nutritional symbiosis. Proceedings of the Royal Society B: Biological Sciences 278:115-121.

Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, Lee T-h, Jin H, Marler B, Guo H. 2012. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic acids research 40:e49-e49.

Ward C, Perry K, Baker G, Powis K, Heckel DG, Baxter S. 2021. A haploid diamondback moth (Plutella xylostella L.) genome assembly resolves 31 chromosomes and identifies a diamide resistance mutation. Insect biochemistry and molecular biology 138:103622.

Ward CM, Baxter SW. 2017. Draft genome assembly of a Wolbachia endosymbiont of Plutella australiana. Genome Announcements 5:10.1128/genomea. 01134-01117.

Ward CM, Ludington AJ, Breen J, Baxter SW. 2020. Genomic evolutionary analysis in R with geaR. bioRxiv:2020.2008. 2006.240754.

Ward CM, Onetto CA, Van Den Heuvel S, Dixon R, Borneman AR. 2023. Metagenomic ecosystem monitoring of soft scale insects and mealybug communities. OENO One 57.

Ward CM, To T-H, Pederson SM. 2020. ngsReports: a Bioconductor package for managing FastQC reports and other NGS related log files. Bioinformatics 36:2587-2588.

Welte CU, de Graaf RM, van den Bosch TJ, Op den Camp HJ, van Dam NM, Jetten MS. 2016. Plasmids from the gut microbiome of cabbage root fly larvae encode SaxA that catalyses the conversion of the plant toxin 2-phenylethyl isothiocyanate. Environmental Microbiology 18:1379-1390.

Wernegreen JJ. 2015. Endosymbiont evolution: predictions from theory and surprises from genomes. Annals of the New York Academy of Sciences 1360:16-35.

Wernegreen JJ. 2011. Reduced selective constraint in endosymbionts: elevation in radical amino acid replacements occurs genome-wide. PLoS One 6:e28905.

Wertheim JO, Murrell B, Smith MD, Kosakovsky Pond SL, Scheffler K. 2015. RELAX: detecting relaxed selection in a phylogenetic framework. Molecular biology and evolution 32:820-832.

Will I, Attardo G, de Bekker C. 2023. Multiomic interpretation of fungus-infected ant metabolomes during manipulated summit disease. Scientific Reports 13:14363.

Will I, Beckerson WC, de Bekker C. 2023. Using machine learning to predict protein–protein interactions between a zombie ant fungus and its carpenter ant host. Scientific Reports 13:13821.

Yang P, Yu S, Hao J, Liu W, Zhao Z, Zhu Z, Sun T, Wang X, Song Q. 2019. Genome sequence of the Chinese white wax scale insect Ericerus pela: The first draft genome for the Coccidae family of scale insects. GigaScience 8:giz113.

You M, Yue Z, He W, Yang X, Yang G, Xie M, Zhan D, Baxter SW, Vasseur L, Gurr GM. 2013. A heterozygous moth genome provides insights into herbivory and detoxification. Nature genetics 45:220-225.

Yu G, Wang L-G, Han Y, He Q-Y. 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics: a journal of integrative biology 16:284-287.

Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC bioinformatics 19:15-30.