Chromosomal resolution reveals symbiotic virus colonization of parasitic wasp genomes

Jérémy Gauthier, Hélène Boulain, Joke J.F.A. van Vugt, Lyam Baudry, Emma Persyn, Jean-Marc Aury, Benjamin Noel, Anthony Bretaudeau, Fabrice Legeai, Sven Warris, et al.

Chromosomal resolution reveals symbiotic virus colonization of parasitic wasp genomes

Jérémy Gauthier[1,2], Hélène Boulain[1,3], Joke J.F.A. van Vugt[4], Lyam Baudry[5,6], Emma Persyn[7], Jean-Marc Aury[8], Benjamin Noel[8], Anthony Bretaudeau[9,10], Fabrice Legeai[9,10], Sven Warris[11], Mohamed Amine Chebbi[1], Géraldine Dubreuil[1]; Bernard Duvic[12], Natacha Kremer[13], Philippe Gayral[1], Karine Musset[1], Thibaut Josse[1], Diane Bigot[1], Christophe Bressac[1], Sébastien Moreau[1], Georges Periquet[1], Myriam Harry[14], Nicolas Montagné[7], Isabelle Boulogne[7], Mahnaz Sabeti-Azad[7], Martine Maïbèche[7], Thomas Chertemps[7], Frédérique Hilliou[15], David Siaussat[7], Joëlle Amselem[16], Isabelle Luyten[16], Claire Capdevielle-Dulac[14], Karine Labadie[8], Bruna Laís Merlin[18], Valérie Barbe[8], Jetske G. de Boer[4,17,20], Martial Marbouty[5,6], Fernando Luis Cônsoli[18], Stéphane Dupas[14], Aurélie Hua Van[14], Gaëlle Le Goff[15], Annie Bézier[1], Emmanuelle Jacquin-Joly[7], James B. Whitfield[19], Louise E.M. Vet[4], Hans M. Smid[17], Laure Kaiser-Arnault[14], Romain Koszul[5], Elisabeth Huguet[1], Elisabeth A. Herniou[1] and Jean-Michel Drezen[1].

Abstract

Most endogenous viruses, an important proportion of eukaryote genomes, are doomed to slowly decay. Little is known, however, on how they evolve when they confer a benefit to their host. Bracoviruses are essential for the parasitism success of parasitoid wasps, whose genomes they integrated ~103 million years ago. Here we show, from the assembly of a parasitoid wasp genome, for the first time at a chromosomal scale, that symbiotic bracovirus genes spread to and colonized all the chromosomes. Moreover, large viral clusters are stably maintained suggesting strong evolutionary constraints. Genomic comparison with another wasps revealed that this organization was already established ~53 mya. Transcriptomic analyses highlight temporal synchronization of viral gene expression, leading to particle production. Immune genes are not induced, however, indicating the virus is not perceived as foreign by the wasp. This recognition suggests that no conflicts remain between symbiotic partners when benefits to them converge.

Main

Cotesia wasps (Hymenoptera, Braconidae) are parasitoids of Lepidoptera. Female wasps lay their eggs into caterpillars and larvae develop feeding on the host hemolymph. Several *Cotesia* species are famous for their use as biological control agents to control insect pests, such as *Cotesia flavipes* which is massively released over several million hectares of sugarcane fields in Brazil[1,2]. Parasitoid wasps evolved several strategies that increase parasitic success, including a sensitive olfactory apparatus to locate their hosts[3,4] and detoxification mechanisms against plant toxic compounds accumulating in their host (Fig. 1). But the most original strategy is the domestication of a bracovirus (BV) shared by over 46,000 braconid wasp species[5]. Bracoviruses originate from a single integration event ~103 million years ago (mya) of a nudivirus in the genome of the last common ancestor of this group[6-9]. Virus domestication nowadays confers a benefit to the wasps that use BVs as virulence gene delivery systems[10]. Indeed, virulence genes are introduced with wasp eggs into their hosts, causing inhibition of host immune defenses[10-12]. Imbedded in wasp DNA, the nudivirus genome has been extensively rearranged[13]. Now BV sequences are differentiated as i.) genes involved in particles production named "nudiviral genes"[6], and ii.) proviral segments packaged as dsDNA circles in viral particles[13], encoding virulence genes involved in successful parasitism.

To understand the genetic underpinning of the endoparasitoid life-style and of BV evolution since the original nudivirus capture event, we used a hybrid sequencing approach combining 454 reads, illumina short reads, and chromosomal contact data (HiC), to obtain a reference genome for *Cotesia congregata* at a chromosomal scale. These approaches allowed the analysis of wasp genes potentially involved in the success of the endoparasitoid lifestyle such as genes implicated in olfaction, detoxification and immunity. Moreover, the results revealed the global distribution of bracoviral sequences in wasp chromosomes and the identification of syntenies with genome scaffolds of *Microplitis demolitor* parasitoid wasp[14,15], which diverged from the *Cotesia* lineage 53 mya[16]. In addition, we sequenced draft genomes of five *Cotesia* species (*C. rubecula*, *C. glomerata*, *C. vestalis*, *C. flavipes*, *C. sesamiae*) to assess bracovirus molecular evolution between closely related species.

We discovered that, in discrepancy with the usual gene decay of endogenous viral elements, bracovirus genes have spread in all wasp chromosomes, reaching a ~2.5-fold higher number of genes than a pathogenic nudivirus. However strong evolutionary constraints have maintained specialized chromosomal regions containing large virus gene clusters, resembling "super gene" regions and concentrating 77% of virulence genes and half of the nudiviral genes,

respectively. Transcriptome analyses point to the major role of viral transcription in maintaining the viral entity, however wasp immune response is not induced, suggesting no conflicts remain in this ancient virus-wasp association.

Results

Genome assembly, annotation and comparison

First, a 207 Mb high-quality genome (scaffolds N50=1.1 Mb and N90=65 kb) was obtained for *Cotesia congregata* using a combination of mate pair 454 pyrosequencing and Illumina sequencing (Supplementary Table 1). Most of the assembly (86%) consisted of 285 scaffolds of over 100 kb. This genome was then re-assembled based on experimentally obtained chromosomal contact maps. The HiC method yielded ten main scaffolds comprising >99% of the previously obtained genome assembly (Supplementary Table 1 and 3; Supplementary Fig. 1 and 2), and corresponding to the ten chromosomes of *C. congregata*[17]. In addition, draft genomes of five related *Cotesia* species – *C. rubecula*, *C. glomerata*, *C. vestalis*, *C. flavipes* and *C. sesamiae* – were sequenced and assembled with Illumina shotgun sequencing reads (Supplementary Table 1). They respectively resulted in scaffold N50 values of 13, 9, 15, 20 and 26 Kb and cumulative sizes of 216, 243, 176, 155 and 166 Mb, in agreement with cytometry data (189-298 Mb) (Supplementary Table 2).

The genome of *C. congregata* comprised 48.7% of repeated DNA sequences including 34.7 % of known Transposable Elements (TEs) (Supplementary Fig. 3). Bracovirus proviral segments comprised other TE sequences that had previously been annotated as bracovirus genes: BV26 gene family corresponded to Sola2 elements (Supplementary Fig. 4; Supplementary Table 5) and CcBV 30.5 to a *hAT*[18]. Moreover, the analysis suggested that the BV7 gene family might also derive from TEs. Interestingly, some of these genes are highly expressed in the host[19] or likely participate in ecological adaptation[20]. This indicates that, contrary to a common paradigm in this field[21], the virulence genes packaged in bracovirus particles do not exclusively originate from the wasp cellular gene repertoire.

We annotated 14,140 genes in the genome of *C. congregata* (Methods and Supplementary Table 7), which include more than 99% of 1,658 conserved insect genes (98 to 99 % of the genes for the other *Cotesia* species, Supplementary Fig. 5).

Gene gains and losses in traits associated to the endoparasitoid lifestyle

Expert annotation of chemoreceptor gene families identified 243 odorant receptor (OR), 54 gustatory receptor (GR) and 105 ionotropic receptor (IR) genes in *C. congregata*. These numbers are in line with those of other parasitoid wasps, but are lower than in ants[22] (Supplementary Table 10). Phylogenetic analyses showed *C. congregata* ORs belong to 15 of the 18 OR lineages (Fig. 2) described in Apocrita[23] and revealed independent OR gene expansions in *N. vitripennis* and in Braconidae (Fig. 2). The most spectacular Braconidae-specific expansions occurred in five clades each harboring at least 25 genes in *C. congregata* (Fig. 2; Supplementary Table 10). Highly duplicated OR genes were found in 6 clusters of at least 10, and up to 19, tandemly arrayed genes (Supplementary Fig. 16). Within Braconidae, many duplications occurred in the ancestors of *Cotesia* but OR copy numbers varied significantly between species (Fig. 2). This illustrates the highly dynamic evolution of OR gene families within parasitoid wasps and between *Cotesia* species, which have different host ranges.

All the genes involved in detoxification in insects were identified in *C. congregata* and are largely conserved within *Cotesia* (Supplementary Table 12 and 13). For instance, each species harbors conserved numbers of UDP-glucosyltransferases (UGTs) and slightly different numbers of gluthatione-S-transferases (GSTs). In contrast, carboxylesterases (CCEs) and cytochrome P450 (P450s) numbers vary widely with *C. flavipes* and *C. sesamiae* harboring few representatives (respectively 22-24 CCEs and 49 P450s), compared to the 32 CCEs of *C. rubecula* and the 70 P450s found in *C. congregata*, which are both exposed to plant toxic compounds (Supplementary Table 12 and 13). *Cotesia* specific P450 families were identified in the clan 3 and 4, both of which are often associated to adaptation to plant compounds and insecticides[24] (Supplementary Table 12 and 13). Altogether *Cotesia* appear equipped for detoxification although no spectacular gene expansion was observed.

The complete genome annotation of *C. congregata* revealed 102 nudiviral gene copies that have colonized all the chromosomes. This number is similar to that of pathogenic nudiviruses[25], which is an unexpected result given that endogenous viral elements usually undergo gene loss in the course of evolution[26]. Here this surprisingly high number of nudiviral genes can be explained by the fact that gene losses have been balanced by expansions of gene families. At least 25 of the 32 nudivirus core genes involved in essential viral functions[27] have been retained in the wasp genome, with the notable absence of the nudiviral DNA polymerase (Supplementary Table 8). The loss of this major replication gene may have prevented viral escape[10,25,28]. The *fen* genes, generally involved in DNA replication, form a gene family with six tandem copies that is found specifically in the *Cotesia* lineage (Fig. 3B). The most spectacular

131 expansion concerns the *odv-e66* gene family with 36 genes in 10 locations (Fig. 2B; Fig. 3C),
132 in contrast to the one or two copies typically found in nudivirus genomes[29]. This expansion
133 occurred both before and after the divergence between *C. congregata* and *M. demolitor* since
134 we found tandemly duplicated copies in homologous loci of the two species or in *C. congregata*
135 only (Fig. 3D). In baculoviruses, *odv-e66* encodes a viral chondroitinase[30] involved in digesting
136 the peritrophic membrane lining the gut, thus allowing access to target cells during primary
137 infection. Different ODV-E66 proteins may similarly allow BVs to cross various host barriers,
138 and BV infection to spread to virtually all Lepidoptera tissues[31,32].

139 *Genomic architecture and synteny of bracovirus genes*

140 Chromosome scale genome assembly of *C. congregata* provides the first comprehensive
141 genomic organization of a bracovirus within the genome of a wasp. Half of the single copy
142 nudiviral genes are located in the nudiviral cluster, which comprises 25 genes and is amplified
143 during bracovirus production but not packaged within the particles[33]. Comparison with the
144 scaffolds of *M. demolitor* showed overall gene content conservation as well as conserved
145 syntenic blocks in the genomic regions flanking nudiviral sequences, over ~53 million years of
146 evolution. This confirms that the nudiviral clusters of both species are orthologous (Fig. 3) and
147 likely derive from a genome fragment of the nudivirus that originally integrated in the ancestor
148 wasp genome. This striking stability suggests that major evolutionary constraints maintain these
149 genes together. The other nudiviral genes are dispersed in the wasp genome, although not
150 evenly, as more loci are located in the 4 smaller chromosomes. Orthology with *M. demolitor*
151 could be identified for 20 nudiviral gene regions (Fig. 3; Supplementary Fig. 9), indicating they
152 were already dispersed in the last common ancestor of both wasps and have stayed in the same
153 loci. Altogether, this showed that nudivirus gene loss and dispersion occurred during the first
154 period of wasp-bracovirus association (100 to 53 mya).

155 The expansion of virulence genes is another aspect of wasp genome colonization. In *C.*
156 *congregata*, the 249 virulence genes encoded by proviral segments are found in 5 regions of
157 the genome located on three chromosomes. However, 77% of these genes clustered in a single
158 region, which comprised four physically linked proviral loci (PL1, PL2, PL4, PL10)
159 interspersed by wasp genes (Fig. 3). This major virulence gene coding region, named
160 "macrolocus"[13], spans half of chromosome 5 short arm and is almost comparable (2 Mb, 177
161 genes) to the Major Histocompatibility Complex (MHC) involved in mammalian immunity (~4
162 Mb, ~260 genes)[34]. Orthology relationships could be inferred between the PL1 in *C. congregata*

macrolocus and the largest proviral region (comprising 11 segments) of *M. demolitor*[14] (Fig. 3) but the macrolocus expanded specifically in the *Cotesia* lineage since *M. demolitor* does not have regions corresponding to PL2, PL4 and PL10. Further syntenies were found between 5 isolated proviral loci, showing they pre-date wasp lineage split (Fig. 3) whereas the 3 proviral loci in the long arm of chromosome C9 and C10 appear as a genuine novelty of the *Cotesia* lineage. The homologous relationships of PLs contrasts with their gene content divergence[35], which show high turnover probably because of their role as virulence factors and coevolution with the hosts[20].

Strong conservation of the bracoviral machinery

The DNA circles packaged in bracovirus particles are produced following the genomic amplification of replication units (RU) that span several proviral segments of PLs[33,36]. Detailed genomic analyses of *C. congregata* data led to the identification of a specific sequence motif at each RU extremity and confirmed the presence of circularization motifs[13,37] on all proviral segments at the origin of packaged circles indicating the conservation of the viral machinery whatever the localization of viral sequences. (Fig. 4; Supplementary Fig. 10).

Over 100 million years of evolution, the conservation of viral functions in microgastroid wasps must have endured through several bouts of natural selection. Synonymous to nonsynonymous substitution ratio analyses on orthologous nudiviral genes in *Cotesia* showed most of the nudiviral genes (65 genes among the 79 tested genes), are evolving under conservative selection, that is however less stringent than on the insect genes used to assess genome completeness (Fig. 4; Supplementary Table 8). This selection is notably strong for genes involved in viral transcription, such as the RNA polymerase subunits (*lef-4*, *lef-8*, *lef-9*, *p47*), which most likely control nudiviral gene expression and, consequently, particle production[6,8]. In contrast, genes involved in infectivity (homologues of baculovirus *pif* genes) appear less conserved (Supplementary Table 8). This might reflect divergence occurring during host shifts, through adaptation of virus envelope proteins to particular host cell receptors. The large *odv-e66* gene family and duplicated genes (*p6.9_2*, *pif-5_2*, *17a*) similarly displayed less stringent to relaxed selection (Fig. 4; Supplementary Table 8), which might be conducive to mutational space exploration for adaptation by neo-functionalization or sub-functionalization[38]. Virulence genes encoded by proviral segments globally displayed lower conservative selection than nudiviral genes (Fig. 4), as expected for genes interacting with host defenses and involved in evolutionary arms race or adaptation to new hosts[20].

Synchronized nudiviral transcription precedes bracovirus production

Bracovirus particle production only occurs in the calyx region of the ovaries, and starts before wasp emergence at day 5. Detailed RNAseq analysis of wasp ovaries and venom glands during pupal development showed 91 (out of 102) nudiviral genes were only expressed in the ovaries and could not be detected in the venom gland, which highlights strong ovarian tissue specificity. The onset of nudiviral gene expression was detected from day 2 of pupal development, with the expression of the RNA polymerase subunits (Fig. 5), reaching a peak on day 4 and declining in later stages (Fig. 5). The other nudiviral genes followed roughly the same expression pattern, low or undetectable at day 2, high at day 3 and reaching a peak at day 5 (Fig. 5). This pattern concurs with the hypothesis that the nudiviral RNA polymerase controls the expression of the other nudiviral genes and therefore bracovirus particle production. Notably, 12 nudiviral genes reached very high mRNA levels (at day 5) and are among the top 50 of most expressed genes in *C. congregata*. Three genes from the nudiviral cluster (including a major capsid component vp39) are by far the 3 most expressed of all wasp genes. Altogether, it is remarkable that the expression of almost all nudiviral genes remains strongly synchronized during pupal ovarian development. Considering the age of the nudivirus wasp association, this suggests strong evolutionary constraints.

