Published in final edited form as:

Mol Phylogenet Evol. 2012 June ; 63(3): 617–624. doi:10.1016/j.ympev.2012.02.003.

Amino acid sequence coevolution in the insect bursicon ligand-receptor system

Austin L. Hughes

Department of Biological Sciences, University of South Carolina, Columbia SC 29205

Abstract

The pattern of amino acid residue replacement in the components of the bursicon signaling system (involving the BURS α /BURS β heterodimer and its receptor BURSrec) was reconstructed across a phylogeny of 17 insect species, in order to test for the co-occurrence of replacements at sets of individual sites. Sets of three or more branches with perfectly concordant changes occurred to a greater extent than expected by chance, given the observed level of amino acid change. The latter sites (SPC sites) were found to have distinctive characteristics: (1) the mean number of changes was significantly lower at SPC sites than that at other sites with multiple changes; (2) SPC sites had a significantly greater tendency toward parallel amino acid changes than other sites with multiple changes, but no greater tendency toward convergent changes; and (3) parallel changes tended to involve relatively similar amino acids, as indicated by relatively low mean chemical distances. The results implicated functional constraint, permitting only a limited subset of amino acids in a given site, as a major factor in causing both parallel amino acid replacement and coordinated amino acid changes in different sites of the same protein and of interacting proteins in this system.

Keywords

bursicon; convergent evolution; functional constraint; parallel evolution; protein-protein interactions

1. Introduction

There is a large theoretical and empirical literature regarding the co-evolution of amino acid sites in proteins, including the co-evolution between sites within the same polypeptide (Afonnikov and Kolchanov 2004; Altschuh et al. 1998; Cordero and Fares 2008; Fitch 1971; Fitch and Markowitz 2007; Garriga et al. 2007; Gruenheit et al. 2008; Pollock et al. 1999; Wang et al. 2007) and the co-evolution between sites in interacting polypeptides, such as ligands and receptors (Bonneton et al. 2003; Collette et al. 2003; Farid and Szudlinski 2004; Hughes and Yeager 1999; Moyle et al. 1994; Park et al. 2002). This topic has been of interest because of its practical consequences for phylogenetic analysis and structure prediction, as well as for understanding of protein function and adaptation (Cordero and Fares 2008; Rokas and Carroll 2009; Wang et al. 2007).

© 2012 Elsevier Inc. All rights reserved.

Austin L. Hughes, Ph.D. Department of Biological Sciences, University of South Carolina, Coker Life Sciences Bldg., 700 Sumter St., Columbia SC 29208 USA, Tel: 1-803-777-9186, Fax: 1-803-777-4002, austin@biol.sc.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

The “covarion” hypothesis of Fitch and Markowitz (1970) was based on the assumption that the focus of purifying selection on a protein can change over evolutionary time, as a consequence of amino acid sequence changes. As a result, when two or more amino acid sites become freed from constraint, neutral amino acid replacements can co-occur in these sites. Another biological process that might give rise to coordinated amino acid changes, either within a single polypeptide or at sites in two interacting polypeptides, is fixation of a compensatory mutation that alleviates the effects of a previously fixed mutation of slightly deleterious effect (Kimura 1985; Stephan 1986). Such a slightly deleterious mutation might occur in any population, but might be fixed by drift during a population bottleneck, when purifying selection becomes less efficient due to low effective population size (Whitlock et al. 2003). On the other hand, a mutation that is selectively favored may have deleterious pleiotropic effects; pathogenic microbes provide abundant evidence of the latter phenomenon in the case of mutants favored because they provide immune escape (Crawford et al. 2007; Friedrich et al. 2004) or drug resistance (Björkman et al. 2000; Levin et al. 2000; Piana et al. 2002). Subsequent compensatory mutations may be selectively favored because they alleviate the pleiotropic deleterious effects of previously favored mutations. Finally coordinated amino acid replacements may result from adaptive protein evolution (Kornegay et al. 1994; McCracken et al. 2009; Stewart et al. 1987; Wood et al. 2005).

Here I examine the pattern of amino acid residue replacement in the components of the bursicon signaling system across the phylogeny of insects, in order to test for the co-occurrence of replacements at sets of individual sites. Bursicon is a protein hormone that initiates the tanning of insect cuticle after molting (Honegger et al. 2008; Loveall and Deitcher 2010; Luo et al. 2005). Bursicon also has effects on the central nervous system, including induction of wing expansion after moulting to the adult stage in *Drosophila* (Peabody et al. 2008). The active form of bursicon is a heterodimer composed of units known as BURS α (or BURS) BURS β (or PBURS, for “partner of bursicon”), each of which is believed to form a cystine knot structure similar to that of certain hormones and other signaling molecules of vertebrates (Luo et al. 2005; Vitt et al. 2001). In *Drosophila*, the heterodimeric hormone has been shown to bind a receptor (BURSrec) belonging to a subfamily of 7-transmembrane domain G protein-coupled receptors characterized by an N-terminal leucine-rich repeat domain (Luo et al. 2005).

