Rôle et évolution de facteurs de virulence impliqués dans une interaction hôte-parasitoïde

Céline Serbielle

▶ **To cite this version:**

Céline Serbielle. Rôle et évolution de facteurs de virulence impliqués dans une interaction hôte-parasitoïde. Ecologie, Environnement. Université François Rabelais - Tours, 2008. Français. <tel-00402832>

HAL Id: tel-00402832
https://tel.archives-ouvertes.fr/tel-00402832

Submitted on 8 Jul 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ FRANÇOIS - RABELAIS
DE TOURS

PUCVL
Pôle Universités
Centre Val de Loire

UNIVERSITÉ FRANÇOIS - RABELAIS TOURS

ÉCOLE DOCTORALE : Santé, Sciences, Technologies

INSTITUT de RECHERCHE sur le BIOLOGIE de l'INSECTE

THÈSE présentée par :

Céline SERBIELLE

soutenue le : **8 décembre 2008**

pour obtenir le grade de : **Docteur de l'université François - Rabelais**

Discipline/ Spécialité : Sciences de la vie

Rôle et évolution de facteurs de virulence impliqués dans une interaction hôte-parasitoïde

THÈSE dirigée par :

Mr DREZEN Jean-Michel	Chargé de recherche, CNRS, Tours
et co-encadrée par :	
Mme HUGUET Elisabeth	Maître de conférences, Université de Tours

RAPPORTEURS :

Mr CASANE Didier	Professeur, Université de Paris 7
Mr MITTA Guillaume	Professeur, Université de Perpignan

JURY :

Mr CASANE Didier	Professeur, Université de Paris 7
Mr DREZEN Jean-Michel	Chargé de Recherche, CNRS, Tours
Mme HUGUET Elisabeth	Maître de conférences, Université de Tours
Mr JIGGINS Francis	Professeur, Université d'Edimbourg, Ecosse
Mr MITTA Guillaume	Professeur, Université de Perpignan
Mme POIRIE Marylène	Professeur, Université de Nice-Sophia Antipolis
Mr RASSCHAERT Denis	Professeur, Université de Tours

Des mercis…

Je voudrais en premier lieu m'adresser aux membres du jury: Didier Casane, Francis Jiggins, Guillaume Mitta, Marylène Poirié et Denis Rasschaert pour les remercier d'avoir accepté de lire et évaluer ce travail.

Je remercie également mes directeurs de thèse Elisabeth Huguet et Jean-Michel Drezen sans qui ce travail n'aurait pu voir le jour.

Elisabeth, tu es une personne d'exception, merci pour ton encadrement sans relâche, ton engouement et ta gentillesse. Tes grandes qualités de scientifique et de pédagogue m'ont amené à toujours plus de rigueur et de précision dans mon travail. Que d'histoires racontées et racontées encore dont les personnages principaux, gènes ou protéines, me paraissaient si lointains il y a à peine trois ans.

Jean-Michel, nos discussions m'ont constamment enrichie de nouvelles connaissances ; Grand scientifique et passionné, je vante vos qualités de directeur de recherche mais il faut quand même que vous compreniez que l'on ne peut pas tout faire !

Merci à vous deux d'avoir eu le désir de m'accompagner tout au long de cette thèse ; je suis très heureuse d'avoir partagé ces trois années avec vous et très fière de vous avoir eu comme directeurs.

Bien sûr je remercie l'ensemble de l'équipe génome et stratégie parasitaire, ce fut un grand plaisir de travailler avec vous.

Catherine, je suis ravie de t'avoir connue et d'avoir partagé ton bureau… Au moins pendant le temps que tu ne passais pas à repeupler la planète. Certains disent que les français n'ont pas beaucoup d'autodérision, avec toi il vaut mieux en avoir !

Annie, merci pour tes précieux conseils et, malgré la montagne de travail que tu dois abattre, d'accorder de ton temps pour aider tes collègues en détresse. Toutes mes félicitations pour ton poste, Elisabeth ne s'est pas trompée en te recrutant.

Georges, vos petites attentions pour chacun et votre don d'adoucir les esprits sont remarquables.

Sébastien, merci d'avoir toujours prêté une oreille attentive à mes questions, de t'être intéressé à mon travail et d'avoir agrémenté le tout d'une bonne dose d'humour.

Juline, ta gentillesse, tes sourires et ton zen sont des denrées rares.

Faustine, encore toutes mes félicitations pour ton allocation, tu ne vas pas le regretter et je te souhaite d'apprécier tes trois années de thèse autant que j'ai pu apprécier les miennes.

Aurore, cela serait un plaisir si tu pouvais assister à ma thèse, je te montrerai que les basques ne pleurent pas. Maxime, ma thèse est terminée, je dois m'y résoudre, je n'ai pas changé le cours de l'histoire…

Bruno, Myriam et Karine, j'espère que vous allez vous plaire dans votre nouvelle équipe mais je ne me fais pas de soucis.

Merci à Carole et Cindy de nous permettre de travailler sur ce modèle comme on le désire et de nous faciliter le travail.

Je tiens aussi à remercier toutes les personnes qui ont participé aux différentes parties de ce travail.

Merci à Stéphane Dupas, d'avoir toujours pris le temps de répondre à mes questions et ce même à des milliers de kilomètres.

Merci à Gilles Lalmanach pour son enthousiasme et ses précieux conseils de biochimiste, je dois maintenant vous avouer que je n'ai pas toujours tout suivi… Merci d'avoir associé Forian Veillard à ce travail qui malgré sa thèse, a été d'un grand soutien pour l'acquisition des données et leurs analyses.

Un grand merci à Elisabeth Herniou, pour m'avoir accueillie à l'Imperial College et d'avoir rendu ces séjours très agréables. Merci d'avoir accepté de répondre avec gentillesse et précision à mes innombrables mails. Grâce à toi, j'ai pu travailler avec Donald Quicke qui m'a beaucoup appris notamment la rigueur, et Mark Shaw qui m'a enseigné tout ce que je sais sur les *Cotesia.*

Enfin ce travail ne serait pas ce qu'il est si Emilien Volvoire et Maria Teresa Claro n'avaient pas mis la main à la pâte

Je remercie bien entendu toutes les personnes qui m'ont accordé un peu de leur temps pour relire ce manuscrit : Franck, Catherine et Marie.

Pour terminer, je remercie l'IRBI et en premier lieu Jérome Casas et Jean-Paul Monge pour m'y avoir accueillie et pour leur gestion admirable de cet institut pas comme les autres.

Un merci tout particulier à tous les membres de cet institut qui ont contribué au plaisir que j'ai eu à venir travailler chaque jour. Christine, merci pour votre efficacité, et Sonia, merci d'avoir accepté avec gentillesse et patience de faire le tri dans mes missions…encore pardon de ne jamais t'avoir facilité la tâche.

Enfin l'IRBI ne serait pas ce qu'il est à mes yeux si Arnold, Wilfried, Sylvain, Thomas, David, Elfie, Laurianne, Aurélie, Fabienne, Appolline, Franck, Marie, Simon et Nadine n'en faisaient pas partie. Merci pour nos soirées diverses et variées. Je pense aussi aux « assimilés » IRBI ; Aniça, François, Hervé et ma petite préférée Rebecca. Merci pour votre gentillesse, vos singularités, vos engagements et bien sûr merci de me faire rire. Je passe un message personnel à Fabienne : mission accomplie, nous avons gardé nos « on », nos « in » et nos « an », nous pouvons être fières de nous !

Laurianne, ce fut un grand plaisir de partager avec toi cette dernière année dans notre petit palace. Tu as été là pour cette fin de thèse pas toujours facile, merci de m'avoir épaulée, conseillée, rassurée…et surtout nourrie !

Enfin, je rends hommage aux amis de longue et de moins longue date pour ce qu'ils sont tout simplement. Vivement qu'on fête ça ensemble !

Mention spéciale à ma famille, mes parents qui se débrouillent toujours par je ne sais quel miracle à me remettre sur pied, à mon frère pour son soutien depuis les quatre coins du globe et mes sœurs Nathalie et Vanessa qui me laissent jouer à la grande sœur. J'associe tout naturellement Cécile et Anelise mes autres sœurs comme on dit !

Mes pensées s'adressent à Frédo…celui qui sait…

Résumé

Les associations mutualistes, en permettant l'acquisition de nouvelles fonctions, ont joué un rôle majeur dans l'évolution des espèces. Pour comprendre en quoi ces associations sont impliquées dans l'adaptation des espèces nous avons étudié un cas unique de mutualisme associant un virus de type polydnavirus et une guêpe parasitoïde. Dans cette association, le virus est injecté dans l'hôte lépidoptère lors de l'oviposition et joue un rôle majeur dans le succès parasitaire en induisant une altération des fonctions physiologiques de l'hôte. En regard du nombre d'espèces de guêpes caractérisées par cette association, le virus doit constituer une innovation adaptative majeure et jouer un rôle déterminant dans l'évolution et la diversification des espèces de guêpes.

En quoi cette association joue t-elle un rôle déterminant dans l'évolution et l'adaptation des guêpes ? Quelles sont les fonctions physiologiques ciblées chez l'hôte ?
Pour répondre à ces questions nous avons étudié l'évolution de deux familles de gènes codant pour des facteurs de virulence potentiels et nous avons exploré les fonctions physiologiques d'une protéine potentiellement ciblée lors du parasitisme.

Nous avons mis en évidence le rôle important de la sélection naturelle dans l'évolution des familles de gènes viraux. Par modélisation de la structure tridimensionnelle d'un facteur de virulence codant pour des cystatines, nous avons montré que cette sélection agissait préférentiellement au niveau des sites d'interaction avec les protéines cibles. De plus, cette étude souligne le caractère dynamique de l'évolution des facteurs de virulence incluant de multiples évènements de duplication, caractérisés par des processus de perte et d'acquisition au cours de l'évolution de l'association. Le caractère adaptatif et dynamique de l'évolution des gènes viraux a aussi été étudié en regard de l'évolution des espèces de guêpes et de leur spectre d'hôte.
Par une approche fonctionnelle, nous avons étudié le rôle physiologique de protéases à cystéine qui constituent des cibles potentielles des cystatines virales. Nous avons montré que ces protéases sont régulées spécifiquement au cours du parasitisme au niveau protéique et transcriptionnel. Nous avons également montré que l'activité de ces protéases est modifiée après parasitisme.

L'évolution adaptative et dynamique des facteurs de virulence reflètent leur rôle important dans le parasitisme. Il reste maintenant à montrer comment ces facteurs interagissent sur la physiologie de l'hôte lépidoptère. Des protéases à cystéine sont spécifiquement ciblées par le parasitisme, en étudiant les mécanismes d'interaction de ces protéases avec les cystatines virales et les processus coévolutifs mis en jeu, nous pourrions comprendre comment le virus intervient dans l'adaptation de la guêpe à son hôte.

Mots clés: polydnavirus, guêpe parasitoïde, cystatine, protéine tyrosine phosphatase, protéases à cystéine, évolution moléculaire.

Table des matières

Des mercis... 1

RESUME ... **4**

TABLE DES MATIERES ... **6**

I-INTRODUCTION ... **10**

A-SYMBIOSES ET CONSEQUENCES ADAPTATIVES ... 14

1-La symbiose : une interaction durable ... *14*

2-Symbiose et acquisition de nouvelles fonctions ... *17*

3-Symbioses et conséquences sur l'évolution des génomes ... *18*

B- UN MUTUALISME VIRUS-GUEPE PARASITOÏDE : ROLES ET CONSEQUENCES ADAPTATIVES ? ... 20

1-Une stratégie parasitaire unique ... *20*

2- Une symbiose obligatoire ... *21*

3-Apparition du virus au cours de l'évolution des guêpes endoparasitoïdes ... *23*

4- Effets physiologiques du virus chez l'hôte parasité ... *24*

5- Le génome des PDV : un cas atypique ... *25*

6- Les facteurs de virulence : une acquisition au cours du temps ... *26*

II-PROBLEMATIQUE ET OBJECTIFS ... **30**

A-PROBLEMATIQUE ... 30

B-OBJECTIFS ... 34

III-LARGE GENE EXPANSION IN MUTUALISTIC POLYDNAVIRUSES: MOLECULAR AND EVOLUTIONARY MECHANISMS AT THE ORIGIN OF PTPs ... **36**

IV-VIRAL CYSTATIN EVOLUTION AND 3D STRUCTURE MODELLING: A CASE OF DIRECTIONAL SELECTION ACTING ON A VIRAL PROTEIN INVOLVED IN A HOST-PARASITOID INTERACTION. ... **60**

V- CYSTEINE PROTEASES: POTENTIAL TARGETS OF PARASITISM ... **84**

A-IDENTIFICATION OF PARASITE-RESPONSIVE CYSTEINE PROTEASES IN *MANDUCA SEXTA* ... 85

B-ATTEMPT TO ISOLATE VIRAL CYSTATIN-HOST CYSTEINE PROTEASE COMPLEXES *IN VIVO* IN THE COURSE OF PARASITISM OF *MANDUCA SEXTA* BY *COTESIA CONGREGATA* ... 110

VI-ADAPTIVE EVOLUTION OF VIRAL GENES AND WASP DIVERSIFICATION ... **120**

VII- DISCUSSION GENERALE ... **140**

A-LES FACTEURS DE VIRULENCE : LES PROCESSUS EVOLUTIFS IMPLIQUES DANS LEUR DIVERSIFICATION ... 141

B-ACQUISITION DE NOUVELLES FONCTIONS ET ADAPTATIONS ... 145

1-Le rôle des facteurs de virulence ... *145*

a) Dans quelles fonctions, les gènes spécifiques des lignées de bracovirus peuvent-ils être impliqués ? ... 146

b) A quelles pressions de sélection ces gènes doivent-ils répondre? 147

2-Conséquences sur l'adaptation et la diversification du parasitoïde *150*

C- Le parasitisme : une course aux armements........ 152

REFERENCES BIBLIOGRAPHIQUES 156

Liste des figures

Figure 1 Cycle du polydnavirus et de la guêpe parasitoïde...22
Figure 2 Représentation schématique de l'histoire évolutive du complexe Microgastroïde et de l'apparition des gènes viraux au cours de l'évolution..24
Figure 3 Un possible scénario évolutif de l'association polydnavirus guêpe braconide. 28
Figure 4 Unrooted PTP phylogenetic tree from Bayesian inferences under the GTR+I+G substitution model...44
Figure 5 PTP alignment, conserved motifs and sites under positive selection. 46
Figure 6 Orthologous genomic regions from CcBV, CvBV and GiBV. 48
Figure 7 Unrooted phylogenetic trees from Bayesian inferences under the GTR+I+G substitution model of the three major PTP clades. 50
Figure 8 Cystatin gene tree obtained by maximum likelihood 70
Figure 9 Molecular model of CcBV cystatin 1 obtained from the average structure.......................73
Figure 10 Graphic representation of variable selective pressures (ω) along the protein sequence. 74
Figure 11 Cystatin sequence tree under local substitution model...76
Figure 12 Branches under positive selection estimated according to the Ga-branch analysis........78
Figure 13 Nucleotide and amino acid sequences of *M. sexta* Cathepsins B (MsCathB), L (MsCathL) and the 26/ 29 kDa C1A protease MsCath1. 95
Figure 14 *M. sexta* C1A cysteine protease alignements..97
Figure 15 Cysteine protease transcriptional regulation in the course of *M. sexta* development by semi-quantitative RT-PCR. 98
Figure 16 Cysteine protease transcriptional regulation after an immune challenge by semi-quantitative RT-PCR...100
Figure 17 Western blot analysis of proteins extracted from *M. sexta* hemocytes and hemolymph (plasma)..102
Figure 18 Cysteine protease activity in hemocyte extracts and plasma sample of *M. sexta*...........104
Figure 19 Immunoblot with anti-cystatin or anti-MsCath1 antibody (A) and Coomassie gel stained in native conditions (B) of non parasitized (NP) and parasitized (P) *M. sexta* plasma. 113
Figure 20 Immunoblot (A) and Coomassie gel in denaturing conditions of excised bands (B)...113
Figure 23 Viral cystatin lineages evolving under positive selection. 134

Figure 24 Certains gènes de virulence sont très conservés au sein des Bracovirus (IkB, PTP, CRP) alors que d'autres facteurs présentent une distribution étroite et sont spécifiques de certains Bracovirus...145

I-INTRODUCTION

Selon la théorie synthétique de l'évolution, la variabilité n'est créée qu'à partir des formes géniques déjà existantes, grâce à deux grands processus évolutifs ; la mutation et la recombinaison génétique homologue. Ces deux mécanismes permettent de faire évoluer pas à pas le génome en créant différentes versions parmi lesquelles, les formes les plus adaptées seront conservées. Pourtant, ces dernières années, les avancées en génomique révèlent que ces deux mécanismes ne sont pas les seuls à permettre au génome d'évoluer. En effet, le séquençage de nombreux génomes a révélé la présence récurrente de formes génétiques exogènes provenant d'organismes issus de différentes lignées évolutives. Ces transferts d'information génétique entre lignées différentes, entraînent de nombreuses modifications ; au niveau génomique puisque la structure du génome n'est plus la même, mais également d'un point de vue phénotypique puisque l'expression de l'information exogène peut entrainer des modifications physiologique et/ou morphologique.

Chez les procaryotes, les transferts de gènes horizontaux sont fréquents et constituent un processus d'évolution majeur. Ils se produisent via les phénomènes de conjugaison et de transformation mais aussi par transduction par l'intermédiaire de bactériophage. Chez les gammaprotéobactéries, par exemple, la plupart des génomes étudiés montre que la majorité des gènes est issue d'acquisition exogènes qui ont eu lieu après la divergence de ces lignées depuis un ancêtre commun (Lerat et al., 2003). De plus, si l'on compare l'ensemble des gènes de différentes souches d'*Escherichia coli* très proches phylogénétiquement (99% d'homologie entre les gènes orthologues), on observe qu'au moins 25% des gènes présents dans un génome sont absents chez les autres souches. Cette différence entre des souches très proches, est le résultat d'une acquisition récente de gènes de sources exogènes et souvent inconnues (Welch et al., 2002)). En permettant l'acquisition d'une information différente suivant les lignées, ces transferts, influencent les processus de diversification et d'adaptation. En effet, chez *E. coli* par exemple, ces transferts confèrent des nouveaux pouvoirs pathogènes, via les ilots de pathogénicité (Welch et al., 2002) et jouent un rôle important dans la résistance des bactéries aux métaux lourds (Liebert et al., 1999) ou à des composés toxiques (Top and Springael, 2003), ce qui permet aux bactéries de coloniser de nouvelles niches.

Chez les protistes, les transferts horizontaux constituent aussi un mécanisme évolutif important et contribuent significativement à l'évolution de leur génome (Morrison et al., 2007; Striepen et al., 2004; Templeton, 2007; Templeton et al., 2004). Chez le parasite humain *Giardia*

lamblia, les transferts de gènes ont façonné la structure du génome et les gènes codant les voies métaboliques sont largement représentés par des gènes d'origine bactérienne (Morrison et al., 2007).

Au contraire des êtres unicellulaires, et malgré certains cas isolés comme les transferts horizontaux massifs chez les Rotifères (Gladyshev et al., 2008), les transferts de gènes ne constituent pas un mécanisme majeur de l'évolution des génomes des organismes eucaryotes, chez qui l'acquisition de nouvelles fonctions est majoritairement expliquée par les duplications de gènes (Arguello et al., 2006; Katju and Lynch, 2003).

Cependant, les transferts physiques d'information génétique ne constituent pas la seule possibilité d'acquérir de nouvelles fonctions et les symbioses en associant de manière durable des organismes non apparentés représentent une alternative efficace.

Les symbioses concernent toutes les branches du vivant impliquant des virus, des bactéries ou des cellules eucaryotes (Moran et al., 2005). Dans ce cas, même si le transfert d'information n'est pas physique et ne modifie pas la structure du génome de l'hôte, l'information du symbiote peut s'exprimer et conférer de nouvelles fonctions à son partenaire. De telles associations, au contraire des interactions proies-prédateurs, impliquent une grande intimité entre les partenaires associés. On peut ainsi se demander quel sera le devenir des partenaires impliqués dans une telle association et comment cette symbiose évolue au cours du temps.

Les enjeux évolutifs impliqués dans ces associations sont en effet considérables. Les données croissantes en génomique permettent de mesurer les conséquences de ces associations sur le devenir des partenaires et de mieux comprendre les raisons de leur maintien au cours de l'évolution. En effet, la génomique a permis d'identifier les gènes qui sous-tendent la symbiose et qui codent des fonctions indispensables à l'hôte. Elle permet aussi de comprendre les modifications génomiques qui rendent l'association obligatoire et conditionnent son devenir. Ainsi, elle a permis de montrer qu'une partie des génomes des mitochondries et des chloroplastes était transférée dans le génome de la cellule hôte, soulignant le rôle de ces échanges dans le maintien de la symbiose et mettant en évidence le rôle des symbioses dans les phénomènes de transfert de gènes.

Les symbioses ont été largement étudiées chez les bactéries ; elles fournissent à leur hôte les métabolites et vitamines indispensables à leur croissance en milieu limité. Nous proposons de mieux comprendre les enjeux évolutifs et fonctionnels engagés dans une symbiose particulière impliquant une guêpe parasitoïde et un virus, qui est le seul mutualisme connu entre un virus et

un organisme eucaryote. Au sein de la symbiose, le virus et la guêpe sont étroitement liés, au niveau fonctionnel puisque l'association est indispensable au succès du parasitisme mais également au niveau génomique puisque le génome du virus est intégré dans le génome de la guêpe.
Au cours de ma thèse, je me suis intéressée aux **fonctions** acquises via cette symbiose, à leur **évolution** au cours de l'histoire de la symbiose et à leur **rôle potentiel** dans le succès du parasitisme.

L'introduction de ce manuscrit a pour but de montrer d'une part en quoi les symbioses ont contribué à l'adaptation des espèces et à l'évolution des génomes et d'autre part le caractère très singulier de la symbiose entre une guêpe parasitoïde et un virus

A-Symbioses et conséquences adaptatives

1-La symbiose : une interaction durable

La symbiose est définie comme étant une association intime et durable entre deux organismes appartenant à deux espèces différentes. On distingue deux grands types de symbioses : le parasitisme où l'un des partenaires tire profit de son associé et le mutualisme où chacun des partenaires bénéficie de l'association. Combes (2001) (Combes, 2001) propose 3 raisons à l'établissement d'une telle association : l'hôte constitue pour son symbiote un habitat, un moyen de dispersion et/ou une source d'énergie. Le maintien de cette association, qu'il s'agisse d'un parasitisme ou d'un mutualisme, dépend du bilan sélectif entre les coûts et les bénéfices s'appliquant à chacun des partenaires de l'association (Thompson, 1988).

Ce type d'interaction diffère du système proie-prédateur dans le sens où la relation entre les partenaires associés s'inscrit dans la durée. La symbiose est qualifiée d'interaction durable dans la mesure où les génomes des partenaires cohabitent sur le long terme (Combes, 2001). Combes insiste sur l'importance de cette notion de durée qui implique de nombreuses conséquences sur l'évolution des partenaires. Premièrement, le génome du symbiote peut s'exprimer à travers le génome de l'hôte, on parle alors de phénotype étendu (Dawkins, 1982). Par exemple, les insectes galligènes manipulent la croissance des végétaux et induisent la formation de galles qui les protègent des agresseurs et constituent une source de nourriture. D'autre part, l'association de deux génomes peut conduire à une redondance des fonctions codées ; il est alors possible qu'une même information soit codée par les deux génomes associés. Si les mutations peuvent être délétères dans le cas d'organismes libres, elles ne le sont plus dans le cas d'une symbiose, ce qui induit une perte de l'expression d'un des gènes par la formation de pseudogènes ou d'une manière plus radicale, par la délétion d'un de ces gènes. Finalement, ces interactions durables peuvent avoir pour conséquence des transferts horizontaux de gènes entre les génomes du symbiote et de l'hôte. Les conséquences évolutives des symbioses sur les partenaires de l'association seront évoquées et illustrées par des exemples précis au cours de cette introduction.

Les symbioses peuvent être classées en deux grandes catégories suivant le degré d'interdépendance entre le symbiote et son hôte : les symbioses obligatoires ou primaires et les symbioses facultatives ou secondaires.

Les symbioses obligatoires

Les symbioses obligatoires sont supposées être le résultat d'associations anciennes et spécialisées, transmises de manière verticale depuis le début de l'association. Ces symbioses sont apparues très tôt au cours de l'évolution ; l'âge de leur apparition variant de 270 à 30 millions d'années. Les partenaires impliqués dans ce type d'association partagent une histoire commune qui a marqué leur génome ainsi que leur physiologie (revue Dale and Moran, 2006).
L'une des signatures les plus évidentes de ces associations, est la réduction du génome du symbiote. Chez le symbiote bactérien *Buchnera aphidicola* qui fournit son hôte, le puceron, en nutriments essentiels, la symbiose s'est accompagnée d'une perte de nombreuses fonctions relatives à la réplication ou au métabolisme (van Ham et al., 2003; Zientz et al., 2004). Ces réductions génomiques empêchent le retour du symbiote à un mode de vie libre et renforce ainsi le lien entre les partenaires. Ces associations sont, en outre marquées par l'existence d'organes spécialisés destinés à l'accueil du symbiote. Le bactériome, par exemple, est un tissu mis en place par l'hôte et réservé à l'accueil de symbiotes bactériens tels que *Buchnera aphidicola* chez le puceron ou *Wigglesworthia glossinidia* chez la mouche tsé tsé. Ces organes constituent de véritables signatures d'une coévolution établie entre le symbiote et son hôte et participent au maintien de la symbiose.
Ces modifications génomiques et morphologiques contribuent à maintenir, voire à renforcer l'interdépendance entre les partenaires. Les signatures de cette interdépendance, se manifestent au niveau macroévolutif par une histoire évolutive commune, comme en témoignent de nombreux exemples de co-phylogénies (Chen et al., 1999; Daniela M. Takiya, 2006; Hosokawa et al., 2006) (voir encadré 1).

Symbioses facultatives

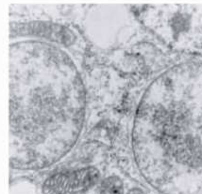
Contrairement aux symbioses obligatoires, les symbioses facultatives sont considérées comme récentes et impliquent des symbiotes dont la transmission n'est pas strictement verticale mais souvent horizontale. En outre, ces symbiotes ne sont pas essentiels du point de vue de l'hôte et ils ne sont jamais logés au niveau de tissus spécialisés tels que les bactériomes. Le terme de l'association impliquant ces symbiotes peut être bénéfique et conférer, par exemple, une protection contre les ennemis naturels (Oliver et al., 2003) ou être délétère comme c'est le cas des

bactéries du genre *Wolbachia* qui envahissent la population hôte en manipulant sa reproduction (Werren et al., 2008).

Le degré d'intimité qui unit le symbiote à son hôte est ainsi variable d'une association à une autre et au sein même d'une association la nature de la relation ainsi que le degré de dépendance peuvent évoluer dans le temps. La bactérie W*olbachia*, en général décrite comme un symbiote facultatif et parasite de la reproduction chez les Arthropodes, est par exemple devenue indispensable à l'ovogenèse chez la guêpe *Asobara tabida* (Dedeine et al., 2001).

Symbiose obligatoire: exemple de l'association Puceron/*Buchnera*

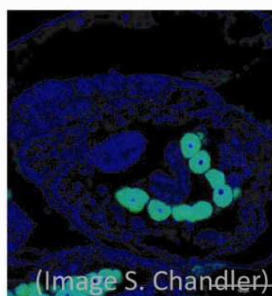
Buchnera dans les cellules hôtes



Acyrthosiphon pisum

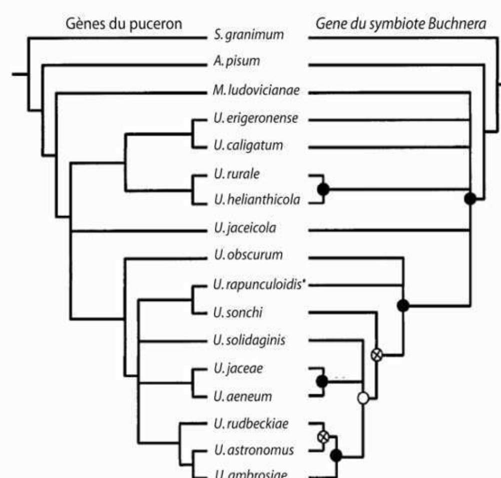
Le symbiote bactérien du genre *Buchnera* fournit a son hôte le puceron des nutriments essentiels, lui permettant de s'alimenter à partir d'une source nutritive limitée; la sève. Cette association ancienne, apparue il y a 100 à 200 millions d'années, est caractérisée par des adaptations (le bactériome) et une histoire évolutive qui témoignent de l'étroite relation établie entre les partenaires.

Le Bacteriome:



Le symbiote est logé dans des cellules spécialisées de l'hôte. Ici, un marquage au FISH permet de visualiser l'endosymbiote (en vert) et un marquage au DAPI permet de visualiser les noyaux des cellules de l'hôte (en bleu).

Une histoire évolutive commune



Les partenaires impliqués dans une symbiose obligatoire sont caractérisés par une histoire évolutive commune marquée par une cospéciation. (D'après Clark, 2000)

Encadré 1 Exemple d'une symbiose obligatoire entre un insecte et une bactérie.

2-Symbiose et acquisition de nouvelles fonctions

Les symbioses sont très répandues dans le monde vivant et ont joué un rôle majeur dans l'émergence des formes de vie et dans la diversification des organismes. L'illustration même du rôle majeur des symbioses dans l'évolution du vivant est l'apparition d'organites cellulaires tels que les mitochondries ou les chloroplastes à partir de cyanobactéries et d'alpha-protéobactéries respectivement. Ces symbioses, à l'origine de l'émergence des mécanismes de respiration et de photosynthèse dans le règne eucaryote, illustrent le rôle majeur de ces associations dans l'adaptation des espèces et leur radiation.

Les symbioses sont aujourd'hui connues chez de nombreux organismes et sont très bien décrites chez les plantes et les insectes.

Chez les plantes, les communautés bactériennes du sol ont largement participé à leur évolution et leur diversification, en fournissant des éléments essentiels à leur croissance.
A titre d'exemple, en permettant la fixation de l'azote du sol, les bactéries symbiotiques ont contribué à l'évolution des plantes légumineuses qui sont aujourd'hui largement distribuées et occupent une grande partie des forêts tropicales (van der Heijden et al., 2008).
Les champignons peuvent également entretenir des relations mutualistes avec les plantes. Il existe par exemple, un champignon mutualiste, présent chez les plantes des régions géothermales, qui confère à la plante hôte une thermo-tolérance (Redman et al., 2002). Récemment, une étude a montré que cette thermo-tolérance était liée à la présence d'un troisième partenaire associé aux champignons mutualistes ; un virus qui interviendrait dans la survie de la plante en détournant les produits oxydatifs libérés par le système de défense de la plante (Marquez et al., 2007). Cependant, la nature obligatoire ou facultative de la relation établie entre le champignon et le virus reste à être déterminée.

Chez les insectes, les symbioses nutritionnelles qui impliquent des symbiotes bactériens ont fait l'objet de plusieurs études. Ces symbioses procurent à leur hôte les éléments nutritionnels essentiels et leur ont permis de diversifier leur habitat. Les symbiotes obligatoires, transmis de manière verticale, peuvent contribuer à la spécialisation de leur hôte, c'est le cas du symbiote associé à l'Hétéroptère du genre *Megacopta* dont le génotype détermine la capacité de son hôte à sa nourrir sur une plante donnée. Ces symbiotes peuvent de cette manière favoriser la diversification de leur hôte (Hosokawa et al., 2007). Au contraire, les symbiotes non spécifiques et transmis de manière horizontale, peuvent permettre d'élargir le spectre d'habitats de son hôte. Dans ce cas au lieu de favoriser la spécialisation de son hôte et contribuer à sa diversification, le symbiote va

homogénéiser la population hôte et jouer un rôle inhibiteur sur la radiation de celle-ci (Kikuchi et al., 2007; Russell and Moran, 2005).

Les symbiotes bactériens peuvent également conférer à leur hôte insecte une protection contre les pathogènes. Par exemple, chez le puceron parasité par le parasitoïde *Aphidius ervi,* des symbiotes bactériens facultatifs permettent de réduire de manière significative le succès du parasitisme en contribuant à l'élimination des larves de parasitoïdes (Oliver et al., 2003).
Inversement, des symbiotes viraux peuvent jouer un rôle essentiel dans le succès parasitaire de certains Hyménoptères. Ces virus, de type polydnavirus, sont injectés lors de l'oviposition dans l'hôte Lépidoptère au sein duquel l'expression des gènes viraux joue un rôle essentiel dans la dérégulation des fonctions physiologiques de l'hôte et assure ainsi le succès du parasitisme (Turnbull and Webb, 2002).

Les organismes mutualistes sont ainsi impliqués dans différents fonctions biologiques et constituent de véritables innovations permettant à leur hôte de coloniser de nouveaux habitats.

3-Symbioses et conséquences sur l'évolution des génomes

La génomique offre des bases solides à la compréhension de l'établissement d'une symbiose et des conséquences sur l'évolution des génomes des partenaires.
Le séquençage des génomes des hôtes et des symbiotes permet en effet de montrer en quoi les symbioses permettent l'acquisition de nouvelles fonctions et comment elles ont façonné le génome des partenaires.

Un exemple qui illustre particulièrement bien les conséquences génétiques associées à une interaction symbiotique concerne l'association entre le nématode parasite filaire *Brugia malayi* et son endosymbiote obligatoire *Wolbachia* (WBm). *B. malayi,* est transmis à l'homme par des moustiques de la sous-famille des Culicinae. L'association entre *B. malayi* et *Wolbachia* est de nature obligatoire puisqu'un traitement antibiotique inhibe la croissance et le développement du nématode (Fenn and Blaxter, 2006). Récemment les séquences complètes des génomes de *B. malayi* et de son endosymbiote ont été publiées offrant la possibilité de mesurer les fonctions impliquées dans le maintien d'une symbiose et les conséquences de cette association sur les deux génomes.

L'analyse du génome de *B. malayi* révèle que ce génome est non seulement façonné par le mode de vie parasitaire de ce nématode mais aussi par l'association symbiotique avec *Wolbachia*. En effet, cette étude montre l'absence totale de neuf enzymes sur dix impliquées dans la synthèse de purine, de six gènes sur sept impliqués dans la biosynthèse de l'hème et de 5 enzymes essentielles pour la biosynthèse des riboflavines. Ces produits métaboliques essentiels, peuvent être ingérés directement par le nématode ou plus vraisemblablement être fournis par l'endosymbiote *Wolbachia,* qui comporte dans son génome les voies de biosynthèse complètes des purines, de l'hème et des riboflavines. L'étude de ces génomes permet de caractériser les bases génétiques expliquant la nature obligatoire de cette association (Foster et al., 2005; Ghedin et al., 2007).

En outre, l'analyse du génome de *Wolbachia* WBm révèle une réduction génomique importante en comparaison du génome de la souche *Wolbachia* Wmel endosymbiote parasite de *Drosophila melanogaster*. Ces réductions reflètent des pertes de gènes liés au mode de vie parasite, puisque contrairement à Wmel, les gènes nécessaires à l'infection des cellules hôtes ou au contournement de la réponse immunitaire, ne sont plus indispensables à WBm devenu un endosymbiote obligatoire (Foster et al., 2005).

L'analyse de plusieurs génomes bactériens montre que ces réductions sont un phénomène fréquent chez les symbiotes. La réduction la plus extrême est illustrée par le génome de *Carsonella rudii,* endosymbiote de l'insecte galligène *Pachypsylla venusta.* En effet, le génome de *C. rudii* réduit à 160 kilobases, ne code plus pour les fonctions indispensables aux processus cellulaires bactériens et a même perdu des fonctions associées au maintien de la symbiose (Nakabachi et al., 2006) ! Ces réductions génomiques spectaculaires impliquent que le symbiote n'est plus capable de retourner à un mode de vie libre et illustrent l'étroite relation des partenaires associés. Dans ce cas, on peut se demander si cette information est totalement perdue par l'association ou si ces gènes ont été transférés dans le génome de l'hôte. Est ce que l'étroite relation établie entre le symbiote et son hôte favorise les échanges d'information génétique expliquant ces pertes inattendues de gènes ?

L'hypothèse supposant une acquisition de gènes via les symbioses doit être considérée. En effet, les génomes des mitochondries ou des chloroplastes, organites cellulaires issus d'évènements d'endosymbioses, sont caractérisés par la perte de gènes transférés dans le génome de l'hôte (Kurland and Andersson, 2000; Martin et al., 2002). Chez *Arabidopsis thaliana*, 18% des gènes codant sont le résultat d'un transfert depuis la cyanobactérie ancêtre des chloroplastes (Martin et al., 2002).

Le génome eucaryote est ainsi capable (1) d'intégrer de l'information génétique provenant d'une autre espèce, (2) de la maintenir de façon stable dans son génome, (3) de l'exprimer et (4) de la transmettre à sa descendance.
Pendant longtemps, les preuves de l'existence de ces transferts de gènes entre endosymbiotes et cellules hôtes étaient limitées aux mitochondries et aux chloroplastes. Pourtant des études récentes, réalisées chez la bactérie *Wolbachia,* prouvent que les transferts de gènes du symbiote vers son hôte eucaryote sont possibles (Hotopp et al., 2007; Kondo et al., 2002). Récemment, une étude a montré qu'une large portion du génome de *Wolbachia pipientis* est présente dans le génome de son hôte *Drosophila ananassae* et transmise de manière mendélienne à la descendance. Cependant, la faible expression des gènes impliqués dans ce transfert ne permet pas encore d'assurer qu'ils sont fonctionnels (Hotopp et al., 2007).
La mise en évidence de ces transferts de gènes suggère qu'ils ont été sous-estimés et qu'il faudra dans le futur, examiner rigoureusement la présence de gènes d'origine procaryote dans des génomes eucaryotes, avant de les considérer comme de potentielles contaminations.

Pour comprendre les avantages adaptatifs d'une symbiose et les conséquences évolutives d'une telle proximité sur le devenir des partenaires, nous avons étudié un mutualisme unique impliquant un virus de type polydnavirus et une guêpe parasitoïde. Cette symbiose atypique, représente un modèle particulièrement adapté à la compréhension des mécanismes fonctionnels et des enjeux évolutifs impliqués dans une symbiose.

B- Un mutualisme virus-guêpe parasitoïde : rôles et conséquences adaptatives ?

1-Une stratégie parasitaire unique

Les interactions entre Hyménoptères endoparasitoïdes et leurs hôtes insectes font partie des interactions les plus complexes du monde des insectes. Dans ce système très spécialisé, les larves de guêpes doivent survivre et se développer à l'intérieur du corps d'autres insectes, en général des larves de Lépidoptères. Les guêpes doivent, en particulier échapper au **système immunitaire** de l'hôte Lépidoptère. Les guêpes endoparasitoïdes ont pour cela développé

différentes **stratégies** afin de contourner ou inactiver cette réponse (Whitfield et al., 2003). Parmi ces stratégies, plusieurs groupes de guêpes ont développé un mécanisme de défense puissant qui consiste à injecter en même temps que leurs œufs un **virus symbiotique** de type **polydnavirus** (PDV). L'expression des gènes viraux dans le Lépidoptère joue un rôle essentiel dans le contrôle des fonctions physiologiques de l'hôte lépidoptère et assure le succès parasitaire de la guêpe (Turnbull and Webb, 2002).

Si d'autres virus mutualistes, tels que les réovirus, participent aussi au succès parasitaire de guêpes parasitoïdes (Renault et al., 2005), l'intimité de l'association entre les PDV et leurs hôtes constitue un cas unique de mutualisme entre un virus et un organisme eucaryote.

2- Une symbiose obligatoire

Les PDV sont des virus très particuliers en raison de leur cycle biologique et de l'intime relation qu'ils ont établie avec leurs hôtes.

Le génome viral est intégré dans le génome de la guêpe sous la forme d'un **provirus** et se transmet ainsi verticalement au cours des générations de guêpes (Desjardins et al., 2007). La réplication de l'ADN viral s'effectue de manière massive dans les ovaires des guêpes femelles au niveau des cellules du calyx, où le génome viral est excisé et circularisé en **multiples cercles d'ADN doubles brins** qui sont ensuite encapsidés. Les particules virales ainsi formées sont libérées dans la lumière de l'ovaire pour former une fraction du fluide du calyx dans lequel se trouvent les œufs matures. Lors de l'oviposition, les particules virales sont injectées en même temps que les œufs de guêpes dans l'hôte Lépidoptère. Le virus ne se réplique pas dans l'hôte parasité mais l'expression des gènes viraux induit une altération des fonctions physiologiques de l'hôte parasité assurant ainsi le développement des larves de guêpes (Asgari et al., 1996; Beckage and Gelman, 2004; Tanaka et al., 2000).

Par conséquent, le système guêpes-polydnavirus est une **symbiose obligatoire** au sein de laquelle le virus, en assurant le succès parasitaire de son hôte, garantit sa propre transmission via le génome de la guêpe (figure 1).

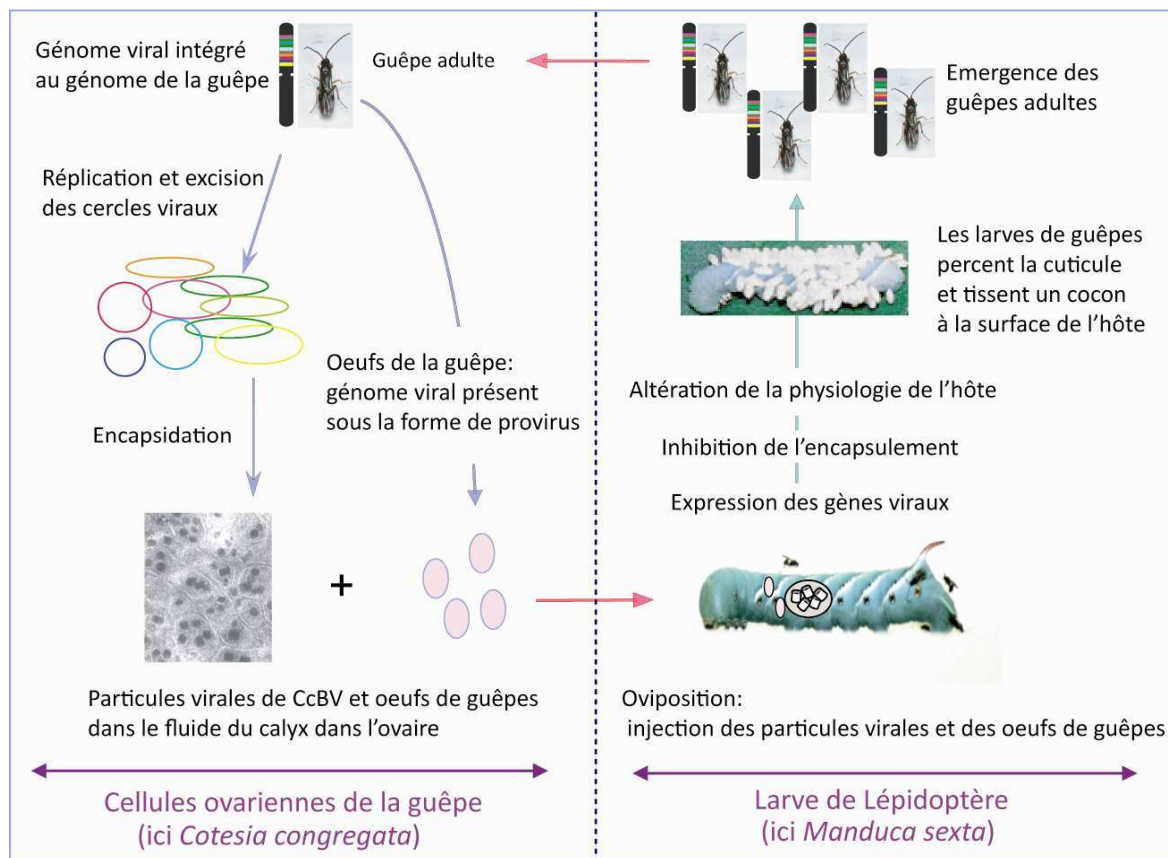


Figure 1 Cycle du polydnavirus et de la guêpe parasitoïde.
L'ADN viral intégré au génome de la guêpe (ici *Cotesia congregata*) est transmis verticalement sous forme de provirus. Dans les cellules ovariennes, la forme intégrée du virus est répliquée et excisée pour ensuite être encapsidée et former les particules virales. Lors de l'oviposition, les particules virales sont injectées en même temps que les œufs de guêpe. Les particules virales entrent alors dans les cellules hôtes où les gènes viraux sont exprimés. L'expression des gènes viraux inhibe la réponse immunitaire de l'hôte et assure le succès du parasitisme et la transmission du provirus à la descendance (D'après Dupuy *et al*, 2006).