Immune gene expression during bracovirus production

After 100 million years of endogenous evolution within the wasp genome, one can question whether the bracovirus is considered as a regular part of the wasp or is still perceived as non-self, in which case its production should trigger an immune response. However, whether viral production interacts in any way with the wasp immune system was totally unknown. Globally, the annotation of immune-related genes indicated *C. congregata* has an arsenal of 258 immune genes that are potentially involved in recognition, signal transduction, different signaling pathways, melanization and effector functions (Supplementary Table 9), in accordance with the recently reported annotation of *C. Vestalis* immune genes[39]. We identified all members of the Toll, IMD, Jak/STAT and RNA interference pathways found in Hymenoptera (Supplementary Table 9).

In contrast to the sharp increase in nudiviral gene expression, no significant changes in immune genes expression could be detected in the ovaries during pupal development (Fig. 6). In particular, expression of the genes involved in antiviral immunity (encoding members of the RNA interference, Jak/STAT or Imd/JNK pathways) was high in ovaries, even at stages before

particle production is observable (ovaries stages 2 to 4), but hardly fluctuated during the course of ovary development, even at day 5, when massive particle production becomes apparent by TEM (Fig. 5A). Thus, no immune response appears to be induced or repressed at the cellular level as a response to high level nudiviral genes expression and particles production. This strongly suggests the bracovirus is not perceived as foreign by wasp ovarian cells (Fig. 1).

Discussion

To investigate the genome features related to the endoparasitoid lifestyle, we sequenced the genomes of 6 *Cotesia* species, and obtained an assembly at the chromosomal scale for *C. congregata*. This approach provided insights into essential functions of the wasps, such as olfaction, detoxification and immunity, as well as on the genomic evolution of the bracovirus. Large OR gene diversifications are often associated to host localization and acceptance. Indeed, female wasps rely on sensitive olfactory apparatus to locate their hosts from volatile cues plants emit in response to herbivore attacks[3,4] and to assess caterpillar quality before oviposition[40]. Interestingly, OR copy numbers varied significantly during the evolution of the genus *Cotesia* (Fig. 2). The high dynamics of OR repertoire might point to the need for more specific recognition of chemical cues from the host and its food plant.

In contrast, no comparable diversification was observed in the detoxification arsenal, even though *Cotesia* larvae can be exposed to various toxic phytochemicals while feeding on the hemolymph of caterpillar hosts (i.e. potential exposure of *C. congregata* to insecticidal nicotine when parasitizing *Manduca sexta* feeding on tobacco; *C. rubecula*, *C. glomerata* and *C. vestalis* are potentially exposed to glucosinolates by developing in hosts consuming crucifers; and *C. flavipes* to phenylpropanoids and flavonoids when parasitizing hosts on sugarcane). This surprisingly low diversification of the detoxification arsenal could suggest that wasp larvae may not be as exposed to toxic compounds as expected due to direct excretion of these chemicals by the host larvae[41,42].

Cotesia wasps face different immune challenges during their lifetime. While feeding on nectar the adult might be exposed to similar environmental challenges as honey bees. Development inside the caterpillar host could on the one hand shield wasp larvae from pathogens but on the other hand expose them to opportunistic infections, because parasitism alters caterpillar immune defenses. Lastly, metamorphosis coincides with the production of bracovirus particles, against which wasp antiviral responses had so far not been investigated.

Insects were recently shown to recognize their obligate bacterial symbionts as foreign and to exert strong immune control, as documented for *Sitophilus oryzae* Primary Endosymbiont (SOPE)[43]. As the immunity gene arsenal of *C. congregata* is comparable to that of the honey bee, this wasp is most probably able to mount an immune response against pathogens, including viruses. However, the transcriptomic approach did not reveal any significant difference in immune gene expression between the ovaries of different pupal stages, although massive amounts of bracovirus particles are produced from day 5. This lack of anti-viral gene induction suggests the wasp no longer perceives the bracovirus as potentially pathogenic but rather considers it as a regular secretion.

With the ancestral integration of a nudivirus genome, the parasitoid wasp gained a series of viral functions: including viral transcription, viral DNA replication, particle morphogenesis and infectivity. Whereas the function of viral DNA replication has been lost[10], thus impeding autonomous virus re-emergence, the other viral functions have been reorganized for virulence gene transfer via bracovirus particles. Chromosomal scale resolution of *C. congregata* genome showed bracovirus genes have colonized all the chromosomes with a nearly 2.5 fold increase of in the total number of virus genes compared to the genome of a pathogenic nudivirus. This contrasts sharply with the decay of most viruses integrated into eukaryote genomes that do not provide a benefit to their host. Bracovirus dispersion occurred between 100 and 53 mya, as 25 viral loci are orthologous between *Cotesia* and *Microplitis* (Fig. 3; Supplementary Fig. 5). Yet, the organization of many bracovirus genes in clusters suggests strong evolutionary constraints maintain these genes together. In the case of the nudiviral cluster, which encodes major capsid components (VP39, 38K) on chromosome 7, DNA amplification, as a single unit[33], might be essential to the mass production of bracoviral particles that are injected into parasitized hosts. This could explain the counter selection on the dispersion or loss of these clustered nudivirus genes since the separation of the *Cotesia* and *Microplitis* lineages. The genomic organization stasis is reminiscent of bacterial symbiont genomes, which underwent major rearrangements specifically during the initial steps of association[44]. In the case of the particularly large macrolocus which comprises 77% of virulence genes in *Cotesia* genome, clustering could facilitate the evolution of new virulence genes copies by duplications[13], and thereby wasp adaptation against host resistance or to new hosts[10,45] This organization may also promote the transmission of bracovirus virulence genes as a block of co-evolved genes as shown for supergenes involved in butterfly wing patterns[46] and ant social behavior[47].

Remarkably, despite their dispersed locations in the wasp genome, bracovirus genes

remain synchronically expressed and under conservative selection, thus enabling the production of infectious virus particles. This striking conservation of the viral machinery highlights the paramount importance of the production of viral particles allowing virulence gene transfer to a host, in the evolutionary history of the wasp. Although high nudiviral gene expression and massive particle production underline the continuing existence of a virus entity, it is now considered as a regular secretion by the wasp. Thus, contrary to a vision of obligatory mutualism as an unstable alliance in which partners can become opponents or cheaters[48,49], this very ancient association shows conflicts disappear when the benefits of the partners converge.

Methods

Sampling

The *C. congregata* laboratory strain was reared on its natural host, the tobacco hornworm, *M. sexta* (Lepidoptera: Sphingidae) fed on artificial diet containing nicotine as previously described[50]. *C. sesamiae* isofemale line came from individuals collected in the field in Kenya (near the city of Kitale) and was maintained on *Sesamia nonagrioides*[51]. *C. flavipes* individuals originated from the strain used for biological control against *Diatraea saccharalis* in Brazil[52]. *C. glomerata*, *C. rubecula* and *C. vestalis* laboratory cultures were established from individuals collected in the vicinity of Wageningen in Netherlands, and reared on *Pieris brassicae*, *Pieris rapae* and *Plutella xylostella* larvae, respectively[53,54]. To reduce the genetic diversity of the samples prior to genome sequencing, a limited number of wasps were pooled; for example, only haploid males from a single virgin female were used for Illumina sequencing of *C. congregata* genome, ten female and male pupae originating from a single parent couple were used to generate *C. glomerata* genome, five male pupae originating from a single *C. rubecula* virgin female for *C. rubecula* genome and 40 adult males and 8 adult females from multiple generations of *C. vestalis* cultured in the Netherlands were used. DNAs were extracted from adult wasps and stored in extraction buffer following two protocols. *C. congregata, C. sesamiae* and *C. flavipes* DNA were extracted using a Qiamp DNA extraction kit (Qiagen) with RNAse treatment following the manufacturer's instructions and eluted in 200 µl of molecular biology grade water (5PRIME). *C. glomerata*, *C. rubecula* and *C. vestalis* DNA was extracted using phenol-chloroform.

Genome sequencing and assembly

Cotesia congregata genome was sequenced combining two approaches: (i) single-end reads and MatePair libraries of 3Kb, 8Kb and 20Kb fragments on 454 GS FLX Titanium platform (Roche) and (ii) paired-end reads of 320bp fragments with HiSeq2000 platform (Illumina). *C. glomerata*, *C. rubecula* and *C. vestalis* DNA libraries were prepared using insert sizes of 300 and 700 bp. For *C. sesamiae* and *C. flavipes* libraries an insert size of 400 bp was selected. These libraries were sequenced in 100 bp paired-end reads on a HiSeq2000 platform (Illumina) at the French National Sequencing Institute (CEA/Genoscope, France) and at the Sequencing Facility of the University Medical Center (Groningen, Netherlands). Reads were then filtered according to different criteria: low quality bases (Q<20) at the read extremities, bases after the second N found in a read, read shorter than 30 bp and reads matching with phiX (Illumina intern control) were removed using in-house software (http://www.genoscope.cns.fr/fastxtend) based on the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) as described in (https://www.nature.com/articles/sdata201793).

The *C. congregata* genome was generated by *de novo* assembly of 454 reads using GS De Novo Assembler from the Newbler software package v2.8[55]. The consensus was polished using the Illumina data as described in (https://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-9-603). Gaps were filled using GapCloser module from the SOAPdenovo assembler (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3626529/). The genomes of *C. sesamiae*, *C. flavipes, C. glomerata*, *C. rubecula* and *C. vestalis* were assembled with Velvet v1.2.07[55] using the following parameters: velveth *k*-mer 91 -shortPaired -fastq -separate, velvetg -clean yes and specific adapted values for -exp_cov and -cov_cutoff for each species. The genome size was estimated using *k*-mer distribution of raw reads on all *Cotesia* species with 17-mer size using Jellyfish v1.1.11 [56].

Chromosome scale assembly of *C. congregata* genome

HiC library preparation

The individual wasps had their gut removed and were immediately suspended after sampling in 30 mL of 1X tris-EDTA buffer and formaldehyde at 3% concentration, then fixed for one hour. 10 mL of glycine at 2.5 M was added to the mix for quenching during 20 min. A centrifugation recovered the resulting pellet for −80°C storage and awaiting further use. The libraries were then prepared and sequenced (2 × 75 bp, paired-end Illumina NextSeq with the first ten bases acting as tags), as previously described[57] using the *Dpn*II enzyme.

Read processing and Hi-C map generation

The HiC read library was processed and mapped onto *Dpn*II fragments of the reference assembly using HiCbox (available at https://github.com/kozullab/HiCbox) with bowtie2[58] on the back-end (option --very-sensitive-local, discarding alignments with mapping quality below 30). Fragments were then filtered according to size and coverage distribution, discarding sequences shorter than 50 bp or below one standard deviation away from the mean coverage. Both trimmed contact maps were then recursively sum-pooled four times by groups of three, yielding bins of 3^4= 81 fragments.

Contact-based re-assembly

The genome was re-assembled using an updated version of GRAAL (dubbed instaGRAAL) for large genomes on the aforementioned contact maps for 1,000 cycles, as described in[59]. Briefly, the program modifies the relative positions and/or orientations of sequences according to expected contacts given by a polymer model. These modifications take the form of a fixed set of operations (swapping, flipping, inserting, merging, etc.) on the 81-fragment bins. The likelihood of the model is maximized by sampling the parameters with an MCMC method. After a number of iterations, the contact distribution converges and the global genome structure ceases to evolve, at which point the genome is considered reassembled. The process yielded eleven main scaffolds comprising >99% of the bin sequences.

Polishing and post-assembly processing

Each scaffold was independently polished by reconstructing the initial contig structure whenever relocations or inversions were found. In addition, previously filtered sequences were reintegrated next to their original neighbors in their initial contig when applicable. The implementation is part of instaGRAAL polishing and available at https://github.com/koszullab/instaGRAAL (run using the –polishing mode).

Assembly validation

We performed the validation with QUAST-LG[60], an updated version of the QUAST analyzer tailored for large genomes. The initial assembly from Illumina short reads was used as reference. The assessed metrics include genomic feature completeness, Nx and NGx statistics as well as global and local misassemblies. In addition, each assembly was assessed for ortholog completeness with BUSCO v3[61]. The reference dataset used for expected gene content was pulled from the OrthoDB (version 9) database for Hymenoptera, comprising 4,415 orthologs.

Genome annotations

Transposable element annotation

Genome annotation was first done on the *C. congregata* reference genome and then exported on the genomes of the five other *Cotesia* species. First, the transposable element annotation was realized following the REPET pipeline comprising a *de novo* prediction (TEdenovo) and an annotation using TE libraries (TEannot)[62]. This annotation was exported into GFF3 files used as mask for the gene annotation.

Automated gene annotation

The automated gene prediction and annotation of *C. congregata* genome was done using Gmove (https://github.com/institut-de-genomique/Gmove) integrating different features based on (i) the mapping of Hymenoptera proteins from all hymenopteran genomes available on NCBI (Supplementary Table 6) and UniProt Hymenoptera, (ii) the mapping of RNA-Seq data from *C. congregata*, *C. glomerata*, *C. vestalis* and *C. rubecula*[63,64], and (iii) *ab initio* genes predictions using SNAP [65]. The automated annotation of the five other *Cotesia* species was performed using MAKER [66] using the same features as for the annotation of *C. congregata* but also including the output automated annotation of *C. congregata*.

Automated gene functional annotation

The functional annotation was performed using blastp from BLAST+ v2.5.0[67] to compare the *C. congregata* proteins to the NCBI non-redundant database (from the 29/06/2014). The ten best hits below an e-value of 1e-08 without complexity masking were conserved. Interproscan v5.13-52.0 [68] was used to analyze protein sequences seeking for known protein domains in the different databases available in Interproscan. Finally, we used Blast2GO[69] to associate a protein with a GO group.

Specialist gene annotation

The automated annotations were followed by manual curations, corrections and annotations realized by specialists of each gene family of interest through Apollo[70]. The genomes were available to this consortium through the web portal: https://bipaa.genouest.org/is/parwaspdb/.

Genome completeness evaluation

The completeness of the genomes and annotations were evaluated using Benchmarking Universal Single-Copy Orthologs BUSCO v3[61] using the insecta_odb9 database composed of 1658 genes.

Contigs were searched for similarities against the nonredundant NCBI nucleotide (nt) (release November 2019) and the Uniref90 protein (release November 2019) databases using respectively blastn v2.7.1+[67] and diamond v0.9.29.130[71]. For both tasks, e-value cutoff was set to 10−25. Taxa were inferred according to the highest-scoring matches sum across all hits for each taxonomic rank in the two databases. Sequencing coverage was deduced after mapping Illumina paired reads to the assembly with Bowtie2 v2.3.4.2[58]. Contigs were then displayed with Blobtools v1.1.1[72] using taxon-annotated-GC-coverage plots.

Orthologous genes identification, alignment and phylogeny

Orthologous genes between all genes annotated in the six *Cotesia* species and the four outgroups (*Microplitis demolitor*, *Nasonia vitripennis*, *Apis mellifera* and *Drosophila melanogaster*) were identified using OrthoFinder v1.14[73]. Universal single-copy ortholog genes from BUSCO v3[61] were extracted for the six *Cotesia* species and the four outgroups, aligned using MAFFT v7.017[74] and concatenated. The species phylogeny was performed on this alignment composed of 1,058 orthologous for a length of 611 kb using PhyML program[75] with the HKY85 substitution model, previously determined by jModelTest v2.1[76] and the branch support were measured after 1,000 bootstrap iterations. The cluster of orthologous genes was used to determine the phylogenetic level of each gene represented in Fig. 2. as follows: genes shared by all species are called shared by all; genes shared by at least nine species among the ten studied species without phylogenetic logical are named "shared by some"; genes shared by only Hymenoptera species and without orthologous gene in *D. melanogaster* are considered as "Hymenoptera specific"; genes shared only by Microgastrinae are named "Microgastrinae specific"; genes shared only by *Cotesia* species and without orthologous genes in any of the outgroup are considered as "Cotesia specific".

