

# Molecular Evolution of Ultraspiracle Protein (USP/RXR) in Insects

Ekaterina F. Hult<sup>1</sup>, Stephen S. Tobe<sup>1</sup>, Belinda S. W. Chang<sup>1,2,3\*</sup>

**1** Department of Cell and Systems Biology, University of Toronto, Toronto, Ontario, Canada, **2** Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, Ontario, Canada, **3** Centre for the Analysis of Genome Evolution and Function, University of Toronto, Toronto, Ontario, Canada

## Abstract

Ultraspiracle protein/retinoid X receptor (USP/RXR) is a nuclear receptor and transcription factor which is an essential component of a heterodimeric receptor complex with the ecdysone receptor (EcR). In insects this complex binds ecdysteroids and plays an important role in the regulation of growth, development, metamorphosis and reproduction. In some holometabolous insects, including Lepidoptera and Diptera, USP/RXR is thought to have experienced several important shifts in function. These include the acquisition of novel ligand-binding properties and an expanded dimerization interface with EcR. In light of these recent hypotheses, we implemented codon-based likelihood methods to investigate if the proposed shifts in function are reflected in changes in site-specific evolutionary rates across functional and structural motifs in insect USP/RXR sequences, and if there is any evidence for positive selection at functionally important sites. Our results reveal evidence of positive selection acting on sites within the loop connecting helices H1 and H3, the ligand-binding pocket, and the dimer interface in the holometabolous lineage leading to the Lepidoptera/Diptera/Trichoptera. Similar analyses conducted using EcR sequences did not indicate positive selection. However, analyses allowing for variation across sites demonstrated elevated non-synonymous/synonymous rate ratios ( $d_N/d_S$ ), suggesting relaxed constraint, within the dimerization interface of both USP/RXR and EcR as well as within the coactivator binding groove and helix H12 of USP/RXR. Since the above methods are based on the assumption that  $d_S$  is constant among sites, we also used more recent models which relax this assumption and obtained results consistent with traditional random-sites models. Overall our findings support the evolution of novel function in USP/RXR of more derived holometabolous insects, and are consistent with shifts in structure and function which may have increased USP/RXR reliance on EcR for cofactor recruitment. Moreover, these findings raise important questions regarding hypotheses which suggest the independent activation of USP/RXR by its own ligand.

**Citation:** Hult EF, Tobe SS, Chang BSW (2011) Molecular Evolution of Ultraspiracle Protein (USP/RXR) in Insects. PLoS ONE 6(8): e23416. doi:10.1371/journal.pone.0023416

**Editor:** Nikolas Nikolaidis, California State University Fullerton, United States of America

**Received:** April 19, 2011; **Accepted:** July 16, 2011; **Published:** August 25, 2011

**Copyright:** © 2011 Hult et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This study was supported by Natural Sciences and Engineering Council of Canada (NSERC) Discovery grants (SST and BSWC) and an Early Researcher Award to BSWC. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: belinda.chang@utoronto.ca

## Introduction

USP/RXR, a group II nuclear receptor and transcription factor which belongs to the steroid receptor superfamily, is found across a diversity of Metazoan species ranging from sponges to mammals [1]. First characterized in *Drosophila melanogaster*, USP is the insect homolog of vertebrate RXR [2,3]. In insects USP/RXR is involved in an array of functions including metamorphosis, reproduction, growth and development (see [4,5,6,7,8] for examples). USP/RXR is known primarily as the partner of the ecdysone receptor (EcR). In response to the binding of 20-hydroxyecdysone (20E) to EcR, heterodimerization between USP/RXR and EcR occurs and the complex then binds DNA to transactivate genes [9,10,11]. However, it is the controversy over whether or not USP/RXR binds juvenile hormone (JH) in insects, and is activated independent of EcR, which has led to the most debate regarding the function and evolution this receptor [12,13,14].

In vertebrates, 9-*cis* retinoic acid (9cRA) was identified as the high affinity ligand of the RXR receptor [15,16,17]. However, in insects no natural ligand has been conclusively identified, and

USP/RXR remains an orphan receptor. JH has been suggested as a candidate ligand [12], but only experimental evidence from the dipteran *D. melanogaster* supports this hypothesis. In cell lines expressing *Drosophila* USP, the application of JH III induces the transcription of a transfected promoter, suggesting that JH binds USP resulting in a functional outcome [13,18,19]. Fluorescence-binding assays have shown that *Drosophila* USP binds not only JH III but also the JH precursors farnesol, farnesoic acid, and methyl farnesoate [20]. However, JH does not appear to directly bind with USP/RXR in less derived insects such as the holometabolous *Tribolium castaneum* or the hemimetabolous *Locusta migratoria* [14,21]. There is also little evidence to suggest a high affinity for retinoids in insects [2,21]. However, recent displacement binding experiments have shown that USP/RXR may bind 9cRA and all-*trans* RA at the high nanomolar range in *L. migratoria* [22]. Given the small quantity of RA found in *Locusta*, the physiological significance of this *in vitro* finding is unclear.

Lineage specific variation is also evident in the structure of USP/RXR. A comparative analysis of structural data demonstrates key differences in the ligand-binding pocket (LBP) of USP/RXR in insects. Crystallography data have shown that Dipteran

and Lepidopteran USP share a conserved ligand-binding domain (LBD) with a large hydrophobic cavity capable of accepting a natural ligand [23,24,25]. In both the Diptera and Lepidoptera, USP copurified from the bacterial expression system with a phospholipid occupying the LBP. However, the identity of the endogenous ligand is unknown. In contrast, crystal structures of USP/RXR from two less derived insects, the red flour beetle *T. castaneum* and the sweet potato whitefly *Bemisia tabaci*, reveal a collapsed LBP [21,26]. Given the taxonomic relationships of insects from which USP/RXR has been crystallized, this data implies a derived open pocket in higher insects. Despite the open LBP, Dipteran and Lepidopteran USP may not function independently as ligand-dependent transcription factors. In these species the ligand-dependent activation function (AF-2) domain in  $\alpha$ -helix H12, is locked in an antagonistic conformation [23,24,25]. As a consequence of interactions between H3, H11–H12 and a highly conserved insertion in the loop connecting helices H1 and H3 (L1–3), residues in H12 occupy the coactivator groove (H3, loop L3–4 and H4), precluding the agonist conformation and preventing the binding of transcriptional coactivators [23,24,27].

These striking lineage-specific differences in both structure and function have sparked an interest in exploring the molecular evolution of insect USP/RXR, in part to understand the evolution of endocrine regulation, but also to aid in the design of selective pesticides which target the USP/RXR–EcR complex. The use of relative-rate tests revealed high divergence rates in the Lepidoptera and Diptera for both USP/RXR and EcR, particularly in the LBD of USP/RXR [21,27]. These results suggested that the heterodimer may have coevolved to become functionally divergent in the Holometabola. It was later shown that this event was unique to the Mecoptera, a suborder of insects which includes the Diptera, Trichoptera, Mecoptera, Siphonaptera and Lepidoptera [28,29]. More recently, this work has been extended to the heterodimer interface between USP/RXR and EcR. By reconstructing and modeling the ancestral Mecoptera heterodimer, Iwema *et al.* [30] demonstrated that an expanded dimerization surface was common to this branch. Furthermore, the data suggested that this enlarged surface was the result of torsion in the structure of USP/RXR caused by the position of loop L1–3 in these insects.

Based on experimental data from *Drosophila*, crystallography work from *Heliothis virescens*, and molecular evolutionary studies, a model for the functional evolution of USP/RXR has been proposed. Iwema *et al.* [21,30] and Tocchini-Valentini *et al.* [31] suggest the following three groups of USP/RXR: 1. The RXR type which evolved retinoid-binding early, a function retained in the Cnidaria, Mollusca, and Chordata; 2. The Non-Mecoptera type which lost ligand-binding function along an ancient arthropod lineage, a state found in insect orders such as the Hymenoptera, Coleoptera, Orthoptera and Dictyoptera; 3. Finally, the Mecoptera USP type which may have gained a novel ligand-binding function and an enlarged dimerization interface with EcR during the evolution of some higher insects.

To investigate these hypotheses of USP/RXR evolution, we implemented codon-based maximum likelihood methods to estimate the ratio of non-synonymous ( $d_N$ ) to synonymous ( $d_S$ ) substitutions across an insect phylogeny and independently along the Mecoptera lineage. This ratio serves as an indicator of selective constraint and such methods have been successfully used to detect positive selection, the signature of functional gain, in a variety of insect gene families and vertebrate nuclear receptors [32,33,34,35,36]. We also examined among-site variation in  $d_N/d_S$  across a dataset of Mecoptera and Non-Mecoptera insects using random-sites models to investigate changes in site-specific

evolutionary rates across key structural and functional motifs. Additionally, several of these analyses were repeated with a dataset of EcR sequences to examine evolutionary constraint in the heterodimer. Finally, we employed newly developed methods which allow the independent estimation of  $d_N$  and  $d_S$  in order to determine the effect among-site variation in  $d_S$  on the patterns of site-specific evolutionary rates observed under random-sites analyses.

## Methods

### Sequence data collection and dataset assembly

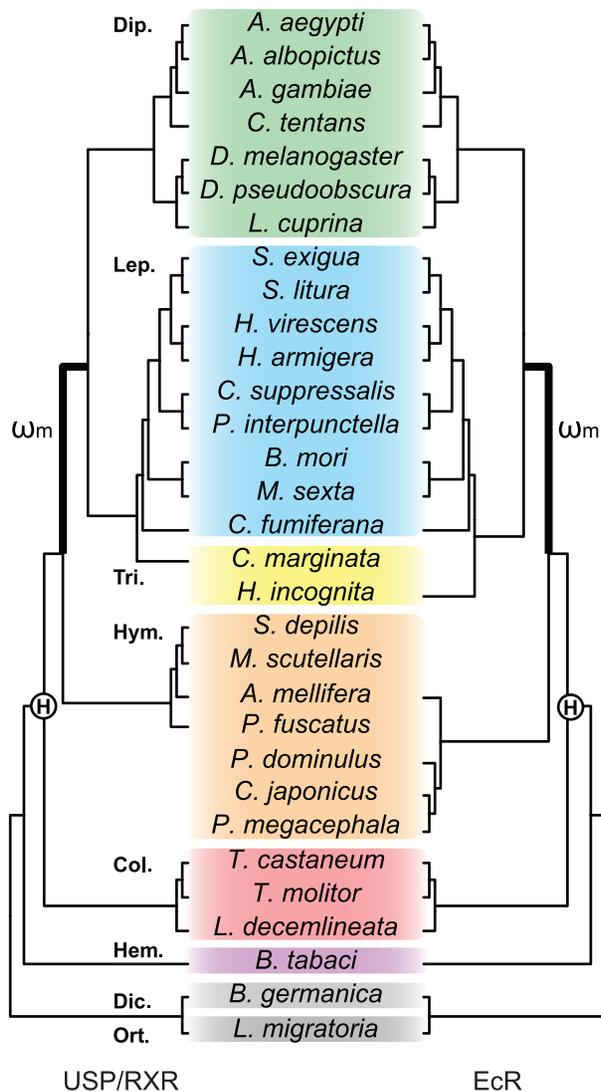
USP/RXR and EcR sequences were collected from literature and GenBank using a combination of BLAST and keyword searches (Table S1). Where possible, EcR-A isoform data were used in order to incorporate a larger portion of the protein in subsequent analyses. Both USP/RXR and EcR sequences were not available for all species, and in such cases sequences from the most closely related insects were used instead. Amino acid multiple sequence alignments for USP/RXR and EcR were constructed using ClustalW [37] as implemented in MEGA 4 [38] and adjusted by eye to ensure structural motifs were maintained. Poorly aligned regions and major gaps were deleted (see supporting Figure S1, S2).