Le fonctionnement de cette association ainsi que l'intime relation établie entre les partenaires suggèrent que la domestication du virus par la guêpe est totale. On peut ainsi supposer que le virus et la guêpe partagent une histoire évolutive commune. Whitfield (2003) a testé l'hypothèse d'une co-divergence chez six espèces de guêpes du genre *Cotesia* en utilisant un gène viral (CrV1) et deux gènes de guêpe (ADNr 16S et NADH). Cette étude révèle une co-phylogénie parfaite entre le virus et son hôte suggérant une évolution conjointe des partenaires de l'association. Cette étude permet, en outre, de montrer qu'au moins chez ce genre, le virus n'est pas transmis de manière horizontale.

Si cette étude reste à être confirmée en utilisant un plus grand nombre d'espèce, elle donne pour la première fois une perspective évolutive à l'intimité de la relation établie entre le virus et la guêpe.

3-Apparition du virus au cours de l'évolution des guêpes endoparasitoïdes

Les PDV sont connus chez plusieurs milliers d'espèces de guêpes de la super-famille des Ichneumonoidea. Sur la base de leur structure génomique et morphologique deux lignées virales sont distinguées : les **Ichnovirus** associés à des guêpes de la famille des Ichneumonidae et les **Bracovirus** associés à des guêpes de la famille des Braconidae. Les associations guêpe-PDV sont ainsi supposées être apparues à la suite de deux évènements symbiotiques indépendants (Turnbull and Webb, 2002). Une troisième lignée de polydnavirus a récemment été décrite chez les guêpes de la sous famille des Banchinae et serait issue d'un troisième événement d'association symbiotique (Lapointe et al., 2007).

Les associations entre les guêpes Braconides et leur bracovirus sont les plus étudiées ; l'histoire évolutive des guêpes a fait l'objet de nombreuses études (Banks and Whitfield, 2006; Michel-Salzat and Whitfield, 2004; Murphy et al., 2008; Whitfield, 2002) et plusieurs génomes de Bracovirus sont complètement ou partiellement séquencés ((Choi et al., 2005; Desjardins et al., 2007; Gundersen-Rindal and Pedroni, 2006; Webb et al., 2006, Pennachio et al. non publié, Choi et al. non publié). Les bracovirus sont distribués dans six sous-familles de guêpes : les Cardiochilinae, les Microgastrinae, les Miracinae, les Khoikhoïnae, les Cheloninae et les Mendesellinae. L'ensemble de ces sous-familles forme un groupe monophylétique appelé complexe Microgastroïde dont l'âge est estimé à 103 millions d'années (Murphy et al., 2008) (Figure 2).

La **monophylie** des guêpes associées aux Bracovirus suggère qu'un seul événement associant une guêpe ancêtre à un virus ancêtre est à l'origine de la distribution actuelle du virus chez les guêpes Braconides (Whitfield, 2002). Parmi les Braconidae, la sous-famille des Microgastrinae, formée de 5000 à 10000 espèces (Mason, 1981) est caractérisée par une **diversification rapide**. Une telle expansion suggère que les guêpes ont subit une radiation adaptative, probablement liée à la diversification des espèces hôtes (Banks and Whitfield, 2006; Murphy et al., 2008).

Si les symbioses sont connues pour leur implication dans l'adaptation des espèces hôtes et si le rôle essentiel joué par les PDV dans le succès parasitaire de la guêpe est incontestable, la contribution du virus dans les **processus adaptatifs** de la guêpe n'a encore jamais été étudiée.

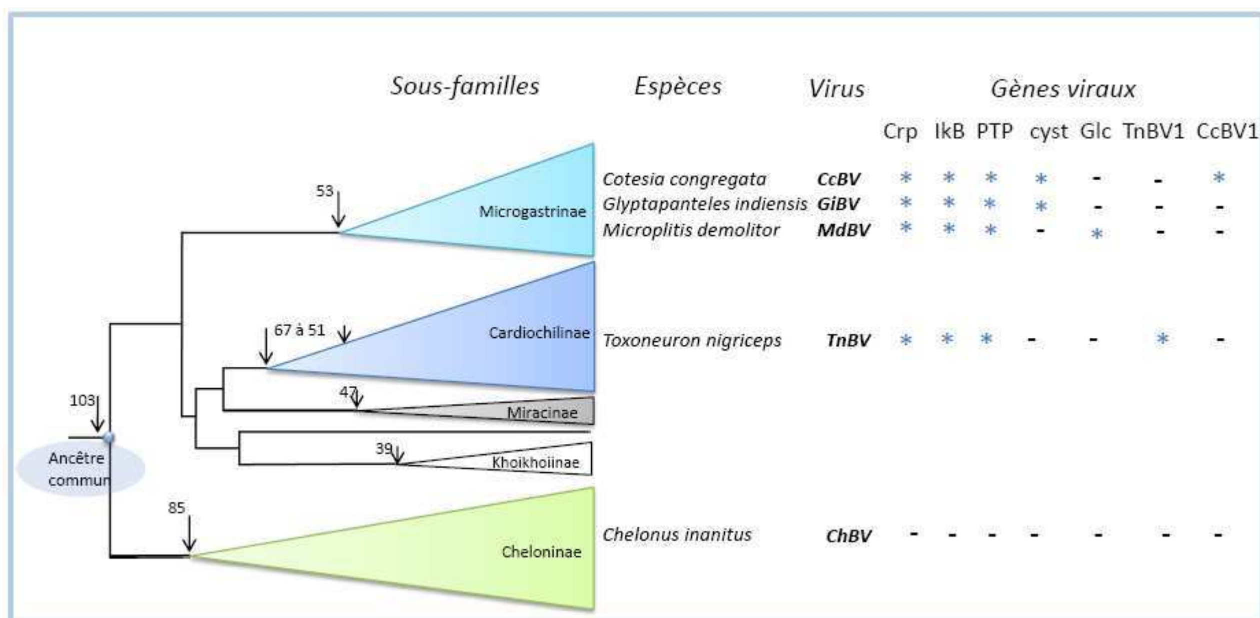


Figure 2 Représentation schématique de l'histoire évolutive du complexe Microgastroïde et de l'apparition des gènes viraux au cours de l'évolution.
Les dates (en millions d'années) indiquent l'apparition des sous-familles de guêpes (D'après Murphy, 2008). Une liste non exhaustive des gènes viraux montrent que certains gènes sont spécifiques de certaines lignées de bracovirus.

4- Effets physiologiques du virus chez l'hôte parasité

Les conséquences les plus évidentes du parasitisme sont le contrôle du développement et l'inhibition des défenses immunitaires de l'hôte Lépidoptère. Ces modifications sont liées à des bouleversements biochimiques et endocrines.

Effets du parasitisme sur le développement de l'hôte

Chez le Lépidoptère *Manduca sexta* par exemple, le développement larvaire se fait en 5 stades interrompus par 4 mues et au terme du 5ème stade larvaire la mue nymphale aboutit à la formation d'une chrysalide qui sera suivie de la métamorphose de la chenille en adulte. Le développement de la chenille parasitée est arrêté au stade L5 et la métamorphose est totalement inhibée. Les mues et la métamorphose des insectes sont sous la dépendance de plusieurs hormones : l'hormone prothoracique, les ecdystéroïdes et l'hormone juvénile. L'hormone prothoracique induit la production d'ecdystéroïdes qui interviennent dans les phénomènes de mue. La concentration de l'hormone juvénile par rapport à la concentration d'ecdystéroïde va déterminer les différents types de mues : mue larvaire, mue nymphale et mue imaginale (Nation, 2002). Chez *M. sexta,* au cours du cinquième stade larvaire, la diminution du taux d'hormone

juvénile corrélée à un faible taux d'hormone de mue entraine la métamorphose (Riddiford et al., 2003).

Chez les chenilles de *M. sexta* injectées par du virus purifié, on observe une suppression partielle du pic pré-pupal des hormones prothoraciques et ecdystéroïdes. Cette suppression, combinée au maintien d'un taux élevé d'hormone juvénile sécrétée par les larves de guêpes empêche la mue pupale et maintient l'hôte au stade larvaire au cours du parasitisme (Beckage and Gelman, 2004; Beckage et al., 1994).

Effets du parasitisme sur les défenses de l'hôte

Lors de l'injection d'un œuf de parasitoïde ou de tout autre macro-organisme étranger, le nombre total d'hémocytes (cellules immunitaires chez les insectes) augmente dans l'hémolymphe (i.e Strand and Noda, 1991). Les hôtes immunocompétents éliminent les parasitoïdes ainsi que les macro-organismes par le mécanisme d'encapsulement, qui consiste en la formation d'une enveloppe formée de plusieurs couches d'hémocytes à la surface du corps étranger. Cette réaction implique 3 étapes : la reconnaissance, l'adhésion des hémocytes pour la formation de la capsule et l'élimination du parasitoïde (Strand and Pech, 1995a). Peu de temps après le début de l'encapsulement, des tâches brunes de mélanine apparaissent à la surface de l'œuf grâce à l'action de la phénoloxydase. Ce processus, appelé mélanisation est un phénomène local participant à la neutralisation de l'œuf grâce à la production de composés réactifs oxygénés et nitrés hautement toxiques (Nappi et al., 2005; Nappi and Vass, 1998).

Lors du parasitisme, l'injection du virus inhibe le processus de prolifération cellulaire des hémocytes et empêche ainsi l'encapsulement (Beck and Strand, 2003; Strand and Noda, 1991). En outre, il se produit des modifications de la composition de l'hémolymphe accompagnée de l'inactivation de la phénoloxydase qui joue un rôle clé dans le processus de mélanisation. (Otto Schmidt, 2001; Strand and Pech, 1995a).

Ainsi, le virus cible deux fonctions clés de la biologie de l'hôte : le système immunitaire et le développement et permet au parasitoïde de se développer dans l'hôte sans être éliminé.

5- Le génome des PDV : un cas atypique

Si les effets physiologiques du virus sur l'hôte Lépidoptère ont fait l'objet de nombreuses études, les mécanismes par lesquels le virus induit la dérégulation des fonctions physiologiques restaient encore mal connus. Afin d'aborder efficacement cette problématique, le séquençage de

la forme libre du virus et la caractérisation des fonctions codées et exprimées par les gènes viraux ont été entrepris.

L'analyse de ces génomes révèle que les polydnavirus font partie des virus dits « géants » avec une taille d'environ 570 et 240 Kbases chez CcBV et MdBV respectivement. Le génome du polydnavirus est segmenté en plusieurs cercles d'ADN double brins dont le nombre est variable suivant l'espèce et dont la taille varie d'un cercle à l'autre. Une particularité importante du génome des polydnavirus est la présence d'introns que l'on retrouve dans 6 à 14% des gènes viraux (Espagne et al., 2004; Webb et al., 2006) suivant les génomes. Etonnamment, le génome contenu dans les particules virales est presque totalement dépourvu de gènes homologues à des gènes viraux pouvant intervenir dans les processus de réplication ou coder des protéines de capside.

En résumé, le génome des PDV ressemble plus à un génome eucaryote qu'à un virus classique, constatation qui a pendant plusieurs années contribué à remettre en question l'origine virale de la symbiose (Whitfield and Asgari, 2003). En effet, les caractéristiques du génome du PDV, proche de l'organisation des génomes eucaryotes, laissent penser que ces particules ne constituent finalement qu'un moyen habile de sécrétion, mis en place par la guêpe, pour introduire dans l'hôte ses propres gènes.
Cependant, par des approches de protéomique et de transcriptomique, réalisées à partir des ovaires de guêpes considérées où les particules sont produites, des gènes de PDV codant pour des protéines impliquées dans la réplication, la formation de la capside et de l'enveloppe ont été isolés. Les gènes isolés présentent des similarités significatives avec des gènes de virus d'insecte appartenant à la famille des Nudiviridae (Bezier et al soumis). Cette découverte a permis de prouver l'origine virale des bracovirus et assure l'existence de l'intégration d'un virus libre dans le génome de la guêpe ancêtre.

Les polydnavirus constituent ainsi des virus très singuliers, et leur structure actuelle suggère qu'ils sont le résultat d'une longue évolution depuis leur intégration dans le génome de la guêpe ancêtre.

6- Les facteurs de virulence : une acquisition au cours du temps

Si le génome des particules virales ne contient pas de gène de structure virale « classique », ils codent cependant pour plusieurs facteurs de virulence potentiels dont la plupart sont organisés en familles multigéniques.

Au sein des Bracovirus, trois familles de gènes sont conservées et trouvées dans tous les génomes viraux étudiés (TnBV, MdBV, GiBV, CpBV et CcBV) (Figure 2).

Parmi ces familles, on trouve des gènes à motifs ankyrines appartenant à la famille des IκB, également présents chez l'Ichnovirus CsIV (Webb et al., 2006). Chez les vertébrés et les insectes, ces gènes codent des inhibiteurs de facteurs de transcription impliqués dans l'immunité (De Gregorio et al., 2001; Hoffmann, 2003). Des études sur l'activité de ces protéines mettent en évidence leur capacité à inhiber des facteurs de transcription de type NFκB, suggérant l'implication des protéines IκB des polydnavirus dans la régulation de la réponse immunitaire de l'hôte (Thoetkiattikul et al., 2005).

La seconde famille commune aux PDV, code pour des protéines riches en cystéines (CRP). Ces protéines seraient impliquées dans le détournement du métabolisme de l'hôte ainsi que dans la régulation du système immunitaire via l'inhibition partielle de l'encapsulement (Kim, 2005; Li and Webb, 1994).

Enfin la dernière famille partagée par les bracovirus code pour des protéines tyrosine phosphatase (PTP). Cette famille constitue la plus grande famille de gènes décrite chez les PDV avec 27 gènes chez CcBV (Espagne et al., 2004). Les PTP sont connues chez les vertébrés pour jouer un rôle majeur dans la régulation des voies de signalisation cellulaire en déphosphorylant les résidus de type tyrosine de protéines régulatrices (Andersen et al., 2001). Chez MdBV, certaines de ces protéines montrent une activité inhibitrice de l'adhésion cellulaire et de la phagocytose et peuvent ainsi jouer un rôle important dans la régulation du système immunitaire de l'hôte (Beck and Strand, 2003, 2005; Pruijssers and Strand, 2007; Suderman et al., 2008)

En dépit du peu de génomes viraux disponibles, nous pouvons supposer que ces gènes sont apparus tôt au cours de l'évolution de l'association. En effet, si chacun de ces gènes partagent avec les autres Bracovirus le même gène ancêtre, leur apparition précède la divergence entre les sous-familles des Microgastrinae et des Cardiochilinae qui s'est produite il y a plus de 67 millions d'années (Murphy et al., 2008).

Au contraire des ankyrines, des CRP et des PTP, certains gènes ne sont partagés que par un nombre restreint de virus associés à des espèces de guêpes apparues récemment (figure 2). C'est le cas par exemple des cystatines qui ne se retrouvent que chez GiBV et CcBV (Desjardins et al., 2007; Espagne et al., 2004). Les cystatines sont des protéines largement répandues dans le monde vivant mais les cystatines des PDV constituent la première description de cystatines virales. Les cystatines sont souvent décrites comme étant impliquées dans des interactions hôte-parasite mais leur rôle dans la symbiose est encore méconnu.

Le génome viral est donc formé de gènes codant pour des facteurs de virulence potentiels permettant le contrôle des fonctions physiologiques de l'hôte. Ces gènes ne présentent aucune signature virale et sont, au contraire, décrits dans des génomes cellulaires, suggérant qu'ils proviennent du génome de la guêpe. L'étude de plusieurs génomes viraux, associés à des guêpes apparues à différentes périodes évolutives, révèle que le génome viral actuel est le résultat de plusieurs événements d'acquisition de gènes qui se sont produits après l'établissement de la symbiose et de manière successive au cours l'évolution des lignées de guêpes (figure 3).

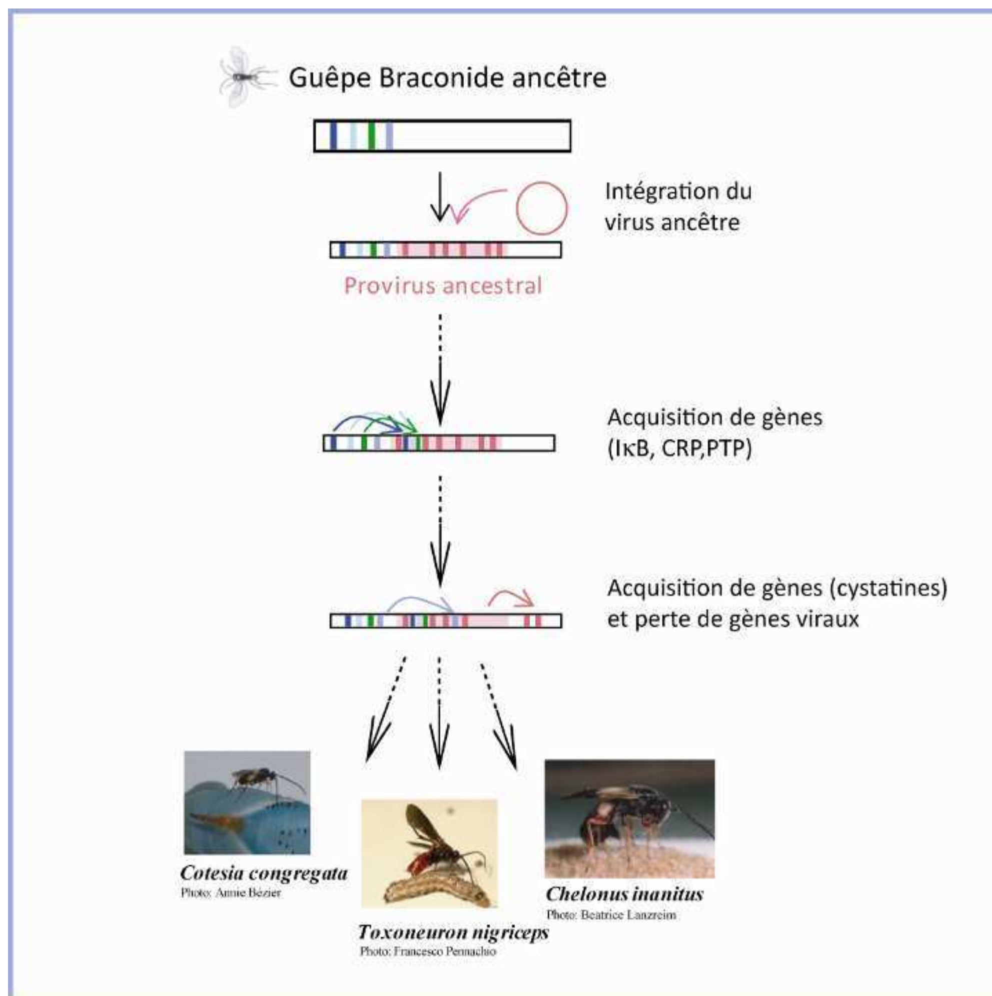


Figure 3 Un possible scénario évolutif de l'association polydnavirus guêpe braconide. Après l'intégration du virus ancêtre dans le génome de la guêpe, le virus a progressivement acquis de nouveaux gènes, probablement depuis le génome de la guêpe. Plusieurs évènements indépendants sont vraisemblablement à l'origine des gènes trouvés dans le génome viral. Inversement, des gènes viraux ont été perdus par le génome viral et transférés dans le génome de la guêpe (D'après Dupuy, 2006**).**

II-PROBLEMATIQUE ET OBJECTIFS

A-Problématique

Les connaissances sur les associations PDV-guêpes parasitoïdes montre qu'elles forment des symbioses uniques dans le sens où elles associent de manière étroite un organisme eucaryote à un virus. On a vu que cette association était essentielle au succès parasitaire de la guêpe. En effet, le virus en contrôlant et en manipulant les fonctions physiologiques de l'hôte, assure le développement de la guêpe et garantit sa propre transmission via le génome de la guêpe.
La particularité du génome viral réside dans sa taille et sa structure, mais aussi dans l'information codée par les gènes viraux. Le génome contenu dans les particules virales est presque totalement dépourvu de gène d'origine virale. Au contraire, il est constitué de gènes codant pour des facteurs de virulence potentiels, probablement acquis depuis le génome de la guêpe.

Comme cela a été illustré en introduction, les symbioses constituent de véritables moteurs de l'évolution dans le sens où elles permettent des évolutions plus rapides par rapport à aux mutations ou aux recombinaisons génétiques. De très nombreuses associations symbiotiques impliquant des micro-organismes bactériens ou mycéliens ont été sélectionnées au cours de l'évolution. Ces symbioses bien connues, interviennent dans des fonctions très différentes mais elles ont toutes contribuées à l'adaptation des organismes à leur milieu.
En revanche, les associations impliquant des virus sont très peu connues, et malgré certains exemples ont décrit la domestication de gènes d'origine virale, réutilisés pour des fonctions physiologiques (Mallet et al., 2004). Mais la domestication d'une machinerie virale complète dans le cas des PDV, représente un cas unique de symbiose virus-eukaryote à ce jour.
En quoi cette symbiose constitue-t-elle un avantage adaptatif ? Comment les fonctions transmises à l'hôte via le virus sont-elles impliquées dans le succès parasitaire de la guêpe ? Comment ces fonctions évoluent-elles au cours de l'histoire de la symbiose ?

Mon travail de thèse s'inscrit dans la compréhension du rôle de cette symbiose dans le succès parasitaire et l'évolution de l'hôte parasitoïde. En étudiant **l'évolution de facteurs de virulence** particuliers et en explorant les **fonctions potentiellement ciblées** chez l'hôte, j'ai voulu comprendre les enjeux évolutifs impliqués dans cette association et le devenir des gènes acquis au cours de l'évolution de l'association.

En regard des gènes codés par le virus, de leur organisation en familles multigéniques et de leur acquisition progressive au cours de l'association, le génome du polydnavirus apparaît comme une entité dynamique, capable de perdre ou d'intégrer de nouvelles fonctions.
Parmi les fonctions acquises par le virus, les PTP qui jouent un rôle clé dans les processus de régulation, sont apparues tôt au cours de l'évolution de l'association. Après son intégration dans le génome de la guêpe, le gène ancêtre des PTP a été hérité au cours de l'évolution et a connu une forte expansion pour former aujourd'hui la famille multigénique la plus large observée chez les PDV. En effet, l'expansion de cette famille constitue un cas exceptionnel dans l'histoire des gènes du polydnavirus. Actuellement, on trouve 13 copies de ce gène chez MdBV, 9 copies chez GiBV encore partiellement séquencé et 27 copies chez CcBV (Espagne et al., 2004; Gundersen-Rindal and Pedroni, 2006; Webb et al., 2006).

Les études menées sur le rôle des PTP au cours du parasitisme ont montré que ces gènes étaient exprimés au cours du parasitisme et que leur expression variait dans le temps et selon le tissu de l'hôte (Provost et al., 2004). Ces protéines régulent les voies de signalisation cellulaire par des réactions de déphosphorylation et jouent ainsi un rôle clé dans la régulation des voies de signalisation en particulier du système immunitaire (Ibrahim et al., 2007; Ibrahim and Kim, 2008; Pruijssers and Strand, 2007). Cependant, parmi les nombreux gènes de cette famille, certains ne présentent pas de site catalytiquement actif. Ces formes particulières pourraient réguler la réponse immunitaire, probablement en séquestrant les protéines phosphorylées, modulant ainsi l'activité phosphatase intracellulaire au niveau des hémocytes par exemple (Ibrahim and Kim, 2008).

Si de nombreux travaux ont été focalisés sur la fonction de ces gènes et sur les processus cellulaires potentiellement ciblés, aucune étude ne permet d'expliquer l'extraordinaire expansion de ce gène en famille multigénique. En effet, comprendre la fonction des PTP et leur rôle au cours de l'évolution de l'association implique aussi de comprendre quelles sont **les forces évolutives** et **les mécanismes moléculaires** qui ont façonné l'évolution de ce gène et probablement sa fonction.

Contrairement aux PTP, les cystatines sont apparues tardivement au cours de l'évolution de l'association et sont ainsi présentes chez un nombre restreint d'espèces. A ce jour une copie unique de cystatine a été isolée chez GiBV et 3 copies sont présentes sur le même cercle d'ADN chez CcBV. Les PTP et les cystatines sont des familles de gènes très différentes en regard de leur distribution dans les espèces de virus et de leur expansion en famille multigénique. On peut alors

supposer qu'elles n'ont pas été soumises aux mêmes pressions de sélection au cours de l'évolution de l'association. En effet la présence des PTP dans plusieurs génomes viraux suggère que ces protéines ciblent des fonctions communes et essentielles pour la défense de l'hôte. Au contraire, si les cystatines sont apparues plus tardivement nous pouvons supposer qu'elles codent pour des fonctions moins capitales mais qui confèrent une plus grande efficacité au parasitisme.

Les cystatines forment un complexe inhibiteur avec des protéases à cystéine de la famille C1, comme la papaine chez les plantes ou les cathepsines chez les animaux. Chez les plantes, les cystatines jouent un rôle important dans la protection contre les insectes phytophages qui libèrent des protéases afin de digérer les tissus de la plante (Arai et al., 2002). Chez les nématodes filaires, les cystatines constituent un facteur de virulence majeur qui inhibe la réponse immunitaire (Maizels et al., 2001; Schierack et al., 2003). Le rôle des cystatines virales dans l'interaction hôte-parasitoïde est encore inconnu. Toutefois, la purification d'une forme recombinante de la cystatine 1 chez CcBV montre que celle-ci est active contre des protéases à cystéine. De plus, nous savons que les cystatines sont exprimées très tôt au cours du parasitisme et en grande quantité. Cette expression est maintenue à un taux élevé 3 jours après le parasitisme. Ce schéma d'expression suggère que les cystatines doivent jouer un rôle important, probablement dès le debut de l'infestation (Espagne et al., 2004).

Pour comprendre le **rôle** des cystatines virales dans le parasitisme, il est nécessaire de comprendre en quoi les cystatines constituent un **facteur de virulence** potentiel et d'étudier la **fonction ciblée** par les cystatines virales chez l'hôte lépidoptère. De plus, en étudiant l'évolution de ce gène nouvellement acquis, il nous sera plus facile de comprendre son **rôle dans l'évolution de l'association et dans le succès parasitaire.**

Modèle expérimental: *Cotesia congregata* / *Manduca sexta*

Le parasitoïde: *Cotesia congregata*

Nous avons travaillé à partir de guêpes du genre *Cotesia,* guêpe de la famille des Braconidae et de la sous-famille des Microgastrinae. La sous-famille des Microgastrinae constitue l'une des sous-familles les plus diversifiées de la famille des Braconidae (Mason, 1981). Tous les Microgastrinae sont des koinobiontes endoparasitoïdes de lépidoptère (Shaw, 1991). Les guêpes du genre *Cotesia* présentent pour la plupart un spectre d'hôte étroit malgré certaines rares exceptions où le parasitoïde peut avoir un spectre d'hôte plus large que le cadre de la famille. Le genre *Cotesia* est un groupe très diversifié en termes de richesse spécifique et de spectre d'hôte (Mason, 1981). Le développement larvaire compte 3 stades dont les 2 premiers se déroulent à l'intérieur de l'hôte lépidoptère. Au 3eme stade, les larves percent la cuticule de l'hôte et tissent un cocon duquel émergeront les guêpes adultes.

L'hôte: *Manduca sexta*

L'étude des fonctions potentiellement ciblées par le virus a été réalisée chez le lépidoptère *Manduca sexta*. *M. sexta* appartient à la famille des Sphingidae et est distribuée sur tout le continent américain. Les larves se nourrissent de feuilles de plantes de la famille des Solanaceae. Le cycle larvaire présente 5 stades séparés chacun par des mues. Au terme du 5eme stade, la larve entre dans le stade pupal qui s'achève par la métamorphose en adulte.
La propagation de cet insecte ravageur est biologiquement contrôlée par les guêpes du genre *Cotesia.*

Encadré 2 Modèle expérimental

B-Objectifs

Ce travail de thèse porte sur les facteurs de virulences codés par le virus et plus particulièrement les PTP et les cystatines. Nous avons étudié les processus évolutifs qui ont façonné leur divergence depuis leur intégration et inféré leur rôle dans l'adaptation des guêpes. De plus, pour comprendre le rôle de ces facteurs, nous avons recherché leurs cibles chez l'hôte.

Mon travail de thèse s'articule ainsi en cinq grands chapitres :

- Le chapitre III présente l'étude des PTP dont l'acquisition s'est produite tôt au cours de l'histoire de l'association. Ce travail, rédigé sous la forme d'un article en préparation, vise à caractériser les **mécanismes moléculaires** et les **forces évolutives** qui ont conduit à l'expansion des **PTP** en **famille multigénique**.

- Dans le chapitre IV, en étudiant l'évolution d'un gène viral récemment acquis au cours de l'histoire de la symbiose, nous avons caractérisé les **marques moléculaires** portées par un **facteur de virulence** potentiel impliqué dans l'interaction avec l'hôte Lépidoptère. Cette étude réalisée sur les **cystatines virales** a fait l'objet d'un article accepté dans BMC Biology.

- Le chapitre V, porte sur les protéases à cystéine qui sont des cibles potentielles du parasitisme. La première partie de ce chapitre porte sur l'étude de la régulation de ces protéines chez l'hôte *M. sexta* au cours du parasitisme (manuscrit en préparation).
La deuxième partie de ce chapitre, présente les travaux réalisés pour isoler les **protéases à cystéine effectivement ciblées** au cours du parasitisme.

- Dans le chapitre VI, nous avons étudié en quoi **l'histoire évolutive de l'hôte** et **du symbiote viral** étaient **corrélées** et en quoi l'évolution de cette fonction au cours de l'histoire de la symbiose pouvait être impliquée dans les **processus de diversification** des guêpes parasitoïdes. Cette étude est en cours de préparation.

- Dans le chapitre VII, l'ensemble des résultats obtenus est résumé et discuté.

III-Large gene expansion in mutualistic polydnaviruses: Molecular and evolutionary mechanisms at the origin of PTPs

Serbielle Céline, Dupas Stéphane, Héricourt François, Huguet Elisabeth, Drezen Jean-Michel.

Abstract

Gene duplications have been proposed to be the main mechanism involved in genome evolution and in acquisition of new functions. They constitute therefore an important source of innovations and adaptations. Gene duplications are found to be particularly common in mutualistic viruses associated with parasitoid wasps. In these systems the virus is integrated into the wasp genome and virions injected in the parasitoids' hosts are essential for parasitism success. These viruses encode virulence factors which are involved in host immune suppression and developmental arrest. How did gene family expansion occur and what are the evolutionary forces inducing gene copy divergence? In order to understand gene duplication and divergence mechanisms which occurred during virus-wasp associations, we studied the protein tyrosine phosphatase (PTP) gene family which is the largest virus gene family described in these viruses.

Here, we show that viral PTPs expansion occurred through three main mechanisms; by duplication of large genomic segments (segmental duplication) and by tandem and dispersed gene duplications within the viral genome. These duplication events became sources of evolutionary innovations conferring to wasps adaptive properties. Indeed, PTP gene copy evolution was shown to undergo conservative evolution along with episodes of adaptive evolution which were correlated with duplication and wasp speciation processes. Altogether duplications and subsequent gene copy evolution likely contributed to the different patterns of PTP gene regulation and activities observed today.

Given the essential role played by the virus in wasp parasitism success and the extraordinary expansion of PTP genes, we can propose that PTP duplication contributed to wasp adaptation and diversification.

INTRODUCTION

Gene duplications have been recognized as an important source of evolutionary innovation and adaptation. The contribution of gene duplication to the evolution of new functional genes has been widely demonstrated in various organisms (Arguello et al., 2006; Katju and Lynch, 2003). Duplicated genes usually can be classified into tandem and dispersed duplicates, and duplicated copies supposedly evolve to improve the ancestral function. Two major questions are addressed when studying gene duplications: what are the molecular mechanisms underlying duplications and how are duplicated copies maintained during evolution? Two main hypotheses have been proposed to explain gene duplication evolution and acquisition of new function. Classical models propose that after gene duplication, one copy evolves under purifying selection and conserves the parental function whereas the extra copy is assumed to be neutral (Force et al., 1999; Hughes, 1994). In contrast, alternative models propose that duplications are adaptive and that duplicated copies diverge under positive selection for acquisition of novel function (Bergthorsson et al., 2007; Des Marais and Rausher, 2008).

Here we present a functional gene family encoded by a mutualistic virus involved in a host-parasitoid interaction. This family has been subjected to particularly strong expansion and constitutes therefore a remarkable model to study evolutionary processes involved in gene duplications.

The organization of genes into families constitutes a common characteristic in viruses of the polydnaviridae family associated with parasitoid wasps. The gene families encode putative virulence factors, some of which were proven to disrupt lepidopteran host physiology (Desjardins et al., 2007; Espagne et al., 2004; Lapointe et al., 2007; Webb et al., 2006). In wasp-PDV associations, PDVs persist as stably integrated proviruses in the genome of their associated wasp (Desjardins et al., 2007 ; Bezier 2008) and replicate in female ovaries only. Virus particles are injected into the lepidopteran host during wasp oviposition at the same time as wasp eggs. PDVs do not replicate in the parasitized host insect, but viral gene products suppress the host immune system and cause physiological alterations ensuring parasitoid development (Asgari et al., 1996; Beckage and Gelman, 2004; Tanaka et al., 2000). This unique example of mutualism between a virus and an eukaryotic organism, constitutes a real evolutionary success in regards to the tens of thousands of parasitoid species which carry PDVs. All wasps carrying PDV are found within the two separate lineages which constitute the Ichneumonoidea wasp superfamily: Ichnoviruses (IV) are associated with Ichneumonid wasps and Bracoviruses (BV) are found in Braconid wasps (Turnbull and Webb, 2002). Recently, PDVs associated with the Banchinae

wasps, previously belonging to the Ichnovirus genera, have been shown to be sufficiently distinct from both Ichnoviruses and Bracoviruses to justify the creation of a third PDV group (Lapointe et al., 2007). Each PDV genera presents distinct morphological and packaging characteristics (Lapointe et al., 2007; Webb et al., 2006) and the absence of PDVs in the basal Ichneumonoids suggests that IV, BV and Banchinae viruses arose independently in the three wasp lineages (Lapointe et al., 2007; Whitfield, 2002). To date, several PDV genomes have been sequenced and have all been shown to have large genomes segmented in multiple dsDNA circles. The other common and original feature of PDVs is that putative genes encoding virulence factors are organized in multigene families (Desjardins et al., 2007; Espagne et al., 2004; Lapointe et al., 2007; Webb et al., 2006). The diversification of virulence genes into families may reflect the adaptive pressures imposed on PDV genome evolution and underline their role in wasp parasitism.

The Braconid wasps carrying PDVs form a monophyletic group called the Microgastroid complex which was estimated, thanks to the calibration of the molecular clock by fossil records, to have arisen 103 Mya ago from an unique BV-braconid ancestral association (Murphy et al., 2008). Some genes encoding IκBs and protein tyrosine phophatases (PTP) respectively are common to most sequenced bracoviruses, suggesting they were acquired early in the course of the wasp-bracovirus evolution. IκB genes found in the three PDV lineages, encode proteins which are inhibitors of nuclear transcriptional factors involved in vertebrate and in *Drosophila* immune responses (De Gregorio et al., 2001; Falabella et al., 2007; Hoffmann, 2003; Thoetkiattikul et al., 2005). PTP genes are not found in IV and form a distinct clade in Banchinae, the lack of evidence of a common ancestor between PTPs from Banchinae and BV suggests that PTPs evolved separately in these two virus lineages (Lapointe et al., 2007). In all Bracovirus genomes described so far, PTPs belong to the largest gene family with 27 members in *Cotesia congregata* Bracovirus (CcBV), 13 members in *Microplitis demolitor* Bracovirus (MdBV) and at least 9 members in *Glyptapanteles indiensis* Bracovirus (GiBV) genome, which is partly sequenced (Desjardins et al., 2007; Espagne et al., 2004; Webb et al., 2006). PTP genes 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). All PDV PTPs studied so far are expressed in virus infected hosts but only a subset of these genes encodes catalytically functional PTPs (Ibrahim and Kim, 2008; Provost et al., 2004; Pruijssers and Strand, 2007). The "inactive" PTPs have been suggested to play a role in trapping phosphorylated proteins to impair cellular PTP activity in a competitive way (Provost et al., 2004). Moreover PTP gene expression is regulated in a tissue specific and time dependant manner (Gundersen-Rindal and Pedroni, 2006; Ibrahim et al., 2007;

Provost et al., 2004; Pruijssers and Strand, 2007). Bracovirus PTPs appear therefore to be important virulence factors which have undergone a high expansion rate and a high functional divergence (Bézier et al., 2007). In this context, Bracovirus PTPs emerge as an interesting gene family model to study the mechanistics and the evolution of gene duplication. To date, complete or partial sequence data for Bracovirus genomes are available giving us a support for understanding the genomic organization and the transmission of duplications in a dynamic interaction between a parasitoid wasp, a virus and a lepidopteran host.

Studying PTP gene family evolution enabled us to determine the molecular and evolutionary mechanisms at the origin of this family, and to highlight viral genome plasticity and evolutionary forces at the basis of PTP diversity. We discuss the critical role of duplications and natural selection in regard to functional divergence and adaptation.

Materials and Methods

Wasp specimens

Fourteen PTP genes previously isolated from Bracoviruses associated with Braconid wasps have been studied: PTP P, Q, Y, K, L, C, α, S, M, E, X, H, R and Δ.
These PTP genes were isolated from nine *Cotesia* species were considered: *C. congregata* (laboratory reared, France, Drezen,J-M), *C. chilonis* (laboratory reared, USA, Wiedenmann,. R), *C. flavipes* (Field collected, Kenya, S. Dupas), *C. glomerata* (laboratory reared, Netherland, Vet,L), *C. melanoscela* (Field collected, France, C. Villemant), *C. marginiventris* (laboratory reared, USA, Joyce,A), *C. vestalis* (Field collected, Benin, Guilloux,T), *C. rubecula* (laboratory reared, Netherland, Smid,H), *C. sesamiae* (Field collected, Kenya, S. Dupas). All specimens were placed in 95% ethanol and preserved at –20°C until DNA was extracted.

DNA extraction, amplification and sequencing

DNA was extracted with the « chelex » method from 2 individuals for each species except for *C. rubecula* where 1 individual was used. Briefly, individuals were ground in a 5 % chelex 100 resin (Biorad) solution with proteinase K (0.12 mg/ml) and incubated at 56°C for 30 min, then incubated at 95°C for 15 min and supernatants were collected. The primers were designed

according to the sequences of the CcBV PTPs: each pair of primers flank motifs 1 to 10 that characterize PTPs and are specific for each PTP. Primer sequences are listed in table 1. PCR conditions varied in stringency depending on whether amplified PTP genes belonged to closely related species (55°C annealing temperature and 1,5 mM of MgCl2) or to distantly related species (annealing at 45°C and 3 mM of MgCl2) of CcBV. One µl of DNA was used for each PCR reaction. The standard PCR program was : 95°C for 2,5 min ; 30 cycles at 95°C for 30 sec, annealing for 45 sec, 72°C for 60 sec ; 72°C for 5 min. The amplimers were purified with the Qiaquick kit (Qiagen) and sequenced directly. For most PTP genes direct sequencing was possible and sequence profiles did not show multiple peaks suggesting allele mix. For PTP Cα and PTP EX, direct sequencing was not possible therefore cloning of PCR products (Qiagen cloning kit) was performed and 10 clones for each gene were sequenced. The sequencing reactions were performed with the BigDye Terminator Sequencing Kit (Perkin Elmer ABI) and analysed on an ABI PRISM 3100 Genetic Analyzer.

Table 1 : Primers used for PTP amplification in different *Cotesia* species

Name	Direction	Sequence
PTPR5	Forward	TGGATCGGACTATTGATGAGC
PTPR3	Reverse	ACAGCTTGTTGGATCGGAGT
PTPΔ5	Forward	GCAAACAAAATGGCACATGA
PTPΔ3	Reverse	TCGGACCGGACTTGTCTTTA
PTPH5	Forward	CCACATTTTTCAAAGTTGGTGA
PTPH3	Reverse	CTGAACAACAAATCCACGTCA
PTPS5	Forward	TGGCTACCAACCTCTCAATG
PTPS3	Reverse	TCCAGCGACAATAAATACGC
PTPM5	Forward	CCGATTTGTTTGCACCTTTT
PTPM3	Reverse	AAGTGCAACAAAACACTGTGC
PTPE5	Forward	TGAGCAAGTAGCCGAATCAAG
PTPE3	Reverse	CGATGACAGAATAATCGTTT
PTPX5	Forward	GAAGCAAGTAGCTGAATCTGA
PTPX3	Reverse	CGATTACAGAGAAATCGGAA
PTPL5	Forward	GAATGCAAAAACTCGCCATT
PTPL3	Reverse	TGCAGGCAATCGTATCTTTG
PTPK5	Forward	TTTTCTGGAGACCTGGGAAA
PTPK3	Reverse	GAACGCGTTAAATAGAAACGAA
PTPQ5	Forward	TGGGTTGTGGCAACTCTAAA
PTPQ3	Reverse	GATATGTCAATGGCGCAGAA
PTPP5	Forward	TTCAAAAACGCTAAGCCGTAA
PTPP3	Reverse	TGCCTTTCCTTTCTTAGATTCG
PTPY5	Forward	TGGGGAGTGGAAATTCTAAGTC
PTPY3	Reverse	TCAACATAACAAAGCAAAACTGC
PTPC5	Forward	CGAAGAGCTATCTGCCGTTG
PTPC3	Reverse	TGTTTTATCGAGTGAGTTCT
PTPα5	Forward	TGAGTACCGAATTCGAAGAGC
PTPα3	Reverse	TGTTTTATCAATTGAGTCCC

PTP duplication patterns

In order to understand molecular mechanisms which induced the high PTP gene numbers we searched for homologous PTP genes in CcBV, CvBV, GiBV and MdBV. Thirty-seven PTP protein sequences from CvBV, 9 protein sequences from GiBV, 13 protein sequences from MdBV and 27 PTP protein sequences from CcBV were blasted and genes were considered to be orthologous when identities were higher than 80%. PTP genes were mapped within virus genomes by using gene positions in Genbank.

Sequence analysis and phylogeny

We isolated 87 PTP genes from *Cotesia* bracoviruses and other orthologous PTP sequences from CcBV, CvBV, GiBV and MdBV PDV genome sequencing projects were joined to the analysis (Espagne et al., 2004; Gundersen-Rindal and Pedroni, 2006; Webb et al., 2006). (CcBV PTP K, L, Q, P, M: AJ632304; CcBV PTP S, E,C:AJ632313; CcBV PTP α, X, Y: AJ632319; CcBV PTPH: AJ632307, CcBV PTPR: AJ632310; CvBV PTP2: AY871265; CvBV PTP3: AY651829; CvBVPTP6: AY651829; CvBV PTP10: AY651828; CvBV PTP11:DQ075354; GiBV PTP4: AY871265; GiBVPTP3:AY871265)
Translated sequences were aligned manually in McClade version 4.03 (Maddison, 2001) based on the PTP conserved motifs (Andersen et al., 2001).

This sequence alignment was used to construct a tree in order to have an overview of PTP evolution in Microgastrinae. Using Modeltest version 2.2 (Posada and Crandall, 1998), the GTR+I+G model of sequence evolution was selected according to the likelihood ratio test (LRT) and the Akaike information criterion (AIC). Bayesian MCMC analyses were performed for the entire data set using MrBayes version 3.12 (Ronquist and Huelsenbeck, 2003). Two independent analysis were run simultaneously for each data set, each consisting of 1000000 generations, sampling every 1000 generations and using four chains and uniform priors. Maximum parsimony (MP) analyses were performed using a heuristic search with stepwise random addition sequence. Support values for internal nodes were estimated using a non-parametric bootstrap resampling procedure after 100 replicates.

