

RESEARCH ARTICLE

Open Access

Molecular adaptation and resilience of the insect's nuclear receptor USP

Arnaud Chaumot², Jean-Luc Da Lage³, Oscar Maestro⁴, David Martin⁴, Thomas Iwema⁵, Frederic Brunet¹, Xavier Belles⁴, Vincent Laudet¹ and François Bonneton^{1*}

Abstract

Background: The maintenance of biological systems requires plasticity and robustness. The function of the ecdysone receptor, a heterodimer composed of the nuclear receptors ECR (NR1H1) and USP (NR2B4), was maintained in insects despite a dramatic divergence that occurred during the emergence of Mecoptera. This receptor is therefore a good model to study the evolution of plasticity. We tested the hypothesis that selection has shaped the Ligand-Binding Domain (LBD) of USP during evolution of Mecoptera.

Results: We isolated *usp* and *cox1* in several species of Drosophilidae, Tenebrionidae and Blattaria and estimated non-synonymous/synonymous rate ratios using maximum-likelihood methods and codon-based substitution models. Although the *usp* sequences were mainly under negative selection, we detected relaxation at residues located on the surface of the LBD within Mecoptera families. Using branch-site models, we also detected changes in selective constraints along three successive branches of the Mecoptera evolution. Residues located at the bottom of the ligand-binding pocket (LBP) underwent strong positive selection during the emergence of Mecoptera. This change is correlated with the acquisition of a large LBP filled by phospholipids that probably allowed the stabilisation of the new Mecoptera structure. Later, when the two subgroups of Mecoptera (Amphiesmenoptera: Lepidoptera, Trichoptera; Antliophora: Diptera, Mecoptera, Siphonaptera) diverged, the same positions became under purifying selection. Similarly, several positions of the heterodimerisation interface experienced positive selection during the emergence of Mecoptera, rapidly followed by a phase of constrained evolution. An enlargement of the heterodimerisation surface is specific for Mecoptera and was associated with a reinforcement of the obligatory partnership between ECR and USP, at the expense of homodimerisation.

Conclusions: In order to explain the episodic mode of evolution of USP, we propose a model in which the molecular adaptation of this protein is seen as a process of resilience for the maintenance of the ecdysone receptor functionality.

Keywords: Nuclear receptors, Ecdysone receptor, ECR, USP, Mecoptera, Selection

Background

During the course of evolution, the maintenance of biological systems requires the conservation of vital functions. At the same time, living beings are facing constant changes in their genome and environment. These variations, which are continuously challenging the fitness of the organism, are key modulators of adaptation. It is therefore essential to maintain a good balance between

plasticity and robustness [1]. How does this equilibrium evolve at the genetic level? In order to tackle this important issue, it is necessary to use a model system with a vital function that has been conserved during a given period of time, despite the divergence of the genetic modules that determine this function. The ecdysone receptor fulfils these requirements. This hormone receptor is a heterodimer composed of two nuclear receptors: ECR (NR1H1) and its obligatory partner USP (NR2B4), the ortholog of mammalian RXR (NR2B1,2,3). The activation of ECR by ecdysteroid hormones triggers a gene cascade that controls development (embryogenesis, molts, metamorphosis) and reproduction in insects [2,3]. Both

* Correspondence: francois.bonneton@ens-lyon.fr

¹Institut de Génétique Fonctionnelle de Lyon (IGFL), Université de Lyon, Université Lyon 1; CNRS; INRA; Ecole Normale Supérieure de Lyon, 32-34 avenue Tony Garnier, Lyon 69007, France

Full list of author information is available at the end of the article

partners of this hormone receptor have undergone a dramatic divergence during the emergence of the Mecoptera, a superorder of holometabolous insects, which includes Diptera (flies, mosquitoes), Lepidoptera (moths, butterflies), Mecoptera (scorpionflies), Siphonaptera (fleas), and Trichoptera (caddisflies) [4-6]. This acceleration of evolutionary rate is correlated with structural and functional changes of the LBP of USP and of the heterodimer interface between ECR and USP [7,8]. These results suggested to us that the maintenance of the ecdysone pathway was achieved, at least in part, through molecular adaptation of the USP LBD. We thus tested the hypothesis that selection has shaped this domain during the emergence of Mecoptera.

Changes in evolutionary rates are commonly used to detect innovations in the function of proteins. It is generally assumed that a slowdown of substitution rate is caused by the action of negative purifying selection that maintains an important function. By contrast, an acceleration of evolutionary rate can be interpreted in two opposite ways. It could be due to a phase of positive selection on a new function, or to neutral evolution of a functionless sequence. This simplistic approach raises several problems. Indeed, the rate of evolution at a given position can change according to its interactions with other sites. Many interactions are compatible with different conformations of a protein with the same function. Therefore, within-site rate variation in time (heterotachy) is not necessarily associated with a change in function, but may simply reflect the epistatic nature of genetic units [9]. In other words, the same site of a protein can follow different evolutionary paths in a phylogenetic tree, and whether these variations are adaptive or neutral cannot be inferred solely from the estimate of evolutionary rates. It is essential to take into account the interdependence of residues, within and between proteins. Consequently, new methods try to incorporate protein structural constraints into evolutionary models [10]. Another important problem is to distinguish between positive selection and relaxation of negative selection [11]. Furthermore, the evolution of protein function involves alternating periods of conservative evolution and of rapid change [12]. As a consequence, evidence for positive selection at the protein level are rare at a short time scale but much more frequent when considering deeper evolutionary time [13]. Finally, non-adaptive processes such as variations of the mutation rate or GC-biased gene conversion, can lead to an increase of evolutionary speed that is independent of natural selection [14,15]. We are trying to tackle some of these issues that are central for a correct understanding of the genetic mechanisms of protein evolution using the ecdysone receptor. Thanks to the availability of crystal structures and functional data, the ECR-USP case provides clear biological hypotheses

regarding the codons upon which positive selection is expected. Using appropriate models, we could test our hypotheses at different time scales by analysing evolutionary constraints either within extant insect families, or during three successive branches of Mecoptera radiation. Our results reveal the dynamics of evolutionary forces acting on the LBD of USP.

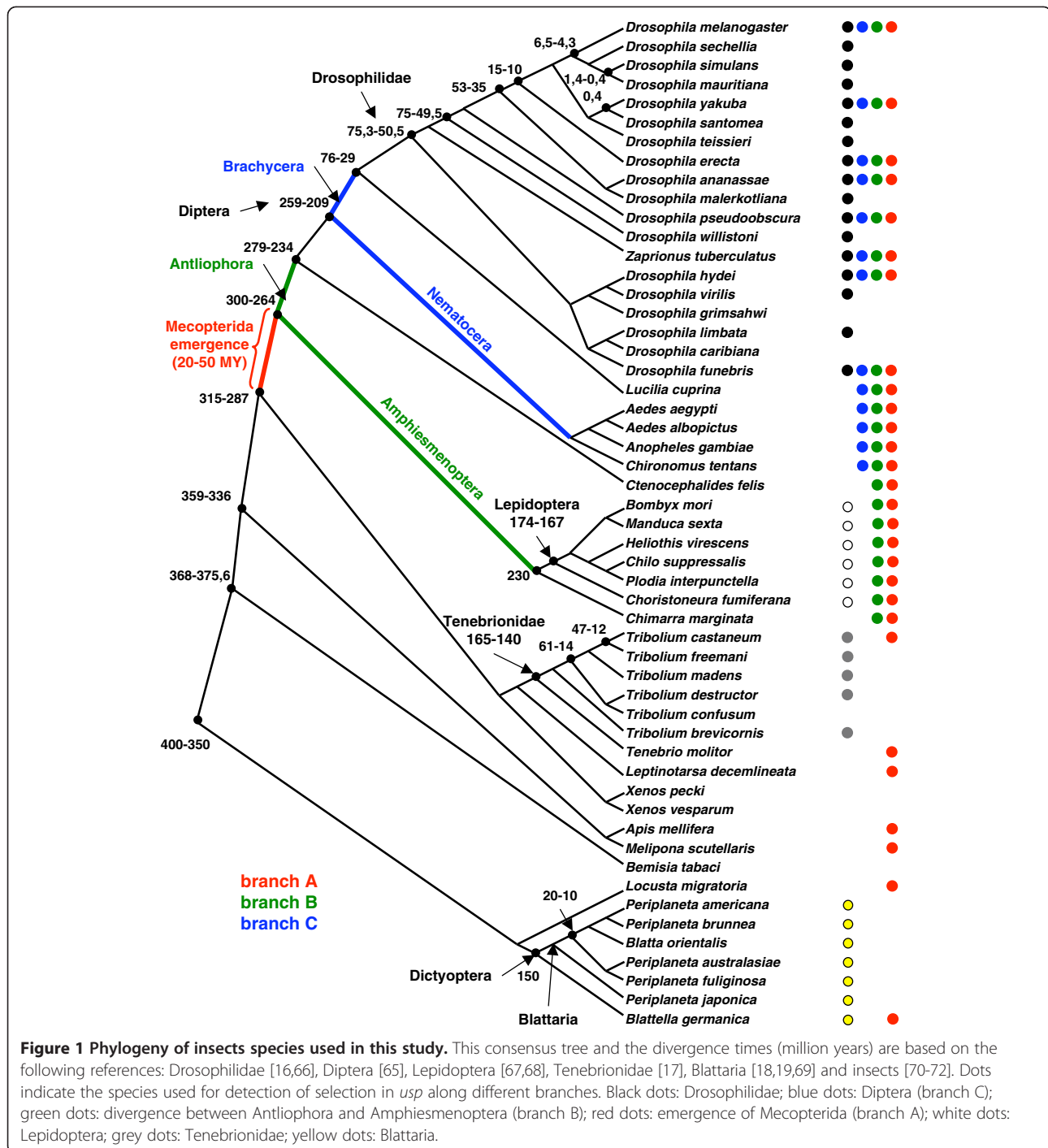
Results

Sequence variation of *usp* and *cox1* in three insect species-groups

In order to study the evolutionary dynamics of USP at different time scales, we isolated sequences from several species of Drosophilidae, Tenebrionidae and Blattaria (Figure 1; Additional file 1: Table S1). We have chosen the first two families as representative of the holometabolous insects, with Drosophilidae showing the Mecoptera “overacceleration” of USP [4,6]. Tenebrionidae belong to Coleoptera, which is the sister group of Mecoptera, while Blattaria are hemimetabolous insects. Both Tenebrionidae and Blattaria have a non-divergent USP [4]. The 19 Drosophilidae species used in this study span an evolutionary time of 0.4 to 75 million years [16]. The most ancient divergence time for the five true *Tribolium* species used in our analysis was estimated between 14 and 61 million years ago [17]. Blattaria species may have diverged between 10 and 20 million years ago [18,19].

