

Adaptive evolution of viral genes and wasp diversification

Serbielle C, Herniou EA, Quicke D, Drezen JM, Huguet E.

Abstract:

Symbioses allow acquisition of new functions and offer their hosts new opportunities to colonize ecological niches. Vertically transmitted symbionts could therefore promote host specialization and trigger their diversification.

Thousands of endoparasitoid wasps have been shown to be tightly associated with polydnviruses essential for wasp parasitism success. Viruses are injected into the host with wasp eggs during oviposition and viral gene expression is involved in the control of both host development and immunity. Moreover, polydnviruses are integrated in the wasp genome and are therefore vertically transmitted. These viruses, as vertically transmitted and obligatory symbionts, could contribute to the speciation of their carrier-wasps by mediating wasp adaptation and specialization to their lepidopteran host.

In this study, we focus on the role of a particular class of virulence genes encoding for cystatins. Cystatins are cysteine protease inhibitors, which have been shown to be highly expressed and at an early stage during parasitism. This suggests they could play a role in the early steps of parasitism. Cystatin genes were acquired recently by the symbioses and are restricted to narrow wasp lineages.

The aim of this study is to characterize cystatin evolutionary patterns in relation to wasp evolutionary history in order to infer their contribution to wasp adaptation, specialization and finally diversification.

Although it is still preliminary, I wanted to outline the great lines of this study which is the first in this field.

Introduction

Symbiotic associations, as illustrated by mitochondria and chloroplast have played a prominent role in evolution. These associations have provided key biological innovations conferring selective advantages for the colonization of new ecological niches and subsequent specialization. The plataspid stinkbug insects for example show plant specializations, which are in fact mediated by the associated bacterial symbiont (Hosokawa *et al.*, 2007) which potentially induce host diversification. In these cases, the genome and the encoded functions of the symbiont are predicted to be directly involved in host specialization processes. What are the functions acquired through symbiosis? How are these functions diverging during host-symbiont evolution and influencing host evolution?

In nutritional obligatory symbioses, illustrated for example by the highly specialized association between the bacteria *Buchnera aphidicola* and their host aphid, the symbiont supplies the host with essential nutrients. In this particular case, the bacteria are vertically transmitted and co-diverge with their host (Clark *et al.*, 2000). Other symbionts provide protection against natural enemies. This is the case of non obligatory bacterial symbionts which confer resistance to aphids against parasitoids (Oliver *et al.*, 2003). The transmission pattern and the degree of interdependence of the partners are factors which can influence host adaptation and its diversification. Vertically transmitted symbionts with no lateral transfer promote divergent selection on the host. A facultative symbiont is not expected to be differentiated between different host populations and could, in contrast, provide a form of phenotypic plasticity, favour generalized functions instead of specialization, and act as an inhibitor on host diversification (for review see (Janson *et al.*, 2008)).

Among endoparasitoid wasps, particular lineages are intimately associated with polydnviruses, which are involved in wasp protection against host immunity and are therefore essential for wasp parasitism success. PDVs are integrated into the wasps' chromosomes, as proviruses, and thus transmitted vertically from one wasp generation to the next (Belle *et al.*, 2002; Desjardins *et al.*, 2007; Fleming and Summers, 1991; Gruber *et al.*, 1996; Xu and Stoltz, 1991). Proviruses are excised from the wasp genome in the female ovaries and are injected in the host caterpillar as multiple double stranded DNA chromosomal circles. The virus does not replicate in the host caterpillar, but viral gene expression and protein production are essential for host immune and developmental alterations leading to successful development of the wasp larvae. Do viruses carry functions promoting wasp adaptation and specialization? How could viruses be involved in wasp radiation?

Polydnaviruses are associated with two wasp families and were therefore described as two virus lineages presenting distinct structural properties; Ichnoviruses are found in Ichneumonidae wasps and Bracoviruses are found in Braconidae wasps (Turnbull and Webb, 2002). Recently, a virus of a Banchinae wasp was found to be distinct enough from ichnoviruses and bracoviruses to constitute a third virus lineage (Lapointe *et al.*, 2007) suggesting that three independent symbioses events are at the origin of the wasp-virus associations visible today. Braconid wasps associated with polydnaviruses form the Microgastroid complex, which diverged from an unique ancestral association estimated to have occurred 103 Mya ago (Murphy *et al.*, 2008). The capture of the virus by the ancestor braconid wasp is thought to have been followed by a rapid wasp diversification probably in response to lepidopteran host speciation (Banks and Whitfield, 2006; Mardulyn and Whitfield, 1999; Murphy *et al.*, 2008).

The PDVs represent an evolutionary novelty which can be predicted to confer a selective advantage to the wasp by promoting new host colonization and therefore act by playing an active part in wasp radiation.

Genomic analyses of several polydnavirus genomes is now starting to provide keys to understand how this association is maintained and what are the physiological functions supplied by the wasp to ensure parasitism success. A recent study, combining transcriptomic and proteomic approaches showed that virus genes homologous to nudivirus genes are encoded in the wasp genome and expressed in the ovaries to form viral particles (Bezier 2008, submitted). However, no genes homologous to viral structural genes (i.e. capsid) were found in the virus genome injected into the host caterpillar. In contrast, the injected virus contains putative virulence factors presenting similarities with cellular genes probably acquired from the wasp genome during the evolutionary history of the symbiosis (Desjardins *et al.*, 2007; Espagne *et al.*, 2004; Lapointe *et al.*, 2007; Webb *et al.*, 2006). Similar to what is generally observed in bacterial symbioses (Foster *et al.*, 2005; Kondo *et al.*, 2002; Nakabachi *et al.*, 2006; Nikoh *et al.*, 2008; Perez-Brocal *et al.*, 2006), these data prove that virus-wasp evolution was accompanied by genome reduction of the viral symbiont and the acquisition of new functions by the wasp host. These patterns are molecular signatures of the long co-evolution of the two partners which have made these viruses and parasitoid wasps obligatory interdependent for capital functions and which have shaped the symbiont genome since its integration into the wasp genome.

Most putative virulence factors transferred to the virus genomes are organized in multigene families suggesting strong adaptive pressures are imposed on PDV genomes, emphasizing their role in parasitism success (Desjardins *et al.*, 2007; Espagne *et al.*, 2004; Lapointe *et al.*, 2007; Webb *et al.*, 2006). Among them, I κ B and protein tyrosine phosphatases

are known to have been acquired early during the virus-wasp association. The I κ B proteins are inhibitors of nuclear transcription factors involved in vertebrate and in *Drosophila* immune responses (De Gregorio *et al.*, 2001; Hoffmann, 2003). The protein tyrosine phosphatases form the largest gene family and are known to play a key role in the control of signal transduction pathways by dephosphorylating tyrosine residues on regulatory proteins (Andersen *et al.*, 2001). These two classes of PDV genes are widely represented in bracoviruses suggesting they have coevolved with the wasp genome for a long time. In contrast, a recently described class of genes encoding cystatins is restricted to narrow virus lineages and is proposed to have appeared recently in the history of the virus-wasp association. Ongoing functional studies are unravelling the role of virulence genes during parasitism, however their divergence during symbiosis evolution and their contribution to wasp adaptation and diversification have never been investigated.

As recently acquired genes, PDV cystatins are good candidate with which to study the early evolutionary steps leading to the acquisition of a new gene function and to investigate their contribution to wasp diversification. Cystatins are widely distributed in living organisms but PDV cystatins constitute the first description of viral cystatins (Espagne *et al.*, 2005). Cystatins and their target proteases have often been shown to be involved in host-parasite interactions with cystatins either playing the role of defence molecules or virulence factors. For example, in parasitic nematodes, cystatins are thought to play a key role in controlling the host immune response (Dainichi *et al.*, 2001; Maizels *et al.*, 2001; Schierack *et al.*, 2003). So far, cystatin genes have been found as a unique gene in the wasp *Glyptapanteles indiensis* (Desjardins *et al.*, 2007) and in several copies in the closely related genus *Cotesia* (Espagne *et al.*, 2005; Serbielle *et al.*, 2008). This suggests that cystatin duplications occurred recently in PDV evolution. In the *Cotesia congregata* Bracovirus (CcBV), cystatins have been shown to be highly expressed during parasitism and to be active against cysteine proteases (Espagne *et al.*, 2005). Moreover, PDV cystatins have been shown to exhibit the same conformational structure than other eukaryotic cystatins and interacting sites were proved to evolve under positive selection (Serbielle *et al.*, 2008). Altogether these results suggest that cystatins are virulence factors important for parasitism.

Cystatin lineages and *Cotesia* wasps probably diverged contemporaneously; we can therefore hypothesize that cystatins could be involved in wasp diversification. This study is still in its preliminary steps but aims at (1) testing *Cotesia* wasp and viral cystatin co-divergence, (2) estimating the evolutionary pressures acting on cystatins and their role in wasp adaptation and diversification.

Material and Methods

Wasp samples

DNA sequence data were obtained from 49 individual wasps belonging to 26 different *Cotesia* species. *Cotesia* parasitoids were sampled fresh from cocoons found in the field or reared from wild collected hosts, which were carefully identified. Some samples were from dried and pinned museum specimens and others were from lab rearing (Table 10). *Glyptapanteles indiensis*, was also included in this analysis as an outgroup.