Synteny analyses

The different synteny analyses were performed on the orthologous genes identified by OrthoFinder v1.14[73] and by reciprocal blastp from BLAST+ v2.2.28[67] on the annotated proteins

(e-value below $10e^{-20}$). The correspondence between the genes, the localizations on the scaffold and the figures were realized thanks to a custom R script (R Core Team 2013).

Evolution of gene families

For OR, P450 and odv-e66 genes manually annotated genes from the reference genome of *C. congregata* were used along with orthologous genes from the five other *Cotesia* species, *M. demolitor*[77], *N. vitripennis*[78], *A. mellifera*[79] to create a phylogeny of each family among Hymenoptera. Protein sequences were first aligned with MAFFT v7.017[74] and the maximum-likelihood phylogeny was performed with PhyML[75] using the JTT+G+F substitution model for OR and using HKY80 substitution model for P450 and odv-e66 genes. The branch support was assess using aLRTfor OR and 1,000 bootstraps for P450 and odv-e66 genes.. The trees were then rooted to Orco (OR-coreceptor) clade for OR and t the midpoint for the other. The gene gains and losses along the phylogeny for the different gene families of interest were identified with NOTUNG v2.9 [80 81].

Evolution of single copy genes

To determine evolutionary rates within *Cotesia* genus, single-copy orthologous gene clusters (BUSCO, nudiviral and virulence genes) were first aligned using MACSE[82] to produce reliable codon-based alignments. From these codon alignments, pairwise dN/dS values were estimated between *C. congregata* and *C. sesamiae*, the two most diverging species in the *Cotesia* phylogeny, with PAML v4.8 using the YN00 program[83]. dN/dS of the different gene categories of interest were then compared using a Kruskal–Wallis test, and Nemenyi-Tests for multiple comparisons were realized with the R package. For the nudiviral genes the dN/dS values were calculated using genes from the 6 species. Orthologous genes from the six *Cotesia* species were aligned as described before and codeml (implemented in PAML v4.8) was used to estimate the M0 dN/dS (free ratio model). This model was compared to a neutral model for which the dN/dS is fixed to 1.

RNA-Seq analyses

Sample preparation, extraction and sequencing

The ovaries and venom glands were extract from females at five pupal stages, i.e. days 2, days 3, days 4, days 5 and at emergence, corresponding to the number of days after the creation of the cocoon and identified following body melanization[84]. Ovaries were pooled by groups of 20 pairs and venom glands by 100 and in duplicates for each condition. Samples were stored in buffer provided in the extraction kit by adding 2-mercaptoethanol to reduce RNA degradation. Extractions were performed using QIAGEN RNeasy kit following manufacturer's recommendations. Library preparation and sequencing were performed at the Genescope using a paired-end strategy with HiSeq2000 platform (Illumina).

Analyses

The pair-end reads from *C. congregata* ovary and venom gland libraries were mapped on the reference genome using TopHat2[85] with default parameters. Then, featureCounts program from the Subread package[86] was used to determine fragment counts per genes using default parameters.

To analyze gene expression, the raw fragment counts of ovaries and venom glands samples were first converted to counts per million (CPM) using the edgeR[87] R-implemented package (R-Core Team 2017). Expressed genes were filtered based on a CPM > 0.4 (corresponding to raw count of 15) in at least two of the libraries incorporated in the analysis (Supplementary Fig. 6 A) and subsequent normalization was performed on CPMs using the edgeR TMM method for Normalization Factor calculation[88] (Supplementary Fig. 6 B). The reproducibility of replicates was then assessed by Spearman correlation of gene expression profiles based on filtered and normalized CPMs (Supplementary Fig. 6 C).

To examine differential expression between ovary stages and with venom glands a quasi-likelihood negative binomial generalized log-linear model was fitted to the data after estimation of the common dispersion using edgeR. Then, empirical Bayes quasi-likelihood F-tests were performed to identify differentially expressed (DE) genes under chosen contrasts[89]. Finally, F-test p-values were adjusted using False Discovery Rate (FDR) method[90]. Genes were considered as DE whether FDR < 0.05 and fold-change (FC) of expressions between compared conditions was higher or equal to 1.5. Four contrasts were designed between the five successive ovary stages and a control contrast was tested between ovaries and venom glands at wasp emergence stage.

Data availability

509 The datasets generated during the current study are available from the National Center for
510 Biotechnology Information (NCBI) at the following BioProject ID: PRJEB36310 for genome
511 raw reads; PRJNA594477 for transcriptome raw reads. Genome database (genomes and
512 annotated genes) are available on the web site BIPAA (Bioinformatic Platform for Agrosystem
513 Arthropods) https://bipaa.genouest.org/is/parwaspdb/. Custom scripts are available from
514 https://github.com/JeremyLGauthier/Script_Gauthier_et_al._2020_NEE.

515

516

517 **Figure Legends**

518 **Fig. 1. A.** Major traits involved in the parasitoid koinobiont lifestyle and genome content of six
519 *Cotesia* species. First OLFACTION plays an important role in the detection of the plant
520 (tobacco) attacked by caterpillars and host (*M. sexta)* larvae acceptance by adult wasps. Once
521 the host is accepted, the wasp injects its eggs bathed in ovarian fluid filled with bracovirus
522 particles. Bracovirus particles infect host cells, from which expression of bracovirus
523 VIRULENCE GENES alter host immune defenses, allowing wasp larvae development (the
524 eggs laid in the host body would otherwise be engulfed in a cellular sheath of hemocytes). As
525 the host ingests plant toxic compounds, such as nicotine, while feeding, wasp larvae consuming
526 the hemolymph containing these compounds rely on DETOXIFICATION to complete their life
527 cycle. But in these species associated with endogenous viruses the most important trait for
528 parasitism success consists in BRACOVIRUS MORPHOGENESIS during wasp
529 metamorphosis, using genes originating from a nudivirus ancestrally integrated in the wasp
530 genome. As massive production of virus particles occurs within wasp ovaries, WASP
531 IMMUNITY may be induced during particles production; d2, d3, d4, d5 refer to developmental
532 stages of *C. congregata* larvae[84]. **B** Pictures of the six *Cotesia* species sequenced (credit H. M.
533 Smid and R. Copeland). **C** Phylogeny of these species based on 1,058 single-copy orthologous
534 insect genes including the Microgastrinae *Microplitis demolitor* and outgroups (*N. vitripennis*,
535 *A. mellifera*, and *D. melanogaster*). Black dots highlight branches with at least 90% support
536 from maximum-likelihood analysis (1,000 bootstraps). **D** Distribution of shared genes at
537 several phylogenetic levels. Full protein-coding gene sets were included to identify orthologous
538 gene groups. The “shared by some” category refers to genes shared by at least nine species
539 among the 10 studied. Note that the lower number of genes for *C. congregata* probably reflects
540 the higher quality of the genome assembly obtained.

Fig. 2. Gene family extensions in *Cotesia.* **A** Maximum-likelihood phylogeny of the OR family in *C. congregata* and four other Hymenoptera species. The dataset included 243 amino acid sequences from *C. congregata* (blue), 203 sequences from *M. demolitor* (red), 216 sequences from *N. vitripennis* (orange), 162 sequences from *A. mellifera* (yellow). The tree was rooted using the Orco (OR-coreceptor) clade. Circles indicate nodes strongly supported by the approximate likelihood-ratio test (black circles aLRT ≥ 0.95; white circles 0.90 ≥ aLRT ≤ 0.95). The scale bar represents 0.5 expected amino acid substitutions per site. ORs of the five Hymenoptera species are distributed into 18 OR subfamilies previously described in [23] delineated in grey. **B** Copy number dynamics of OR (olfaction) P450 (detoxification) and Odv-e66 genes, note that the later are found specifically in bracovirus associated wasps since they derive from the ancestrally integrated nudivirus. Estimated numbers of gene gain and loss events are shown on each branch of the species tree. The size of OR repertoires in common ancestors is indicated in the boxes. The lack of phylogenetic resolution for closely related Cotesia OR genes precluded any comprehensive analysis of gene gains and losses.

Fig. 3. Synteny of nudiviral genes loci and Proviral Loci (PL) between *C. congregata* and *M. demolitor*. **A** *C. congregata* chromosome map with the position of 24 nudiviral genes loci. **B** Comparisons between nudiviral gene regions of *C. congregata* and *M. demolitor*. Synteny between the two species has been characterized by at least two hymenopteran (non-viral) orthologous genes in the vicinity of homologous nudiviral gene(s) of both species. Genome scaffolds are represented in black. Red boxes indicate nudiviral genes and white boxes refer to hymenopteran genes. 1. the *vp91* region is orthologous indicating the position of this gene was inherited from their common ancestor 53 mya; 2.: the fen region is also orthologous but an expansion occurred specifically in *Cotesia* lineage giving rise to six copies; 3.: the organization of the nudiviral cluster encoding in particular capsid genes has remained strikingly similar with the same viral genes in the same order (except p6-9-2) in both species indicating strong evolutionary constraints. **C** *C. congregata* chromosome map with the position of gene loci corresponding to the highly expanded *odv-e66* nudiviral gene family. **D** Comparison of two odv-e66 loci showing that expansion occurred before (cluster 7) and after (cluster 4) the separation of both species **E** *C. congregata* chromosome map with the position of Proviral Loci (PL) encoding virulence genes packaged in bracovirus particles. Note the concentration of loci (successively PL10-PL1-PL2 and PL4) in a 2 Mb region termed "macrolocus" and representing half of the chromosome 5 short arm. **F** Comparison of *C. congregata* and *M. demolitor* PL. Numbers 1 to 37 and letters correspond to the different dsDNA circles present in CcBV and

574 MdBV particles produced from the PL. Blue boxes indicate virulence genes while white boxes
575 refer to hymenopteran genes and the red box to a nudiviral *odv-e66* gene located between PL1
576 and PL2. Ø indicates the absence of orthologs PL in the *M. demolitor* genome.

577 **Fig. 4. A** *C. congregata* chromosome map with the location of bracovirus loci: nudiviral gene
578 loci are shown in red, nudiviral *odv-e66* gene loci in hatched red and Proviral Loci (PL) in blue.
579 The sizes of the circles correspond to the relative number of genes in each locus **B** taxon-
580 annotated GC content-coverage plot of *C. congregata* scaffolds. Each circle represents a
581 scaffold in the assembly, scaled by length, and colored by taxonomy assignation to Braconidae
582 and Polydnaviridae. The x-axis corresponds to the average GC content of each scaffold and the
583 y-axis corresponds to the average coverage based on alignment of illumina reads. **C** Measure
584 of selection pressure on hymenopteran conserved genes, nudiviral genes and virulence genes.
585 Pairwise evolutionary rates (dN/dS) of single-copy orthologous BUSCO genes, nudiviral genes,
586 different copies from the expanded odv-e66 nudiviral gene family and virulence genes of *C.*
587 *congregata* and *C. sesamiae*. Letters above boxes indicate significant differences determined
588 by Kruskal–Wallis test ($H= 296.8$, 2 d.f., $P<0.001$) followed by *post-hoc* comparisons. **C**
589 Schematic representation of the genomic amplification during the production of viral particles
590 in the wasp ovaries. Replication Unit Motifs (RUM) are the motifs that constitute the
591 extremities of the molecules amplified during particle production. Direct Repeat Junctions
592 (DRJ), at the extremities of each segment are used during the excision/circularization process
593 to produce packaged dsDNA circles from the amplified molecules. Host Integration Motifs
594 (HIM) are motifs used during the integration of bracovirus circles in host genome. For each of
595 these motifs an alignment of a representative set of sequence comprising five motifs from *C.*
596 *congregata* and *M. demolitor* are represented (all alignments are reported in Supplementary
597 Fig. 10).

598 **Fig. 5. Gene expression of nudiviral genes in the ovaries during *C. congregata* nymphal**
599 **development**. **A** Pictures of wasp developmental stages studied and characteristic electron
600 micrographs of ovarian cells involved in particle production. From d2 to d4 stage, cells that
601 will produce particles show enlarged nuclei with chromatin condensation (left panel). From d5
602 massive particle production begins, particle assembly occurs at the periphery of a zone of
603 electron dense material in the nucleus named "virogenic stroma" (middle panel). In newly
604 emerged wasp nuclei are completely filled with bracovirus particles (right panel). Credit: Juline
605 Herbinière. **B** Unsupervised hierarchical clustering based on gene expression in ovaries and

venom glands. Ov2, Ov3, Ov4, Ov5 refer to the different ovaries across wasp developmental stages. Ove and vg refer to ovaries and venom glands of adult wasps. The colored squares associated with the clustering tree indicate the viral functions to which different nudiviral genes are supposed to contribute based on those of their baculovirus homologues. Heatmap of expression levels of 95 nudiviral genes is shown in the middle panel. Bold names highlight the genes that are validated as significantly differentially expressed between two consecutive stages using the statistical analysis and dots represent the four different comparisons studied between ovary stages (Ov2 *vs.* Ov3, Ov3 *vs*. Ov4, Ov4 *vs*. Ov5 and Ov5 *vs.* Ove). Black, red and green dots indicate similar, increased and reduced expressions between consecutive developmental stages respectively. The increase of some nudivirus genes expression between d2 and d3 visualized on the heat map was not validated statistically for all of them because in one of the d2 duplicates (shown in C) nudiviral genes expression had already reached high levels. Underlined genes show higher expression in venom glands compared to ovaries (Ove) note that 27a and 17b are not nudiviral genes but wasp genes, the products of which have been identified in Chelonus inanitus bracovirus particles. **C** Unsupervised hierarchical clustering of gene expression from the two replicates of Ov2 ovary stage, that are very different regarding nudiviral gene expression levels, although dissected nymphae presented a similar coloration pattern, the left one representing a slightly earlier stage from the analysis of the whole set of wasp genes. Note that the genes within the box are already expressed at a high level in the earlier stage, including all of the genes involved in nudiviral transcription (shown in green) except lef5, in accordance with the hypothesis that the nudiviral RNA polymerase complex controls the expression of the other genes (lef-5 is associated to the complex but not a part of it).

Fig. 6. Gene expression of immune genes in the ovaries during *C. congregata* nymphal development. Gene expression of antiviral immunity genes during *C. congregata* development. The heatmaps show the expression of genes involved in **A** RNAi, **B** Imd, **C** Toll and **D** Jak-STAT pathways across the developmental stages of ovaries (Ov2, Ov3, Ov4, Ov5, Ove) and in venom glands (vg). The trees on the left are unsupervised hierarchical clustering of expression values. Boxplots represent overall expression of each pathway in ovaries and venom glands. Bold names highlight the genes that are differentially expressed between two stages and dots represent the four different comparisons studied between consecutive ovary stages (Ov2 *vs.* Ov3, Ov3 *vs*. Ov4, Ov4 *vs*. Ov5 and Ov5 *vs.* Ove). Black, red and green dots indicate similar, increased and reduced expressions between consecutive developmental stages

639 respectively. Note that no particular trend appears correlated to bracovirus particles production
640 which occurs massively from Ov5 onward.

641

642 **Acknowledgements**

643 We thank Paul André Catalayud for providing the pictures of *C. sesamiae* and *C. flavipes*
644 Germain Chevignon for the pictures of *C. congregata* nymphal stages and Juline Herbinière for
645 TEM images of *Cotesia congregata* ovaries. We thank the ADALEP (Adaptation of
646 Lepidoptera) network for the involvement of its members and access to bioinformatic facilities
647 for genome annotation.