In order to compare the variable rates of USP evolution between these different insect groups, we used the gene *cox1*, which encodes the mitochondrial cytochrome oxidase subunit 1. This gene is known to evolve under a relatively uniform rate in most taxa of insects [20]. Overall, our data set consists of protein sequences for both USP and COX1 in 32 species of Drosophilidae, Tenebrionidae and Blattaria (Additional file 1: Table S1). Among these 64 sequences, 43 are new and were isolated for this study. The alignment of these proteins provides 218 positions (including gaps) for USP LBD and 410 positions (no gap) for COX1. As expected, the COX1 protein sequences are very well conserved between these three insect groups (>85% identity), whereas the USP LBD of Drosophilidae is highly divergent from the other species (≈40% identity).

The phylogenies that we obtained using alignments of concatenated sequences (1042 positions) of USP and COX1 proteins are similar to those already published (data not shown). For example, we observed that *Tribolium brevicornis* is not closely related to the other *Tribolium* species, as shown by Angelini *et al.* [17], who suggest that it should be included into the genus *Aphanotus*. We also observed the paraphyly of *Periplaneta spp.*, with *Blatta orientalis* located within this clade, as shown by other authors [18,19].

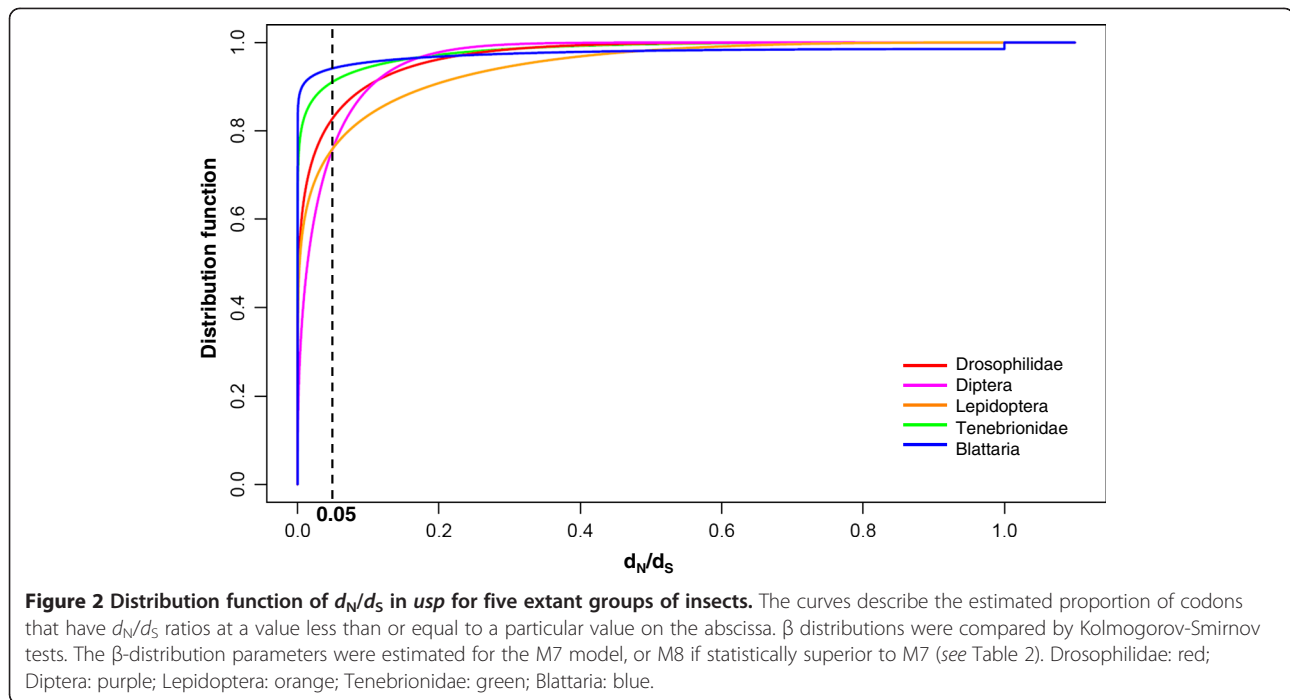


The negative selection on the *usp* gene is weaker in Mecoptera families

We estimated non-synonymous/synonymous rate ratios (d_N/d_S or ω) by whole tree analyses using maximum-likelihood methods and codon-based substitution models [21,22] (Additional file 2: Table S3). We analysed three groups of species that are characterized by a derived *usp* (Drosophilidae, Diptera, Lepidoptera) and two groups

with a less modified *usp* (Tenebrionidae and Blattaria) (Figure 1).

If we consider each group separately, the best model indicates that the main force acting on *usp* sequences is negative selection: d_N/d_S is less than 0.05 for at least 75% of the sites (Figure 2). We could detect rare neutral codons in Blattaria (3/279), but not in the other groups (Table 1). Likelihood ratio tests contrasting positive



selection models with nearly neutral models (M2a versus M1a or M8 versus M7) never conclude that positive selection ($d_N/d_S > 1$) is acting on *usp* within these groups (Table 1). This notably explains the difficulty to adjust parameters of models M2a and M8, as shown by standard errors associated to the estimation of p_1 , ω_2 and ω_s (Additional file 3: Table S4). The highly constrained pattern also gives rise to an imprecise estimation (large

relative standard errors) of the two shape parameters of the β distribution for the two non-Mecoptera groups (Additional file 3: Table S4). This is coherent with the very weak improvement of the likelihood between M1a and M7/M8 models for these two groups. Kolmogorov-Smirnov tests indicate significant differences in the distribution of site-specific d_N/d_S between insect groups (p-value $< 10^{-7}$). As an illustration, the percentage of

Table 1 Likelihood analysis of the *usp* sequence datasets of five extant insect groups

Model		Drosophilidae	Diptera	Lepidoptera	Tenebrionidae	Blattaria
M1a	LnL (np)	-5649.5 (31)	-6946.6 (25)	-4376.2 (12)	-2113.5 (10)	-1945.6 (13)
	$\omega_0 \omega_1$	0.02 1.00	0.03 1.00	0.03 1.00	0.01 1.00	0.00 1.00
	$p_0 p_1$	0.98 0.02	0.98 0.02	0.95 0.05	0.99 0.01	0.99 0.01
M2a	LnL (np)	-5649.5 (33)	-6946.6 (27)	-4376.2 (14)	-2113.5 (12)	-1945.6 (15)
	$\omega_0 \omega_1 \omega_2$	0.02 1.00 1.00	0.03 1.00 1.00	0.03 1.00 1.00	0.01 1.00 1.00	0.00 1.00 1.24
	$p_0 p_1 p_2$	0.98 0.01 0.01	0.98 0.01 0.01	0.95 0.03 0.02	0.99 0.01 0.00	0.99 0.01 0.00
M7	LnL (np)	-5610.8 (31)	-6859.5 (25)	-4339.9 (12)	-2112.9 (10)	-1950.0 (13)
	$p q$	0.16 5.05	0.40 10.7	0.15 2.58	0.06 3.13	0.01 0.31
M8	LnL (np)	-5610.8 (33)	-6857.1 (27)	-4338.8 (14)	-2112.9 (12)	-1946 (15)
	$p q$	0.16 5.05	0.43 12.9	0.16 2.85	0.06 3.13	0.02 2.09
	ω_s	1.00	1.00	4.46	1.00	1.00
	p_s	0.00	0.01	0.00	0.00	0.01
LRT (p-value)	test of positive selection					
M2a vs M1a		1	1	1	1	1
M8 vs M7		1	0.08	0.32	1	0.02

LnL: log likelihood. np: number of parameters. ω_0 , ω_1 , ω_2 : non-synonymous/synonymous rate ratio that characterize a negative, a neutral, and a positive selection regime, respectively (with corresponding proportions of sites p_0 , p_1 , p_2). p and q : shape parameters of a β function describing the distribution of non-synonymous/synonymous rate ratios. ω_s : ratio affecting a p_s proportion of sites under positive selection in model M8. LRT: Likelihood ratio test.

codons with a d_N/d_S above 0.05 is 20% in Mecoptera and only 5% in non-Mecoptera groups (Figure 2). Since saturation was detected among Diptera *usp*, we performed a re-analysis of alignments simulated according to the estimated characteristics of the Diptera dataset (13 species, 346 codons, same topology and branch lengths for the phylogenetic tree, same codon usage, transition/transversion rate ratio of 1.89, repartition of sites between ten categories of d_N/d_S according to the estimated β distribution). The results showed a good accuracy in the estimate of the β distribution function (Additional file 4: Figure S1), indicating that the detection of this pattern is robust to the saturation in Diptera sequences. The confidence interval for the proportion of codons with $d_N/d_S > 0.05$ was less than $\pm 1\%$. In the same way, we confirmed that the weak precision in the estimation of the β distribution parameters for the Tenebrionidae and Blattaria does not alter their contrasted pattern with Mecoptera groups (data not shown). The Mecoptera specific regions are among the less constrained of *usp* because, when they are removed from the analysis, the Mecoptera curves shift slightly towards Tenebrionidae and Blattaria curves (data not shown). To compare the evolutionary dynamics of *usp* between groups, we used *cox1* as a reference (410 codons) (Additional file 5: Table S5). In Drosophilidae, the mean rate of non-synonymous substitutions of *usp* is 3.4 times higher than for *cox1*. By contrast, *usp* accumulates less non-synonymous substitutions than *cox1* in the non-Mecoptera families (2.5 times less in Tenebrionidae and 1.3 times less in Blattaria). Overall, by employing *cox1* as a relative molecular clock, the Drosophilidae *usp* gene evolved 8.6 times faster than the Tenebrionidae *usp* and 4.4 times faster than the Blattaria homolog.

These results suggest that negative selection on *usp* is relaxed in Mecoptera families, when compared to non-Mecoptera families.

Relaxation of negative selection on the external surface of USP protein

In order to identify precisely the sites that present contrasted patterns in selective constraints between these groups, we projected the site-specific d_N/d_S onto the available 3D structures of USP protein (Figure 3). Three functional domains of USP are highly constrained in every group of insects: the DBD (Additional file 2: Table S3), the LBP and the dimerisation interface. By contrast, other parts of the protein have experienced more relaxed evolutionary pressure, especially among the Mecoptera families.

The LBD of non-Mecoptera USPs are homogeneously constrained. Only 9 sites (5%), located on the external surface of the LBD, have an estimated $d_N/d_S \geq 0.1$ in Tenebrionidae and Blattaria. The uncertainty in parameter estimates (Additional file 3: Table S4) for these two groups may hinder the computation of site-specific rates,

but these sites are not homologous in these two insect groups, except one position in helix H4 (Ser246 in Blattaria and Thr241 in Tenebrionidae) (Figure 3A, B).