Analysis of orthologous amino acid sequences of BURS α , BURS β , and BURSrec across the insect phylogeny was used to reconstruct co-occurring amino acid replacements. A computer simulation approach was used to test whether certain sites showed co-occurring amino acid changes on multiple branches of the phylogeny to a greater extent than expected by chance, given the degree of amino acid sequence change observed in the phylogeny. Finally, examination of the characteristics of such sites provided insights into the mechanisms responsible for coordinated amino acid sequence evolution in these three interacting proteins.

2. Methods

2.1. Sequences and Phylogeny

Orthologous sequences of BURS α , BURS β , BURSrec were identified by BLASTP homology search in the NCBI sequence databases (Table 1). A sequence for each of these proteins was obtained for 17 species of insects and one arachnid, the tick *Ixodes scapularis* (Table 1). Amino acid sequences were aligned using CLUSTALW (Thompson et al. 1997). Using *I. scapularis* as an outgroup to root the tree, the organismal phylogeny (Figure 1) was constructed on the basis of published phylogenies of the holometabolous insect orders (Wiegmann et al. 2009) and *Drosophila* species (*Drosophila* 12 Genomes Consortium 2007).

I used the maximum parsimony (MP) method (Swofford 1999) to reconstruct amino acid replacements in BURS α , BURS β , and BURSrec on the 32 branches of the tree, exclusive of the outgroup (“the ingroup branches”; Figure 1). Any site at which the alignment postulated a gap in any sequence was excluded from the analysis. The data set used for analysis (Supplementary Figure S1) included 105 aligned amino acid sites in BURS α , 107 aligned amino acid sites in BURS β , and 799 aligned amino acid sites in BURSrec (1011 total sites). Of these sites, 531 showed one or more changes in the ingroup branches; and a total of 1562 changes were reconstructed at these sites. Thus, given 32 branches and 1011 sites, there were 32,352 possible changes along the phylogeny; and the proportion of possible changes that were reconstructed to have occurred was 4.8%.

In the absence of detailed three-dimensional structures of BURS α , BURS β , or BURSrec, changes were mapped with regard to structural features reported in the literature (Eriksen et al. 2000; Luo et al. 2005). In the case of BURS α and BURS β , the position of changes was described relative to the 11 cysteine residues (C1-C11) that form the predicted cystine knot structure (Luo et al. 2005). BURSrec has an N-terminal leucine rich repeat region (LRR) domain, believed to be involved in ligand binding; seven transmembrane domains (TM1-TM7); and a cytoplasmic C-terminal domain (Eriksen et al. 2000).

2.2. Concordance Measure

Consider an alignment of n sites evolving along m branches of a phylogenetic tree. This alignment might correspond to a single protein or (as in the present study) a concatenated alignment of more than one protein. The changes occurring on the branches of the phylogenetic tree can be represented as an $n \times m$ matrix $\mathbf{B}$. Elements of $\mathbf{B}$ (b_{xy}) takes values of 0 or 1. If there is a change at the x th site on the y th branch of the tree, $b_{xy} = 1$; otherwise $b_{xy} = 0$. Let $\mathbf{A} = \mathbf{B}\mathbf{B}'$. For two sites i and j , define the coefficient of concordance (CC_{ij}) by the following:

$$CC_{ij} = a_{ij} / [(1/2)(a_{ii} + a_{jj})]$$

where a_{ij} , a_{ii} , and a_{jj} are elements of $\mathbf{A}$.

If $CC_{ij} = 1$, the two sites i and j are said to be *perfectly concordant*. Sets of two or more sites all of which are perfectly concordant with each other are called perfectly concordant sets. Perfect concordance is relatively trivial if just one branch is involved; i.e., if two or more sites all show changes on just one branch of the phylogeny. Instead, I used calculation of CC_{ij} to identify perfectly concordant sets involving two or more branches of the phylogeny.

I used computer simulation to estimate the probability of occurrence of two or more sites perfectly concordant on a given number of branches. Each simulation involved the creation of simulated 1000 populations of m branches and 1011 sites; separate simulations were run with $m = 2, 3, 4$, and 5. Changes were assigned to sites in the branches at random, using the observed frequencies of change on branches in the data to generate expected values. In the actual data, there was a wide range of branch lengths (ranging from 0 to 188 changes, with a mean of 48.1 ± 9.9 SE). In order to reflect this range of branch lengths in the simulations, four separate expected values were generated in each simulation, each of which was based on a category of branch lengths in the actual data. The 32 branches in the actual data were assigned to four categories with regard to the number of changes: (1) 8 branches with mean 1.5 ± 0.4 changes (range 0–3 changes); (2) 10 branches with mean 17.2 ± 1.3 changes (range 10–22 changes); (3) 9 branches with mean 76.8 ± 3.7 changes (range 61–93); and (4) 5 branches with mean 137.4 ± 13.8 changes (range 113–188). In each population of simulated branches, these categories occurred in the same proportions in which they were found in the

actual data. The branches in simulations ranged from 0 to 176 changes, with a mean of 48.3, which was very close to the mean number of changes observed in the 32 branches in the actual data. As a conservative test of the statistical significance of perfect concordance observed at a given number of sites (m) along b branches, I used the proportion of 1000 simulated samples having at least one set of m or more number of sites which were perfectly concordant along b branches.