648

649 **Author information**

650 **Affiliations**

651 1. UMR 7261 CNRS, Institut de Recherche sur la Biologie de l'Insecte, Faculté des Sciences
652 et Techniques, Université de Tours, France
653 2. Geneva Natural History Museum, 1208 Geneva, Switzerland
654 3. EAWAG, Swiss Federal Institute of Aquatic Science and Technology, Dübendorf,
655 Switzerland
656 4. Department of Terrestrial Ecology, Netherlands Institute of Ecology (NIOO-KNAW),
657 Wageningen, The Netherlands
658 5. Institut Pasteur, Unité Régulation Spatiale des Génomes, UMR 3525, CNRS, Paris, F-
659 75015, France
660 6. Sorbonne Université, Collège Doctoral, Paris, F-75005 France
661 7. Sorbonne Universités, UPMC University Paris 06, Institute of Ecology and Environmental
662 Sciences of Paris, Paris, 75005, France
663 8. Genoscope, Institut de biologie François-Jacob, Commissariat à l'Energie Atomique (CEA),
664 Université Paris-Saclay, Evry, , F-91057, France
665 9. IGEPP, Agrocampus Ouest, INRA, Université de Rennes, 35650 Le Rheu, France
666 10. Université de Rennes 1, INRIA, CNRS, IRISA, 35000, Rennes, France
667 11. Applied Bioinformatics, Wageningen University & Research, Wageningen, The
668 Netherlands
669 12. DGIMI, Univ Montpellier, INRA, Montpellier, France

13. Laboratoire de Biométrie et Biologie Evolutive, Université de Lyon, Claude Bernard University Lyon 1, CNRS, UMR 5558, 69100 Villeurbanne, France

14. UMR EGCE (Evolution, Génome, Comportement, Ecologie), CNRS-IRD-Univ. Paris-Sud, Université Paris-Saclay, Gif-sur-Yvette Cedex, France

15. Université Côte d'Azur, INRA, CNRS, Institut Sophia Agrobiotech, 06903 Sophia-Antipolis, France

16. URGI, INRA, Université Paris-Saclay, 78026, Versailles, France

17. Laboratory of Entomology, Wageningen University, Wageningen, The Netherlands

18. Insect Interactions Laboratory, Department of Entomology and Acarology, Luiz de Queiroz College of Agriculture (ESALQ), University of São Paulo, São Paulo, Brazil

19. Department of Entomology, University of Illinois, Urbana, IL 61801, USA

20. Evolutionary Genetics, Centre for Ecological and Evolutionary Studies, University of Groningen, Groningen, The Netherlands

Contributions

DNA/RNA preparation: J. Gauthier, C. Capdevielle-Dulac, K. Labadie, H. Boulain; F. Consoli, B. L. Merlin; **Sequencing assembly and automatic annotation:** B. Noël, J.-M. Aury, V. Barbe, J. Gauthier, A. Bretaudeau; F. Legeai; ***C. glomerata*, *C. rubecula*, *C. vestalis* sequencing and first assembly:** J. Van Vugt, H. Smid, J. de Boer, S.Warris, L. Vet; **HiC approach:** L. Baudry, M. Marbouty, R. Koszul; **TEs annotation:** A. Hua-Van (group leader), J. Amselem, I. Luyten; **Bracovirus annotation:** A. Bézier (group leader), J. Gauthier, P. Gayral, K. Musset, T. Josse, D. Bigot, C. Bressac, S. Moreau, E.A. Herniou; **Bracovirus gene evolution and synteny analyses:** J. Gauthier; H. Boulain; **Immune genes annotation:** E. Huguet (group leader), H. Boulain; G. Dubreuil; B. Duvic, J.-M. Escoubas, N. Kremer; **Conserved bracovirus regulatory sequences:** G. Periquet; **Chemosensory genes annotation:** E. Jacquin-Joly (group leader), M. Harry, E. Persyn, N. Montagné, I. Boulogne, M. Sabety, M. Maibeche, T. Chertemps; **Detoxification genes annotation:** G. le Goff (group leader), F. Hilliou, D. Saussiat; **Web annotation online platform:** A. Bretaudeau; F. Legeai; **RNA-Seq analyses**: H. Boulain, J. Gauthier; **Wasp phylogenetic and biological background:** J. B. Whitfield; **Project writing for funding**: S. Dupas, C. L. Kaiser-Arnault, E. Herniou, J.-M. Drezen, H. Smid, L. Vet; **Project coordination:** J. Gauthier and J.-M. Drezen; **Manuscript writing:** J. Gauthier, J.-M. Drezen with contributions from all authors.

Corresponding author

Correspondence to *drezen@univ-tours.fr*

Funding

C. congregata, *C. sesamiae*, *C. flavipes* genomes sequencing were funded by French National Research Agency ANR (ABC Papogen project ANR-12-ADAP-0001 to L. Kaiser-Arnaud). *C. rubecula*, *C. glomerata*, *C. vestalis* genomes sequencing were funded by NWO EcoGenomics grant 844.10.002 to L.E.M. Vet, NWO VENI grant 863.07.010 and Enabling Technology Platform Hotel grant to L.E.M. Vet. HiC approach was funded by ERC project 260822 to R. Koszul. *C. congregata* transcriptomic analysis was funded by APEGE project (CNRS-INEE) to J.-M. Drezen. J. Gauthier thesis was funded by ANR and Region Centre-Val de Loire. Collaboration between French and Netherland laboratories was funded by French ministry of foreign affairs and "Nuffic" ("VanGogh" Project to J-M Drezen and L.E.M. Vet).

Ethics declarations

Competing interests

The authors declare no competing interests.

Additional information

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary information

Cotesia_genomes_Supplementary_data.pdf

1. Parra, J.R.P. Biological Control in Brazil: An overview. *Scientia Agricola* **71**, 420-429 (2014).
2. Parra, J.R.P. & Coelho, A. Applied Biological Control in Brazil: From Laboratory Assays to Field Application *Journal of Insect Science* **19**, 1-6 (2019).
3. Dicke, M. Behavioural and community ecology of plants that cry for help. *Plant, Cell and Environment* **32**, 654-666 (2009).
4. Poelman, E.H. *et al.* Hyperparasitoids use herbivore-induced plant volatiles to locate their parasitoid host. *PLoS Biol* **10**, e1001435 (2012).
5. Rodriguez, J.J. *et al.* Extrapolations from field studies and known faunas converge on dramatically increased estimates of global microgastrine parasitoid wasp species richness (Hymenoptera: Braconidae). *Insect Conservation and Diversity* **6**, 530–536 (2013).
6. Bézier, A. *et al.* Polydnaviruses of braconid wasps derive from an ancestral nudivirus. *Science* **323**, 926-30 (2009).

7. Drezen, J.-M., Herniou, E.A. & Bézier, A. Evolutionary progenitors of bracoviruses. in *Parasitoid viruses symbionts and pathogens* (eds. Beckage, N.E. & Drezen, J.-M.) 15-31 (Elsevier, San Diego, 2012).
8. Burke, G.R., Thomas, S.A., Eum, J.H. & Strand, M.R. Mutualistic polydnaviruses share essential replication gene functions with pathogenic ancestors. *PLoS Pathog* **9**, e1003348 (2013).
9. Thézé, J., Bézier, A., Periquet, G., Drezen, J.M. & Herniou, E.A. Paleozoic origin of insect large dsDNA viruses. *Proc Natl Acad Sci U S A* **108**, 15931-5 (2011).
10. Gauthier, J., Drezen, J.M. & Herniou, E.A. The recurrent domestication of viruses: major evolutionary transitions in parasitic wasps. *Parasitology* **145**, 713-723 (2018).
11. Beckage, N.E., Tan, F., Schleifer, K.W., Lane, R.D. & Cherubin, L.L. Characterization and biological effects of *Cotesia congregata* polydnavirus on host larvae of the tobacco hornworm, *Manduca sexta*. *Arch. Insect Biochem. Physiol.* **26**, 165-195 (1994).
12. Strand, M.R. Polydnavirus gene products that interact with the host immune system. in *Parasitoid viruses symbionts and pathogens* (eds. Beckage, N.E. & Drezen, J.-M.) 149-161 (Elsevier, San Diego, 2012).
13. Bézier, A. *et al.* Functional endogenous viral elements in the genome of the parasitoid wasp Cotesia congregata: insights into the evolutionary dynamics of bracoviruses. *Philos Trans R Soc Lond B Biol Sci* **368**, 20130047 (2013).
14. Burke, G.R., Walden, K.K., Whitfield, J.B., Robertson, H.M. & Strand, M.R. Widespread genome reorganization of an obligate virus mutualist. *PLoS Genet* **10**, e1004660 (2014).
15. Burke, G.R., Walden, K.K.O., Whitfield, J.B., Robertson, H.M. & Strand, M.R. Whole Genome Sequence of the Parasitoid Wasp Microplitis demolitor That Harbors an Endogenous Virus Mutualist. *G3 (Bethesda)* **8**, 2875-2880 (2018).
16. Murphy, N., Banks, J.C., Whitfield, J.B. & Austin, A.D. Phylogeny of the parasitic microgastroid subfamilies (Hymenoptera: Braconidae) based on sequence data from seven genes, with an improved time estimate of the origin of the lineage. *Mol. Phylogenet. Evol.* **47**, 378-95 (2008).
17. Belle, E. *et al.* Visualization of polydnavirus sequences in a parasitoid wasp chromosome. *J Virol* **76**, 5793-6 (2002).
18. Zhang, H.H. *et al.* Unexpected invasion of miniature inverted-repeat transposable elements in viral genomes. *Mob DNA* **9**, 19 (2018).
19. Chevignon, G. *et al.* Functional annotation of Cotesia congregata bracovirus: identification of the viral genes expressed in parasitized host immune tissues. *J Virol* **88**, 8795-8812 (2014).
20. Gauthier, J. *et al.* Genetic footprints of adaptive divergence in the bracovirus of Cotesia sesamiae identified by targeted resequencing. *Mol Ecol* **27**, 2109-2123 (2018).
21. Gundersen-Rindal, D., Dupuy, C., Huguet, E. & Drezen, J.-M. Parasitoid Polydnaviruses: Evolution, Pathology and Applications. *Biocontrol Science and Technology* **23**, 1-61 (2013).
22. Robertson, H.M. Molecular Evolution of the Major Arthropod Chemoreceptor Gene Families. . *Annual Review of Entomology* **64**, 227-242 (2019).
23. Zhou, X. *et al.* Phylogenetic and transcriptomic analysis of chemosensory receptors in a pair of divergent ant species reveals sex-specific signatures of odor coding. *PLoS Genet* **8**, e1002930 (2012).
24. Wang, H. *et al.* CYP6AE gene cluster knockout in Helicoverpa armigera reveals role in detoxification of phytochemicals and insecticides. *Nat Commun* **9**, 4820 (2018).
25. Leobold, M. *et al.* The Domestication of a Large DNA Virus by the Wasp Venturia canescens Involves Targeted Genome Reduction through Pseudogenization. *Genome Biol Evol* **10**, 1745-1764 (2018).

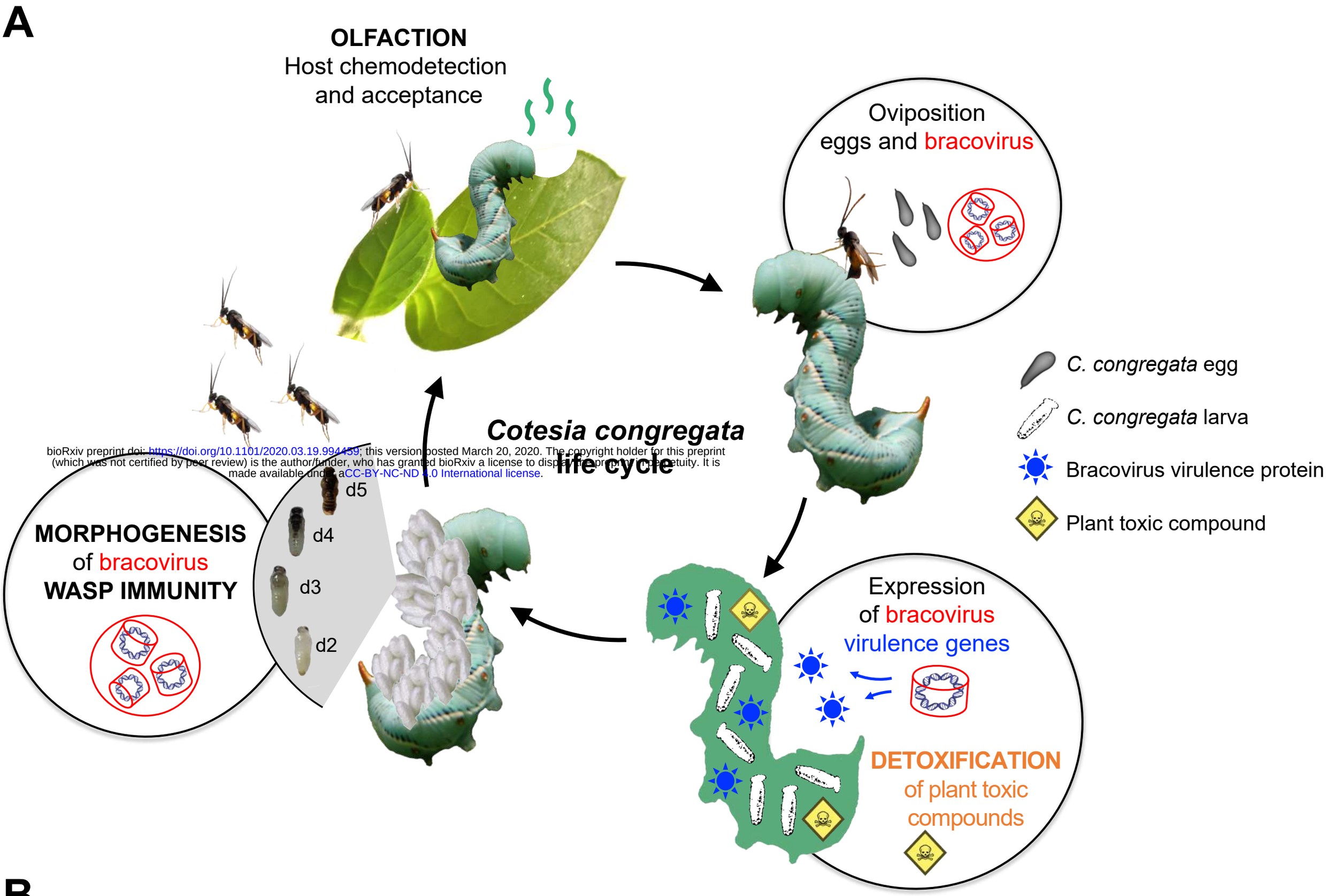
26. Katzourakis, A. & Gifford, R.J. Endogenous viral elements in animal genomes. *PLoS Genet* **6**, e1001191 (2010).
27. Harrison, R.L. *et al.* ICTV Virus Taxonomy Profile: Nudiviridae, Journal of General Virology. *Journal of General Virology*, in press (2020).
28. Burke, G.R., Simmonds, T.J., Sharanowski, B.J. & Geib, S.M. Rapid Viral Symbiogenesis via Changes in Parasitoid Wasp Genome Architecture. *Mol Biol Evol* **35**, 2463-2474 (2018).
29. Bézier, A. *et al.* The genome of the nucleopolyhedrosis-causing virus from *Tipula oleracea* sheds new light on the Nudiviridae family. *J Virol* **89**, 3008-25 (2015).
30. Sugiura, N. *et al.* Chondroitinase from baculovirus Bombyx mori nucleopolyhedrovirus and chondroitin sulfate from silkworm Bombyx mori. *Glycobiology* **23**, 1520-30 (2013).
31. Wyder, S., Blank, F. & Lanzrein, B. Fate of polydnavirus DNA of the egg-larval parasitoid *Chelonus inanitus* in the host *Spodoptera littoralis*. *J. Insect Physiol.* **49**, 491-500 (2003).
32. Beck, M.H., Inman, R.B. & Strand, M.R. Microplitis demolitor bracovirus genome segments vary in abundance and are individually packaged in virions. *Virology* **359**, 179-89 (2007).
33. Louis, F. *et al.* The bracovirus genome of the parasitoid wasp *Cotesia congregata* is amplified within 13 replication units, including sequences not packaged in the particles. *J Virol* **87**, 9649-60 (2013).
34. Trowsdale, J. & Knight, J.C. Major histocompatibility complex genomics and human disease. *Annu Rev Genomics Hum Genet* **14**, 301-23 (2013).
35. Bézier, A. *et al.* Bracovirus gene products are highly divergent from insect proteins. *Arch Insect Biochem Physiol* **67**, 172-87 (2008).
36. Drezen, J.M., Chevignon, G., Louis, F. & Huguet, E. Origin and evolution of symbiotic viruses associated with parasitoid wasps. *Current Opinion in Insect Science* **6**, 35-43 (2014).
37. Desjardins, C.A. *et al.* Comparative genomics of mutualistic viruses of Glyptapanteles parasitic wasps. *Genome Biol* **9**, R183 (2008).
38. Francino, M.P. An adaptive radiation model for the origin of new gene functions. *Nat Genet* **37**, 573-7 (2005).
39. Shi, M. *et al.* The genomes of two parasitic wasps that parasitize the diamondback moth. *BMC Genomics* **20**, 893 (2019).
40. Bichang'a, G. *et al.* Alpha-amylase mediates host acceptance in the Braconid parasitoid *Cotesia flavipes*. *Journal of Chemical Ecology* **44**, 1030-1039 (2018).
41. Pentzold, S. *et al.* Metabolism, excretion and avoidance of cyanogenic glucosides in insects with different feeding specialisations. *Insect Biochem Mol Biol* **66**, 119-28 (2015).
42. Kumar, P., Pandit, S.S., Steppuhn, A. & Baldwin, I.T. Natural history-driven, plant-mediated RNAi-based study reveals CYP6B46's role in a nicotine-mediated antipredator herbivore defense. *Proc Natl Acad Sci U S A* **111**, 1245-52 (2014).
43. Masson, F., Zaidman-Remy, A. & Heddi, A. Antimicrobial peptides and cell processes tracking endosymbiont dynamics. *Philos Trans R Soc Lond B Biol Sci* **371**(2016).
44. Moran, N.A. & Mira, A. The process of genome shrinkage in the obligate symbiont *Buchnera aphidicola*. *Genome Biol* **2**, 1--12 (2001).
45. Herniou, E.A. *et al.* When parasitc wasps hijacked viruses: genomic and functionnal evolution of polydnaviruses. *Phil. Transac. R. Soc. B* **368**, 1-13 (2013).
46. Joron, M. *et al.* Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature* **477**, 203-6 (2011).