The LBD of Mecoptera USPs shows very different patterns, with 63 sites (25%) having a $d_N/d_S \geq 0.1$. Overall, the relaxation of evolution is more pronounced in Lepidoptera than in Diptera. Only one-third of these sites are similar between these two insect orders. Most of the relaxed residues are scattered on the surface of the LBD (Additional file 6: Figure S2), a few of them being located in regions with a known or putative function. For instance, the hinge D domain and the following helix H1 of the LBD exhibit a high proportion of relaxed positions in Diptera and Lepidoptera (17/30 and 8/28, respectively), when compared to Tenebrionidae and Blattaria (1/31 and 0/31, respectively) (Additional file 2: Table S3). This region is known to have an allosteric effect on the corepressor-binding surface of RXR and other nuclear receptors [23,24]. A cluster of seven relaxed positions is shared by both groups of insects in helix H4-H5, which is included into the putative coactivation surface (Figure 3C, D). In helix H12, the Lepidoptera sequences are totally divergent from all other USP-RXR proteins. Contrary to ECR, the helix H12 of USP has no transactivation activity because it adopts a structure similar to the antagonist conformation observed for RXR [7,25-28]. In *Drosophila* and in *Heliothis*, helix H12 is locked into this position by the Mecoptera specific domain present in the loop L1-3, between helices H1 and H3. Interestingly, several residues of this specific insertion are evolutionarily relaxed in Diptera, and even more in Lepidoptera (Figure 3C, D). A similar pattern of relaxation seems to occur in the loop between helix H5 and the first beta turn (Additional file 2: Table S3), where there is a second Mecoptera specific insertion of 20–30 amino acids. This divergent region, only partly stabilized in the crystal structures of *Heliothis* and *Drosophila*, is possibly implicated in contacts with ECR [25,29]. By contrast, the residues that constitute the heterodimerisation interface are constrained in Diptera and Lepidoptera.

These results show that the main feature of the recent evolution of USP is a relaxation of purifying selection in Mecoptera families. The LBP and the dimerisation interface, which underwent important Mecoptera specific changes in structure and function [7,8], are highly constrained at this time scale and positive selection was not detected. By contrast, an evolutionary relaxation might have affected external surfaces involved in protein-protein interactions between USP and its coregulators.

Positive selection on the *usp* gene during the emergence of Mecoptera

Since USP LBD underwent important modifications in Mecoptera, we wondered whether some of these changes

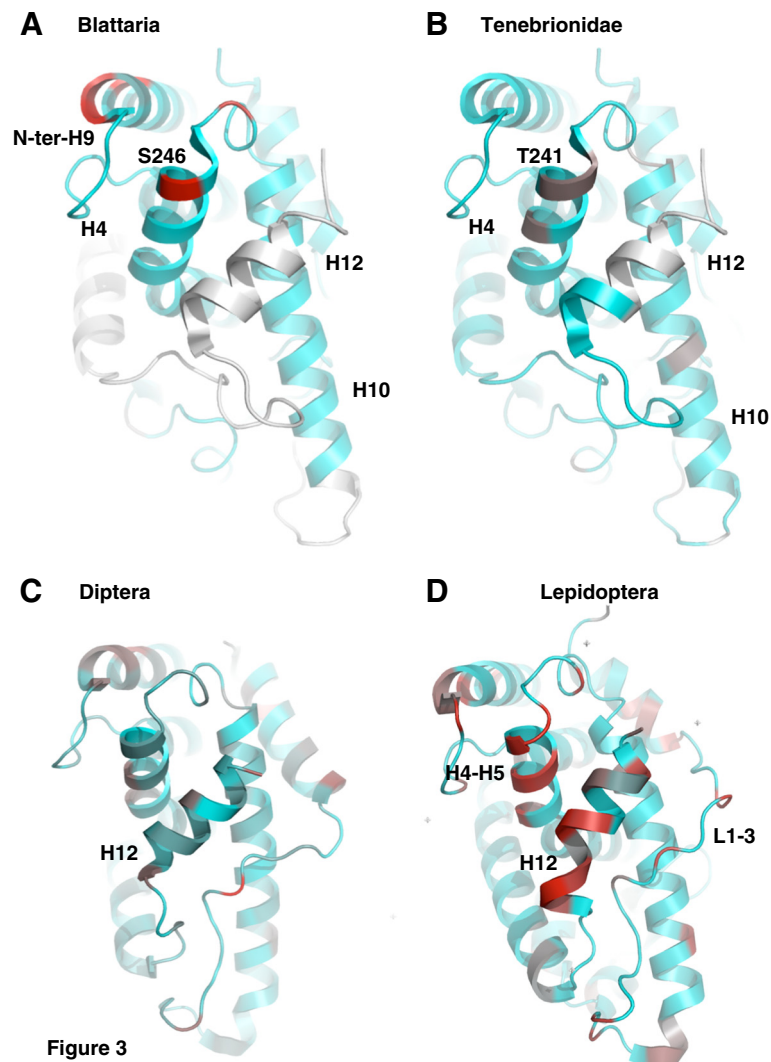


Figure 3

Figure 3 Mapping of evolutionary rates on USP structures in four extant groups of insects. Site-specific d_N/d_S were projected onto the crystal structure of the USP LBD domain of *Tribolium* (2NXX) for Blattaria (A) and Tenebrionidae (B), of *Drosophila* (1HG4) for Diptera (C) and of *Heliothis* (1R1K) for Lepidoptera (D). The values are distributed along a colour scale from blue (low d_N/d_S) to red (high d_N/d_S). Sequences not available for the estimation of evolutionary rates are in white.

involved functional shifts during the emergence of this group. Assuming that evolutionary pressure is changing over time, we used branch-site models [30] to detect potential site-specific changes in selective constraints (positive selection or relaxation of negative selection) along three successive branches of Mecoptera radiation. Branch A is the stem lineage of Mecoptera, branch B is the subdivision between Amphiesmenoptera and Antliophora and branch C is the subdivision of Diptera into the suborders Brachycera and Nematocera (Figure 1).

In order to assess whether relaxation or positive selection occurred along these branches, we applied tests 1 and 2 from Zhang *et al.* [30]. While test 1 is significant for all the three branches, test 2, a direct conservative test of positive selection, is significant only for the two

most ancient branches A and B (Table 2). It appears therefore that positive selection has occurred on specific sites at the origin of the Mecoptera clade (21% of *usp* sites) and during the divergence between Amphiesmenoptera and Antliophora (11% of *usp* sites). Subsequently, some sites (8%) evolved neutrally before becoming under purifying selection within Dipteran suborders (Nematocera and Brachycera). Since the specificity of these tests could be affected by long evolutionary distances, false positive rates were computed, as a control, from the re-analysis of alignments simulated with Evolver in PAML [12]. These rates were less than 1% for test 1 for the three datasets. For test 2, we obtained false positive rates of 5% for branch A and 2% for branch B. Therefore, our detection of positive selection is not significantly biased by

Table 2 Likelihood analysis of the *usp* sequence datasets along three successive branches of the Mecoptera radiation

	Model		Likelihood Ratio Test			
	No Shift	Branch relaxation	Branch positive selection		Relaxation	
	M1a (nearly neutral)	modified branch-site model A $\omega_2 = 1$ fixed	modified branch-site model A		test 1	
	LnL (np)	LnL (np)	LnL (np)	ω_2	p-value	p-value
Branch A	-11732.4 (55)	-11713.2 (56)	-11711.1 (57)	999	$6.6 \cdot 10^{-11}$	$4.2 \cdot 10^{-2}$
Branch B	-9769.8 (41)	-9754.5 (42)	-9750.9 (43)	97.3	$8.0 \cdot 10^{-9}$	$6.8 \cdot 10^{-3}$
Branch C	-6946.6 (25)	-6938.1 (26)	-6937.5 (27)	4.75	$1.1 \cdot 10^{-4}$	0.56

LnL: log likelihood. np: number of parameters.

saturation of synonymous sites during the large evolutionary times considered in this analysis. The initial episode of positive selection was dramatic, if we consider the percentage of sites involved and the high values of d_N/d_S for the site class under positive selection (ω_2 of branch B in Table 2). Note that the value of 999 computed for ω_2 along the stem lineage of Mecoptera (branch A) indicates the absence of predicted synonymous substitutions for this short branch ($d_S=0$), so the d_N/d_S ratio (ω_2) is undefined.

The sites that experienced a functional shift along the three branches were identified by computing the probability for each site to belong to the site-class that experienced positive selection (along branch A or branch B) or neutral evolution (along branch C), before becoming under negative selection in extant groups (Additional file 7: Table S6). The projection of these posterior probabilities onto the 3D structure of the LBD of *Drosophila* USP reveals that most of these sites under positive selection, or relaxation, are on the external surface of the LBD, notably on branch A. There is no known function for the majority of these sites. A minority of the positively selected amino acids are located on two functional domains of the LBD that are the LBP and the heterodimerisation interface: 36% on branch A, 15% on branch B and 21% on branch C. Among the residues with a known function, there is a high fraction which is positively selected or relaxed on branch A (18/36), and less for branch B (7/20) and branch C (6/14).

The large LBP of Mecoptera USP is composed of two parts: a channel extending inside from the opening and a deeper pocket, which is similar to the LBP of RXR. The tail of the phospholipid is buried inside this pocket, while its phosphate head group is at the entrance. The amino acids of the channel, from where the ligand penetrates inside the LBP, are under negative selection in the three branches. By contrast, four residues that make contacts with the ligand at the bottom of the internal surface underwent strong positive selection during the emergence of Mecoptera (Figure 4A). This change of evolutionary pressure is correlated with the specific acquisition of a

large LBP filled by phospholipids in Mecoptera. In other insects, residues of the loops L6 and L11 fill the deep unliganded pocket. Later, when Amphipsectera and Antliophora diverged (branch B), the same positions became under purifying selection (Figure 4B). Subsequently, one of them (Tyr372), followed a new path of relaxed evolution during the subdivision of Dipteran suborders (branch C) (Figure 4C). All Brachycera share a tyrosine at this position, while Nematocera and other Mecoptera have a leucine. In the homologous position of RXR there is a valine, which does not interact with the ligand.

A similar evolutionary dynamic is observed for the heterodimerisation interface, where several positions experienced positive selection during the emergence of Mecoptera, followed by a phase of constrained evolution (Figure 5). Importantly, among these positions, arginine Arg426 in helix H9 (Figure 5A) is known to be involved in the Mecoptera specific enlargement of the heterodimerisation surface [8]. On the same branch, positive selection is detected on three other residues that, albeit not interacting with ECR, are located within the canonical interaction surface. The arginine Arg399 in helix H7 was predicted to form an electrostatic interaction with ECR in the ancestral Mecoptera model [8]. This is interesting because this bond is not seen in the current USP structure of *Drosophila* or *Heliothis* [25,26]. Another noticeable residue is the glutamine Gln457 in helix H10, which is variable in Mecoptera. Indeed, its homologous position in non-Mecoptera USP and in vertebrate RXR is a strictly conserved lysine that interacts with the partner of heterodimerisation. This bond was observed in human and mouse RXR heterodimers as well as in the ECR-USP structure of the Hemiptera *Bemisia tabaci* white flies, but not in *Tribolium* [7,28].