In preliminary analyses, the maximum likelihood (ML) method in MEGA 5 (Tamura et al. 2011) was also used to reconstruct changes. In general, the ML results showed good agreement with the results of MP. At the 511 variable sites, the numbers of changes reconstructed by the two methods were strongly positively correlated ($r = 0.977$; $P < 0.001$). In addition, all perfectly concordant sets identified by MP were also detected by ML. However, the ML method predicted a number of concordant sets that were not seen in the MP results. Therefore, MP appeared to be more conservative with regard to the detection of perfect concordance, and only the MP results are reported below.

2.3. Chemical Distances

The chemical similarity between pairs of amino acid was measured by the chemical distances of Miyata et al. (1979) and of Xia and Xie (2002). Since the results using the two distances were essentially identical, only the former is reported here. That distance has an arbitrary scale ranging from 0.06 to 5.23, with a median value for the 190 possible amino acid pairs of 2.365.

3. Results

3.1. Inferred Amino Acid Changes

Amino acid changes were inferred in BURSa, BURSB, and BURSec across the phylogeny given in Figure 1. In the case of in BURSa, the length of the tree was 96 changes, with a consistency index of 0.844. For BURSB, the length of the tree was 276 changes, with a consistency index of 0.808; and for BURSec, the length of the tree was 1505 with a consistency index of 0.819. Thus, the level of homoplasy (i.e., inferred parallel and convergent amino acid changes) was relatively low and similar in the case of each of the three proteins.

Of the 1011 sites analyzed, 480 (47.5%) were invariant across the phylogeny. Of the 531 sites with at least one amino acid change, 358 (74.6%) showed 2 or more inferred changes. The highest number of changes inferred at a site was 10, which occurred at a site in the LRR of BURSec (site 456 in BURSec of *Drosophila melanogaster*). Overall, the number of sites showing a given number of changes was negatively correlated with that number ($r_s = -0.998$; $P < 0.001$; Figure 2A). For the 531 sites with at least one inferred change, the number of cases of perfect concordance involving a given site was negatively correlated with the number of changes at that site ($r_s = -0.836$; $P < 0.001$; Figure 2B). Only sites with 1, 2, 3, 4, or 5 changes were involved in relationships of perfect concordance (Figure 2B).

3.2. Sites with Perfect Concordance

Computer simulation was used to identify sets of sites at which perfect concordance occurred to a greater extent than expected by chance (these were referred to as “significant perfectly concordant sites” or SPC sites; Table 2). Simulations indicated that, in the case of 2 branches, there were no significant sets. In the case of 3 branches, sets of 2 sites ($P = 0.019$) or 3 sites ($P = 0.009$) were significant (Table 2). In the case of 4 or 5 branches, perfect concordance at 2 sites was significant at $P < 0.001$ (Table 2). In the data, there was one set of 3 SPC sites and 9 sets of 2 SPC sites perfectly concordant on 3 branches (Table

2). There were two sets of 2 SPC sites perfectly concordant on 4 branches, and one set of 2 sites perfectly concordant on 5 branches (Table 2). Overall, there were 27 SPC sites, at which 91 amino acid changes were inferred (Table 2).

3.3. Parallel Amino Acid Changes

The amino acid changes at SPC sites included a substantial number of parallel changes; that is, cases where the same ancestral amino acid was replaced by the same derived amino acid independently on two or more branches of the phylogeny (Table 2). For example, at site 123 (numbered as in *D. melanogaster*) of BURS α , the change R $\rightarrow$ K occurred on the branch from Node 10 $\rightarrow$ *A. gambiae*, on the branch from Node 13 to Node 12, and on the branch from Node 15 $\rightarrow$ *Acyrtosiphon pisum* (Table 2). Of 91 inferred changes at SPC sites, 36 (43.8%) were parallel changes. This pattern was compared with that at non-SPC sites with more than one inferred change; there were 331 such sites, with 1298 inferred changes. Of the latter changes only 230 (17.7%) were parallel.

Of the 27 SPC sites, 16 (59.3%) showed parallel changes, whereas of the other sites with more than one inferred change only 104/331 (31.4%) showed parallel changes. The difference in proportions was significant ($\chi^2 = 8.7$; 1 d.f. $P = 0.003$). A similar pattern was seen when SPC sites were compared with other sites showing perfect concordance but not belonging to significant sets. There were 62 of the latter sites, 16 of which (25.8%) showed parallel changes. Again this proportion differed significantly from that at SPC sites ($\chi^2 = 9.1$; 1 d.f. $P = 0.002$). Likewise, there were parallel changes at 88 of 269 (32.7%) sites without perfect concordance, and this proportion differed significantly from that at SPC sites ($\chi^2 = 7.6$; 1 d.f. $P = 0.006$).

Note that the higher frequency of parallel changes at SPC sites could not be attributed to a higher overall level of change at SPC sites. The mean number of changes at the 27 SPC sites (3.37 ± 0.10) was in fact significantly lower than that at the 331 non-SPC sites (3.92 ± 0.12 ; t-test; $P < 0.001$). The difference was even more striking when SPC sites were compared with sites that lacked perfect concordance, where the mean number of changes was 4.36 ± 0.10 , again significantly different from the value for SPC sites (t-test; $P < 0.001$).