PTP genes studied could be separated in 3 monophyletic subclades that shared conserved motifs. Each subclade was studied independently, the tree topologies were obtained using maximum parsimony (MP) in PAUP 4.0b10 (Swofford, 2002) and Bayesian inference in MrBayes 3.12 (Ronquist and Huelsenbeck, 2003). For MP analysis, we performed a heuristic search starting with stepwise addition trees replicated 10 times and using a simple input order of

sequences to get the initial tree. Robustness of MP topologies was assessed by bootstrap with 1000 replicates (full heuristic search) of 10 random stepwise additions. For Bayesian inference, the best substitution model was selected using Modeltest 3.7 (Posada and Crandall, 1998). When this model was not available in MrBayes, the closest generalisation of the selected model was used, which for all clades happened to be the GTR+I+G model. The data were partitioned by codon position. We performed a 1000000 generation run sampled every 1000 generations on 4 incrementally heated chains. The burnin period was estimated by plotting likelihood values against generation time and after 20000 generations all were stabilized.

Branch and site selection analyses

The bayesian consensus trees were chosen as a phylogenetic hypothesis for the estimation of nonsynonymous to synonymous substitution rate ratio (ω = dN/dS) models on each clade using PAML 3.14. Six different models of site- and/or branch-specific ω ratios (Yang and Nielsen, 2002; Zhang et al., 2005) were optimised using Bayesian methods in PAML 3.14 (Yang, 1997). The maximum likelihoods of each model tested were compared between all models by the Akaike information criterion (AIC) and between nested models by the Likelihood ratio test (LRT).

Site specific positive selection was tested by comparing the selective model M8 (beta + ω>1) to the non selective model M8a (beta + ω = 1) in a likelihood ratio test (LRT) (Swanson et al., 2003). Branch specific selection was tested by comparing models M0b (branch specific selection and no variation among sites) to M0 (no branch or site specific selection) using a LRT.
The sites under selection were determined on different clades or subclades. For each gene clade or gene subfamily clade, the best model selected by AIC was considered and the percentage attributed to each class of sites (conserved ω<0.2, nearly neutral 0.2<ω<1, neutral ω=1 and selected ω>1) was recorded. When branch specific selection was shown to better explain lineage evolution by AIC and LRT tests, the site attribution was performed separately on positively selected branches (ω>1 in model M0b) and non positively selected branches (ω<1 in model M0b) using the branch and site model (MA new) (Zhang et al., 2005). In this model, two classes of branches are considered; the foreground branches where positive selection is allowed ($0<\omega_0<1$; $\omega_1=1$; $\omega_2>1$) and the background branches where sites are conserved ($0<\omega_0<1$; ω_1 =1). In our case we assigned the branches selected in M0b to the foreground branches and the branches not selected in M0b to the background branches.

Results

The PTP genes studied form three different clades

PTP genes can be organized by clusters based on phylogeny and on conserved motifs. PTP phylogeny revealed three major clades well supported among the Bracoviral PTPs studied (Figure 4). The PQLKYCα clade is formed by P, Q, L, K, Y, C and α PTP genes in which all PTP genes are separated in distinct and well supported monophyletic groups of homologues belonging to different species. Three orthologous PTPs were found in GiBV; PTP2, 3 and 4. The RΔ clade formed by R and Δ PTP gene families is also monophyletic. And finally the MHSEX clade, where the H-S, M, and EX PTP genes formed distinct subclades. No MdBV PTPs were found to group with the PTPs studied suggesting that MdBV PTPs are too distantly related to CcBV and GiBV PTPs to be considered as orthologous genes.

Bracovirus PTP sequences carry the 10 conserved motifs that define the protein tyrosine phosphatase family (Andersen et al., 2001) (Figure 5). Two kinds of motifs can be described, the structural motifs which are involved in PTP secondary or tertiary structure (motifs 2, 3, 4, 5, 6, 7) and motifs which are involved in phosphotyrosine recognition and activity (motifs 1, 8, 9 and 10). Conservation of these motifs in bracovirus PTPs differ depending on PTP clades considered. In the PTP RΔ cluster all motifs are strongly conserved except for a few differences in motif 4 where the conserved residues Gln (Q)-Gly (G) are mutated to Glu(E)-Ala(A) in PTP Δ and to Gln(Q)-Ala(A) in PTP R.

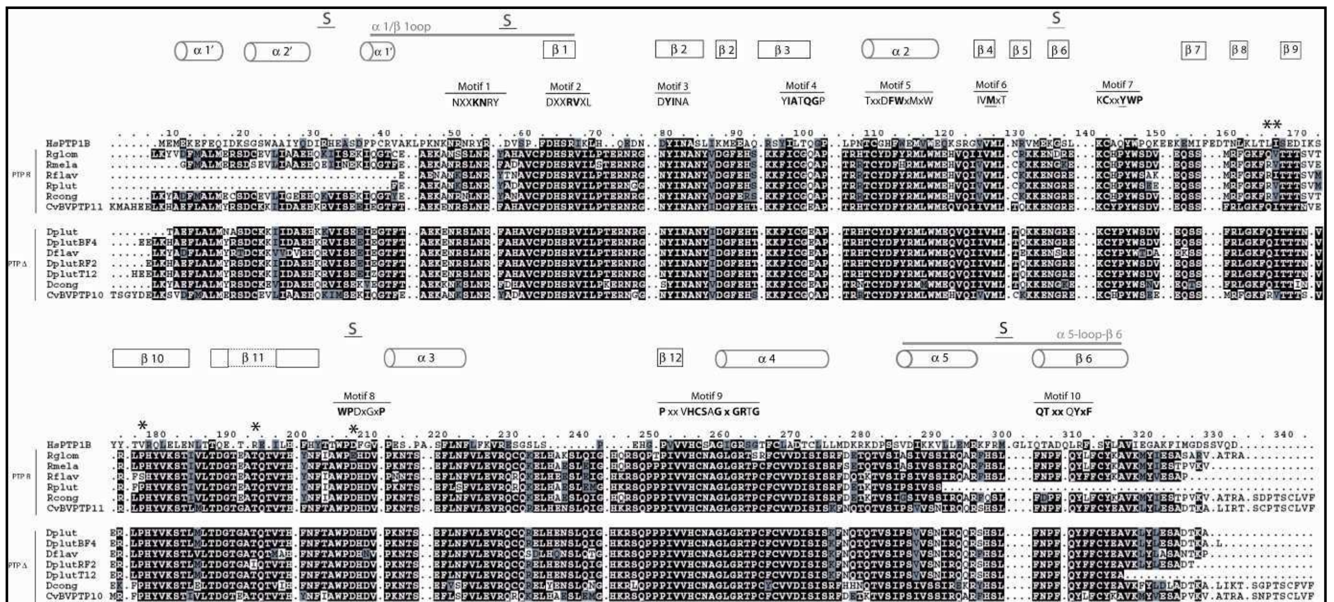
In the MHSEX cluster most motifs are well conserved. Amino acid differences occurring within motifs are usually conserved among species within a same PTP gene. For example, in motif 4 the conserved residues Gln(Q)-Gly(G) are mutated to Glu(E)- Gln(Q) in PTPM or to Gln(Q)- Glu(E) in PTP HS and PTP EX, in motif 5 the conserved residues are mutated to Tyr(Y)-Trp(W) in PTP EX, and in motif 8 the conserved Pro(P) is mutated to Thr(T) or Ala(A) in PTP EX. In all gene families of the MHSEX cluster, the most radical difference occurs in motif 9 where the Cys(C) residue which confers the PTP activity is mutated to Gly (G) or Ser(S). Numerous differences are observed in the PQLKYCα cluster, but again these differences are mostly conserved within PTP gene families. Despite high divergence among these genes, motif 9 carrying the active site shows a high conservation. In conclusion, PTP domains are overall well conserved in Bracovirus PTPs and when differences are observed they are shared within the

same PTP gene cluster.

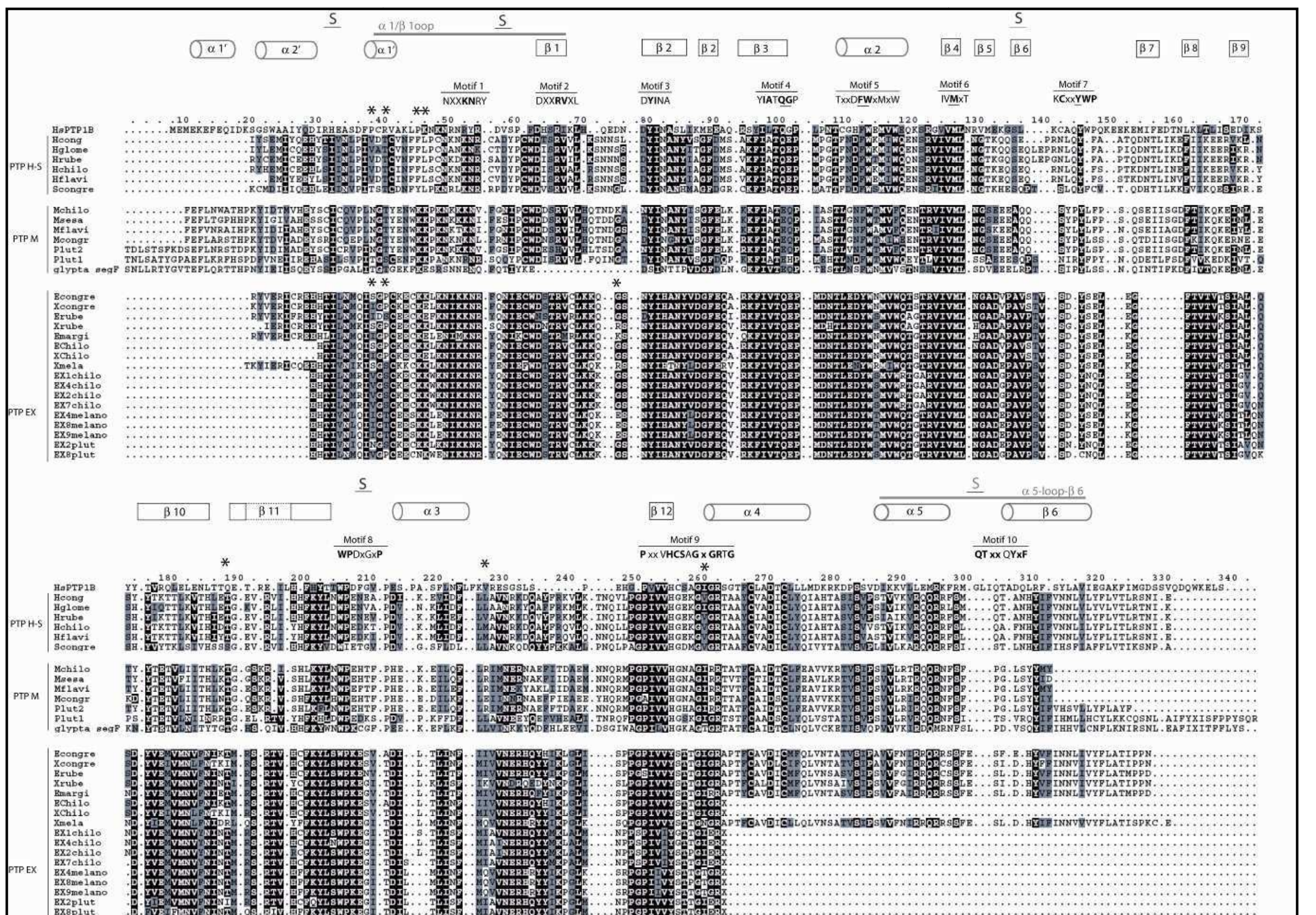
All together, conserved motifs and phylogenetic analyses allow us to separate the bracovirus PTP studied in three main clusters each constituted by PTP gene subclades.

Figure 4 Unrooted PTP phylogenetic tree from Bayesian inferences under the GTR+I+G substitution model
Posterior probabilities and non parametric bootstraps are indicated between major clades. Bootstrap values inside groups of orthologs are not shown for readability. Sequence names are given in abbreviated form.

A)



B)



C)

Figure 5 PTP alignment, conserved motifs and sites under positive selection.
(A) RΔ clade, (B) PTP MHSEX clade and (C) PTP PQLKYCα clade. PTP sequences are represented with the same amino acids numbering to facilitate comparisons. Stars indicate sites under positive selection for each clade and for PTP Cα and EX genes. Structural motifs are indicated. Variable regions α1/ β1 loop, α5-loop-α6 and sites called S involved in peptide binding are indicated.

Duplication patterns studied by comparison of three Bracoviruses species

Cross species BLAST analysis using PTPs from CcBV, CvBV, GiBV as queries allowed us to infer homologies between PTPs and to determine orthology between bracovirus genome segments (Figure 6).

27 genes present in CcBV, 23 CvBV PTP genes and 6 GiBV PTP genes were found to be orthologous and among them the synteny was conserved. Indeed the GiBV segment F, CvBV segment 28 and CcBV circle 1 present orthologous PTP families organized in the same order and are therefore considered as homologous segments. The same was shown between CvBV segment 36, 27, 30, 26 and 50 and CcBV circle 17, 10, 7, 2 and 4 respectively (Figure 6).

Moreover this analysis allowed us to infer molecular mechanisms which occurred during evolution resulting in PTP gene diversity observable today.

In CcBV circle 1 and its potential orthologs in other species, 6 orthologous PTP genes are found in the three species. The genes P Q and L group together in a monophyetic group (Figure 4). Each gene is found to group with an ortholog in GiBV segment F (genes 3, 2 and 4) or in CvBV segment 28 (gene 4, 3, 6), suggesting the duplications that led to circle 1 copy occurred before the divergence between *Glyptapanteles* and *Cotesia* bracoviruses. After the split between *Cotesia* and *Glyptapanteles* genus, GiBV PTP9 ancestral form was lost and the orthologous PTP 2 and D genes were acquired by the *Cotesia* bracoviruses. CcBV lost the orthologous genes of CvBV PTP 8 and 9 but acquired PTP K which is closely related to PTP L and thus most probably results from a recent tandem duplication event.

The CvBV segment 36 and 27 and the CcBV circle 17 and 10 were shown to be orthologous between the two species respectively, and within species each segment pair are paralogues that have evolved from duplication. Indeed, CvBV segment 36 and 27 correspond to the CcBV circle 17 and 10 respectively and moreover CvBV segment 36 is a near replicate of segment 27 and CcBV circle 10 is a near replicate of circle 27. These replicates were shown to be tandemly associated in the proviral form of CcBV (Bezier, unpublished). This pattern of circles duplicated in tandem was probably produced by segmental duplications. After this duplication the orthologous genes PTP 21 and Y and the orthologous genes PTP 34 and N were acquired in ancestral *Cotesia* bracovirus. Finally PTP 35 and PTP 19 were lost in CcBV circle 17 and circle 10 respectively.

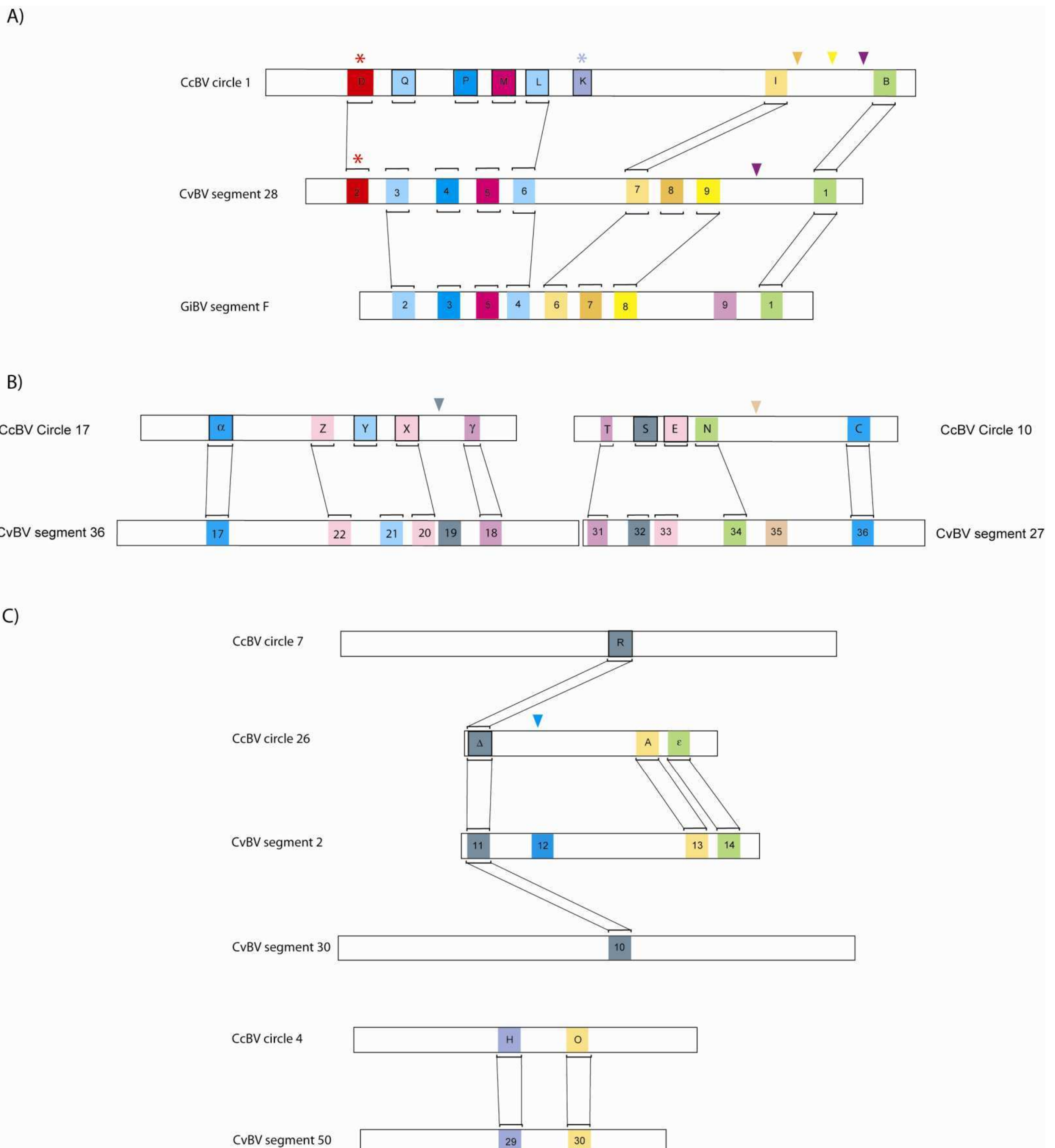


Figure 6 Orthologous genomic regions from CcBV, CvBV and GiBV.
PTP gene names correspond to those used in GenBank. (A) Orthologous genes between CcBV circle 1, CvBV segment 28 and GiBV segment F, (B) Orthologous genes between CcBV circle 17 and 10 and CvBV segment 36 and 27, (C) Orthologous genes between CcBV circle 7, 26 and 4 and CvBV segment 2, 30 and 50. Stars indicate gene acquisition and triangles gene loss. The same color is used between orthologous genes and between common events of gene acquisition or gene loss. Genes outlined are CcBV genes studied in PAML and phylogeny analyses.

Bracovirus genome comparisons allowed us to pin point two molecular mechanisms at the origin of PTP families. First, segmental and tandem gene duplications which either occurred in ancestral lineages or more recently in specific bracovirus lineages. Second, loss and acquisition of genes which occurred between genus but also within bracovirus species. Extension of the PTP gene family can also be explained through dispersed duplications. Indeed, according to PTP phylogeny PTP M, H, S, E, X and PTP R and Δ (and their homologs in CvBV) respectively form monophyletic clades which implies each set of genes shares a common ancestor. However, these PTP genes are found in different genome segments, for example R and Δ genes are found in circles 7 and 26 respectively and PTP M genes related to PTP E and X are found in circle 1 whereas PTP E and X genes are found on circles 10 and 17.

As a conclusion, duplications appear as a major molecular mechanism involved in PTP diversification including gene and segmental duplication as well as tandem and dispersed duplications. These duplications occurred at different times during PTP diversification after or before wasp divergence and were acompagnied by gene loss particularly in CcBV.

Episodes of positive selection during PTP evolution

To determine if selection pressures are acting on PTP gene evolution we measured branch specific selection in the 3 PTP clusters independently. Figure 7 shows phylogenetic relationships obtained from Bayesian analysis along with branch specific selection obtained in PAML and represented by branch width. MP (maximum parcimony) and BI (Bayesian inference) majority rule trees were generally congruent and when differences occurred, the bayesian tree was chosen to provide phylogenetic hypothesis for selective pressure inferences. Each of the 3 PTP gene family clusters are shown to diverge under varying selection pressures depending on the branch. Indeed the comparison of the M0 and M0b models by LRT indicates that a model in which branches evolve under different selective pressures (M0b) explains better the evolution of the PTP gene family rather than a model which does not (M0) (Table 2). These analyses enabled us to determine and to define the selective pressures which governed PTP gene evolution before and after duplication events.

For the PQLKYCα cluster, M0b model fit significantly better our data compared to M0 model ($p<0.05$). Branches supporting PTP genes of the PQ and KLCα clades were found to be highly selected ($\omega \gg 10$), suggesting that natural selection acted after a duplication event. In contrast, after this initial duplication, branches supporting the PTP K, L and Cα on one hand and Q and P on the other hand were not followed by episode of positive selection but evolved

rather under purifying and nearly neutral evolution (0.2<ω<0.7). A second episode of positive selection was visible within PTP Cα, P, Q and Y clades, this is particularly true for PTP C and α genes which arose from a segmental duplication, for which more than 40% of branches show ω>1 (Table 3).

PTP RΔ cluster evolution was also better explained by the M0b evolutionary model. The ancestral copy of the PTP R cluster diverged under strong positive selection (ω>10) in contrast to the ancestral copy of PTP Δ cluster (ω<0.03). Positive selection was also detected within PTP R and Δ genes where 16.9 % and 2% of branches were observed to be positively selected (ω>2) respectively.

Figure 7 Unrooted phylogenetic trees from Bayesian inferences under the GTR+I+G substitution model of the three major PTP clades.
Posterior probabilities and non parametric bootstraps are indicated on the left of clades branchs. Branch widths are proportional to the ω ratio estimated under the branch-specific model in PAML (A) PTP PQLKYCα clade; (B) PTP MHSEX clade, (C) RΔ clade.

Finally, the MHSEX cluster also presented branches evolving under different selective pressures. In this case, branches supporting each PTP gene were not showed to have evolved under positive selection but were better explained by evolution under purifying selection (ω<0.5). The M0b model is probably statistically significant thanks to the EX cluster, in which 26.5% of branches evolved under positive selection (Table3). In the EX cluster, numerous PTP forms are found and they appear to have diverged under positive selection between and within species (1.7<ω< 7 and ω>10). PTP E and X were shown to have arisen from a segmental duplication and positive selection would have acted to induce PTP E and X divergence.

To conclude, except for the MHSEX cluster where PTP clade ancestral lineages principally evolved under purifying selection, duplication events in other PTP clades were followed by episodes of positive selection. A second episode of positive selection acted more recently within PTP clades and is still active in PTP R and Δ and in PTP C and α, as well as in EX clade. Interestingly these three clades correspond to those in the data set analyzed that originated from whole segment duplication instead of tandem gene duplication. This suggests a particular role of segment duplication in evolutionary dynamic f PDVs.

Positive selection acted on specific residues

To determine whether the positive selection observed acted on specific amino acid residues, we measured site selection in the three PTP clusters. We tested site selection model (M8) and branch-site selection model (MA), as shown in table 2.

Evolutionary model comparisons showed that the clusters MHSEX and PKLKYCα are better explained by selective models. The model selected by AIC and LRT was M8 for MHSEX and PKLKYCα. In contrast, the RΔ cluster is better explained by the M8a model which is non selective. When site selection analysis was performed within Cα and EX PTP families, amino acids in position 83, 94 and 281 for PTP Cα and 38, 40, 75 for PTP EX were shown to be positively selected.

Using the M0/M0b test, we showed that some PTP lineages evolved under different selective pressures (previous section). When M0b proved to better explain PTP lineage evolution, we used the MA/MAnull model comparison to test wether particular sites evolved under positive selection. The MA model proved to better explain PQLKYCα, MHSEX and RΔ cluster evolution. Site analysis on the entire PQLKYCα clade revealed 7 amino acids under positive selection (at positions 87, 88, 120, 136, 158, 173 and 219) none of these sites are common to

positively selected sites identified specifically in the Cα clade. These results suggest that a first episode of positive selection induced divergence between PTP families and that a second episode acted on sites within PTP Cα. Within the MHSEX cluster, 9 sites were detected to be positively selected (positions 38, 40, 45, 46, 188, 227, 260) and 2 of these sites are common to positively selected sites identified in PTP EX (positions 38, 40, 75). These results suggest that PTP EX clade divergence was accompanied by changes of particular amino acids by positive selection, and a second episode of positive selection acted on particular amino acids within the PTP EX. Within the RΔ cluster, 5 sites are under positive selection (positions 165, 166, 177, 193, 207).

Table 2 Branch, Site and Branch-site model comparisons and position of positively selected sites.

Cluster	***AIC rank order***	***LRT***			***Sites selected (site number)***
		M8/M8a	***M0b/M0***	***MA/MAnull***	
PQLKYCα	**M8<M8a<MA<MAnull<M0b<M0**	**$P<10^{-4}$**	**$P<0.05$**	**$P<0.0005$**	**87, 88, 120, 136, 158, 173, 219**
P	M8a<M8< M0b<M0	NS	NS	-	
Q	M8a<M8<M0<M0b	NS	NS	-	
K	M8a<M8<M0b<M0	NS	NS	-	
L	M8<M8a<M0<M0b	NS	NS	-	
Y	M8a<M8<M0b<M0	NS	NS	-	
Cα	M8<M8a<M0B<M0	$P<0.01$	NS	-	83, 94, 281
MHSEX	**MA<M8<M8a<MAnull<M0b<M0**	**NS**	**$P<10^{-4}$**	**$P<10^{-5}$**	**38, 40, 45, 46, 188, 227, 260**
M	MA<MAnull<M8a<M8<M0b<M0	NS	$p<10^{-4}$	NS	
H	MA<M0b<M8a<MAnull<M8<M0	NS	$P<0.05$	NS	
EX	M8<M8a<M0<M0b	$P<10^{-5}$	NS	-	38, 40, 75
RΔ	**MA<MAnul<M8A<M8<M0b<M0**	**NS**	**$P<10^{-2}$**	**$P<0.05$**	**165, 166, 177, 193, 207**
R	MA<MAnull<M0b<M8a<M8<M0	NS	$P<0.01$	NS	
Δ	M0b<M0<M8a<M8	NS	NS	-	

Model comparisons were performed using Akaike Information Criterion (AIC) between all models and Likelihood Ratio Test (LRT) between nested models. Model descriptions: Branch selection models: M0=one class of ω ratio and M0b=tree branches have different ω ratio. Site selection models (ω ratio varies according to two classes); M8a:beta distribution of ω_0 and $\omega_1=1$ and M8: beta distribution of ω_0 and $\omega_1>1$. Branch-site selection models (ω ratio varies among sites in specific lineages): MA: two classes of sites; $0<\omega_0<1$ and $\omega_1=1$ and MA: three classes of sites; $0<\omega_0<1$, $\omega_1=1$ and $\omega_2>1$.

Table 3 Percentage of branch length for 4 classes of ω ratio for the different genes.

PTP genes	% of clade branch length with :				
	ω<0.2	0.2<ω<0.5	0.5<ω<1	1<ω<2	ω>2
P	31.9%	30.1%	30.3%	0.0%	7.7%
Q	0.0%	98.6%	0.0%	0.0%	1.4%
L	0.0%	78.5%	21.5%	0.0%	0.0%
K	2.5%	12.8%	84.6%	0.0%	0.0%
Y	5.0%	68.7%	12.1%	14.1%	0.0%
M	8.7%	45.5%	41.37%	4.4%	0.0%
H	0.0%	75.0%	4.14%	0.0%	0.0%
EX	1.7%	53.6%	18.15%	14.0%	12.5%
R	52.9%	30.3%	0.00%	0.0%	16.9%
Δ	18.7%	30.0%	49.19%	0.0%	2.0%
Cα	0.0%	11.6%	44.35%	22.5%	21.6%

The ω ratio for each branch was estimated using branch specific codon substitution models in PAML.

In conclusion, positive selection is shown to have acted between PTP families of a same cluster to fix particular amino acids but also between and within closely related PTP families (PTP Cα and PTP EX) inducing divergence between PTP lineages from different species or within a same species.

Positively selected sites are different according to the PTP cluster which means that each cluster diverged by selection acting on different amino acids. Most selected sites are in the vicinity of PTP conserved domains, and two positively selected sites are within motif 8 and 9. The amino acid in position 207 is positively selected in the RΔ cluster; the aspartic acid commonly found in other PTPs is mutated to glutamic acid in PTP R of *C. glomerata* bracovirus. Mutations in this site are expected to modify PTP activity efficiency (Andersen et al., 2001). In the MHSEX cluster, one positively selected site is found in motif 9 which carries the catalytic site. Interestingly, by co-cristallography and Cα-regiovariation in human PTP 1B, two regions (α1/ β1 loop and α5-loop-α6) and four specific areas located in proximity to the active site were shown to be involved substrate specificity. (Andersen et al., 2001). In MHSEX we found that positively selected residues number 38, 40, 45 and 46 fall in the α1/β1loop region and in Cα, residue 281 is found

near α5-loop-α6 (see figure 5). In these areas, the combination of residues is unique and could consequently represent a region determining protein specificity.

Discussion

The PTP gene family is known to be the most diversified family found in polydnaviruses associated with Braconid wasps (Bailey and Eichler, 2006; Desjardins et al., 2007; Espagne et al., 2004; Webb et al., 2006). Our study focused essentially on PTP genes from Bracoviruses associated with *Cotesia* genus wasps, for which 27 genes have been found in CcBV. Furthermore, PTP genes are annotated in the PDV genome offering interesting support to understand molecular duplication mechanisms which conduced to this large gene family. Therefore, bracovirus PTPs are a particularly interesting gene family to study molecular and evolutionary mechanisms involved in duplications and to understand the role of these duplications in a biological context.

PTP duplication mechanisms

Three major mechanisms are suggested to be involved in PTP diversification; first, large genomic regions have been duplicated (segmental duplication) and secondly individual genes are duplicated either in tandem or dispersed. Segmental duplications are clearly involved in PTP diversification. Indeed CcBV circle 10 and 17 were shown to be linked in the wasp genome (Bezier, unpublished) and harbour 5 homologous PTP genes suggesting they arose from a segmental duplication. These processes have previously been proposed to play a critical role in primate evolution in creating new genes and shaping human genetic variation (Bailey and Eichler, 2006). They seem particularly important in stimulating evolutionary changes since most of the recently selected Bracovirus PTP lineages were located in clades that emerged from segment duplication.

Tandem duplications are thought to be the major mechanism for the creation of new genes and have been documented in several organisms (Fan et al., 2008; Ganko et al., 2007; Hoffmann et al., 2008; Hooper and Berg, 2003b). In bracoviruses this pattern was observed for PTP K, L, P, Q genes found in CcBV circle 1. According to PTP phylogeny we can suggest that these genes were produced after two rounds of duplications which occurred at different periods.

First, the PTP P, Q and L genes emerged from an ancient duplication shared by *Glyptapanteles* and *Cotesia* bracoviruses whereas PTP K which is the recent PTP L duplicate is only shared by some *Cotesia* Bracoviruses. These results emphasize that PTP tandem duplications constitute dynamic processes which appear to be lineage specific. Tandem and segmental duplications are expected to produce genes or genomic regions closely associated in the genome. However, dispersed duplications which produced closely related genes were also shown to be involved in PTP diversification. How could these particular duplication patterns arise? It has been proposed that dispersed duplications arise from RNA mediated retrotransposition. This process is allegedly mediated by retrotransposons and produces intronless genes. Bracovirus PTPs studied to date are all characterized by their lack of introns (Espagne et al., 2004; Gundersen-Rindal and Pedroni, 2006; Webb et al., 2006) but other traces of RNA mediated retrotransposition have not been investigated to date. These patterns of duplications are important to consider in regard to gene function because tandem, segmental and dispersed duplications have different consequences on gene regulation. Indeed, genes originating from tandem or segmental duplications tend to maintain a similar function to their parental copy due to their sharing the same regulatory elements (Arisue et al., 2007; Darbo et al., 2008; Ponce and Hartl, 2006). In contrast, dispersed copies are expected to develop different functions since they are separated from their original regulatory elements (Wang et al., 2002). Interestingly, Weber and colleagues showed in *Chelonus inanitus* Bracovirus that genes found in the same PDV segment are similarly expressed suggesting that genome segmentation plays a role in gene regulation (Beck et al., 2007; Weber et al., 2007). Moreover, several studies suggest that DNA circles are not produced at the same rate in CcBV (Provost et al., 2004). Therefore, gene dispersion following gene duplications would tend to change gene replication rate and gene regulation pattern. The different duplication processes which led to the expansion of the Bracovirus PTPs probably played an important role in PTP function diversification and consequently in wasp parasitism success. Moreover, Bracovirus PTP gene expansion was shown to be accompanied by gene loss, implying that PTP gene evolution was conducted by the “Birth and Death” model described by Nei and colleagues (Nei and Rooney, 2005). According to this model, genes arise continually by duplication and are lost by deletion or by mutational events. Therefore, some PTP ancestral lineages were lost while others were created by duplications and were transmitted in particular lineages. As it has been shown in primates or in *Drosophila*, gene expansion and contraction could explain important adaptive traits illuminating the physiological adaptations of their host species (Babushok et al., 2007; McBride, 2007). By studying bracovirus PTP genes, we showed that genome reorganisation occurred at a very fine evolutionary scale with gene acquisition and loss occurring between species and

pseudogenization was also observed for some PTPs in particular bracovirus species (data not shown). Braconid wasps associated with polydnaviruses have been shown to be a highly diversified group with a very narrow host range (Smith et al., 2008) and virus genome plasticity could be viewed as a powerful mechanism allowing wasp adaptive radiation. Indeed PTP gene expansion and more generally virus genome expansion may be a source of evolutionary innovations offering wasps dynamic adaptive properties.

How did PTP family divergence occur?

PTP genes underwent several duplication events which appear to be lineage specific. Gene duplication has been considered as the most important mechanism in creating new genes. It is therefore essential to understand evolutionary forces which could explain how duplicated copies evolve new functions.

Our analysis emphasizes that PTP gene families did not evolve under the same selective pressures and PTP evolution underwent two episodes of positive selection. The first episode of positive selection occurred after a duplication event and before the speciation processes. The second episode occurred between and within PTP families after speciation processes. Understanding evolutionary processes underlying divergence of duplicated copies is of major interest to determine how genes can be innovated and how new functions could appear. For classical models, duplications are selectively neutral and maintenance of duplicated copies depends on a beneficial mutation which appears randomly and increases in frequency in the population (Hugues, 1994; Ohno, 1970). In contrast both in the "escape from the adaptive conflict" (EAC) (Des Marais and Rausher, 2008) and in "the innovation, amplification, divergence" (IAD) (Bergthorsson et al., 2007) models, duplications are immediately advantageous allowing the increase of protein amount.

The fact that positive selection is shown to be involved in bracovirus PTP copy divergence is concordant with the second class of models. How could PTP function be improved and innovated after duplications? The EAC model assumes that a novel function arises in the ancestral gene and after duplication each copy is selected to improve ancestral or novel function. For the IAD model, the novel function existed as minor activity in the ancestral gene and its duplication offered multiple targets to improve this function. The particular characteristics of polydnavirus life cycle may favour the EAC model. The virus is transmitted by the wasp but expressed by another organism. Therefore the wasp does not pay directly the cost of gene expression. This low cost of functional gene duplication may favour acquisition of new function.

During the period of gene change, the gene may escape from the adaptive conflict, and not be lost due to the limited cost payed by the wasp. Some Bracovirus PTP genes play an important role in host immune alteration particularly by modulating PTP cell activity in hemocytes (Pruijssers and Strand, 2007). Interestingly, one particular class of PTP have been identified for not carrying PTP activity, they were proved to reduce PTP cells activity probably through competition with host PTPs (Ibrahim and Kim, 2008; Pruijssers and Strand, 2007). Furthermore, it has been shown that PTPs are differentially expressed in the course of parasitism suggesting they performed different functions (Gundersen-Rindal and Pedroni, 2006; Ibrahim et al., 2007; Pruijssers and Strand, 2007). Interestingly, in a baculovirus infecting lepidopteran, PTP play a major role in host behaviour manipulation. They were shown to enhance locomotory activity and thus increase baculovirus transmission (Kamita et al., 2004). Nevertheless we are still unable to determine the precise role of each bracovirus PTP and correlate PTP evolution with functional innovations. For that we should link the different functions of PTPs with mutational events.

Our analyses identified particular amino acids targeted in the different PTP families to evolve under positive selection. We showed that selected amino acids were different in the three PTP clusters and have been fixed within PTP families. One of these mutations was shown to occur in a conserved motif known to be important for PTP efficiency (Andersen et al., 2001). Furthermore, some amino acids were found in regions shown to be involved in PTP specificity. PTPs are known to be highly specific proteins (Andersen 2001), these residues may play a role in PTP specificity and they could be therefore related to a particular function.

The challenge will now be to determine how PTP function has been innovated through duplications by studying the role and regulation of PTPs in relation with mutational events occurring after duplications.

Conclusion

The virus-wasp association is the only mutualism involving a virus known so far. Viruses confer to wasps new essential functions ensuring wasp parasitism success. Our study reflects the dynamic evolution of virus genes undergoing multiple gene duplication events and several episodes of natural selection. In other viruses, genomes are normally limited to a few genes, in our system gene expansion is a common process observed for most virulence factors and may be related to the essential role played by the virus in wasp parasitism success. Gene duplications are

likely to have offered new sources of innovation allowing wasps to colonize new host environments.

Acknowledgment

Cindy Ménoret and Carole Labrousse are gratefully acknowledged for taking care of the insects. We are grateful to Jérôme Lesobre for help in processing sequence reactions.

This work was funded by an EC grant (QLK3-CT-2001-01586 "Bioinsecticides from insect parasitoids") and by the ANR grant "EVPARASITOID". CS was supported by a PhD grant from the French ministry de l'Enseignement Supérieur.

IV-Viral cystatin evolution and 3D structure modelling: a case of directional selection acting on a viral protein involved in a host-parasitoid interaction.

Céline Serbielle, Shafinaz Chowdhury, Samuel Pichon, Stéphane Dupas, Jérôme Lesobre, Enrico O. Purisima, Jean-Michel Drezen and Elisabeth Huguet

Abstract

In pathogens, certain genes encoding proteins that directly interact with host defences coevolve with their host and are subject to positive selection. In the lepidopteran host-wasp parasitoid system, one of the most original strategies developed by the wasps to defeat host defences is the injection of a symbiotic polydnavirus at the same time as the wasp eggs. The virus is essential for wasp parasitism success since viral gene expression alters the host immune system and development. As a wasp mutualist symbiont, the virus is expected to exhibit a reduction in genome complexity and evolve under wasp phyletic constraints. However as a lepidopteran host pathogenic symbiont, the virus is likely undergoing strong selective pressures for the acquisition of new functions by gene acquisition or duplication.

To understand constraints imposed by this particular system on virus evolution, we studied a polydnavirus gene family encoding cyteine protease inhibitors of the cystatin superfamily.

We show that *cystatins* are the first bracovirus genes proven to be subject to strong positive selection within a host-parasitoid system. A generated 3-dimensional model of *Cotesia congregata* bracovirus cystatin 1 provides a powerful framework to position positively selected residues and reveal that they are concentrated in the vicinity of actives sites which directly interact with cysteine proteases. In addition, phylogenetic analyses reveal two different *cystatin* forms which evolved under different selective constraints and are characterized by independent adaptive duplication events.

Positive selection acts to maintain *cystatin* gene duplications and induces directional divergence presumably to ensure the presence of efficient and adapted cystatin forms. Directional selection has acted on key cystatin active sites, suggesting that cystatins coevolve with their host target. We can strongly suggest that cystatins constitute major virulence factors, as was already proposed in previous functional studies.

Introduction

In a host-parasite interaction the associated partners can influence each others evolution (Woolhouse et al., 2002). Molecular signatures of these complex evolutionary processes can be detected in the genomes of both organisms involved in such associations. Indeed, genes encoding pathogenicity factors directly involved in counteracting host defences or vice-versa are expected to be subject to positive selection, driven by an arms race between the two partners. Such coevolutionary processes have been well described in certain plant-pathogen interactions, where the host resistance genes and corresponding avirulence genes in the pathogen show evidence of positive selection (Dodds et al., 2006). In the *Xanthomonas*-pepper interaction, the Hrp pilus, a filamentous structure allowing bacteria to directly inject toxins into plant cells, also evolves under positive selection, thereby avoiding the plant defence surveillance system (Weber and Koebnik, 2006). Positive selection has also been detected in insect-pathogen interactions. For example, in *Drosophila*, RNA interference molecules involved in anti-viral defence are among the fastest evolving genes in this insect. This rapid evolution is due to strong positive selection, illustrating that the host pathogen arms race between RNA viruses and host antiviral RNAi genes is very active and significant in shaping RNAi function (Obbard et al., 2006).

We are interested in characterising the evolutionary processes underlying the insect host-parasite interactions between lepidopteran hosts and parasitoid wasps. In these systems, the endoparasitoid wasp larvae develop inside the lepidopteran host despite the hostile environment this habitat represents. One of the most original strategies developed by these wasps to defeat these defences is the injection of a symbiotic polydnavirus (PDV) at the same time as the wasp eggs (Beckage, 1998; Beckage and Alleyne, 1997; Stoltz et al., 1984). PDVs are divided in two genera, ichnoviruses and bracoviruses, which are associated with tens of thousands of endoparasitoid wasps belonging to two different families, Ichneumonidae and Braconidae, respectively (Fleming and Krell, 1993). PDVs are found in these wasps as proviruses which are 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 after replication, are injected in the host caterpillar as multiple double stranded DNA circles packaged in capsids. 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.

In this biological system, the virus plays key roles both in the mutualistic association with the wasp and in the parasitic association between the wasp and the caterpillar. PDVs are therefore likely to display molecular signatures which reflect constraints imposed both by the wasp and the host caterpillar. So far however, reports have principally concentrated on the influence of wasp evolution on viral genomes. Braconid wasps carrying PDV form a monophyletic lineage, suggesting an unique event of association between the wasp ancestor and the virus ancestor and a vertical transmission of the virus along wasp lineages (Whitfield, 2002). Accordingly, a phylogenetic study of *Cotesia* spp. and their associated viruses has shown a codivergence between the two mutualists (Whitfield and Asgari, 2003). Finally, recent data on the genome sequence of several PDVs has revealed that these viruses harbour a large number of eukaryotic genes likely picked up from the wasp genomes. These genes form multigene families that are good candidates to be involved in alteration of host caterpillar physiology (Espagne et al., 2005; Espagne et al., 2004; Provost et al., 2004; Strand and Pech, 1995b; Webb et al., 2006). Surprisingly, very few studies have focused on the potential influence of the host caterpillar on viral gene evolution despite the strong selective pressure this habitat represents. In this paper, we report the molecular evolution of a viral gene family considering both wasp evolution and the selective pressure imposed by the caterpillar hosts.

Our model system is the interaction between the braconid wasp *Cotesia congregata* and its lepidopteran host, the tobacco hornworm, *Manduca sexta*. The PDV associated with *C. congregata* (CcBracovirus, CcBV) has been sequenced, revealing the presence of numerous genes possibly involved in host deregulation (Espagne et al., 2004). Among these viral genes, one gene family encoding cystatins constitutes an interesting candidate system to study the influence of the host-parasitoid association at the viral molecular level. Cystatins are tightly binding reversible inhibitors of papain-like cysteine proteases, and are widespread in plants and animals (Rawlings et al., 2004). They are characterized by three conserved domains forming the site of interaction with C1 cysteine proteases: an N-terminal glycine, a glutamine-X-valine-X-glycine motif and a C-terminal proline-tryptophane amino acid pair (Bode et al., 1988; Stubbs et al., 1990). 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). Remarkably, plant cystatins acting as defence proteins have been shown to evolve under strong positive selection in response to cysteine proteases released by phytophagous insects. In this system, it has been

suggested that plant cystatins and insect cysteine proteases are involved in a coevolutionary process (Kiggundu et al., 2006).