By contrast to the LBP and to the dimerisation interface, there is no clear evidence of positive selection in the putative coactivation surface, simply defined here by homology with RXR (Additional file 8: Figure S3). Although we could not estimate the evolutionary rate for the whole surface in the three branches, it seems that this domain is rather homogeneously constrained. Only two

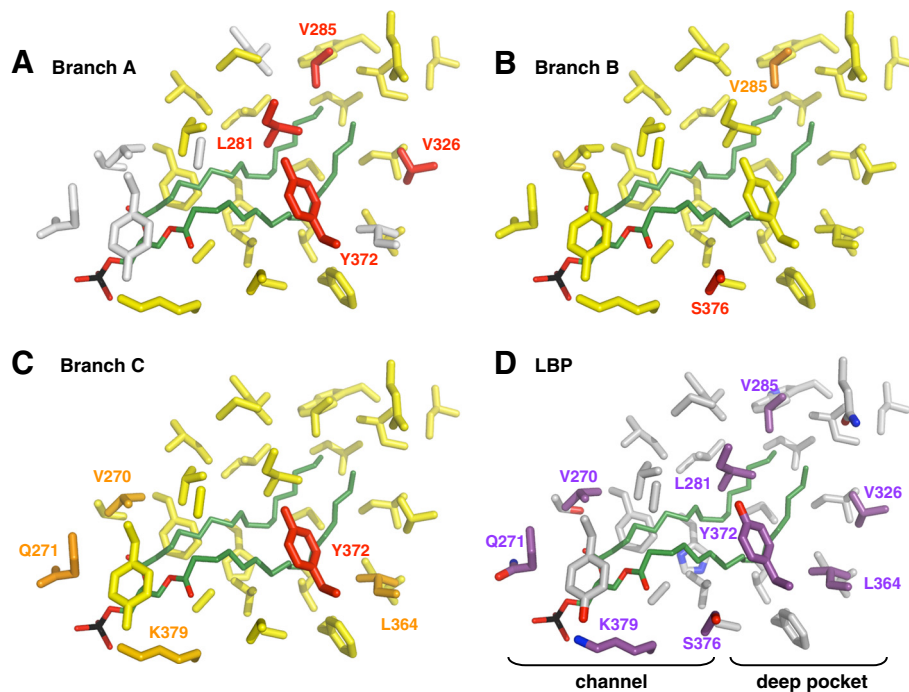


Figure 4 Selection in the LBP during the radiation of Mecoptera. Posterior probabilities for each site to belong to the site-class under positive selection (along branch **A** or branch **B**) or relaxed evolution (along branch **C**) were projected onto the crystal structure of the USP LBD domain of *Drosophila* (1HG4). Probabilities are distributed along a colour scale from yellow ($p=0$) to red ($p=1$). Sequences not available for the estimation of evolutionary rates are in white. The phospholipid ligand (LPP) is in green, with its tail on the right in the deep pocket. **(A)** Branch **A**, stem lineage of Mecoptera. **(B)** Branch **B**, subdivision between Amphiesmenoptera (Lepidoptera, Trichoptera) and Antliophora (Diptera, Mecoptera, Siphonaptera). **(C)** Branch **C**, subdivision of Diptera into the suborders Brachycera and Nematocera. **(D)** All the 32 amino acids of the LBP in the USP structures of *Drosophila* and *Heliothis*. Residues under positive selection or relaxed evolution in at least one of the three branches are shown in purple.

residues in helix H3 have experienced positive selection during the emergence of Mecoptera (branch A), one of which being also selected on branch B (subdivision between Amphiesmenoptera and Antliophora).

In conclusion, the structural analysis of the sites under selection shows that both the LBP and the dimerisation interface of USP LBD have experienced modifications of selective constraints during the emergence of Mecoptera.

Did biased gene conversion contribute to the early evolution of *usp* in Mecoptera?

Acceleration of substitution rates can result from other processes than selection, such as GC-biased gene conversion (gBGC), which enrich the G+C content [15,31]. We thus measured the G+C content of *usp* gene and found that it is higher in Drosophilidae (59.6%), when compared to Tenebrionidae and Blattaria (48.7%). The highest average G+C content is at third codon positions, where the Mecoptera G+C bias is stronger (Table 3). As a result, the preference for codons ending with a G or a C in *usp* is higher in Mecoptera than in non-Mecoptera for most of the amino acids (Additional file 9: Table S2). Therefore, the nucleotide composition of

usp gene in Mecoptera is biased. By contrast, the average *cox1* G+C content are similar in the three families (30-37%) and fall into the range of insect's whole mitochondrial genomes (13-35%, mean=24%) [32]. Since most of the synonymous sites are at the third codon position, such a bias in G+C content may affect the estimation of evolutionary rates of synonymous and non-synonymous substitutions of the *usp* gene. We therefore tested whether the increase of substitution rate observed in Mecoptera *usp* could be due, at least in part, to GC-biased gene conversion along branch A. First, the two ancestral sequences for the internal nodes at each extremity of branch A were determined by marginal reconstruction performed with Codeml within PAML software (Yang et al. 1997). Then, we identified the predicted substitutions along this branch by comparing these two ancestral sequences. We could not detect any bias in favour of AT→GC substitutions (p -value = 0.4). This means that the current bias in the G+C content of *usp* gene was probably acquired later during evolution. It is therefore unlikely that a phase of GC-biased gene conversion had a significant effect in the global increase of evolutionary rates observed in *usp* during the emergence of Mecoptera.

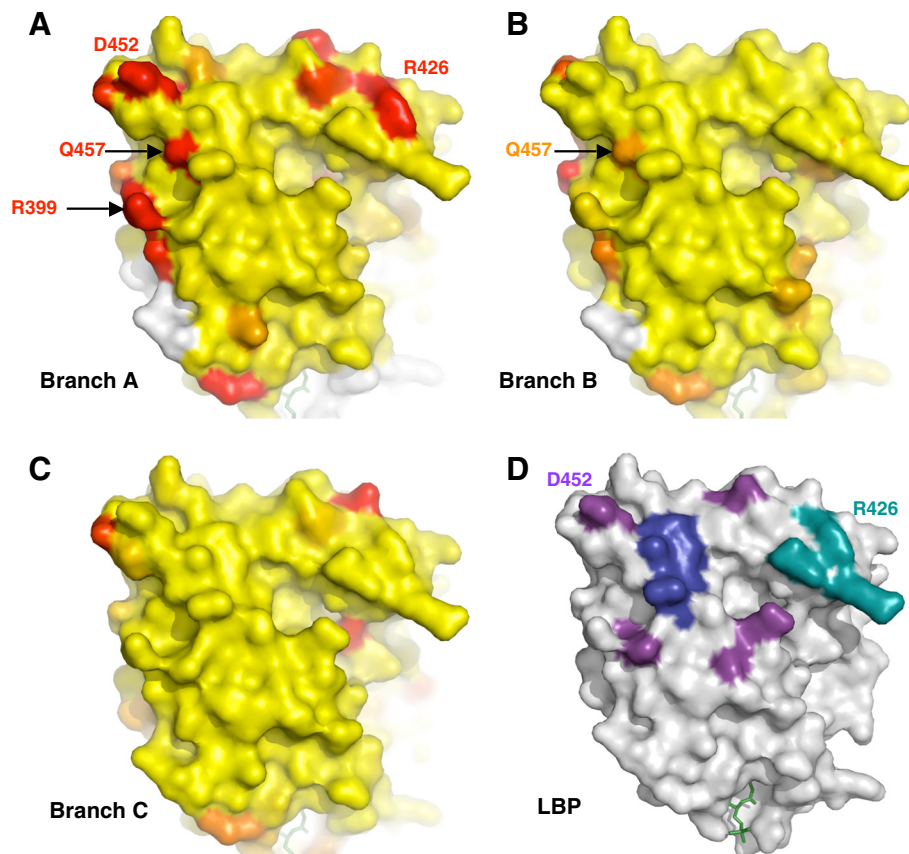


Figure 5 Selection in the heterodimerisation interface during the radiation of Mecoptera. Posterior probabilities for each site to belong to the site-class under positive selection (along branch **A** or branch **B**) or relaxed evolution (along branch **C**) were projected onto the crystal structure of the USP LBD domain of *Drosophila* (1HG4). Probabilities are distributed along a colour scale from yellow ($p=0$) to red ($p=1$). Sequences not available for the estimation of evolutionary rates are in white. The lipid ligand (LPP) is in green. **(A)** Branch **A**, stem lineage of Mecoptera. **(B)** Branch **B**, subdivision between Amphiesmenoptera (Lepidoptera, Trichoptera) and Antliophora (Diptera, Mecoptera, Siphonaptera). **(C)** Branch **C**, subdivision of Diptera into the suborders Brachycera and Nematocera. **(D)** The hydrophobic core (helix H10) is shown in blue, the insect canonical interaction surface is in purple and the Mecoptera specific surface is in cyan (for details, see Additional file 1: Table S1 of: [8]).

Discussion

The ecdysone receptor (ECR-USP) has undergone a dramatic phase of mutations during the emergence of the Mecoptera superorder [4-6]. This acceleration of evolutionary rate is correlated with structural and functional changes of the LBP of USP and of the heterodimer interface between ECR and USP [7,8]. These results suggested that the maintenance of the ecdysone pathway was

achieved, at least in part, through molecular adaptation of the USP LBD. We thus tested the possibility that this domain was under positive selection during the emergence of Mecoptera.

Constraints and relaxation on USP LBD

Overall, USP is highly constrained. The majority of the sites are under negative selection on all branches. Similar conclusions have been obtained with other nuclear receptors, including RXR, the vertebrate homolog of USP [33-35]. This result is compatible with the idea that USP could be a hub protein. Indeed, proteins with a large number of interactions are usually under purifying selection because any substitution may have pleiotropic effects. At the same time, adaptations are allowed by relaxation in the interactive surfaces that establish transient interactions between protein domains [36].

Table 3 Average percentage of G+C content in codons of usp

	Diptera	Lepidoptera	<i>Tribolium</i>	Blattaria
Codons	346	398	283	279
First position	57,5%	59,0%	49,7%	57,9%
Second position	38,7%	42,2%	35,3%	40,4%
Third position	67,5%	72,8%	52,7%	48,9%
All sites	54,6%	58,0%	45,9%	49,1%

In the case of the LBD of USP, we detected a relaxation of negative selection within the Mecoptera, relative to non-Mecoptera insects. Most of the relaxed sites, which are on the external surfaces of the protein, may be involved in protein-protein interactions between USP and its coregulators. It is therefore possible that the evolutionary plasticity observed on the surface of the Mecoptera USP LBD reflects adaptability to its different partners. Unfortunately, the network of protein-protein interactions is largely unknown for USP and experimental data are needed to go beyond these speculations. In fact, the complexity of this matter concerns all nuclear receptors, for which the analysis of protein interactions are largely restricted to the dimerisation interface and to the small coregulatory groove [37]. The use of original evolutionary methods can yield a better understanding of these interactive sites. For instance, difference evolutionary trace analysis has revealed a new functional site for steroid receptor located on a surface opposite to currently known protein-protein interaction [38].