3.4. Convergent Amino Acid Changes

A convergent amino acid change was defined to occur when, at a given site, the same derived amino acid arose on two or more branches, starting from two or more different ancestral amino acids. For example, at the SPC site BURS β 61 showed the change Q $\rightarrow$ S on the branch Node 13 $\rightarrow$ Node 12 and the change A $\rightarrow$ S on the branch Node 15 $\rightarrow$ *P. humanus* (Table 2). To analyze the frequency of convergent changes, I considered only sites at which three or more different amino acids occurred in the data, because a minimum of three different amino acids is necessary for convergent change. SPC sites included 23 such sites, only 2 of which (8.7%) showed convergent amino acid changes. Of 289 non-SPC sites meeting the same criteria, 62 (21.5%) had at least one convergent change. The difference in proportions was not significant (Fisher's exact test).

Changes at all sites with at least three different amino acids were classified as neither convergent nor parallel; convergent; parallel; or both convergent and parallel. Note that a change might be both convergent and parallel if both convergent and parallel changes occurred at the same site; for example, at the SPC site BURSrec 417 (Table 2). The highest mean amino acid chemical distance was seen in the case of sites with convergent but not parallel changes (1.70 ± 0.08 S.E.; Figure 3A). The next highest mean amino acid chemical distance was seen at sites with neither convergent nor parallel changes (1.52 ± 0.05 ; Figure 3A). The lowest mean amino acid chemical distance was seen at sites with both convergent

and parallel changes (1.22 ± 0.10 ; Figure 3A). The mean amino acid distance at sites with parallel but not convergent changes (1.25 ± 0.06) was similar (Figure 3A). By a factorial ANOVA, there was a significant effect of parallel change on amino acid distance ($F_{1, 308} = 17.73$; $P < 0.001$). However, there was not a significant effect of convergent change on mean amino acid chemical distance ($F_{1, 308} = 0.74$; n.s.); and there was no significant interaction ($F_{1, 308} = 1.30$; n.s.).

When SPC sites with 3 or more amino acids were analyzed separately, mean chemical distance was significantly lower at sites with parallel changes (1.13 ± 0.11) than at sites without parallel changes (1.81 ± 0.22 ; t-test for unequal variances; $P = 0.014$). On the other hand, mean chemical distance did not differ significantly between sites with convergent changes (1.48 ± 0.31) than at sites without convergent changes (1.45 ± 0.15 ; t-test for unequal variances; n.s.).

Among sites at least three different amino acids, the mean number of changes was highest in sites with both convergent and parallel changes (5.56 ± 0.34) and only slightly lower at sites with convergent changes but no parallel changes (5.37 ± 0.26 ; Figure 3B). The mean number of changes was lowest at sites with neither parallel nor convergent changes (3.54 ± 0.11), and somewhat higher at sites with parallel changes but not convergent changes (4.42 ± 0.16 ; Figure 3B). By a factorial ANOVA, there were significant effects of parallel change ($F_{1, 308} = 5.18$; $P = 0.024$) and of convergent change ($F_{1, 308} = 40.06$; $P < 0.001$). However, there was not a significant interaction ($F_{1, 308} = 2.20$; n.s.).

3.5. Rapidly Evolving Branches

Across the phylogeny of Figure 1, the number of inferred amino acid changes in BURS α was positively correlated with that in BURS rec ($r = 0.684$; $P < 0.001$; Figure 4A). Likewise, the number of inferred amino acid changes in BURS β was positively correlated with that in BURS rec ($r = 0.816$; $P < 0.001$; Figure 4B). Studentized t-residuals (equivalent to a t-test of the improvement in the regression due to elimination of one point) were used to identify outliers from the regression in each case (Figure 4). In the regression of the number of changes in BURS α against the number of changes in BURS rec (Figure 4A), there were two significant outliers, both with significantly greater numbers of changes in BURS α than expected based on the number of changes in BURS rec : (1) the branch from Node 13 $\rightarrow$ Node 12 ($t = 4.40$; $P < 0.001$); and (2) the branch from Node 10 $\rightarrow$ *Anopheles gambiae* ($t = 2.55$; $P < 0.05$). In the regression of the number of changes in BURS β against the number of changes in BURS rec (Figure 4B), two branches showed significantly greater numbers of changes in BURS β than expected: (1) the branch from Node 12 $\rightarrow$ *Apis mellifera* ($t = 2.10$; $P < 0.05$); and (2) the branch from Node 13 $\rightarrow$ *Nasonia vitripennis* ($t = 2.72$; $P < 0.050$). On the other hand, the branch Node 10 $\rightarrow$ Node 9 (*Drosophila*) showed fewer changes in BURS β than expected ($t = -3.73$; $P < 0.001$; Figure 4B).

4. Discussion

Reconstruction of amino acid changes in BURS α , BURS β , and BURS rec across a phylogeny of insects showed that sets of two or more branches with perfectly concordant changes occurred to a greater extent than expected by chance, given the observed level of amino acid change. Sets of sites showing perfect concordance on more sites and branches than expected by chance (SPC sites) were found to have distinctive characteristics. First, the mean number of changes was significantly lower at SPC sites than that at other sites with multiple changes. Second, SPC sites underwent significantly fewer amino acid changes than other sites with multiple changes but showed a greater tendency toward parallel amino acid changes than did other sites. Both parallel and convergent changes tended to occur at sites with high numbers of changes. Finally, parallel changes tended to involve relatively similar

amino acids, as indicated by significantly lower mean chemical distances, a tendency not seen in the case of convergent changes.