To explore selective constraint across the entire phylogeny, both USP/RXR and EcR alignments were truncated such that only the well conserved LBD was included for use in branch and branch-sites analyses (termed USP/RXR LBD and EcR LBD datasets). To examine the variation in evolutionary rates across sites and across functional domains within each group, the USP/RXR and EcR alignments were each split into Mecoptera and Non-Mecoptera datasets (termed USP/RXR A/B-LBD and EcR A/B-LBD datasets for Mecoptera and Non-Mecoptera respectively).

### Estimation of evolutionary rates

In order to test for positive selection and examine changes in site-specific evolutionary rates codon-based likelihood methods were used to estimate  $d_N/d_S$  ( $\omega$ ) ratios across the USP/RXR gene. Under no selective pressure, sequences evolve neutrally, and this is indicated by  $\omega = 1$ , whereas  $\omega < 1$  indicates purifying selection, and  $\omega > 1$  positive selection [39,40]. To estimate  $\omega$  several branch-specific [41,42], branch-site [43,44], and random-sites [45,46] models were implemented using the codeml program of the PAML software package (version 4.2b; [47]). Random-sites models which allow for variation in  $d_S$  [48] were also implemented using the HyPhy software package (version 1.0; [49]). In addition, many of these analyses were also carried out for EcR to examine evolutionary constraint in the USP/RXR–EcR heterodimer.

The amino acid alignments described above for each of the six datasets were converted into nucleotide data. A tree reflecting current understanding among major insect lineages was used for both USP/RXR and EcR (Figure 1; [50,51] Insecta; [52] Coleoptera; [53] Lepidoptera; [54] Diptera; [55,56] Hymenoptera). Gene trees generated with our USP/RXR and EcR alignments, using both maximum-likelihood [57] and neighbor-joining [38,58] methods, were generally consistent with these known inter-species relationships (see supporting Figure S3). However, low branch support for many of the ordinal relationships resulted in trees with poor resolution of the Non-Mecoptera, therefore all analyses were only performed using the species trees. For analyses carried out on the USP/RXR and EcR A/B-LBD datasets, the trees were modified to include only Mecoptera or Non-Mecoptera taxa. However, this resulted in



**Figure 1. Phylogeny of insect species used in our analysis.** The topology of the trees used for the USP/RXR (left) and EcR (right) datasets are based on known taxonomic relationships. Species contained in the Diptera (Dip.) are shaded green, Lepidoptera (Lep.) blue, Trichoptera (Tri.) yellow, Hymenoptera (Hym.) orange, Coleoptera (Col.) red, Hemiptera (Hem.) purple, Dictyoptera (Dic.) and Orthoptera (Ort.) grey. The origin of holometabolous insects is indicated by an encircled bold H. The highlighted branch is the foreground lineage for detecting positive selection, where a separate  $d_N/d_S$  ratio was estimated for the Mecoptera ( $\omega_m$ ).  
doi:10.1371/journal.pone.0023416.g001

a tree bifurcated at the root with two clades when only the Mecoptera were considered, so *T. castaneum* was added as an outgroup to form a tripartition tree. Correspondingly, the *T. castaneum* sequence was added to both USP/RXR and EcR A/B-LBD Mecoptera datasets.

It is important to note that our dataset was composed of insect taxa which diverged over 140 million years ago [59], and that the saturation of synonymous substitution rates over such large distances may result in the estimation of  $d_N/d_S$  ratios higher than the actual value [60]. However, this would affect all estimations across all datasets, and our primary objective was to examine relative differences and patterns in substitution rates.

## Branch and branch-site models

To investigate the gain in function hypothesized to have occurred in Mecoptera USP/RXR, branch models were implemented which allowed for an additional  $\omega$  parameter along the lineage leading to the Mecoptera (Two-ratios), or alternatively along the lineage leading to the ancestor of Mecoptera and Hymenoptera. Codon frequencies were estimated using the F3x4 matrix and models were run from varying starting  $\omega$  values ranging above and below 1, or  $\kappa$  (relative transition to transversion rate ratio) values ranging from 0 to 5 in order to test for convergence. Likelihood ratio tests (LRTs; [61]) were then conducted to determine statistically significant differences between nested models. Models with the additional  $\omega$  parameter were compared against M0 where only one  $\omega$  value was estimated across the phylogeny. Since coevolution of USP/RXR and EcR has been suggested, the above branch models were also applied to the EcR LBD dataset.

Given the elevation in  $\omega$  identified by branch models, variation in selective pressure among codon sites in Mecoptera USP/RXR was examined using branch-site models applied to the USP/RXR LBD dataset. Branch-site models allow the use of a Bayes Empirical Bayes (BEB) analysis [62] to identify specific positively selected sites within the gene along a given branch. A model with Mecoptera designated as foreground, Model A alt, was compared against a stringent model for positive selection, Model A null, and a less stringent model, M1a. Positive selection is detected if 'Model A alt' is a better fit than 'Model A null' where classes of positively selected sites have  $\omega$  set to 1, whereas relaxed purifying selection (possibly suggestive of positive selection) is indicated if 'Model A alt' is a better fit than M1a which does not allow a class of sites with  $\omega > 1$  (referred to as test 2 and test 1 of positive selection by [44]). Sites identified as under positive selection by BEB analysis with a posterior probability (P) > 0.95 were subsequently mapped onto the 1.65 Å crystal structure of *H. virescens* USP (PDB 1G2N) or the 2.90 Å crystal structure of the *H. virescens* USP-EcR heterodimer (PDB 1R1K) using the VMD software package [63].

## Random-sites models

In order to compare among-site variation in  $d_N/d_S$  across functional domains of USP/RXR and EcR between Mecoptera and Non-Mecoptera insects, random-sites models were implemented using PAML [45,46]. Codon frequencies and initial starting values were the same as those used in the above branch and branch-site models. Comparison between M3, with three discrete classes of  $\omega$ , and null M0 tests for variation in selection pressure among sites, but does not explicitly incorporate positive selection. However, comparisons between both M2a and null M1a (discrete distribution), and M8 and null M7 (where  $\omega$  is drawn from a beta distribution), test for positive selection by allowing for an additional class of sites with  $\omega > 1$ . Posterior probabilities for site classes under sites models can be calculated either by the Naive Empirical Bayes (NEB) approach [45,46] or by the BEB approach [62] which considers sampling errors in the estimation of  $\omega$ . For models such as M8 where BEB data was available, the posterior mean  $\omega$  at each site was plotted using site class assignments as calculated by BEB, not NEB.

## Random-sites models that incorporate variation in $d_S$

To account for the effect of variation of synonymous substitution rates across sites, we implemented several codon selection analyses using the dNdSRateAnalysis program of the HyPhy software package [49]. Similar to M3 in PAML the variable nonsynonymous rates model, or 'Nonsynonymous',

implemented in HyPhy assumes a constant  $d_S$  ( $\alpha_S = 1$ ), but samples  $d_N$  ( $\beta_S$ ) values from a given rate distribution. However, the dual variable rates model, or ‘Dual’, estimates  $d_S$  ( $\alpha_S$ ) and  $d_N$  ( $\beta_S$ ) independently, sampling both from a given rate distribution [48]. In this study, both models were run using the MG94×REV core rate matrix with the GDD (general discrete distribution) rate distribution method using the ‘Independent Discrete’ setting. For each model three synonymous and nonsynonymous rate classes (3×3) were specified. All models were run several times using the randomized initial value option to find the global optimum. To test for site-to-site variation in  $d_S$  across both USP/RXR and EcR A/B-LBD datasets for the Mecoptera and Non-Mecoptera, nested Dual and Nonsynonymous models were compared using LRTs.

## Results

### Branch models

The current theory of USP/RXR functional evolution proposes that the ligand-binding function was lost in an ancient arthropod lineage followed by a subsequent gain in function along the Mecoptera lineage. Additionally, it has been suggested that Mecoptera USP/RXR acquired an expanded dimerization interface with EcR. We tested these hypotheses of functional gain by estimating evolutionary rates across a dataset composed of insect USP/RXR ligand-binding domain sequences (USP/RXR LBD dataset), using codon-based models of substitution (Figure 1). Likelihood scores and  $\omega$  ( $d_N/d_S$ ) values, as calculated by PAML, are shown in Table 1. Branch models implemented in PAML allow  $\omega$  to be freely estimated along specified foreground branches while all other background branches are constrained to the same  $\omega$  across the phylogeny (Two-ratios model). A branch model analysis, for which the Mecoptera lineage was set as foreground, demonstrated an elevated value ( $\omega_m = 0.166$ ) compared to the background ( $\omega = 0.040$ ), but cannot distinguish between relaxed constraint or positive selection at a subset of sites (Table 1). However, the likelihood ratio test (LRT) indicated that the additional parameter did not yield a significantly better fit for the data than the null model M0 where one  $\omega$  is estimated across the entire phylogeny ( $p = 0.347$ ).