Regions of USP LBD under positive selection during Mecoptera emergence

Our previous results suggested that the main functional shifts affecting USP LBD of insects occurred during the emergence of Mecoptera, approximately 280–300 million years ago (early Permian). We therefore used branch-site models to detect changes in selective constraints along this branch. Importantly, we also analysed two more recent branches within the Mecoptera clade in order to decipher the successive state of evolutionary forces acting on USP LBD. Indeed, the interpretation of changes in evolutionary rates must take into account all the biological complexity of the protein, such as the dynamic of the process or the distinction between functional modules. Using this approach, we were able to detect positive selection at amino acids located on the LBP and on the heterodimerisation interface of USP LBD.

The episode of positive selection is correlated with the acquisition of a large LBP filled by fatty acids. In the basal insect *Locusta migratoria* (Orthoptera), USP is able to bind 9-cis retinoic acid, like homologous vertebrate RXR, which natural ligands are fatty acids [39]. In *Bemisia* (Hemiptera) and *Tribolium* (Coleoptera), USP is unable to be activated by RXR ligands and acts as a constitutively silent partner of ECR. In fact, in these species, residues of the loops L6 and L11 fill the deep unliganded pocket and, therefore, USP lacks a proper LBP [7,28]. All the four residues that underwent strong positive selection during the emergence of Mecoptera are located at the bottom of the deep pocket, which is similar to the LBP of RXR. These residues make contacts with the tails

of the phospholipid in Mecoptera structures. Later during the diversification of Mecoptera, the same positions became under purifying selection. These results suggest that the Mecoptera pocket was positively selected for a new function, which then became highly constrained and remained largely unchanged until now. What was this innovation? We can hypothesize that the presence of phospholipids allowed the stabilisation of the new Mecoptera LBD. Indeed, this structure is characterised by the acquisition of a new loop L1-3 that locks the LBD in a conformation similar to the one seen for antagonist-bound nuclear receptors [25,26,29]. In that perspective, phospholipids would be analogous to pharmacological chaperone, which are small molecules that help to maintain the correct folding of mutant proteins [40]. This interpretation is in agreement with a previous proposition, based on an information-theoretic approach, according to which the ligand-contacting residues perform a dual role: a functional role in ligand recognition and a structural role as core residues [41]. Thus, the structure of the LBP of USP was highly plastic during insect evolution, adopting three of the four states that are known for nuclear receptors: nutritional sensor (basal insects), real orphan (Hemiptera, Coleoptera) and receptor with a constitutive activity (Mecoptera). The first evolutionary transition involved a loss of function, from a sensor to an orphan receptor [8]. Our results highlight the second transition, from an orphan receptor to a constitutive receptor, which could be seen as a gain of function.

We also detected positive selection in the dimerisation interface during the emergence of Mecoptera. This episode of positive selection is correlated with a loss of ability to homodimerise for both ECR and USP. Non-Mecoptera ECRs are able to homodimerise and are active without USP [42–44]. By contrast, even in presence of their ligand, the ECR LBD monomers of *Heliothis* and *Drosophila* are very unstable in solution and need USP for solubilisation and stabilisation [45,46]. Similarly, non-Mecoptera USPs are able to form homodimers, whereas Mecoptera USPs are not able to do so [7,42,47,48]. It seems therefore that Mecoptera USP acts as a chaperone-like partner for ECR stabilisation. We have shown previously that, during the emergence of Mecoptera, a relaxation of evolutionary constraints occurred at a site that affects homodimerisation but not heterodimerisation [8,49]. This residue is the glutamine Gln457 in helix H10, for which we provide here evidence for strong positive selection on the branch leading to Mecoptera (Figure 5A). This result confirms our first hypothesis proposing that substitution of this site located in the core of the dimerisation surface may have induced a loss of homodimerisation for USP. Positive selection was found at those three sites in an independent study [50].

Importantly, if the homodimerisation ability has been lost in Mecoptera, on the contrary, the heterodimerisation between ECR and USP has been conserved by all arthropods and was even reinforced in Mecoptera. How can we explain the maintenance of the functional interaction between ECR and USP, despite the challenging burst of mutations that occurred in the stem lineage of Mecoptera? The heterodimerisation between ECR and USP depends largely on a core of residues localised in helices H9 and H10 of the LBDs. While the sequence of this interface is well conserved in all insects, in Mecoptera, the loop L8-9 of USP interacts with the helix H7 of ECR, creating a symmetric and larger heterodimerisation surface [8]. This novel surface is a consequence of the Mecoptera specific structural divergences of USP, where the loop L1-3 and the presence of a phospholipid, which is in tight contact with L1-3 and H3, induce a 15°-torsion of a sub-domain of USP LBD towards ECR. If the maintenance of heterodimerisation depends on this new surface, then positive selection should be detected at this specific enlargement on the stem lineage of Mecoptera. We have confirmed this prediction for two sites of USP: Arg426 in helix H9 and Arg399 in helix H7 (Figure 5A). The arginine Arg426 is involved in the Mecoptera specific enlargement of the heterodimerisation surface, while the Arg399 was predicted to form an electrostatic interaction with ECR in the ancestral Mecoptera model [8]. In addition, we detected positive selection at the site Asp452, which is located in the core interface and makes contacts with two residues of the helix H9 of ECR. It is thus possible that this substitution played a role in the reinforcement of the ECR-USP heterodimerisation. If we assume that the reinforcement of ECR-USP heterodimerisation is a selectively favoured phenotype, then every amino acid replacement causing a slight increase in interaction would be selectively favoured. However, once binding is optimum, no further changes will be favoured. Thus, a phase of purifying selection would follow an initial phase of positive selection. We have confirmed this expectation by showing that the same four positions described above became under purifying selection during the diversification of Mecoptera. In conclusion, the dimerisation interface of USP has experienced modifications of selective constraints during the emergence of Mecoptera, followed by a phase of constrained evolution.

Although our results suggested that coevolution has maintained the interaction between ECR and USP, we could not detect any significant modification of evolutionary constraints in the dimerisation surface of ECR [8]. Recent random-sites model analyses confirmed the absence of positive selection in ECR [50]. It seems therefore that the helix H7 of ECR, which is implicated in the new Mecoptera interface with USP, was able to adapt to its changing partner in a manner independent of

selection. In order to explain the evolution of this domain, it may be necessary to explore the role of molecular drive, which includes all the DNA turnover mechanisms that change the genetic composition of a population, independently of selection and drift [51]. There is indeed a renewed interest in the evolutionary forces, other than natural selection or drift, that can induce coevolution between functionally interacting genetic systems in order to maintain essential functions [15,52]. However, interestingly, six residues with elevated ω were identified in the core interface of Mecoptera ECR, suggesting that the dimerisation core of ECR, like that of USP/RXR, may also be under relaxed constraint in the Mecoptera clade [50].

Within the family of nuclear receptors, evidence for positive selection has been reported in a few other cases, such as PXR (NR1I2), CAR (NR1I3) and ER α (NR3A1) [33,53,54]. For example, positive selection was detected at amino acids involved in ligand binding of the frog PXR, which have lost broad specificity for ligands and gained efficient activation by endogenous amphibian-specific benzoates [55]. A recent study has shown that RXR, the USP homolog, was under positive selection immediately following the first round (1R) and second round (2R) of whole genome duplications in vertebrates [34]. Most of the positive selected sites were located either in the LBD or in the N-terminal region, which contain activation domains (AF-1 and AF-2) and phosphorylation sites that control RXR expression. During this evolutionary transition, it seems that positive selection did not affect the LBP or the dimerisation interface of RXR (except for one site of helix H9 on the post-duplication branch leading to RXR γ). The authors thus conclude: "The presence of positive selection/accelerated rates in the paralogs, coupled with the evidences of altered expression of paralogous genes explains the preservation of the additional RXR copies following genome duplication" [34]. It would be interesting to analyse the evolution of the N-terminal region of USP and ECR, but the structure and the function of these highly divergent domains are largely unknown in insects.

Episodic evolution of USP

In conclusion, our results allow to define three phases of evolution for the LBD of USP in Mecoptera. We can now propose a model to explain this episodic mode of evolution. We have previously proposed that the ancestral state of USP LBD was characterised by the absence of a LBP and the ability to form homodimers as well as heterodimers with ECR [7,8]. During the emergence of Mecoptera, an unexplained acceleration of evolutionary rate was probably responsible for the insertion of 11 residues in the loop L1-3. This insertion is the primordial event in our hypothesis, because it changed the

structure of the LBD in two ways. First, it increased the size of the heterodimer interface. Second, it restored a LBP. However, this new pocket is very large and filled with phospholipids, which is unusual for a nuclear receptor of the NR2B group. These modifications allowed the exploration of new functions, as revealed by the detection of positive selection at these two regions during the first phase of Mecoptera evolution. The new LBP probably allowed the maintenance of a correct folding for the whole LBD, while the new extended surface of dimerisation was associated with a reinforcement of the obligatory partnership between ECR and USP, at the expense of homodimerisation. Therefore, the correlation between positive selection and functional innovations suggests that the USP LBD underwent molecular adaptation during the emergence of Mecoptera. One limit of this kind of analysis is the difficulty to distinguish between positive selection and relaxation of negative selection [11]. However, the availability of crystal structures and functional data provide *a priori* biological hypotheses regarding the amino acids upon which positive selection is expected. Once the new functions are effective, no further changes will be favoured. Thus, the initial phase of positive selection should be followed by a second phase of purifying selection. In agreement with this hypothesis, we found that the new functional domains were under strong negative selection during the subdivision of Mecoptera between Amphiesmenoptera and Antliophora. This second phase of evolution explains the conservation of the new surfaces among Mecoptera species. Finally, we could characterise a third phase of evolution during the subdivision of Diptera into Brachycera and Nematocera. On this branch, USP LBD evolved mainly under a relaxation of negative selection, with neutrality at some sites. However, the new Mecoptera functional domains remained highly constrained (except the tyrosine Y372 in the LBP of Diptera). The alternation of selective and conservative periods during the transformation of USP LBD in insects provides an example of the episodic model of protein evolution, a major mechanism recently revealed by application of new methods at the genomic level [12,56,57].