Table 10 provides a list of 49 specimens belonging to 26 *Cotesia* species, collected for this study, along with the localities and lepidoteran host species. Viral cystatin genes isolated from associated bracoviruses associated are included.

| Wasp species name | Host species and families | Localities | Viral cystatin gene abbreviations | Number of cystatin copies |
|--------------------------------|---|-------------|-----------------------------------|---------------------------|
| <i>C. neustriala</i> | <i>Malacosoma castraise</i> | Greece | Cyst neustriala | 1 |
| <i>C. glabrata</i> (56) | <i>Carcharodus baeticus</i> (Hesperidae) | Spain | Cyst glabrata56 | 1 |
| <i>C.sp near glabrata</i> (60) | <i>Pyrgus cirsii</i> (Hesperidae) | Spain | Cyst glabrata60 | 1 |
| <i>C.sp near glabrata</i> (7) | <i>Carcharodus alceae</i> (Hesperidae) | Spain | Cyst glabrata7 | 2 |
| <i>C. glabrata</i> (6) | <i>Carcharodus tripolina</i> (Hesperidae) | Maroc | Cyst glabrata6 | 2 |
| <i>C. glabrata</i> (1) | <i>Carcharodus alceae</i> (Hesperidae) | France | Cyst glabrata1 | 1 |
| <i>C. glabrata</i> (2) | <i>Carcharodus alceae</i> (Hesperidae) | France | Cyst glabrata2 | 2 |
| <i>C. pieridis</i> | <i>Aporia crataegi</i> (Nymphalidae) | Greece | Cyst pieridis | 1 |
| <i>C. tenebrosa</i> | <i>Polyommatus icarus</i> (Lycaenidae) | Scotland | Cyst tenebrosa | 1 |
| <i>C. salebrosa</i> (8) | <i>Epirrita autumnata</i> (Geometridae) | Finland | Cyst salebrosa8 | 2 |
| <i>C. salebrosa</i> (9) | <i>Epirrita autumnata</i> (Geometridae) | Finland | Cyst salebrosa9 | 1 |
| <i>C.astrarches</i> (?) (14) | <i>Aricia artaxerxes</i> (Lycaenidae) | England | Cyst astrarches14 | 1 |
| <i>C. astrarches</i> (15) | <i>Aricia agestis</i> (Lycaenidae) | England | Cyst astrarches15 | 2 |
| <i>C. astrarches</i> (16) | <i>Polyommatus thersites</i> (Lycaenidae) | Greece | Cyst astrarches16 | 3 |
| <i>C. astrarches</i> (37) | <i>Aricia agestis</i> (Lycaenidae) | England | Cyst astrarches 37 | 1 |
| <i>C. acuminata</i> (23) | <i>Euphydryas maturna</i> | France | Cyst acuminata23 | 1 |
| <i>C. acuminata</i> -agg (29) | <i>Melitaea telona</i> (Nymphalidae) | Greece | Cyst acuminata29 | 2 |
| <i>C. acuminata</i> -agg (27) | <i>Melitaea phoebe</i> (Nymphalidae) | Italy | Cyst acuminata27 | 2 |
| <i>C. acuminata</i> (52) | <i>Melitaea phoebe</i> (Nymphalidae) | Spain | Cyst acuminata52 | 1 |
| <i>C. acuminata</i> -agg (31) | <i>Melitaea telona</i> (Nymphalidae) | Greece | Cyst acuminata31 | 1 |
| <i>C. acuminata</i> -agg (28) | <i>Melitaea telona</i> (Nymphalidae) | Greece | Cyst acuminata28 | 2 |
| <i>C. acuminata</i> -agg (30) | <i>Melitaea telona</i> “argigia”(Nymphalidae) | Greece | Cyst acuminata30 | 3 |
| <i>C. kariyai</i> | <i>Pseudaletia separata</i> (Noctuidae) | Japan | Cyst kariyai | 6 |
| <i>C. vanessae</i> (1) | <i>Cynthia cardui</i> (Nymphalidae) | Spain | Cyst vanessae1 | 3 |
| <i>C.vanessae</i> (3) | <i>Vanessae atalanta</i> (Nymphalidae) | England | Cyst vanessae3 | 3 |
| <i>C. vanessae</i> (45) | <i>Aglais urticae</i> (Nymphalidae) | Belgium | Cyst vanessae45 | 4 |
| <i>C. ruficus</i> | <i>Autographa gamma</i> (Noctuidae) | Austria | Cyst ruficus | 2 |
| <i>C. ofella</i> | <i>Acronicta rumicis</i> (Noctuidae) | France | Cyst ofella | 3 |
| <i>C. melitaeorum</i> (41) | <i>Melitaea deione</i> (Nymphalidae) | Spain | Cyst melitaeorum41 | 4 |
| <i>C. melitaeorum</i> (43) | <i>Melitaea deione</i> (Nymphalidae) | Spain | Cyst melitaeorum43 | 3 |
| <i>C. vestalis</i> (2) | <i>Plutellae xylostella</i> | lab rearing | Cyst vestalis2 | 1 |
| <i>C. vestalis</i> (3) | <i>Plutellae xylostella</i> | Benin | Cyst vestalis3 | 3 |
| <i>C. melanoscela</i> | <i>Lymantria dispar</i> (Nymphalidae) | France | Cyst melanoscela | 5 |
| <i>C. flavipes</i> (2) | <i>Busseola fusca</i> (Noctuidae) | Kenya | Cyst flavipes2 | 4 |
| <i>C. flavipes</i> (1) | - | - | Cyst flavipes1 | 2 |
| <i>C. chilonis</i> | <i>Chilo suppressalis</i> | | Cyst chilonis | 4 |
| <i>C. sesamiae</i> (1) | <i>Busseola fusca</i> (Noctuidae) | Kenya | Cyst sesamiae1 | 1 |
| <i>C. sesamiae</i> (2) | <i>Busseola fusca</i> (Noctuidae) | Kenya | Cyst sesamiae2 | 3 |
| <i>C. glomerata</i> (11) | <i>Aporia crataegi</i> (Nymphalidae) | Spain | Cyst glomerata11 | 2 |
| <i>C. glomerata</i> (2) | <i>Pieris brassicae</i> | lab reared | Cyst glomerata2 | 2 |
| <i>C. glomerata</i> (3) | - | - | Cyst glomerata2 | 2 |
| <i>C.sp near glabrata</i> (61) | <i>Carcharodus baeticus</i> (Hesperidae) | Spain | Cyst glabrata61 | 1 |
| <i>Cotesia sp</i> (2) | <i>Cerura vinula</i> (Noctuidae) | france | Cyst sp2 | 3 |
| <i>Cotesia sp</i> (3) | <i>Cerura vinula</i> (Noctuidae) | France | Cyst sp3 | 2 |
| <i>C. sibyllarum</i> | <i>Limenitis reducta</i> (Nymphalidae) | Spain | Cyst sibyllarym | 3 |
| <i>C. rubecula</i> | <i>P. rapae</i> (Pieridae) | | Cyst rubecula | 4 |
| <i>C. congregata</i> (1) | - | - | Cyst congregata1 | 4 |
| <i>C. congregata</i> (2) | <i>Manduca sexta</i> (Sphingidae) | Lab | Cyst congregata2 | 3 |

DNA extraction and sequencing

For DNA isolation, insects were ground in Wilson buffer (0,1M Tris, 10mM EDTA, 0,1M NaCl, 1mM SDS, 50mM DTT) and incubated for 1-2h with proteinase K (0.6 mg/ml) at 37°C. One phenol/chloroform extraction was followed by ethanol precipitation and resuspension in ultrapure water.

A fragment of about 1482 bp of the cytochrome oxydase I (COI) was amplified by using the universal primers HCO1490 and LCO2198 (Former et al 1994) and CI-J-2183 and TL2-N-3014 (Simon et al 1994). In addition, a fragment of about 500 bp from the NADH mitochondrial gene was isolated using primers ND1-F and ND1-R (Smith et al 1999). PCR amplifications were performed in a total volume of 20µl containing 1X Taq buffer, 2 mM of MgCl₂, 200mM of dNTP, 0.2 µl Taq polymerase (Goldstar, Eurogentec) and 30 pmol of each primer. PCR conditions consisted of an initial denaturation step at 95°C for 2 min followed by 35 cycles of a denaturation step at 94°C for 1 min, annealing step for 1 min at 49°C for COI or 48°C for NADH and polymerization step at 72°C for 1min 30s and final elongation at 72°C for 10 min (Kankare and Shaw, 2004; Michel-Salzat and Whitfield, 2004). PCR products were purified using the Exosap-IT kit (GA-Healthcare). Sequences were obtained in both directions using the Big Dye^R terminator v3.1 cycle sequencing Kit and analysed using a capillary DNA sequencer (ABI prism).

Cystatin gene amplifications were performed using two primers; Cyst15 and Cyst103 (Serbielle et al, 2008). The amplified product is around 400 nucleotides long and encodes a 93 amino acid sequence containing the first two conserved domains of the cystatins. PCR amplification was performed in a 50 µl volume containing 1X Taq buffer, 3mM of MgCl₂, 2.5mM of dNTP, 0.3 µl Taq polymerase (Goldstar, Eurogentec) and 50 pmol of each primer. PCR conditions consisted of an initial denaturation step at 94°C for 2 min followed by 30 cycles of a denaturation step at 94°C for 45 s, annealing step at 45°C for 1 min and polymerization step at 72°C for 45 s and final elongation at 72°C for 10 min. PCR products were cloned into the pDrive-cloning vector (Qiagen cloning kit). For each wasp sample, 12 clones were sequenced for each bracovirus species as described above. In order to avoid analytical biases generated by PCR errors, only clones obtained at least twice were included in our dataset except if sequences were very dissimilar.

Phylogenetic analyses

NADH and COI genes were used to infer wasp phylogeny. These two genes were initially aligned using Clustal W implemented in Bioedit (Hall, 1999) and subsequently optimised manually in McClade with respect to codons. For each gene, Modeltest 2.2 (Posada and Crandall, 1998) selected the evolutionary model GTR+I+G using Aikaike Information Criteria and hierarchical likelihood ratio test.

Initially, phylogenetic trees were built with COI and NADH separately and these two data sets were eventually assembled as a single dataset. Maximum likelihood phylogenetic analyses were conducted in Phyml (Guindon and Gascuel, 2003) and support values for internal nodes were estimated using a bootstrap resampling procedure with 1000 replicates. Using MrBayes (Huelsenbeck and Ronquist, 2001), Monte Carlo Markov chains (MCMC) were run for 2000000 and 3000000 generations for separate and combined data set respectively, sampling every 1000 generations with the first 200000 and 300000 generations discarded as burn-in respectively.

Cystatin genes alignment was performed as described above. MrModeltest ver2.2 (Posada and Crandall, 1998) selected the evolutionary model GTR+G. Phylogenetic analyses were estimated by maximum likelihood using Phyml (Guindon and Gascuel, 2003) with 1000 replications and Bayesian analyses using 10000000 generations, sampling every 100 generations with the first 100000 generations discarded as burn-in. The cystatin tree was rooted using cystatin from partial genome sequencing of GiBV.