Since coevolution of USP/RXR and EcR has been suggested, a branch model for which the Mecoptera lineage was set as foreground was also fit to a dataset composed of EcR ligand-binding domain sequences (EcR LBD dataset). Unlike USP/RXR, the foreground lineage demonstrated a decreased evolutionary rate ( $\omega_m = 0.008$ ) suggestive of stronger purifying selection (Table 1). However, the LRT again indicated that the additional parameter did not yield a significantly better fit for the data than M0 ( $p = 0.143$ ).

In a separate analysis, the lineage leading to the ancestor of Hymenoptera and Mecoptera was also freely estimated across both USP/RXR and EcR LBD sequences to determine if  $\omega$  was elevated in the branch preceding the Mecoptera. However, in both datasets,  $\omega$  was found to be below background ( $\omega = 0.002$  and  $\omega = 0.0001$ , respectively; data not shown). The added parameter along that branch also did not yield a statistically better fit than M0 in either dataset ( $p = 0.072$  and  $p = 0.084$ , respectively; data not shown). Overall, the branch model results, which only allow for a constant  $\omega$  parameter across sites, suggest a weak trend where evolutionary rates might be elevated in USP/RXR along the Mecoptera lineage.

### Branch-sites models

The insignificance of the branch models tests may have been a consequence of a lack of positive selection, or positive selection

acting only on a few sites within the gene along a given branch. Therefore branch-site models were implemented to estimate  $\omega$  for each site in the USP/RXR LBD dataset along the specified foreground lineage. Branch-site Model A alt demonstrated that a proportion of sites in the USP/RXR gene have  $\omega > 1$  in the Mecoptera lineage (Table 1). When compared to both M1a and branch-site Model A null, a more stringent test for positive selection, the LRT showed that the result was statistically significant ( $p = 2.174 \times 10^{-8}$  and  $p = 0.016$ , respectively; Table 1).

Using a Bayes Empirical Bayes (BEB) analysis in Model A alt, a class of sites with  $\omega > 1$  was identified (Table S2). Nine sites showed  $\omega > 1$  at a posterior probability (P)  $0.99 > P > 0.95$ , and four at  $P > 0.99$  (Table 1). These sites mapped onto important structural and functional regions of the *H. virescens* USP [24,25] crystal structure (Figure 2A). Of the 13 sites, one directly interacted with the ligand, and three others lay near ligand-binding sites (Figure 2B). Several sites lay within loop L1–3, and were involved in a hydrogen bond network with loop L11–12 and H3 which stabilizes the position of structural elements in Mecoptera USP/RXR (Figure 2C). Although only one site is known to form direct contact with EcR in the crystal structure of the heterodimer, several others fell within H9, a component of the dimerization core (Figure 2D). One site was located in the coactivator groove in loop L3–4, and two others lay in the region immediately adjacent to loop L5–S1. In addition, several of the sites with  $P < 0.95$  also participated in interactions with EcR or were located beside sites which do so (supporting Table S2). Overall, these results indicate significantly elevated  $\omega$  values along the Mecoptera lineage, indicating positive selection in regions of the protein important for ligand-binding, structural stability and dimerization.

### Random-sites models

In order to compare among-site variation in selection pressure across USP/RXR functional domains between Mecoptera and Non-Mecoptera insects, random-sites models were conducted using PAML with datasets for each group which spanned all domains (USP/RXR A/B-LBD datasets) (Table 2). M3 was the best fit for the Mecoptera and Non-Mecoptera datasets ( $p = 2.231 \times 10^{-127}$  and  $p = 5.063 \times 10^{-47}$ , respectively), suggestive of significant among-site variation in  $\omega$ . However, M3 is generally a better fit than the null M0 for most proteins as a single  $\omega$  value across all sites is unlikely. For both the Mecoptera and Non-Mecoptera USP/RXR A/B-LBD datasets, tests of positive selection were not significant. Neither M2a nor M8 were a better fit than the respective null models M1a and M7 ( $p = 1.0$  in both cases). Although no positively selected sites were identified in either Mecoptera or Non-Mecoptera USP/RXR,  $\omega$  was in fact, elevated at particular regions of the gene (Figure 3A, B). Site class assignments were generally consistent across models, with peaks in  $\omega$  occurring at the same codon sites, so only the results for M8 are presented (for M3 results see supporting Figure S4). Overall, more sites with elevated  $\omega$  values were observed in the Mecoptera compared to the Non-Mecoptera.

In the Mecoptera, there were a dramatic number of sites with increased  $\omega$  in the LBD and D domain compared to the A/B and C domains (Figure 3A). These sites clustered within H1, the amino-terminal of loop L1–3, loop L3–4, H4, H7, H9 and H12 of the LBD (see supporting Table S3 for details). Two sites were located beside ligand-binding sites and six other sites lay near residues which make dimerization contacts with EcR. Three sites in H12 lay near residues that interact with loop L1–3, H3 and the coactivator binding groove. Four sites lay in the coactivator binding groove, a region blocked in the Mecoptera as a consequence of contacts with H12 and loop L1–3 [23,24]. Several

**Table 1.** Parameter estimates for branch and branch-site models (LBD).

Model	np	lnL	$\kappa$	Parameter Estimates	Positively Selected Sites	LRT		
						Null	df	p-value
<b>USP/RXR</b>								
M0: one-ratio	53	-11110.02	1.537	$\omega = 0.041$	-			
Branch-specific models:								
Two-ratios	54	-11109.58	1.533	$\omega_0 = 0.040, \omega_1 = 0.166$	-	M0	1	0.347
Branch-site models:								
M1a <sup>a</sup>	54	-11097.30	1.544	$\omega_0 = 0.040, \omega_1 = 1.000$ $p_0 = 0.990, p_1 = 0.010$	-			
Model A null <sup>b</sup>	55	-11082.55	1.535	$\omega_0 = 0.039, \omega_1 = 1.000$ $\omega_{2a} = 1.000, \omega_{2b} = 1.000$ $p_0 = 0.814, p_1 = 0.009$ $p_{2a} = 0.175, p_{2b} = 0.002$	-			
Model A alt	56	-11079.65	1.536	$\omega_0 = 0.040, \omega_1 = 1.000$ $\omega_{2a} = 8.122, \omega_{2b} = 8.122$ $p_0 = 0.796, p_1 = 0.009$ $p_{2a} = 0.194, p_{2b} = 0.002$	30 sites: 222, 230, 231, 252, 253, 272, 302, 403, 411 (at $P > 0.95$ ) 296, 301, 324, 398 (at $P > 0.99$ )	M1a Br-s A null <sup>b</sup>	2 1	$2.174 \times 10^{-08*}$ 0.016*
<b>EcR</b>								
M0: one-ratio	53	-9701.18	1.661	$\omega = 0.029$				
Branch-specific models:								
Two-ratios	54	-9700.11	1.664	$\omega_0 = 0.029, \omega_1 = 0.008$		M0	1	0.143

NOTE – np is the number of parameters for each model, df is the degrees of freedom, while  $p$  is the proportion of sites in a given site class. Sites under positive selection are listed according to the site numbering of the *H. virescens* reference sequence, accession number AX383958.

<sup>a</sup>Null model for test 1 of positive selection from Zhang *et al.* [44].

<sup>b</sup>Null model for test 2 from Zhang *et al.* [44], a more stringent test of positive selection.

\*Significant ( $p$ -value < 5%).

doi:10.1371/journal.pone.0023416.t001

of the above sites also overlapped with those identified to be under positive selection along the Mecoptera lineage by branch-site analysis (Figure 3A, red squares). However, many sites differed between these types of analyses as branch-sites models compare the specified foreground branch to the background lineages, whereas random-sites models detect among-site rate variation within a clade. Unlike the Mecoptera, only a handful of sites with an elevated  $\omega$  were observed in the Non-Mecoptera USP/RXR dataset (Figure 3B; supporting Table S3). These were primarily located in loop L1–3, a region with extensive sequence variation among Non-Mecoptera species [14,64,65]. There were also a few sites with elevated  $\omega$  in H4, H9, H12 and the carboxy-terminal DNA-binding domain (C domain). Overall, among-site variation in the USP/RXR suggests that regions involved in dimerization and the blockage of the coactivator binding groove may be under relaxed constraint within the Mecoptera clade.

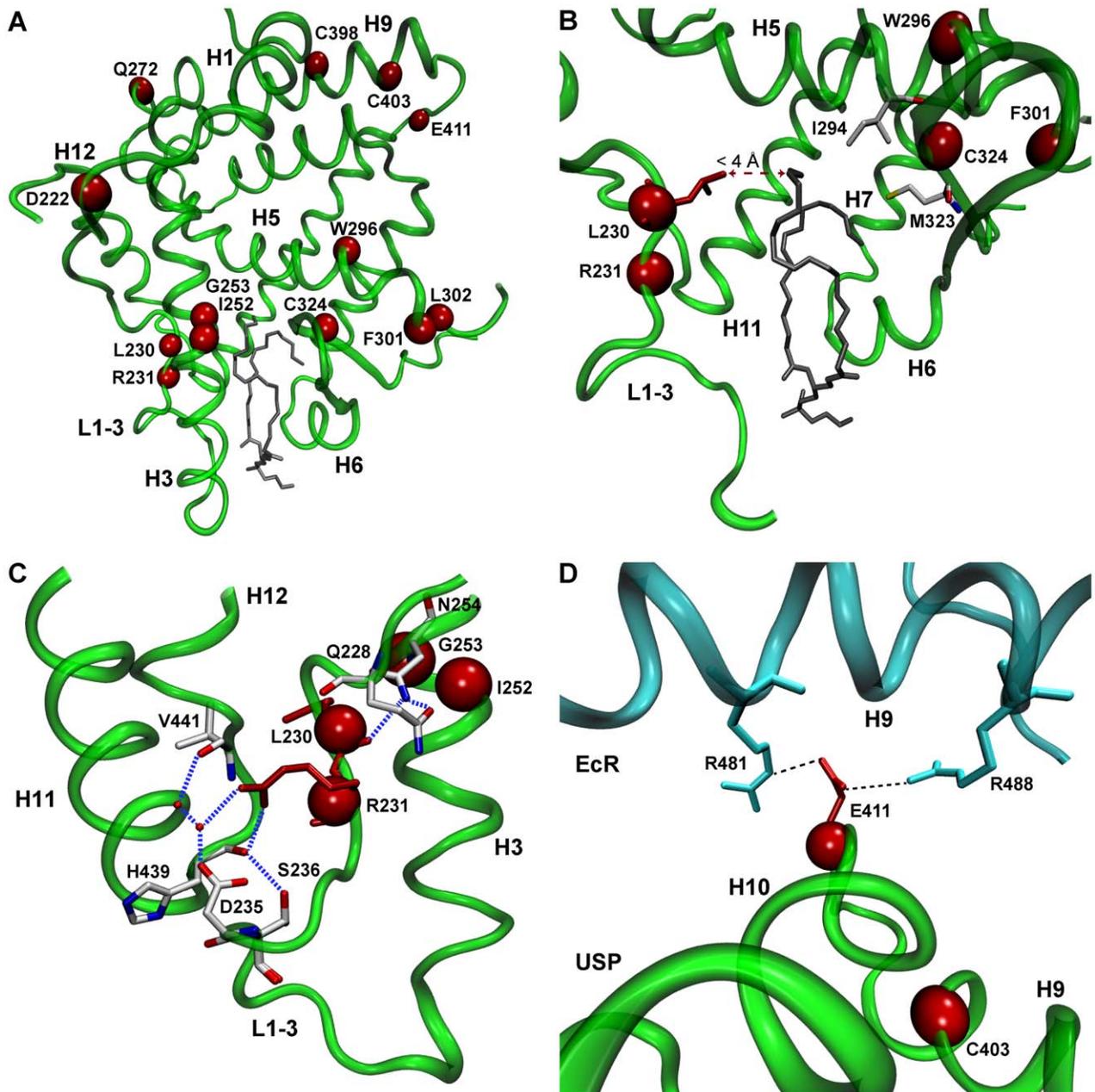
We also sought to examine among-site variation across EcR functional domains (EcR A/B-LBD datasets), as shifts in USP/RXR structure and function may also have affected the molecular evolution of EcR (Table 3). As with USP/RXR, M3 also yielded the best fit for both Mecoptera and Non-Mecoptera datasets ( $p = 3.320 \times 10^{-97}$  and  $9.273 \times 10^{-32}$ , respectively). No positively selected sites were identified and neither M2a nor M8 were a better fit than the null models M1a and M7 for either dataset ( $p = 1.0$  in both cases). As in the USP/RXR analyses,  $d_N/d_S$  site-profiles were plotted for EcR based on the M8 result (Figure 4A, B; see supporting Figure S4 for M3 results). Again, more sites with