Conclusions

The function of the heterodimer ECR-USP in the ecdysone response is conserved in all insects, including Mecoptera [58]. Therefore, although we cannot exclude the possibility that other biological functions have changed with USP, it seems that adaptation of this protein allowed maintenance rather than innovation. It is an important property of living organisms to remain constant in the face of change. The concept of canalisation describes the capacity of a population to produce the same phenotype despite genetic or environmental

variability. At the molecular level, the capacity of a protein to restore functionality after a mutational event could be qualified as “molecular resilience”. In the case of USP, the insertion in the loop L1-3 was likely the initial “trauma”, induced by a phase of increased mutation rate at the origin of Mecoptera. Resilience was probably dependent on the existence of pre-existing functionally neutral mutations that accumulated in the LBD of USP. Once the insertion occurred, only a specific combination of these permissive mutations became adaptive, because they could compensate for the structural changes induced by the insertion. Epistatic contacts between and within proteins are essential to allow a genetic network to assimilate new genotypes while remaining functional, because any loss of function could be deleterious [1,59]. Like canalisation, molecular resilience would therefore allow evolvability by accumulation of genetic diversity without modification of the phenotype. In that perspective, the molecular adaptation of USP could be seen as a process of resilience for the ecdysone receptor.

Methods

Insects

Drosophilidae species from the Gif laboratory were obtained from different origins: *Drosophila simulans* (La Réunion, France), *Drosophila mauritiana* (Mauritius), *Drosophila santomea* (Sao Tomé), *Drosophila teissieri* (Congo), *Drosophila malerkotliana* (Madagascar), *Drosophila (Zaprionus) tuberculatus* (Ivory Coast), *Drosophila limbata* (France), *Drosophila caribiana* (Martinique, France), *Drosophila funebris* (France).

The species of *Tribolium* were reared on wheat flour with 5% of dried yeast extract at room temperature. The USDA-ARS Grain Marketing and Production Research Centre (Manhattan, Kansas) provided all the species, except *Tribolium destructor*, which was provided by Durdica Ugarkovic (Ruder Bosković Institute, Zagreb, Croatia).

Concerning Blattaria species, *Blattella germanica* was from the strain maintained in the Institute of Evolutionary Biology (CSIC-UPF), in Barcelona, whereas *Blatta orientalis*, *Periplaneta americana*, *Periplaneta japonica*, *Periplaneta australasiae*, *Periplaneta brunnea* and *Periplaneta fuliginosa*, were kindly provided by Pierre Deleporte (Station Biologique de Paimpont, Université de Rennes). Blattarian specimens were maintained in the dark at $30 \pm 1^\circ\text{C}$ and 60-70% RH prior to RNA extraction.

Cloning and sequencing

We isolated 43 new sequences using PCR with degenerated primers. Most of the partial sequences for USP encompass regions from the beginning of the DNA binding domain (zinc-finger 1 or zinc-finger 2) to the last helix (H12) of the LBD. Sequences of the mitochondrial

cytochrome oxidase subunit 1 (COX1, used as a reference) are almost complete, including the twelve membrane-spanning helices. Available sequences were obtained from GenBank. All accession numbers are shown on Additional file 1: Table S1.

Genomic DNAs of Drosophilidae were prepared as described previously [60] from either live stocks, or frozen or ethanol-preserved flies from the Gif LEGS laboratory. Partial *usp* (intronless in *Drosophila*) and *cox1* sequences were amplified by PCR using a set of primers listed in Additional file 10: Table S7. For *usp*, two rounds of nested amplifications were performed with Amplitaq Gold (Applied Biosystems) in a Trioblock device (Biometra) as follows: first round: primers U50dbis/U31dbis; 45 cycles; annealing temperature 60°C; elongation for 1mn45 at 72°C; second round: primers U51dbis/U33dbis; annealing temperature 60°C; elongation for 1mn30 at 72°C. Amplicons of the correct size were purified and cloned into the pGEM-T easy vector (Promega). For *cox1*, samples were amplified with primer COIDIR (forward) combined with various reverse primers using the following conditions: 45 cycles, annealing temperature 50°C; elongation for 1mn30 at 72°C. Purified amplicons were sequenced directly without cloning. For some species, the data were completed with predicted sequences provided by the “*Drosophila* 12 Genomes Consortium” [61].

Sequences for *Tribolium* USP were obtained exactly as described previously [5], using RT-PCR with degenerated primers. Sequences for cytochrome oxidase I (COX1) were obtained by PCR on total DNA. Primers are listed in Additional file 10: Table S7. All combinations of primers were used in semi-nested PCR amplifications in 50 μ l volume with 10 mM Tris-HCl pH = 8.3, 50 mM KCl, 1.5 mM MgCl₂, 0.25mM of each dNTP, 2.5 U AmpliTaq Gold DNA polymerase (Applied Biosystems) and 300 ng of each primer. Reactions were performed in a Biometra T3 Thermocycler, using the following protocol: 2 min at 95°C; cycles 1–5: 94°C 1 min, 45°C 1 min, 72°C 1 min; cycles 6–10: 94°C 1 min, 50°C 1 min, 72°C 1 min; cycles 11–30: 94°C 1 min, 55°C 1 min, 72°C 1 min; followed by terminal elongation for 10 min at 72°C. PCR products were cloned into a TA cloning vector (Invitrogen) and transformed into competent cells according to the manufacturer’s instructions. Sequencing reactions were performed using the ABI PRISM BigDye Terminator v1.1 cycle sequencing kit (Applied Biosystems).

In the case of Blattaria, total RNA was isolated with the GenElute™ Mammalian Total RNA kit (Sigma) and cDNAs were obtained as described [62]. For RXR amplifications, primers based on the amino acid motif conserved in all RXR/USP sequences located upstream from the DNA-binding domain (DBD) and from the ligand-binding domain (LBD) were used. PCR amplifications

were performed using, as a template, cDNA generated by reverse transcription from polyA+ RNA obtained from the different species under study. Amplifications were carried out for 30 cycles at 94°C for 30 s, 52°C for 1 min, 72°C for 1 min, followed by a terminal elongation for 5 min at 72°C. The amplified fragments were sub-cloned into the pSTBlue-1 vector (Novagen) and sequenced. For cytochrome oxidase I (COX1) amplifications, the strategy was the same as before. Primers are listed in Additional file 10: Table S7.

Sequence dataset and insect’s phylogeny

In order to estimate the site-specific ratio of non-synonymous/synonymous substitutions (d_N/d_S or ω) through whole tree analyses, we used Maximum Likelihood methods and codon substitution models, implemented in the Codeml program from the PAML package [21,63]. Codons of insect sequences of *usp* and *cox1* genes were aligned using the program Muscle implemented in SeaView [64]. Positions with gaps were removed. All accession numbers are supplied in Additional file 1: Table S1. The predefined topology of insects phylogeny, imposed for the ML computations, is shown on Figure 1. This consensus tree is based on classical taxonomic data, as well as specific references concerning the following groups: Diptera [65], Drosophilidae [16,66], Lepidoptera [67,68], Tenebrionidae [17], Blattaria [18,19,69] and insects [70-72]. We carried out statistical analyses with the R software [73].

Detection of selective constraints

In a first set of analyses, we characterized the pattern of site-specific d_N/d_S ratios along *usp* sequences within two non-Mecopterida (5 Tenebrionidae, 7 Blattaria) and three Mecopterida groups (16 Drosophilidae, non-Drosophilidae 13 Diptera and 6 Lepidoptera), based on the alignment of 283, 279, 377, 346 and 398 codons, respectively. In order to detect positive selection within these groups, we compared the likelihood of two pairs of site models: M1a/M2a and M7/M8 [22,74]. The null model M1a (nearly neutral) assumes two classes of sites, with either $\omega_0 \leq 1$ (negative or purifying selection) or $\omega_1 = 1$ (neutral). The alternative model M2a (selection) adds a third class of sites with $\omega_2 \geq 1$ (positive selection). The null model M7 assumes that ω ratios vary among codons, following a β distribution in the interval $0 < \omega < 1$. The alternative model M8 adds an extra class of sites under neutral evolution or positive selection with $\omega_S \geq 1$. We fixed ten categories of sites for the discrete approximate of the β distribution. This high number of categories allows to use the Bayesian approach to calculate the posterior probabilities for site classes and expected site-specific d_N/d_S values [74]. The best codon model for each group was run ten times, using different initial values

for parameters, in order to test for convergence during the process of likelihood maximization by the Codeml algorithm. β distributions were compared by Kolmogorov-Smirnov tests. The average evolutionary rate of USP protein was compared between Drosophilidae, Tenebrionidae and Blattaria by using mean d_N of COX1 as a reference [20]. Site-specific d_N/d_S values estimated along the *usp* sequence were projected onto the 3D protein structure of the USP LBD domain of *Drosophila* (1HG4), *Heliothis* (1R1K), or *Tribolium* (2NXX). Because saturation was detected among Diptera *usp*, we assessed the accuracy of the estimate of the β distribution function for this clade. For this, we re-estimated the β distribution parameters with the same Codeml analysis (under model M7) from 50 simulated alignments, which were generated with the Evolver program of the PAML package using the global parameters established for the real *usp* Diptera dataset (number of sequences, sequence length, topology and branch lengths of the tree, β distribution function for d_N/d_S , Ts/Tv ratio, codon usage) (Additional file 4: Figure S1).

In a second type of analyses performed with branch-site models from the Codeml program, we defined three sets of *usp* sequences in order to detect site-specific changes in selective constraints along three internal branches that are relevant to the early history of Mecoptera: branch A is the stem lineage of Mecoptera, branch B delineates the two subgroups of Mecoptera (Amphiesmenoptera: Lepidoptera, Trichoptera and Antliophora: Diptera, Mecoptera, Siphonaptera), and branch C follows the subdivision of Diptera into Brachycera and Nematocera (Figure 1). We applied tests 1 and 2 from Zhang *et al.* [30] to assess whether relaxation or positive selection occurred along these branches. For this, the likelihood of three models were computed: the nearly neutral model (M1a), assuming constrained and neutral residues with no shifts along the tested branch; the modified branch-site model A, allowing some sites to undergo a $\omega_2 > 1$ for the predefined branch, thus considering possible positive selection; a third branch model with site-specific relaxation built in the same way as the modified branch-site model A but with $\omega_2 = 1$ fixed. As recommended by Zhang *et al.* [30], we considered χ^2 distributions with two and one degree(s) of freedom for the null distribution of the likelihood ratio tests 1 and 2 (respectively), which guarantees a conservative test of positive selection (test 2). Although tests 1 and 2 are claimed to be robust to the large evolutionary distances [21], we controlled false positive rate with a simulation methodology as previously proposed by Studer *et al.* [13]. For this, tests 1 and 2 were applied on 100 alignments generated with Evolver under the null model of the test (M1a and modified branch-site model A with $\omega_2 = 1$) using the global parameters established for the real dataset (branch A, B, or C). Lastly, we

identified sites that experienced a selective shift along the three branches of interest. To this end, we used the Bayesian procedure implemented in Codeml to calculate the posterior probability that a given site is under relaxation or under positive selection [74]. We projected these probabilities onto the 3D structures of USP DBD (2HAN) and LBD domains of *Drosophila* (1HG4).

Additional files

Additional file 1: Table S1. Accession numbers of sequences used for alignments.

Additional file 2: Table S3. d_N/d_S of *usp* in Drosophilidae, Diptera, Lepidoptera, Tenebrionidae and Blattaria.