Co-phylogeny study

In this section we want to reconcile the viral cystatin tree with the wasp tree based on COI and NADH recombined analyses. A reconciliation describes how the gene tree has evolved within a species. For that, we will use the primeGEM program, which infers reconciliation by probabilistic approaches. MCMC algorithm estimates, for a given gene tree and species tree, the posterior distribution of reconciliation and estimates the probability of two genes being orthologs.

Practically, we need to use the (1) cystatin tree, (2) a file mapping cystatin genes and wasp species and (3) a wasp tree with divergence time. We used the Bayesian cystatin tree and divergence times on the wasp tree estimated with BEAST v1.4.6 (Drummond *et al.*, 2006). This program uses a Bayesian inference and a MCMC procedure to estimate the posterior distribution of rates. Given no fossil calibration was yet possible, we estimated relative divergence times by fixing root time to one and time divergence were then estimated relatively to the root age. We used a relaxed molecular clock and a constant-rate Yule (speciation process) prior and as starting

tree the maximum posterior probability topology obtained from the Bayesian analysis, which remained fixed throughout the MCMC. Two independent chains were run for 20000000 generations each, and convergence checked by comparing these runs after a burn-in of 2000000 generations was discarded.

These three files are then used to run PrimeGEM analyses. PrimeGEM gives a posterior probability of the duplication and loss parameters which are used to get the orthology probabilities of the internal nodes of the gene tree or to enumerate reconciliations. A particular algorithm is then used to extract from all possible reconciliation the maximum probability reconciliation.

Selective pressures acting on cystatin genes

In order to infer selective pressures acting on cystatin lineages we calculated the ω (dN/dS) value, which is the ratio of non-synonymous / synonymous substitution rates. We used a naïve approach to detect branches specifically under positive selection in the tree using Ga-Branch analysis implemented in the HyPhy package (Pond and Frost, 2005; Pond et al., 2005). The basic principle of this method is to assign each branch of a phylogenetic tree to a particular ω class. Different models assigning branches into different ω classes were tested and compared using the Akaike information criterion (AIC_c). To search the space of possible models HyPhy employs a genetic algorithm (Ga) that measures the fitness of each model by its AIC_c score. Ga-branch analysis enables the assignment of lineages in a phylogeny to a fixed number of different classes of ω , thus allowing variable selection pressure without *a priori* specification of particular lineages. The evolutionary codon model used for this analysis was determined from the AnalyzecodonData program implemented in the Hyphy package.

Beast and Apes: Molecular dating and estimating diversification dates

The aim of this analysis is to know if cystatin and wasp genes diverged concomitantly in time or if cystatin diversification occurs after or before wasp genes, which are here considered as markers of wasp speciation.

Divergence rates were estimated using a log-normal distributed relaxed molecular clock using BEAST v1.4.6 (Drummond *et al.*, 2006). Given cystatin genes were undergoing positive selection (Serbielle *et al.*, 2008), time divergence for both wasps and cystatins were estimated from the third codon position, which are supposed to be less constrained.

For cystatin divergence time estimation, we used a birth-death process prior and as starting tree we used the Bayesian cystatin tree and we constrained the tree topology throughout the MCMC analyses. Two independent chains were run for 20000000 generations each, and convergence checked by comparing these runs after a burn-in of 2000000 generations was discarded.

The resulting ultrametric trees obtained for wasp genes and cystatins will be imported into the APE package (Paradis et al., 2004) to generate semi logarithmic LLT plots calculated from sampled trees. In this analysis, the log of lineages is plotted against the relative time of each node since the root node, giving us the tendency of diversification rates through time.

Preliminary results

Phylogenetic analyses

Wasp gene phylogeny

We obtained COI gene sequences for all wasp samples tested whereas we were unable to isolate NADH sequences for particular wasp species. The ML and Bayesian trees based on NADH and COI genes studied separately were congruent but contrary to NADH, COI analyses failed to resolve ancestral nodes. NADH is a slower evolving gene useful for determining ancient relationships while COI is more particularly adapted to resolve evolutionary history at the species level. In order to improve wasp tree resolution we build ML and Bayesian trees with a combined dataset. The trees obtained were better resolved with both ML and Bayesian analyses except for three ancestral nodes (figure 21). The root is supported by a weak value probably due to the close relation between *Glyptapanteles* and *Cotesia* genus. The support value for the branch sustaining the *C. glabrata* and *Csp near glabrata* species clade is weak, suggesting that *C.sp near glabrata* 56 and 60 are not *Cotesia glabrata* species. The last weakly supported node presents a posterior probability value of 0.60 and a bootstrap value of 534 and constitutes also an ancestral node.

Reconciliation analyses require well supported trees, in spite of the three weak supported nodes, the COI-NADH tree is overall well supported and will be used for the following analyses.

Viral cystatin gene phylogeny

We constructed a phylogeny with 121 cystatin sequences isolated from bracoviruses associated with 49 wasp samples. For almost each sample we obtained different cystatin clones which could refer to different gene copies or alleles. The viral cystatin sequence from GiBV was used to root the tree.

ML and Bayesian analyses gave the same tree topology for cystatin gene phylogeny. Some nodes present weak support and are mostly found in the ancestral part of the tree but cystatin tree topology is overall robust. We can observe two patterns of cystatin lineage evolution. The first pattern is characterized by clades formed by viral cystatin lineages which were isolated from the same wasp species, as it is observed for Cyst sp55, Cyst glomerata, Cyst salebrosa, Cyst melanoscela, Cyst vestalis, Cyst sesamiae2, Cyst flavipes1, Cyst melitearum and Cyst vanessae. In contrast, some cystatin lineages are not grouped by wasp species like Cyst sp3, Cyst sp2, Cyst sybillarum, Cyst congregata, Cyst sesamiae1, Cyst flavipes2, Cyst chilonis, Cyst rubecula, Cyst acuminata, Cyst ofella and Cyst ruficus. In the first pattern, cystatin copies and/or alleles isolated from bracoviruses associated with the same wasp specimen form individual clades whereas in the second pattern cystatin lineages from the same specimen are distributed in several clades and group with cystatins from different bracovirus species. Three different clades are concerned by the second pattern; a clade formed by all Cyst glabrata and most of Cyst acuminata, a second clade formed by Cyst congregata, Cyst rubecula, Cyst chilonis, Cyst flavipes2 and Cyst sesamiae1 and a third clade formed by Cyst sp2, Cyst sp3, Cyst acuminata30 and Cyst sybillarum (see figure 22 and colored sub clades).

In conclusion some cystatin lineages are likely to co-diverge with wasp genes whereas other do not show the same pattern. To ensure that wasp and cystatin genes share a common evolutionary history, it is necessary to test statistically this hypothesis.

Viral cystatin lineages under selection

In order to understand cystatin lineage evolution and characterize selective pressures acting on cystatin lineages, we measured the ω ratio for each branch of the cystatin tree. Hyphy analyses showed that a large majority of cystatin lineages evolved under positive selection. Indeed, Ga-branch analyses selected a model where cystatin tree branches evolved under two different ω ratio; 60% of branches diverged under positive selection with $\omega=2.916$ while 40% evolved under purifying selection with $\omega=0.375$.

Three different kinds of branches can be described to evolve under positive selection (Figure 23).

The first class refers to cystatin lineage evolution after wasp speciation. These branches correspond to the most external branches and refer to the last steps of cystatin alleles/copies divergence. We can effectively observe that the evolution of these branches is mainly explained by positive selection.

The second kind of branches refers to cystatin lineage evolution before wasp speciation. For example, the clade formed by *C. vestalis*, *C. melitearum*, *C. sesamiae*, *C. flavipes* in the wasp tree presents the same topology as the clade formed by Cyst vestalis, Cyst melitearum, Cyst flavipes and Cyst sesamiae in the cystatin tree, and all branches sustaining these cystatin lineages (or sub-clades) evolved under positive selection. This pattern suggests that cystatin diversification processes which preceded wasp speciation occurred under positive selection.

The third class of branches concerns more internal cystatin branches and corresponds to ancestral cystatin lineage evolution during the early steps of wasp diversification processes. For these particular branches, it is difficult to distinguish between cystatin copy evolution it from cystatin evolution during wasp diversification processes. They constitute ancestral cystatin lineages of particular cystatin forms inherited by some bracovirus species. This pattern is observed in particular in the clade where cystatins are not grouped by bracovirus species (see clade colored in figure 22).