47. Wang, J. *et al.* A Y-like social chromosome causes alternative colony organization in fire ants. *Nature* **493**, 664-8 (2013).
48. Wernegreen, J.J. Endosymbiosis: lessons in conflict resolution. *PLoS Biol* **2**, E68 (2004).
49. Douglas, A.E. Conflict, cheats and the persistence of symbioses. *New Phytol* **177**, 849-58 (2008).
50. Harwood, S.H., Grosovsky, A.J., Cowles, E.A., Davis, J.W. & Beckage, N.E. An abundantly expressed hemolymph glycoprotein isolated from newly parasitized *Manduca sexta* larvae is a polydnavirus gene product. *Virology* **205**, 381-392 (1994).
51. Gitau, C.W., Gundersen-Rindal, D., Pedroni, M., Mbugi, P.J. & Dupas, S. Differential expression of the CrV1 haemocyte inactivation-associated polydnavirus gene in the African maize stem borer Busseola fusca (Fuller) parasitized by two biotypes of the endoparasitoid Cotesia sesamiae (Cameron). *J Insect Physiol* **53**, 676-84 (2007).
52. Veiga, A.C.P., Vacari, A.M., Volpe, H.X.L., de Laurentis, V.L. & De Bortoli, S.A. Quality control of Cotesia flavipes (Cameron) (Hymenoptera: Braconidae) from different Brazilian bio-factories. *Biocontrol Science and Technology* **23**, 665-673 (2013).
53. Geervliet, J.B.F., Vet, L.E.M. & Dicke, M. Volatiles from damaged plants as major cues in long-range host-searching by the specialist parasitoid *Cotesia rubecula*. *Entomologia Experimentalis et applicata* **73**, 289-297 (1994).
54. Smid, H.M. *et al.* Species-specific acquisition and consolidation of long-term memory in parasitic wasps. *Proc Biol Sci* **274**, 1539-46 (2007).
55. Margulies, M. *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376-80 (2005).
56. Marcais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764-70 (2011).
57. Marbouty, M. *et al.* Metagenomic chromosome conformation capture (meta3C) unveils the diversity of chromosome organization in microorganisms. *Elife* **3**, e03318 (2014).
58. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-9 (2012).
59. Marie-Nelly, H. *et al.* Filling annotation gaps in yeast genomes using genome-wide contact maps. *Bioinformatics* **30**, 2105-13 (2014).
60. Mikheenko, A., Prjibelski, A., Saveliev, V., Antipov, D. & Gurevich, A. Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics* **34**, i142-i150 (2018).
61. Waterhouse, R.M. *et al.* BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. *Mol Biol Evol* **35**, 543-548 (2018).
62. Flutre, T., Duprat, E., Feuillet, C. & Quesneville, H. Considering transposable element diversification in de novo annotation approaches. *PLoS One* **6**, e16526 (2011).
63. Gao, F. *et al.* Cotesia vestalis teratocytes express a diversity of genes and exhibit novel immune functions in parasitism. *Sci Rep* **6**, 26967 (2016).
64. van Vugt, J.J. *et al.* Differentially expressed genes linked to natural variation in long-term memory formation in Cotesia parasitic wasps. *Front Behav Neurosci* **9**, 255 (2015).
65. Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59 (2004).
66. Cantarel, B.L. *et al.* MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res* **18**, 188-96 (2008).
67. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
68. Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236-40 (2014).

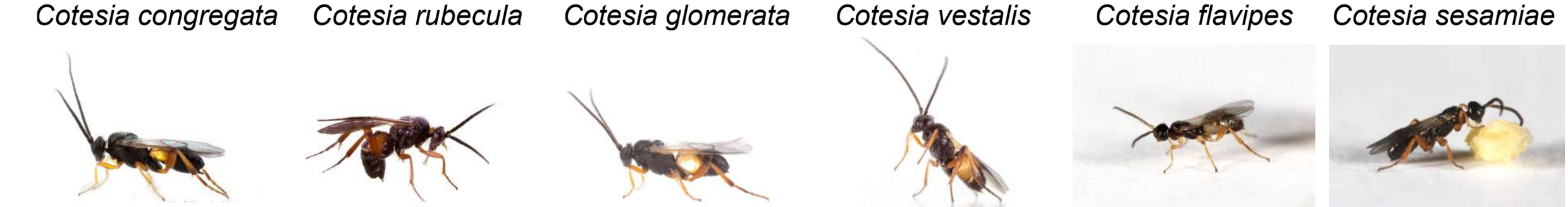
69. Conesa, A. *et al.* Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674-6 (2005).
70. Dunn, N.A. *et al.* Apollo: Democratizing genome annotation. *PLoS Comput Biol* **15**, e1006790 (2019).
71. Buchfink, B., Xie, C. & Huson, D.H. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* **12**, 59-60 (2015).
72. Kumar, S., Jones, M., Koutsovoulos, G., Clarke, M. & Blaxter, M. Blobology: exploring raw genome data for contaminants, symbionts and parasites using taxon-annotated GC-coverage plots. *Front Genet* **4**, 237 (2013).
73. Emms, D.M. & Kelly, S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology* **16**, 157 (2015).
74. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**, 3059-66 (2002).
75. Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**, 696-704 (2003).
76. Darriba, D., Taboada, G.L., Doallo, R. & Posada, D. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* **9**, 772 (2012).
77. Zhou, X. *et al.* Chemoreceptor Evolution in Hymenoptera and Its Implications for the Evolution of Eusociality. *Genome Biol Evol* **7**, 2407-16 (2015).
78. Robertson, H.M., Gadau, J. & Wanner, K.W. The insect chemoreceptor superfamily of the parasitoid jewel wasp Nasonia vitripennis. *Insect Mol Biol* **19 Suppl 1**, 121-36 (2010).
79. Robertson, H.M. & Wanner, K.W. The chemoreceptor superfamily in the honey bee, Apis mellifera: expansion of the odorant, but not gustatory, receptor family. *Genome Res* **16**, 1395-403 (2006).
80. Chen, K., Durand, D. & Farach-Colton, M. NOTUNG: a program for dating gene duplications and optimizing gene family trees. *J Comput Biol* **7**, 429-47 (2000).
81. Stolzer, M. *et al.* Inferring duplications, losses, transfers and incomplete lineage sorting with nonbinary species trees. *Bioinformatics* **28**, i409-i415 (2012).
82. Ranwez, V., Harispe, S., Delsuc, F. & Douzery, E.J. MACSE: Multiple Alignment of Coding SEquences accounting for frameshifts and stop codons. *PLoS One* **6**, e22594 (2011).
83. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**, 1586-91 (2007).
84. Pasquier-Barre, F. *et al.* Polydnavirus replication: the EP1 segment of the parasitoid wasp *Cotesia congregata* is amplified within a larger precursor molecule. *J. Gen. Virol.* **83**, 2035-2045. (2002).
85. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**, R36 (2013).
86. Liao, Y., Smyth, G.K. & Shi, W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res* **41**, e108 (2013).
87. Robinson, M.D., McCarthy, D.J. & Smyth, G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-40 (2010).
88. Robinson, M.D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* **11**, R25 (2010).

939 89. Lund, S.P., Nettleton, D., McCarthy, D.J. & Smyth, G.K. Detecting differential
940 expression in RNA-sequence data using quasi-likelihood with shrunken dispersion
941 estimates. *Stat Appl Genet Mol Biol* **11**(2012).
942 90. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and
943 Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series*
944 *B (Methodological)* **57**, 289-300 (1995).
945

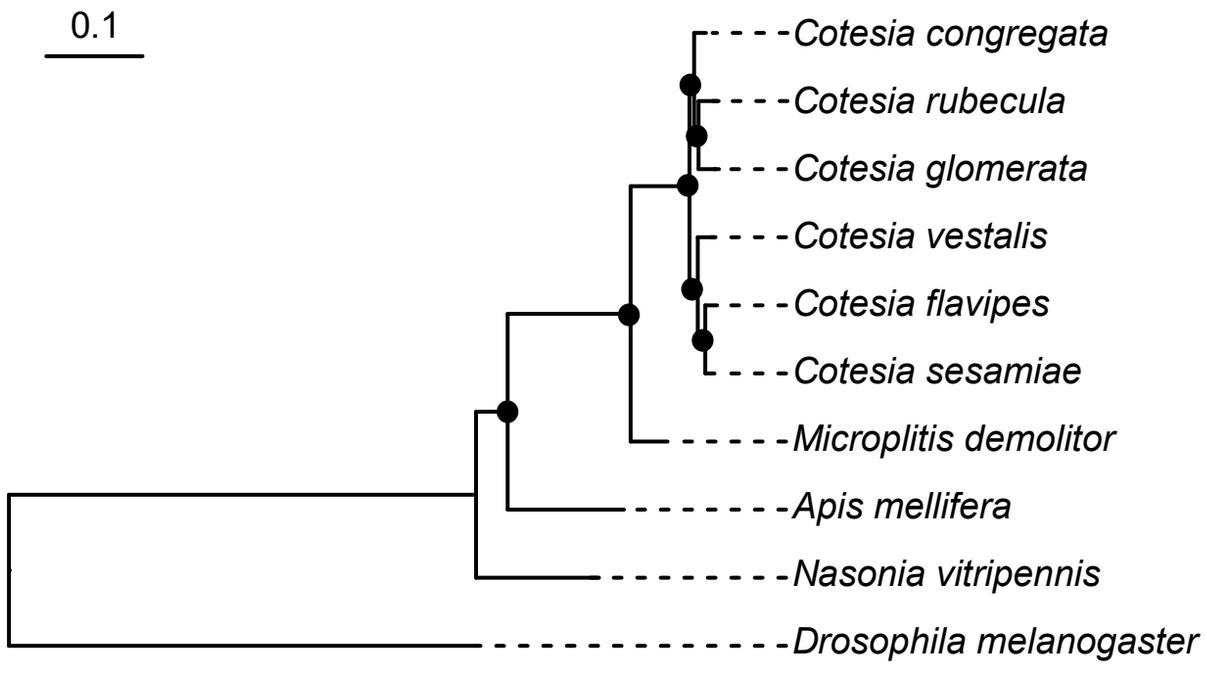
A



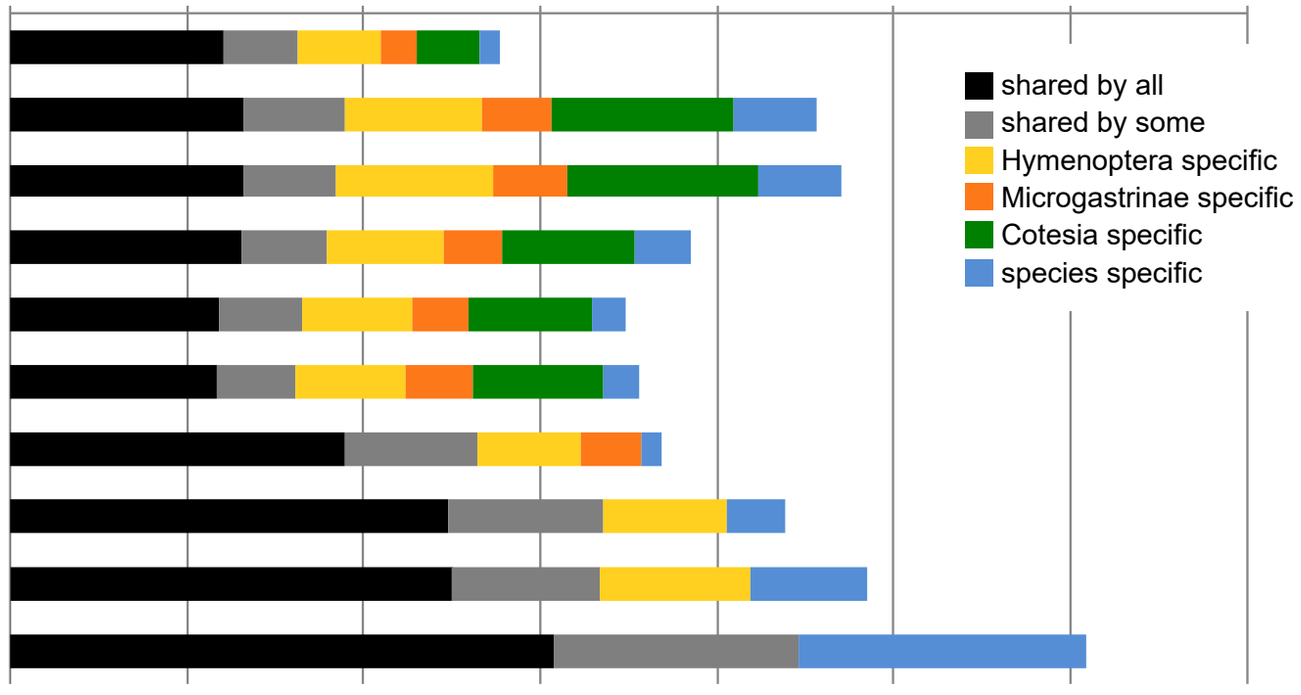
B



C



D



A

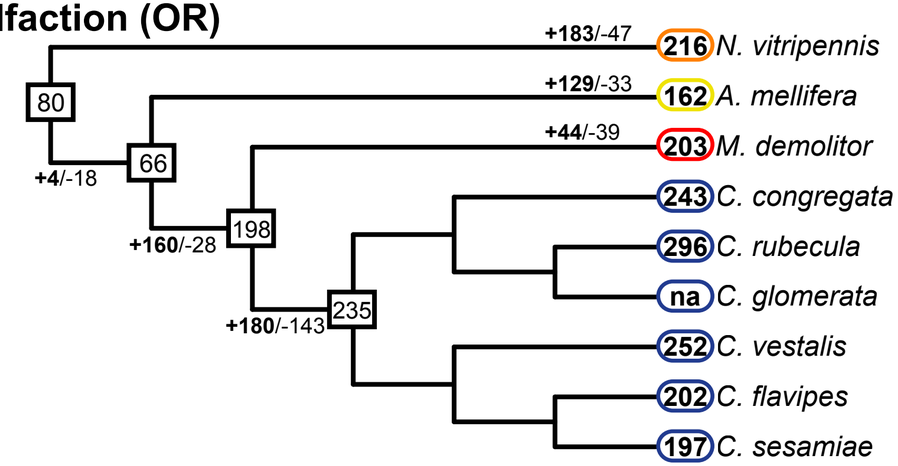
V
TU
MNOP
GHX
QRS
K
L
9-exon
A
B
J
Z
F
ZB
CDE
0.5