Taken together, these findings suggested that SPC sites were in general more subject to functional constraint than other sites with multiple changes. Not only did SPC sites show fewer changes on average than did other sites with multiple changes, but the predominance of parallel changes at SPC sites implied a strong tendency toward conservative amino acid changes. The present analysis of SPC sites thus pointed to functional constraint, permitting only a limited subset of amino acids in a given site, as a major factor in accounting for both parallel amino acid replacement and coordinated amino acid changes in different sites of the same protein and of interacting proteins in the bursicon ligand-receptor system. These results suggest caution with regard to the assumption that parallel and convergent amino acid changes, especially when they occur in highly conserved proteins, necessarily reflect adaptive evolution (Castoe et al. 2009).

Two branches showed relatively rapid evolution of BURS α in comparison to BURSrec: the branch from Node 13 to Node 12 and the branch from Node 10 to *Anopheles gambiae* (Figure 4A). Interestingly, these branches were involved in a set of SPC sites including a parallel amino acid change (R $\rightarrow$ K) at site 123 of BURS α (Table 2). Likewise there were two branches showing relatively rapid evolution of BURS β : the branch from Node 12 to *Apis mellifera* and the branch from Node 13 to *Nasonia vitripennis* (Figure 4B). The latter branch was involved in a set of SPC sites including site 103 of BURS β (Table 2). But the branch from Node 12 to *Apis mellifera* was not involved in any SPC sites. Thus, though some SPC sites occurred on rapidly evolving branches, the majority of SPC sites did not occur on such branches; and rapid evolution did not seem to be a necessary factor for coordinated change in the present data.

It is well known that most amino acid sequence data sets reveal some degree of homoplasy, much of it due to chance convergent or parallel evolution not apparently related to adaptive function (Bollback and Huelsenbeck 2009; Rokas and Carroll 2008; Stewart 1993). The present analyses support the hypothesis that sites subject to functional constraints that limit the range of acceptable amino acid replacements may be an important factor contributing to the occurrence of parallel and convergent amino acid replacements because of limitation on acceptable options. This conclusion is consistent with that of a multi-genome study concluding that excess homoplasy is due mainly to substitutions between chemically similar amino acids (Rokas and Carroll 2009). On the other hand, there is evidence in the literature that in some cases parallel or convergent amino acid replacements may occur when homologous proteins adapt independently to similar functions (Kornegay et al. 1994; McCracken et al. 2009; Stewart et al. 1987; Wood et al. 2005; Yeager et al. 1997).

The present analyses revealed several cases of more than one parallel change occurring together on two or more different branches of the phylogeny. For example, three branches showed the R $\rightarrow$ K change at position 123 of BURS α ; and two of those branches (from Node 10 to *A. gambiae* and from Node 15 to *A. pisum*) also showed the change I $\rightarrow$ V at site 385 of BURSrec (Table 2). Because these coordinated changes involved both BURS α and portions of BURSrec (especially the LRR) believed to be involved in interactions with the ligand, it seems a plausible hypothesis that these changes are functionally related.

Available data do not make it possible to decide whether this example of parallel evolution of BURS α and BURSrec represents adaptive evolution in response to some environmental factor shared by *A. gambiae* and *A. pisum*. It is hard to imagine an environmental factor that might affect two species as different in lifestyle as *A. gambiae* and *A. pisum*. Moreover, the fact that the amino acids involved in both of these changes are chemically very similar

supports the hypothesis that this case involves selectively neutral mutations at functionally constrained sites. Alternatively, if a slightly deleterious amino acid replacement were fixed by chance in one of the two interacting proteins, a subsequent compensatory change might have been favored that improved the receptor-ligand interaction. Whatever evolutionary forces are responsible for coordinated amino acid replacements in this system, the results suggest that non-random patterns of coordinated amino acid replacement can provide information regarding potential interactions between amino acids both within a single polypeptide chain and between interacting polypeptide chains.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by grant GM43940 from the National Institutes of Health.