elevated  $\omega$  values were observed in Mecoptera compared to the Non-Mecoptera.

In the Mecoptera there was a concentration of sites with elevated  $\omega$  in carboxy-terminal D domain and the dimerization core (Figure 4A; supporting Table S4). Several sites were located in loop L2–3, a region which varies in sequence and is not well modeled in insect crystal structures [21,25,26]. One site in the  $\beta$  sheet region was located between several ligand-binding sites. Six sites in H7, H9 and H10 were located near residues which make contact with USP/RXR. Three sites at the amino-terminus of H10 fell into a region associated with dimerization and ligand-binding specificity. Substitutions in this region confer *Aedes aegypti*-like sensitivity to ecdysone to *D. melanogaster* EcR, which is normally only responsive to 20E [66]. As in USP/RXR, there was far less among-site variation in  $\omega$  across Non-Mecoptera EcR (Figure 4B; supporting Table S4). Three sites in loop L9–10 and H10 align with the aforementioned region in *A. aegypti* which affects sensitivity to ecdysone. In addition, three other sites were located near ligand-binding sites. Overall, these results indicate that the dimerization core of EcR, like that of USP/RXR, may also be under relaxed constraint in the Mecoptera clade.

#### Random-sites models that incorporate variation in $d_S$

Traditional  $d_N/d_S$  methods estimate  $d_N$  and  $d_S$  as a ratio, and assume a constant  $d_S$  [45,46]. However, this may be an unrealistic assumption for some datasets, and synonymous substitutions are thought to be under selection in flies, nematodes and yeast [67].

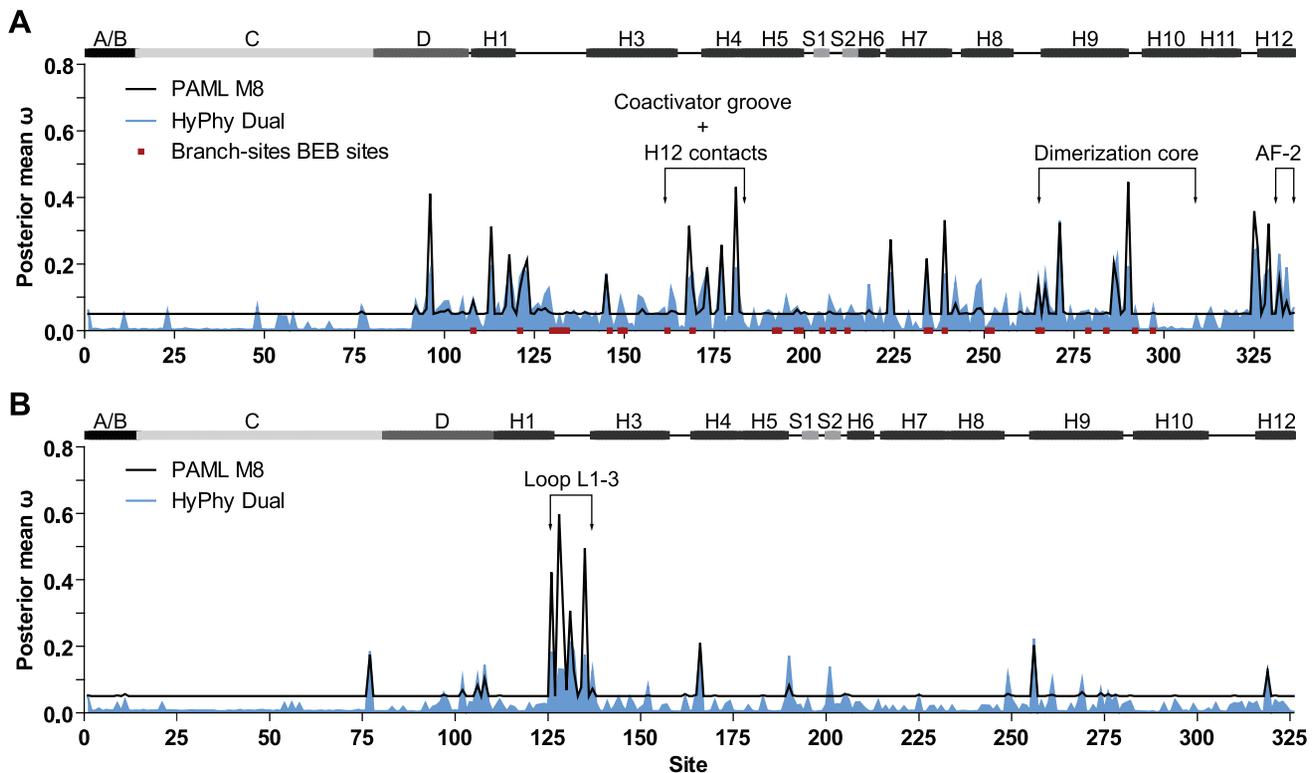


**Figure 2. Location of putative positively selected sites in Mecoptera USP/RXR.** (A) The distribution of sites inferred by PAML to be under positive selection ( $P > 0.95$ ) along the Mecoptera branch (red spheres) across the crystal structure of *H. virescens* USP (green) with the ligand shown in grey. (B) Positively selected sites (red spheres) located near ligand binding sites (white sidechains). Only L230 directly interacts with the phospholipid ligand (grey) via van der Waals (red dashes). (C) The involvement of sites L203 and R231 in a hydrogen bond network between H3, loop L1–3 and H12. Hydrogen bonds are indicated by dotted blue lines and waters by small red dots (adapted from [24]). The backbone and sidechains of residues not under positive selection are shown in white, with blue and red indicating nitrogen and oxygen respectively. (D) Polar interactions (dashed black lines) between the positively selected site E411 in USP loop L9–10 (green) and two arginine residues in H9 of the EcR (cyan). The images were created using PDB 1G2N and 1R1K, site numbering according to *H. virescens* USP (AX383958) and EcR (Y09009). doi:10.1371/journal.pone.0023416.g002

Recently, models have been developed to allow for the independent estimation of  $d_N$  and  $d_S$ . To examine the effect of among-site variation in  $d_S$  on the estimation of  $\omega$  in our datasets, we implemented several such models in the HyPhy software package (see supporting Tables S5, S6, S7 for parameter estimates). Overall, the results of HyPhy are similar to those of PAML (Figure 3 and 4, shaded blue). For all datasets the regions with elevated  $\omega$  were consistent between both methods, with some

variation in peak height. The baseline of the  $d_N/d_S$  site-profile plots appears elevated for the PAML M8 results compared to the HyPhy results because the former employs BEB methods, which account for sampling error in the maximum likelihood  $d_N/d_S$  (and proportion) estimates, whereas the latter does not.

For both USP/RXR and EcR A/B-LBD datasets in the Mecoptera site-to-site variation in  $d_S$  was significant when the Dual model compared to the Nonsynonymous null model where



**Figure 3. Posterior mean  $\omega$  at each amino acid site across the USP/RXR gene.** The values of  $\omega$  as estimated by M8 in PAML (black line) and the Dual model of HyPhy (shaded blue) are shown for each codon site across the gene for the Mecoptera (A) and Non-Mecoptera (B). All sites inferred to be under positive selection by branch-site analysis the along the Mecoptera branch are shown as red boxes on the x-axis in of the Mecoptera plot. A schematic of USP/RXR secondary structure is shown above both plots to illustrate the position of each functional domain (A/B, C, and D) as well as the helices (H1–H12) and  $\beta$  sheets (S1–S2) of the ligand-binding domain. The schematic for the Mecoptera and Non-Mecoptera genes are based of the crystal structures of *H. virescens* and *B. tabaci* USP/RXR, respectively. Site numbering is based on the USP/RXR alignment given in supporting figure S1.

doi:10.1371/journal.pone.0023416.g003

$d_S$  is fixed at all sites ( $p = 0.037$  and  $0.025$  respectively; Table 4). However, this was not the case with either dataset for the Non-Mecoptera in which significant variation in  $d_S$  among sites was not detected. Given the small size of the Non-Mecoptera datasets and the limited taxonomic sampling, we may lack the statistical power to estimate  $d_N$  and  $d_S$  accurately in this group. Overall the HyPhy results indicate that despite significant among-site variation in  $d_S$  within some datasets, the patterns in substitution rates across both the USP/RXR and EcR A/B-LBD datasets are the same when  $d_N$  and  $d_S$  are estimated independently.