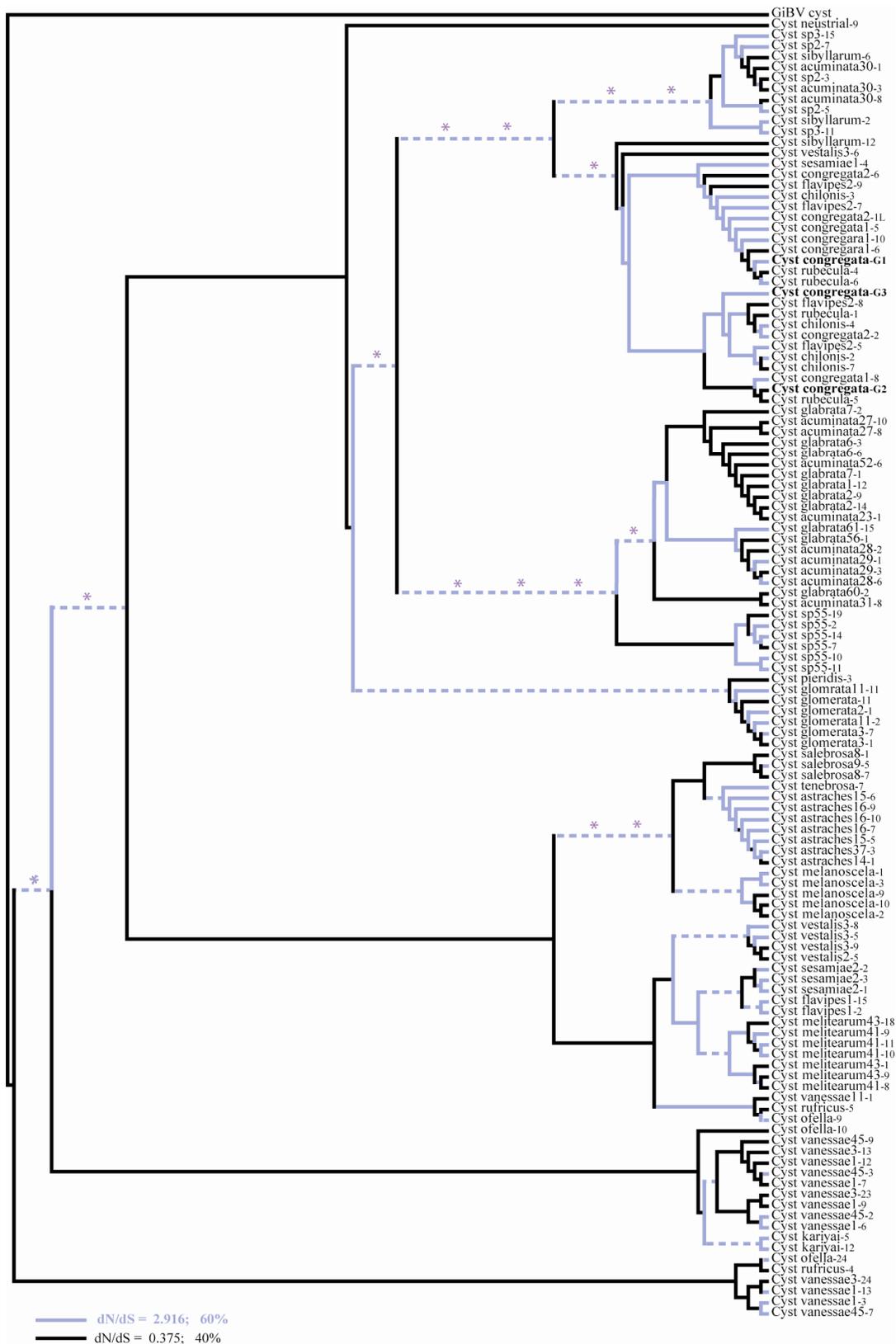


Figure 21 Viral cystatin lineages evolving under positive selection.

Branches significantly evolving under positive selection are in purple. 3 types of branches are mentioned: black line indicate branches corresponding to cystatin copy evolution after wasp speciation, dashed line indicate cystatin lineage evolution before wasp speciation and stars associated to purple lines indicate cystatin lineages evolution during the early steps of wasp diversification processes.

Reconciliation analysis

This part is still in preparation. Testing reconciliation analysis is really important to evaluate how cystatin genes could contribute to wasp adaptation and diversification. Indeed, it is a prerequisite to link cystatin gene adaptation to wasp speciation and to compare cystatin genes divergence to wasp diversification.

Cystatin and wasp divergence

This section is also in preparation and strongly depends on reconciliation analysis results. To run this analysis and to compare cystatin gene and wasp genes divergences we have first to estimate if cystatin genes and wasp genes share a common evolutionary history.

Preliminary discussion

The extraordinary diversity of the world's parasite is often attributed to the intimate relationship they share with a limited number of host species and the potential for disruptive selection pressures associated with different hosts driving ecological divergence (Ackermann and Doebeli, 2004). In insects, the parasitoid lifestyle implies an intimate relationship with hosts because a single host harbors the parasitoid's offspring until maturity. Although little is known about the mechanisms that mediate population and species divergence in parasitoids, evidence suggest hosts may be responsible for driving differentiation in lineages (Dubuffet *et al.*, 2006; Dupas *et al.*, 2008; Dupas *et al.*, 2003b).

Among parasitoid wasps associated with polydnviruses, the Microgastrinae wasps are one of the most diversified groups, characterized by high host specificity (Banks and Whitfield, 2006; Murphy et al., 2008; Smith et al., 2008). It has been proposed that the high diversification rate of these wasps is a response to host speciation (Banks and Whitfield, 2006). We can therefore suppose that symbiotic polydnviruses, by playing an important role in wasp parasitism success, can be involved in wasp adaptation, specialization and in their subsequent diversification. We would like to know, if in PDV-wasp associations, the symbiosis has been involved in wasp adaptation, specialization and diversification.

Cystatins genes were acquired recently by wasp species and are so far almost restricted to the *Cotesia* genus. These genes constitute interesting candidates to study these processes because cystatin and *Cotesia* lineage diversification probably occurred contemporaneously.

What cystatin evolutionary characteristics could be related to wasp adaptive properties?

Viral cystatins form a multigene family found in *Cotesia* bracoviruses (Espagne *et al.*, 2005; Serbielle *et al.*, 2008). Cystatin gene family amplification could have increased protein dosage and therefore allowed a better inhibition of host cysteine proteases. However, our analysis reveals cystatin evolution is principally explained by adaptive diversification. Interestingly, a previous study pointed that positive selection acted preferentially on cystatin interacting sites suggesting that cystatin adaptive evolution is a direct response to selective pressure imposed by host cysteine proteases (Serbielle *et al.*, 2008). Therefore, cystatin lineage divergence under adaptive selection could mean that selection acted to obtain different cystatins forms. Why is cystatin diversifying divergence selected?

Cystatin copies could be transitory and respond to a particular selective pressure which would later act to conserve a unique and efficient cystatin copy. In contrast, cystatin copies could be permanent and be under diversification in response to several selective pressures.

According to the first hypothesis, cystatin copy diversification could be part of the amplification phenomenon described by Bergthorsson and co-workers (2007). For them, gene expansion into families first occurs by a step of gene amplification, which multiplies genetic support for mutations and increases probabilities to obtain a new function. In this model gene amplification and mutation rate which enhance gene function innovation are undergoing adaptive selection. Amplification and gene innovations constitute therefore transitory phenomenon which occur only under particular selective pressures.

In contrast, cystatin divergence could not be transitory and promote the acquisition of permanent copies able to respond to different host cysteine proteases. Cystatin divergence could then be a response to selective pressures imposed by different cysteine proteases in a same host or different cysteine proteases in different hosts. In the first case, cystatin divergence would be probably involved in wasp specialization to one particular host whereas in the second case cystatine divergence would promote host shift for further specialization or enlarge wasp host range.

Indeed, in *Manduca sexta*, cysteine proteases co-localized with cystatins during parasitism. This suggests cystatin could effectively inhibit host cysteine proteases (Chapter V). Only one

cysteine protease was shown to be released but we can hypothesize that it could be the same for other cysteine proteases (Chapter V). Therefore viral cystatins would have to inhibit different cysteine proteases in a same host species. These data associated with the high host specificity of *Cotesia* wasps, suggest that cystatin copies divergence is directly a response to different targets in the same host species.

Can cystatins be implicated in wasp specialization?

Several lines of evidences show that *Cotesia* species present a very narrow host range as illustrated by *Cotesia congregata* which is only able to parasitize a few sphingid hosts (Harwood *et al.*, 1998). Indeed, even if the polydnavirus is a powerful instrument for host physiological disruption, parasitism success is largely dependent on host- wasp-virus genotype combinations. For example, the sphingidae *Manduca sexta* is permissive to *C. congregata* whose eggs successfully develop inside the host whereas the same host is able to encapsulate *C. flavipes* eggs which never succeed to complete their development (Rodríguez-Pérez *et al.*, 2005). Many examples emphasize the considerable genetic variation in parasitoid virulence and host resistance in endoparasitoid systems (Dubuffet *et al.*, 2006; Dupas *et al.*, 2003a; Dupas *et al.*, 2008). Interestingly, a recent study showed the correlation between the divergence of the CrV1 viral gene with host resistance to *C. sesamiae* virulent and avirulent strains and emphasized that genetic polymorphism could be related to local host resistance (Dupas *et al.*, 2008). Altogether, these examples suggest firstly that the intimate relationship could promote co-evolutionary dynamics, generating disruptive selective pressures associated with specialization on different host species and secondly that specialization could be related to symbiotic virus gene polymorphisms. In this context cystatin adaptive evolution could also be related to wasp adaptation to particular host species and promotes wasp specialization. This specialization and the intimate relationship between host and parasitoid could then trigger wasp diversification.

Could cystatins be involved in wasp diversification?

To answer this question it is first important to show that cystatins could be involved in wasp adaptation but also that cystatin genes and wasp share a common evolutionary history. Phylogenetic analyses reveal that diversification of particular cystatin lineages present common patterns with wasp speciation processes suggesting that at least these lineages could share a common evolutionary history with wasps. We can expect, according to virus transmission, that

viral cystatins co-diverge with wasp lineages, as cystatin genes are integrated into the wasp genome and are therefore transmitted in a mendelian fashion through wasp generations (Desjardins et al., 2007). Moreover, a previous co-phylogenetic study performed on certain *Cotesia* species did not show any evidence of horizontal transfer (Whitfield et al., 2003). However it is necessary to confirm by statistical analyses this co-diversification to ensure the close relation between virus and wasp.

Co-diversification between cystatins and wasp is expected because phyletic constraints imposed by wasp diversification would trigger viral genes divergence. However if cystatins are involved in wasp adaptation we can reasonably expect that cystatin divergence was not a consequence of wasp divergence but on the contrary that cystatin diversification precedes wasp speciation. By studying relative time divergence of both cystatin and wasp genes we will know if cystatin genes have effectively diverged before wasps.

Altogether, these analyses help us determine if a new gene acquisition can be related to diversification and will provide a strong evolutionary perspective to explain the important diversification observed in *Cotesia* wasps and more generally in the microgastroid complex.

VII- DISCUSSION GENERALE

Au cours de mon travail de thèse, nous nous sommes principalement intéressés aux facteurs de virulence, produits par un virus symbiotique impliqué dans une interaction hôte-parasitoïde. Nous avons étudié les processus qui ont dicté l'évolution de ces facteurs de virulence en mettant en évidence une forte composante adaptative qui souligne le rôle essentiel de ces facteurs au cours du parasitisme. Si le rôle des PTP a fait l'objet de plusieurs études (Falabella et al., 2006; Ibrahim et al., 2007; Ibrahim and Kim, 2008; Provost et al., 2004; Pruijssers and Strand, 2007), le rôle des cystatines est encore inconnu. Pour cela, nous avons voulu caractériser les cibles potentielles de cystatines chez l'hôte et étudier leur régulation suite à une infection et plus particulièrement au cours du parasitisme. Enfin, afin de comprendre si les facteurs de virulence des PDV pouvaient jouer un rôle dans l'adaptation et l'évolution des espèces, nous avons étudié les conséquences de l'acquisition des cystatines sur l'évolution des guêpes. Cette étude encore préliminaire, suggère que l'évolution des cystatines est étroitement liée à l'évolution des guêpes, et que l'évolution adaptative des cystatines aurait joué un rôle dans la spéciation des guêpes.