References

- Afonnikov DA, Kolchanov NA. CRASP: a program for analysis of coordinated substitutions in multiple alignments of protein sequences. *Nucl Acids Res.* 2004; 32:W64–W68. [PubMed: 15215352]
- Altschuh D, Vernet T, Berti P, Moras D, Nagai K. Coordinated amino acid changes in homologous protein families. *Protein Engineering.* 1988; 2:193–199. [PubMed: 3237684]
- Björkman J, Nagaev I, Berg OG, Hughes D, Andersson DI. Effects of environment on compensatory mutations to ameliorate costs of antibiotic resistance. *Science.* 2000; 287:1479–1482. [PubMed: 10688795]
- Bollback JP, Huelsenbeck JP. Parallel genetic evolution within and between bacteriophage species of varying degrees of divergence. *Genetics.* 2009; 181:225–234. [PubMed: 19001294]
- 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; 20:541–553. [PubMed: 12654933]
- Castoe TA, de Koning AP, Kim H-M, Gu W, Noonan BP, Naylor G, Jiang ZJ, Parkinson CL, Pollock DD. Evidence for an ancient adaptive episode of convergent molecular evolution. *Proc Natl Acad Sci USA.* 2009; 106:8986–8991. [PubMed: 19416880]
- Collette Y, Gilles A, Pontarotti P, Olive D. A co-evolution perspective of the TNFSF and TNFRSF families in the immune system. *Trends Immunol.* 2003; 24:387–394. [PubMed: 12860530]
- Cordoñer FM, Fares MA. Why should we care about molecular coevolution? *Evol Bioinf.* 2008; 4:29–38.
- Crawford H, Prado JG, Leslie A, Hué S, Honeybourne I, Reddy S, van der Stock M, Mncube Z, Brander C, Rousseau C, Mullins JI, Kaslow R, Goepfert P, Allen S, Hunter A, Hunter E, Mulenga J, Kiepiela P, Walker BD, Goulder PJ. Compensatory mutation partially restores fitness and delays reversion of escape mutation within the immunodominant HLA-B*5703-restricted Gag epitope in chronic human immunodeficiency virus type 1 infection. *J Virol.* 2007; 81:8346–8351. [PubMed: 17507468]
- Drosophila 12 Genomes Consortium. Evolution of genes and genomes on the Drosophila phylogeny. *Nature.* 2007; 450:203–218. [PubMed: 17994087]
- Eriksen KK, Hauser F, Schiøtt M, Pedersen K-M, Søndergaard L, Grimmelikhuijzen CJ. Molecular cloning, genomic organization, developmental regulation, and a knock-out mutant of a novel Leu-rich repeats-containing G protein-coupled receptor (DLGR-2) from *Drosophila melanogaster*. *Genome Res.* 2000; 10:924–938. [PubMed: 10899142]
- Farid NR, Szudlinski MW. Minireview: structural and functional evolution of the thyrotropin receptor. *Endocrinol.* 2004; 145:4048–4057.

- Fitch WM. Rate of change of concomitantly variable codons. *J Mol Evol.* 1971; 1:84–96. [PubMed: 4377447]
- Fitch WM. An improved method for determining codon variability in a gene and its application to the rate of fixation of mutations in evolution. *Biochem Genet.* 1970; 4:579–593. [PubMed: 5489762]
- Friedrich TC, Frye CA, Yant LJ, O'Connor DH, Kriewaldt NA, Benson M, Vojnov L, Dodds EJ, Cullen C, Rudersdorf R, Hughes AL, Wilson N, Watkins DI. Extraepitopic compensatory substitutions partially restore fitness to simian immunodeficiency virus variants that escape from an immunodominant cytotoxic-T-lymphocyte response. *J Virol.* 2004; 78:2581–2585. [PubMed: 14963161]
- Garriga C, Pérez-Elías MJ, Delgado R, Ruiz L, Nájera R, Pumarola T, Alonso-Socas MD, García-Bujalance S, Menéndez-Arias L. Mutational patterns and correlated amino acid substitutions in the HIV-1 protease after virological failure to nelfinavir- and lopinavir/ritonavir-based treatments. *J Med Virol.* 2007; 79:1617–1628. [PubMed: 17854027]
- Gruenheit N, Lockhart PJ, Steel M, Martin W. Difficulties in testing for covarion-like properties of sequences under the confounding influences of changing proportions of variable sites. *Mol Biol Evol.* 2008; 25:1512–1530. [PubMed: 18424773]
- Honegger HW, Dewey EM, Ewer J. Bursicon, the tanning hormone of insects: recent advances following the discovery of its molecular identity. *J Comp Physiol A.* 2008; 194:989–1005.
- Hughes AL, Yeager M. Coevolution of the mammalian chemokines and their receptors. *Immunogenetics.* 1999; 49:115–124. [PubMed: 9887348]
- Kimura M. The role of compensatory neutral mutations in molecular evolution. *J Genet.* 1985; 64:7–19.
- Kornegay JR, Schilling JW, Wilson AC. Molecular adaptation of a leaf-eating bird: stomach lysozyme of the hoatzin. *Mol Biol Evol.* 1994; 11:921–928. [PubMed: 7815930]
- Levin BR, Perrot V, Walker N. Compensatory mutations, antibiotic resistance and the population genetics of adaptive evolution in bacteria. *Genetics.* 2000; 154:985–997. [PubMed: 10757748]
- Loveall BJ, Deitcher DL. The essential role of bursicon during *Drosophila* development. *BMC Devel Biol.* 2010; 10:92. [PubMed: 20807433]
- Luo CW, Dewey EM, Sudo S, Ewer J, Hsu SY, Honegger HW, Hsueh AJ. Bursicon, the insect cuticle-hardening hormone, is a heterodimeric cysteine knot protein that activates G protein-coupled receptor LGR2. *Proc Natl Acad Sci USA.* 2005; 102:2820–2825. [PubMed: 15703293]
- McCracken KG, Barger CP, Bulgarella M, Johnson KP, Sonsthagen SA, Trucco J, Valqui TH, Wilson RE, Winkler K, Sorenson MD. Parallel evolution in the major haemoglobin genes of eight species of Andean waterfowl. *Mol Ecol.* 2009; 18:3992–4005. [PubMed: 19754505]
- Miyata T, Miyazawa S, Yasunaga T. Two types of amino acid substitutions in protein evolution. *J Mol Evol.* 1979; 12:219–236. [PubMed: 439147]
- Moyle WR, Campbell RK, Myers RV, Bernard MP, Han Y, Wang X. Co-evolution of ligand-receptor pairs. *Nature.* 1994; 368:251–255. [PubMed: 8145825]
- Park Y, Kim Y, Adams ME. Identification of G protein-coupled receptors for *Drosophila* PRXamide peptides, CCAP, corazonin, and AKH supports a theory of ligand-receptor coevolution. *Proc Natl Acad Sci USA.* 2002; 99:11423–11428. [PubMed: 12177421]
- Peabody NC, Diao F, Luan H, Wang H, Dewey EM, Honegger H-W, White BH. Bursicon functions within *Drosophila* CNS to modulate wing expansion behavior, hormone secretion, and cell death. *J Neurosci.* 2008; 28:14379–14391. [PubMed: 19118171]
- Pollock DD, Taylor WR, Goldman N. Coevolving protein residues: maximum likelihood identification and relationship to structure. *J Mol Biol.* 1999; 287:187–198. [PubMed: 10074416]
- Piana S, Carloni P, Rothlisberger U. Drug resistance in HIV-1 protease: flexibility-assisted mechanism of compensatory mutations. *Protein Sci.* 2002; 11:2393–2402. [PubMed: 12237461]
- Rokas A, Carroll SB. Frequent and widespread parallel evolution of protein sequences. *Mol Biol Evol.* 2008; 25:1943–1953. [PubMed: 18583353]
- Stephan W. The rate of compensatory evolution. *Genetics.* 1996; 144:419–426. [PubMed: 8878705]
- Stewart CB. The powers and pitfalls of parsimony. *Nature.* 1993; 361:603–607. [PubMed: 8437621]