CcBV *cystatins* constitute the first description of *cystatin* genes in a virus and are organized in a multigene family, composed of three genes present on the same circle (Espagne et al., 2005; Espagne et al., 2004). To date, there is no evidence of *cystatin* genes in *Microplitis demolitor* bracovirus (MdBV) which has been fully sequenced (Webb et al., 2006) and they have only been identified in one other polydnavirus (GiBV) from the braconid wasp *Glyptapanteles indiensis* (Desjardins et al., 2007). Both genomic and physiological features of cystatins suggest that these viral proteins could play an important role in the host-parasite association. First, the genomic organisation in a multigene family could be indicative of selective pressures acting on these genes. Indeed, Francino (2005) (Francino, 2005) suggested that gene duplications that can lead to an increase in protein dosage are favored by selective pressures. Secondly, *cystatin* genes are expressed rapidly and at an extremely high level during parasitism. This early and prolonged expression could be indicative of a role of cystatins in the early steps of host physiological disruption, as well as in the maintenance of this perturbed state. Finally a recombinant viral cystatin (Cystatin 1) was shown to be a functional and specific cysteine protease inhibitor (Espagne et al., 2005).

In this study we checked for molecular signatures associated with positive selection that may act on the viral *cystatin* gene family. We demonstrate strong and lineage specific adaptive evolution acting on these genes. Using homology modelling and molecular dynamics simulation techniques we obtained the three dimension (3D) structure of CcBV cystatin 1. The predicted model of the 3D structure of CcBV cystatin provides a framework to position the positively selected residues, and reveals that these are situated in key sites which are important for the interaction with target proteases. This particular selection, which is probably imposed by host defences, emphasizes the potential role of cystatins as pathogenic factors and suggests that cystatins coevolve with host cysteine proteases.

Methods

Wasp specimens

Cystatin genes were isolated from nine viruses associated with the following *Cotesia* species: *C. congregata, C. flavipes, C. chilonis, C. melanoscela, C. vestalis, C. rubecula, C. sesamiae, C. kariyai and C. glomerata* (Table 4). These species provide a good representation of *Cotesia* species diversity based on the *Cotesia* phylogeny (Michel-Salzat and Whitfield, 2004).

Table 4 Wasp samples and primers used for cystatin gene amplification

Wasp species	Location	Collections	Primers	Species Abbreviations
C. congregata	Lab reared	Drezen, J.M (Fr)	Cyst15/Cyst93	*CcBV*
C. chilonis	Lab reared	Wiedenmann,. R (USA)	Cyst15/Cyst93	*CchBV*
C. flavipes	Kenya	Dupas, S (Fr)	Cyst15/Cyst93	*CfBV*
C. glomerata	Lab reared	Vet, L (NL)	Cyst15/Cyst93	CgBV
C. kariyai	Japan	Tanaka, T (J)	Cyst15/Cyst103	*CkBV*
C. melanoscela	France	Villemant, C (Fr)	Cyst15/Cyst93	*CmBV*
C. vestalis	Benin	Guilloux, T (Fr)	Cyst15/Cyst93	*CvBV*
C. rubecula	Lab reared	Smid, H (NL)	Cyst15/Cyst103	*CrBV*
C. sesamiae	Kenya	Dupas, S (Fr)	Cyst15/Cyst103	*CsBV*

DNA extraction, amplification, cloning and sequencing

DNA extractions were performed using the Chelex method from a whole individual wasp. Wasp tissues were disrupted in a 5% Chelex solution including proteinase K (0.12 mg/ml). Three primers for *cystatin* gene amplification were designed based on an alignment of the three *cystatin* genes from *C. congregata* bracovirus [EMBL: AJ632321] and one *cystatin* gene from *Glyptapanteles indiensis* bracovirus [genbank: AC191960]; one forward primer Cyst15 5'-ATGGGCAAGGAATATCGAGTG-3' and two reverse primers Cyst93 5'-GTAAGGACAGTTTTTATCTAG-3', Cyst103 5'-GTAAGGACGACTTTTATCTAG-3'. The amplified product is composed of 279 nucleotides and encodes a 93 amino acid sequence containing the first two conserved domains of cystatins. PCR amplification was performed in a 50 µl volume containing 1X Taq buffer, 3mM of $MgCl_2$, 2.5mM of dNTP, 0.3 µl Taq polymerase (Goldstar, Eurogentec) and 50 pmol of each primer. Goldstar polymerase displays a very good fidelity of one error every 5.10^{-5} bases. 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 species, 12 positive clones were sequenced in order to isolate all the *cystatin* gene copies and to obtain a minimum of two identical clones per sequence. Only CcBV21L and CcBV21I correspond to unique sequences. However excluding these sequences from the data set does not change the results of the analysis on PSS. Cloned inserts were sequenced in both directions using the Big Dye^{R} Terminator v3.1 cycle sequencing kit and the sequenced products were analysed using a capillary DNA sequencer (ABI PRISM 3100).

Sequence analysis and phylogeny

Cystatin sequence obtained and sequences already available from viral genome sequencing (CcBVcyst1, CcBVcyst2 and CcBVcyst3) were aligned using ClustalW implemented in Bioedit version 5.06 (Hall, 1999). We estimated the intraspecies and interspecies *cystatin* gene divergence using MEGA ver3.1 (Kumar et al., 2004). Divergence was calculated by a pairwise distance under the Kimura 2-parameter substitution model.

Recombination can mislead phylogenetic estimation and positive selection analysis. In order to avoid this bias we tested the *cystatin* gene family for recombination using a Genetic Algorithm Recombination Detection (GARD) implemented in Hyphy (Pond et al., 2005).

The program MrModeltest ver2.2 (Posada and Crandall, 1998) was used to determine the appropriate model of DNA substitution by the hierachical likelihood ratio test (hLRTs). Phylogenetic trees were obtained by Maximum Likelihood (ML) using PHYML program (Guindon and Gascuel, 2003) and by Bayesian inference in Mr Bayes 3.12 (Ronquist and Huelsenbeck, 2003). Modeltest chose the Kimura 80 model with a gamma distribution of parameter shape α=0.7875, a transition/transversion ratio of 1.12 and a proportion of invariables sites equal to 0. These parameter estimations were used as initial parameter values for ML and Bayesian inference. The topology and branch length estimation by ML was repeated 1000 times and for Bayesian analysis we performed 1000000 generations until the standard deviation was below 0.01.

Positive selection among sites

All the analyses on the rate of protein evolution among taxa and tests of positive selection were conducted using the codeml program in the PAML package v3.14 (Yang, 1997). Pairwise estimates of the number of non synonymous substitutions per synonymous site (d_N) and the

number of synonymous sites (d_S) were calculated using maximum likelihood (Goldman and Yang, 1994).

To test for evidence of positively selected sites, we performed different models allowing evolutionary rates ($\omega = d_N/d_S$) to vary across codon sites (Models M0, M3, M8A and M8) (Yang, 2000). M0 (one ratio model) assumes that all branches in the phylogeny and all sites have the same ω. The model M3 classifies sites in the sequence into three discrete classes with ω estimated from the data (Yang, 2000). M8A assumes a β-distribution of d_N/d_S ratio constrained to lie between 0 and 1.0 and adds to the β-distribution a point mass at $\omega = 1$ (Swanson et al., 2003) whereas the selection model M8 permits one additional d_N/d_S ratio to be above 1. Nested models (i.e., M0 vs M3 and M8A vs M8) (non positively selected vs positively selected models) were compared using the likelihood ratio test: 2X the log-likelihood difference between the two models can be compared to a χ^2 distribution, with the number of degrees of freedom equal to the difference between the two models. Codon sites under positive selection were identified using the Bayes Empirical Bayes (BEB) calculation of posterior probability for site classes (Yang et al., 2005) that analyses the sites under positive selection identified by the selective models. The numbers of substitutions between *cystatin* genes were counted using "Codeml" program in the PAML package (Yang and Swanson, 2002), with the F1X4 model of codon frequencies. Four sequences containing a stop codon were eliminated from the analysis. Each analysis was repeated ten times with different initial ω values to avoid problems of multiple local optima.

Positive selection among lineages

To test for evidence of positive selection among sites but also among lineages we performed a branch-site analysis using the codeml program in the PAML package v3.14. In this analysis, the branches under positive selection are called "foreground" branches and all other branches are called "background" branches. Sites changing in the foreground lineage are permitted to have $\omega>1$. Yang and Nielsen (2002) (Yang and Nielsen, 2002) implemented two versions of branch-site models called MA and MB. In MA, ω_0 is estimated from the data under the constraint $0<\omega_0<1$; hence positive selection is permitted only in the foreground branch. This model is compared with model M1a. In MB ω_0 and ω_1 are free parameters. Thus some sites evolve by positive selection across the entire phylogeny, whereas other sites evolve by positive selection in the foreground branch only. MB is compared with M3. Parameters used to perform this analysis are the same as those used in the site analysis.

The branch-site analysis was used to gain information on the possibility of different evolutionary constraints in different lineages. A problem with this method is that it assumes an *a priori* hypothesis. Indeed we have to specify foreground and background lineages with no knowledge on lineage history or on the type of substitutions that occur.
To determine the precise lineage analysis we used a local codon model implemented in HyPhy (Pond et al., 2005) able to estimate non synonymous and synonymous substitutions per site for each branch. This analysis informs us about the kind of substitutions that occurred during *cystatin* lineages divergence.

In addition we used a naïve approach to detect branches specifically under positive selection in the tree. 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. Ga-branch analysis as most molecular evolution programs is computationally challenging and imposed that the number of sample sequences be reduced to 25. We therefore removed from our sample all nearly identical sequences and pseudogenes. The evolutionary codon model used for this analysis was determined from the AnalyzecodonData implemented in the Hyphy package.

Structure prediction and model building

The available structure of chicken egg white cystatin (pdb code 1cew) and human cystatin D (pdb code 1roa) were used as templates (Alvarez-Fernandez et al., 2005; Bode et al., 1988). The sequence of mature CcBV cystatin1 shares 28% and 24% identity with chicken egg white cystatin [Swissprot: P81061] and human cystatin D [Swissprot: P28325], respectively. Despite the relatively modest level of sequence identity, a reasonable alignment could be made. In particular, the "wedge" region containing the conserved QxVxG motif could be readily aligned. CcBV cystatin 1 corresponds to a type 2 cystatin, which has two conserved disulfide bridges. For the inter-beta-strand disulfide bond, the sequence alignment and template structure place the Cys residues within disulfide bonding distance. The second pair of Cys residues in the initial model were too distant to form a disulfide bond, and had to be brought closer together through energy refinement . The homology modelling was carried out using the program COMPOSER in

SYBYL 7.3 (Tripos Inc, St Louis, MO) on residues 6 to 108 of the mature protein. Three structurally conserved regions (SCRs) were used to build an initial model of CcBV cystatin 1, with three deletions and no insertion relative to the template.

Molecular model refinement and molecular dynamics (MD) simulations

Structural refinement of the complex was done by stepwise energy minimization in Sybyl using the AMBER all atom force field (Cornel et al., 1995) to a gradient of 0.05 kcal/mol/Å. First, only the side chains of the SCRs were energy-minimized, followed by energy minimization of the entire structure. The energy-minimized model was then used as the starting point for molecular dynamics (MD) simulations using the AMBER ff03 force field in the AMBER 9 suite of programs (Case et al., 2005). The protein was solvated in a truncated octahedron TIP3P water box (Jorgensen et al., 1983). The distance between the wall of the box and the closest atom of the solute was 12.0 Å, and the closest distance between the solute and solvent atoms was 0.8 Å. Counterions (Cl^-) were added to maintain electroneutrality of the system. The solvated system was energy-minimized with harmonic restraints of 10 kcal mol^{-1} $Å^{-2}$ on all solute atoms, followed by heating from 100 K to 300 K over 25 ps in the canonical ensemble (NVT). Then, the solvent density was adjusted by running a 25 ps isothermal isobaric ensemble (NPT) simulation under 1 atm pressure. The harmonic restraints were then gradually reduced to zero with four rounds of 25 ps NPT simulations. After an additional 25 ps simulation, a 10 ns production NPT run was carried out with snapshots collected every 1 ps. For all simulations, a 2 fs time-step and 9 Å non-bonded cutoff were used. The Particle Mesh Ewald (PME) method was used to treat long-range electrostatic interactions (Darden et al., 1993), and bond lengths involving bonds to hydrogen atoms were constrained by SHAKE (Ryckaert et al., 1977).

Results

Cystatin genes from polydnaviruses associated with *Cotesia* species exhibit weak genetic divergence

To study the molecular evolution of viral *cystatin* genes, we isolated 48 sequences from polydnaviruses associated with nine *Cotesia* species, revealing that several *cystatin* forms exist in a same species. Accession numbers are provided in Additional File 1. The divergence of the third domain prevented amplification of this region, therefore the *cystatin* sequences isolated contain

only the first two interactive sites. It is extremely unlikely that endogenous wasp cystatins could be amplified by this approach given that polydnavirus cystatins show a low level of relatedness to insect cystatins, and are no more related to insect cystatins than to mammalian inhibitors (Espagne et al., 2005).

Four alleles isolated from *C. glomerata* correspond to a pseudogene with a stop codon situated in the same position for all sequences obtained. Genetic divergence estimated by pairwise distance, which gives the mean number of substitutions per site, ranges from 0.007 to 0.31; these weak values suggest that *cystatin* genes are very similar. Finally, GARD detected no evidence of recombination, allowing us to estimate phylogenies and test for positive selection on *cystatin* genes.

Cystatin **phylogeny shows two main** ***cystatin*** **forms**

Cystatin phylogeny was studied using Bayesian inference and maximum likelihood analysis. Both methods gave the same tree topology. The best tree obtained by maximum likelihood is presented in Figure 8 with bootstrap scores and posterior probabilities. The tree presented was unrooted because there is no suitable outgroup for this study. Phylogenetic analysis revealed the presence of two major cystatin forms supported by high bootstraps and posterior probabilities. The form A cystatins are constituted by CkBV, CmBV, CgBV, CvBV sequences and CfBVD and F, CsBV1, 2 and 3. The form B cystatin are constituted by CcBV, CchBV, CrBV sequences and CfBV 5, 7, 8 and 9, CsBV4 and 6. The form A, in which each clades is supported by high scores, matches the wasp phylogeny (Michel-Salzat and Whitfield, 2004). Indeed, in this case, *cystatin* sequences from a same species group together in the same way as in the wasp phylogeny (Michel-Salzat and Whitfield, 2004). In form B the organisation is different and does not match wasp phylogeny (Michel-Salzat and Whitfield, 2004). Indeed, we do not find a preferential association between sequences from the same wasp species, and the internal branches of this clade are not well supported. The second important difference concerns the branch length: form A *cystatins* exhibits higher overall branch length than form B, suggesting different rates of evolution for these two cystatin forms. This phylogeny strongly suggests the existence of two main ancestral *cystatin* gene forms which have evolved under different constraints to give form A and B. Indeed in form A *cystatins*, long branch lengths are exhibited and follow wasp speciation, as opposed to the form B *cystatins*, which exhibit shorter branch lengths and seem to evolve independently of the wasp speciation process.

Among *cystatin* sequences isolated from a same species some are likely to correspond to allelic forms like *CgBV cystatin* sequences whereas others seem to be different *cystatin* copies such as *CsBV*1, *CsBV*2 *and CsBV*3 (form A). *Cystatin* copies obtained from the CcBV genome sequencing project (*CcBV*cyst1, *CcBV*cyst2 and *CcBV*cyst3) are found in form B and therefore do not seem to have any orthologous sequences in *Cotesia melanoscela*, *Cotesia glomerata* and *Cotesia kariyai* bracoviruses. In form B *cystatins*, these three *cystatin* copies are not grouped together, indicating that duplications occurred before or at the same time as wasp speciation. On the contrary, *cystatin* copies or *cystatin* alleles in form A are grouped by wasp species, suggesting that duplications occurred after wasp speciation.

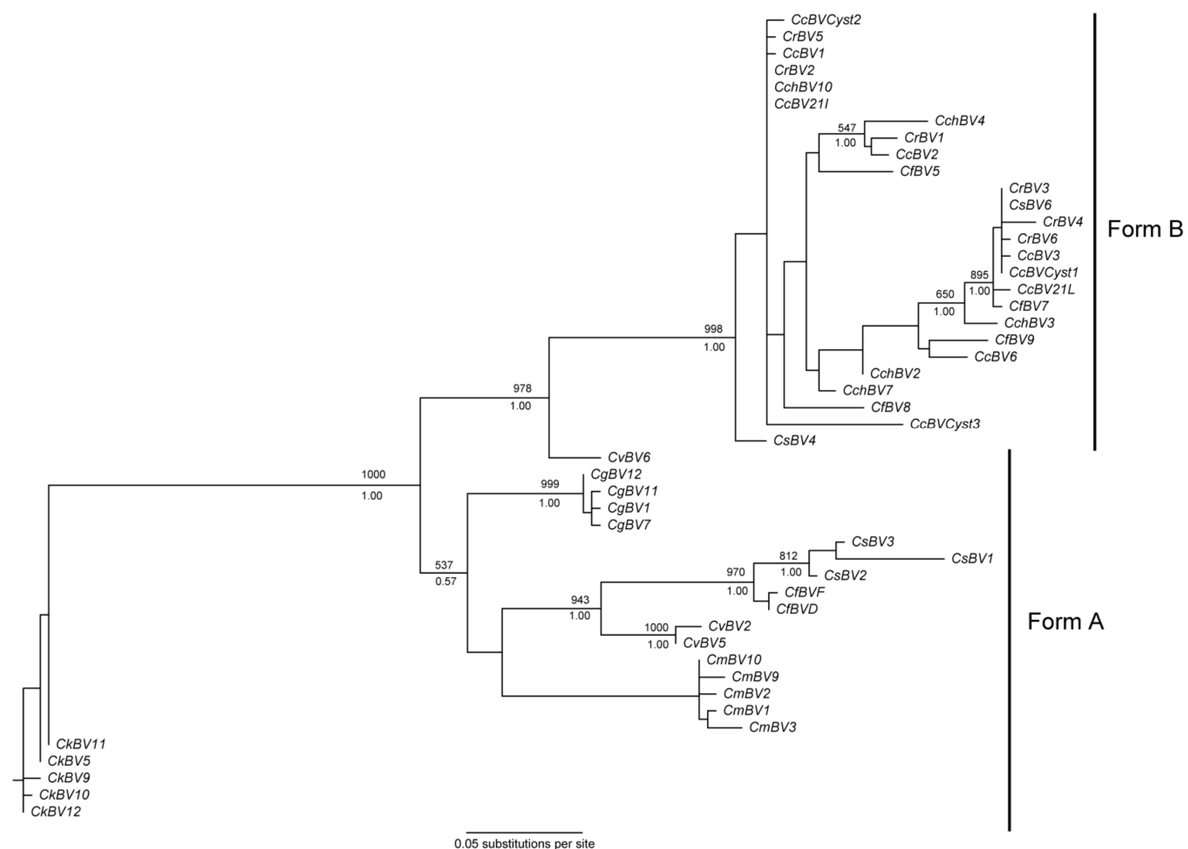


Figure 8 Cystatin gene tree obtained by maximum likelihood
Node supports are shown by bootstraps and by posterior probabilities from Bayesian inferences respectively above and below each branch. Bootstrap scores or posterior probabilities lower than 50% are not represented. Sequences were obtained from bracoviruses of *Cotesia congregata* (CcBV), *Cotesia flavipes* (CfBV)*, Cotesia chilonis* (CchBV)*, Cotesia melanoscela* (CmBV)*, Cotesia vestalis* (CvBV)*, Cotesia rubecula* (CrBV)*, Cotesia sesamiae* (CsBV)*, Cotesia karyai* (CkBV) *and Cotesia glomerata* (CgBV). Cystatin sequences from *CcBV* genome are noted *CcBV*cyst1, *CcBV*cyst2 and *CcBV*cyst3.

Cystatin genes evolve under positive selection

In order to analyse protein evolution and test for positive selection in *cystatins*, ML of different substitution models were determined and compared using chi-squared statistics. Model M0 assumes that all sites have the same ω value whereas M3 distributes amino acids into three classes allowing sites to evolve under different evolutionary constraints. M8A model constrains amino acids to have ω values equal or under 1 whereas the M8 model adds a supplementary class of ω allowing sites to evolve under positive selection. LRTs indicated that selected models M3 and M8 fit the data better than M0 and M8A respectively with *P* values < 0.001 (Table 5). These results suggest firstly that all amino acids are not constrained by the same selective pressures and secondly that cystatin sequences, with an average ω value of 1.2 over all sites and branches, evolve under positive selection. A class-specific site selection analysis was performed to determine the heterogeneity of selection regimes relative to the amino acid position. This analysis indicates that more than 30% of all amino acids are under strong diversifying selection (Table 5).

Table 5 Positive selection analysis among sites and lineages of viral cystatins from *Cotesia* spp. parasitoid wasps

Site analysis						
Models	χ^2 value	df	P value for best model	Gobal ω	ω>1, parameters	PSS
M0 vs M3	125.00	6	< 0.001	1.31	ω=2.65, p= 0.310	26, 5*, 11**
M8A vs M8	19.57	1	< 0.001	1.21	ω=2.85, p= 0.289	26, 7*, 5**
Branch site analysis						
Models	χ^2 value	df	P value for best model		ω>1, parameters (foreground lineage)	
M1a versus MA	126.81	2	< 0.01		ω=11.93, p=0.232	
M3 versus MB	73.20	3	< 0.01		ω=12.04, p=0.233	

Notes: PSS is the number of positive selected sites; * corresponds to a posterior probability>95% of having ω>1 and ** corresponds to a posterior probability >99% of having ω>1.

Modelling by molecular dynamics simulations reveals that the overall folding of known cystatin structures are preserved in CcBV cystatin 1

We wanted to determine whether PDV cystatins adopt a similar 3D structure to chicken cystatin and human cystatins for which the 3D structures have been resolved by crystallography (Alvarez-Fernandez et al., 2005; Bode et al., 1988; Janowski et al., 2001; Stubbs et al., 1990). This constitutes an important prerequisite to be able to interpret the potential consequences of the position of the positively selected sites with respect to the function and the evolution of function of PDV cystatins.

In a previous study, a multiple sequence alignment of CcBV cystatin 1 was performed with insect, chicken, mouse and human cystatins (Espagne et al 2005). Although there is only a modest level of sequence identity among CcBV cystatin 1, human cystatin D and chicken egg white cystatin, a reasonable alignment could be found that permitted a homology model to be built. A 10 ns molecular dynamics simulation was carried out to check the stability of the modelled structure. The energy of the system levelled off after about 800 ps, indicating that an equilibrium state had been reached (data not shown). The overall structure was stable during the simulation. Visual inspection of the trajectory showed that the global fold remained essentially intact. The PROCHECK program (Laskowski et al., 1993) did not flag any conformational problems with the structure. Figure 9A shows a superposition of three average structures during three different time frames in the trajectory. We see that the structures of L1, L2 and L4 are very stable during the simulation. L3 shows somewhat greater structural variability.

The modelled structure preserves the overall fold of solved cystatin structures – a five stranded anti-parallel β-sheet wrapped around a five-turn α-helix (Figure 9). However, □1 maintains its beta strand conformation for only part of the MD simulation. The protease binding site shows a wedge shaped area formed by N-terminal residues (Glycine 6), the first hairpin loop L1 (QxVxG motif positions 50 to 54) and the second hairpin loop L2 (PW). The two conserved type 2 cystatin disulfide bonds are also preserved in this 3-D model of CcBV cystatin 1. Importantly, the 3D model shows that the three conserved domains in CcBV cystatin 1 form the typical tripartite 'wedge' which was shown in the crystal structure of human cystatin B in complex with papain to slot into the protease's active site (Stubbs et al., 1990). These domains therefore display a correct conformation in CcBV cystatin 1, consistent with previous data showing that cystatin 1 is a functional cysteine protease inhibitor (Espagne et al., 2004).

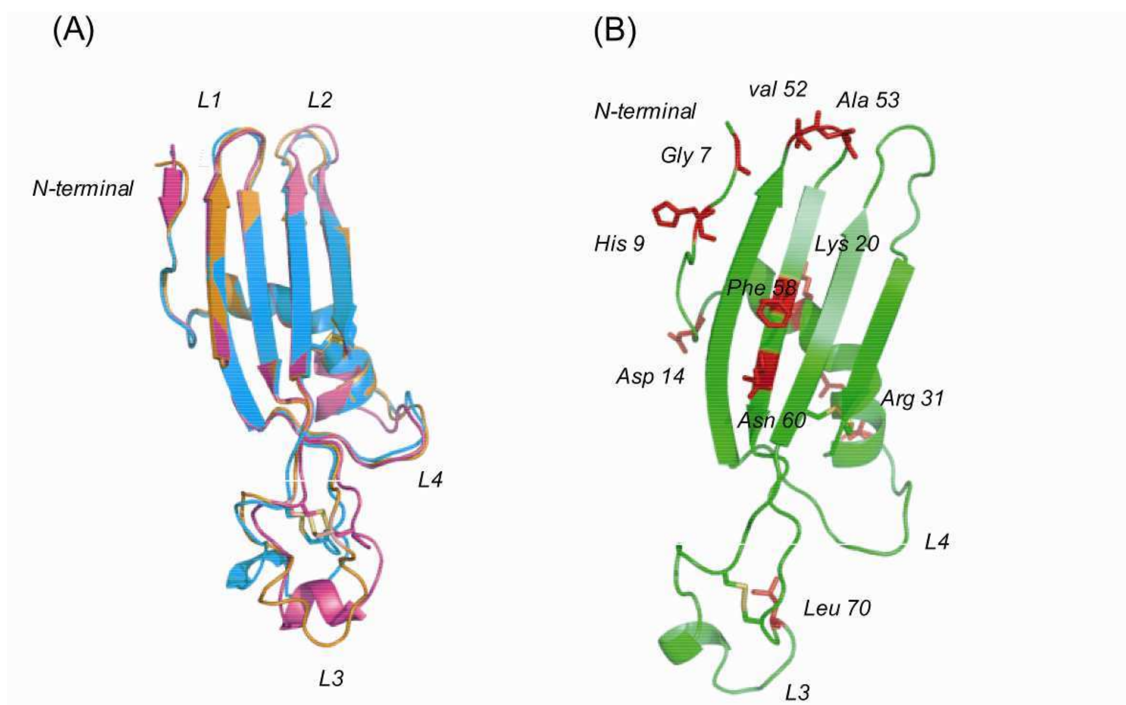


Figure 9 Molecular model of CcBV cystatin 1 obtained from the average structure
(A) The superimposed MD average structure of CcBV cystatin 1 orange (1-5ns), cyan (5-7ns) and purple (8-10ns) of 10ns MD simulation trajectory. (B) Positively selected residues (probability 95%) are represented as a red colour capped stick model on the secondary structure (green) of the final model of CcBV cystatin 1 average structure (1-10ns). Glycine in N-terminal and Valine and Alanine in the L1 are important for C1 protease binding. CcBV mature cystatin 1 amino acid numbering is used.

Most positively selected sites are situated in the vicinity of the cystatin active sites

Sites showing a significant probability (p>95%) of being positively selected in viral cystatins were mapped onto the primary sequence (Figure 10) and on the structural model of CcBV cystatin 1 (Figure 9B). Out of the 12 positively selected sites identified in the mature protein, four are situated in the N-terminal segment containing the conserved Glycine 6 residue (residues Lysine 5, Glycine 7, Histidine 9 and Aspartic acid 14) and two residues are within the first hairpin loop L1 containing the QxVxG motif (residues Valine 52, Alanine 53) (Figure 9B, Figure 10). Lysine 20 and Arginine 31 are located in the α-helix and Phenylalanine 58 and Asparagine 60 at the β3 sheet. Leucine 70 is located at loop 3 between β3 and β4 near the first disulfide bond.

Analysis of the viral cystatin protein alignment among the different wasp species revealed that out of the 12 positively selected sites, 8 (corresponding to Lysine 5, Glycine 7, Histidine 9, Aspartic Acid 14, Lysine 20, Arginine 31, Asparagine 60 and Leucine 70) undergo radical changes in biochemical properties which could induce changes in protein conformation and specificity

(see Additional File 2). For example, Lysine 5, which is a polar and hydrophilic amino acid, can be replaced in other viral cystatin lineages by a leucine which is a hydrophobic residue.

Two amino acids under strong positive selection are also found in the signal peptide. These residues are located in the central, commonly hydrophobic part of the signal peptide, and they do not undergo changes in hydrophobicity. Although selection on signal peptides has rarely been analysed it has already been described in virulence proteins (Liu et al., 2005) and it is thought that variations in the signal peptide could affect exportation of proteins (Byun-McKay and Geeta, 2007; Fujiwara and Asogawa, 2001). In our biological system, viral cystatins are secreted by the host secretory system, therefore we could speculate that the modification of the signal peptide composition could ensure more efficient secretion.

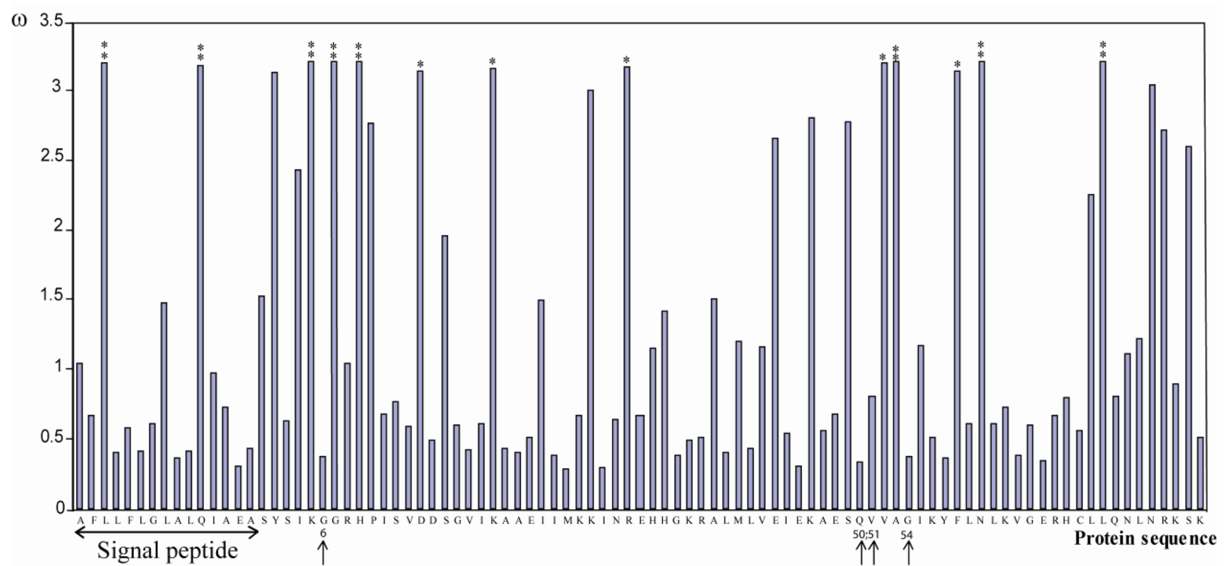


Figure 10 Graphic representation of variable selective pressures (ω) along the protein sequence.

The * indicates the posterior probability >95% of having ω>1 and ** indicates the posterior probability >99% of having ω>1. Conserved amino acids implicated in the interaction with target proteases are indicated by arrows and are numbered according to the mature protein.

Two main *cystatin* lineages show different evolutionary histories

To test for evidence of positive selection among lineages we performed a branch site analysis. In MA and MB models we assigned a $\omega \leq 1$ (ω_0) for form A *cystatins* (background branches) which is congruent with the wasp tree and should evolve under purifying selection and a $\omega > 1$ (ω_1) for form B *cystatins* (foreground branches) which is not congruent with wasp phylogeny and therefore should evolve under positive selection. LRTs indicate that MA and MB fit the data better than models M1a and M3 respectively, with p values < 0.01. Furthermore, these analyses suggest that in foreground lineages about 23% of sites evolve under strong positive selection with ω values around 12 (Table 5). Branch-site analysis results therefore suggest that form A *cystatins* are mainly undergoing purifying selection, whereas form B *cystatins* are mainly evolving under positive selection.

Because this PAML analysis did not allow us to determine the nature of selective pressures acting on each branch, we constructed trees in which branch length represents the expected number of substitutions per codon. The tree in Figure 11A is based on nonsynonymous substitutions, whereas the tree in Figure 11B represents the expected number of synonymous substitutions in *cystatins*. These representations clearly showed a difference in the type of substitutions occurring in the two *cystatin* forms and suggested that divergence between *cystatin* sequences from the form A are particularly due to synonymous substitutions which occur principally in the internal branches, whereas divergence in form B is principally explained by nonsynonymous substitutions. A similar analysis conducted with a nuclear wasp gene (COI) did not reveal differences in synonymous and non synonymous substitutions between wasp species (data not shown) suggesting that the different evolutionary patterns observed above are specific to viral cystatins.

Figure 11 Cystatin sequence tree under local substitution model
(A) Tree scaled on expected number of nonsynonymous substitutions per site and (B) Tree scaled on expected number of synonymous substitutions per site.

To gain further insight into the nature of selective pressures acting on each branch we performed a Ga-branch analysis that confirmed that all lineages are not constrained by the same evolutionary forces. Ga-branch analysis selected a model with two classes of ω. In total, 49 % of branches are assigned to a ω of 0.6 and 51% to a ω class of 5.7 (Figure 12). Both types of branches are present in form A and B, however their position in the tree differs. In form A, positive selection occurs in terminal branches between intraspecies *cystatin* copies. This analysis emphasizes that divergence between *cystatin* copies from the same wasp species occurred by positive selection. Internal branches in form A *cystatins* are characterized by purifying selection, indicating that *cystatin* genes evolved under conservative selection during wasp speciation. A different pattern is observed in form B *cystatins*, where positive selection occurred preferentially in internal branches of the tree. Indeed positive selection occurred in the original branch and in almost all internal branches of this clade, thereby diluting the effect of wasp speciation on *cystatin* divergence. In conclusion, PDV *cystatin* divergence has been driven by positive selection, which has acted at different levels either before, during or after the wasp speciation process.

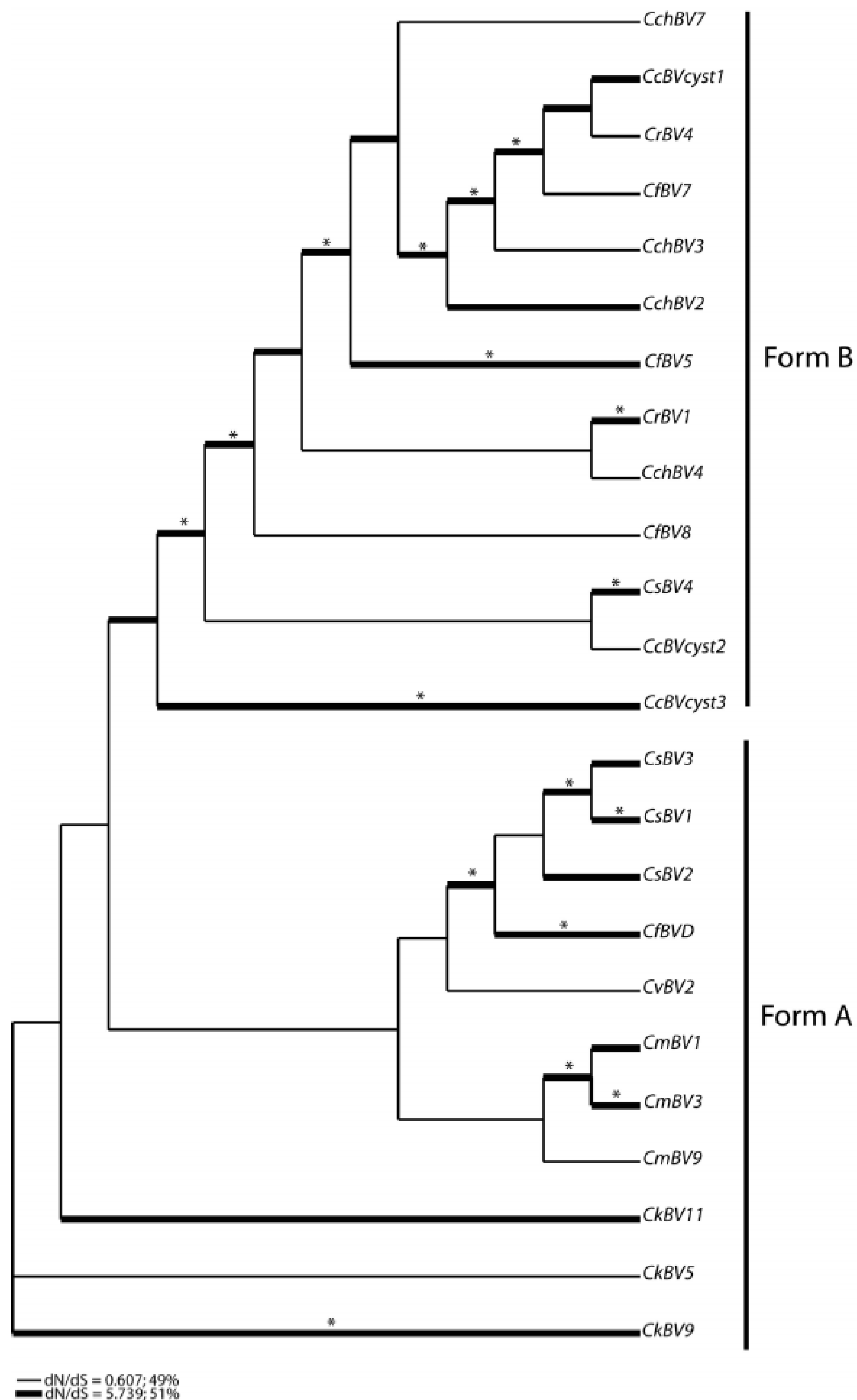


Figure 12 Branches under positive selection estimated according to the Ga-branch analysis.

Percentages for branch classes in the legend reflect the proportion of total tree length (measured in expected substitutions per site per time unit) and evolving under the corresponding value of dN/dS. The * indicates the posterior probabilities >95% of having ω>1.

Discussion

Cystatin genes constitute a young multigene family compared to the other *Cotesia* bracovirus genes

Cystatin genes appear to be unique compared to the other gene families found in the viruses associated with *Cotesia* genus. First *cystatin* divergence, which gives the mean number of substitutions per site, is very weak ranging from 0.007 to 0.31, whereas divergence between CcBV copies of other viral genes like protein tyrosine phosphatases (PTP) or IκB-like proteins range from 0.56 to 0.832 (Bézier et al., 2007). In contrast to PTP or IκB-like proteins, which are both widely distributed in the Bracoviruses carrying PDV (Webb et al., 2006), *cystatin* genes are so far restricted to *Glyptapanteles* and *Cotesia* (Desjardins et al., 2007; Espagne et al., 2004). Furthermore *G. indiensis cystatin* is found in a single copy, whereas three copies are found in *C. congregata* (Desjardins et al., 2007; Espagne et al., 2004), suggesting that the *C. congregata cystatin* gene family resulted from a recent duplication event. The weak divergence between *cystatin* lineages as well as their narrow phylogenetic distribution constitute evidence of the recent acquisition of *cystatin* genes by the bracovirus.

As a consequence, studying *cystatin* gene evolution might allow us to understand the preliminary evolutionary processes involved in the diversification of a young multigene family. The recent events of acquisition and duplication of *cystatin* genes might explain the lack of divergence between *cystatin* copies and our inability to distinguish orthologous and paralogous relationships between copies. For this reason in our analysis, all *cystatin* copies that might include orthologs and paralogs were analysed together.

Are cystatin genes codivergent with wasp species?

PDVs are integrated into wasp chromosomal DNA as a provirus which is inherited exclusively in a Mendelian fashion (Stoltz, 1990). There is no evidence that PDVs can be transferred horizontally between parasitoids and PDVs do not replicate in the host caterpillar. In view of this particular virus life-cycle we can hypothesize that PDV gene evolution is in part determined by evolutionary constraints acting on wasps, such as a phyletic constraints. Nevertheless, viral genes, which are likely to be involved in parasitism success, also have to adapt to caterpillar defences.

A study comparing wasp phylogeny of seven *Cotesia* species based on mitochondrial DNA and viral evolution using the CrV1 gene, has shown a perfect congruence between wasp and viral phylogenies (Whitfield and Asgari, 2003). In our study the *cystatin* gene tree also shows perfect codivergence between wasp and viral genes for some *cystatin* gene lineages. The evolution of these *cystatin* forms appears therefore to be constrained by wasp phylogeny and the molecular constraints acting on the wasp genome. However, in contrast to the results obtained using CrV1, not all *cystatin* lineages follow wasp evolution; instead, some *cystatin* genes are submitted to other constraints since their phylogeny does not match wasp phylogeny.

Cystatins are under strong selective pressure acting on key sites

The study of selective pressures acting on *cystatin* genes confirms that *cystatin* genes are not simply constrained by wasp evolutionary history. Indeed, we showed that *cystatin* gene evolution is driven by a strong positive selection. The global ω value of 1.2 obtained through analysis of viral *cystatins* is similar to the value obtained with plant cystatins (Kiggundu et al., 2006). Plant cystatins are involved in a plant-phytophagous interaction, but in that case cystatins play a role in defence against digestive cysteine proteases of herbivorous insects. Plant cystatins and their targets are thought to be involved in a coevolutionary process. Other examples of positive selection are also available with pathogen molecules. A previous study performed on an Ichnovirus protein involved in host immune inhibition has shown that positive selection was only detected at particular protein sites (Dupas et al., 2003b). Our study constitutes the first example of a major impact of positive selection in the evolution of a bracovirus protein.

The identification of the position of positively selected sites in PDV cystatins in the primary sequence and in the 3D-model revealed that 70 % of sites are situated within or proximal to the N-terminal segment harbouring the conserved Glycine and the first hairpin loop containing the QxVxG motif. These two domains, together with the C-terminal PW sequence, make up the « wedge » in the cystatin1 model, shown by crystallography in cystatin B and chicken cystatin to interact directly with the active-site cleft of target C1 proteases (Bode et al., 1988; Stubbs et al., 1990). These results suggest that diversifying selection could be acting on viral cystatins to modify the inhibitor's sites of interaction with host target proteases, which could translate into an increased or reduced affinity towards these enzymes. Interestingly, modifications in inhibitor affinity have been reported in engineered cystatin proteins carrying deletions or mutations in the N-terminal segment or the first hairpin loop (Abrahamson et al., 2003; Hall et al., 1995; Kiggundu et al., 2006; Machleidt et al., 1989). In chicken cystatin, the removal of the

residues preceding the conserved Glycine leads to a 5000 fold decrease in affinity towards papain (Machleidt et al., 1989). Furthermore, a site-directed mutagenesis approach used to pin-point which residues contribute the most to target enzyme affinity in human cystatin C revealed that the -1 residue (with respect to Glycine) is responsible for the major part of this affinity (Hall et al., 1995). In PDV cystatins it is noteworthy to stress that the equivalent site (corresponding to Lysine 5 preceding the conserved Glycine 6 in CcBV cystatin 1) is under positive selection. This suggests that PDV cystatins have evolved under diversifying selection possibly to produce inhibitors of varying affinity for caterpillar proteases, just as cystatin C laboratory engineered mutants have been developed that have discriminating affinities for mammalian cysteine proteases (Mason et al., 1998). We can predict that the other sites under positive selection in the N-terminal region of viral cystatins are also likely to influence the interaction with proteases. Indeed, comparison of positions of positively selected sites of PDV cystatins and plant cystatins revealed that 2 of these sites are in equivalent positions with respect to the conserved Glycine residue in both sets of inhibitors (positions -1, +3). Furthermore, in plant cystatins, independent mutations in these sites lead to variations in inhibitory activity towards papain and cathepsin B (Kiggundu et al., 2006) (Goulet et al., 2008).

Two positively selected sites have also been identified in the first hairpin loop of PDV cystatins including the central valine of the QxVxG motif. These sites, corresponding to Valine 52 and Alanine 53 in cystatin 1, are inside this region with one affecting the central Valine. However this central site is not absolutely conserved in all cystatins. In the chicken egg white cystatin the hairpin loop motif is QLVSG and an increase in binding affinity to cysteine proteinases was obtained when this motif was mutagenized to QVVAG (Auerswald et al., 1995) indicating that variation in central residues of this loop affects binding with target proteases.

In summary, the majority of positively selected sites identified in PDV cystatins are located in the vicinity of the two inhibitory sites analysed in this study. Furthermore, these sites affected by positive selection have been shown experimentally in other cystatins to be important for affinity with target proteases. Taken together these results suggest that positive selection is acting presumably to modulate viral cystatin affinity for caterpillar protease targets.