## Discussion

Our results indicate that USP/RXR is under positive selection along the branch leading to the Mecoptera, with positively selected sites tending to be located in regions involved in ligand-binding, interactions with loop L1–3 and dimerization (Figure 2). Furthermore, random-sites analyses showed that  $\omega$  was elevated across the Mecoptera clade compared to the Non-Mecoptera for both USP/RXR and EcR (Figure 3, 4). Sites with elevated  $\omega$  tended to be concentrated in components of the dimerization interface, suggesting relaxed constraint in this region among the Mecoptera (Figure 3A, 4A). There were also several sites with elevated  $\omega$  in both proteins located near sites important for ligand-binding. Overall, these results are consistent with the acquisition of

functional gains with respect to ligand-binding and dimerization in USP/RXR in the Mecoptera.

Although higher relative substitution rates have been shown in the Mecoptera, our study represents the first instance of codon-based likelihood phylogenetic methods being used to estimate  $d_N/d_S$  ratios across the USP/RXR gene [21,27,28]. The elevated  $d_N/d_S$  values we observed along the Mecoptera lineage are consistent with high amino acid substitution rates reported previously [28]. Strikingly, the regions identified as under either positive selection or relaxed constraint play a major role in defining the structure of USP/RXR in the Mecoptera. In order to function as ligand-dependent transcription factors, nuclear receptors must undergo a shift in the position of H12 upon ligand binding to generate an interface for coactivators [68]. Residues in the AF-2 domain and the coactivator binding groove become accessible, allowing the binding of transcriptional coactivators [68,69,70]. However, the position of loop L1–3 in Mecoptera USP/RXR prevents shifts in the H12 position, regardless of ligand binding, thereby locking the receptor in an antagonist conformation [23,24].

These conformational changes are evident in the evolutionary history of USP/RXR. Studies using reconstruction and homology modeling of the ancestral Mecoptera USP/RXR have investigated the origin of an expanded interface with EcR; these novel contacts were a result of intramolecular epistasis whereby the position of loop L1–3 created torsion in the protein, shifting the

**Table 2.** Parameter estimates for USP/RXR gene (A/B-LBD).

Model	np	lnL	$\kappa$	Parameter Estimates	LRT		
					Null	df	p-value
<b>Mecoptera:</b>							
M0: one-ratio	35	-10594.82	1.532	$\omega = 0.035$			
M1a: neutral	36	-10531.81	1.622	$\omega_0 = 0.029, \omega_1 = 1.000$ $p_0 = 0.950, p_1 = 0.050$			
M2a: selection	38	-10531.81	1.622	$\omega_0 = 0.029, \omega_1 = 1.000, \omega_2 = 1.000$ $p_0 = 0.950, p_1 = 0.002, p_2 = 0.048$	M1a	2	1.000
M3: discrete ( $K=3$ )	39	-10297.50	1.564	$\omega_0 = 0.002, \omega_1 = 0.046, \omega_2 = 0.194$ $p_0 = 0.457, p_1 = 0.450, p_2 = 0.092$	M0	4	$2.231 \times 10^{-127*}$
M7: beta	36	-10299.37	1.557	$p = 0.417, q = 9.184$			
M8: beta& $\omega$	38	-10299.37	1.557	$p = 0.417, q = 9.185$ $p_0 = 0.99999, (p_1 = 0.00001)$ $\omega = 1.000$	M7	2	1.000
<b>Non-Mecoptera:</b>							
M0: one-ratio	19	-5537.53	1.625	$\omega = 0.016$			
M1a: neutral	20	-5496.97	1.789	$\omega_0 = 0.013, \omega_1 = 1.000$ $p_0 = 0.967, p_1 = 0.033$			
M2a: selection	22	-5496.97	1.789	$\omega_0 = 0.013, \omega_1 = 1.000, \omega_2 = 1.000$ $p_0 = 0.967, p_1 = 0.019, p_2 = 0.014$	M1a	2	1.000
M3: discrete ( $K=3$ )	23	-5426.21	1.735	$\omega_0 = 0.001, \omega_1 = 0.020, \omega_2 = 0.119$ $p_0 = 0.521, p_1 = 0.394, p_2 = 0.085$	M0	4	$5.063 \times 10^{-47*}$
M7: beta	20	-5429.44	1.728	$p = 0.300, q = 13.893$			
M8: beta& $\omega$	22	-5429.44	1.728	$p = 0.300, q = 13.895$ $p_0 = 0.99999, (p_1 = 0.00001)$ $\omega = 1.000$	M7	2	1.000

NOTE – np is the number of parameters for each model, df is degrees of freedom, while  $p_0$ – $p_2$  is the proportion of sites in a given site class, and  $p, q$  describe the beta distribution.

\*Significant (p-value < 5%).

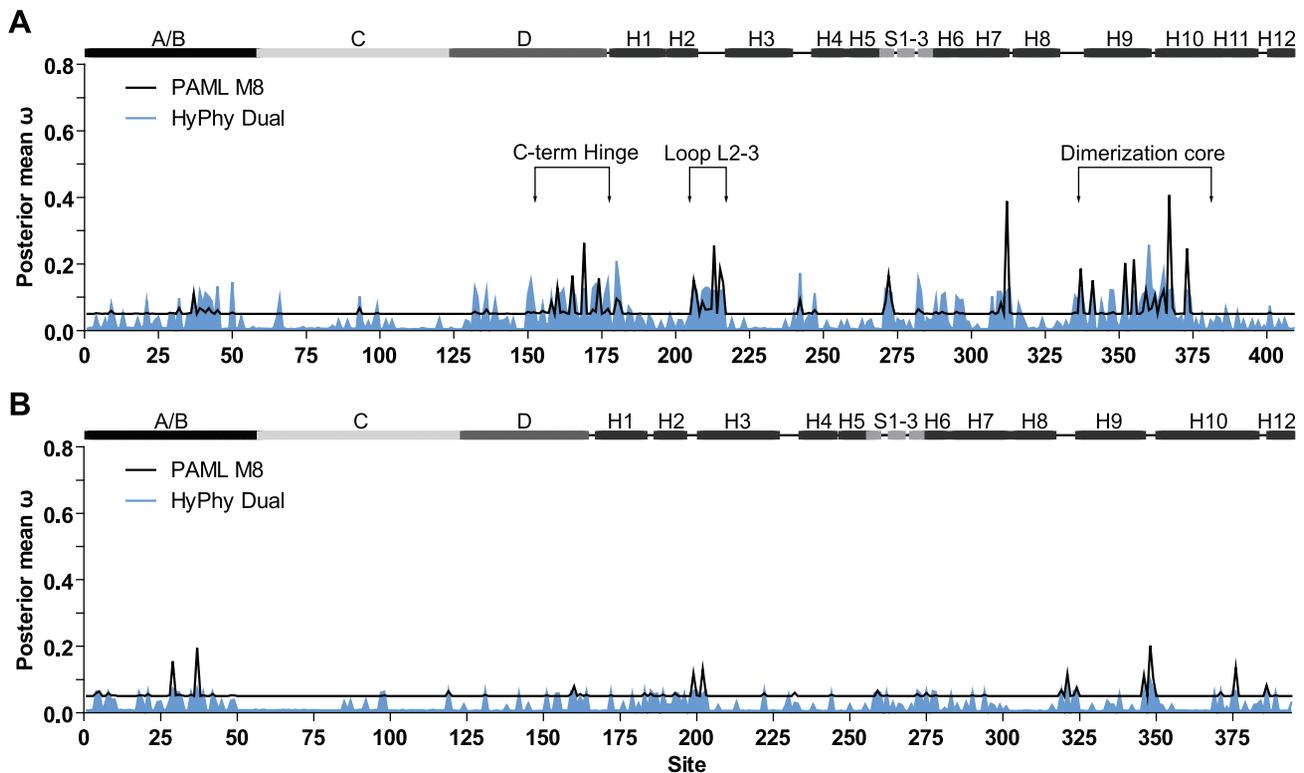
doi:10.1371/journal.pone.0023416.t002

position of loop L8–9 closer to EcR [30]. We identified sites under positive selection along the Mecoptera lineage in the loop L1–3 region which interact with the ligand and contribute to a hydrogen bond network with H3 and loop L11–12 (Figure 2). Site R385 (according to *H. virescens*) in H9, a Mecoptera-specific contact with EcR, was also shown to be under positive selection along the Mecoptera lineage by branch-site analysis. Furthermore, sites surrounding R385 appear to be under relaxed constraint within the Mecoptera in our random-sites analysis (Figure 3A). This was also the case at sites near S447 (according to *H. virescens*) in EcR H7, which establishes contact with USP/RXR R385 (Figure 4A). This is consistent with the acquisition of an expanded dimer interface between USP/RXR and EcR in the Mecoptera.

It has been postulated that the expansion of this interface may reflect a strengthened inter-dependency between USP/RXR and EcR for activation [30]. Based on structural studies, Diptera and Lepidoptera USP/RXR cannot bind the canonical NR-box LXXLL motif of coactivators because H12 sits within the binding site [23,24,71,72]. Our random-sites analyses revealed sites with elevated  $\omega$  within both H12 and the coactivator binding groove (loop L3–4, H4) of Mecoptera USP/RXR (Figure 3A). These results are consistent with a change in the function of these regions, leading to a relaxation of evolutionary constraint at these sites. Without a novel interaction interface, such shifts may render

USP/RXR more dependent on partner proteins for coactivator recruitment [23,73]. Interestingly, similar shifts in cofactor interfaces may also have occurred in EcR. An elevation in  $\omega$  was observed in the carboxy-terminal D domain of Mecoptera EcR, a region implicated in corepressor binding in the mammalian thyroid hormone receptor [74]. The biological significance of an increased dependence on partner proteins remains unclear, but there are some possibilities.