Cotesia congregata
Microplitis demolitor
Nasonia vitripennis
Apis mellifera

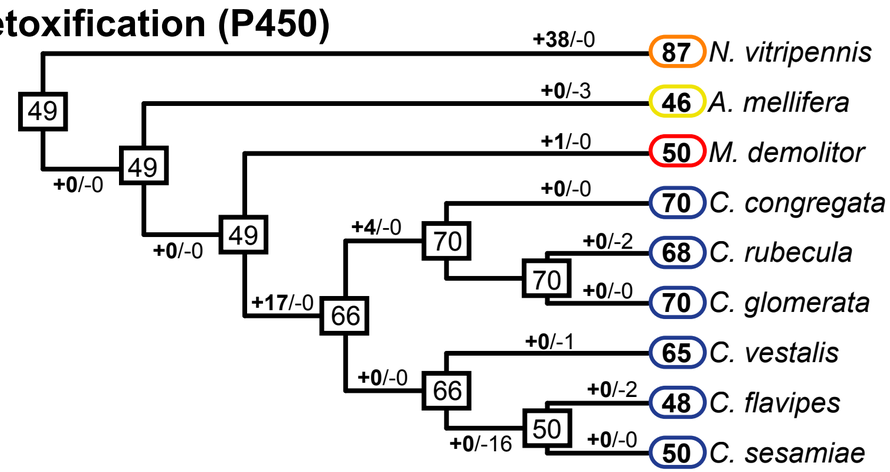
aLRT ≥ 0.95
0.90 ≥ aLRT ≥ 0.95

B

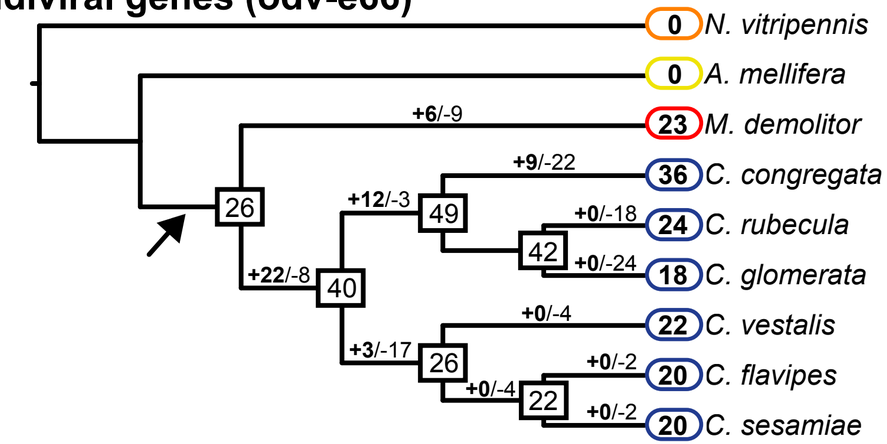
Olfaction (OR)

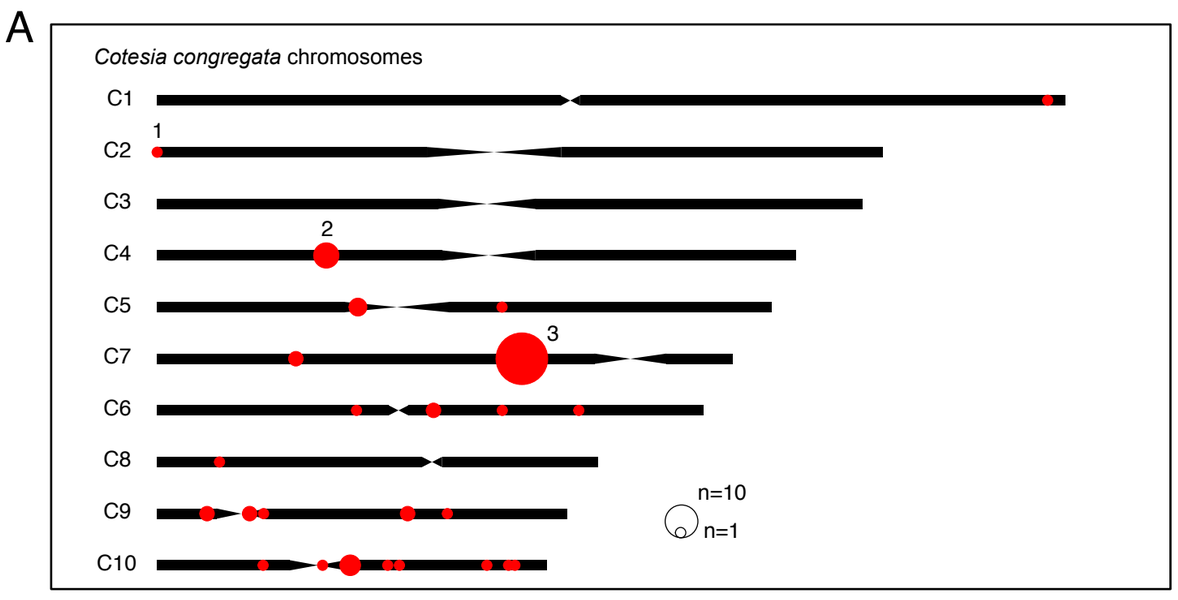


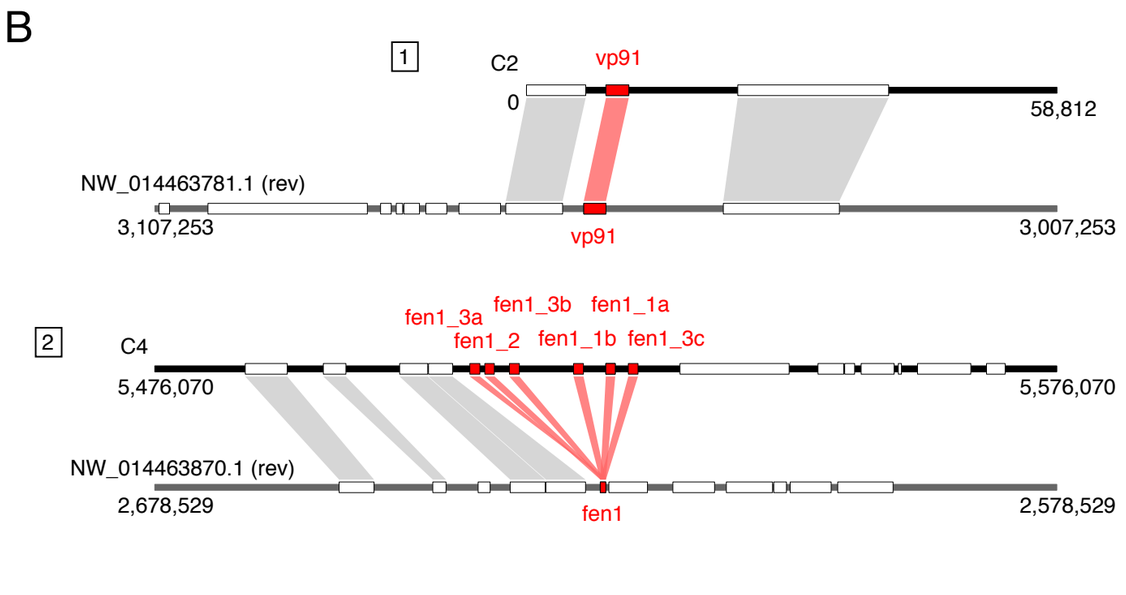
Detoxification (P450)

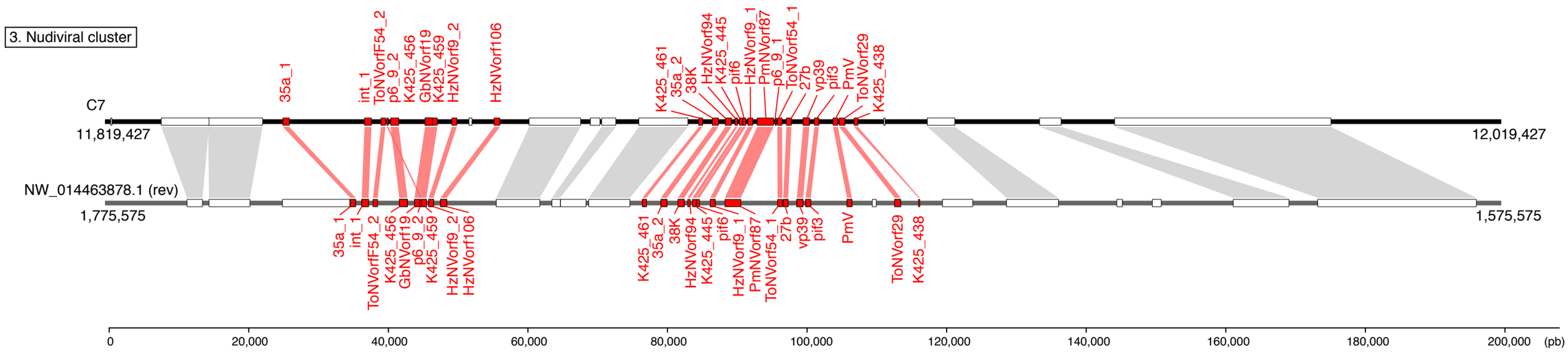


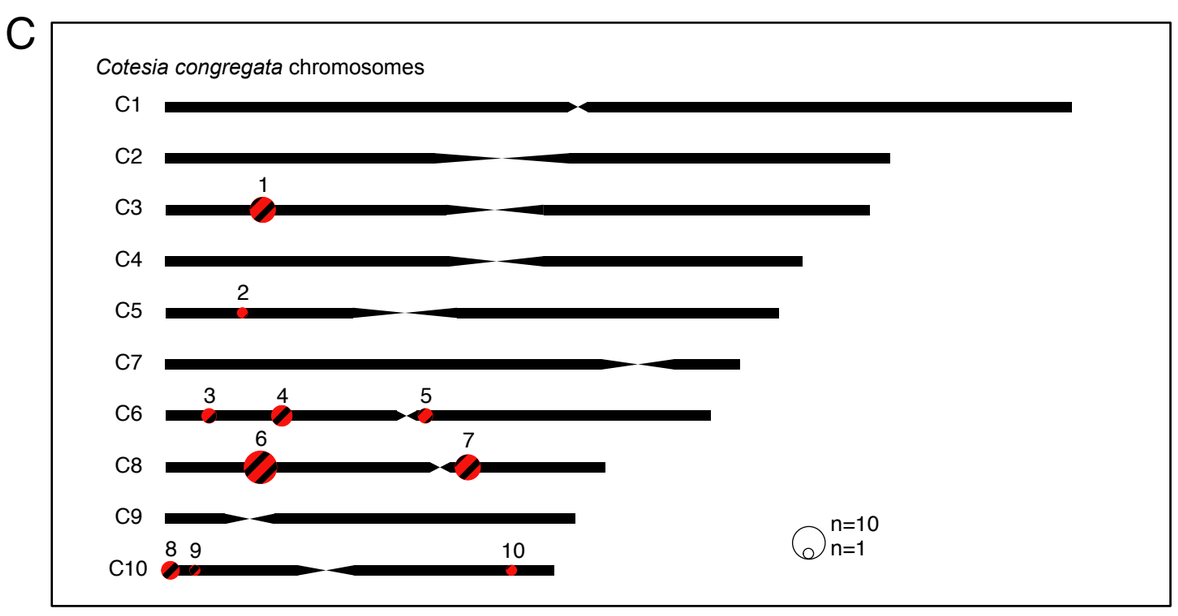
Nudiviral genes (odv-e66)