It will now be interesting to determine what could be the role of the positively selected sites which are more distant from the cystatin inhibitor sites (Lysine 20, Arginine 31, Phenylalanine 58, Asparagine 60 and Leucine 70 in cystatin1). Phenylalanine 58 and Asparagine 60 may still be influencing the L1 loop at position 50-54. Leucine 70 is located near the disulphide bond and variations in this position may affect the structure of the protein. These sites could also be unmasking new sites of interactions with proteases, indeed in chicken cystatin it

was suggested that others regions or sites of the protein could be important for the strong interaction with the cysteine protease cathepsin L (Auerswald et al., 1995).

Scenario for cystatin gene evolution

The strong selective pressure observed emphasizes the important role of cystatins in the host-parasitoid interaction. These results suggest that cystatins have to continuously evolve in order to adapt to their target in the host caterpillar. Given the potential pathogenic role of viral cystatins and also the probable involvement of cysteine proteases in insect immunity (Saito et al., 1992), these results can be interpreted by integrating cystatins in a coevolutionary context. Nevertheless, this diversifying evolutionary pattern could also be explained by wasp host switches and the subsequent necessity for cystatins to evolve rapidly to respond to new biochemical targets.

Our analysis reveals the existence of two viral cystatin forms which display different evolutionary patterns in regard to wasp evolution. In more classical non-obligate mutualist associations, horizontal gene transfer can explain incongruences between host and symbiont phylogenies. However, in this case, virus and wasp have a long and stable relationship since more than 100 MYA (Murphy et al., 2008) and artificial infection of wasps by PDV is not possible. Therefore we propose and our analyses strongly suggest that adaptive constraints have contributed to the different evolutionary patterns observed in the two *cystatin* forms.

Moreover, for both of these two forms duplication events occurred independently in the different *Cotesia* bracoviruses studied and are fixed by positive selection which is also responsible of the ensuing divergence of *cystatin* copies.

Interestingly, Francino (2005) (Francino, 2005) proposed in the "radiation adaptive model" that duplications are fixed to their selective advantage and that gene copies evolve under natural selection before new functions appear. This mode of evolution, particularly for functional genes, could be a response to specific environmental pressures such as new biochemical niches. Therefore, the particular evolution of the *cystatin* gene family could be a response to particular cystatin targets in a specific host-parasitoid system.

Conclusions

Unravelling the molecular evolution of proteins can lead to a better understanding of their function. For the first time in a host-parasite interaction system, we show that viral *cystatins*

are subject to strong positive selection. 3D Modelling of a viral cystatin revealed that most of the positively selected residues are in the vicinity of the inhibitory active sites, suggesting that adaptive selection acted to improve the inhibitory activity of viral cystatins. Furthermore two different *cystatin* forms have been identified, each of them evolving under different selective constraints probably imposed by different host cysteine proteases.

In order to better explain *cystatin* gene family evolution, we have now to consider the host range of each wasp species studied. For this purpose, studying the Melitaeini-*Cotesia* system appears clearly adapted since their ecology in terms of host range is well characterized (Kankare and Shaw, 2004). Such a study would precise the potential coevolutionary processes involved between viral *cystatins* and host cysteine proteases.

Authors' contributions

CS participated in the data acquisition and performed the sequence alignment, the phylogenetic and selection analysis and wrote the main draft of the manuscript. SC performed the 3-D structure modelling and participated in drafting the manuscript. SP was involved in the main acquisition of data and participated in the critical revision of the manuscript. SD participated in selection analysis and in intellectual revision of the manuscript. JL participated in experimental discussions and sequence acquisitions. EP was involved in the cystatin 3-D structure modelisation and in intellectual discussions. JMD was involved in obtaining funding, in project conception and in critical revision of the manuscript. EH was responsible of the project conception and of the experiments and was actively involved in drafting the manuscript. All authors have given their final approval of the version to be published.

Acknowledgments

We are extremely grateful to Cindy Ménoret and Carole Labrousse for insect rearing. We thank Franck Dedeine for critical reading of the manuscript and Gilles Lalmanach for fruitful discussions. We thank the three anonymous reviewers for corrections and helping to improve the manuscript. We sincerely thank Louise E.M Vet, Robert N. Wiedenmann, Claire Villemant, Thomas Guilloux, Tomoyuki Tanaka and Hans M. Smid for providing specimens for this study. CS was supported by a PhD grant from the French ministry de l'Enseignement Supérieur. Financial support was provided by the ANR grant EVPARASITOID.

V- Cysteine proteases: potential targets of parasitism

Cystatins are C1 cysteine protease inhibitors, to understand the role of cystatins in the course of parasitism we studied potential cystatin targets in *M. sexta*.

In part A, we present the isolation of cysteine proteases in *M. sexta* and their regulation pattern after immune challenge.

In part B, we present our attempt to isolate host proteins specifically targeted by cystatins during parasitism.

A-Identification of parasite-responsive cysteine proteases in *Manduca sexta*

Serbielle.C, Moreau. S, Veillard.F, Voldoire.E, Drezen, JM, Lalmanach.G, Huguet.E

Abstract

In the course of parasitism, the immune system is one of the principal barriers involved in host resistance. However, a multitude of parasite virulence strategies have evolved which could target unsuspected host physiological functions. In *Manduca sexta* parasitized by the wasp *Cotesia congregata,* the main alterations induced by parasitism are host developmental arrest and disruption of the immune system. In this host-parasitoid interaction, successful parasitism is ensured by a symbiotic virus (CcBV), associated with the wasp, and injected at the same time as wasp eggs into the lepidopteran host. The CcBV genome encodes different virulence factors which are predicted to play an important role in parasitism success. One of the most highly expressed CcBV gene families, encodes cystatins which are tight binding inhibitors of C1A cysteine proteases. Cystatins are suspected to play an important role in the interaction not only because of the high level of their expression but also because of the high diversifying selection acting at the level of the protease-interacting active sites in these genes. *In vitro* studies showed that CcBV cystatin1 is a functional inhibitor active against both mammalian and dipteran C1A cysteine proteases. So far, the *in vivo* targets in *M. sexta* were unknown. In the present work, we identified for the first time in *M. sexta*, four C1 cysteine proteases; a cathepsin L and B and two different "26-29 kDa" cysteine proteases. Among them, our analyses revealed that "26-29 kDa" proteases are transcriptionally down regulated in the course of parasitism. In addition, we showed that viral Cystatin1 and the 26/29 kDa host protease MsCath1 colocalized in the plasma following parasitism, suggesting they have the potential to interact. We also show that parasitism induces a general increase of cysteine protease activity which is later controlled.

Our study characterizes for the first time C1A cysteine proteases in *M. sexta.* The potential involvement of particular cysteine proteases in defence against parasitoids is discussed with respect to the known physiological roles of cysteine proteases both in mammals and in insects.

Introduction

Insects depend entirely on innate immune responses for protection against pathogens. In the course of infection, pathogens and parasites of insects have to counteract both humoral and cellular defences. Humoral factors such as antimicrobial peptides serve as defences against a wide range of microbes (Lemaitre and Hoffmann, 2007). Circulating blood cells are involved in phagocytosis of microbes and melanization and encapsulation of parasitoid eggs (Strand and Pech, 1995b). However, in host-pathogen/parasite interactions, the immune system is not the only physiological process targeted and parasites can control alternative physiological functions to ensure their infection. For example, the Drosophila C virus requires clathrin mediated endocytosis to complete their life cycle (Cherry and Perrimon, 2004). Parasitoid wasps affect development of their host, thereby ensuring the completion of their own life-cycle (Beckage and Gelman, 2004). Understanding host-parasite interactions and infectious processes not only requires knowledge on the host immune system but also more widely on host physiological functions.

In the present study we focus on a particular biological system where a lepidopteran host is parasitized by a wasp associated with a mutualistic virus. In this association, polydnaviruses (PDV) persist as stably integrated proviruses in the genomes of their associated wasps (Desjardins *et al.*, 2007) and replicate in female ovaries only. Virus particles are injected into the lepidopteran host during wasp oviposition at the same time as wasp eggs and play an essential role in wasp parasitism success. PDVs do not replicate in the parasitized host insect, but viral gene products suppress the host immune system and cause physiological alterations ensuring parasitoid development (Asgari et al., 1996; Beckage and Gelman, 2004; Tanaka et al., 2000; Turnbull and Webb, 2002). PDV have been proved to be involved in the deregulation of several host physiological functions such as development, immunity, feeding and pigmentation (Beckage *et al.*, 1994). However, mechanisms by which the PDVs target hosts physiological functions are not yet totally resolved.

Three families of PDV have been described, ichnoviruses (IV) associated with Ichneumonidae wasps, bracoviruses (BV) associated with Braconidae and recently PDVs associated with the Banchinae wasps have been shown to potentially constitute a third virus lineage (Lapointe et al., 2007; Turnbull and Webb, 2002). In the last few years several PDV genomes have been sequenced which offers strong support to understand how PDVs are involved in wasp parasitism success. PDVs have large genomes containing numerous genes

possibly involved in host physiological disruption. Many of these genes encode proteins with domains showing similarity to cellular proteins suggesting they were probably picked up by the virus from the wasp genome during their evolution. PDVs from different families and even different sub-families do not carry exactly the same set of genes. Some genes are common to both ichnoviruses and bracoviruses such as ankyrins and cysteine rich proteins, whereas others, like PTP are present in bracoviruses and viruses associated with Banchinae wasps but are absent from ichnoviruses (Espagne et al., 2004; Gundersen-Rindal and Pedroni, 2006; Lapointe et al., 2007; Webb et al., 2006). Certain genes are only present in specific viral subfamilies, for example, viral cystatins are only found in PDVs associated with *Cotesia spp*. and *Glyptapanteles indiensis*, a closely related species (Desjardins *et al*., 2007; Serbielle *et al*., 2008). These differing patterns of gene acquisition events and their conservation throughout virus evolution are likely to reflect the role of these genes in regard to the multitude of hostile environments represented by different lepidopteran hosts.

The recent acquisition of viral cystatins by bracoviruses associated with *Cotesia spp*. appears as a particular interesting case to understand the role of viral gene acquisition in parasitism success. Cystatins are inhibitors of C1A cysteine proteases with which they form a tight-binding reversible inhibitory complex (Abrahamson *et al*., 2003; Turk *et al*., 1997) They are found in plants and animals and bracovirus cystatins constitute the first description of viral cystatins. In *Cotesia congregata* Bracovirus (CcBV), 3 genes encode cystatins which are probably the result of a recent duplication event. Previous studies showed that CcBV cystatins are highly expressed during parasitism and that one recombinant CcBV cystatin was proved to be active against C1A cysteine proteases *in vitro* (Espagne *et al*., 2005). Moreover cystatin active sites were shown to be directionally and strongly selected (Serbielle *et al*., 2008). All together these data strongly suggest an active and important role of cystatins during parasitism. In other organisms, cystatins are involved in host-pathogen/parasite interactions and play either the role of defence molecules as in plants or the role of virulence factors as in filarial nematodes (Arai et al., 2002; Gregory and Maizels, 2008; Maizels et al., 2001; Schierack et al., 2003).

What could be the role of cystatins during parasitism:? Whether they are involved in host immune system disruption or they are targeting non immune host physiological functions to ensure wasp development remains to be determined.

A prerequisite to understand the role of cystatins in parasitism is to investigate the presence of their targets in the host and to evaluate the parasite-responsive nature of their expression.

Natural cystatin targets are C1 A cysteine proteases corresponding to papain-like cathepsins in mammals and in plants. Cathepsins are synthesized as inactive proenzymes with N-terminal pro-peptide regions which are removed upon activation. The catalytic triad of cathepsins is formed by conserved cysteine, histidine and asparagine residues (Cys25, His159, and Asn175 in papain numbering) where Cys and His are believed to exist as a thiolate-imidazolium ion pair which is stabilized by Asn via a hydrogen bond with His (Lecaille *et al.*, 2002). Four main types of C1A cysteine proteases have been described in insects which correspond to Cathepsins B, L, F-like and 26/29 kDa proteins (Attardo *et al.*, 2006; Kurata *et al.*, 1992). 26/29 kDa type proteases have so far only been described in insects. In insects, cathepsins are implicated in different physiological processes such as digestion, development and immunity. Herbivorous insects utilize C1A cysteine proteases to break down dietary proteins but also to function as defence proteins against toxins or protease inhibitors potentially produced by plants (Koo et al., 2008; Shindo and Van Der Hoorn, 2008). Insect cathepsins also play an important role in development during embryogenesis (Cho *et al.*, 1999; Uchida *et al.*, 2001), moulting (Liu *et al.*, 2006) and metamorphosis (Hegedus *et al.*, 2002; Homma *et al.*, 1994). Most studies have focused on the role of insect cysteine proteases in digestion or development but there is also less well highlighted data which suggests a potential role of cathepsins in insect defences. Indeed in the flesh fly *Sarcophaga peregrina*, a 26/29 kDa protease was shown to be specifically released by hemocytes after an immune challenge (Saito *et al.*, 1992). Similarly, in *Drosophila*, cathepsin L protein is induced in the hemolymph of adults after fungal infection (Levy *et al.*, 2004). Moreover, Cathepsins B, F and L were recovered from an EST library constructed from immune-stimulated Glossina fat body (Attardo *et al.*, 2006). In *Drosophila,* bacterial infection induces an up-regulation of cathepsin L (De Gregorio *et al.*, 2001).

Our approach to understand the role of polydnavirus cystatins in the host-parasitoid interaction involving *M. sexta* and *C. congregata* was 1) to characterize host cathepsins potentially targeted by these inhibitors and 2) to investigate for the first time in *M. sexta*, the effects of parasitism by *C. congregate* , expression and activity of these proteases.

Materials and methods

Insect rearing

The tobacco hornworm *M. sexta* (Lepidoptera, Sphingidae) and the parasitoid wasp *C. congregata* (Hymenoptera, Braconidae, Microgastrinae) were reared as previously described (Harwood and Beckage 1994, Provost 2004). *M. sexta* life cycle comprises five larval instars followed by pupal and adult stages. Parasitism interrupts the larval-pupal molt.

cDNA Cysteine protease isolation

Four complete C1A cysteine protease sequences were obtained by RT-PCR (MsCath1) and RACE-PCR (MsCath1, MsCath2, MsCathB and MsCathL) from a pool of RNA extracted from fat body and hemocytes of non parasitized L4 larvae. For fat body isolation, larvae were dissected in a physiological saline solution. Hemocytes were collected by bleeding larvae from a cut proleg and pelleting the cells by gentle centrifugation (900g for 5 min). Total RNA was extracted form these tissues according to the manufacturer's instructions (Total RNA isolation, Macherey-Nagel France). Reverse transcription of 1µg RNA was carried out using the Omniscript RT kit as described by the manufacturer (Qiagen, France). For isolation of MsCath1, a primer (MsCath1For1) based on a *Spodoptera frugiperda* EST sequence encoding a C1A cysteine protease (Sf1H05182-5-1) was used in combination with an oligo dT primer to amplify the 3'end of the equivalent sequence from *M. sexta* cDNA. The Takara polymerase was used in a reaction composed of 2 minutes denaturation at 95°C, followed by 30 cycles at 45 seconds 95°C, 45 seconds at 45°C and 1 minute at 72°C, followed by a final extension for 10 minutes at 72°C. 5' RACE-PCR was then performed using primers MsCath1RACE5-1 and MsCath1RACE5-2 as described by the manufacturer (SMART-RACE cDNA Amplification Kit, Clontech, France). For isolation of MsCath2, MsCathB and MsCathL specific primers were designed based on *M. sexta* Expressed Sequence Tags (EST) (MSP00278_1, MSP01261_1 and MSP00859_1 respectively) obtained from Butterflybase (Papanicolaou et al 2008) encoding C1A cysteine proteases (Table 6). These primers were used in RACE-PCR amplification to amplify the 5' or the 3' part of cysteine proteases. All resulting fragments were cloned in the pDrive-cloning vector (Qiagen PCR Cloning Kit). Multiple cloned inserts were sequenced in both directions using the Big Dye^{R} Terminator v3.1 cycle sequencing kit and the sequenced products were analysed using a capillary DNA sequencer (ABI PRISM 3100).

Table 6 Cysteine proteases primers for RT-PCR and RACE-PCR

Primer names		Nucleotide Sequences
MsCath1	For1	5'- GCT GGC CAT CAA CCA GTT TGC TGA-3'
	RACE5-1	5' - GTC AGT GTG AGG GAA GGG - 3'
	RACE5-1	5'- CAA TAA CCC TTT GTA TCT CTT CAT CTC C -3'
	RACE5-2	5'- CTT CAT CGC ATC CAT AAT CAT -3'
MsCath2	RACE5-1	5' - GTC AGT GTG AGG GAA GGG - 3'
	RACE5-2	5' - CGT CAG TCC AGT CTG CAT ATT TG - 3'
	RACE5-3	5' – GTG ATT TGA ACG GGT ACG TAG TAG TC - 3'
	RACE3-1	5' - ACT GGA CAA CAG CGA TAG CTT - 3'
	RACE3-2	5' – GGA GTT CCA ACA GAT ATG G - 3'
	RACE3-3	5'- AAG AAC TCC TGG GGT GAG GAC TGG- 3'
MsCathB	RACE5-1	5' - TGA CAC AAT ACC GAA GTG TT - 3'
	RACE5-2	5' – CTT GTA GCG CTC CCA TGA GCT T - 3'
	RACE3-1	5' - GCC GAT ACA CCT CAT CCG CTC - 3'
	RACE3-2	5' – GCGAGGACCACATCAGAG- 3'
MsCathL	RACE5-1	5' - GGA CTT GCG GAA GTG CTG - 3'
	RACE5-2	5' – TTC GGG CCG AGC TTG TAC GA-3'
	RACE3-1	5' - GTCGACTGGACCAAGAAG- 3'

Cysteine protease and control gene transcriptional analysis

Cysteine protease expression was studied at different stages of development of non-parasitized larvae and after immune challenge. Samples were obtained and analysed in triplicate. In order to follow cysteine protease expression during larval development, fat body was dissected at six different stages: 24h before the L4 stage (TB), at the L4 stage (L4), 24h after the L4 stage (L4-24h), 48h after the L4 stage (L4-48H) and at the stage of dorsal pigmentation (VD). To follow cysteine protease regulation after immune challenge, L4 larvae were subjected to either no treatment or injection of 10 µl of PBS (controls), or to parasitism (2 ovipositions were checked visually) or injection of 10 µl *Escherichia coli* (10^6 cfu/ml) in PBS. Fat body was extracted 2, 18 or 24 hours after treatment and total RNA isolation was performed as described above. First-strand cDNA was synthesized using 1µg of total RNA using the Omniscript RT kit as described by the manufacturer (Qiagen). Expression of cysteine protease genes and control genes was analyzed by semi-quantitative PCR using specific primers (see Table 7). Amplification of a house-keeping gene EF1α (Elongation Factor α) was used to adjust initial amounts of cDNA of each sample using Goldstar polymerase (Eurogentec) with the following PCR cycling conditions: initial step at 94°C 3 minutes, 20 cycles with 45 seconds denaturation at 94°C, 45 seconds hybridization at appropriate temperature, 45 seconds extension at 72°C, followed by 10 minutes of final extension at 72°C. cDNA dilutions were then used to analyze cysteine protease, cystatin 1 and attacine 1

expression using the same PCR conditions as above with either 20, 25 or 30 cycles depending on the level of expression of each gene. PCR reactions were performed on 1µl of adjusted quantities of cDNA in a 25µl final volume with 2.5µl of 10X buffer, 2µl of dNTP (2.5mM), 1.5µl of MgCl2 (25mM), 1µl of forward and reverse primer (20pmol/µl each) and 0.1µl of the Goldstar Taq Polymerase (Eurogentec). No amplification was detected from samples which had not undergone reverse transcription indicating they were not contaminated by genomic DNA. Specific primers and annealing temperatures are shown in table 7. Amplification products were cloned and sequenced to ensure the specificity of PCR reactions (data not shown).

Table 7 PCR primers used in semi-quantitative PCR

Primers	Nucleotide sequences	Sequence sizes	Tm (°C)
EF1α-For1 EF1α-Rev1	5'- TTTCAGGACACCAGTCTCAACTCT - 3' 5'- CTTGTGGAGCGCAAGGAGG - 3'	177 bp	62
Cystatine1-For1 Cystatine1-Rev1	5'- AATTTGTTTCCAGTTTTCTCAACCAT - 3' 5'- TGTAGCCGGAATCAAGTATTTTCTT - 3'	161 bp	60
MsCath2qPCR-F1 MsCath2qPCR-R1	5'-ACCAAAGTTGTCAAAGTTAAAAC-3' 5'-GATTCCAACCGCTCCTTATC-3'	174 bp	60
Attacine-For1 Attacine-Rev1	5'- CGATGTGGTGTTGGTACCGATG - 3' 5'- TGGCGCTAAAGACGTTCTTGTC - 3'	141 bp	58.5
Mscath1-For1 Mscath1-Rev1	5'- AGATAATATTGACCACGTTAAGAGGAAAA - 3' 5'- CAATAACCCTTTGTATCTCTTCATCTCC - 3'	156 bp	58.5
Cathepsine B-For1 Cathepsine B-Rev1	5'- GGTAGCTGTGGCTGCCGATA - 3' 5'- CTTGTAGCGCTCCCATGAGCTT - 3'	158 bp	60
Cathepsine L-For1 Cathepsine L-Rev1	5'- GGTCGCGTTCAAGATGCAA - 3' 5'- TTCGGGCCGAGCTTGTACGA - 3'	150 bp	60

Western blot analysis

Western blot analysis was conducted on hemocytes and plasma isolated from L4 larvae 24 hours after being subjected to different treatments: no treatment, parasitization (2 ovipositions), injection of 1µl of sephadex beads (36g/L), injection of bacteria (as above). Non parasitized and parasitized larvae were also analyzed 2 and 48 hours after treatment.

Hemolymph and hemocytes were extracted by bleeding larva as described above and were diluted in an equal volume of PBS solution containing 0.02 % of phenylthiourea (PTU) to prevent hemolymph melanization. After gentle centrifugation (5min, 900g), plasma and hemocytes were separated and conserved at –20°C. For each sample, protein concentration was calculated using the Bradford reagent following the manufacturer's instructions (Sigma). 5 µg of proteins in Laemmli buffer (4X) were deposited on a 12% polyacrylamide gel in denaturing conditions. After migration, the gel was incubated in transfer buffer (Tris-Base 48mM, Glycine 39mM, SDS 0.0375%, Methanol 20%) and blotted one hour at 38 mA on a PVDF membrane

(Bio-Rad). Immunodetection was conducted with two different antibodies. Two polyclonal antibodies were raised against specific peptides of the *Cotesia congregata* Bracovirus viral cystatin 1 protein (YSIKGGRHPISVDDS starting at residue 24 of the preprotein) and the catalytic domain of Mscath1 (PTEAEYGPYLNRDNYC starting at residue 413 of the pre-pro enzyme), respectively (Eurogentec). The membrane was blocked overnight with a TNT solution (Tris-Base 48mM, NaCl 140mM, Tween 20 0.05%) with 5% skimmed milk, incubated with rabbit antiserum to cystatin/MsCath1 (1:1000 dilution) for 2 hours and then incubated for 1 hour with anti-rabbit IgG conjugated with alkaline phosphatase (1:10000 dilution, Sigma). Immunostained bands were detected with the BCIP/NBT liquid substrate system (Sigma).

Cysteine protease activity assays

Plasma and hemocytes were extracted as described above from nine to ten parasitized larvae 24 or 48 hours after oviposition and from nine to ten synchronous non parasitized caterpillars. 150 µl of extracts were conserved at –80°C in a same volume of PBS with 0.02% of PTU until activity assays. Protein concentrations were measured using the Bradford reagent (Sigma). To ensure larvae had been parasitized successfully a Western blot analysis using anti-cystatin antibody was performed on an aliquot of each sample (data not shown).

5 µl of hemocyte lysates were incubated in 175µl of activation buffer (100 mM sodium phosphate buffer, pH 6.0, containing 1 mM EDTA and 4mM DDT) for 10 min at 37°C, prior addition of 20µl of Z-Phe-Arg-AMC (final concentration : 50 µM), a broad-spectrum substrate of cysteine cathepsins. Alternatively Z-Arg-Arg-AMC (50 µM), a specific substrate of mammalian cathepsin B-related enzyme, and Z-Leu-Arg-AMC (50 µM) were also used as AMC-derived fluorigenic substrates to measure proteolytic activities. Fluorecence release was continuously monitored with a Gemini spectrofluorimeter (Molecular Devices, Saint-Grégoire, France) with λexc =350 nm and λem =460 nm, using 96-well Dynex microtitre plates (Elvetec, Genas, France), at 37°C under gentle agitation. The same procedure was repeated and carried out with hemolymph samples (20 µl/assay).

Control experiments were performed by pre-incubating hemocyte and hemolymph samples with E-64 (10µM) prior enzymatic assays as described above to ensure the fluorescence release was specifically due to cysteine protease activities.

Specific material:

DTT (DL-dithiothreitol) came from Bachem (Weil am Rhein, Germany). E-64 (L-3-carboxy-trans-2.3-epoxypropionyl-leucylamido-(4-guanidino) butane) was from Sigma-Aldrich

(Saint-Quentin Fallavier, France. Z-Phe-Arg-AMC and Z-Arg-Arg-AMC were purchased from Bachem, while Z-Leu-Arg-AMC was supplied by ICN Biochemical (Aurora, USA). All other reagents were of analytical grade.

Results

Four C1A cysteine proteases were isolated from *M. sexta*

Four different C1A cysteine proteases were isolated from *M. sexta* by RT-PCR and RACE-PCR corresponding to cathepsin B-like (MsCathB) and cathepsin L-like (MsCathL) proteases and to two 26/29 kDa-like proteases named MsCath1 and MsCath2. All proteases contain a signal peptide (preregion) and a pro-domain and the catalytic domain. Moreover MsCath1 and MsCath2 carry a supplementary “26 kDa”domain situated before the pro-domain (Figure 1). All mature proteases contain the catalytic triad formed by the Cys, His and Asn residues (Figures 2 and 3).

MsCathB ORF is 1013 bp long and encodes a 337 residue protein of 37 kDa. By analysing conserved domains the mature protease corresponds to the 254 residues between positions 84 and 330 (28 kDa). The twelve cysteine residues known to participate in the formation of six disulphide bridges in other cathepsins are all conserved in MsCathB (Yan et al., 2002) (Figure 13A). MsCath B also harbours residues corresponding to the occluding loop (residues 190-199 in Figure 13A), conserved in all cathepsins B. The ability of cathepsin B to function as a carboxypeptidase has been attributed to the presence of the two Histidine residues present within the occluding loop (Cygler *et al.*, 1996; Krupa *et al.*, 2002; Musil *et al.*, 1991). MsCathB shows a high level of sequence conservation with other insect and mammalian Cathepsin B proteins (Figure 14A). For example, overall amino acid identity/similarity with *Bombyx mori* (NP_001036850) and Human Cathepsin B (AAH95408) are 80 % identity /87 % similarity, and 54% identity / 66% similarity, respectively. If the mature cathepsin B is compared, these scores reach 86 % identity / 92 % similarity with silkmoth and 63% identity/ 73 % similarity with the human protein.

A) MsCathB

```
  1 M  T  L  S  R  A  A  C  V  A  L  V  C  A  V  A  V  A  A  D  T  P  H  P  L  S  D  A  F  I
  1 ATGACTCTGTCGCGTGCAGCTTGCGTCGCGCTCGTGTGCGCGGTAGCTGTGGCTGCCGATACACCTCATCCGCTCTCCGACGCGTTTATA

 31 R  L  I  N  S  K  Q  N  T  W  R  A  G  R  N  F  P  T  T  T  P  F  A  H  I  N  K  L  M  G
 91 AGACTCATCAACAGCAAGCAGAACACATGGAGGGCCGGCCGCAACTTCCCAACGACGACGCCGTTCGCTCACATCAACAAGCTCATGGA

                                                          ▼
 61 A  L  Q  D  D  N  V  A  K  M  P  K  V  E  H  D  A  D  L  I  A  S  L  P  E  N  F  D  P  R
181 GCGCTACAAGATGACAACGTAGCGAAAATGCCGAAAGTTGAACACGACGCCGACCTAATTGCAAGTCTACCCGAAAATTTCGATCCTCGC
                   1                             2☆
 91 D  K  W  P  D  C  P  T  L  N  E  I  R  D  Q  G  S  C  G  S  C  W  A  F  G  A  V  E  A  M
271 GACAAATGGCCCGATTGTCCCACTTTAAATGAAATTAGGGATCAGGGGTCTTGCGGTAGTTGCTGGGCGTTCGGCGCAGTTGAAGCGATG
                1                                                        3  4        4
121 T  D  R  Y  C  T  Y  S  N  G  T  K  H  F  H  F  S  S  E  D  L  L  S  C  C  P  I  C  G  L
361 ACTGACCGATACTGTACATACTCGAATGGAACTAAACATTTTCATTTCTCATCAGAGGATCTTTTGAGTTGCTGTCCGATCTGCGGGCTC
       2
151 G  C  N  G  G  I  P  S  L  A  W  E  Y  W  K  H  F  G  I  V  S  G  G  N  Y  N  S  T  Q  G
451 GGGTGCAACGGTGGTATTCCGTCGCTGGCCTGGGAATATTGGAAACACTTCGGTATTGTGTCAGGTGGGAATTACAACTCAACCCAGGGC
    5                       6                                            6                       3
181 C  R  P  Y  E  I  P  P  C  E  H  H  V  P  G  N  R  M  P  C  S  G  D  T  K  T  P  K  C  Q
541 TGCAGACCGTACGAAATTCCTCCTTGTGAGCACCACGTCCCTGGTAACAGAATGCCGTGCAGTGGTGACACCAAAACACCTAAATGCCAG
          5
211 K  N  C  E  N  G  Y  N  V  M  Y  K  K  D  K  R  Y  G  K  H  V  Y  S  V  S  A  G  E  D  H
631 AAAAACTGCGAAAATGGATACAACGTTATGTACAAAAAAGACAAGCGGTACGGTAAACATGTGTACAGTGTGTCTGCGGGCGAGGACCAC

241 I  R  A  E  L  Y  K  N  G  P  V  E  G  A  F  T  V  Y  A  D  L  L  A  Y  K  S  G  V  Y  K
721 ATCAGAGCCGAGCTGTATAAGAACGGGCCGGTAGAAGGCGCGTTTACTGTTTACGCAGATTTGCTCGCATACAAGAGTGGTGTTTACAAG

                             ☆                                                               ☆
271 H  I  Q  G  D  A  L  G  G  H  A  I  K  I  L  G  W  G  V  E  N  D  N  K  Y  W  L  V  A  N
811 CATATACAAGGCGACGCTCTCGGCGGCCACGCAATTAAGATTTTAGGTTGGGGCGTGGAAAACGACAACAAATATTGGCTTGTTGCTAAC

301 S  W  N  T  D  W  G  D  N  G  F  F  K  I  L  R  G  E  N  H  C  G  I  E  G  S  I  I  A  G
901 TCATGGAATACAGATTGGGGAGACAATGGTTTCTTCAAAATCCTGCGTGGAGAGAATCATTGCGGGATTGAGGGCTCGATCATCGCCGGT

331 E  P  L  L  V  D  D  *
991 GAGCCTTTGTTAGTGGATGATTA
```

B)MsCathL

```
  1 M  K  C  L  V  F  L  L  C  A  V  A  A  S  A  S  A  V  S  F  F  D  L  V  K  E  E  W  V  A
  1 ATGAAGTGTTTGGTGTTCCTGCTGTGTGCGGTGGCGGCCAGCGCCAGCGCCGTCTCGTTCTTCGACCTCGTCAAGGAGGAATGGGTCGCG

 31 F  K  M  Q  H  D  K  K  Y  D  S  E  V  E  D  R  F  R  M  K  I  Y  A  E  N  K  H  K  I  A
 91 TTCAAGATGCAACACGACAAGAAGTACGATAGCGAGGTGGAGGACCGGTTCCGCATGAAGATCTACGCGGAGAACAAGCACAAGATCGCG

 61 K  H  N  Q  L  Y  E  Q  G  L  V  S  Y  K  L  G  P  N  K  Y  T  D  M  L  H  H  E  F  I  Q
181 AAGCACAACCAGCTGTACGAGCAGGGACTCGTCTCGTACAAGCTCGGCCCGAACAAGTACACCGACATGCTGCACCACGAGTTCATACAG

 91 A  M  N  G  Y  N  R  T  A  K  H  N  K  G  L  Y  G  K  K  H  D  V  R  G  A  T  F  I  P  P
271 GCCATGAACGGGTACAACAGGACCGCGAAACACAACAAGGGTCTTTACGGCAAGAAGCACGACGTCCGCGGCGCCACGTTCATCCCGCCG

             ▼                                                                               ☆
121 A  H  V  K  Y  P  D  H  V  D  W  T  K  K  G  A  V  T  E  V  K  D  Q  G  K  C  G  S  C  W
361 GCGCACGTGAAGTACCCCGACCACGTCGACTGGACCAAGAAGGGCGCCGTCACCGAGGTCAAGGACCAGGGCAAGTGCGGCTCCTGCTGG

151 A  F  S  T  T  G  A  L  E  G  Q  H  F  R  K  S  G  Y  L  V  S  L  S  E  Q  N  L  I  D  C
451 GCCTTCAGCACCACGGGCGCTCTGGAAGGCCAGCACTTCCGCAAGTCCGGATACCTGGTGTCGCTGTCGGAGCAGAACCTGATCGACTGC

181 S  S  T  Y  G  N  N  G  C  N  G  G  L  M  D  N  A  F  K  Y  I  K  D  N  G  G  I  D  T  E
541 TCGTCGACGTACGGCAACAACGGCTGCAACGGCGGCCTCATGGACAACGCCTTCAAGTACATCAAGGACAACGGCGGCATCGACACGGAG

211 K  T  Y  P  Y  E  G  V  D  D  K  C  R  Y  N  P  K  N  S  G  A  E  D  V  G  F  V  D  I  P
631 AAGACCTACCCGTACGAGGGTGTCGACGACAAGTGCAGGTACAACCCGAAGAACTCCGGCGCCGAAGACGTGGGCTTCGTGGACATCCCC

241 S  G  D  E  E  K  L  M  Q  A  V  A  T  V  G  P  V  S  V  A  I  D  A  S  Q  N  S  F  Q  F
721 TCAGGTGACGAGGAGAAGCTGATGCAGGCGGTGGCCACCGTGGGCCCCGTGTCCGTCGCCATCGACGCCTCGCAGAACAGCTTCCAGTTC

                                                  ☆
271 Y  S  G  G  V  Y  Y  D  T  E  C  S  S  T  D  L  D  H  G  V  L  V  V  G  Y  G  T  D  E  A
811 TACTCGGGCGGCGTCTACTACGACACGGAGTGCTCCTCCACTGACCTCGACCACGGCGTGCTGGTGGTGGGGTACGGCACGGACGAGGCC

                          ☆
301 G  G  D  Y  W  L  V  K  N  S  W  S  R  T  W  G  E  L  G  Y  I  K  M  A  R  N  R  D  N  H
901 GGCGGTGACTACTGGCTCGTGAAGAACTCGTGGAGCCGCACGTGGGGCGAGCTCGGATACATCAAGATGGCGCGCAACCGCGACAACCAC

331 C  G  I  A  T  D  A  S  Y  P  L  V  *
991 TGCGGCATCGCCACCGACGCCTCCTACCCGCTCGTCTAA
```

C) MsCath1

```
   1 M  L  P  V  W  Y  I  L  S  F  L  I  A  A  I  S  A  Q  D  I  W  N  D  D  S  E  L  S  W  P
   1 ATGTTACCGGTTTGGTACATACTGTCGTTTTTGATCGCCGCGATTTCCGCCCAAGATATATGGAATGACGATTCCGAACTGTCATGGCCC
  31 E  K  Y  T  V  E  A  I  R  I  S  L  T  A  G  L  V  E  N  H  K  A  W  R  T  N  V  S  S  R
  91 GAAAAATACACTGTGGAGGCAATTCGGATCTCTCTAACTGCAGGGCTTGTTGAAAACCACAAAGCCTGGAGAACAAATGTGTCGTCGCGA
  61 I  D  Y  N  N  G  V  V  K  S  I  T  K  E  N  S  P  N  F  N  K  Y  E  I  H  P  E  F  T  K
 181 ATTGACTACAACAATGGTGTTGTAAAGTCCATCACTAAAGAAAACTCTCCGAATTTTAATAAATATGAGATCCATCCTGAGTTCACGAAA
  91 N  S  S  N  V  V  K  C  N  A  A  E  L  D  P  S  W  Q  S  V  S  L  E  Q  I  L  P  D  T  N
 271 AACTCATCTAACGTTGTGAAGTGTAACGCAGCGGAACTTGACCCCAGTTGGCAGTCGGTATCACTAGAACAAATTCTGCCTGATACAAAT
 121 D  F  E  F  V  D  L  D  P  V  N  D  K  D  T  W  K  F  F  K  K  E  S  L  E  G  S  D  I  R
 361 GACTTCGAGTTTGTCGACCTGGATCCTGTCAATGACAAAGATACTTGGAAATTTTTCAAAAAGGAATCACTAGAAGGAAGTGACATTAGA
 151 R  T  V  W  A  E  F  D  N  E  T  E  V  W  I  P  V  R  Y  E  V  Y  E  Y  N  E  W  L  G  F
 451 AGAACCGTGTGGGCAGAGTTTGACAATGAAACAGAAGTTTGGATTCCAGTCAGATATGAAGTGTATGAATACAACGAGTGGTTGGGTTTC
 181 L  S  K  H  D  I  W  D  F  F  E  F  S  I  D  F  D  E  D  V  F  E  T  D  D  Y  G  C  D  E
 541 TTGAGCAAACACGATATTTGGGATTTCTTTGAATTCTCCATTGATTTCGACGAAGATGTTTTTGAAACCGATGATTATGGATGCGATGAA
 211 E  Y  T  M  H▼ E  M  D  S  E  T  V  T  K  H  L  M  F  I  D  P  E  N  D  K  H  V  D  H  V
 631 GAATACACAATGCATGAGATGGACAGTGAGACGGTGACCAAACACTTAATGTTCATCGATCCTGAAAACGATAAACACGTCGATCATGTT
 241 F  N  A  F  K  S  K  H  N  K  V  Y  K  D  N  I  D  H  V  K  R  K  N  I  F  Q  N  K  M  R
 721 TTCAACGCGTTCAAAAGCAAACACAATAAAGTATACAAAGATAATATTGACCACGTTAAGAGGAAAAACATATTCCAAAATAAAATGAGG
 271 L  I  R  A  T  N  R  K  N  L  G  Y  K  L  A  V  N  Q  F  A  D  R  T  P  E  E  M  K  R  Y
 811 TTGATAAGAGCCACAAACCGAAAGAATCTTGGCTACAAACTTGCTGTAAATCAGTTTGCTGATCGCACTCCTGAGGAGATGAAGAGATAC
 301 K  G  L  L  R  R  P  E  G  K  T  G  N  I  A  F  P  Y  K  D  S  D  I  K  N  I  E  T▼ D  L
 901 AAAGGGTTATTGAGACGTCCTGAAGGAAAAACTGGCAATATTGCCTTTCCGTACAAGGATTCTGACATTAAAAATATAGAAACTGATCTA
                                                                       ☆
 331 P  D  E  Y  D  A  R  L  L  G  L  V  S  A  V  K  N  Q  D  N  C  G  S  C  W  T  F  G  T  T
 991 CCTGATGAGTACGACGCAAGGCTATTGGGATTAGTATCCGCGGTTAAAAATCAGGATAACTGTGGTTCCTGCTGGACTTTCGGCACCACC
 361 A  A  V  E  G  A  L  A  Q  H  N  G  G  K  L  L  S  L  S  N  Q  A  L  I  D  C  A  W  P  F
1081 GCTGCCGTGGAGGGAGCTCTCGCTCAACATAACGGAGGGAAACTTCTAAGTCTGAGCAACCAGGCACTAATTGACTGTGCTTGGCCTTTT
 391 G  V  R  G  C  D  G  G  S  D  N  A  A  Y  E  W  M  M  E  Y  G  L  P  T  E  A  E  Y  G  P
1171 GGTGTACGCGGCTGTGACGGAGGCAGTGACAACGCAGCGTATGAGTGGATGATGGAATACGGCTTACCCACTGAAGCGGAGTACGGACCC
 421 Y  L  N  R  D  N  Y  C  H  I  D  N  V  T  V  K  Y  P  I  K  G  F  T  D  V  T  P  Y  S  V
1261 TACCTTAACAGGGATAACTACTGCCACATCGATAATGTGACAGTGAAATACCCAATCAAAGGCTTCACTGATGTTACACCATACAGTGTC
 451 G  A  L  K  V  A  L  V  N  H  G  P  L  S  V  S  V  H  A  S  E  A  F  T  L  Y  S  G  G  I
1351 GGGGCGCTGAAGGTTGCGCTGGTGAACCACGGCCCGCTGTCAGTGTCTGTCCACGCAAGTGAGGCGTTCACACTGTATTCCGGAGGTATA
                                       ☆
 481 F  Y  D  T  E  C  D  P  T  S  L  D  H  E  V  A  L  V  G  Y  G  T  R  D  E  D  T  F  W  I
1441 TTCTACGACACAGAATGCGATCCTACAAGTTTAGACCATGAAGTGGCACTGGTGGGATATGGAACGCGAGACGAAGATACATTCTGGATC
           ☆
 511 L  K  N  S  W  G  P  M  W  G  I  D  G  Y  I  L  I  S  S  R  D  N  N  C  G  V  A  T  E  P
1531 CTGAAGAACTCCTGGGGACCCATGTGGGGAATCGATGGATACATCCTAATATCTTCGAGGGACAACAACTGCGGTGTTGCCACTGAACCT
 541 T  Y  V  N  F  *
1621 ACTTACGTGAATTTTTAA
```

Figure 13 Nucleotide and amino acid sequences of *M. sexta* Cathepsins B (MsCathB), L (MsCathL) and the 26/ 29 kDa C1A protease MsCath1.

Dark arrows indicate cleavage sites of signal peptides, the 26kDa region (in the case of MsCath1), pro-regions and the « 29 kDa » protease regions. Stars indicate the conserved catalytic sites. Cysteine residues high-lighted in red correspond to the twelve residues known to participate in the formation of six disulphide bridges in MsCathB. Residues high-lighted in purple indicate the occluding loop of MsCathB. Amino acids high-lighted in blue indicate the characteristic cathepsin L-like motifs (E-R-F-N-I-N) and (R-N-Y-D) that are present in the pro-regions of MsCathL and MsCath1. Amino acids high-lighted in green correspond to the peptide used to generate the anti-MsCath1 antibody.

MsCathL gene sequence (1029 bp) encodes a 342 aa protein (38 kDa) and the mature enzyme has an expected molecular weight of 24 kDa. Two characteristic cathepsin L-like motifs (E-R-F-N-I-N) and (R-N-Y-D) are present in the pro-region of MsCathL (E44-R48-Y52-N55-I59-N63) and (G66-N68-Y70-D82) (Karrer *et al.*, 1993; Liu *et al.*, 2006) (Figure 13B). MsCathL displays excellent sequence conservation with other cathepsin L proteins, the mature protein displays 85% identity / 92 % similarity with *Bombyx mori* fibrionase (NP_001037464.2) and 62 % identity / 75 % similarity with human Cathepsin L1 (NP_666023.1) (Figure 13B).