One mechanism which may have led to increased evolutionary rates in the Mecoptera dimer interface is the increased recruitment of USP/RXR as a protein hub. In *D. melanogaster* and *A. aegypti*, USP/RXR is known to interact with many other proteins such as Seven-up (Svp), hormone receptors 38 and 78 (HR38, 78), as well as the *Methoprene-tolerant* gene product (MET) [75,76,77,78]. Hubs at major branch points in protein-protein interaction networks tend to display evidence of elevated  $\omega$  rates and positive selection [79]. The strong positive selection we observed could be congruent with such a role for USP/RXR. Indeed network level analyses suggest that several proteins in the ecdysone cascade, including USP/RXR and EcR, experienced an increased evolutionary rate at the base of the Mecoptera [80]. However, it is unclear if this is due to functional divergence or simply parallel accelerations across the network in order to maintain interactions. Similarly, the elevated evolutionary rates we



**Figure 4. Posterior mean  $\omega$  at each amino acid site across the EcR gene.** The values of  $\omega$  as estimated by M8 in PAML (black line) and the Dual model of HyPhy (shaded blue) are shown for each codon site across the gene for the Mecoptera (A) and Non-Mecoptera (B). A schematic of EcR secondary structure is shown above both plots to illustrate the position of each functional domain (A/B, C, and D) as well as the helices (H1–H12) and  $\beta$  sheets (S1–S3) of the ligand-binding domain. The schematic for the Mecoptera and Non-Mecoptera genes are based on the crystal structures of *H. virescens* and *B. tabaci* EcR, respectively. Site numbering is based on the EcR alignment given in supporting figure S2. doi:10.1371/journal.pone.0023416.g004

detected within the dimerization interface of the Mecoptera clade may be the result of a combination of strong positive selection and purifying selection, not simply relaxed constraint. Substitutions in this region may affect both the affinity and specificity of protein-protein interaction, but might not necessarily improve both of these attributes. In fact, there may be some trade-offs, whereby some substitutions confer a tighter affinity for one protein, but reduce the specificity of interactions with other partner proteins. However, at this point the residues involved in the interface between USP/RXR and alternate partner proteins remain unknown. Furthermore, the network of USP/RXR protein interactions is unknown in lower insects, and a lack of sequence data from outside of the Holometabola prevented the analysis of alternate USP/RXR partners in our study. Therefore, further work is needed to clarify the protein-protein interactions of USP/RXR among insect lineages.

The link between shifts in USP/RXR molecular evolution and shifts in physiology at the organismal level has been difficult to establish. Studies have shown that the role of USP/RXR in molting and metamorphosis seems conserved across insect lineages [6,7]. Differences in the reproductive biology of insect lineages have been largely unexplored, but may hold more promise [28]. Although there are some exceptions, JH is the major hormone regulating reproductive events such as vitellogenesis in most insect groups [81]. In the Lepidoptera, there is a decreased dependence on JH for vitellogenesis, whereas in the Diptera, vitellogenesis is largely regulated by ecdysteroids. In many Diptera, JH is instead required during previtellogenesis in order for tissues to acquire

competence to 20E [82]. The site of vitellogenin production also varies between insect lineages. In many basal lineages, the fat body appears to be the sole site of synthesis, in others, both the ovary and fat body contribute, whereas in some Diptera, the ovary may be the major source [83,84]. Thus, a stronger relationship between USP/RXR and EcR for heterodimerization is also consistent with the increased dependence on ecdysteroids for the regulation of reproductive events in some higher insects.

USP/RXR is involved in many reproductive events in higher insects such as egg chamber formation and chorion gene expression in *Drosophila* and regulation of the cyclicity of vitellogenesis in *A. aegypti* [75,85,86]. Ovarian morphology and structure also varies among insect groups, particularly in lineages between the basal Non-Mecoptera and Mecoptera, such as the Paraneoptera, Coleoptera and Neuropterida [87,88]. The genes which control reproductive events differ among ovary types as well. For example, the set of genes responsible for chorion formation in meroistic ovaries differs from those in the panoistic type ovary, found in the basal Non-Mecoptera [87,89]. In addition, USP/RXR and EcR are highly expressed in nurse cells, a cell type unique to the meroistic ovary [85,90]. Thus, shifts in ovarian morphology, development and the endocrine control of reproduction may have also required plasticity in nuclear receptors. However, the role of USP/RXR in the reproductive events of basal insects is not known, as is its role in highly derived insects contained within the Mecoptera.

Finally, it is important to consider caveats of  $d_N/d_S$  based methods. In recent years a debate has emerged regarding the use

**Table 3.** Parameter estimates for EcR gene (A/B-LBD).

Model	np	lnL	κ	Parameter Estimates	LRT		
					Null	df	p-value
<b>Mecoptera:</b>							
M0: one-ratio	35	-10215.15	1.690	$\omega = 0.029$			
M1a: neutral	36	-10198.25	1.737	$\omega_0 = 0.028, \omega_1 = 1.000$ $p_0 = 0.988, p_1 = 0.012$			
M2a: selection	38	-10198.25	1.737	$\omega_0 = 0.028, \omega_1 = 1.000, \omega_2 = 1.000$ $p_0 = 0.988, p_1 = 0.008, p_2 = 0.003$	M1a	2	1.000
M3: discrete (K=3)	39	-9987.57	1.721	$\omega_0 = 0.000, \omega_1 = 0.023, \omega_2 = 0.107$ $p_0 = 0.404, p_1 = 0.393, p_2 = 0.203$	M0	4	$3.320 \times 10^{-97*}$
M7: beta	36	-9988.17	1.725	$p = 0.357, q = 10.575$			
M8: beta& $\omega$	38	-9988.17	1.719	$p = 0.357, q = 10.575$ $p_0 = 0.99999, (p_1 = 0.00001)$ $\omega = 1.000$	M7	2	1.000
<b>Non-Mecoptera:</b>							
M0: one-ratio	19	-6357.75	2.022	$\omega = 0.016$			
M1a: neutral	20	-6357.74	2.029	$\omega_0 = 0.016, \omega_1 = 1.000$ $p_0 = 0.999, p_1 = 0.001$			
M2a: selection	22	-6357.74	2.029	$\omega_0 = 0.016, \omega_1 = 1.000, \omega_2 = 1.000$ $p_0 = 0.999, p_1 = 0.001, p_2 = 0.000$	M1a	2	1.000
M3: discrete (K=3)	23	-6281.96	2.085	$\omega_0 = 0.000, \omega_1 = 0.013, \omega_2 = 0.0737$ $p_0 = 0.412, p_1 = 0.435, p_2 = 0.153$	M0	4	$9.273 \times 10^{-32*}$
M7: beta	20	-6282.43	2.085	$p = 0.326, q = 17.365$			
M8: beta& $\omega$	22	-6282.43	1.906	$p = 0.326, q = 17.365$ $p_0 = 0.99999, (p_1 = 0.00001)$ $\omega = 1.000$	M7	2	1.000

NOTE – np is the number of parameters for each model, df is the degrees of freedom, while  $p_0-p_2$  is the proportion of sites in a given site class, and  $p, q$  describe the beta distribution.

\*Significant (p-value<5%).

doi:10.1371/journal.pone.0023416.t003

of these methods to reliably detect positive selection because of potential false positives identified in some genes [91,92,93]. However, in our dataset the positively selected sites identified along the Mecoptera lineage by branch-site analysis were consistent with the structural and experimental work of other researchers. Furthermore, we used patterns of  $d_N/d_S$  variation across sites, not merely positive selection, to draw conclusion

regarding USP/RXR evolution. Finally, we utilised new models which allow for among-site variation in  $d_S$  and observed strikingly similar patterns in site-specific evolutionary rates even across these different methods of estimating  $d_N/d_S$ .

While computational methods, such as those used here, are useful in identifying patterns of selective constraint acting on USP/RXR across insect lineages, the functional changes generated by

**Table 4.** Synonymous rate variation among sites (HyPhy).

Dataset	MG94×REV Nonsynonymous GDD 3		MG94×REV Dual GDD 3×3		LRT	
	np	Log L	np	Log L	df	p-value
Mecoptera USP/RXR (A/B-LBD)	43	-10237.10	47	-10232.01	4	0.037*
Non-Mecoptera USP/RXR (A/B-LBD)	27	-5415.51	31	-5410.81	4	0.052
Mecoptera EcR (A/B-LBD)	43	-9946.36	47	-9940.77	4	0.025*
Non-Mecoptera EcR (A/B-LBD)	27	-6269.43	31	-6269.40	4	0.999

NOTE – np is the number of parameters for each model, Log L is the logarithm of the maximum likelihood value, and df is the degrees of freedom used in the LRT calculation when comparing Dual versus Nonsynonymous models.

\*Significant (p-value<5%).

doi:10.1371/journal.pone.0023416.t004

the substitutions identified here remain speculative. For example, we cannot clarify whether or not sites under positive selection near the ligand-binding pocket are related to the potential acquisition of JH binding. Further experimental evidence is required to understand the physiological implications of these evolutionary shifts. In particular, examining the role of USP/RXR in the reproductive system of more basal insects, for example, with the use of techniques such as RNAi, may shed light on these issues. Additionally, the apparent increase in the rigidity of Mecoptera USP/RXR makes this an excellent system for the use molecular dynamics studies to investigate subtle shifts in structure. Future work in this area may yield more insights into the role of the positively selected sites identified in this study which impact the conformation of the receptor complex. Overall, our results highlight the need for more comparative physiological work in the basal Non-Mecoptera and provide critical site-specific information for the design of future site directed mutagenesis studies on USP/RXR.

## Supporting Information

**Figure S1 Alignment of insect USP/RXR sequences.** The green and orange bars indicate the Mecoptera and Non-Mecoptera taxa, respectively, and the schematic below the alignment shows the USP/RXR domain structure. For ease of viewing one alignment is shown. However, the complete alignment was not used for analysis as some regions do not align (e.g. D domain). Only the carboxy-terminal E/F (\*), or ligand-binding domain, was used for the branch and branch-sites analyses reported in table 1. Arrows indicate where poorly aligned regions and major gaps were deleted. Full length sequences were used for the random-sites and HyPhy analyses where the larger dataset was split into Mecoptera and Non-Mecoptera only datasets in order to compare evolutionary rates between the two groups. For clarity, site numbering for the full length Mecoptera (green) and Non-Mecoptera (orange) datasets is shown above and below the alignment, respectively. Note that species names have been abbreviated to six characters, complete names can be found in supporting table S1. (PDF)

**Figure S2 Alignment of insect EcR sequences.** The green and orange bars indicate the Mecoptera and Non-Mecoptera taxa, respectively, and the schematic below the alignment shows the EcR domain structure. For ease of viewing one alignment is shown. However, the complete alignment was not used for analysis as some regions do not align (e.g. D domain). Only the carboxy-terminal E (\*), or ligand-binding domain, was used for the branch analyses reported in table 1. Arrows indicate where poorly aligned regions and major gaps were deleted. Full length sequences were used for the random-sites and HyPhy analyses where the larger dataset was split into Mecoptera and Non-Mecoptera only datasets in order to compare evolutionary rates between the two groups. For clarity, site numbering for the full length Mecoptera (green) and Non-Mecoptera (orange) datasets is shown above and below the alignment, respectively. Note that species names have been abbreviated to six characters, complete names can be found in supporting table S1. (PDF)

## References

1. Wiens M, Batel R, Korzhov M, Müller WEG (2003) Retinoid X receptor and retinoic acid response in the marine sponge *Suberites domuncula*. J Exp Biol 206: 3261–3271.