- Stewart CB, Schilling JW, Wilson AC. Adaptive evolution in the stomach lysozymes of foregut fermenters. *Nature*. 1987; 330:401–404. [PubMed: 3120013]
- Swofford, DL. PAUP*: Phylogenetic Analysis Using Parsimony (*and other methods). Sinauer; Sunderland MA: 1999.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*. 28:2731–2739. [PubMed: 21546353]
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Diggins DG. The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res*. 1997; 25:4876–4882. [PubMed: 9396791]
- Vitt UA, Hsu Y, Hsueh AJ. Evolution and classification of cystine knot-containing hormones and related extracellular signal molecules. *Mol Endocrinol*. 2001; 15:681–694. [PubMed: 11328851]
- Wang H-C, Spencer M, Susko E, Roger AJ. Testing for covarion-like evolution in protein sequences. *Mol Biol Evol*. 2007; 24:294–305. [PubMed: 17056642]
- Whitlock MC, Griswold CK, Peters AD. Compensating for the meltdown: the critical effective size of a population with deleterious and compensatory mutations. *Ann Zool Fennici*. 2003; 40:169–183.
- Wiegmann BM, Trautwein MD, Kim J-W, Cassel BK, Bertone MA, Winterton SL, Yeates DK. Single-copy nuclear genes resolve the phylogeny of the holometabolous insects. *BMC Biol*. 2009; 7:34. [PubMed: 19552814]
- Wood TE, Burke JM, Rieseberg LH. Parallel genotypic adaptation: when evolution repeats itself. *Genetica*. 2005; 123:157–170. [PubMed: 15881688]
- Xia X, Xie Z. Protein structures, neighbor effect, and a new index of amino acid similarities. *Mol Biol Evol*. 2002; 19:58–67. [PubMed: 11752190]
- Yeager M, Kumar S, Hughes AL. Sequence convergence in the peptide-binding region of primate and rodent MHC class Ib molecules. *Mol Biol Evol*. 1997; 14:1035–1041. [PubMed: 9335143]

Highlights

The pattern of amino acid residue replacement in insect bursicon signaling proteins.

Sets of branches with perfectly concordant changes occurred more than expected by chance.

Functional constraint on amino acids was the major factor causing coordinated amino acid changes.

Figure 1.
Schematic phylogeny of species used in analyses.

A)

B)

Figure 2.

(A) The numbers of amino acid sites showing a given number of reconstructed changes in BURS α , BURS β , and BURS rec ($r = -0.761$; $P = 0.001$). (B) For 531 sites with at least one inferred change, the number of cases of perfect concordance as a function of the number of changes ($r = -0.555$; $P < 0.001$).

A)

B)

Figure 3.

(A) Mean chemical distances at sites with at least three different amino acids categorized by the occurrence of convergent and parallel changes; numbers of sites in each category are shown. (B) Mean number of amino acid changes at sites with at least three different amino acids categorized by the occurrence of convergent and parallel changes.

A)

B)

Figure 4.

Numbers of reconstructed changes along a given branch in (A) BURS α vs. that in BURSrec BURS α ; and (B) in BURS β vs. that in BURSrec. Open circles indicate branches with significant deleted-t residuals: * $P < 0.05$; *** $P < 0.001$.

Table 1

Sequences used in analyses.