Dans ce dernier chapitre, les résultats obtenus au cours de cette thèse seront discutés autour de trois questions principales :

- Quelles sont les caractéristiques évolutives communes des gènes viraux étudiés et comment ces caractéristiques sont-elles être liées à leur statut de **facteurs de virulence** ?
- Quel est le **rôle** des facteurs de virulence dans le succès parasitaire ? Ces gènes ont-ils influencé l'**adaptation** du parasitoïde à son hôte, et contribué à sa **diversification** ?
- Ces gènes sont-ils engagés dans des processus **coévolutifs** avec leur cible chez l'hôte ? Comment ces gènes pourraient-ils évoluer au sein de l'interaction hôte-parasitoïde ?

A-Les facteurs de virulence : les processus évolutifs impliqués dans leur diversification

Nous avons étudié deux familles de gènes viraux codant pour des facteurs de virulence potentiels ; les protéines tyrosine phosphatases (PTP) et les cystatines.

Les PTP constituent une famille de gènes largement rencontrée chez les brucovirus ce qui suggère que ce gène a été acquis très tôt au cours de l'évolution du virus. De plus, plusieurs études fonctionnelles ont montré que ces gènes jouaient un rôle important dans l'altération du système immunitaire (Ibrahim *et al.*, 2007; Ibrahim and Kim, 2008; Pruijssers and Strand, 2007) et dans le contrôle du développement de l'hôte (Falabella *et al.*, 2006). Au cours de son évolution, ce gène a connu une très forte expansion pour former chez tous les brucovirus étudiés, la famille de gènes la plus importante.

Au contraire, les cystatines appartiennent à une classe de gène viraux apparue récemment au cours de l'évolution des brucovirus et qui, jusqu'à présent n'a été identifiée que chez deux genres de guêpe très proches d'un point de vue phylogénétique (Desjardins *et al.*, 2007; Espagne *et al.*, 2004). La fonction de ces gènes au cours du parasitisme reste encore méconnue, cependant ils sont exprimés très tôt, à un haut niveau et sont donc susceptibles d'intervenir au cours des premières étapes du parasitisme (Espagne *et al.*, 2005). L'expansion de ces gènes en famille multigénique est aussi très récente, en effet une seule copie est trouvée chez GiBV alors que 3 copies sont rencontrées chez CcBV (Desjardins *et al.*, 2007; Espagne *et al.*, 2004).

L'acquisition plus ou moins récente des gènes de virulence, est une donnée importante à considérer, puisque de cette date dépend la durée pendant laquelle ces gènes ont évolué avec la guêpe et plus largement au sein de l'interaction hôte-parasitoïde. Les processus évolutifs impliqués dans la divergence des PTP ont conduit à une famille de gènes où chaque membre est particulier et bien différencié, alors que les différentes copies de cystatines sont encore très similaires.

Dans la mesure où ces deux gènes viraux « coopèrent » en vue d'un objectif commun, à savoir le succès du parasitisme ; ces gènes présentent-ils des **patrons évolutifs similaires** ?

Au cours de ce travail de thèse, nous avons montré que l'évolution de ces gènes était dictée par la sélection naturelle. En effet les cystatines et les PTP sont caractérisés par une évolution soumise à une sélection positive. Ceci suppose que l'évolution de ces gènes n'est pas

soumise au hasard et que seules les mutations avantageuses vont être sélectionnées. Cependant, si l'évolution des PTP est caractérisée par des épisodes de sélection positive qui se sont en partie produits dans le passé, les processus adaptatifs qui dictent la divergence des cystatines sont encore actifs.

On peut toutefois noter que certains aspects de l'évolution des PTP sont tout à fait comparables à ce que l'on observe chez les cystatines. En effet, certaines PTP (PTPEX et PTP C α) présentent les mêmes patrons d'évolution que les cystatines ; les gènes isolés sont très peu divergents, ils sont présents en plusieurs copies et leur évolution est largement soumise à la sélection positive. On retrouve le même patron évolutif pour des gènes nouvellement acquis par le génome d'*E. coli*. En effet, ces gènes sont pour la plupart peu efficaces et leur amplification en de multiples copies permet en théorie d'augmenter la quantité de protéine produite (Hooper and Berg, 2003a).

Les processus évolutifs qui marquent l'évolution de ces **gènes nouvellement acquis**, sont probablement représentatifs des premières étapes de l'expansion des gènes en famille multigénique. Bergthorsson et ses collaborateurs (2007) qualifient ce mécanisme « d'amplification de gènes ». Ils proposent que cette amplification répond à des pressions de sélection qui agissent pour conserver plusieurs copies et permet ainsi d'augmenter la probabilité d'acquérir de nouvelles fonctions en multipliant les supports génétiques des mutations. Ce phénomène est particulièrement bien illustré par l'augmentation du nombre de gènes qui se produit dans une région du génome de *Plasmodium falciparum* impliquée dans la résistance aux traitements antipaludéens. Dans cette région, l'augmentation du nombre de copies de gènes est sélectionnée lorsque le parasite est soumis à un traitement alors qu'en l'absence de traitement les copies de gènes sont très rapidement perdues (Nair *et al.*, 2007). Ces observations suggèrent que le phénomène d'amplification est transitoire et répond à des pressions de sélection particulières et qu'en leur absence les copies de gènes sont probablement contre sélectionnées (Nair *et al.*, 2007).

Ainsi, ces exemples indiquent que les premières étapes de l'évolution d'une famille de gènes, qui se caractérisent par des phénomènes d'amplification de gènes et l'accumulation de mutations, s'effectuent en réponse à des pressions de sélection particulières.

Ces mécanismes d'amplification permettent d'expliquer les patrons de duplication observés aujourd'hui chez les cystatines et certaines PTP.

Les processus évolutifs qui dictent l'évolution des gènes codant pour les PTP ou les cystatines, -l'expansion en famille multigénique et la sélection adaptative-, doivent être mis en relation avec leur qualité de facteurs de virulence et le rôle de ces gènes dans **l'interaction hôte-**

parasitoïde. Par exemple, chez *Trypanosoma brucei*, parasite humain, de nombreux gènes impliqués dans la virulence sont organisés en famille multigénique et leur évolution est soumise à la sélection naturelle. Ces gènes codent pour des protéines majoritairement rencontrées à la surface des cellules ce qui suggère qu'elles sont en interaction avec l'environnement hôte (Emes and Yang, 2008). Cependant, si les gènes de virulence sont directement en interaction avec des molécules de l'environnement, il en est de même pour les molécules impliquées dans le système immunitaire qui doivent être capables de reconnaître les pathogènes afin de répondre de manière efficace aux infections. Ainsi, de nombreuses études ont montré que les protéines du CMH, impliquées dans les mécanismes de reconnaissance chez les vertébrés, forment une famille multigénique dont les membres évoluent également sous sélection positive (Axtner and Sommer, 2007; Darbo et al., 2008).

Contrairement à d'autres classes de gènes, les acteurs moléculaires impliqués dans une interaction hôte-pathogène subissent des pressions de sélection réciproques qui imposent aux partenaires de l'association une évolution dynamique. Les enjeux des interactions hôte-pathogène sur l'évolution des gènes impliqués dans l'association, sont particulièrement bien mis en évidence par l'étude des ARN interférents (ARNi) chez la Drosophile. Ces gènes sont impliqués dans la dégradation des ARN double brins, et une étude a montré que suivant leur rôle, ces gènes n'évoluaient pas de la même manière. En effet, les ARNi impliqués dans l'immunité antivirale évoluent très rapidement en comparaison de leurs homologues, dont le rôle est alloué à l'élimination des ARN endogènes, et qui sont très conservés (Obbard *et al.*, 2006). Ainsi, même si des molécules sont identiques du point de vue de leur structure et des mécanismes cellulaires dans lesquels elles sont engagées, leur évolution est directement liée à leur rôle physiologique et les pressions de sélection engagées dans une interaction hôte-pathogène façonnent de manière particulière l'évolution des gènes.

Cependant, les mécanismes de duplication et l'évolution adaptative des gènes impliqués dans une interaction ne répondent pas forcément de concert aux pressions de sélection. Une revue très intéressante a mis en évidence que les gènes impliqués dans l'immunité cellulaire ne présentaient pas les mêmes patrons évolutifs suivant leur rôle dans la réponse au pathogène. En effet, alors que l'évolution des peptides antimicrobiens (AMP) est principalement expliquée par des duplications, l'évolution des voies de signalisation, de type Imd par exemple, n'impliquent pas de duplication mais une évolution rapide des gènes (Lazzaro, 2008).

Ces résultats très surprenants sont certainement liés à la fonction de ces gènes dans la cascade de réactions mise en place pour répondre au pathogène. Les gènes codant pour les AMP sont organisés en familles de gènes dont l'évolution est très dynamique, puisqu'au sein d'une

même espèce le nombre de copies de gènes peut être très variable. Ces protéines agissent de manière non spécifique sur les pathogènes en détruisant les membranes cellulaires. Même si la résistance des pathogènes à ces attaques peut évoluer, elle est probablement limitée par la multiplicité des AMP et leur activité différente (Lazzaro, 2008).

Au contraire, les gènes codant pour les protéines NFκB, qui sont des facteurs de transcription de la voie Imd, ne sont pas organisés en famille multigénique mais présentent des taux de mutations importants. L'évolution rapide de ces gènes suggère qu'ils sont engagés dans une course aux armements avec les pathogènes capables d'inhiber directement ces facteurs de transcription (Apidianakis *et al.*, 2005; Thoetkiattikul *et al.*, 2005). L'évolution rapide des gènes NFκB imposerait au reste des acteurs moléculaires de la voie Imd d'évoluer en parallèle par des mutations compensatrices. C'est l'interdépendance de ces facteurs qui expliquerait l'évolution rapide observée sur l'ensemble de cette voie. L'évolution rapide des gènes impliqués dans la voie Imd serait une conséquence de l'évolution en parallèle de l'ensemble des acteurs qui fonctionnent de concert dans cette voie (Lazzaro, 2008).