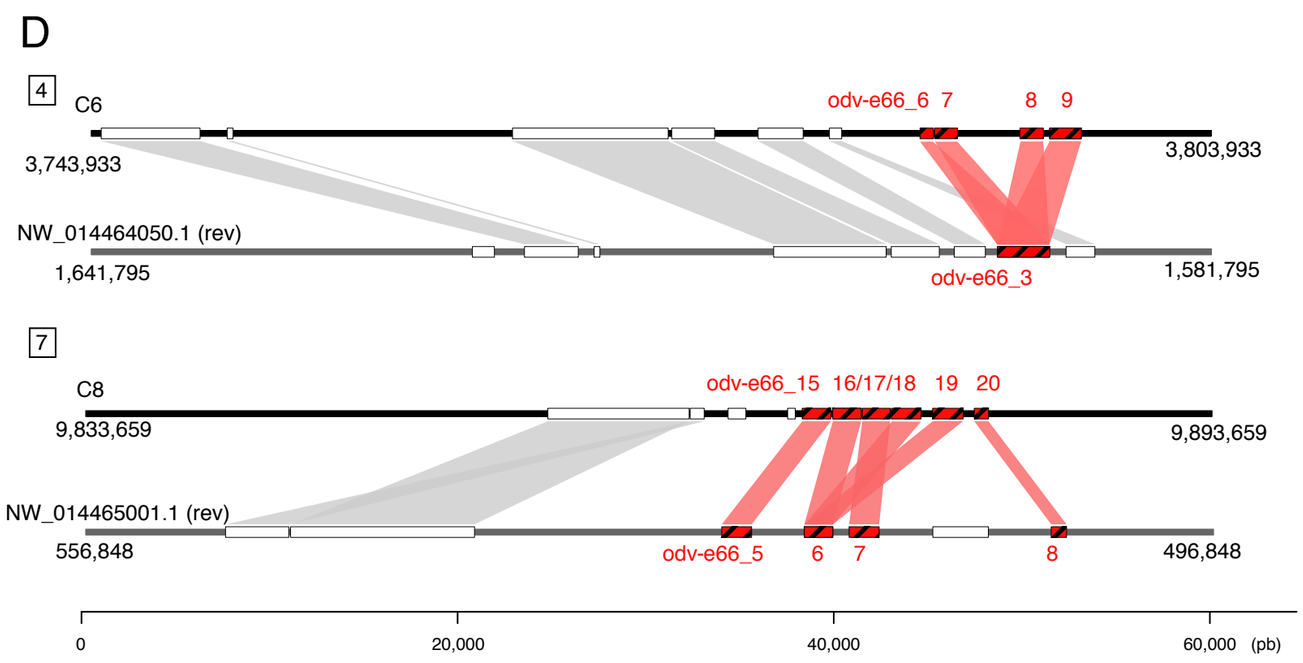


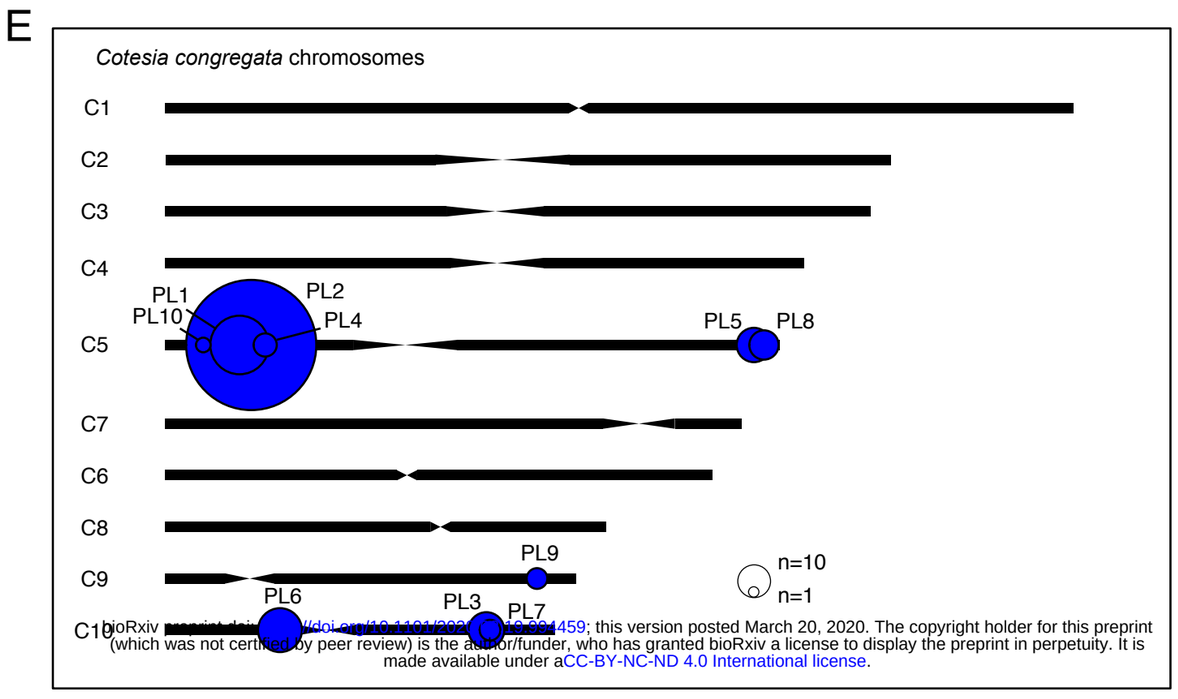


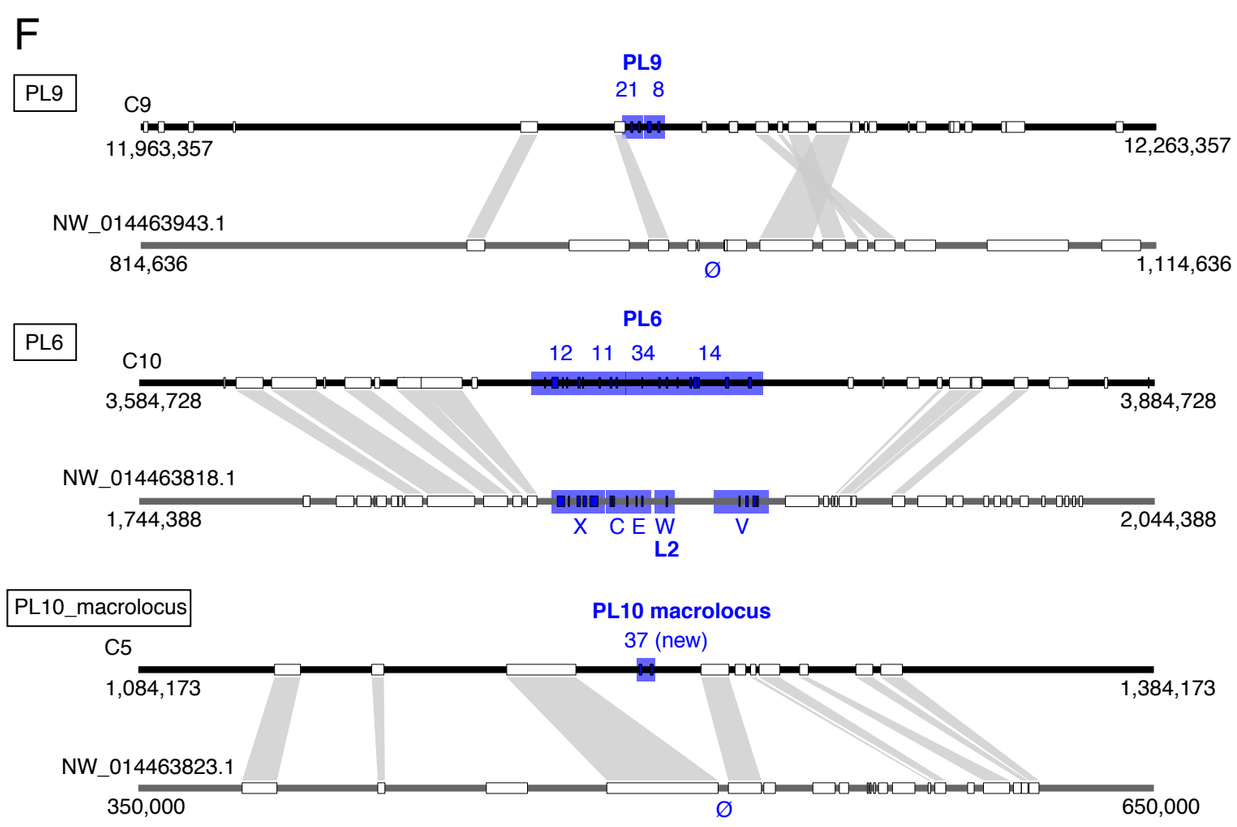


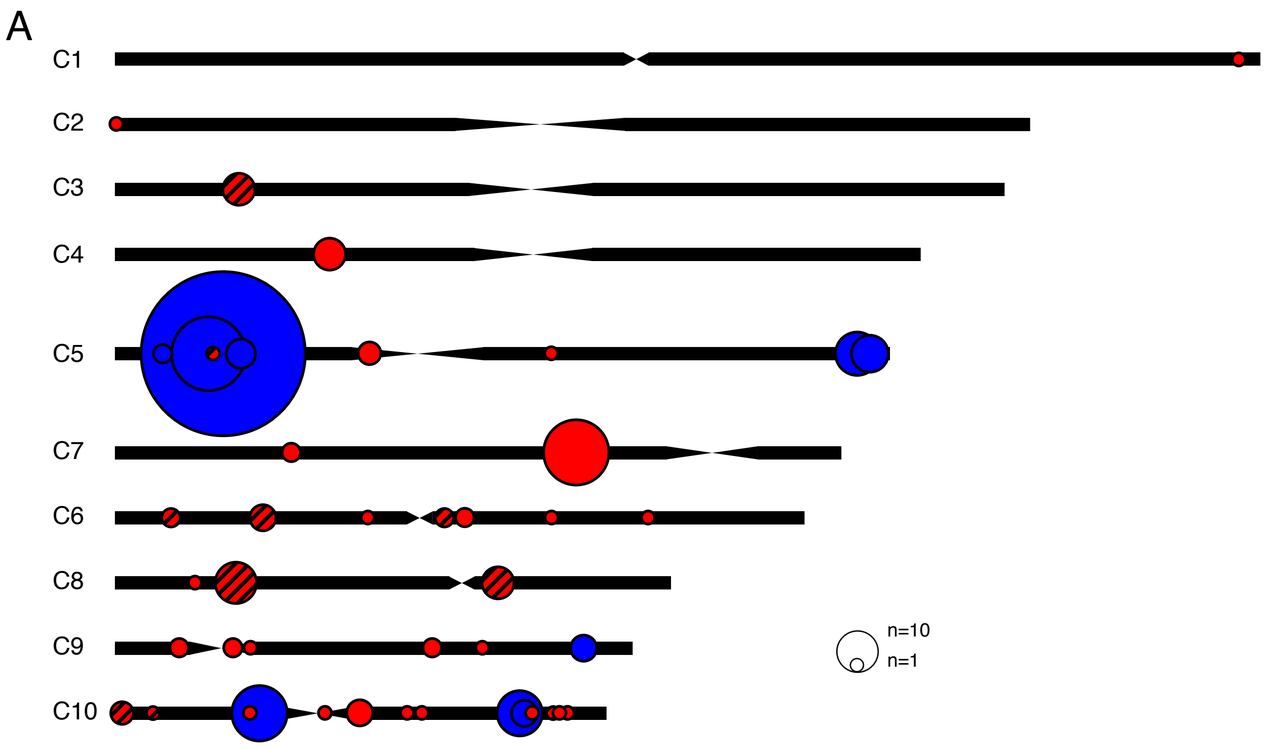


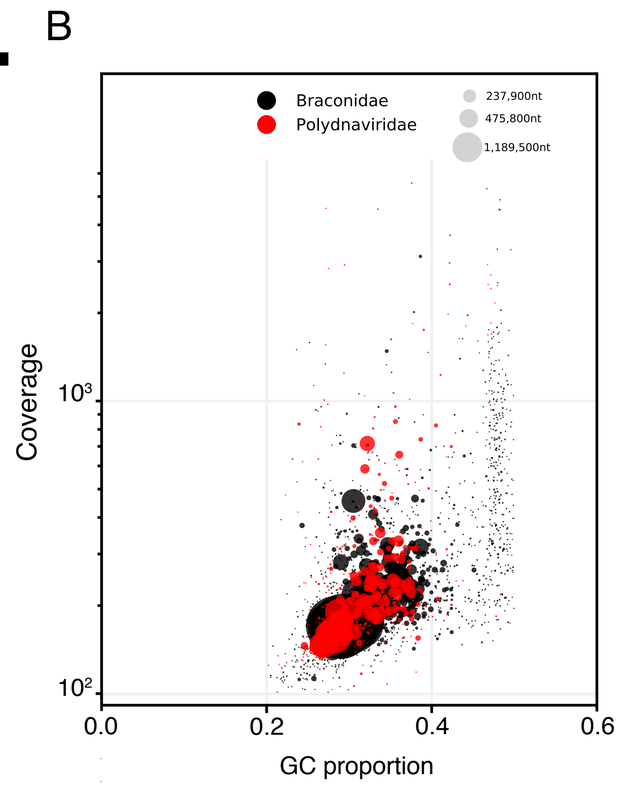


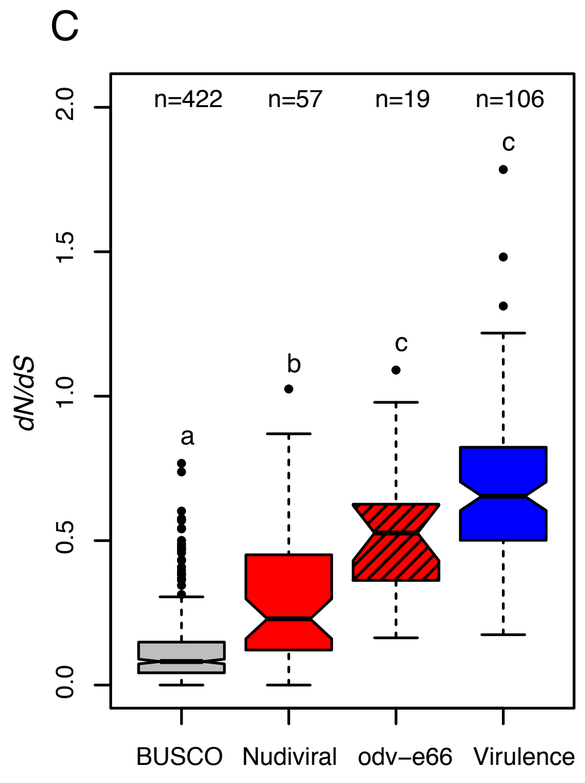












D

Replication Unit Motif Type 1 Head

```
                                10        20        30        40        50        60
CC_RUM_T1_H_LP3_SEG-17   - CGGATTCGGATATGTGGAAACTTACTGTAAGA------ATAATTTATGAATAAATTATATCAC
CC_RUM_T1_H_LP4_SEG-7    - AGAATCCGTTTTAGCGATTACTTACTGTAAACGTGATATAATTTATTCATATTGTATAT-
CC_RUM_T1_H_LP5_SEG-1    - GGCATACGGATATGCGGAAGTTTACTGTAGTAGTGATATAATTTATTCATAAATTATTCTTT
CC_RUM_T1_H_LP8_SEG-26   - TATTCAGATTAACGAGACTTTACTG---TAGCGATATAATTAATTCATAAATTATCCTT-
MD_RUM_T1_H_LP3_SEG-J    - AACATCCGGATACGCGGTATTTACTGTAGTAGTGATATAATTTATTCATAAATTATTCGTT
MD_RUM_T1_H_LP3_SEG-N    - TACATTCGGATATGCGGTACTTTACTGTAGTAGTGATATAATTTATTCATAAGTTACTGATT
```

Replication Unit Motif Type 2 Head

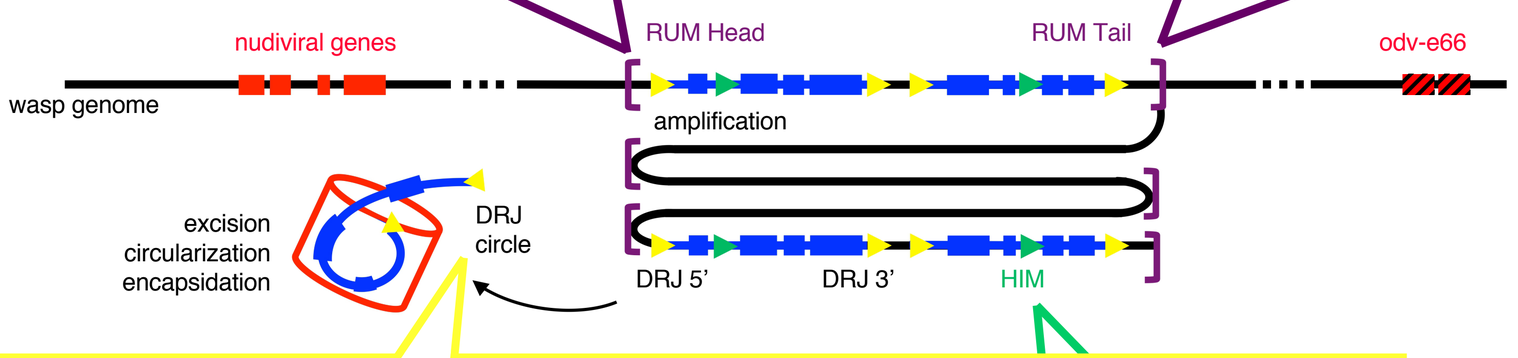
```
                                10        20        30        40        50
CC_RUM_T2_H_PL1_PART-2   TTTGGGATGATTGGTAACCCTA--GTAATATATTATAATTTGAATCATATT--
CC_RUM_T2_H_PL2_PART-1   TTTGGGATGATTGGTAACCCTA--GTAAGATATTGTAATTTAAGTCATAAT--
CC_RUM_T2_H_PL2_PART-2   TCTGGGACTCATAGTAAGCCTA--CTTATATAATAAATAATCTTAAAACATAA
CC_RUM_T2_H_PL2_PART-3   TTTGGGATGATTGGTACCCCTA--TTAGTATATCGTAATTTAAGTCATAAT--
CC_RUM_T2_H_PL9_PART-1   TTTGGGACGATTAGTAACCCTATAGTAATATATTGTAATTTGAGTCACAAT--
MD_RUM_T2_H_PL1_SEG-O    TTATGGATTGATAGCAAGCCTA--GTTTTTATTAACAATTAGGTGAG------
MD_RUM_T2_H_PL1_SEG-P    AACTGGACTGATGGCAAGCTTA--CATTAAATAATATGTTCAAGCAA------
```

Replication Unit Motif Type 1 Tail

```
                                10        20        30        40        50        60
CC_RUM_T1_T_LP3_SEG-17   TAACAATATTTGTCATAATTTATGACAAGTATTGATTATATCAATATATTTATACTTTTTTTCACTC
CC_RUM_T1_T_LP4_SEG-7    TTAAAAAACGTGATATAATTTATTCATGA------CCGTACAGTAAATTCTCACGAATCCGAATCATCGG
CC_RUM_T1_T_LP5_SEG-1    ACTAAAGTAGTGATATAATTTATTCATAAATTATTCTTACAGTAAAACTTCTTATATCCGAATTTAACT
CC_RUM_T1_T_LP8_SEG-26   ACTATAGTAGTGATATAATTTATTCATAAATAATCCTTACAGTAGAAGCTCGCATATCCGAATCAA-TT
CC_RUM_T1_T_PL7_SEG-4    TATTTAATAGTGATATAATTTATTCATAAATTATTCTTACAGTAAGGAACCGTTCATCCGTATTTTATA
MD_RUM_T1_T_LP3_SEG-J    ACTGTAGTAGTGATATAATTTATTCATAAATTATCCTTACAGTAAAACTCCGATTATCCGGATTCAACT
MD_RUM_T1_T_LP3_SEG-N    ACTGTAGTAGTGATATAATTTATTCATAAATTATCCGTACAGTAGCACTCCGATTATCCGGATCAAATT
```

Replication Unit Motif Type 2 Tail

```
                                10        20        30        40
CC_RUM_T2_T_PL2_PART_6   TAATTATTATAATTATAATTTAATAGGCTTACCATAAGCCCTTAT
CC_RUM_T2_T_PL6_PART-1   AAATTTTAGTTTAAGTAATGACTAGGCTTGCCATAAATTTGAAA
CC_RUM_T2_T_PL6_PART-2   ATGAATGAATTTATTTTGATACTTAGACTCTGTCTCATTCAATAG
CC_RUM_T2_T_PL6_PART-3   TAAATATAAATTTTAGAAATGATTAGTATTCATATTTCGACAAAT
CC_RUM_T2_T_PL6_PART-4   GGGTATATATATATAGATATATAATCTTGCCATAAGTCCAGAT
MD_RUM_T2_T_PL1_SEG-G    AATTGTATCTATTATTAAATAAATAGACTTGCCATGAGTCTTAGA
MD_RUM_T2_T_PL2_SEG-X    T-AATAATAATAATAAGATTAACTAGACTTGCCATAATTTTGAAT
```