MsCath1 (1638 bp) and MsCath2 (1626 bp) encode proteins of 545 aa and 541 aa, respectively, and mature proteases of 24 kDa based on analysis of conserved domains with the 26/29 kDa protease from *S. peregrina*. MsCath1 and MsCath2 mature proteases show 55 % identity / 70 % similarity and 46% identity / 65% similarity with mature 26/29 kDa *S. peregrina* protease. MsCath1 displays less similarity with MsCath2 (50 % identity / 68 % similarity) compared to the *S. peregrina* protease. Globally, the mature 26/29 kDa proteases display higher sequence conservation with mature Cathepsin L proteases than with Cathepsin B. For example mature MsCath1 shares 41% identity /56% similarity with MsCathL and only 19% identity /35% similarity with MsCathB. Furthermore if the MsCath1 and MsCath2 pro-proteins are compared to MsCathL pro-protein, the two characteristic L-like motifs (see above) can be detected in MsCath1 and 2 (see for MsCath1 in Figure 13C). No significant sequence similarity was found between the 26 kDa subunit and other proteins, only slight similarity (48 %) was found with a portion of a protein of *Pichia stipitis* which harbours domains suggesting this protein could be involved in actin organization and endocytosis. However these domains do not overlap with the 26 kDa homologous region. Analysis of proteins showing sequence similarity to 26/29 kDa proteins revealed that these proteases have now also been isolated from the starlet sea anemone *Nematostella vectensis*, the placazoan *Trichoplax adhaerens* and the vertebrate zebrafish *Danio rerio,* indicating these proteases are no longer restricted to insects. Multiple sequence alignment of 26/29 kDa proteases from various insects and these new sequences enabled us to define residues and domains in the 26 kDa region which are highly conserved Y 33, (S 59, R 60, I 61, D 62, Y 63), G 66, P 86, N 94, C 98, (L 116, P 117), and P 166 (MsCath1 numbering) (Figure 13C).

Figure 14 *M. sexta* C1A cysteine protease alignements

A) MsCath B, (B) MsCathL and (C) MsCath1 and MsCath2. Arrows indicate cleavage sites and stars indicate conserved sites in the protease domains.

Cysteine protease expression at different stages of larval development

Using semi-quantitative RT-PCR the temporal patterns of *M. sexta* cathepsin mRNA were analysed from fat body of larvae at different stages between L3 and L5. All samples were checked for the absence of genomic DNA contamination. PCR amplifications of the house-keeping gene EF1-α at 20 cycles enabled calibration of cDNA samples (Figure 15). After calibration the same conditions were kept for cysteine protease amplifications (see Materials and Methods). MsCathL, MsCathB and MsCath2 showed a constant level of expression for all larval stages studied. MsCath 1 expression was overall constant except at the L4 stage where a weaker level of expression was observed.

These results showed that the four proteases are constitutively expressed at the larval stages studied with a weak decrease of MsCath1 expression at the fourth instar.

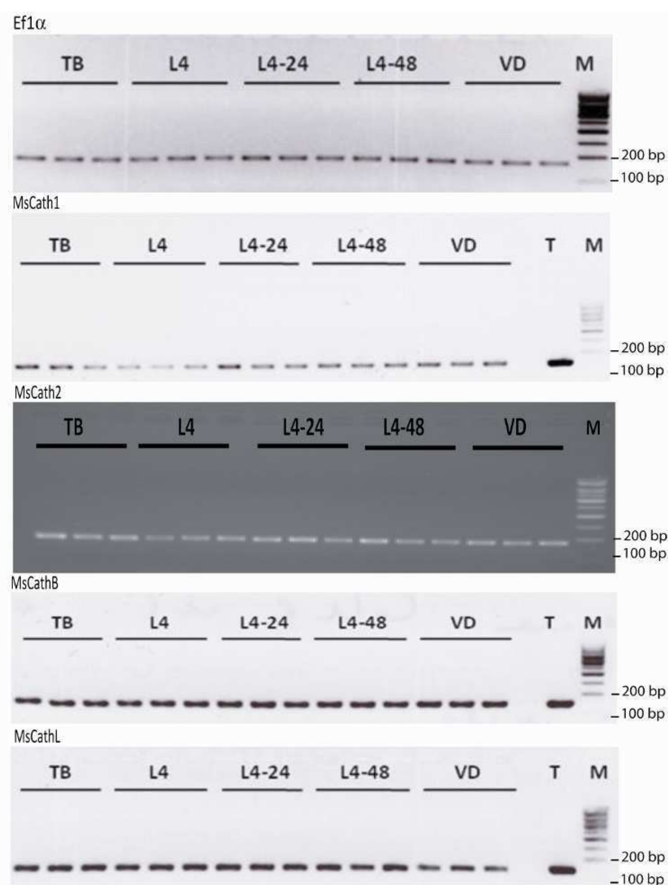


Figure 15 Cysteine protease transcriptional regulation in the course of *M. sexta* development by semi-quantitative RT-PCR.
EF1α was used to calibrate cDNA concentrations. MsCath1, B and L regulation was studied at the following larval stages: L4: 4th instar larvae; TB: 24h before the 4th instar; L4-24: 24h after the 4th instar; L4-48: 48h after the 4th instar; VD: dorsal pigmentation stage.

Cysteine protease regulation after immune challenge

Semi-quantitative RT-PCR analysis was used to determine the expression of *M. sexta* C1A cysteine proteases from fat body of L4 larvae at 2, 18 and 24 hours after immune challenge. cDNA samples were calibrated as described above (Figure 16). Immune challenges corresponded to parasitized larvae or larvae injected with *E.coli* in PBS, corresponding controls were non parasitized larvae and larvae injected with PBS alone, respectively. To ensure parasitism had been successful, larvae were checked for the expression of the viral cystatin1 gene in parasitized caterpillars (Figure 16). Successful stimulation of the immune system via *E. coli* was verified by monitoring the expression of the antimicrobial peptide gene attacin 1 (Figure 16). 2 hours after treatment, all cysteine proteases showed a weak and constant expression pattern whatever the stress applied (Figure 16A). In contrast, after 18 h profiles differed: MsCathB and L showed a constant level of expression whereas MsCath1 and 2 were down regulated after bacterial injection or parasitism (Figure 16B). In parasitized larvae MsCath1 expression was no longer detectable and expression was barely visible for MsCath2. 24 hours after treatment, MsCathB and L expression remained constant and MsCath1 and 2 expressions were inhibited both in parasitized and bacteria injected caterpillars (data not shown). MsCath1 and 2 are therefore specifically down-regulated after an immune challenge. This down regulation cannot be explained by a general affect on host gene transcription or by the destruction of the tissue during infection since MsCath B and L expression patterns are unaffected. qRT PCR experiments performed on hemocytes of non parasitized and parasitized larvae at 2, 24 and 48 hours also showed down regulation of MsCath1 24 and 48 hours after parasitism (data not shown).

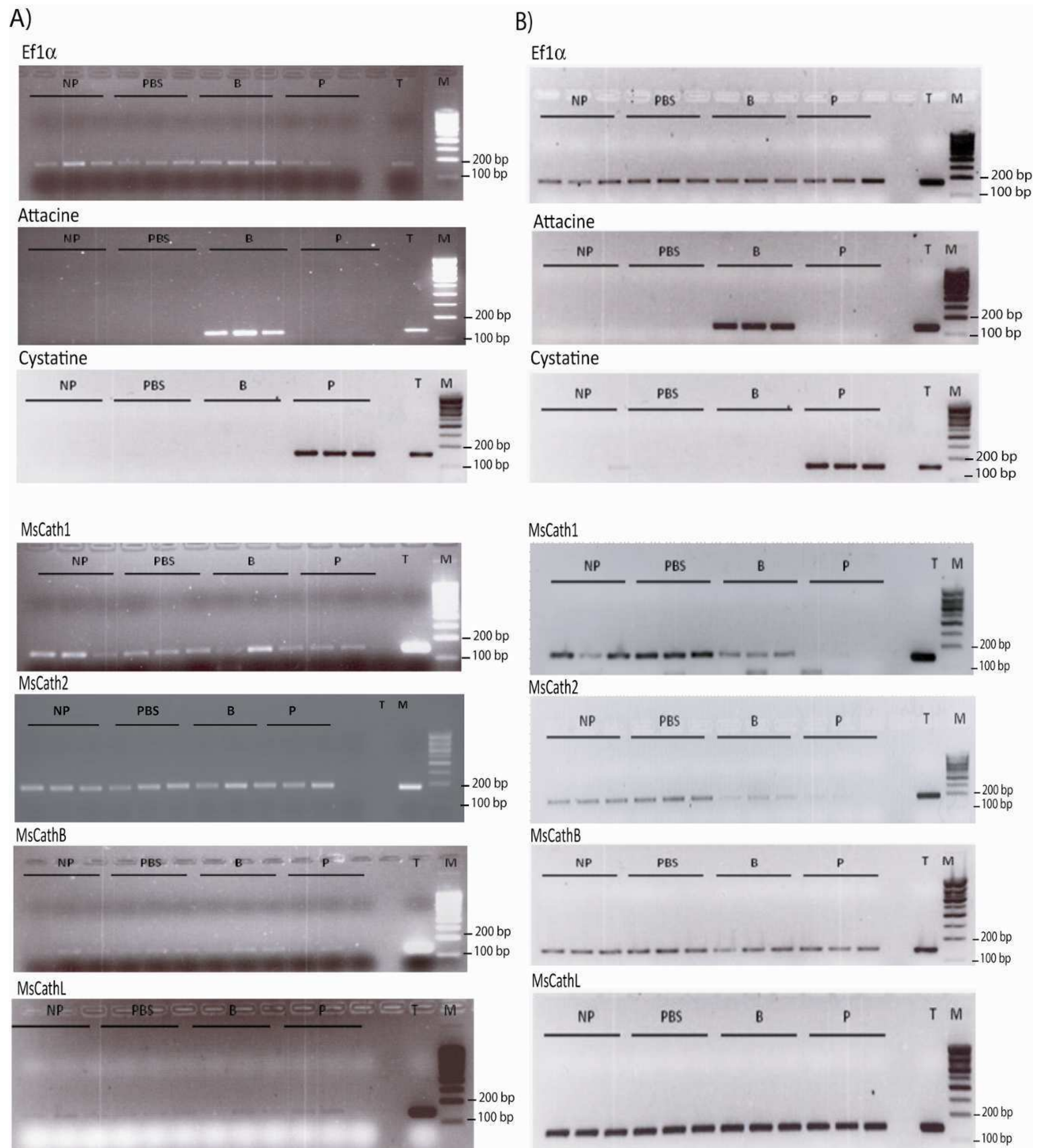


Figure 16 Cysteine protease transcriptional regulation after an immune challenge by semi-quantitative RT-PCR.

EF1α was used to calibrate cDNA concentrations, attacine expression enabled to control for successful bacterial infection and successful parasitism was controlled using the CcBV cystatin 1 gene.MsCath1, 2, B and L expression was monitored 2 hours (A) and 18 hours (B) after no treatment (NP), parasitism (P), injection of PBS (PBS), or injection of bacteria (B).

Cystatins and MsCath1 co-localize in the hemolymph after parasitisation

To determine whether viral cystatins and *M. sexta* cathepsins have the potential to interact we studied the localization of these proteins in hemocytes and plasma. Two polyclonal antibodies were raised against selected epitopes of cystatin or MsCath1 peptide sequences. Hemocytes and plasma were collected from larvae at 2, 24 and 48 hours after parasitism (P) and from synchronous non parasitized larvae (NP). Western analysis using the anti-cystatin1 antibody revealed that cystatins could be detected in plasma of parasitized larvae at 24 and 48 hours. Occasionally a faint signal could also be detected in hemocytes (Figure 17A). MsCath1 antibody detected a protein of approximate molecular weight 29 kDa in hemocytes of both NP and P larvae at 2 and 24 hours, but only in NP larvae at 48 hours (Figure 17B). This protein was also detected in the hemolymph of 24 hour parasitized larvae only (Figure 17C). To determine if other types of stress could induce the release of this protein in plasma, we analysed samples from larvae which had been subjected to PBS injection, bacterial infection and injection of beads. Proteins were extracted from larvae 24 hours after treatment and analysed by western. MsCath1 protein could not be detected in the plasma of these samples (Figure 17C). MsCath1 protein appears therefore to be specifically released after parasitism into hemolymph after 24 hours. At this time point cystatin proteins are also expressed and present in the plasma, suggesting that the proteins have the potential to interact. At 48 hours after parasitism MsCath1 protein is no longer detectable probably because of the downregulation observed earlier on at the RNA level (Figure 17B).

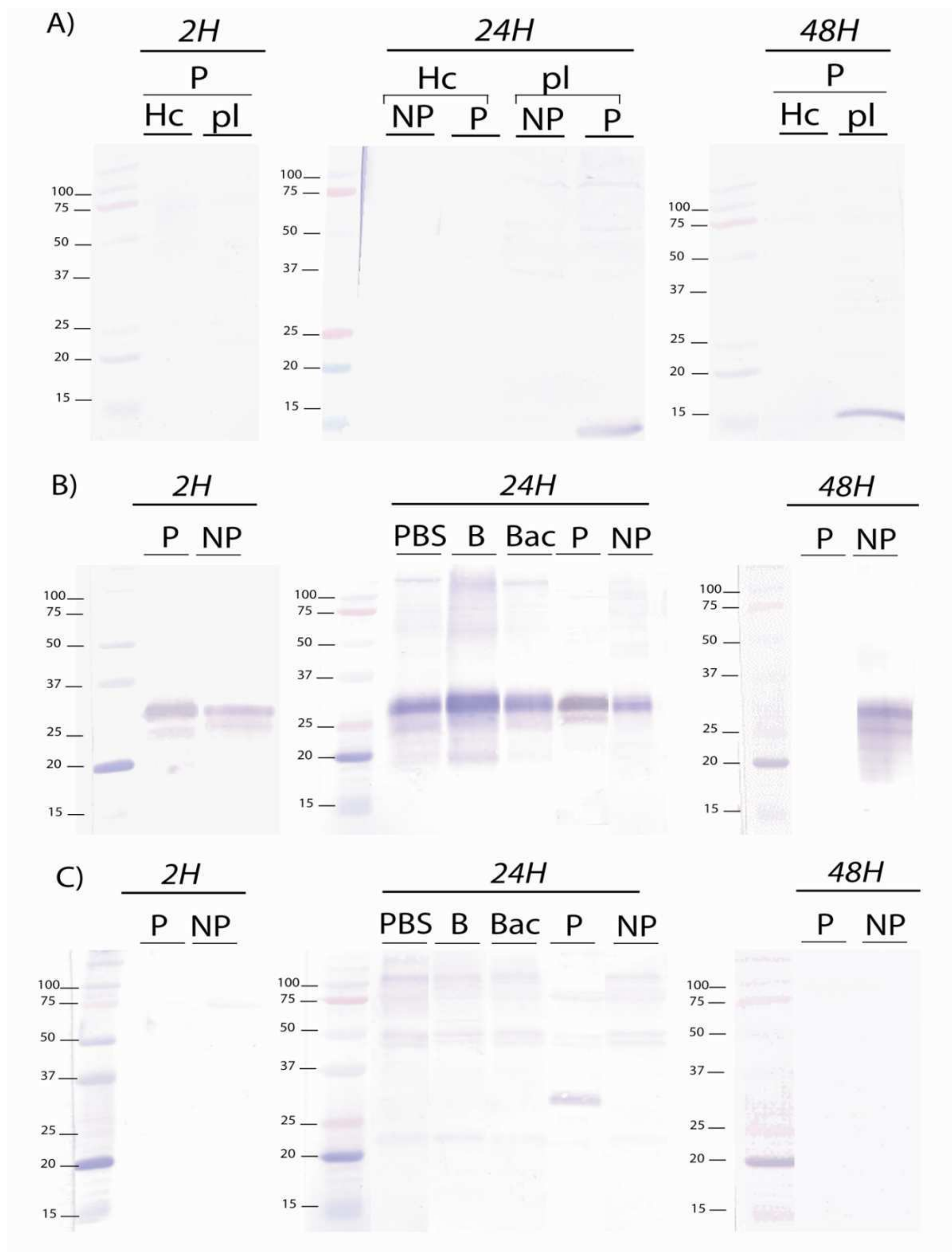


Figure 17 Western blot analysis of proteins extracted from *M. sexta* hemocytes and hemolymph (plasma).

Western blot analyses were performed at 2h, 24h and 48h post-treatment and from synchronous non-treated caterpillars using (A) cystatin antibody against hemocytes (Hc) and plasma (pl), (B) MsCath1 antibody against hemocytes and (C) plasma. Hemocytes and plasma were studied after different treatments: parasitism (P), no treatment (NP), after PBS injection (PBS), after injection of beads (B) and after bacterial injection (Bac).

Cysteine protease activity varies during parasitism

Cysteine protease activities and medians were reported graphically (Figure 18). Normal distributions were not demonstrated for all samples, we therefore performed the Mann-Whitney test with a unimodal density to compare pairs of independent samples. Statistical tests related to cysteine protease activity are shown in Table 8.

Protein concentrations were not significantly different between 24h parasitized and non parasitized hemocyte extracts ($p=0.091$, $U=25$, $n=19$), between 24h parasitized and non parasitized plasma samples ($p=0.067$, $U=26$, $n=19$) and between 48h parasitized and non parasitized plasma samples (data not shown). In contrast, in 48h hemocytes, protein concentration was found to be significantly lower in parasitized than in non parasitized samples ($p=0.003$, $U=10.5$, $n=19$) (data not shown). Whatever the time point, protein concentration in hemocyte extracts was always found to be lower than concentration in plasma samples (data not shown). Cysteine protease activities were therefore presented according to protein concentration for each sample.

Cysteine protease activities recorded in hemocytes extracts were higher than in plasma samples (Figure 18), suggesting that the ratio between cysteine proteases and total protein concentration is high in hemocytes compared to plasma. Furthermore the low level of cathepsin activity in plasma could partly depend on the presence of an endogenous multicystatin which may partially inhibit cysteine proteases (Miyaji et al., 2007). The residual cathepsin inhibitory capacity was further evaluated by measuring the ability of hemolymph to inhibit exogenous E-64-titrated papain, according to Assfalg-Machleidt and co-workers (1992) (Assfalg-Machleidt *et al.*, 1992). Indeed this hypothesis was confirmed by detection of a papain inhibitory activity (data not shown), which was abolished by heating plasma at 100°C for more than 10 min prior activity assays.

By comparing medians we notice that cysteine protease activity is overall higher in parasitized hemocyte samples compared to non parasitized, one exception is observed in 48h hemocytes where cysteine protease activity is higher in non parasitized samples when the ZRR-AMC substrate is used. More precisely, cysteine protease activity is significantly higher in parasitized hemocytes at 24h for all substrates and at 48h for ZFR-AMC and ZLR-AMC. In plasma, cysteine proteases activity was significantly higher at 24h in parasitized samples compared to non parasitized samples for ZLR-AMC and ZRR-AMC substrates. Finally, cysteine protease activities were not found to be significantly different between 48h parasitized and non parasitized plasma.

In conclusion, these results show that cysteine protease activity is significantly higher after 24h in hemocytes and plasma from parasitized hosts compared to non parasitized caterpillars. However, at 48h cysteine protease activity is statistically the same between parasitized and non parasitized plasma suggesting that cysteine protease activity is regulated between 24 and 48h after parasitism.

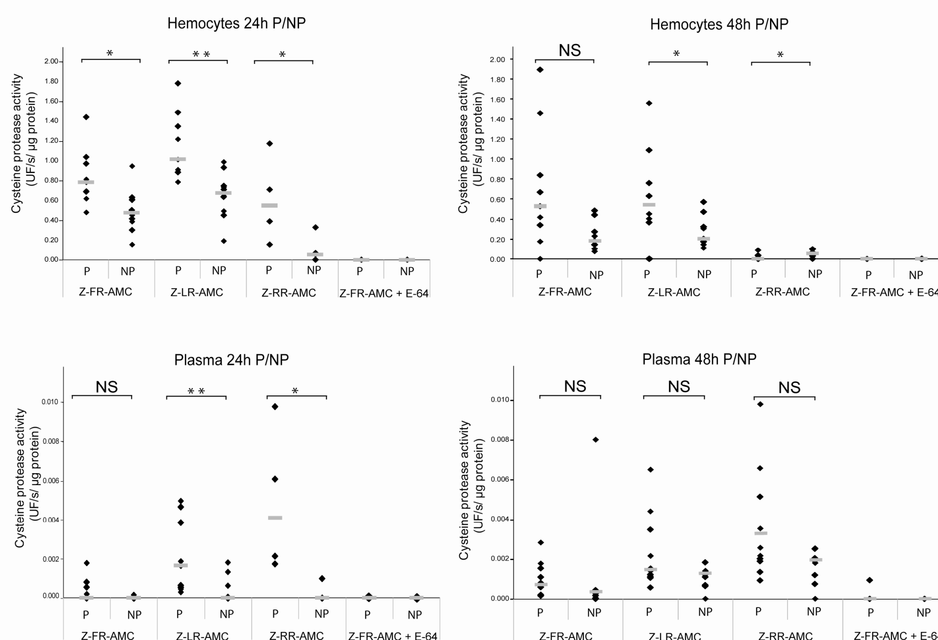


Figure 18 Cysteine protease activity in hemocyte extracts and plasma sample of *M. sexta*
Cysteine protease activities were measured in hemocytes and in plasma from parasitized and non parasitized hosts and were performed on 9 to 10 samples with three different substrates (ZFR-AMC, ZLR-AMC,ZRR-AMC). Specificity of the C1A cysteine protease activity was controlled by using the cysteine protease inhibitor E-64. Medians were calculated and are indicated by grey horizontal bars. * corresponds to p value<0.05 and ** corresponds to p value<0.01

Table 8 Comparison of parasitized and non parasitized samples of *M. sexta*

Sample	P value	U-Statistic	Number of measures
Hemocytes 24h P/NP			
ZFR-AMC	0.002	11	19
ZLR-AMC	0.001	8	19
ZRR-AMC	0.016	1	19
Hemocytes 48h P/NP			
ZFR-AMC	0.078	27	19
ZLR-AMC	0.047	24	19
ZRR-AMC	0.027	21	19
Plasma 24h P/NP			
ZFR-AMC	0.111	30	19
ZLR-AMC	0.003	12	19
ZRR-AMC	0.008	0	19
Plasma 48h P/NP			
ZFR-AMC	0.0912	28	19
ZLR-AMC	0.156	32	19
ZRR-AMC	0.1214	30	19

P:parasitized; NP:non parasitized

Discussion

During parasitism of *M. sexta*, the endoparasitoid braconid wasp *C. congregata*, injects its eggs, venom and polydnavirus particles into its host. The eggs injected by *C. congregata* can be recognized as non self by *M. sexta* and be encapsulated. However the venom components and more particularly expression of the polydnavirus CcBV play a critical role in controlling host defence responses and modifying host development to the advantage of the wasp.

Among potential virulence factors encoded by CcBV, cystatins, which are cysteine protease inhibitors are suspected to play a prominent role in parasitism success. In order to identify the role of viral cystatins in parasitism we present here the first characterization of C1A cysteine proteases in *M. sexta*. Interestingly we could show that a particular class of "26/29 kDa" protease is subject to transcriptional control in the course of parasitism and more generally after immune challenge. Furthermore, antibodies raised against Cystatin1 and the 26/29 kDa protease MsCath1 revealed that these proteins are both found in the extracellular compartment following parasitism, suggesting they have the potential to interact. We also show an increase in cysteine protease activity 24 hours after parasitism in both hemocytes and plasma, which is no longer detected 48h after parasitism in plasma.

Four cysteine proteases were isolated from *M. sexta*

RT-PCR and RACE approaches on *M. sexta* hemocytes and fat body enabled us to isolate 4 different C1A cysteine proteases corresponding to a Cathepsin B, a Cathepsin L and to two "26/29kDa class" cysteine proteases, this latter class had so far only been described in insects (Fujimoto et al., 1999; Saito et al., 1992). All proteases contain a signal peptide (preregion) and a pro-domain which is cleaved after protease activation to release the C-terminal protease domain. 26/29 kDa proteases carry a supplementary 26 kDa domain situated before the pro-domain which is also cleaved (Fujimoto et al., 1999; Saito et al., 1992). *M. sexta* cathepsins B and L (MsCathB, MscathL) were shown to have a high level of sequence conservation with other insect and mammalian B and L Cathepsins, respectively, with levels of similarity between mature proteases ranging from 73 to 92 %. The 26/29kDa class cysteine protease, MsCath1, displays in its mature form more sequence conservation with the *Sarcophaga peregrina* homolog than to MsCath2 suggesting the two *M. sexta* protease genes result from a duplication event predating *M. sexta* divergence. Previously, only one other 26/29 kDa protease had been characterized in *Sarcophaga peregrina* (Saito et al., 1992) and homologs had only been identified in insects (Fujimoto

et al., 1999). Sequence comparisons in data banks now reveal that 26/29 kDa protease related sequences have been identified in various organisms, including a vertebrate, suggesting these particular proteases must have a larger distribution in animals. The 26 kDa region constitutes a supplementary domain in comparison to other proteases and its function remains unknown. This region shows no real sequence homology to known proteins and is overall weakly conserved between proteases. However the growing number of available sequences in the databanks enabled us to define conserved motifs and amino acids which may be interesting targets for future functional studies. In the *S.peregrina* protease this 26 kDa domain was suggested to be involved in non self recognition (Saito *et al.*, 1992), but no functional studies have been performed to support this hypothesis.

What could be the role of *M. sexta* C1A cysteine proteases and are they likely to be targeted by PDV cystatins ?

Considering that in insects, these proteases are implicated in different physiological processes such as digestion, development and immunity and that PDV target these last two functions, we wanted to determine whether these proteases showed developmental or immune challenge regulation.

At the different larval stages studied no obvious regulation could be observed at the transcriptional level, except a slight down regulation of MsCath1 at the fourth larval instar. In *Helicoverpa armigera* caterpillars, Liu and collaborators (2006) reported stage dependant transcriptional regulation of Cathepsin L, making this protease an excellent candidate to be involved in larval moult. Thse authors could show that injection of a cysteine protease inhibitor reduced cathepsin L activity and delayed fifth to sixth instar moulting. The possibility that some of the proteases we have identified in *M. sexta* are involved in the caterpillar's development cannot completely be ruled out.

Our study however provides evidence that *M. sexta* 26/29kDa cysteine proteases could be involved in physiological functions related to immunity. In contrast to MsCathB and MsCathL genes, which are not affected by immune challenge, we showed that MsCath1 and MsCath2 were down regulated in fat body, at the transcriptional level 18 hours after bacterial injection and this effect was even more drastic 18 hours after parasitism. Parasitism-induced transcriptional down regulation led to visible protein down regulation at 48 hours. Indeed MsCath1 protein could no longer be detected in 48 hour parasitized hemocytes. This protein down regulation can also be related to the decrease in cysteine protease activity observed 48h after parasitism.

Several studies have already reported cysteine protease transcriptional regulation after an immune challenge however in all cases the outcome is upregulation of these proteases. In the tsetse fly, *Glossina morsitans morsitans*, a blood meal containing trypanosomes induces an up-regulation of cathepsin B (Yan *et al.*, 2002). Cathepsins B, F and L cDNAs are present in a cDNA library constructed from immune challenged tsetse flies (Attardo et al., 2006). Cathepsin L was also shown to be transcriptionally up-regulated in *Drosophila* infected by *E. coli* (De Gregorio et al., 2001; Levy et al., 2004). Cysteine protease transcriptional responsiveness to immune challenge therefore appears to differ between *M. sexta* and these two dipterans species and whether this is a general trend in lepidopterans would be interesting to investigate.

Protease transcriptional regulation observed after an immune challenge suggests a potential involvement of proteases in physiological functions related to immunity, however this is not the only level at which immune related regulation was observed. Indeed, MsCath1, which in naïve *M. sexta* is restricted to hemocytes was relocalized to the plasma in 24 hour parasitized larvae only. Several reports of cysteine protease release into insect hemolymph following immune challenge can be found in the literature. After injection of sheep red cells in *S. peregrina* larvae, more than 80 % of the 26/29 kDa enzyme is released into the hemolymph, which led authors to speculate that this protease may be involved in non-self immune reactions (Saito et al., 1992). Recently, Cathepsin B protein was shown to be up-regulated in *Plutella xylostella* plasma following parasitism by *C. plutellae*, a closely related species to *C. congregata* (Song et al., 2008). A hemolymph proteomic analysis of the *Drosophila* systemic immune response revealed that Cathepsin L is induced after infection by the fungal pathogen *Beauveria bassiana* (Levy et al., 2004). These results can lead us to speculate that the other *M. sexta* cysteine proteases may also be released in the plasma after parasitism. This release, whether passive or active, attracted our attention for two reasons, first it could constitute, as postulated by Saito and co-workers (1992), an immune response, secondly since viral cystatins were also expected to be secreted into the hemolymph, they could therefore colocalize with their targets. An antibody raised against cystatin1 confirmed that viral cystatins are secreted into the hemolymph and that both cystatin1 inhibitor and MsCath1 protease proteins could be detected extracellularly at 24 hours post parasitism. The fact that cysteine protease activities are no longer significantly different between non-parasitized and parasitized 48 hours plasma samples, compared to 24 hours plasma samples, could therefore be related to inhibition by viral cystatins

Although mammalian cathepsins (CP) are primarily located in lysosomes, under certain circumstances these enzymes can also be released from the cell (Brix *et al.*, 2008). Phagocytes can

release these proteases during phagocytosis, and in the case of "frustrated phagocytosis", in which the cell is unable to ingest a particle (ie: a bacterial colony), large quantities of active enzymes are delivered. Extracellular C1A CP can then be involved in remodelling/damaging extracellular matrix and tissues but also in cleaving cytokines and complement to produce chemoattractants (Dickinson, 2002). In some cancer cells cathepsins are secreted or associated with the plasma membrane and participate in the degradation of the extracellular matrix during tumor progression (Mohamed and Sloane, 2006; Palermo and Joyce, 2008). Once outside the cell these enzymes can potentially be harmful to the organism and can be (or not in a pathological situation) kept in check by endogenous cystatins which will control their aggressiveness. In *M. sexta*, an endogenous cysteine protease inhibitor, MsCPI, has been isolated from larval hemolymph and is likely to play a role in the control of "accidental" or programmed release of C1A cysteine proteases (Miyaji et al., 2007). We suggest that upon parasitization physiologically larger quantities of cysteine proteases may be released and are kept in check by the action of both viral cystatins and MsCPI. It would however be necessary to verify that MsCPI is indeed expressed during parasitism.

What could be the consequences of cysteine protease release during parasitism in Lepidopterans ?

An answer could be found in studies relating to infection of *Heliothis virescens* larvae by recombinant baculovirus expressing the *S. peregrina* Cathepsin L (SCathL). SCathL is a cathepsin L-like cysteine protease that functions in its natural host in basement membrane remodelling during insect development (Homma *et al.*, 1994). Interestingly, recombinant baculovirus expressing SCathL induce faster death of the tobacco budworm, *H. virescens*, compared to wild-type baculovirus and trigger tissue fragmentation and melanization of larvae shortly before their death (Harrison and Bonning, 2001). Purified ScathL however was not able to activate pro-PO *in vitro* showing that the melanisation cascade is activated indirectly by this enzyme (Li et al., 2008). Transmission and scanning electron microscopy revealed that application of low levels of ScathL leads to degradation of the basement membrane, while high concentrations results in complete loss of tissue (Tang et al., 2007). Furthermore, hemocytes were visible at the sites of damage, indicating they may have been recruited to the sites of injury for repair. Since modifications in basement membrane in insects can lead tissue to be recognized as non-self, it has been proposed that ScathL acts by degrading the basement membrane, which causes in turn the host's immune system to recognize the modified basement membrane as non-self, leading to melanisation (Li et

al., 2008). Remarkably, *H. virescens* larvae parasitized by *Cotesia marginiventris*, a close relative of *Cotesia congregata*, when infected with recombinant baculovirus expressing ScathL do not melanise, probably due to the action of the associated PDV (Nusawardani et al., 2005). Collectively these data can enable us to propose a working model for cysteine protease (CP) release during parasitism in *M. sexta*: (1) parasitism leads to CP release by active or passive mechanisms (2) CP release and activity is kept in check by viral and endogenous inhibitors which in consequence limit basement membrane degradation and thereby impede melanization and hemocyte recruitment. One way of testing this model would be to analyze CP release and consequences in circumstances where the parasitoid is encapsulated in *M. sexta*. Rodriguez-Pérez and co-workers (2005) showed that *Cotesia flavipes* eggs are encapsulated by tobacco hornworm, either because CfBV is partially denied entry into the host cells or is not fully expressed. It would be interesting to determine whether parasitism in this system also induces CP release, and if CP activity, which in this case would be partially controlled by endogenous inhibitors only, can be linked to subsequent immune reactions such as melanisation and hemocyte recruitment.

The fact that MsCath1 and MsCath2 are specifically down regulated at the transcriptional level during parasitism suggests they could be involved in different specific functions which do not involve Cathepsin B and L. In mammalian systems certain cathepsins have been reported to be involved in Toll-like receptor signalling (Asagiri et al., 2008) and virus entry into host cells (Chandran et al., 2005; Ebert et al., 2002; Qiu et al., 2006), two processes which could also be relevant in insect cells. RNAinterference of MsCath1 and 2 may help to determine their involvement in such processes.

In conclusion, this work reveals that C1A cysteine proteases are likely to be important targets of CcBV polydnavirus in the course of parasitism of *M. sexta* by *C. congregata*. A first, unexpected level of regulation was observed for a specific class of protease at the transcriptional level. The 26/29 kDa proteases are specifically down-regulated during parasitism suggesting their inactivation is essential. A second level of regulation which is probably not specific involves the release of proteases in the plasma. The finding that viral inhibitors and at least one protease are present at the same time in the same place suggest these proteins have the potential to interact, and therefore CP activity is likely to be kept under control during parasitism. This work shows that the identification of virulence proteins can lead to investigation of potentially novel and so far unsuspected pathways involved in insect immunity.

B-Attempt to isolate viral cystatin-host cysteine protease complexes *in vivo* in the course of parasitism of *Manduca sexta* by *Cotesia congregata*

Introduction

C1A cyteine proteases are the natural targets of cystatins with which they form tight-binding reversible complexes (Abrahamson *et al.*, 2003; Turk *et al.*, 1997). These complexes can be visualized by agarose or polyacrylamide gel electrophoresis as has been shown for example by mixing cystatin C and papain (Abrahamson *et al.*, 1987). Furthermore, several studies have characterized with precision the sites of interaction between the two partners by obtaining the crystal structure of the inhibitor-protease complexes (Jenko *et al.*, 2003; Ljunggren *et al.*, 2007; Stubbs *et al.*, 1990).

In the *Cotesia congregata-Manduca sexta* interaction, several lines of evidence suggest that viral cystatins target caterpillar host C1A cysteine proteases (1) recombinant CcBV cystatin 1 was shown to be a functional inhibitor *in vitro* (Espagne *et al.*, 2005) (2) the 3D model of CcBV cystatin1 revealed the presence of the typical tripartite "wedge" which is important for the interaction with C1A cysteine proteases, (3) strong diversifying selection is acting in the vicinity of at least two of the conserved active sites in polydnavirus cystatins (Serbielle et al. 2008) (4) we showed that at least one viral cystatin (cystatin 1) and one host C1A cysteine protease (MsCath1) colocalize in the hemolymph plasma specifically during parasitism indicating they have the potential to interact (Chapter V). We therefore aimed to isolate the potential complex formed by viral cystatins and host proteases *in vivo*. Here we describe our attempt to isolate the natural targets of PDV cystatins *in vivo* in the course of parasitism of *M. sexta* by *C. congregata*.

Material and Methods

Sample preparation

Hemolymph from three 24 hour super-parasitized larvae (4 to 5 ovipositions checked visually) was collected by cutting a proleg and was diluted in an equal volume of PBS solution

containing 0.02% of phenylthiourea. Hemocytes were then discarded from the sample after a gentle centrifugation (5 min, 900g). Microcon centrifugal filters (Millipore) were used to concentrate plasma samples following the manufacturer's instructions. Plasma samples concentrated 10 fold were then conserved at -80°C.

Identification of potential cystatin-protease complexes

Our aim was to isolate the expected complex formed by cystatins and cysteine protease targets. This analysis was carried out by electrophoresis under non denaturing conditions followed by SDS-PAGE according to Moreau (2004). Hemolymph (plasma) protein samples were separated into three. 15 µl of each sample was subjected to electrophoresis in non denaturing conditions on 10 % polyacrylamide gels. One gel was stained with Coomassie Blue G-250 (Matsui *et al.*, 1999) and the other two gels were transferred onto PVDF membranes and immunoblotted with cystatin or MsCath1 antibody respectively. Bands which cross-reacted with both antibodies were excised from the Coomassie stained gel, resuspended in Laemmli buffer and heated for 15 min at 50°C. These samples were then subjected to three 12 % SDS-PAGE in parallel. Two gels were transferred onto PVDF membranes and blotted with anti-cystatin or anti-MsCath1 antibody. Bands reacting with antibodies were excised from an equivalent Coomassie stained gel and subjected to analysis by LC/MS/MS spectrometry.

LC/MS/MS spectrometry analysis

Samples were first digested by trypsin (Trypsin Sequencing Grade; Roche) for 16 hours at 37°C. Peptides produced by digestion were then extracted from the gel and analyzed. The LC/MS/MS was done on a nanoHPLC CapLC system (Waters, Manchester, UK). The samples were concentrated on a precolumn (Monolithic trap, 0.3mm I.D, Dionex) and separated on a 15cm x 75µm i.d column in reverse phase. The MS and MS/MS analyses were performed with a two hybrid time-of-flight (Q-TOF) mass spectrometer (Micromass, Manchester, UK) equipped with a Z-spray ion source. LC/MS/MS raw data were processed automatically with the ProteinLynx Process program. Data analysis was then performed with Mascot (Matrix Science Ltd., London, UK) against NCBI and EST databases but also against cDNA C1A cysteine protease sequences obtained in the laboratory (Chapter V). Peptide precision was set to 0.3 Da for MS and MS/MS. Double and triple charged peptides were used for the research in databases. Two missed cleavage by trypsin were accepted and carbamidomethylated cysteine and oxidized methionine were set as variable modifications and all taxa were used.

Results

Detection by Western analysis of protein complexes cross-reacting with Cystatin1 and MsCath1 antibodies

This study was performed according to Moreau (2004) in order to isolate *in vivo* potential host cysteine proteases targeted by viral cystatins during parasitism.

We first performed an electrophoresis of hemolymph proteins in native (non-denaturing) conditions to conserve protein complexes, followed by immunoblotting with MsCath1 and viral cystatin antibodies. Western analysis revealed that each antibody cross-reacted with two protein bands in parasitized hemolymph of approximate molecular weight 150 kDa (Figure 19, Band 1) and 60 kDa (Figure 19, Band 2).

The corresponding bands were excised from both parasitized and non parasitized lanes of the equivalent Coomassie stained gel and were subjected to electrophoresis in denaturing conditions, followed by immunoblotting using cystatin and MsCath1 antibodies (Figure 20). Western analysis using anti-cystatin antibody on bands 1 and 2 showed that no cross-reacting proteins could be detected in non parasitized samples and that a band of approximately 15 kDa, corresponding to the expected size of viral cystatins, could be detected in parasitized samples (Figure 20). These results suggest that the bands 1 and 2 which cross-reacted with cystatin and MsCath1 antibodies do contain viral cystatins. Western analysis using MsCath1 antibody on excised band 1 revealed the presence of two high molecular weight proteins in parasitized samples, with one band also cross-reacting in non-parasitized samples (Figure 20). In excised band 2, two proteins reacted to anti-MsCath1 antibody at approximate molecular weights 40 kDa and 17 Kda (Figure 20). None of the bands detected correspond in size to mature MsCath1 (see Chapter V), but given the particular sample preparation and electrophoresis conditions, modifications in mobility behaviour may be expected. The 40 kDa or 17 kDa bands represent the best candidates to correspond to MsCath1.

Proteins identified by viral cystatin antibody and MsCath1 antibody constitute potential proteins involved in the viral cystatin-host target protein complex. The corresponding immunoreactive bands (1, 10, 2, 3, 4, 6, 7) were excised from the equivalent Coomassie gel for Mass spectrometry analysis. Additional bands only visible in parasitized samples were also analyzed at the same time (8, 9, 11) and one band from a non parasitized sample (5). A total of 11 bands were therefore analyzed by LC/MS/MS.

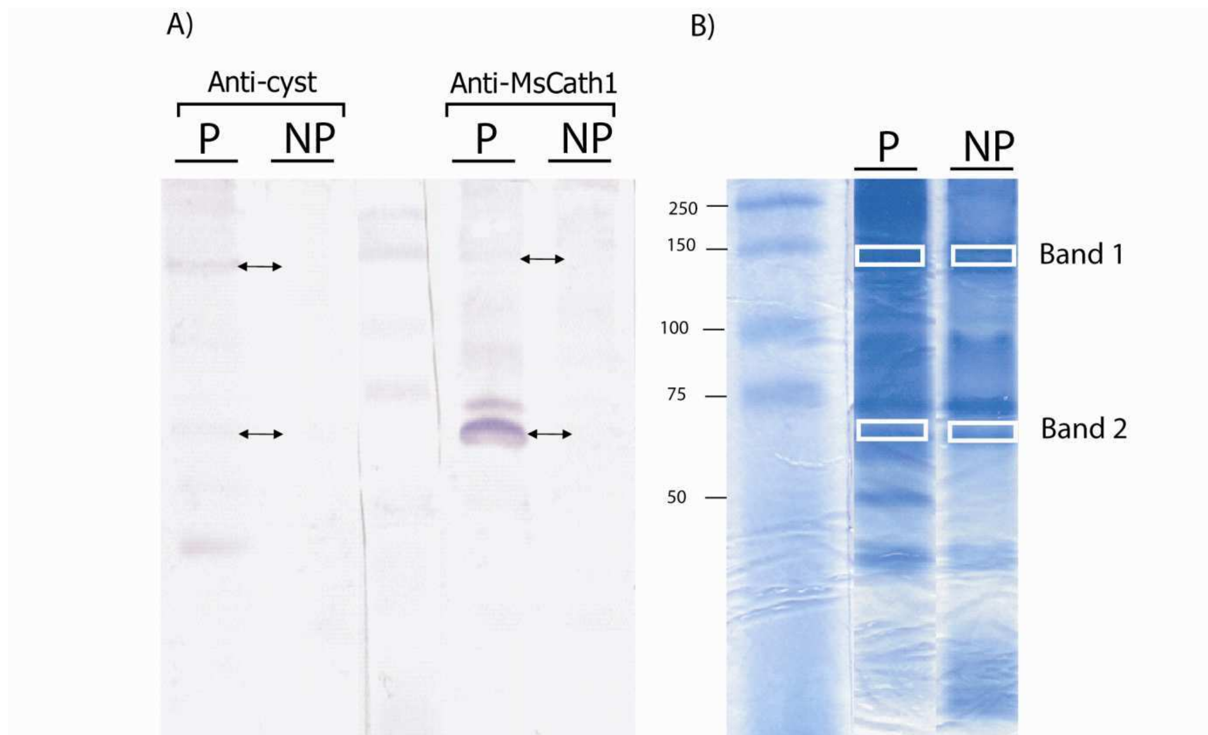


Figure 19 Immunoblot with anti-cystatin or anti-MsCath1 antibody (A) and Coomassie gel stained in native conditions (B) of non parasitized (NP) and parasitized (P) *M. sexta* plasma.
Band 1 and Band 2 were cut from the gel for migration in denaturing conditions

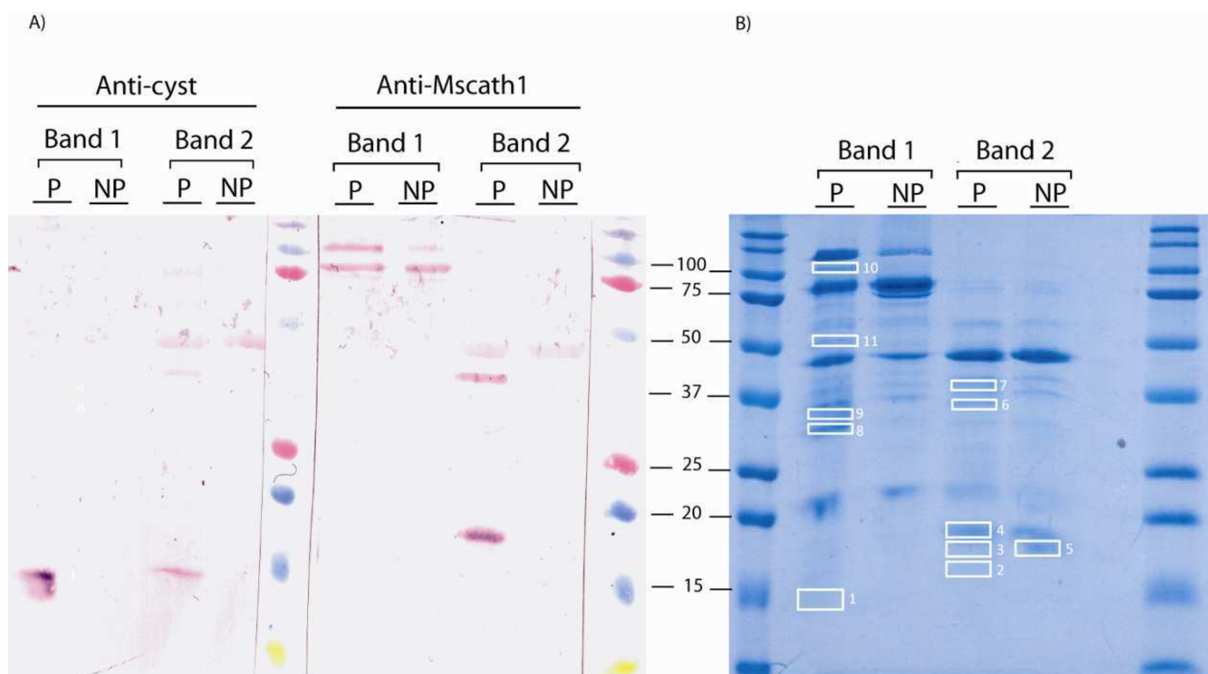


Figure 20 Immunoblot (A) and Coomassie gel in denaturing conditions of excised bands (B).
Spots 1-11 were cut from the gel for MS/MS analyses. P:parasitized sample, NP:non-parasitized sample.