La mise en évidence de ces différents patrons évolutifs dans les gènes impliqués dans une interaction est surprenante et montre que les mécanismes de l'évolution sélectionnés pour générer de la diversité ne seront pas les mêmes suivant le rôle et la fonction des protéines d'intérêt.

Dans notre système biologique, l'évolution des PTP et des cystatines est caractérisée par des duplications et une évolution des copies sous sélection positive. Comment peut-on associer ces patrons évolutifs avec la fonction de ces protéines dans l'interaction ? A quelles pressions de sélection ces protéines répondent-elles ? Pour véritablement répondre à ces questions, il est essentiel de décrire précisément le mode d'action de ces protéines et surtout les fonctions ciblées chez l'hôte.

B-Acquisition de nouvelles fonctions et adaptations

1-Le rôle des facteurs de virulence

Le séquençage de plusieurs génomes de bracovirus révèle que les gènes de virulence n'ont pas été acquis en même temps au cours de l'évolution de l'association (Desjardins et al., 2007; Espagne et al., 2004; Gundersen-Rindal and Pedroni, 2006; Lapointe et al., 2007; Webb et al., 2006) (figure 2, Chapitre I). Cependant, il semble que la majorité des bracovirus présentent très tôt un ensemble de gènes codant pour les PTP, les protéines IκB et les CRP. On peut ainsi penser que ces gènes font partie de l'arsenal de base essentiel à la réussite parasitaire dans une larve de Lépidoptère. En revanche, certains gènes ne sont présents que dans certaines lignées de bracovirus, c'est le cas par exemple du gène *glc1.8* chez MdBV, TnBV1 chez TnBV ou encore des gènes codant les cystatines présents chez GiBV et chez les bracovirus associés aux guêpes du genre *Cotesia* (figure 24). La plupart de ces gènes interviennent pourtant dans le parasitisme en contrôlant des fonctions physiologiques particulières de l'hôte (Beck and Strand, 2003, 2005; Lapointe et al., 2005).

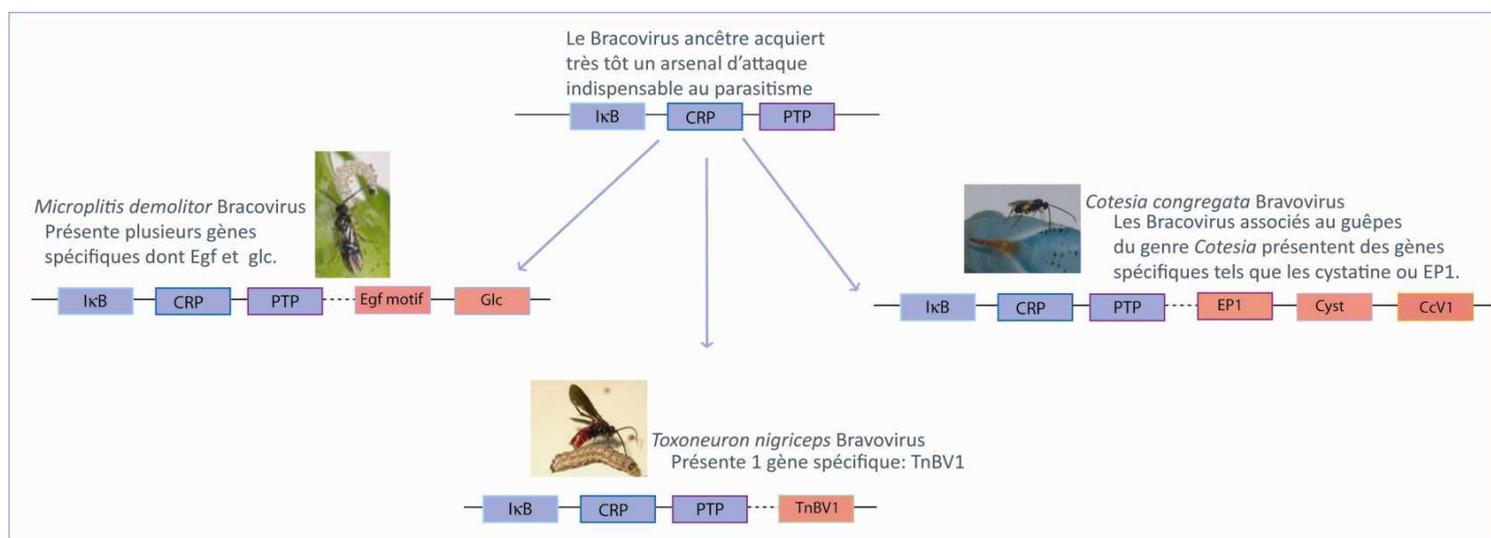


Figure 22 Certains gènes de virulence sont très conservés au sein des Bracovirus (IκB, PTP, CRP) alors que d'autres facteurs présentent une distribution étroite et sont spécifiques de certains Bracovirus.

a) *Dans quelles fonctions, les gènes spécifiques des lignées de bracovirus peuvent-ils être impliqués ?*

Parmi l'arsenal d'attaque de base trouvé chez la plupart des bracovirus, les gènes codant pour les PTP sont probablement apparus il y a plus de 50 millions d'années et depuis leur intégration, ces gènes ont été transmis et ont connu une très forte expansion. De plus, ces gènes se retrouvent aussi chez les PDV associés aux guêpes de la sous famille des Banchinae et forment une lignée évolutive indépendante suggérant que ces gènes ont été acquis de manière indépendante à 2 reprises au cours de l'évolution des PDV. Ces éléments suggèrent que ces gènes doivent cibler des fonctions conservées chez les hôtes.

Les travaux réalisés sur les PTP ont en effet montré que ces protéines ciblent des fonctions clés des défenses de l'hôte en inhibant la phagocytose (Pruijssers and Strand, 2007), en induisant l'apoptose des cellules de l'immunité (Suderman *et al.*, 2008) ou encore en contrôlant le développement de l'hôte (Falabella *et al.*, 2006). L'ensemble de ces travaux suggèrent fortement que les PTP régulent des voies de signalisation impliquées dans l'immunité comme dans le développement et jouent un rôle essentiel dans le succès parasitaire.

Au contraire, d'autres gènes comme les cystatines sont apparus il y a 15 à 20 millions d'années (Whitfield, communication personnelle) et ne sont présents que chez certaines lignées de guêpes. Nous savons que les gènes codant pour les cystatines sont exprimés très tôt au cours du parasitisme et à un niveau élevé (Espagne *et al.*, 2005). De plus, les sites d'interaction des cystatines avec les protéases à cystéine sont fortement sélectionnés (Chapitre IV). L'ensemble de ces caractéristiques suggèrent fortement que le rôle des cystatines est important et qu'elles pourraient intervenir dans les premières étapes du parasitisme. Si les cystatines sont effectivement des protéines essentielles au parasitisme, pourquoi ne présentent-elles pas une distribution plus large au sein des bracovirus ?

Plusieurs hypothèses peuvent être émises quant à l'étroite distribution de certains gènes viraux :

Hypothèse 1 : Ces gènes ciblent des fonctions non conservées chez les espèces hôtes.

Ainsi, les guêpes porteuses de virus codant pour des cystatines auraient colonisé une nouvelle ressource et se retrouveraient en majorité sur un type d'hôte particulier.

Les genres *Cotesia* et *Glyptapanteles*, formés de guêpes associées à des bracovirus dont le génome code pour des cystatines, sont très diversifiés et parasitent des hôtes de familles différentes. Il semble peu probable que des hôtes de familles différentes présentent des caractéristiques communes entre elles et absentes chez d'autres lépidoptères. Récemment, une étude du

transcriptome de *Manduca sexta* après une infection bactérienne, montre une grande conservation des acteurs de l'immunité avec les autres insectes étudiés (*Drosophila melanogaster*, *Bombyx mori* et *Tribolium castaneum*) (Zou et al 2008). Pour tester cette hypothèse il faudrait cependant avoir une meilleure connaissance de la phylogénie des hôtes et de leur système de défense.

Hypothèse 2 : Ces gènes n'agissent pas seuls mais en synergie avec d'autres facteurs de virulence et leur action permettrait une nette amélioration du succès parasitaire. C'est le cas de la protéine glc1.8 présente chez MdBV qui agit en synergie avec la protéine PTP-H2 et qui est indispensable à une inhibition efficace de la phagocytose (Beck and Strand, 2003, 2005; Pruijssers and Strand, 2007).

Hypothèse 3 : Ces gènes codent des protéines de séquences différentes et aux fonctions biochimiques particulières, mais elles ciblent les mêmes voies de régulation chez l'hôte. Ces protéines pourraient également cibler différents acteurs de l'immunité ou du développement qui agissent chez l'hôte en synergie. Par exemple, les protéines glc1.8 chez MdBV et CcV1 chez CcBV ont pour effet d'inhiber la phagocytose mais en agissant à différents niveaux sur les hémocytes (Beck and Strand, 2003, 2005; Vassiliki Labropoulou, 2008).

Cependant, les hypothèses 2 et 3 ne sont pas exclusives l'une de l'autre. En effet ces facteurs peuvent cibler un niveau particulier d'une cascade de réaction et agir en synergie avec d'autres facteurs pour un contrôle efficace de la fonction ciblée. Ces hypothèses supposent également que l'action de ces gènes dépend probablement d'un contexte biochimique particulier matérialisé par l'hôte lépidoptère mais aussi par l'association virus-guêpe à un moment précis de leur histoire évolutive. En effet, si ces gènes peuvent cibler des fonctions similaires, le fait qu'ils soient différents et non conservés au sein des bracovirus, suggère qu'ils ne sont efficaces que dans un contexte biochimique particulier qui est fonction des pressions de sélection agissant à un moment précis de l'évolution de l'association. Dans le cas des cystatines, qui sont des inhibiteurs de protéases à cystéine, cela signifierait que les protéases chez l'hôte exerceraient de telles pressions que le gène cystatine acquis à un moment donné de l'évolution a été conservé au sein de lignées particulières de Bracovirus.

b) A quelles pressions de sélection ces gènes doivent-ils répondre?

Nous discuterons cette question en prenant le cas particulier des cystatines.