**Figure S3 USP/RXR and EcR gene trees.** Gene trees for USP/RXR and EcR were generated using the alignment of LBD sequences given in supporting figures S1 and S2. Maximum-likelihood trees for USP/RXR (A) and EcR (C) were constructed in PhyML [57] using the WAG substitution model, with four rate categories to estimate the gamma parameter shape. Neighbor-joining [58] trees for USP/RXR (B) and EcR (D) were constructed in MEGA 4 [38] using the Poisson correction model, with the pair-wise deletion of gaps. For all analyses 100 bootstrap replicates were performed, and nodes with values less than 60 were later collapsed. Each tree was then rooted along the branch leading to *B. germanica* and *L. migratoria*. Note that species names have been abbreviated, see supporting table S1. (PDF)

**Figure S4  $d_N/d_S$  site-profile plots for PAML model M3.** The values of  $\omega$  as estimated by M3 in PAML using the NEB method are shown for each codon site across Mecoptera USP/RXR (A), Non-Mecoptera USP/RXR (B), Mecoptera EcR (C) and Non-Mecoptera EcR (D). A schematic of USP/RXR and EcR secondary structure is shown above each plot to illustrate the position of each functional domain (A/B, C, and D) as well as the helices (H1–H12) and  $\beta$  sheets of the ligand-binding domain. These schematics are based on the crystal structure of each gene in *H. virescens* and *B. tabaci*. Site numbering is the same as figures 3 and 4. (PDF)

**Table S1 Sequence data information.** (DOC)

**Table S2 Putative positively selected sites.** (DOC)

**Table S3 Sites with elevated  $\omega$  in the USP/RXR gene.** (DOC)

**Table S4 Sites with elevated  $\omega$  in the EcR gene.** (DOC)

**Table S5 Statistics for rate distributions inferred by HyPhy.** (DOC)

**Table S6 Rate class assignment for MG94×REV Non-synonymous GDD 3.** (DOC)

**Table S7 Rate class assignment for MG94×REV Dual GDD 3×3.** (DOC)

## Acknowledgments

We thank Jingjing Du and Cameron J. Weadick of the department of Ecology and Evolutionary Biology at the University of Toronto for their insightful comments, technical expertise and critiques of earlier drafts.

## Author Contributions

Conceived and designed the experiments: EFH SST BSWC. Performed the experiments: EFH BSWC. Analyzed the data: EFH SST BSWC. Contributed reagents/materials/analysis tools: SST BSWC. Wrote the paper: EFH SST BSWC.

2. Oro AE, McKeown M, Evans RM (1990) Relationship between the product of the *Drosophila ultraspiracle* locus and the vertebrate retinoid X receptor. Nature 347: 298–301.

3. Mangelsdorf DJ, Ong ES, Dyck JA, Evans RM (1990) Nuclear receptor that identifies a novel retinoic acid response pathway. *Nature* 345: 224–229.
4. Hall BL, Thummel CS (1998) The RXR homolog Ultraspiracle is an essential component of the *Drosophila* ecdysone receptor. *Development* 125: 4709–4717.
5. Barchuk AR, Figueiredo VLC, Simões ZLP (2008) Downregulation of *ultraspiracle* gene expression delays pupal development in honeybees. *J Insect Physiol* 54: 1035–1040.
6. Martín D, Maestro O, Cruz J, Mané-Padrós D, Bellés X (2006) RNAi studies reveal a conserved role for RXR in molting in the cockroach *Blattella germanica*. *J Insect Physiol* 52: 410–416.
7. Tan A, Palli SR (2008) Ecdysone receptor isoforms play distinct roles in controlling molting and metamorphosis in the red flour beetle, *Tribolium castaneum*. *Mol Cell Endocrinol* 291: 42–49.
8. Xu JJ, Tan AJ, Palli SR (2010) The function of nuclear receptors in regulation of female reproduction and embryogenesis in the red flour beetle, *Tribolium castaneum*. *J Insect Physiol* 56: 1471–1480.
9. Thomas HE, Stunnenberg HG, Stewart AF (1993) Heterodimerization of the *Drosophila* ecdysone receptor with retinoid X receptor and *ultraspiracle*. *Nature* 362: 471–475.
10. Yao TP, Forman BM, Jiang ZY, Cherbas L, Chen JD, et al. (1993) Functional ecdysone receptor is the product of *Ecr* and *Ultraspiracle* genes. *Nature* 366: 476–479.
11. Yao TP, Segraves WA, Oro AE, McKeown M, Evans RM (1992) *Drosophila* *ultraspiracle* modulates ecdysone receptor function via heterodimer formation. *Cell* 71: 63–72.
12. Jones G, Sharp PA (1997) Ultraspiracle: An invertebrate nuclear receptor for juvenile hormones. *Proc Natl Acad Sci U S A* 94: 13499–13503.
13. Xu Y, Fang F, Chu YX, Jones D, Jones G (2002) Activation of transcription through the ligand-binding pocket of the orphan nuclear receptor ultraspiracle. *Eur J Biochem* 269: 6026–6036.
14. Hayward DC, Dhadialla TS, Zong ST, Kuiper MJ, Ball EE, et al. (2003) Ligand specificity and developmental expression of RXR and ecdysone receptor in the migratory locust. *J Insect Physiol* 49: 1135–1144.
15. Levin AA, Sturzenbecker LJ, Kazmer S, Bosakowski T, Huselton C, et al. (1992) 9-*Cis* retinoic acid stereoisomer binds and activates the nuclear receptor RXR $\alpha$ . *Nature* 355: 359–361.
16. Heyman RA, Mangelsdorf DJ, Dyck JA, Stein RB, Eichele G, et al. (1992) 9-*Cis* retinoic acid is a high affinity ligand for the retinoid X receptor. *Cell* 68: 397–406.
17. Mangelsdorf DJ, Borgmeyer U, Heyman RA, Zhou JY, Ong ES, et al. (1992) Characterization of three RXR genes that mediate the action of 9-*cis* retinoic acid. *Genes Dev* 6: 329–344.
18. Wozniak M, Chu YX, Fang F, Xu Y, Riddiford L, et al. (2004) Alternative farnesoid structures induce different conformational outcomes upon the *Drosophila* ortholog of the retinoid X receptor, ultraspiracle. *Insect Biochem Mol Biol* 34: 1147–1162.
19. Fang F, Xu Y, Jones D, Jones G (2005) Interactions of ultraspiracle with ecdysone receptor in the transduction of ecdysone- and juvenile hormone-signaling. *FEBS J* 272: 1577–1589.
20. Jones G, Jones D, Teal P, Sapa A, Wozniak M (2006) The retinoid-X receptor ortholog, ultraspiracle, binds with nanomolar affinity to an endogenous morphogenetic ligand. *FEBS J* 273: 4983–4996.
21. Iwema T, Billas IML, Beck Y, Bonneton F, Nierengarten H, et al. (2007) Structural and functional characterization of a novel type of ligand-independent RXR-USP receptor. *EMBO J* 26: 3770–3782.
22. Nowickiy SM, Chithalen JV, Cameron D, Tyshenko MG, Petkovich M, et al. (2008) Locust retinoid X receptors: 9-*Cis*-retinoic acid in embryos from a primitive insect. *Proc Natl Acad Sci U S A* 105: 9540–9545.
23. Clayton GM, Peak-Chew SY, Evans RM, Schwabe JWR (2001) The structure of the ultraspiracle ligand-binding domain reveals a nuclear receptor locked in an inactive conformation. *Proc Natl Acad Sci U S A* 98: 1549–1554.
24. Billas IML, Moulinière L, Rochel N, Moras D (2001) Crystal structure of the ligand-binding domain of the ultraspiracle protein USP, the ortholog of retinoid X receptors in insects. *J Biol Chem* 276: 7465–7474.
25. Billas IML, Iwema T, Garnier JM, Mischler A, Rochel N, et al. (2003) Structural adaptability in the ligand-binding pocket of the ecdysone hormone receptor. *Nature* 426: 91–96.
26. Carmichael JA, Lawrence MC, Graham LD, Pilling PA, Epa VC, et al. (2005) 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* 280: 22258–22269.
27. Bonneton F, Zelus D, Iwema T, Robinson-Rechavi M, Laudet V (2003) Rapid divergence of the ecdysone receptor in Diptera and Lepidoptera suggests coevolution between ECR and USP-RXR. *Mol Biol Evol* 20: 541–553.
28. Bonneton F, Brunet FG, Kathirithamby J, Laudet V (2006) The rapid divergence of the ecdysone receptor is a synapomorphy for Mecoptera that clarifies the Strepsiptera problem. *Insect Mol Biol* 15: 351–362.
29. Kristensen NP (1999) Phylogeny of endopterygote insects, the most successful lineage of living organisms. *Eur J Entomol* 96: 237–253.
30. Iwema T, Chaumot A, Studer RA, Robinson-Rechavi M, Billas IML, et al. (2009) Structural and evolutionary innovation of the heterodimerization interface between USP and the ecdysone receptor ECR in insects. *Mol Biol Evol* 26: 753–768.
31. Tocchini-Valentini GD, Rochel N, Escrivá H, Germain P, Peluso-Itis C, et al. (2009) Structural and functional insights into the ligand-binding domain of a nonduplicated retinoid X nuclear receptor from the invertebrate chordate amphioxus. *J Biol Chem* 284: 1938–1948.
32. Kelleher ES, Markow TA (2009) Duplication, selection and gene conversion in a *Drosophila mojavensis* female reproductive protein family. *Genetics* 181: 1451–1465.
33. Wu DD, Wang GD, Irwin DM, Zhang YP (2009) A profound role for the expansion of trypsin-like serine protease family in the evolution of hematophagy in mosquito. *Mol Biol Evol* 26: 2333–2341.
34. Smadja C, Shi P, Butlin RK, Robertson HM (2009) Large gene family expansions and adaptive evolution for odorant and gustatory receptors in the pea aphid, *Acyrtosiphon pisum*. *Mol Biol Evol* 26: 2073–2086.
35. Bulmer MS, Crozier RH (2006) Variation in positive selection in termite GNBP and Relish. *Mol Biol Evol* 23: 317–326.
36. Chen CY, Opazo JC, Erez O, Uddin M, Santolaya-Forgas J, et al. (2008) The human progesterone receptor shows evidence of adaptive evolution associated with its ability to act as a transcription factor. *Mol Phylogenet Evol* 47: 637–649.
37. Thompson JD, Higgins DG, Gibson TJ (1994) Clustal W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22: 4673–4680.
38. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24: 1596–1599.
39. Kimura M (1977) Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature* 267: 275–276.
40. Yang ZH, Bielawski JP (2000) Statistical methods for detecting molecular adaptation. *Trends Ecol Evol* 15: 496–503.
41. Yang ZH (1998) Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol Biol Evol* 15: 568–573.
42. Yang ZH, Nielsen R (1998) Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *J Mol Evol* 46: 409–418.
43. Yang ZH, Nielsen R (2002) Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol Biol Evol* 19: 908–917.
44. Zhang JZ, Nielsen R, Yang ZH (2005) Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol* 22: 2472–2479.
45. Nielsen R, Yang ZH (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 148: 929–936.
46. Yang ZH, Nielsen R, Goldman N, Pedersen AMK (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155: 431–449.
47. Yang ZH (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* 13: 555–556.
48. Pond SK, Muse SV (2005) Site-to-site variation of synonymous substitution rates. *Mol Biol Evol* 22: 2375–2385.
49. Pond SLK, Frost SDW, Muse SV (2005) HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21: 676–679.
50. Whiting MF, Carpenter JC, Wheeler QD, Wheeler WC (1997) The Strepsiptera problem: phylogeny of the holometabolous insect orders inferred from 18S and 28S ribosomal DNA sequences and morphology. *Syst Biol* 46: 1–68.
51. Wheeler WC, Whiting M, Wheeler QD, Carpenter JM (2001) The phylogeny of the extant hexapod orders. *Cladistics* 17: 113–169.
52. Hunt T, Bergsten J, Levkanicova Z, Papadopoulou A, John OS, et al. (2007) A comprehensive phylogeny of beetles reveals the evolutionary origins of a superradiation. *Science* 318: 1913–1916.
53. Weller SJ, Friedlander TP, Martin JA, Pashley DP (1992) Phylogenetic studies of ribosomal RNA variation in higher moths and butterflies (Lepidoptera: Ditrysia). *Mol Phylogenet Evol* 1: 312–337.
54. Yeates DK, Wiegmann BM (1999) Congruence and controversy: toward a higher-level phylogeny of Diptera. *Annu Rev Entomol* 44: 397–428.
55. Brothers DJ (1999) Phylogeny and evolution of wasps, ants and bees (Hymenoptera, Chrysoidea, Vespoidea and Apoidea). *Zool Scr* 28: 233–249.
56. Castro LR, Dowton M (2005) The position of the Hymenoptera within the Holometabola as inferred from the mitochondrial genome of *Perga condei* (Hymenoptera: Symphyta: Pergidae). *Mol Phylogenet Evol* 34: 469–479.
57. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, et al. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 59: 307–321.
58. Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4: 406–425.
59. Gao TP, Ren D, Shih CK (2009) The first Xyelotomidae (Hymenoptera) from the Middle Jurassic in China. *Ann Entomol Soc Am* 102: 588–596.
60. Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 16: 111–120.
61. Felsenstein J (1981) Evolutionary trees from DNA sequences: a maximum-likelihood approach. *J Mol Evol* 17: 368–376.
62. Yang ZH, Wong WSW, Nielsen R (2005) Bayes empirical bayes inference of amino acid sites under positive selection. *Mol Biol Evol* 22: 1107–1118.
63. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph* 14: 33–38.