Additional file 3: Table S4. Standard errors [SE] associated to the parameters of the models fitted in the likelihood analysis reported in Table 1.

Additional file 4: Figure S1. Robustness of d_N/d_S estimates to saturation in Diptera *usp* dataset. The accuracy of the estimate of the β distribution function was tested by re-performing the Codeml analysis (under model M7) on 50 simulated alignments, which were generated with Evolver in the PAML package using the global parameters established for the real *usp* Diptera dataset (number of sequences, sequence length, topology and branch lengths of the tree, β distribution function for d_N/d_S , Ts/Tv ratio, codon usage). Bold lines are the same as in Figure 2: Drosophilidae: red; Diptera: purple; Lepidoptera: orange; Tenebrionidae: green; Blattaria: blue. A thin pink line was added for each one of the 50 re-analysis of the simulated alignments.

Additional file 5: Table S5. Comparison between *usp* and *cox1* substitution rates.

Additional file 6: Figure S2. Evolutionary rates on the surface of USP structures in four extant groups of insects. Site-specific d_N/d_S were projected onto the crystal structure of the USP LBD domain of *Tribolium* (2NXX, ECR-USP) for Blattaria (A) and Tenebrionidae (B), of *Drosophila* (1HG4, USP) for Diptera (C) and of *Heliothis* (1R1K, ECR-USP) for Lepidoptera (D). The values are distributed along a colour scale from blue (low d_N/d_S) to red (high d_N/d_S). Sequences not available for the estimation of evolutionary rates are in white. Views on the left show the side of the LBD that is near ECR, while views on the right show the same structure after a rotation of 180°.

Additional file 7: Table S6. Posterior probabilities for each site to belong to the site-class under positive selection (along branch A or branch B) or relaxation (along branch C).

Additional file 8: Figure S3. Selection in the putative coactivation surface during the radiation of Mecoptera. Posterior probabilities for each site to belong to the site-class under positive selection (along branch A or branch B) or relaxed evolution (along branch C) were projected onto the crystal structure of the USP LBD domain of *Drosophila* (1HG4). Probabilities are distributed along a colour scale from yellow ($p=0$) to red ($p=1$). Sequences not available for the estimation of evolutionary rates are in white. (A) Branch A, stem lineage of Mecoptera. (B) Branch B, subdivision between Amphiesmenoptera (Lepidoptera, Trichoptera) and Antliophora (Diptera, Mecoptera, Siphonaptera). (C) Branch C, subdivision of Diptera into the suborders Brachycera and Nematocera. (D) The coactivation surface, defined by homology with RXR, is shown in pink. The helix H12 is shown in green.

Additional file 9: Table S2. Relative synonymous codon usage (RSCU).

Additional file 10: Table S7. Degenerated primers used to clone *usp* and *cox1* genes.

Abbreviations

USP: Ultraspiracle; ECR: Ecdysone receptor; LBD: Ligand Binding Domain; LBP: Ligand Binding Pocket; DBD: DNA Binding Domain.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

AC, Fbo and VL designed the study. JLDL, OM, DM and FBo performed the cloning and sequencing experiments. AC carried out the detection of selective constraints. FBR and FBO participated in the analysis of sequences. TI participated in the analysis of 3D structures. AC and FBO analyzed the data, interpreted the results and wrote the manuscript. JLDL, TI, XB and VL critically revised the manuscript. All authors read and approved the final manuscript.

Acknowledgments

Durdica Ugarkovic for *Tribolium destructor* and *Tribolium brevicornis*. USDA-ARS Grain Marketing and Production Research Center (Manhattan, KS) for *Tribolium freemani*, *Tribolium confusum* and *Tribolium madens*. Pierre Deleporte (Station Biologique de Paimpont, Université de Rennes) for *Blatta orientalis* and the species of *Periplaneta*, *P. americana*, *P. japonica*, *P. australasiae*, *P. brunnea* and *P. fuliginosa*. Support for this research was provided by CNRS, ENS de Lyon, ANR, the Spanish Ministry of Science and Innovation (projects CGL2008-03517/BOS to X.B., and BFU2006-13212 and BFU2009-10571 to D.M.) and by the CSIC (project 2010TW0019 to X.B.). O.M. is recipient of a pre-doctoral research grant from Generalitat de Catalunya.

Author details

¹Institut de Génomique Fonctionnelle de Lyon (IGFL), Université de Lyon, Université Lyon 1; CNRS; INRA; Ecole Normale Supérieure de Lyon, 32-34 avenue Tony Garnier, Lyon 69007, France. ²Irstea, UR MALY, Lyon F-69336, France. ³UPR9034, Laboratoire Evolution, génomes et spéciation (LEGS), CNRS, Gif sur Yvette 91198, France. ⁴Institute of Evolutionary Biology (CSIC-UPF), Passeig Marítim de la Barceloneta 37, Barcelona 08003, Spain. ⁵Groupe de recherche "immunopathologie et maladies infectieuses (GRI), Université de la Réunion, Centre CYROI, Cyclotron Réunion Ocean Indien, Sainte Clotilde Ile de la Réunion 97491, France.

Received: 4 June 2012 Accepted: 25 September 2012

Published: 5 October 2012

References

1. Wagner A: **Robustness, evolvability, and neutrality.** *FEBS Lett* 2005, **579**(8):1772-1778.
2. Henrich VC: **The ecdysteroid receptor.** In *Comprehensive Molecular Insect Science*. Edited by Gilbert LI, Iatrou K, Gill SS. Oxford: Elsevier; 2005:243-285. vol. 3.
3. Spindler KD, Honl C, Tremmel C, Braun S, Ruff H, Spindler-Barth M: **Ecdysteroid hormone action.** *Cell Mol Life Sci* 2009, **66**(24):3837-3850.
4. Bonneton F, Zelus D, Iwema T, Robinson-Rechavi M, Laudet V: **Rapid divergence of the ecdysone receptor in Diptera and Lepidoptera suggests coevolution between ECR and USP-RXR.** *Mol Biol Evol* 2003, **20**(4):541-553.
5. Bonneton F, Brunet FG, Kathirithamby J, Laudet V: **The rapid divergence of the ecdysone receptor is a synapomorphy for Mecoptera that clarifies the Strepsiptera problem.** *Insect Mol Biol* 2006, **15**(3):351-362.
6. Bonneton F, Chaumot A, Laudet V: **Annotation of Tribolium nuclear receptors reveals an increase in evolutionary rate of a network controlling the ecdysone cascade.** *Insect Biochem Mol Biol* 2008, **38**(4):416-429.
7. Iwema T, Billas IM, Beck Y, Bonneton F, Nierengarten H, Chaumot A, Richards G, Laudet V, Moras D: **Structural and functional characterization of a novel type of ligand-independent RXR-USP receptor.** *EMBO J* 2007, **26**(16):3770-3782.
8. Iwema T, Chaumot A, Studer RA, Robinson-Rechavi M, Billas IML, Moras D, Laudet V, Bonneton F: **Structural and Evolutionary Innovation of the Heterodimerization Interface between USP and the Ecdysone Receptor ECR in Insects.** *Mol Biol Evol* 2009, **26**(4):753-768.
9. Philippe H, Casane D, Gribaldo S, Lopez P, Meunier J: **Heterotachy and functional shift in protein evolution.** *IUBMB Life* 2003, **55**(4-5):257-265.
10. Anisimova M, Liberles DA: **The quest for natural selection in the age of comparative genomics.** *Heredity* 2007, **99**(6):567-579.
11. Hughes AL: **Looking for Darwin in all the wrong places: the misguided quest for positive selection at the nucleotide sequence level.** *Heredity* 2007, **99**(4):364-373.
12. Studer RA, Robinson-Rechavi M: **Evidence for an episodic model of protein sequence evolution.** *Biochem Soc Trans* 2009, **37**(Pt 4):783-786.
13. Studer RA, Penel S, Duret L, Robinson-Rechavi M: **Pervasive positive selection on duplicated and nonduplicated vertebrate protein coding genes.** *Genome Res* 2008, **18**(9):1393-1402.
14. Baer CF, Miyamoto MM, Denver DR: **Mutation rate variation in multicellular eukaryotes: causes and consequences.** *Nat Rev Genet* 2007, **8**(8):619-631.
15. Galtier N, Duret L: **Adaptation or biased gene conversion? Extending the null hypothesis of molecular evolution.** *Trends Genet* 2007, **23**(6):273-277.
16. Tamura K, Subramanian S, Kumar S: **Temporal patterns of fruit fly (Drosophila) evolution revealed by mutation clocks.** *Mol Biol Evol* 2004, **21**(1):36-44.
17. Angelini DR, Jockusch EL: **Relationships among pest flour beetles of the genus Tribolium (Tenebrionidae) inferred from multiple molecular markers.** *Mol Phylogenet Evol* 2008, **46**(1):127-141.
18. Grandcolas P, D'haese C: **The phylogeny of cockroach families: is the current molecular hypothesis robust?** *Cladistics-the Int J Willi Hennig Soc* 2001, **17**(1):48-55.
19. Mukha D, Wiegmann BM, Schal C: **Evolution and phylogenetic information content of the ribosomal DNA repeat unit in the Blattodea (Insecta).** *Insect Biochem Mol Biol* 2002, **32**(9):951-960.
20. Gaunt MW, Miles MA: **An insect molecular clock dates the origin of the insects and accords with palaeontological and biogeographic landmarks.** *Mol Biol Evol* 2002, **19**(5):748-761.
21. Yang Z: **PAML: a program package for phylogenetic analysis by maximum likelihood.** *Comput Appl Biosci* 1997, **13**(5):555-556.
22. Yang Z, Bielawski JP: **Statistical methods for detecting molecular adaptation.** *Trends Ecol Evol* 2000, **15**(12):496-503.
23. Zhang J, Hu X, Lazar MA: **A novel role for helix 12 of retinoid X receptor in regulating repression.** *Mol Cell Biol* 1999, **19**(9):6448-6457.
24. Marimuthu A, Feng W, Tagami T, Nguyen H, Jameson JL, Fletterick RJ, Baxter JD, West BL: **TR surfaces and conformations required to bind nuclear receptor corepressor.** *Mol Endocrinol* 2002, **16**(2):271-286.
25. Billas IML, Mouliniér L, Rochel N, Moras D: **Crystal structure of the ligand-binding domain of the ultraspiracle protein USP, the ortholog of retinoid X receptors in insects.** *J Biol Chem* 2001, **276**(10):7465-7474.
26. Clayton GM, Peak-Chew SY, Evans RM, Schwabe JW: **The structure of the ultraspiracle ligand-binding domain reveals a nuclear receptor locked in an inactive conformation.** *Proc Natl Acad Sci U S A* 2001, **98**(4):1549-1554.
27. Hu X, Cherbas L, Cherbas P: **Transcription activation by the ecdysone receptor (EcR/USP): identification of activation functions.** *Mol Endocrinol* 2003, **17**(4):716-731.
28. Carmichael JA, Lawrence MC, Graham LD, Pilling PA, Epa VC, Noyce L, Lovrecz G, Winkler DA, Pawlak-Skrzecz A, Eaton RE, et al: **The X-ray structure of a hemipteran ecdysone receptor ligand-binding domain - Comparison with a Lepidopteran ecdysone receptor ligand-binding domain and implications for insecticide design.** *J Biol Chem* 2005, **280**(23):22258-22269.
29. Billas IM, Iwema T, Garnier JM, Mitschler A, Rochel N, Moras D: **Structural adaptability in the ligand-binding pocket of the ecdysone hormone receptor.** *Nature* 2003, **426**(6962):91-96.
30. Zhang J, Nielsen R, Yang Z: **Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level.** *Mol Biol Evol* 2005, **22**(12):2472-2479.
31. Kostka D, Hubisz MJ, Siepel A, Pollard KS: **The role of GC-biased gene conversion in shaping the fastest evolving regions of the human genome.** *Mol Biol Evol* 2012, **29**(3):1047-1057.
32. Clare EL, Kerr KC, von Konigslow TE, Wilson JJ, Hebert PD: **Diagnosing mitochondrial DNA diversity: applications of a sentinel gene approach.** *J Mol Evol* 2008, **66**(4):362-367.
33. Krasowski MD, Yasuda K, Hagey LR, Schuetz EG: **Evolutionary selection across the nuclear hormone receptor superfamily with a focus on the NR11 subfamily (vitamin D, pregnane X, and constitutive androstane receptors).** *Nucl Recept* 2005, **3**:2.
34. Philip S, Castro LF, da Fonseca RR, Reis-Henriques MA, Vasconcelos V, Santos MM, Antunes A: **Adaptive evolution of the Retinoid X receptor in vertebrates.** *Genomics* 2012, **99**:81-89.