Pour comprendre l'implication des cystatines au cours du parasitisme, nous avons isolé des protéases à cystéine chez *Manduca sexta* et étudié leur régulation (chapitre V). Nous avons

montré que des protéases particulières, MsCath1 et MsCath2 étaient préférentiellement ciblées lors d'une infection et particulièrement au cours du parasitisme. En effet, l'expression de ces protéases, de type 26/29kDa, est significativement inhibée 18h après le parasitisme mais aussi après une infection bactérienne. Nous avons également montré que MsCath1 était libérée dans l'hémolymphe au cours du parasitisme et pouvait potentiellement être ciblée par les cystatines virales. Nous ne sommes pas parvenus à mettre en évidence *in vivo* l'interaction moléculaire entre des protéases à cystéine et les cystatines virales, cependant nous avons montré que le parasitisme induisait une augmentation de l'activité des protéases à cystéine qui était régulée négativement 48h après le parasitisme.

Ces résultats convergent dans le sens d'une inactivation des protéases à cystéine au cours du parasitisme et soulignent le rôle clé que ces protéases doivent jouer lors d'une infection de ce type. Si ces résultats ne permettent pas d'identifier la fonction remplie par les protéases ; ils permettent par contre, de formuler des hypothèses solides pour envisager la ou les potentielle(s) fonction(s) de ces protéases.

Aux vues de nos résultats et de ce qui est connu sur la fonction des protéases dans les autres systèmes, quelles fonctions pourrait-on attribuer aux protéases dans le contexte du parasitisme ?

La libération des protéases à cystéines dans le milieu extracellulaire a été décrite dans plusieurs situations qui peuvent nous apporter des éléments permettant de comprendre ce qui se passe au cours du parasitisme.

En effet chez les mammifères, la libération de protéases a été décrite soit au cours de la réponse immunitaire soit dans des situations pathologiques lors de la propagation de tumeurs invasives. Dans les deux cas cette libération conduit à une dégradation plus ou moins contrôlée de la membrane basale des tissus (Dickinson, 2002).

Chez les insectes, ces enzymes sont essentiellement décrites pour leur rôle dans le développement. Dans ce cas ces protéases sont libérées et digèrent de façon contrôlée les matrices extracellulaires qui sont remodelées au cours de l'embryogenèse, de la métamorphose ou au cours de la mue (Cho *et al.*, 1999; Hegedus *et al.*, 2002; Homma *et al.*, 1994; Liu *et al.*, 2006). Cependant, chez les insectes ces protéases peuvent aussi être libérées dans le milieu extracellulaire à la suite d'une infection par un pathogène (Levy *et al.*, 2004; Song *et al.*, 2008) ou d'un corps étrangers (Saito *et al.*, 1992) mais les conséquences de cette libération sont encore inconnues.

Récemment, plusieurs études ont montré une relation entre la libération de cathepsine L dans l'hémolymphe et l'induction du processus de mélanisation. En effet, un baculovirus recombinant exprimant une cathepsine L et introduit dans *Heliothis virescens*, induit une fragmentation des tissus suivie de leur mélanisation, avant de provoquer une mort précoce du lépidoptère (Harrison and Bonning, 2001). Dans cette étude, il est supposé que la surexpression de cathepsine L, induit une dégradation de la membrane basale et de cette manière recrute les hémocytes sur le lieu de dégradation et active la mélanisation (Li *et al.*, 2008; Pech *et al.*, 1995).

On peut ainsi proposer l'hypothèse suivante : l'inhibition des protéases à cystéine libérées lors du parasitisme par *C. congregata* permettrait de contrôler l'amplification de la réponse immunitaire et de maintenir en vie la larve de lépidoptère pendant le développement des guêpes.

On peut envisager de tester cette hypothèse chez *M. sexta* par différentes approches :

- 1- Est-ce que la libération des protéases est une réponse spécifique au parasitisme ou est-elle une réponse générale dans un contexte pathogène ?
- 2- La libération des protéases est-elle sous la dépendance de mécanismes passifs ou actifs ?
- 3- Quel est l'effet de la surexpression de ces protéases dans l'hôte?
- 4- Quelles sont les conséquences de l'inhibition de ces protéases?

Dans chacune de ces situations, il serait nécessaire de comparer des situations de pathologie (Parasitisme par *C. congregata*, injection de bactéries pathogènes de types *Photobacterium*, injection de virus pathogènes) avec des situations où l'hôte est capable de se défendre (Parasitisme par *C. flavipes*, injection d'*E.coli*, injection de baculovirus) (Tableau 11).

Table 11 Perspectives expérimentales pour identifier le rôle des protéases à cystéine au cours du parasitisme.

| Tests | Méthodes |
|--|---|
| 1) La régulation des protéases est-elle spécifique au parasitisme ou est-elle une réaction non spécifique ? | Injection de pathogènes bactériens et de pathogènes viraux. Suivi de la régulation transcriptionnelle et de la libération des protéases dans l'hémolymphe |
| 2) La libération des protéases dans l'hémolymphe est-elle passive ou active ? | Suivi de la localisation d'une protéine intracellulaire dans l'hémolymphe au cours du parasitisme par western blot |
| 3) Quel est l'effet de la surexpression des protéases ? | Surexpression transitoire des protéases par l'injection de vecteur d'expression recombinants ou de protéines purifiées dans des conditions d'infection par un pathogène ou non (voir plus haut). Observations : - étude de la survie et du niveau de mélanisation des tissus de l'hôte; - étude de la survie et du niveau de mélanisation du pathogène ; - étude de la croissance du pathogène ou du succès de l'infection |
| 4) Quel est l'effet de l'inhibition des protéases ? | Inhibition des protéases par RNAi ou par injection d'inhibiteurs dans des conditions d'infection par un pathogène ou non (voir plus haut). Observations : - étude de la survie et du niveau de mélanisation des tissus de l'hôte; - étude de la survie et du niveau de mélanisation du pathogène ; - étude de la croissance du pathogène ou du succès de l'infection |

2-Conséquences sur l'adaptation et la diversification du parasitoïde

Le virus est essentiel au succès parasitaire de la guêpe parasitoïde ; nous pouvons supposer que l'acquisition de nouveaux gènes au cours de son évolution a permis la colonisation de nouveaux hôtes et joué un rôle dans les processus de diversification des guêpes.

L'acquisition d'un gène peut-elle suffire à la colonisation de nouvelles niches et à la diversification des espèces ?

Chez les singes de l'ancien monde, le régime alimentaire des membres de la famille des Colobinae est constitué de feuilles alors que chez les autres singes les fruits et les insectes dominant. On trouve chez ces primates « ruminants » un gène codant pour une enzyme de digestion, la RNASE 1B, issue d'un phénomène de duplication de la RNASE 1 qui est elle, présente chez tous les primates. Ce gène a connu une évolution rapide et code pour une protéine aux propriétés biochimiques différentes de la RNASE 1, la rendant plus efficace dans des conditions de pH particulières compatibles avec une alimentation à base de feuilles. L'acquisition de ce gène via le phénomène de duplication est étroitement liée à la radiation adaptative des Colobinae suggérant un lien de cause à effet entre ces deux processus (Zhang *et al.*, 2002).

Les exemples permettant de lier l'acquisition d'un gène aux processus de radiation des espèces sont encore très rares et les preuves de ce lien sont souvent indirectes. Cependant, d'après l'exemple précédent, l'acquisition d'un gène peut suffire à expliquer la colonisation de nouvelles niches et la diversification des espèces.

Chez les guêpes Braconides, l'évolution des guêpes du complexe microgastroïde est caractérisée par une diversification très rapide (Banks and Whitfield, 2006; Murphy et al., 2008). La diversification rapide des guêpes et leur spécificité d'hôte suggèrent que la radiation des guêpes est une réponse à la radiation des espèces hôtes (Banks and Whitfield, 2006; Smith et al., 2008).

Le virus est essentiel au succès du parasitisme ; est il pour autant impliqué dans la spécificité de la relation entre le parasitoïde et son hôte ?

Pour répondre à cette question, nous voulions évaluer si les cystatines virales dont l'acquisition par la symbiose est probablement contemporaine de l'apparition du genre *Cotesia*, ont pu jouer un rôle dans l'adaptation des guêpes à leur hôte et participer à leur diversification.

Les premiers résultats de cette étude, encore très préliminaire, montrent que la divergence des cystatines virales est principalement expliquée par une évolution adaptative. Ces pressions de sélection ont agi avant la spéciation des guêpes, suggérant que la divergence des cystatines est une composante nécessaire à la spéciation des guêpes (Chapitre VI). De plus, dans le chapitre IV, nous avons vu que la sélection agissait préférentiellement au niveau des sites d'interaction avec les protéines cibles, ces résultats supposent que l'évolution des cystatines est étroitement liée à leur interaction avec les protéases à cystéine. Le fait que ces pressions restent actives après les processus de spéciation suggère que l'adaptation des espèces de guêpes à leur hôte est un processus dynamique au sein duquel les acteurs moléculaires de l'interaction subissent des pressions de sélection réciproques.

L'expansion en famille multigénique et la divergence des copies de gènes pourraient fournir en théorie à la guêpe les outils nécessaires à la colonisation de nouvelles niches écologiques.

Pour conforter cette hypothèse il est toutefois nécessaire de savoir si les cystatines et les guêpes partagent la même histoire évolutive. De plus, en datant de manière relative les événements de diversification des cystatines et des guêpes, nous pourrions tester si la divergence des cystatines est antérieure à la divergence des guêpes et évaluer ainsi le rôle potentiel joué par les cystatines dans l'évolution des guêpes.

Cette étude permettra de comprendre si l'acquisition d'un nouveau gène chez les guêpes parasitoïde peut être corrélée à une diversification et d'évaluer de manière plus générale la contribution du virus dans l'extraordinaire expansion des guêpes du complexe microgastroïde.