Table 9 Identification of plasma proteins isolated in the 11 spots

Polydnavirus proteins						
Function	**Identity/homology**	**Spot**	**coverage(%)**	**Species**	**MW**	**acc number**
Reduction of circulating hemocytes	EP1	6	3	*CcBV*	30219	BAA06195
	EP1-like	8	27	*CcBV*	30654	BAA06195
		9	32	*CcBV*	30654	BAA06195
		10	10	*CcBV*	30654	BAA06195
Unknown functions	Hypothetical Protein 2.13	1	16	*CcBV*	17574	YP184818
		2	12.5	*CcBV*	17574	YP184818
	Hypothetical Protein 22.3	2	4	*CcBV*	15572	YP184855
		3	8	*CcBV*	15572	YP184855
		4	23	*CcBV*	15572	YP184855
	Hypothetical Protein 1.5	6	5	*CcBV*	32918	YP184761
		7	8	*CcBV*	32918	YP184761
	Cystatin 1	1	6.6	*CcBV*	15457	YP184846
Insect proteins						
Function	**Identity/homology**	**Spot**	**coverage(%)**	**Species**	**MW**	**acc number**
Pattern recognition						
	Peptidoglycan recognition protein 1B	3	5	*M. sexta*	21517	AAO21502
	Immunolectin A precursor	7	3.5	*M. sexta*	35279	AAC33576
Phenoloxidase						
		10	4	*M. sexta*	40710	AAV91006
	Hemolymph proteinase 6	7	5.6	*M. sexta*	38981	AAV91004
	Serine proteinase-like protein 4	11	10	*M. sexta*	43122	AAV91027
	Serine proteinase-like protein 2	11	4.3	*M. sexta*	43461	AAM69353
	Scolexin B (serine protease)	8	6.5	*M. sexta*	30003	AAC33576
Phenoloxidase						
	Serpin 3a	10	7.2	*M. sexta*	23290	AOO21505
	Alaserpin (Serpin 1)	6	5.9	*M. sexta*	43483	P14754
		9	5.9	*M. sexta*	43483	P14754
		10	3.3	*M. sexta*	43483	P14754
		11	5.1	*M. sexta*	43483	P14754
Iron sequestering	Transferrin precursor	9	1.6	*M. sexta*	75175	P22297
		10	14	*M. sexta*	75175	P22297
Lipid metabolism	Apolypophorin3-precursor	4	29	*M. sexta*	20781	P13276
		5	14.3	*M. sexta*	20781	P13276
Heat shock protein						
	Hsc70Cb	10	1.7	*A.mellifera*	91146	XP623199
Other						
	AGAP000889	3	9	*A. gambiae*	16513	XP316668
		4	8.9	*A. gambiae*	16513	XP316866

The immunoreactive proteins present similarities with viral cystatins but not with cysteine proteases

Spot number 1, which corresponds to proteins recognized by viral cystatin antibody, was found to indeed present similarities with viral cystatin 1, confirming that the complex of proteins in excised band 1 contains cystatin 1. Cystatin 1 is also likely to be present in excised band 2 given the results of the Western analysis in denaturing conditions (Figure 20). However
proteins at the expected size for cystatins in band 2 could not be detected in the equivalent Coomassie gel, probably because they are at concentrations below detection level in this particular sample (Figure 20).

In spots number 2, 3, 4, 6, 7, 8, 9, and 10, which reacted with anti-MsCath1 antibody (Figure 20), no peptides were found to present similarities with cysteine proteases.

Proteins specifically detected in parasitized samples

Two classes of proteins were identified corresponding to polydnavirus proteins and to host proteins (Table 9) and possibly one wasp protein.

Peptides isolated from 4 spots were identified to present homologies with the CcBV EP1 and EP1-like early expressed protein which has an expected molecular weight of 30 KDa. The other viral proteins identified correspond to Cystatin 1 (see above) and to hypothetical proteins of unknown function.

Among host proteins identified, most of them belong to proteins involved in insect immunity. Indeed 3 spots contain peptides with homologies to pattern recognition receptors, 5 spots are found to contain peptides similar to serine proteases which participate in phenoloxydase cascade activation and 4 spots present peptides similar to serpins which are regulators of the phenoloxydase cascade. Finally, 2 spots contain peptides similar to transferrin involved in iron sequestering.
Peptides correspond to a protein of unknown function in *Anopheles* and a heat shick protein in *A. mellifera* were also identified in parasitized samples only. They could correspond to an induced host protein or to a wasp venom protein.

Proteins detected in parasitized and non-parasitized samples

Peptides corresponding to a protein involved in insect immunity (serpin 1) and to a protein involved in lipid metabolism (Apolypophorin 3-precursor) were detected in bands of both non-parasitized and parasitized samples.

Discussion

Our attempt to isolate polydnavirus cystatin and host cysteine protease complexes *in vivo* during natural parasitism of *Manduca sexta* by its parasitoid wasp did not succeed. We could however observe the presence of proteins cross-reacting with both anti-cystatin1 and anti-MsCath1 antibody. From these cross-reacting proteins, we could demonstrate that at least one partner of the interaction is present which corresponds to cystatin 1, we could not however recover MsCath1 cysteine protease.

The failure to identify MsCath1 cysteine protease in these samples could have several explanations:

1. Hemolymph samples have a high protein concentration particularly during parasitism. Cysteine proteases may not be present in sufficient quantity to be identified and are masked by contaminating proteins present in overwhelming quantities. The fact that the protein band corresponding to cystatin in the excised band 2 was not visible in the Coomassie gel although it was detectable by Western (Figure 20) is an indication that low levels of the potential complex were present in this sample (Figure 20). Furthermore several of the proteins that we identified have previously been shown to be abundant or induced during parasitism or infection. The polydnavirus early-expressed protein (EP1) constitutes one of the major proteins, representing 5% of the total hemolymph protein 24h after parasitism. This was the first virus gene shown to be highly expressed in parasitized *M. sexta* hosts (Harwood *et al.*, 1994). Apolipophorin-III was also shown to be induced in hemolymph of parasitized *Plutella xylostella* (Song et al. 2008). Serine proteases, serpins and transferrins were shown to be induced in hemolymph of *Drosophila melanogaster* infected by the fungal pathogen *Beauveria bassiana* (Levy *et al.*, 2004). Transferrins were also shown to be up-regulated in mosquitoes after immune system activation (Paily *et al.*, 2007).

2. Cysteine proteases are known to be associated by non covalent binding to other proteins such as transferrins and serpins (Gettins, 2002; Xing and Mason, 1998). The fact that these proteins were retrieved by our analysis may be an indication that we have the correct “cortège” of

proteins, but since transferrins and serpins are probably present in large quantity in the plasma (see above) they may "mask" cysteine proteases and impede their detection.

3. In our analysis the antibody used was raised against a specific peptide of MsCath1 and is not likely to cross react with other *Manduca* cysteine proteases. MsCath1 may not be the privileged target of viral cystatins. Song *et al.* 2008, for example, were able to detect a Cathepsin B in the hemolymh of *P. xylostella* after parasitism by *Cotesia plutellae*. Using a similar approach, but with a wider panel of antibodies against newly characterized *M. sexta* cathepsins (see chapter V), may help to identify complexes.

4. There is still relatively limited sequence data in the data banks concerning *M. sexta*. The genome of this insect is not sequenced and a few EST libraries have been constructed (Papanicolaou *et al.*, 2008; Zou *et al.*, 2008). There is still therefore very limited sequence data on *Manduca* cathepsins and it is possible that in our analysis we do have peptides corresponding to cathepsins but which we cannot identify.

5. Biochemical properties of proteins may render their analysis difficult because of a) limited transfer on membranes b) refractoriness to analysis by Mass spectrometry. C1A cysteine proteases are apparently among a category of proteins which are not so easy to identify by these techniques (G. Lalmanach, personal communication).

6. C1A cysteine proteases may not be the only targets of cystatins in this interaction. It may be necessary to investigate whether the proteins we identified could also targeted by cystatins.

This small scale analysis did however enable us to identify polydnavirus proteins with unknown function that are present in host hemolymph, which suggests these hypothetical proteins play a role in the host-parasite interaction. Approximately half of the coding sequences detected by genomic analysis of CcBV encode proteins of unknown function (Espagne *et al.*, 2004), it is therefore of great interest to have information on the expression of these genes in the course of the interaction.

This analysis also enabled us to identify a subset of host proteins which are potentially upregulated at 24 hours after parasitism. Interestingly we identified proteins which were not detected in the proteomic analysis *of P. xylostella* parasitized by *C. plutellae* (Song et al., 2008), most likely because of the different time frame used. Host proteins which were isolated in this study are known to take part in insect immunity and/or metabolism.

The pattern recognition receptors (PRRs) include receptors making part of the immune system, which recognize molecules called pathogen associated molecular patterns (PAMPs). PRRs

stimulated by PAMPs trigger signaling cascades to activate immune cells or antimicrobial peptides to isolate or kill pathogens (Dimopoulos et al., 2002; Pancer and Cooper, 2006). Serine proteases and serpins, which are respectively activators and inhibitors of the phenoloxidase cascade, are involved in melanization processes (Jiang *et al.*, 2003; Nappi *et al.*, 2005).
Apolipophorin III belongs to the functionally important class of lipoproteins which are responsible for lipid transport and lipoprotein metabolism in various animal classes. (Weers and Ryan, 2006). This protein has also been shown to function as an activator of several immune responses (Kim *et al.*, 2004; Whitten *et al.*, 2004).

Finally, the last class of host proteins isolated belongs to transferrins which sequester iron. Transferrins are proteins up-regulated in mosquitoes and *Drosophila* after immune system activation (Levy *et al.*, 2004; Paily *et al.*, 2007). The up-regulation reported, suggests that transferrin is an active component of the *Drosophila* and mosquito defense system against invading parasites or pathogens. They may act by sequestering iron and limiting parasite access to iron resources (Paily *et al.*, 2007). They have also been proposed in *Drosophila* to act as anti-microbial peptides or as inducers of the immune response (Levy *et al.*, 2004).

Functional analysis of PDV genes is a difficult task considering that reverse genetics approaches are impossible in these systems. So far functional characterizations have been mainly performed by *in vitro* biochemical assays and transient over-expression in non-host and sometimes host cell lines *in vitro*, and more rarely *in vivo*. Although it can be discussed whether over-expression of genes in non host cells is a relevant functional assay, it is oubvious that these experiments enable to identify targeted pathways in the host. It is likely that *in vivo* certain viral gene products have a more subtle effect both in time and in space and that in the natural system a blend of genes and their products are necessary to ensure perfect host control.
Here our aim was to try and identify complexes which would form in the context of natural parasitism without over-expression of either partner. This approach did not, however, enable us to isolate the complex probably because of the limitations that we have already discussed above.

What alternative strategies could be used ?

To continue in the context of an *in vivo* approach not too distant from the natural context we could try to obtain larger quantities of hemolymph proteins by harvesting larger quantities of hemolymph and concentrating the samples. Finer separation of proteins by 2D electrophoresis should also enable to limit the presence of co-migrating contaminating bands. Detection of a

wider panel of *M. sexta* cysteine proteases is also necessary either by raising specific antibodies against newly characterized proteases (see Chapter V)

Other approaches which consist in "trapping" or "pulling down" the interacting partners are also possible:

1. The Biacore system for example allows studying real time molecular interactions. One of the interacting partners has to be purified and immobilized on a chip (ie: viral cystatins). Hemolymph can then be injected and molecules interacting with viral cystatins can be recovered.
2. In the same vein, GST-pull down experiments could be developed. In this system hybrid GST-cystatin recombinant proteins could be incubated with parasitized hemolymph. Sepharose beads which have a high affinity for Glutathione-S-transferase can then be used to "pull-down" the GST-cystatin-interacting protein complexe. A similar approach was used by Lu and co-workers (2008) to isolate proteins interacting with the polydnavirus MdBV Egf1.0 protein.
3. Co-immunoprecipitation experiments could also be developed. In this approach anti-sera raised against cystatins and proteases are used to precipitate proteins in the complexe. This approach enabled to demonstrate the interaction between a cystatin-like protease inhibitor of the plant pathogen *Phytophthora infestans* and a papain-like extracellular cysteine protease from the tomato host plant (Tian *et al.*, 2007).

In order to understand the role of viral cystatins during parasitism, the isolation of their targets in hosts is of great interest because (1) it will enable to determine which proteases are targeted by cystatins during parasitism, (2) we will then be able to study the physiological role of these proteases and the consequences of their inhibition and (3) we will be able to study the evolutionary processes involved between the two molecular partners and the evolutionary consequences in host-parasitoid association.

Our experience shows that *in vivo* functional approaches are difficult in these systems. However, this experiment has enabled us to unravel PDV and host proteins which are present in the hemolymph at the time of parasitism, and gives us a basis for improvement of the approach.

VI-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 endoparasitoïd wasps have been shown to be tightly associated with polydnaviruses 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, polydnaviruses 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 polydnaviruses, 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 Inchneumonidae 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 caspide) 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, IκB and protein tyrosine phosphatases

are known to have been acquired early during the virus-wasp association. The IκB 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
C. neustrial	*Malacosoma castraise*	Greece	Cyst neustrial	1
C. glabrata (56)	*Carcharodus baeticus* (Hesperidae)	Spain	Cyst glabrata56	1
C.sp near glabrata (60)	*Pyrgus cirsii* (Hesperidae)	Spain	Cyst glabrata60	1
C.sp near glabrata (7)	*Carcharodus alceae* (Hesperidae)	Spain	Cyst glabrata7	2
C. glabrata (6)	*Carcharodus tripolina* (Hesperidae)	Maroc	Cyst glabrata6	2
C. glabrata (1)	*Carcharodus alceae* (Hesperidae)	France	Cyst glabrata1	1
C. glabrata(2)	*Carcharodus alceae* (Hesperidae)	France	Cyst glabrata2	2
C. pieridis	*Aporia crataegi* (Nymphalidae)	Greece	Cyst pieridis	1
C. tenebrosa	*Polyommatus icarus* (Lycaenidae)	Scotland	Cyst tenebrosa	1
C. salebrosa (8)	*Epirrita autumnata* (Geometridae)	Finland	Cyst salebrosa8	2
C. salebrosa (9)	*Epirrita autumnata* (Geometridae)	Finland	Cyst salebrosa9	1
C.astrarches (?) (14)	*Aricia artaxerxes* (Lycaenidae)	England	Cyst astrarches14	1
C. astrarches (15)	*Aricia agestis* (Lycaenidae)	England	Cyst astrarches15	2
C. astrarches (16)	*Polyommatus thersites* (Lycaenidae)	Greece	Cyst astrarches16	3
C. astraches (37)	*Aricia agestis* (Lycaenidae)	England	Cyst astrarches 37	1
C. acuminata (23)	*Euphydryas maturna*	France	Cyst acuminata23	1
C. acuminata-agg (29)	*Melitaea telona* (Nymphalidae)	Greece	Cyst acuminata29	2
C. acuminata-agg (27)	*Melitaea phoebe* (Nymphalidae)	Italy	Cyst acuminata27	2
C. acuminata (52)	*Melitae phoebe* (Nymphalidae)	Spain	Cyst acuminata52	1
C. acuminata-agg (31)	*Melitaea telona* (Nymphalidae)	Greece	Cyst acuminata31	1
C. acuminata-agg (28)	*Melitaea telona* (Nymphalidae)	Greece	Cyst acuminata28	2
C. acuminata-agg (30)	*Melitaea telona* "argigia"(Nymphalidae)	Greece	Cyst acuminata30	3
C. kariyai	*Pseudaletia separata* (Noctuidae)	Japan	Cyst kariyai	6
C. vanessae (1)	*Cynthia cardui* (Nymphalidae)	Spain	Cyst vanessae1	3
C.vanessae (3)	*Vanessae atalanta* (Nymphalidae)	England	Cyst vanessae3	3
C. vanessae (45)	*Aglais urticae* (Nymphalidae)	Belgium	Cyst vanessae45	4
C. rufricus	*Autographa gamma* (Noctuidae)	Austria	Cyst rufricus	2
C. ofella	*Acronicta rumicis* (Noctuidae)	France	Cyst ofella	3
C. melitaearum (41)	*Melitaea deione* (Nymphalidae)	Spain	Cyst melitearum41	4
C. melitaearum (43)	*Melitaea deione* (Nymphalidae)	Spain	Cyst melitearum43	3
C. vestalis (2)	*Plutellae xylostella*	lab rearing	Cyst vestalis2	1
C. vestalis (3)	*Plutellae xylostella*	Benin	Cyst vestalis3	3
C. melanoscela	*Lymantria dispar* (Nymphalidae)	France	Cyst melanoscela	5
C. flavipes (2)	*Busseola fusca* (Noctuidae)	Kenya	Cyst flavipes2	4
C. flavipes (1)	-	-	Cyst flavipes1	2
C. chilonis	*Chilo suppressalis*		Cyst chilonis	4
C. sesamiae (1)	*Busseola fusca* (Noctuidae)	Kenya	Cyst sesamiae1	1
C. sesamiae (2)	*Busseola fusca* (Noctuidae)	Kenya	Cyst sesamiae2	3
C. glomerata (11)	*Aporia crataegi* (Nymphalidae)	Spain	Cyst glomerata11	2
C. glomerata (2)	*Pieris brassicae*	lab reared	Cyst glomerata2	2
C. glomerata (3)	-	-	Cyst glomerata2	2
C.sp near glabrata (61)	*Carcharodus baeticus* (Hesperidae)	Spain	Cyst glabrata61	1
Cotesia sp (2)	*Cerura vinula* (Noctuidae)	france	Cyst sp2	3
Cotesia sp (3)	*Cerura vinula* (Noctuidae)	France	Cyst sp3	2
C. sibyllarum	*Limenitis reducta* (Nymphalidae)	Spain	Cyst sibyllarym	3
C. rubecula	*P. rapae* (Pieridae)		Cyst rubecula	4
C. congregata (1)	-	-	Cyst congregata1	4
C. congregata (2)	*Manduca sexta* (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 $MgCl_2$, 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 $MgCl_2$, 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.

Figure 21 Bayesian wasp tree obtained with NADH and COI combined data set.

Node supports are shown by bootstraps and by posterior probabilities from Bayesian inferences respectively below and above each branch. Bootstrap scores or posterior probabilities lower than 50% are not represented

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 rufricus. 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.

Figure 22 Bayesian cystatin tree.
Node supports are shown by bootstraps and by posterior probabilities from Bayesian inferences respectively below and above each branch. Bootstrap scores or posterior probabilities lower than 50% are not represented. In bold the three cystatins genes from *C. congregata* sequenced genome. In pink, cystatins lineages not grouped according to bracovirus species and/or *Cotesia* wasps phylogeny.

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 ω=2.916 while 40% evolved under purifying selection with ω=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).

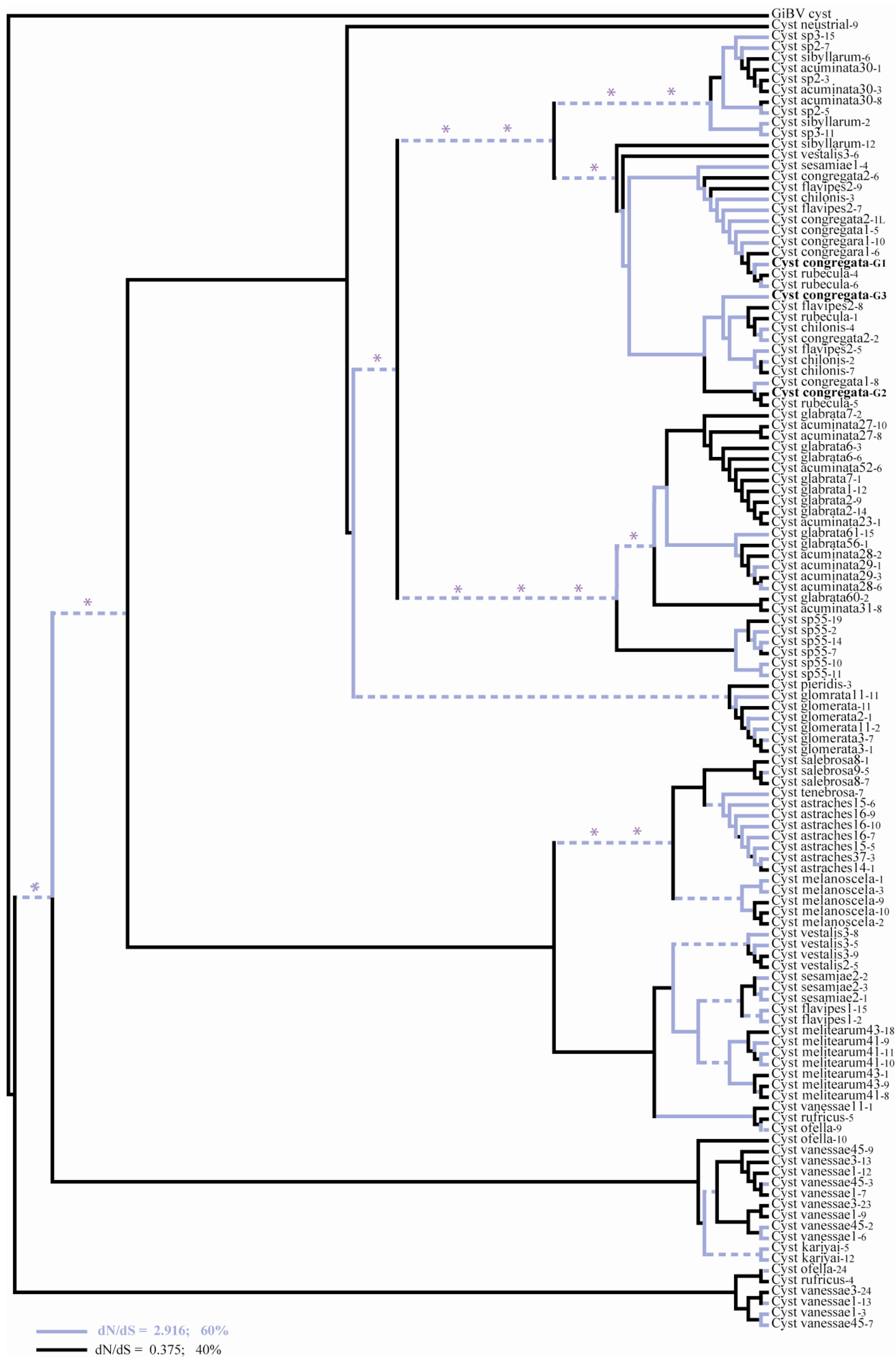


Figure 21 Viral cystatin lineages evolving under positive selection.
Branches significantly evolving under positive selection are in purple. 3 types of branches are mentioned: black line indicate branches corresponding to cystatin copy evolution after wasp speciation, dashed line indicate cystatin lineage 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 polydnaviruses, 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 polydnaviruses, 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 occurres 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 parasite 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 develope 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 bracovirus 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 bracovirus 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 bracovirus 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 Cα) présentent les mêmes patrons d'évolution que les cystatines ; les gènes isolés sont très peu divergents, ils 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 des 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 IkB 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).

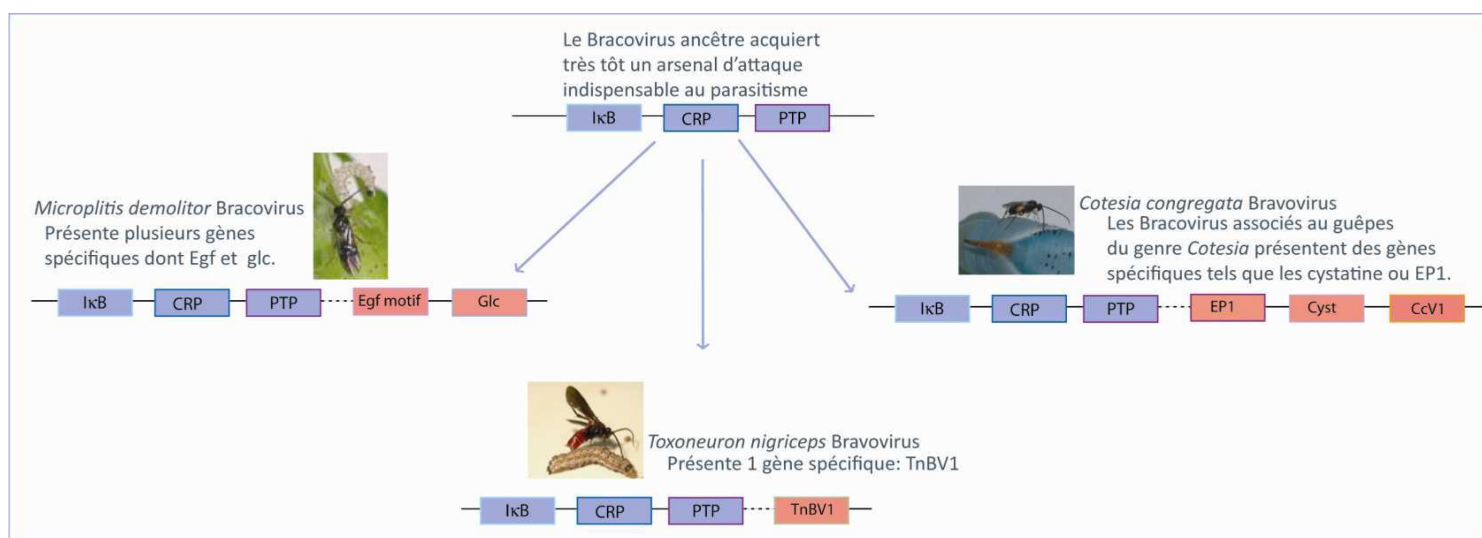


Figure 22 Certains gènes de virulence sont très conservés au sein des Bracovirus (IkB, PTP, CRP) alors que d'autres facteurs présentent une distribution étroite et sont spécifiques de certains Bracovirus.

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 ciblaient 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 tissues 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 Photorhabdus, 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 bacoluvirus) (Tableau 11).

Table 11 Perspectives expérimentales pour identifier le rôle des protéases à cystéine ou 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 transcriptionelle 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 proteine 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 dominent. 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 evaluer 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 évènements de diversification des cystatines et des guêpes, nous pourrons 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 entrainent 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.

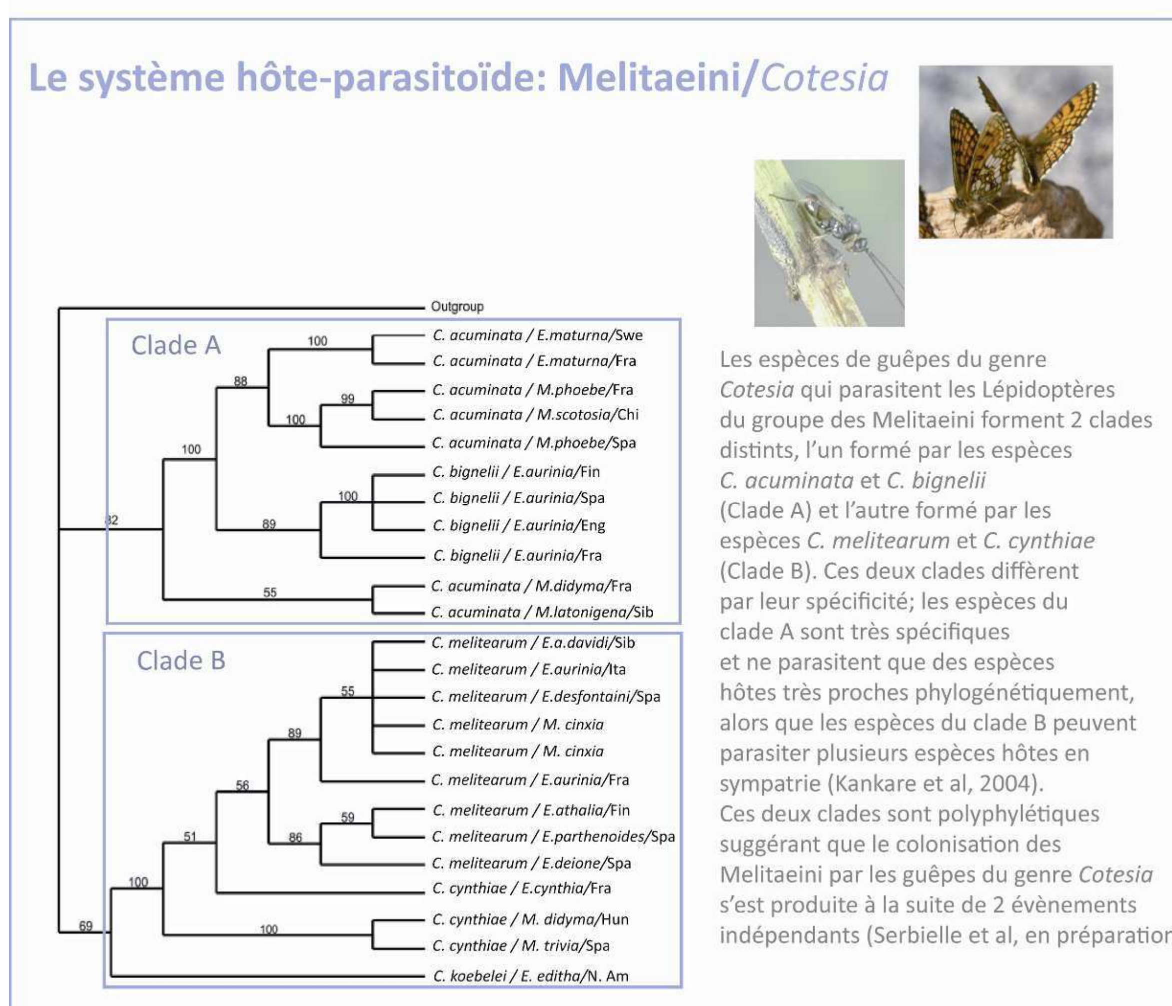
Aucune étude ne permet encore de mettre en évidence cette coévolution, cependant, comme cela a été observé chez *L. boulardi*, il est aussi possible d'observer chez les guêpes porteuses de PDV des variations de virulence. En effet, il existe deux souches de *C. sesamiae* dont l'une dite virulente est 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 ou 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

évènements 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.



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

References bibliographiques

Abrahamson, M., Alvarez-Fernandez, M., Nathanson, C-M., 2003. Cystatins. Biochemical society 70, 179-199.

Abrahamson, M., Ritonja, A., Brown, M.A., Grubb, A., Machleidt, W., Barrett, A.J., 1987. Identification of the probable inhibitory reactive sites of the cysteine proteinase inhibitors human cystatin C and chicken cystatin. J. Biol. Chem. 9688-9694.

Ackermann, M., Doebeli, M., 2004. Evolution and niche width and adaptive diversification. Evolution 58, 2599-2612.

Alvarez-Fernandez, M., Liang, Y.-H., Abrahamson, M., Su, X.-D., 2005. Crystal structure of human cystatin D, a cysteine peptidase inhibitor with restricted inhibition profile. J. Biol. Chem. 280, 18221-18228.

Andersen, J.N., Mortensen, O.H., Peters, G.H., Drake, P.G., Iversen, L.F., Olsen, O.H., Jansen, P.G., Andersen, H.S., Tonks, N.K., Moller, N.P.H., 2001. Structural and evolutionary relationships among protein tyrosine phosphatase domains. Mol. Cell. Biol. 21, 7117-7136.

Apidianakis, Y., Mindrinos, M.N., Xiao, W., Lau, G.W., Baldini, R.L., Davis, R.W., Rahme, L.G., 2005. Profiling early infection responses: *Pseudomonas aeruginosa* eludes host defenses by suppressing antimicrobial peptide gene expression. Proc. Natl. Acad. Sci. U. S. A. 102, 2573-2578.

Arai, S., Matsumoto, I., Emori, Y., Abe, K., 2002. Plant seed cystatins and their target enzymes of endogenous and exogenous origin. J. Agric. Food Chem. 50, 6612-6617.

Arguello, J.R., Chen, Y., Yang, S., Wang, W., Long, M., 2006. Origination of an X-linked testes chimeric gene by illegitimate recombination in *Drosophila*. Plos Genetics 2, e77.

Arisue, N., Hirai, M., Arai, M., Matsuoka, H., Horii, T., 2007. Phylogeny and evolution of the SERA multigene family in the genus *Plasmodium*. J. Mol. Evol. 65, 82-91.

Asagiri, M., Hirai, T., Kunigami, T., Kamano, S., Gober, H.-J., Okamoto, K., Nishikawa, K., Latz, E., Golenbock, D.T., Aoki, K., et al., 2008. Cathepsin K-dependent Toll-Like receptor 9 signaling revealed in experimental arthritis. Science 319, 624-627.

Asgari, S., Hellers, M., Schmidt, O., 1996. Host haemocyte inactivation by an insect parasitoid: transient expression of a polydnavirus gene. J. Gen. Virol. 77, 2653-2662.

Assfalg-Machleidt, I., Rothe, G., Klingel, S., Banati, R., Mangel, W., Valet, G., Machleidt, W., 1992. Membrance permeable fluorogenic rhodamine substrates for selective determination of cathepsin L. Biol. Chem. 373, 433-440.

Attardo, G.M., Strickler-Dinglasan, P., Perkin, S.A.H., Caler, E., Bonaldo, M.F., Soares, M.B., El-Sayeed, N., Aksoy, S., 2006. Analysis of fat body transcriptome from the adult tsetse fly, *Glossina morsitans morsitans*. Insect Mol. Biol. 15, 411-424.

Auerswald, E.A., Nagler, D.K., Assfalg-Machleidt, I., Stubbs, M.T., Machleidt, W., Fritz, H., 1995. Hairpin loop mutations of chicken cystatin have different effects on the inhibition of cathepsin B, cathepsin L and papain. FEBS Lett. 361, 179-184.

Axtner, J., Sommer, S., 2007. Gene duplication, allelic diversity, selection processes and adaptive value of MHC class II DRB genes of the bank vole, *Clethrionomys glareolus*. Immunogenetics 59, 417-426.

Babushok, D., Ostertag, E., Kazazian, H., 2007. Current topics in genome evolution: Molecular mechanisms of new gene formation. Cell. Mol. Life Sci. 64, 542-554.

Bailey, J.A., Eichler, E.E., 2006. Primate segmental duplications: crucibles of evolution, diversity and disease. Nat. Rev. Genet. 7, 552-564.

Banks, J.C., Whitfield, J.B., 2006. Dissecting the ancient rapid radiation of Microgastrinae wasp genera using additional nuclear genes. Mol. Phylogenet. Evol. 41, 690-703.

Beck, M., Strand, M.R., 2003. RNA interference silences *Microplitis demolitor* bracovirus genes and implicates glc1.8 in disruption of adhesion in infected host cells. Virology 314, 521-535.

Beck, M., Strand, M.R., 2005. Glc1.8 from *Microplitis demolitor* bracovirus induces a loss of adhesion and phagocytosis in insect high five and S2 cells. J. Virol. 79, 1861-1870.

Beck, M.H., Inman, R.B., Strand, M.R., 2007. *Microplitis demolitor* bracovirus genome segments vary in abundance and are individually packaged in virions. Virology 359, 179-189.

Beckage, N.E., 1998. Modulation of immune responses to parasitoids by polydnaviruses. Parasitology 116, 57-64.

Beckage, N.E., Alleyne, M., 1997. Parasitism-induced effects on host growth and metabolic efficiency in tobacco hornworm larvae parasitized by *Cotesia congregata*. J. Insect Physiol. 43, 407-424.

Beckage, N.E., Gelman, D.B., 2004. Wasp parasitoid disrutpion of host development: implications for new biologically based strategies of pest control. Annu. Rev. Entomol. 49, 299-330.

Beckage, N.E., Tan, F.T., Schleifer, K.W., Lane, R.D., Cherubin, L.L., 1994. Characterization and biological effects of *Cotesia congregata* Polydnavirus on host larvae of the tobacco hornworm, *Manduca sexta*. Arch. Insect Biochem. Physiol. 26, 165-195.

Belle, E., Beckage, N.E., Rousselet, J., Poirie, M., Lemeunier, F., Drezen, J.-M., 2002. Visualization of polydnavirus sequences in a parasitoid wasp chromosome. J. Virol. 76, 5793-5796.

Bergthorsson, U., Andersson, D.I., Roth, J.R., 2007. Ohno's dilemma: Evolution of new genes under continuous selection. Proc. Natl. Acad. Sci. U. S. A 104, 17004-17009.

Bézier, A., Herbinière, J., Serbielle, C., Lesobre, J., Wincker, P., Huguet, E., Drezen, J.M., 2007. Bracovirus gene products are highly divergent from insects protein. Arch. Insect Biochem. Physiol. 66, 172-187.

Bode, W., Engh, R., Musil, D., Thiele, U., Huber, R., Karshikov, A., Brzin, J., Kos, J., Turk, V., 1988. The 2.0 A X-ray crystal structure of chicken egg white cystatin and its impossible mode of interaction with cysteine proteinases. The EMBO journal 7, 2593-2599.

Brix, K., Dunkhorst, A., Mayer, K., Jordans, S., 2008. Cysteine cathepsins:Cellular roadmad to different functions. Biochimie 90, 194-207.

Byun-McKay, S.A., Geeta, R., 2007. Protein subcellular relocalization: a new perspective on the origin of novel genes. Trends Ecol. Evol. 22, 338-344.

Case, D.A., Cheatham, T.E., Darden, T., Gohlke, H., Luo, R., Merz, K.M., Onufriev, A., Simmerling, C., Wang, B., Woods, R.J., 2005. The Amber biomolecular simulation programs. J. Comput. Chem. 26, 1668-1688.

Chandran, K., Sullivan, N.J., Felbor, U., Whelan, S.P., Cunningham, J.M., 2005. Endosomal Proteolysis of the Ebola Virus Glycoprotein Is Necessary for Infection. Science 308, 1643-1645.

Chen, X., Li, S., Aksoy, S., 1999. Concordant evolution of a symbiont with its host insect species: molecular phylogeny of genus *Glossina* and its bacteriome-associated endosymbiont, *Wigglesworthia glossinidia*. J. Mol. Evol. 48, 49-58.

Cherry, S., Perrimon, N., 2004. Entry is a rate-limiting step for viral infection in a Drosophila melanogaster model of pathogenesis. Nat. Immunol. 5, 81-87.

Cho, W.-L., Tsao, S.-M., Hays, A.R., Walter, R., Chen, J.-S., Snigirevskaya, E.S., Raikhel, A.S., 1999. Mosquito cathepsin B-like protease involved in embryonic degradation of vitellin is produced as a latent extraovarian precursor. J. Biol. Chem. 274, 13311-13321.

Choi, J.Y., Roh, J.Y., Kang, J.N., Shim, H.J., Woo, S.D., Jin, B.R., Li, M.S., Je, Y.H., 2005. Genomic segments cloning and analysis of *Cotesia plutellae* polydnavirus using plasmid capture system. Biochem. Biophys. Res. Commun. 332, 487-493.

Clark, M.A., Moran, N.A., Baumann, P., Wernegreen, J.J., 2000. Cospeciation between bacterial endosymbionts (*Buchnera*) and a recent radiation of aphids (*Uroleucon*) and pitfalls of testing for phylogenetic congruence. Evolution 54, 517-525.

Combes, C., 2001. Parasitism: The ecology and evolution of intimate interactions. The unviversity of Chicago press.

Cornel, W.D., Cieplak, P., Bayly, C., Gould, I.R., Merz, K.M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J., Kollman, P.A., 1995. A second generation force field for the simulation of proteins, nucleic acids and organic molecules. J. Am. Chem. Soc. 117, 5179-5197.

Cygler, M., Sivaraman, J., Grochulski, P., Coulombe, R., Storer, A.C., Morts, J.S., 1996. Structure ot rat procathepsin B: model for inhibition of cysteine protease activity by the proregion. Structure 15, 405-416.

Dainichi, T., Maekawa, Y., Ishii, K., Zhang, T., Nashed, B.F., Sakai, T., Takashima, M., Himeno, K., 2001. Nippocystatin, a cysteine protease inhibitor from *Nippostrongylus brasiliensis*, inhibits antigen processing and modulates antigen-specific immune response. Infect. Immun. 69, 7380-7386.

Dale, C., Moran, N.A., 2006. Molecular interactions between bacterial symbionts and their hosts. Cell 126, 453-465.

Takiya, D.M., Tran, P.L., Dietrich, C.H., Moran, N.A., 2006. Co-cladogenesis spanning three phyla: leafhoppers (Insecta: Hemiptera: Cicadellidae) and their dual bacterial symbionts. Mol. Ecol. 15, 4175-4191.

Darbo, E., Danchin, E., Mc Dermott, M., Pontarotti, P., 2008. Evolution of major histocompatibility complex by "en bloc" duplication before mammalian radiation. Immunogenetics 60, 423-438.

Darden, T., York, D., Pedersen, L., 1993. Particle mesh Ewald: An N [center-dot] log(N) method for Ewald sums in large systems. J. Chem. Phys. 98, 10089-10092.

Dawkins, R., 1982. The extended phenotype. Oxford university press, New York.

De Gregorio, E., Spellman, P.T., Rubin, G.M., Lemaitre, B., 2001. Genome-wide analysis of the Drosphila immune response by using oligonucleotide microarrays. Proc. Natl. Acad. Sci. U. S. A. 98, 12590-12595.

Dedeine, F., Vavre, F., Fleury, F.d.r., Loppin, B., Hochberg, M.E., Boulétreau, M., 2001. Removing symbiotic Wolbachia bacteria specifically inhibits oogenesis in a parasitic wasp. Proc. Natl. Acad. Sci. U. S. A. 98, 6247-6252.

Des Marais, D.L., Rausher, M.D., 2008. Escape from adaptive conflict after duplication in an anthocyanin pathway gene. Nature, 762-765.

Desjardins, C., Gundersen-Rindal, D., Hostetler, J., Tallon, L., Fuester, R., Schatz, M., Pedroni, M., Fadrosh, D., Haas, B., Toms, B., et al., 2007. Structure and evolution of a proviral locus of *Glyptapanteles indiensis* bracovirus. BMC Microbiol. 7, 1-17.

Dickinson, D.P., 2002. Cysteine peptidases of mammals; their biological roles and potential effects in the oral cavity and other tissues in health and disease. Crit. Rev. Oral Biol. Med. 13, 238-275.

Dimopoulos, G., Christophides, G.K., Meister, S., Schultz, J.r., White, K.P., Barillas-Mury, C., Kafatos, F.C., 2002. Genome expression analysis of *Anopheles gambiae*: Responses to injury, bacterial challenge, and malaria infection. Proc. Natl. Acad. Sci. U. S. A. 99, 8814-8819.

Dodds, P., Lawrence, G.J., Catanzariti, A.-M., Teh, T., Wang, C.-I.A., Ayliffe, M.A., Ellis, J., 2006. Direct protein interaction underlies gene-for-gene specificity and coevolution of the flax resistance genes and flax rust avirulence genes. Proc. Natl. Acad. Sci. U. S. A. 103, 8888-8893.

Drummond, A.J., Ho, S.Y., Phillips, M.J., Rambaut, A., 2006. BEAST 1.3
Relaxed phylogenetics and dating with confidence. BMC Evol. Biol. p. e88.

Dubuffet, A., Dupas, S., Frey, F., Drezen, J.M., Poirie, M., Carton, Y., 2006. Genetic interactions between the parasitoid wasp *Leptopilina boulardi* and its *Drosophila* hosts. Heredity 98, 21-27.