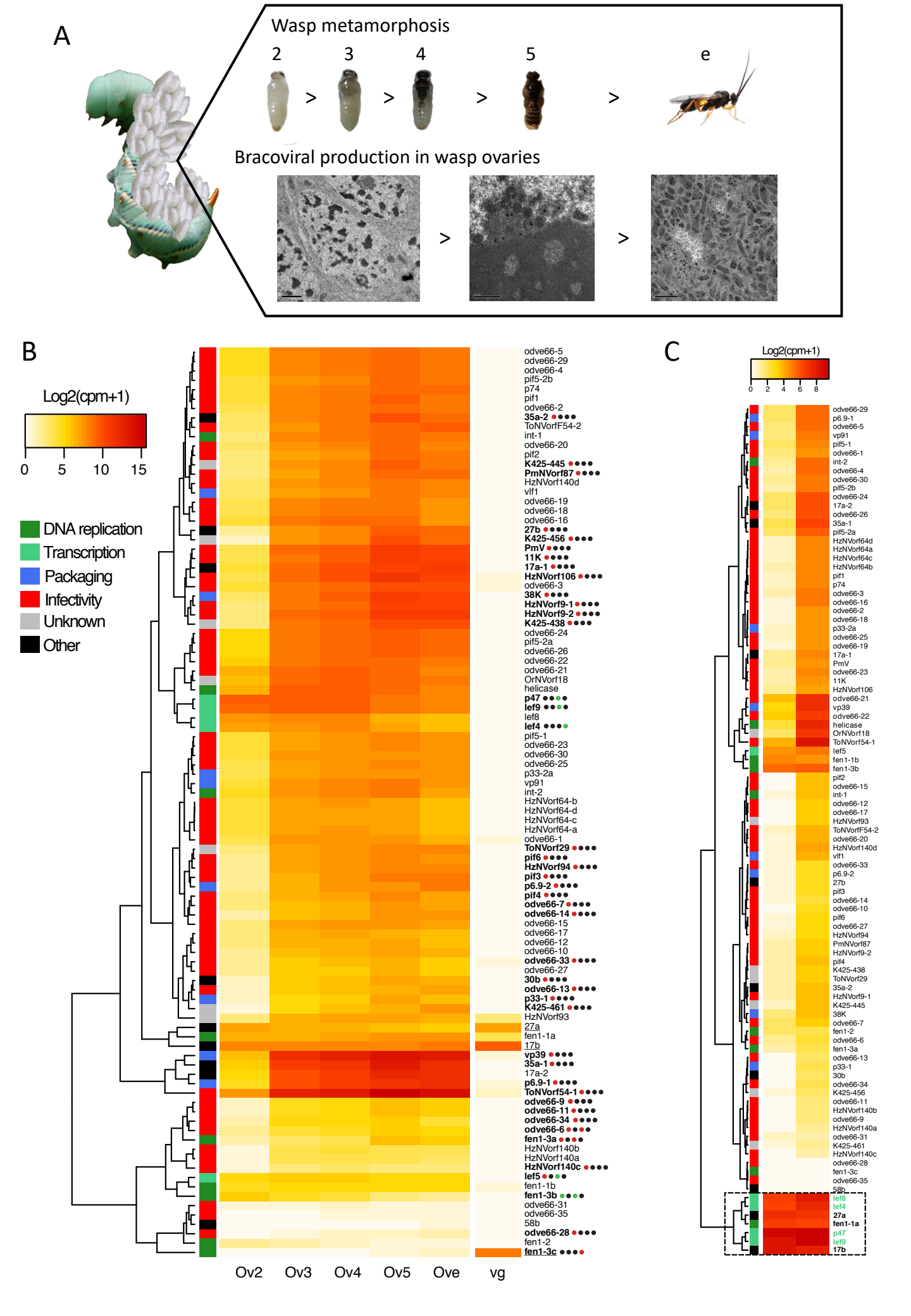


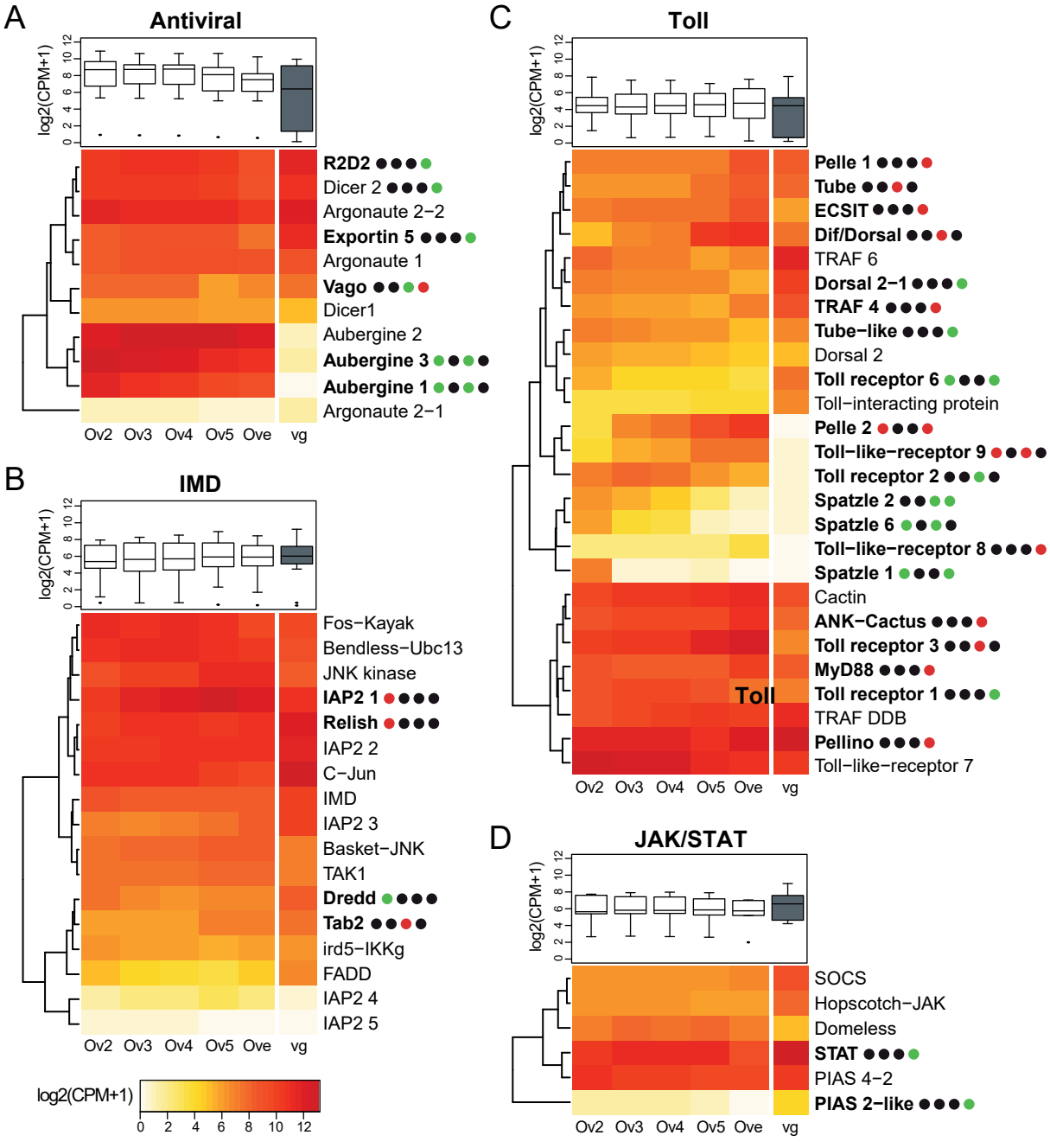
DRJ Circle

```
                                10        20        30        40        50        60        70        80        90        100       110
CC_DRJ_CIRCLE-5          TGATTTTTAGTAGTATATTAATTCTTGGTAAAGAGTGAATTTTTTATGATCCAAGAATTGACAAATGTGAGTTGCTTATAGCTTTTTGAATAATGAATTGAAATAACGATCATAT
CC_DRJ_CIRCLE-6          TGAATAAGAATGATGATTAAATCACTATCGAAAAAGGATCAATAGATTATCAAATCATCTGTTTATGTCACAATCAAGTAGCTTTTTAAATATTGAACAAAATGACTGATCATAG
CC_DRJ_CIRCLE-7          TGAATTGTATGACTGAATAAATTAAAAACTAAAAAGTTTTCGAGTATGTATTTAATTTGTACTTA-GTTCTAATTGTATAGCTTTTCTAGATCTGAACAAAAATGGTGAACAAAA
CC_DRJ_CIRCLE-8          TGATTTTTAGTGGTGTATTAATTTTGGTAATGAGTGAATAGTTGATGACCTAAGATTAAATAAATGTGAGTTTCTTATAGCTTTTTGAATAATGAATTGAAATAACGATCAGAT
CC_DRJ_CIRCLE-9          TGAATTTAGAAAGTAATCTAATTGAGACTTTATTGCGGTTAAAATTGTTTTTAATCATTAACAAACTTCATAAGCAAATAGCTTTTTGAATAGTGAACAAAAATAGTGATCAACT
MD_DRJ_CIRCLE-B          TGAATTAAATGACACCTCATATGATGACTTATTTAAGTATTTAATAATCATCAGTCCATTATTGATTCATCTATATTAAAGCTTTCTGAATAATGAACAAAAATATTGAACAAAG
MD_DRJ_CIRCLE-C          TGAATTCGAATAACACGCAAATTATGAACAAAGAATGCCTACAAGATGACTTAGTCATCTTTTTATCTGGAAGTCTAATAGCTTTTTGAATGATGAACAAAAATACTGAACAAAG
```

Host Integration Motif

```
                                10        20        30        40        50        60        70        80        90        100       110       120
CC_HIM_CIRCLE-17         TTAAATTTCGACATTGAAC-AGGTCCCTGGTCGACGCTCAGTACAT----------TTTG-TATAGGAAATCGCATTACCCTAACAGTAATGGTCGCGATCTGGAAATTTATAGACTTA
CC_HIM_CIRCLE-26         TGAAATTTCGACATGGAAA-CCTTTACTAGTATGCCCGCCATACAA----------ATCA-GTAGGATTTTAGACCAAGCACAACTGTAAAGTACGCGAACTGAAAGTTAAGGGACTT
CC_HIM_CIRCLE-35         ATAAATTTCTACATTCCAA-ACTCTCCTGGTTGGATCGTCGTAGAT-TTTCATACTAATTGT-ATAAGATTTAAGACTGTTGCTACCAGTAAAGTTTGCGATCTCGAAATCTAAGGAACGT
CC_HIM_CIRCLE-4          CTAAATTTCGACATCGAGA-TGAGCCCTAGTGGGCGCTCAGTACAT----------TTTG-TATAGGAAATCGCATGATCTTAACAGTAATAGTCGCGATCTGGAAATTTATGGACTTA
CC_HIM_CIRCLE-7          ATAATTTTCGACATGAAAA-CTGTTCCTGGTCTGAATTGGTTAGGA-ATGTATGGGATTTGT--AGGGATGCTGGTCTTTTTGTACCAGTTATAGTAGCGATTGTGAAACTAACGGGACTT
MD_HIM_CIRCLE-B          GAAAATTTCAACATAGCGA-TCGTCCCTAGTACGAACTTCACATGG-CTCTATACAAATTGC-AAAGCTTTTGTTAAAATACTTACCAGGAGTGATTTCTTAGTAGAAATTTTCAATTTAA
MD_HIM_CIRCLE-C          GAAATTTTCTGCCTGACGACTTATGCCTAGTAGGCTTTAACAAGTA-ATGTATGGAATCTTC-GAACAAACTGTTCTGGCTCGTACTAGGGGCGATCGTCAGGTCGAAAATTTCAATTAGC
```


A
Wasp metamorphosis
2
3
4
5
e
Bracoviral production in wasp ovaries
B
Log2(cpm+1)
0
5
10
15
DNA replication
Transcription
Packaging
Infectivity
Unknown
Other
odve66-5
odve66-29
odve66-4
pif5-2b
p74
pif1
odve66-2
35a-2
ToNVorfF54-2
int-1
odve66-20
pif2
K425-445
PmNVorf87
HzNVorf140d
vlf1
odve66-19
odve66-18
odve66-16
27b
K425-456
PmV
11K
17a-1
HzNVorf106
odve66-3
38K
HzNVorf9-1
HzNVorf9-2
K425-438
odve66-24
pif5-2a
odve66-26
odve66-22
odve66-21
OrNVorf18
helicase
p47
lef9
lef8
lef4
pif5-1
odve66-23
odve66-30
odve66-25
p33-2a
vp91
int-2
HzNVorf64-b
HzNVorf64-d
HzNVorf64-c
HzNVorf64-a
odve66-1
ToNVorf29
pif6
HzNVorf94
pif3
p6.9-2
pif4
odve66-7
odve66-14
odve66-15
odve66-17
odve66-12
odve66-10
odve66-33
odve66-27
30b
odve66-13
p33-1
K425-461
HzNVorf93
27a
fen1-1a
17b
vp39
35a-1
17a-2
p6.9-1
ToNVorf54-1
odve66-9
odve66-11
odve66-34
odve66-6
fen1-3a
HzNVorf140b
HzNVorf140a
HzNVorf140c
lef5
fen1-1b
fen1-3b
odve66-31
odve66-35
58b
odve66-28
fen1-2
fen1-3c
Ov2
Ov3
Ov4
Ov5
Ove
vg
C
Log2(cpm+1)
0
2
4
6
8
odve66-29
p6.9-1
odve66-5
vp91
pif5-1
odve66-1
int-2
odve66-4
odve66-30
pif5-2b
odve66-24
17a-2
odve66-26
35a-1
pif5-2a
HzNVorf64d
HzNVorf64a
HzNVorf64c
HzNVorf64b
pif1
p74
odve66-3
odve66-16
odve66-2
odve66-18
p33-2a
odve66-25
odve66-19
17a-1
PmV
odve66-23
11K
HzNVorf106
odve66-21
vp39
odve66-22
helicase
OrNVorf18
ToNVorf54-1
lef5
fen1-1b
fen1-3b
pif2
odve66-15
int-1
odve66-12
odve66-17
HzNVorf93
ToNVorfF54-2
odve66-20
HzNVorf140d
vlf1
odve66-33
p6.9-2
27b
pif3
odve66-14
odve66-10
pif6
odve66-27
HzNVorf94
PmNVorf87
HzNVorf9-2
pif4
K425-438
ToNVorf29
35a-2
HzNVorf9-1
K425-445
38K
odve66-7
fen1-2
odve66-6
fen1-3a
odve66-13
p33-1
30b
odve66-34
K425-456
odve66-11
HzNVorf140b
odve66-9
HzNVorf140a
odve66-31
K425-461
HzNVorf140c
odve66-28
fen1-3c
odve66-35
58b
lef8
lef4
27a
fen1-1a
p47
lef9
17b


A
Antiviral
log2(CPM+1)
R2D2
Dicer 2
Argonaute 2−2
Exportin 5
Argonaute 1
Vago
Dicer1
Aubergine 2
Aubergine 3
Aubergine 1
Argonaute 2−1
Ov2 Ov3 Ov4 Ov5 Ove vg
B
IMD
log2(CPM+1)
Fos−Kayak
Bendless−Ubc13
JNK kinase
IAP2 1
Relish
IAP2 2
C−Jun
IMD
IAP2 3
Basket−JNK
TAK1
Dredd
Tab2
ird5−IKKg
FADD
IAP2 4
IAP2 5
Ov2 Ov3 Ov4 Ov5 Ove vg
log2(CPM+1)
0 2 4 6 8 10 12
C
Toll
log2(CPM+1)
Pelle 1
Tube
ECSIT
Dif/Dorsal
TRAF 6
Dorsal 2−1
TRAF 4
Tube−like
Dorsal 2
Toll receptor 6
Toll−interacting protein
Pelle 2
Toll−like−receptor 9
Toll receptor 2
Spatzle 2
Spatzle 6
Toll−like−receptor 8
Spatzle 1
Cactin
ANK−Cactus
Toll receptor 3
MyD88
Toll receptor 1
TRAF DDB
Pellino
Toll−like−receptor 7
Toll
Ov2 Ov3 Ov4 Ov5 Ove vg
D
JAK/STAT
log2(CPM+1)
SOCS
Hopscotch−JAK
Domeless
STAT
PIAS 4−2
PIAS 2−like
Ov2 Ov3 Ov4 Ov5 Ove vg