C- Le parasitisme : une course aux armements

Les pressions de sélection impliquées dans les relations hôte-parasitoïde sont considérables étant donné que l'issue de l'interaction se solde par la mort d'au moins un des partenaires. Chez les endoparasitoïdes, ces pressions sont d'autant plus grandes que le parasite se développe à l'intérieur de l'hôte et doit ainsi faire face aux acteurs du système immunitaire. Ce sont vraisemblablement ces pressions de sélection qui ont généré la diversité des stratégies parasitaires que l'on observe chez les endoparasitoïdes. Ces systèmes d'attaque comprennent la formation de cellules spécialisées de guêpes qui circulent dans l'hémolymphe et jouent un rôle important important dans la nutrition des larves de guêpes (tératocytes), les associations symbiotiques avec les polydnavirus ou encore la grande diversité des facteurs de virulence contenus dans les venins (Pennacchio and Strand, 2006). Ces stratégies ne sont pas exclusives et peuvent fonctionner en synergie.

En outre, la diversité des stratégies parasitaires employées par les endoparasitoïdes peut s'observer à des échelles évolutives très fines comme l'illustre le cas de *Leptopilina boulardi* parasite de *Drosophila melanogaster* chez qui les variations de stratégie sont observées entre les populations (Dupas et al., 2003a) et sont liées aux propriétés qualitatives et quantitatives des glandes à venin (Labrosse et al., 2005a; Labrosse et al., 2005b). Ces variations subtiles de virulence constituent probablement une réponse aux variations des systèmes de résistance chez l'hôte. Effectivement, dans ces systèmes on observe également un polymorphisme de résistance chez l'hôte dont le déterminisme est génétique (Dubuffet et al., 2006).

Considérant les fortes pressions de sélection engagées dans l'interaction hôte-parasitoïde, nous pouvons penser que la diversité de stratégies parasitaires et le polymorphisme de résistance, sont le reflet d'une course aux armements engagée entre le parasitoïde et son hôte.

Dans ce cas, la relation intime établie entre le parasitoïde et son hôte implique des pressions de sélection réciproques ; où des changements évolutifs observés chez l'un des partenaires entraînent des changements évolutifs chez l'autre (Van Valen, 1973). Ces évolutions réciproques, résumées sous le terme de co-évolution, sont des mécanismes évolutifs dynamiques qui supposent par exemple qu'un gène peut être efficace à un moment et dans un environnement

particuliers et peut au contraire être inefficace dans d'autres conditions (Woolhouse *et al.*, 2002). Les théories génétiques de la coévolution, comme la théorie du « gène pour gène » ou la théorie du « matching genotype », supposent que l'issue d'une interaction hôte-pathogène dépend de la combinaison des génotypes de l'hôte et du parasite. Par exemple, chez les plantes, la théorie du gène pour gène prédit qu'un gène R (résistance à un pathogène) codant pour un récepteur reconnaît spécifiquement les produits d'un gène Avr (avirulent) chez le pathogène. De manière réciproque, des modifications du gène Avr empêchent sa reconnaissance par les récepteurs de la plante et permettent au pathogène d'éviter la résistance par la plante (Stahl and Bishop, 2000). Des phénomènes d'épistasie et l'effet de l'environnement (facteurs biotiques et abiotiques) contribuent aux variations observées dans ces interactions (Wade, 2007).

En regard de la spécificité d'hôtes des guêpes endoparasitoïdes et des patrons évolutifs des gènes viraux, nous pouvons penser que les gènes de virulence des PDV sont eux aussi engagés dans des processus coévolutifs avec leur partenaire chez l'hôte.

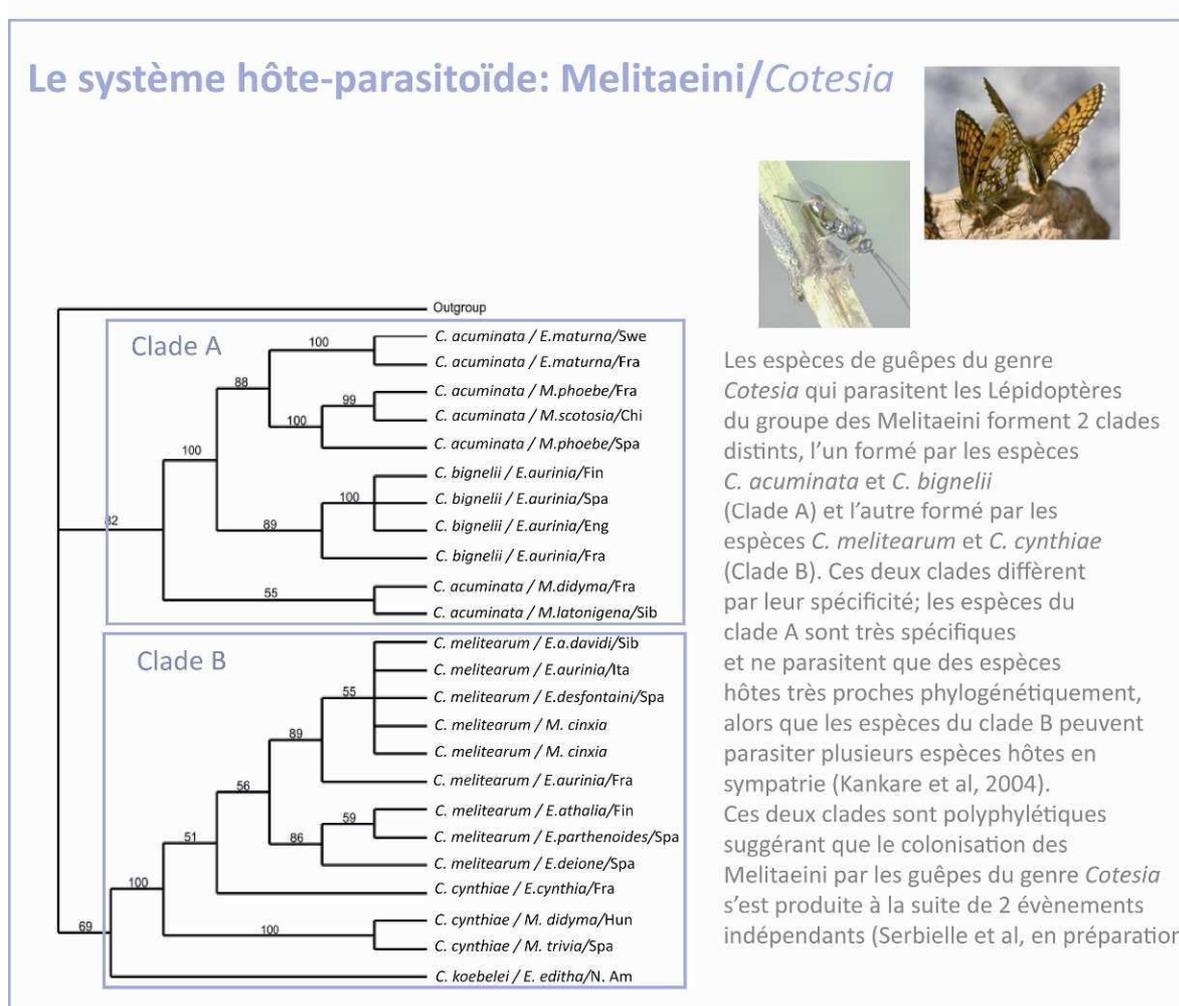
Aucune étude ne permet encore de mettre en évidence cette coévolution, cependant, comme cela a été observé chez *L. boulandi*, il est aussi possible d'observer chez les guêpes porteuses de PDV des variations de virulence. En effet, il existe deux souches de *C. sesamiae* dont l'une dite virulente est capable de parasiter l'hôte lépidoptère *Busseola fusca*, alors que la souche dite avirulente en est incapable. Cette variation de virulence intra-spécifique est associée à des variations du gène viral CrV1 (Dupas *et al.*, 2008). Cette étude suggère, d'une part que la relation associant *C. sesamiae* à son hôte est très spécifique et sous entend que l'hôte et son parasitoïde sont engagés dans des processus co-évolutifs.

L'évolution rapide des gènes cystatines notamment au niveau de sites actifs suggère que ces gènes sont engagés dans des processus de coévolution avec leur cible chez l'hôte. Il serait ainsi très intéressant de tester l'hypothèse de coévolution, en étudiant les coadaptations entre les protéines impliquées dans l'interaction et en considérant le contexte historique au sein duquel ces gènes évoluent.

Pour tester cette hypothèse il est nécessaire d'avoir une bonne connaissance du modèle biologique. Le système hôte-parasitoïde Melitaeini-*Cotesia*, est un modèle très intéressant dans le sens où il associe des guêpes ainsi que des hôtes proches phylogénétiquement, ce qui facilite l'isolement des gènes. De plus, une étude en cours, montre que les guêpes parasitant les lépidoptères de la famille des Melitaeini, forment 2 clades distincts et polyphylétiques, suggérant que la colonisation des Melitaeini par les guêpes parasitoïdes s'est produite à la suite de 2

événements indépendants (Serbielle, en préparation). Cette association est très diversifiée, représentée par de nombreuses espèces hôtes et parasitoïdes, mais elle est aussi très spécifique puisque le spectre d'hôte des guêpes est limité à nombre restreint d'espèces (encadré 3).

Une étude de coévolution à l'échelle des espèces et à l'échelle des gènes impliqués dans l'interaction semble particulièrement intéressante dans ce système. Cette étude, permettra de comprendre les processus de spéciation impliqués dans ces associations et leur influence réciproque mais aussi les processus de coadaptations potentiels entre les gènes directement impliqués dans l'interaction. De plus, il sera aussi intéressant d'étudier si la colonisation des hôtes par deux lignées différents de guêpes a conduit à une évolution convergente chez les cystatines.



Les espèces de guêpes du genre *Cotesia* qui parasitent les Lépidoptères du groupe des Melitaeini forment 2 clades distincts, l'un formé par les espèces *C. acuminata* et *C. bignellii* (Clade A) et l'autre formé par les espèces *C. melitearum* et *C. cynthiae* (Clade B). Ces deux clades diffèrent par leur spécificité; les espèces du clade A sont très spécifiques et ne parasitent que des espèces hôtes très proches phylogénétiquement, alors que les espèces du clade B peuvent parasiter plusieurs espèces hôtes en sympatrie (Kankare et al, 2004). Ces deux clades sont polyphylétiques suggérant que la colonisation des Melitaeini par les guêpes du genre *Cotesia* s'est produite à la suite de 2 événements indépendants (Serbielle et al, en préparation)

Encadré 3 Un modèle intéressant pour une étude des processus co-évolutifs engagés dans l'interaction hôte-parasitoïde