64. Hayward DC, Bastiani MJ, Trueman JWH, Turman JW, Riddiford LM, et al. (1999) The sequence of *Locusta* RXR, homologous to *Drosophila* Ultraspiracle, and its evolutionary implications. *Dev Genes Evol* 209: 564–571.
65. Maestro O, Cruz J, Pascual N, Martin D, Bellés X (2005) Differential expression of two RXR/ultraspiracle isoforms during the life cycle of the hemimetabolous insect *Blattella germanica* (Dictyoptera, Blattellidae). *Mol Cell Endocrinol* 238: 27–37.
66. Wang SF, Ayer S, Segraves WA, Williams DR, Raikhel AS (2000) Molecular determinants of differential ligand sensitivities of insect ecdysteroid receptors. *Mol Cell Biol* 20: 3870–3879.
67. Sharp PM, Averof M, Lloyd AT, Matassi G, Peden JF (1995) DNA sequence evolution: the sounds of silence. *Philos Trans R Soc Lond B Biol Sci* 349: 241–247.
68. Wurtz JM, Bourguet W, Renaud JP, Vivat V, Chambon P, et al. (1996) A canonical structure for the ligand-binding domain of nuclear receptors. *Nat Struct Biol* 3: 87–94.
69. Bourguet W, Ruff M, Chambon P, Gronemeyer H, Moras D (1995) Crystal structure of the ligand-binding domain of the human nuclear receptor RXR- $\alpha$ . *Nature* 375: 377–382.
70. Bourguet W, Vivat V, Wurtz JM, Chambon P, Gronemeyer H, et al. (2000) Crystal structure of a heterodimeric complex of RAR and RXR ligand-binding domains. *Mol Cell* 5: 289–298.
71. Darimont BD, Wagner RL, Apriletti JW, Stallcup MR, Kushner PJ, et al. (1998) Structure and specificity of nuclear receptor-coactivator interactions. *Genes Dev* 12: 3343–3356.
72. Nolte RT, Wisely GB, Westin S, Cobb JE, Lambert MH, et al. (1998) Ligand binding and co-activator assembly of the peroxisome proliferator-activated receptor- $\gamma$ . *Nature* 395: 137–143.
73. Moras D, Gronemeyer H (1998) The nuclear receptor ligand-binding domain: structure and function. *Curr Opin Cell Biol* 10: 384–391.
74. Hörlein AJ, Näär AM, Heinzl T, Torchia J, Gloss B, et al. (1995) Ligand-independent repression by the thyroid hormone receptor mediated by a nuclear receptor co-repressor. *Nature* 377: 397–404.
75. Zhu JS, Miura K, Chen L, Raikhel AS (2003) Cyclicity of mosquito vitellogenic ecdysteroid-mediated signaling is modulated by alternative dimerization of the RXR homologue *Ultraspiracle*. *Proc Natl Acad Sci U S A* 100: 544–549.
76. Hirai M, Shinoda T, Kamimura M, Tomita S, Shiotsuki T (2002) *Bombyx mori* orphan receptor, BmHR78: cDNA cloning, testis abundant expression and putative dimerization partner for *Bombyx* ultraspiracle. *Mol Cell Endocrinol* 189: 201–211.
77. Baker KD, Shewchuk LM, Kozlova T, Makishima M, Hassell A, et al. (2003) The *Drosophila* orphan nuclear receptor DHR38 mediates an atypical ecdysteroid signaling pathway. *Cell* 113: 731–742.
78. Li YP, Zhang ZL, Robinson GE, Palli SR (2007) Identification and characterization of a juvenile hormone response element and its binding proteins. *J Biol Chem* 282: 37605–37617.
79. Alvarez-Ponce D, Aguadé M, Rozas J (2009) Network-level molecular evolutionary analysis of the insulin/TOR signal transduction pathway across 12 *Drosophila* genomes. *Genome Res* 19: 234–242.
80. Bonneton F, Chaumot A, Laudet V (2008) Annotation of *Tribolium* nuclear receptors reveals an increase in evolutionary rate of a network controlling the ecdysone cascade. *Insect Biochem Mol Biol* 38: 416–429.
81. Raikhel AS, Brown MR, Bellés X (2005) Hormonal Control of Reproductive Processes. In: Gilbert LI, Iatrou K, Gill SS, eds. *Comprehensive Molecular Insect Science*. Oxford: Pergamon. pp 433–491.
82. Zhu JS, Chen L, Raikhel AS (2003b) Posttranscriptional control of the competence factor  $\beta$ FTZ-F1 by juvenile hormone in the mosquito *Aedes aegypti*. *Proc Natl Acad Sci U S A* 100: 13338–13343.
83. Valle D (1993) Vitellogenesis in insects and other groups - a review. *Mem Inst Oswaldo Cruz* 88: 1–26.
84. Huebner E, Tobe SS, Davey KG (1975) Structural and functional dynamics of oogenesis in *Glossina austeni*: vitellogenesis with special reference to the follicular epithelium. *Tissue Cell* 7: 535–558.
85. Carney GE, Bender M (2000) The *Drosophila* ecdysone receptor (*EcR*) gene is required maternally for normal oogenesis. *Genetics* 154: 1203–1211.
86. Bernardi F, Romani P, Tzertzinis G, Gargiulo G, Cavaliere V (2009) EcR-B1 and Usp nuclear hormone receptors regulate expression of the VM32E eggshell gene during *Drosophila* oogenesis. *Dev Biol* 328: 541–551.
87. Büning J (1993) Germ cell cluster formation in insect ovaries. *Int J Insect Morphol Embryol* 22: 237–253.
88. Trauner J, Büning J (2007) Germ-cell cluster formation in the telotrophic merostic ovary of *Tribolium castaneum* (Coleoptera, Polyphaga, Tenebrionidae) and its implication on insect phylogeny. *Dev Genes Evol* 217: 13–27.
89. Irls P, Bellés X, Piulachs MD (2009) Identifying genes related to choriogenesis in insect panoistic ovaries by Suppression Subtractive Hybridization. *BMC Genomics* 10: 206.
90. Oro AE, McKeown M, Evans RM (1992) The *Drosophila* retinoid X receptor homolog *ultraspiracle* functions in both female reproduction and eye morphogenesis. *Development* 115: 449–462.
91. Nozawa M, Suzuki Y, Nei M (2009) Reliabilities of identifying positive selection by the branch-site and the site-prediction methods. *Proc Natl Acad Sci U S A* 106: 6700–6705.
92. Yang Z, Nielsen R, Goldman N (2009) In defense of statistical methods for detecting positive selection. *Proc Natl Acad Sci U S A* 106: E95.
93. Yang Z, dos Reis M (2011) Statistical properties of the branch-site test of positive selection. *Mol Biol Evol* 28: 1217–1228.