35. Zhang Z, Burch PE, Cooney AJ, Lanz RB, Pereira FA, Wu J, Gibbs RA, Weinstock G, Wheeler DA: **Genomic analysis of the nuclear receptor family: new insights into structure, regulation, and evolution from the rat genome.** *Genome Res* 2004, **14**(4):580–590.
36. Pal C, Papp B, Lercher MJ: **An integrated view of protein evolution.** *Nat Rev Genet* 2006, **7**(5):337–348.
37. Perissi V, Rosenfeld MG: **Controlling nuclear receptors: the circular logic of cofactor cycles.** *Nat Rev Mol Cell Biol* 2005, **6**(7):542–554.
38. Raviscioni M, He Q, Salicru EM, Smith CL, Lichtarge O: **Evolutionary identification of a subtype specific functional site in the ligand binding domain of steroid receptors.** *Proteins* 2006, **64**(4):1046–1057.
39. Nowickij SM, Chithalen JV, Cameron D, Tyshenko MG, Petkovich M, Wyatt GR, Jones G, Walker VK: **Locust retinoid X receptors: 9-Cis-retinoic acid in embryos from a primitive insect.** *Proc Natl Acad Sci U S A* 2008, **105**(28):9540–9545.
40. Arakawa T, Ejima D, Kita Y, Tsumoto K: **Small molecule pharmacological chaperones: From thermodynamic stabilization to pharmaceutical drugs.** *Biochim Biophys Acta* 2006, **1764**(11):1677–1687.
41. Nagl SB, Freeman J, Smith TF: **Evolutionary constraint networks in ligand-binding domains: an information-theoretic approach.** *Pac Symp Biocomput* 1999, **4**:90–101.
42. Minakuchi C, Ogura T, Miyagawa H, Nakagawa Y: **Effects of the structures of ecdysone receptor (EcR) and ultraspiracle (USP) on the ligand-binding activity of the EcR/USP heterodimer.** *J Pestic Sci* 2007, **32**(4):379–384.
43. Nakagawa Y, Sakai A, Magata F, Ogura T, Miyashita M, Miyagawa H: **Molecular cloning of the ecdysone receptor and the retinoid X receptor from the scorpion *Liocheles australasiae*.** *Febs J* 2007, **274**(23):6191–6203.
44. Ogura T, Minakuchi C, Nakagawa Y, Smaghe G, Miyagawa H: **Molecular cloning, expression analysis and functional confirmation of ecdysone receptor and ultraspiracle from the Colorado potato beetle *Leptinotarsa decemlineata*.** *Febs Journal* 2005, **272**(16):4114–4128.
45. Henrich VC: **The ecdysteroid receptor.** In *Insect endocrinology*. Edited by Gilbert LI. Amsterdam, Boston: Academic Press, Elsevier; 2012:177–218.
46. Li C, Schwabe JWR, Banayo E, Evans RM: **Coexpression of nuclear receptor partners increases their solubility and biological activities.** *Proc Natl Acad Sci U S A* 1997, **94**(6):2278–2283.
47. Graham LD, Pilling PA, Eaton RE, Gorman JJ, Braybrook C, Hannan GN, Pawlak-Skrzecz A, Noyce L, Lovrecz GO, Lu L, et al: **Purification and characterization of recombinant ligand-binding domains from the ecdysone receptors of four pest insects.** *Protein Expr Purif* 2007, **53**(2):309–324.
48. Lezzi M, Bergman T, Henrich VC, Vogtli M, Fromel C, Grebe M, Przbilla S, Spindler-Barth M: **Ligand-induced heterodimerization between the ligand binding domains of the *Drosophila* ecdysteroid receptor and ultraspiracle.** *Eur J Biochem* 2002, **269**(13):3237–3245.
49. Vivat-Hannah V, Bourguet V, Gottardis M, Gronemeyer H: **Separation of retinoid X receptor homo- and heterodimerization functions.** *Mol Cell Biol* 2003, **23**(21):7678–7688.
50. Hult EF, Tobe SS, Chang BS: **Molecular evolution of ultraspiracle protein (USP/RXR) in insects.** *PLoS One* 2011, **6**(8):e23416.
51. Dover GA, Strachan T, Coen ES, Brown SD: **Molecular drive.** *Science* 1982, **218**(4577):1069.
52. Dover GA: **Evolution of genetic redundancy for advanced players.** *Curr Opin Genet Dev* 1993, **3**(6):902–910.
53. Krasowski MD, Yasuda K, Hagey LR, Schuetz EG: **Evolution of the pregnane x receptor: adaptation to cross-species differences in biliary bile salts.** *Mol Endocrinol* 2005, **19**(7):1720–1739.
54. Shyu C, Brown CJ, Ytreberg FM: **Computational study of evolutionary selection pressure on rainbow trout estrogen receptors.** *PLoS One* 2010, **5**(3):e9392.
55. Krasowski MD, Ni A, Hagey LR, Ekins S: **Evolution of promiscuous nuclear hormone receptors: LXR, FXR, VDR, PXR, and CAR.** *Mol Cell Endocrinol* 2011, **334**(1–2):39–48.
56. Domazet-Loso T, Tautz D: **An evolutionary analysis of orphan genes in *Drosophila*.** *Genome Res* 2003, **13**(10):2213–2219.
57. Savard J, Tautz D, Lercher MJ: **Genome-wide acceleration of protein evolution in flies (Diptera).** *BMC Evol Biol* 2006, **6**:7.
58. Bonneton F, Laudet V: **Evolution of nuclear receptors in insects.** In *Insect endocrinology*. Edited by Gilbert LI. Amsterdam, Boston: Academic Press; Elsevier; 2012:219.
59. Kondrashov AS, Sunyaev S, Kondrashov FA: **Dobzhansky-Muller incompatibilities in protein evolution.** *Proc Natl Acad Sci U S A* 2002, **99**(23):14878–14883.
60. Gloor G, Engels W: **Single fly DNA preps for PCR.** *Drosophila Information Service* 1992, **71**:148–149.
61. Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA, Kaufman TC, Kellis M, Gelbart W, Iyer VN, et al: **Evolution of genes and genomes on the *Drosophila* phylogeny.** *Nature* 2007, **450**(7167):203–218.
62. Cruz J, Martin D, Pascual N, Maestro JL, Piulachs MD, Belles X: **Quantity does matter. Juvenile hormone and the onset of vitellogenesis in the German cockroach.** *Insect Biochem Mol Biol* 2003, **33**(12):1219–1225.
63. Yang Z: **PAML 4: phylogenetic analysis by maximum likelihood.** *Mol Biol Evol* 2007, **24**(8):1586–1591.
64. Galtier N, Gouy M, Gautier C: **SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny.** *Comput Appl Biosci* 1996, **12**(6):543–548.
65. Yeates DK, Wiegmann BM: **Congruence and controversy: toward a higher-level phylogeny of Diptera.** *Annu Rev Entomol* 1999, **44**:397–428.
66. Remsen J, O'Grady P: **Phylogeny of Drosophilinae (Diptera: Drosophilidae), with comments on combined analysis and character support.** *Mol Phylogenet Evol* 2002, **24**(2):249–264.
67. Weller SJ, Friedlander TP, Martin JA, Pashley DP: **Phylogenetic studies of ribosomal RNA variation in higher moths and butterflies (Lepidoptera: Ditrysia).** *Mol Phylogenet Evol* 1992, **1**(4):312–337.
68. Regier JC, Mitter C, Friedlander TP, Peigler RS: **Re: Phylogenetic relationships in Sphingidae (Insecta: Lepidoptera): initial evidence from two nuclear genes.** *Mol Phylogenet Evol* 2001, **20**(2):311–316.
69. Lo N, Bandi C, Watanabe H, Nalepa C, Beninati T: **Evidence for cocladogenesis between diverse dictyopteran lineages and their intracellular endosymbionts.** *Mol Biol Evol* 2003, **20**(6):907–913.
70. Kristensen NP: **Phylogeny of insect orders.** *Annu Rev Entomol* 1981, **26**:135–157.
71. Whiting MF, Carpenter JC, Wheeler QD, Wheeler WC: **The streptisera problem: phylogeny of the holometabolous insect orders inferred from 18S and 28S ribosomal DNA sequences and morphology.** *Syst Biol* 1997, **46**:1–68.
72. Wiegmann BM, Trautwein MD, Kim JW, Cassel BK, Bertone MA, Winterton SL, Yeates DK: **Single-copy nuclear genes resolve the phylogeny of the holometabolous insects.** *BMC Biol* 2009, **7**:34.
73. R Development Core Team: **R: A language and environment for statistical computing.** Vienna, Austria: R Foundation for Statistical Computing; 2009.
74. Yang Z, Wong WS, Nielsen R: **Bayes empirical bayes inference of amino acid sites under positive selection.** *Mol Biol Evol* 2005, **22**(4):1107–1118.

doi:10.1186/1471-2148-12-199

Cite this article as: Chaumot et al.: Molecular adaptation and resilience of the insect's nuclear receptor USP. *BMC Evolutionary Biology* 2012, **12**:199.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