Dupas, S., Carton, Y., Poirie, M., 2003a. Genetic dimension of the coevolution of virulence-resistance in *Drosophila* - parasitoid wasp relationships. Heredity 90, 84-89.

Dupas, S., Gitau, C.W., Branca, A., Le Ru, B.P., Silvain, J.-F., 2008. Evolution of a polydnavirus gene in relation to parasitoid host species immune resistance. J. Hered. 99, 204-211.

Dupas, S., Matthew, W.T., Bruce, A.W., 2003b. Diversifying selection in a parasitoid's symbiotic virus among genes involved in inhibiting host immunity. Immunogenetics 55, 351-361.

Dupuy, C., Huguet, E., Cattolico, L., Drezen, J.M., 2006. Vers la compréhension de l'histoire évolutive des polydnavirus. Virologie 10, 109-118.

Ebert, D.H., Deussing, J., Peters, C., Dermody, T.S., 2002. Cathepsin L and Cathepsin B mediate reovirus disassembly in murine fibroblast cells. J. Biol. Chem. 277, 24609-24617.

Emes, R.D., Yang, Z., 2008. Duplicated paralogous subject to positive selection in the genome of *Trypanosoma brucei*. PLoS ONE 3, e2295.

Espagne, E., Douris, V., Lalmanach, G., Provost, B., Cattolico, L., Lesobre, J., Kurata, S., Iatrou, K., Drezen, J.M., Huguet, E., 2005. A virus essential for insect host-parasite interactions encodes cystatins. J. Virol. 79, 9765-9776.

Espagne, E., Dupuy, C., Huguet, E., Cattolico, L., Provost, B., Martins, N., Poirie, M., Periquet, G., Drezen, J.M., 2004. Genome sequence of a polydnavirus: insights into symbiotic virus evolution. Science 306, 286-289.

Falabella, P., Caccialupi, P., Varricchio, P., Malva, C., Pennacchio, F., 2006. Protein tyrosine phosphatases of *Toxoneuron nigriceps* Bracovirus as potential disrupters of host prothoracic gland function. Arch. Insect Biochem. Physiol. 61, 157-169.

Falabella, P., Varricchio, P., Provost, B., Espagne, E., Ferrarese, R., Grimaldi, A., De Eguileor, M., Fimiani, G., Ursini, M.V., Malva, C., et al., 2007. Characterization of the IkB-like gene family in polydnaviruses associated with wasps belonging to different Braconid subfamilies. J. Gen. Virol. 88, 92-104.

Fan, C., Chen, Y., Long, M., 2008. Recurrent tandem gene duplication gave rise to functionally divergent genes in *Drosophila*. Mol. Biol. Evol. 25, 1451-1458.

Fenn, K., Blaxter, M., 2006. Wolbachia genomes: revealing the biology of parasitism and mutualism. Trends Parasitol. 22, 60-65.

Fleming, J.G.W., Krell, P., 1993. Polydnavirus genome organization. In: Beckage, N.E., Thompson, S., Federici, B. (Eds.), Parasites and pathogens of insects Academic press, New York, pp. 189-225.

Fleming, J.G.W., Summers, M.D., 1991. Polydnavirus DNA is integrated in the DNA of its parasitoid wasp host. Proc. Natl. Acad. Sci. U. S. A. 88, 9770-9774.

Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y.-l., Postlethwait, J., 1999. Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151, 1531-1545.

Foster, J., Ganatra, M., Kamal, I., Ware, J., Makarova, K., Ivanova, N., Bhattacharyya, A., Kapatral, V., Kumar, S., Posfai, J., et al., 2005. The Wolbachia genome of *Brugia malayi*: endosymbiont evolution within a human pathogenic nematode. PLoS Biol. 3, e121.

Francino, M.P., 2005. An adaptive radiation model for the origin of new gene functions. Nat. Genet. 37, 573-578.

Fujimoto, Y., Kobayashi, A., Kurata, S., Natori, S., 1999. Two subunits of the insect 26/29-kDa proteinase are probably derived from a common precursor protein. pp. 566-573.

Fujiwara, Y., Asogawa, M., 2001. Prediction of subcellular localizations using amino acids composition and order. Genome Bioinformatics 12, 103-102.

Ganko, E.W., Meyers, B.C., Vision, T.J., 2007. Divergence in expression between duplicated genes in *Arabidopsis*. Mol. Biol. Evol., 2298-2309.

Gettins, P.G.W., 2002. Serpin structure, mechanism, and function. Chem. rev. 102, 4751-4803.

Ghedin, E., Wang, S., Spiro, D., Caler, E., Zhao, Q., Crabtree, J., Allen, J.E., Delcher, A.L., Guiliano, D.B., Miranda-Saavedra, D., et al., 2007. Draft genome of the filarial nematode parasite *Brugia malayi*. Science 317, 1756-1760.

Gladyshev, E.A., Meselson, M., Arkhipova, I.R., 2008. Massive horizontal gene transfer in Bdelloid rotifers. Science 320, 1210-1213.

Goldman, N., Yang, Z., 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol. Biol. Evol. 11, 725-736.

Goulet, M.-C., Dallaire, C., Vaillancourt, L.-P., Khalf, M., Badri, A.M., Preradov, A., Duceppe, M.-O., Goulet, C., Cloutier, C., Michaud, D., 2008. Tailoring the specificity of a plant cystatin toward herbivorous insect digestive cysteine proteases by single mutations at positively selected amino acid sites. Plant Physiol. 146, 1010-1019.

Gregory, W.F., Maizels, R.M., 2008. Cystatins from filarial parasites: Evolution, adaptation and function in the host-parasite relationship. Int. J. Biochem. Cell Biol. 40, 1389-1398.

Gruber, A., Stettler, P., Heiniger, P., Schumperli, D., Lanzrein, B., 1996. Polydnavirus DNA of the braconid wasp *Chelonus inanitus* is integrated in the wasp's genome and excised only in later pupal and adult stages of the female. J. Gen. Virol. 77, 2873-2879.

Guindon, S., Gascuel, O., 2003. A simple, fast and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52, 696-704.

Gundersen-Rindal, D.E., Pedroni, M.J., 2006. Characterization and transcriptional analysis of protein tyrosine phosphatase genes and an ankyrin repeat gene of the parasitoid *Glyptapanteles indiensis* polydnavirus in the parasitized host. J. Gen. Virol. 311-322.

Hall, A., Håkansson, K., Mason, R.W., Grubb, A., Abrahamson, M., 1995. Structural basis for the biological specificity of cystatin C. JBC 270, 5115-5121.

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids. Symp. Ser. 41, 95-98.

Harrison, R.L., Bonning, B.C., 2001. Use of proteases to improve the insecticidal activity of Baculoviruses. Biol. Control 20, 199-209.

Harwood, S.H., Grosovsky, A.J., Cowles, E.A., Davis, J.W., Beckage, N.E., 1994. An abundantly expressed hemolymph glycoprotein isolated from newly parasitized *Manduca sexta* larvae is a polydnavirus gene product. Virology 205, 381-392.

Harwood, S.H., McElfresh, J.S., Nguyen, A., Conlan, C.A., Beckage, N.E., 1998. Production of early expressed parasitism-specific proteins in alternate Sphingid hosts of the Braconid wasp *Cotesia congregata*. J. Invertebr. Pathol. 71, 271-279.

Hegedus, D., O'Grady, M., Chamankhah, M., Baldwin, D., Gleddie, S., Braun, L., Erlandson, M., 2002. Changes in cysteine protease activity and localization during midgut metamorphosis in the crucifer root maggot (*Delia radicum*). Insect Biochem. Mol. Biol. 32, 1585-1596.

Hoffmann, F.G., Opazo, J.C., Storz, J.F., 2008. Rapid rates of lineage-specific gene duplication and deletion in the {alpha}-globin gene family. Mol. Biol. Evol. 25, 591-602.

Hoffmann, J.A., 2003. The immune response of *Drosophila*. Nature 42, 33-38.

Homma, K., Kurata, S., Natori, S., 1994. Purification, characterization, and cDNA cloning of procathepsin L from the culture medium of NIH-Sape-4, an embryonic cell line of *Sarcophaga peregrina* (flesh fly), and its involvement in the differentiation of imaginal discs. J. Biol. Chem. 269, 15258-15264.

Hooper, S., Berg, O., 2003a. Duplication is more common among laterally transferred genes than among indigenous genes. Genome Biol. 4, R48.

Hooper, S.D., Berg, O.G., 2003b. On the nature of gene innovation: Duplication patterns in microbial genomes. Mol. Biol. Evol. 20, 945-954.

Hosokawa, T., Kikuchi, Y., Nikoh, N., Shimada, M., Fukatsu, T., 2006. Strict host-symbiont cospeciation and reductive genome evolution in insect gut bacteria. PLoS Biol. 4, e337.

Hosokawa, T., Kikuchi, Y., Shimada, M., Fukatsu, T., 2007. Obligate symbiont involved in pest status of host insect. Proc.R. Soc. B, 274, 1979-1984.

Hotopp, J.C.D., Clark, M.E., Oliveira, D.C.S.G., Foster, J.M., Fischer, P., Torres, M.C.M., Giebel, J.D., Kumar, N., Ishmael, N., Wang, S., et al., 2007. Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. Science 317, 1753-1756.

Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics. 754 - 755.

Hughes, A.L., 1994. The Evolution of functionally novel proteins after gene duplication. Proc.R. Soc. B 256, 119-124.

Hugues, A.L., 1994. The evolution of functionally novel proteins after gene duplication. Proc.R. Soc. B 256, 119-124.

Ibrahim, A.M.A., Choi, J.Y., Je, Y.H., Kim, Y., 2007. Protein tyrosine phosphatases encoded in *Cotesia plutellae* bracovirus: Sequence analysis, expression profile, and a possible biological role in host immunosuppression. Dev. Comp. Immunol. 31, 978-990.

Ibrahim, A.M.A., Kim, Y., 2008. Transient expression of protein tyrosine phosphatases encoded in *Cotesia plutellae* bracovirus inhibits insect cellular immune responses. Naturwissenschaften 95, 25-32.

Janowski, R., Kozak, M., Jankowska, E., Grzonka, Z., Abrahamson, M., Jaskolski, M., 2001. Human cystatin C, an amyloidogenic protein, dimerizes through three-dimensional domain swapping. Nat. Struct. Biol. 8, 316-320.

Janson, E.M., Stireman, J.O., Singer, M.S., Abbot, P., Mauricio, R., 2008. Phytophagous insect-microbe mutualisms and adaptive evolutionary diversification. Evolution 62.

Jenko, S., Dolenc, I., Guncar, G., Dobersek, A., Podobnik, M., Turk, D., 2003. Crystal Structure of Stefin A in Complex with Cathepsin H: N-terminal Residues of Inhibitors can Adapt to the Active Sites of Endo- and Exopeptidases. J. Mol. Biol. 326, 875-885.

Jiang, H., Wang, Y., Yu, X.-Q., Zhu, Y., Kanost, M., 2003. Prophenoloxidase-activating proteinase-3 (PAP-3) from *Manduca sexta* hemolymph: a clip-domain serine proteinase regulated by serpin-1J and serine proteinase homologs. Insect Biochem. Mol. Biol. 33, 1049-1060.

Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., Klein, M.L., 1983. Comparison of simple potential functions for simulating liquid water. The Journal of Chem. Phys. 79, 926-935.

Kamita, S.G., Nagasaka, K., Chua, J.W., Shimada, T., Mita, K., Kobayashi, M., Maeda, S., Hammock, B.D., 2004. A baculovirus-encoded protein tyrosine phosphatase gene induces enhanced locomotory activity in a lepidopteran host. Proc. Natl. Acad. Sci. U. S. A. 102, 2584-2589.

Kankare, M., Shaw, M.R., 2004. Molecular phylogeny of *Cotesia* Cameron, 1891 (Insecta: Hymenoptera: Braconidae: Microgastrinae) parasitoids associated with Melitaeini butterflies (Insecta: Lepidoptera: Nymphalidae: Melitaeini). Mol. Phylogenet. Evol. 32, 207-220.

Karrer, K.M., Peiffer, S.L., DiTomas, M.E., 1993. Two distinct gene subfamilies within the family of cysteine protease genes. Proc. Natl. Acad. Sci. U. S. A. 90, 3063-3067.

Katju, V., Lynch, M., 2003. The structure and early evolution of recently arisen gene duplicates in the *Caenorhabditis elegans* genome. Genetics 165, 1793-1803.

Kiggundu, A., Goulet, M.-C., Goulet, C., Dubuc, J.-F., Rivard, D., Benchabane, M., Pepin, G., Vyver, C.v.d., Kunert, K., Michaud, D., 2006. Modulating the proteinase inhibitory profile of a plant cystatin by single mutations at positively selected amino acid sites. The Plant Journal 48, 403-413.

Kikuchi, Y., Hosokawa, T., Fukatsu, T., 2007. Insect-microbe mutualism without vertical transmission: a stinkbug acquires a beneficial gut symbiont from the environment every generation. Appl. Environ. Microbiol. 73, 4308-4316.

Kim, H.J., Je, H.J., Park, S.Y., Lee, I.H., Jin, B.R., Yun, H.K., Yun, C.Y., Han, Y.S., Kang, Y.J., Seo, S.J., 2004. Immune activation of apolipophorin-III and its distribution in hemocyte from *Hyphantria cunea*. Insect Biochem. Mol. Biol. 34, 1011-1023.

Kim, Y., 2005. Identification of host translation inhibitory factor of *Campoletis sonorensis* Icnhovirus on the Tobacco Budworm, *Heliotis virescens*. Arch. Insect Biochem. Physiol. 59, 230-244.

Kondo, N., Nikoh, N., Ijichi, N., Shimada, M., Fukatsu, T., 2002. Genome fragment of Wolbachia endosymbiont transferred to X chromosome of host insect. Proc. Natl. Acad. Sci. U. S. A. 99, 14280-14285.

Koo, Y.D., Ahn, J.E., Salzman, R.A., Moon, J., Chi, Y.H., Yun, D.J., Lee, S.Y., Koiwa, H., Zhu-Salzman, K., 2008. Functional expression of an insect cathepsin B-like counter-defence protein. Insect Mol. Biol. 17, 235-245.

Krupa, J.C., Hasnain, S., Nägler, D.K., Ménard, R., Mort, J.S., 2002. S2' substrate specificity and the role of His110 and His111 in the exopeptidase activity of human cathepsin B. Biochem. J. 361, 613-619.

Kumar, S., Tamura, K., Nei, M., 2004. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Brief. Bioinform. 5, 150-163.

Kurata, S., Saito, H., Natori, S., 1992. Purification of a 29-kDa hemocyte proteinase of *Sarcophaga peregrina*. Eur. J. Biochem. 204, 911-914.

Kurland, C.G., Andersson, S.G.E., 2000. Origin and evolution of the mitochondrial proteome. Microbiol. Mol. Biol. Rev. 64, 786-820.

Labrosse, C., Eslin, P., Doury, G., Drezen, J.M., Poirié, M., 2005a. Haemocyte changes in *D. melanogaster* in response to long gland components of the parasitoid wasp *Leptopilina boulardi*: a Rho-GAP protein as an important factor. J. Insect Physiol. 51, 161-170.

Labrosse, C., Stasiak, K., Lesobre, J., Grangeia, A., Huguet, E., Drezen, J.M., Poirie, M., 2005b. A RhoGAP protein as a main immune suppressive factor in the *Leptopilina boulardi* (Hymenoptera, Figitidae)-*Drosophila melanogaster* interaction. Insect Biochem. Mol. Biol. 35, 93-103.

Lapointe, R., Tanaka, K., Barney, W.E., Whitfield, J.B., Banks, J.C., Béliveau, C., Stoltz, D., Webb, B., Cusson, M., 2007. Genomic and morphological features of a Banchine polydnavirus: comparison with Bracovirus and Ichnovirus. J. Virol. 81, 6491-6501.

Lapointe, R., Wilson, R., Vilaplana, L., O'Reilly, D.R., Falabella, P., Douris, V., Bernier-Cardou, M., Pennacchio, F., Iatrou, K., Malva, C., et al., 2005. Expression of a *Toxoneuron nigriceps* polydnavirus-encoded protein causes apoptosis-like programmed cell death in lepidopteran insect cells. J. Gen. Virol. 86, 963-971.

Laskowski, R.A., Moss, D.S., Thornton, J.M., 1993. Main-chain bond lengths and bond angles in protein structures. J. Mol. Biol. 231, 1049-1067.

Lazzaro, B.P., 2008. Natural selection on the *Drosophila* antimicrobial immune system. Curr. Opin. Microbiol. 11, 284-289.

Lecaille, F., Kaleta, J., Brömme, D., 2002. Human and parasitic papain-like cysteine proteases:their role in physiology and pathology and recent developments in inhibitor design. Chem. Rev. 102, 4459-4488.

Lemaitre, B., Hoffmann, J., 2007. The host defense of *Drosophila melanogaster*. Annu. Rev. Immunol. 25, 697-743.

Lerat, E., Daubin, V., Moran, N.A., 2003. From gene trees to organismal phylogeny in prokaryotes: the case of the gamma-Proteobacteria. PLoS Biol. 1, E19.

Levy, F., Rabel, D., Charlet, M., Bulet, P., Hoffmann, J.A., Ehret-Sabatier, L., 2004. Peptidomic and proteomic analyses of the systemic immune response of *Drosophila*. Biochimie 86, 607-616.

Li, H., Tang, H., Sivakumar, S., Philip, J., Harrison, R.L., Gatehouse, J.A., Bonning, B., 2008. Insectividal activity of a basement membrane-degrading protease against *Heliothis virescens* (Fabricius) and *Acyrthosiphon pisum* (Harris). J. Insect Physiol. 54, 777-789.

Li, X., Webb, B., 1994. Apparent functional role for a cysteine-rich polydnavirus protein in suppression of the insect cellular immune response. J. Virol. 68, 7482-7489.

Liebert, C.A., Hall, R.M., Summers, A., 1999. Transposon Tn21, flagship of the floating genome. Microbiol. Mol. Biol. Rev. 63, 507-522.

Liu, J., Shi, G.P., Zhang, W.Q., Zhang, G.R., Xu, W.H., 2006. Cathepsin L function in insect moulting: molecular cloning and functional analysis in cotton bollworm, *Helicoverpa armigera*. Insect Mol. Biol. 15, 823-834.

Liu, Z., Bos, J.I.B., Armstrong, M., Whisson, S.C., da Cunha, L., Torto-Alalibo, T., Win, J., Avrova, A.O., Wright, F., Birch, P.R.J., et al., 2005. Patterns of diversifying selection in the phytotoxin-like scr74 Gene Family of *Phytophthora infestans*. 22, 659-672.

Ljunggren, A., Redzynia, I., Alvarez-Fernandez, M., Abrahamson, M., Mort, J.S., Krupa, J.C., Jaskolski, M., Bujacz, G., 2007. Crystal structure of the parasite protease inhibitor chagasin in complex with a host target cysteine protease. J. Mol. Biol. 371, 137-153.

Machleidt, W., Thiele, U., Laber, B., Assfalg-Machleidt, I., Esterl, A., Wiegand, G., Kos, J., Turk, V., Bode, W., 1989. Mechanism of inhibition of papain by chicken egg white cystatin : Inhibition constants of N-terminally truncated forms and cyanogen bromide fragments of the inhibitor. FEBS Lett. 243, 234-238.

Maizels, R.M., Blaxter, M.L., Scott, A.L., 2001. Immunological genomics of *Brugia malayi*: filarial genes implicated in immune evasion and protective immunity. Parasite Immunol. 23, 327-344.

Mallet, F.o., Bouton, O., Prudhomme, S., Cheynet, V.r., Oriol, G., Bonnaud, B., Lucotte, G.r., Duret, L., Mandrand, B., 2004. The endogenous retroviral locus ERVWE1 is a bona fide gene involved in hominoid placental physiology. Proc. Natl. Acad. Sci. U. S. A. 101, 1731-1736.

Mardulyn, P., Whitfield, J.B., 1999. Phylogenetic signal in the COI, 16S, and 28S genes for inferring relationships among genera of Microgastrinae (Hymenoptera; Braconidae): evidence of a high diversification rate in this group of parasitoids. Mol. Phylogenet. Evol. 12, 282-294.

Marquez, L.M., Redman, R.S., Rodriguez, R.J., Roossinck, M.J., 2007. A virus in a fungus in a plant: three-way symbiosis required for thermal tolerance. Science 315, 513-515.

Martin, W., Rujan, T., Richly, E., Hansen, A., Cornelsen, S., Lins, T., Leister, D., Stoebe, B., Hasegawa, M., Penny, D., 2002. Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. Proc. Natl. Acad. Sci. U. S. A. 99, 12246-12251.

Mason, R.W., Sol-Church, K., Abrahamson, M., 1998. Amino acid substitutions in the N-terminal segment of cystatin C create selective protein inhibitors of lysosomal cysteine proteinases. Biochem. J. 330, 833-838.

Mason, W.R.M., 1981. The polyphyletic nature of *Apanteles* Foerster (Hymenoptera: Braconiade): a phylogeny and reclassification of Microgastrinae. Memoirs of the Entomological Society of Canada.

Matsui, N.M., Smith-Beckerman, D.M., Epstein, L.B., 1999. Staining of preparative 2-D gels. Coomassie blue and imidazole-zinc negative staining. Methods Mol. Biol. 112, 307-311.

McBride, C.S., 2007. Rapid evolution of smell and taste receptor genes during host specialization in Drosophila sechellia. Proc. Natl. Acad. Sci. U. S. A. 104, 4996-5001.

Michel-Salzat, A., Whitfield, J.B., 2004. Preliminary evolutionnary relationships within the parasitoids wasp genus *Cotesia* (Hymenoptera: Braconidae: Microgastrinae): combined analysis of four genes. Syst. Entomol. 29, 371-382.

Miyaji, T., Kouzuma, Y., Yaguchi, J., Matsumoto, R., Kanost, M.R., Kramer, K.J., Yonekura, M., 2007. Purification of a cysteine protease inhibitor from larval hemolymph of the

tobacco hornworm (*Manduca sexta*) and functional expression of the recombinant protein. Insect Biochem. Mol. Biol. 37, 960-968.

Mohamed, M.M., Sloane, B.F., 2006. Cysteine cathepsins: multifunctional enzymes in cancer. Nat. Rev. Cancer 6, 764-775.

Moran, N.A., Degnan, P.H., Santos, S.R., Dunbar, H.E., Ochman, H., 2005. The players in a mutualistic symbiosis: Insects, bacteria, viruses, and virulence genes. Proc. Natl. Acad. Sci. U. S. A. 102, 16919-16926.

Moreau, S.J.M., Cherqui, A., Doury, G., Dubois, F., Fourdrain, Y., Sabatier, L., Bulet, P., Saarela, J., Prévost, G., Giordanengo, P., 2004. Identification of an aspartylglucosaminidase-like protein in the venom of the parasitic wasp *Asobara tabida* (Hymenoptera: Braconidae). Insect Biochem. Mol. Biol. 34, 485-492.

Morrison, H.G., McArthur, A.G., Gillin, F.D., Aley, S.B., Adam, R.D., Olsen, G.J., Best, A.A., Cande, W.Z., Chen, F., Cipriano, M.J., et al., 2007. Genomic minimalism in the early diverging intestinal parasite *Giardia lamblia*. Science 317, 1921-1926.

Murphy, N., Banks, J.C., Whitfield, J.B., Austin, A.D., 2008. 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, 378-395.

Musil, D., Zucic, D., Turk, D., Engh, R.A., Mayr, I., Huber, R., Popovic, T., Turk, V., Towatari, T., Katunuma, N., et al., 1991. The refined 2.15 A X-ray cristal structure of human liver cathepsin B:the structural basis for its specificity. EMBO Journal 10, 2321-2330.

Nair, S., Nash, D., Sudimack, D., Jaidee, A., Barends, M., Uhlemann, A.-C., Krishna, S., Nosten, F., Anderson, T.J.C., 2007. Recurrent gene amplification and soft selective sweeps during evolution of multidrug resistance in malaria parasites. Mol. Biol. Evol. 24, 562-573.

Nakabachi, A., Yamashita, A., Toh, H., Ishikawa, H., Dunbar, H.E., Moran, N.A., Hattori, M., 2006. The 160-Kilobase genome of the bacterial endosymbiont *Carsonella*. Science 314, 259-260.

Nappi, A.J., Frey, F., Carton, Y., 2005. *Drosophila* serpin 27A is a likely target for immune suppression of the blood cell-mediated melanotic encapsulation response. J. Insect Physiol. 51, 197-205.

Nappi, A.J., Vass, E., 1998. Hydrogen peroxide production in immune-reactive *Drosophila melanogaster*. J. Parasitol. 84 1150-1157.

Nation, J.L., 2002. Hormone and development. Insect physiology and biochemistry. CRC Press, Boca Raton, pp. 119-155.

Nei, M., Rooney, A.P., 2005. Concerted and birth and death evolution families. Annu. Rev. Genet. 39, 121-152.

Nikoh, N., Tanaka, K., Shibata, F., Kondo, N., Hizume, M., Shimada, M., Fukatsu, T., 2008. Wolbachia genome integrated in an insect chromosome: Evolution and fate of laterally transferred endosymbiont genes. Genome Res. 18, 272-280.

Nusawardani, T., Ruberson, J.R., Obrycki, J.J., Bonning, B., 2005. Effects of a protease-expressing recombinant baculovirus insecticide on the parasitoid *Cotesia marginiventris* (Cresson). Biol. Control 46.

Obbard, D.J., Jiggins, F.M., Halligan, D.L., Little, T.J., 2006. Natural selection drives extremely rapid evolution in antiviral RNAi genes. Curr. Biol. 16, 580-585.

Ohno, S., 1970. Evolution by gene duplication.

Oliver, K.M., Russell, J.A., Moran, N.A., Hunter, M.S., 2003. Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. Proc. Natl. Acad. Sci. U. S. A. 100, 1803-1807.

Otto Schmidt, U.T.M.S., 2001. Innate immunity and its evasion and suppression by hymenopteran endoparasitoids. BioEssays 23, 344-351.

Paily, K., Kumar, B., Balaraman, K., 2007. Transferrin in the mosquito, *Culex quinquefasciatus* Say (Diptera: Culicidae), up-regulated upon infection and development of the

filarial parasite, *Wuchereria bancrofti* (Cobbold) (Spirurida: Onchocercidae). Parasitol. Res. 101, 325-330.

Palermo, C., Joyce, J.A., 2008. Cysteine cathepsin proteases as pharmacological targets in cancer. Trends Pharmacol. Sci. 29, 22-28.

Pancer, Z., Cooper, M.D., 2006. The evolution of adaptive immunity. Annu. Rev. Immunol. 24, 497-518.

Papanicolaou, A., Gebauer-Jung, S., Blaxter, M.L., McMilla, W.O., Jiggins, C.D., 2008. ButterflyBase: a platform for lepidopteran genomics Nucleic Acids Res. 36, D582-D587.

Paradis, E., Claude, J., Strimmer, K., 2004. APE: Analyses of phylogenetics and evolution in R language. Bioinformatics 20, 289-290.

Pech, L.L., Trudeau, D., Strand, M.R., 1995. Effects of basement membranes on the behavior of hemocytes frome *Pseudoplusia includens* (Lepidoptera; Noctuidae):development of an *in vitro* encapsulation assay. J. Insect Physiol. 41, 801-807.

Pennacchio, F., Strand, M.R., 2006. Evolution and developmental strategies in parasitic hymenoptera. Annu. Rev. Entomol. 51, 233-258.

Perez-Brocal, V., Gil, R., Ramos, S., Lamelas, A., Postigo, M., Michelena, J.M., Silva, F.J., Moya, A., Latorre, A., 2006. A small microbial genome: the end of a long symbiotic relationship? Science 314, 312-313.

Ponce, R., Hartl, D.L., 2006. The evolution of the novel Sdic gene cluster in *Drosophila melanogaster*. Gene 376, 174-183.

Pond, S.L.K., Frost, S.D.W., 2005. A genetic algorithm approach to detecting lineage-specific variation in selection pressure. Mol. Biol. Evol. 22, 478-485.

Pond, S.L.K., Frost, S.D.W., Muse, S.V., 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics 21, 676-679.

Posada, D., Crandall, K.A., 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14, 817-818.

Provost, B., Varricchio, P., Arana, E., Espagne, E., Falabella, P., Huguet, E., La Scaleia, R., Cattolico, L., Poirie, M., Malva, C., et al., 2004. Bracoviruses contain a large multigene family coding for protein tyrosine phosphatases. J. Virol. 78, 13090-13103.

Pruijssers, A.J., Strand, M.R., 2007. PTP-H2 and PTP-H3 from *Microplitis demolitor* Bracovirus localize to focal adhesions and are antiphagocytic in insect immune cells. J. Virol. 81, 1209-1219.

Qiu, Z., Hingley, S.T., Simmons, G., Yu, C., Das Sarma, J., Bates, P., Weiss, S.R., 2006. Endosomal proteolysis by cathepsins is necessary for murine coronavirus mouse hepatitis virus type 2 spike-mediated entry. J. Virol. 80, 5768-5776.

Rawlings, N.D., Tolle, D.P., Barrett, A.J., 2004. Evolutionary families of peptidase inhibitors. Biochem. J. 378, 705-716.

Redman, R.S., Sheehan, K.B., Stout, R.G., Rodriguez, R.J., Henson, J.M., 2002. Thermotolerance generated by Plant/Fungal symbiosis. Science 298, 1581-.

Renault, S., Stasiak, K., Federici, B., Bigot, Y., 2005. Commensal and mutualistic relationships of reoviruses with their parasitoid wasp hosts. J. Insect Physiol. 51, 137-148.

Riddiford, L.M., Hiruma, K., Zhou, X., Nelson, C.A., 2003. Insights into the molecular basis of the hormonal control of molting and metamorphosis from *Manduca sexta* and *Drosophila melanogaster*. Insect Biochem. Mol. Biol. 33, 1327-1338.

Rodríguez-Pérez, M.A., Dumpit, R.F., Lenz, J.M., Powell, E.N., Tam, S.Y., Beckage, N.E., 2005. Host refractoriness of the tobacco hornworm, *Manduca sexta*, to the braconid endoparasitoid *Cotesia flavipes*. Arch. Insect Biochem. Physiol. 60, 159-171.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572-1574.

Russell, J.A., Moran, N.A., 2005. Horizontal transfer of bacterial symbionts: heritability and fitness effects in a novel aphid host. Appl. Environ. Microbiol. 71, 7987-7994.

Ryckaert, J.P., Ciccotti, G., Berendsen, H.J.C., 1977. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J. Comput. Phys. 23 327-341.

Saito, H., Kurata, S., Natori, S., 1992. Purification and characterization of a hemocyte proteinase of *Sarcophaga*, possibly participating in elimination of foreign substances. FEBS 209, 939-944.

Schierack, P., Lucius, R., Sonnenburg, B., Schilling, K., Hartmann, S., 2003. Parasite-specific immunomodulatory functions of filarial cystatin. Infect. Immun. 71, 2422-2429.

Serbielle, C., Chowdhury, S., Pichon, S., Dupas, S., Lesobre, J., Purisima, E.O., Drezen, J.M., Huguet, E., 2008. Viral cystatin evolution and three-dimensional structure modelling: a case of directional selection acting on a viral protein involved ina host-parasitoid interaction. BMC Biology 6.

Shindo, T., Van Der Hoorn, R.A., 2008. Papain-like cysteine proteases: key players at molecular battlefields employed by both plants and their invaders. Mol. Plant Pathol. 9, 119-125.

Smith, M.A., Rodriguez, J.J., Whitfield, J.B., Deans, A.R., Janzen, D.H., Hallwachs, W., Hebert, P.D.N., 2008. Extreme diversity of tropical parasitoid wasps exposed by iterative integration of natural history, DNA barcoding, morphology, and collections. Proc. Natl. Acad. Sci. U. S. A. 105, 12359-12364.

Song, K.-H., Jung, M.-K., Eum, J.-H., Hwang, I.-C., Han, S.S., 2008. Proteomic analysis of parasitized *Plutella xylostella* larvae plasma. J. Insect Physiol. 54, 1271-1280.

Stahl, E.A., Bishop, J.G., 2000. Plant-pathogen arms races at the molecular level. Curr. Opin. Plant Biol. 3, 299-304.

Stoltz, D.B., 1990. Evidence for chromosomal transmission of polydnavirus DNA. J. Gen. Virol. 71, 1051-1056.

Stoltz, D.B., Krell, P., Summers, M.D., Vinson, S.B., 1984. Polydnaviridae-a proposed family of insect viruses with segmented, double-stranded, circular DNA genomes. Intervirology 21, 1-4.

Strand, M.R., Noda, T., 1991. Alterations in the haemocytes of *Pseudoplusia includens* after parasitism by *Microplitis demolitor*. J. Insect Physiol. 37, 839-850.

Strand, M.R., Pech, L.L., 1995a. Immunological basis for compatibility in parasitoid-host relationships. Annu. Rev. Entomol. 40, 31-56.

Strand, M.R., Pech, L.L., 1995b. *Microplitis demolitor* polydnavirus induces apoptosis of a specific haemocyte morphotype in *Pseudoplusia includens*. J. Gen. Virol. 76, 283-291.

Striepen, B., Pruijssers, A.J.P., Huang, J., Li, C., Gubbels, M.-J., Umejiego, N.N., Hedstrom, L., Kissinger, J.C., 2004. Gene transfer in the evolution of parasite nucleotide biosynthesis. Proc. Natl. Acad. Sci. U. S. A. 101, 3154-3159.

Stubbs, M.T., Laber, B., Bode, W., Huber, R., Jerala, R., Lenarcic, B., Turk, V., 1990. The refined 2.4 A X-ray crystal structure of recombinant human stefin B in complex with cysteine proteinase papain: a novel type proteinase inhibitor interaction. The EMBO journal 9, 1939-1947.

Suderman, R.J., Pruijssers, A.J., Strand, M.R., 2008. Protein tyrosine phosphatase-H2 from a polydnavirus induces apoptosis of insect cells. J. Gen. Virol. 89, 1411-1420.

Swanson, W.J., Nielsen, R., Yang, Q., 2003. Pervasive adaptive evolution in mammalian fertilization proteins. Mol. Biol. Evol. 20, 18-20.

Tanaka, K., Tsuzuki, S., Matsumoto, H., Hayakawa, Y., 2000. Expression of *Cotesia kariyai* polydnavirus genes in lepidopteran hemocytes and Sf9 cells. J. Insect Physiol. 49, 433-440.

Tang, H., Li, H., Meng Lei, S., Harrison, R.L., Bonning, B.C., 2007. Tissue specificity of a baculovirus-expressed, basement membrane-degrading protease in larvae of *Heliothis virescens*. Tissue and Cell 39, 431-443.

Templeton, T.J., 2007. Whole-genome natural histories of apicomplexan surface proteins. Trends Parasitol. 23, 205-212.

Templeton, T.J., Iyer, L.M., Anantharaman, V., Enomoto, S., Abrahante, J.E., Subramanian, G.M., Hoffman, S.L., Abrahamsen, M.S., Aravind, L., 2004. Comparative analysis of Apicomplexa and genomic diversity in eukaryotes. Genome Res. 14, 1686-1695.

Thoetkiattikul, H., Beck, M.H., Strand, M.R., 2005. Inhibitor IKB-like proteins from a polydnavirus inhibit NF-IkB activation and suppress the insect immune response. Proc. Natl. Acad. Sci. U. S. A. 102, 11426-11431.

Thompson, J.N., 1988. Variation in interspecific interactions. Annu. Rev. Ecol. Syst. 19, 65-87.

Tian, M., Win, J., Song, J., van der Hoorn, R., van der Knaap, E., Kamoun, S., 2007. A Phytophthora infestans Cystatin-like protein targets a novel tomato papain-like apoplastic protease. Plant Physiol. 143, 364-377.

Top, E.M., Springael, D., 2003. The role of mobile genetic elements in bacterial adaptation to xenobiotic organic compounds. Curr. Opin. Biotechnol. 14, 262-269.

Turk, B., Turk, V., Turk, D., 1997. Structural and functional aspects of papain-like cysteine proteinases and their protein inhibitors. Biol. Chem. 378, 141-150.

Turnbull, M., Webb, B., 2002. Perspectives on polydnavirus origins and evolution. Adv. Virus Res. Academic Press, pp. 203-254.

Uchida, K., Ohmori, D., Ueno, T., Nishizuka, M., Eshita, Y., Fukunaga, A., Kominami, E., 2001. Preoviposition activation of cathepsin-like proteinases in degenerating ovarian follicles of the mosquito *Culex pipiens pallens*. Dev. Biol. 237, 68-78.

van der Heijden, M.G., Bardgett, R.D., van Straalen, N.M., 2008. The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. Ecol. Lett. 11, 296-310.

van Ham, R.C., Kamerbeek, J., Palacios, C., Rausell, C., Abascal, F., Bastolla, U., Fernandez, J.M., Jiménez, L., Postigo, M., Silva, F.J., et al., 2003. Reductive genome evolution in *Buchnera aphidicola*. Proc. Natl. Acad. Sci. U. S. A. 100, 581-586.

Van Valen, L., 1973. A new evolutionary law. Evolutionary Theory 1, 1-30.

Labropoulou, V., Douris, V., Stefanou., Magrioti., Swevers, L., Iatrou, K., 2008. Endoparasitoid wasp bracovirus-mediated inhibition of hemolin function and lepidopteran host immunosuppression. Cell. Microbiol. 10, 2118-2128.

Wade, M.J., 2007. The co-evolutionary genetics of ecological communities. Nat. Rev. Genet. 8, 185-195.

Wang, W., Brunet, F.d.r.G., Nevo, E., Long, M., 2002. Origin of sphinx, a young chimeric RNA gene in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. U. S. A. 99, 4448-4453.

Webb, B.A., Strand, M.R., Dickey, S.E., Beck, M.H., Hilgarth, R.S., Barney, W.E., Kadash, K., Kroemer, J.A., Lindstrom, K.G., Rattanadechakul, W., et al., 2006. Polydnavirus genomes reflect their dual roles as mutualists and pathogens. Virology 347, 160-174.

Weber, B., Annaheim, M., Lanzrein, B., 2007. Transcriptional analysis of polydnavirus genes in the course of parasitization reveals segment-specific patterns. Arch. Insect Biochem. Physiol. 66, 9-22.

Weber, E., Koebnik, R., 2006. Positive Selection of the Hrp Pilin HrpE of the Plant Pathogen *Xanthomonas*. J. Bacteriol. 188, 1405-1410.

Weers, P.M.M., Ryan, R.O., 2006. Apolipophorin III: Role model apolipoprotein. Insect Biochem. Mol. Biol. 36, 231-240.

Welch, R.A., Burland, V., Plunkett, G., Redford, P., Roesch, P., Rasko, D., Buckles, E.L., Liou, S.R., Boutin, A., Hackett, J., et al., 2002. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic Escherichia coli. Proc. Natl. Acad. Sci. U. S. A. 99, 17020-17024.

Werren, J.H., Baldo, L., Clark, M.E., 2008. Wolbachia: master manipulators of invertebrate biology. Nat Rev Micro 6, 741-751.

Whitfield, J.B., 2002. Estimating the age of the polydnavirus/braconid wasp symbiosis. Proc. Natl. Acad. Sci. U. S. A. 99, 7508-7513.

Whitfield, J.B., Asgari, S., 2003. Virus or not? Phylogenetics of polydnaviruses and their wasp carriers. J. Insect Physiol. 49, 397-405.

Whitfield, J.B., Baker, M.R., Littlewood, 2003. Phylogenetic Insights into the Evolution of Parasitism in Hymenoptera. Adv. Parasitol. Academic Press, pp. 69-100.

Whitten, M.M.A., Tew, I.F., Lee, B.L., Ratcliffe, N.A., 2004. A novel role for an insect Apolipoprotein (Apolipophorin III) in {beta}-1,3-Glucan pattern recognition and cellular encapsulation reactions. J. Immunol. 172, 2177-2185.

Woolhouse, M.E.J., Webster, J.P., Domingo, E., Charlesworth, B., Levin, B.R., 2002. Biological and biomedical implications of the co-evolution of pathogens and their hosts. Nat. Genet. 32, 569-577.

Xing, R., Mason, R.W., 1998. Design of a transferrin-proteinase inhibitor conjugate to probe for active cysteine proteinases in endosomes. Biochem. J. 336, 667-673.

Xu, D.M., Stoltz, D., 1991. Evidence for a chromosomal location of polydnavirus DNA in the ichneumonid parasitoid *Hyposoter fugitivus*. J. Virol. 65, 6693-6704.

Yan, J., Cheng, Q., Li, C.B., Aksoy, S., 2002. Molecular characterization of three gut genes from *Glossina morsitans morsitans*: cathepsin B, zinc-metalloprotease and zinc-carboxypeptidase. Insect Mol. Biol. 11, 57-65.

Yang, Z., 1997. PAML: A program package for phylogenetic analysis by maximum likelihood Comput. Appl. Biosci. 13, 555-556.

Yang, Z., 2000. Maximum likelihood estimation on large phylogenies and analysis of adaptive evolution in human influenza virus A. J. Mol. Evol. 51, 423-432.

Yang, Z., Nielsen, R., 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Mol. Biol. Evol. 19, 908-917.

Yang, Z., Swanson, W.J., 2002. Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. Mol. Biol. Evol. 19, 49-57.

Yang, Z., Wong, W.S.W., Nielsen, R., 2005. Bayes empirical bayes inference of amino acid sites under positive selection. Mol. Biol. Evol. 22, 1107-1118.

Zhang, J., Nielsen, R., Yang, Z., 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol. Biol. Evol. 22, 2472-2479.

Zhang, J., Zhang, Y.-p., Rosenberg, H.F., 2002. Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. Nat. Genet. 30, 411-415.

Zientz, E., Dandekar, T., Gross, R., 2004. Metabolic interdependence of obligate intracellular bacteria and their insect hosts. Microbiol. Mol. Biol. Rev. 68, 745-770.

Zou, Z., Najar, F., Wang, Y., Roe, B., Jiang, H., 2008. Pyrosequence analysis of expressend sequence tags for *Manduca sexta* hemolymph proteins involved in immune responses. Insect Biochem. Mol. Biol. 38, 677-682.

PUCVL
Pôle Universités
Centre Val de Loire

Céline SERBIELLE

Rôle et évolution de facteurs de virulence impliqués dans une interaction hôte-parasitoïde

UNIVERSITÉ
FRANÇOIS - RABELAIS
TOURS

Résumé

Les symbioses, en permettant l'acquisition de nouvelles fonctions, ont joué un rôle majeur dans l'évolution des organismes. Cette thèse, vise à comprendre les mécanismes moléculaires et évolutifs qui font des polydnavirus des symbiontes indispensables au succès parasitaire de guêpes parasitoïdes. Pour cela, nous avons étudié l'évolution des gènes viraux et les fonctions physiologiques potentiellement ciblées chez l'hôte.

Nous avons montré que l'évolution des gènes viraux était caractérisée par des duplications et une divergence rapide répondant à des pressions de sélection positive.

Chez l'hôte, nous avons montré que des protéases à cystéine, étaient régulées au cours du parasitisme au niveau de la transcription et de la traduction, suggérant que ces protéines doivent être impliquées dans le parasitisme.

Il reste maintenant à déterminer quelle est la fonction de ces protéines et leur influence sur l'évolution des gènes viraux.

Mots clés: polydnavirus, guêpe parasitoïde, cystatine, protéine tyrosine phosphatase, protéases à cystéine, évolution moléculaire.

Abstract

Symbioses have largely contributed to the evolution of species by providing new function. Here, our aim was to study the molecular evolution of virulence factors of polydnaviruses, essential symbiont for wasp parasitism success. We studied both PDV genes and their Lepidopteran host targets to determine how these genes evolved and the host functions targeted.

We showed that natural selection has largely contributed to virus genes evolution. We also underlined the dynamic evolution of PDV genes mainly explained by gene duplication processes.

In the host, cysteine proteases, were shown to be regulated during parasitism both at the gene and the protein level. These results emphasized the important role played by these proteins against invaders.

To understand both evolutionary and mechanistic processes involved in this interaction, it is now necessary to determine the function of cysteine proteases and to study how virus gene evolution could be shaped by host defence factors.

Key words: polydnavirus, parasitoid wasp, cystatin, protein tyrosine phosphatise, cysteine proteases, molecular evolution.