Class	Order/species	BURSa	BURS β	BURSrec
Insecta	Hemiptera:			
	<i>Acyrtosiphon pisum</i>	XM_001946306	XM_001946263	XM_002423422
	Phthiraptera:			
	<i>Pediculus humanus</i>	XM_002430737	XM_00243076	XM_002423422
	Hymenoptera:			
	<i>Nasonia vitripennis</i>	NM_001162360	XM_001601112	XM_001605787
	<i>Apis mellifera</i>	NM_001098234	NM_00104262	XM_395206
	<i>Harpognathus saltator</i>	GL80834	GL450531	GL449658
	Coleoptera:			
	<i>Tribolium castaneum</i>	NM_001114307	NM_0012114308	XM_970421
	Diptera:			
	<i>Anopheles gambiae</i>	AY735443	XM_313804	XM_317111
	<i>Drosophila virilis</i>	XM_002053896	XM_002057554	XM_002057463
	<i>Drosophila grimshawi</i>	XM-001989717	XM_001992828	XM_001988339
	<i>Drosophila willistoni</i>	XM_002073879	XM_002066409	XM_002064731
	<i>Drosophila pseudoobscura</i>	XM_001358106	XM_001357344	XM_001356183
	<i>Drosophila ananassae</i>	XM_001953944	XM_01961495	XM_001962712
	<i>Drosophila erecta</i>	XM_001980987	XM_001968851	XM_002089223
	<i>Drosophila yakuba</i>	XM_002098151	XM_002089257	XM_002089223
	<i>Drosophila melanogaster</i>	NM_142726	NM_135868	NM_057354
	<i>Drosophila simulans</i>	XM_002104285	XM_002079417	XM_002079386
	<i>Drosophila sechellia</i>	XMJ_002032067	XM_002035792	XM_002035757
Arachnida	Ixodida:			
	<i>Ixodes scapularis</i>	XM_002407468	XM_002407649	XM_002407568

Table 2

Amino acid changes at significant perfectly concordant (SPC) sites. Sites concordant on three branches:

Protein	Site ¹	Region	Branch ²		
			Node 10 → Node 9	Node 11 → <i>T. castaneum</i>	Node 15 → <i>A. pisum</i>
BURS _{rec}	380	LRR	N → S	N → G	N → G
BURS _{rec}	640	LRR	T → S	T → D	T → A
BURS _{rec}	647	LRR	S → T ³	S → N	S → T
			Node 10 → <i>A. gambiae</i>	Node 13 → Node 12	Node 15 → <i>A. pisum</i>
BURSa	123	Between C8 and C9	R → K	R → K	R → K
BURS _{rec}	385	LRR	I → V	I → L	I → V
			Node 10 → Node 9	Node 13 → Node 12	Node 15 → <i>A. pisum</i>
BURSa	151	C-terminal	V → I	V → I	I → V
BURS _{rec}	417	LRR	R → K	R → K	N → K
			Node 11 → Node 10	Node 14 → Node 13	Node 15 → <i>A. pisum</i>
BURSβ	59	C-terminal	E → D	E → D	E → D
BURS _{rec}	1284	C-terminal cytoplasmic	G → P	G → S	G → N
			Node 14 → Node 13	Node 13 → <i>N. vitripennis</i>	Node 12 → <i>H. saltator</i>
BURSβ	103	Between C7 and C8	T → H	H → N	H → S
BURS _{rec}	239	LRR	L → A	A → I	A → W
			Node 11 → <i>T. confusum</i>	Node 14 → Node 13	Node 15 → <i>P. humanus</i>
BURSβ	134	Between 10 and C11	F → Y	F → I	F → S
BURS _{rec}	673	LRR	D → T	D → E	D → E
			Node 10 → Node 9	Node 14 → Node 13	Node 15 → <i>P. humanus</i>
BURS _{rec}	271	LRR	F → L	F → A	F → L
BURS _{rec}	400	LRR	K → R	K → R	A → V
			Node 10 → <i>A. gambiae</i>	Node 14 → Node 11	Node 15 → <i>A. pisum</i>
BURS _{rec}	440	LRR	A → S	S → A	S → A
BURS _{rec}	922	LRR	T → A	A → T	A → V
			Node 11 → Node 10	Node 10 → Node 9	Node 14 → Node 13
BURS _{rec}	590	LRR	N → H	H → K	N → A

Protein	Site ¹	Region	Branch ²	
BURSec	740	LRR	L → E	E → M
			Node 10 → Node 9	Node 13 → <i>N. vitripennis</i>
BURSec	1134	C-terminal cytoplasmic	Q → P	Q → R
BURSec	1153	C-terminal cytoplasmic	V → I	V → M

Sites concordant on four branches:				
Protein	Site	Region	Branch	
			Node 10 → <i>A. gambiae</i>	Node 13 → <i>H. saltator</i>
BURSec	267	LRR	D → E	D → K
BURSec	698	LRR	M → P	M → V
			Node 10 → <i>A. gambiae</i>	Node 13 → <i>H. saltator</i>
BURSec	631	LRR	S → F	S → T
BURSec	633	LRR	V → I	V → L

Sites concordant on five branches:				
Protein	Site	Region	Branch	
			Node 11 → Node 10	Node 14 → Node 13
BURSec	61	Between C2 and C3	A → I	A → Q
BURSec	LRR		H → Q	H → L
			Q → R	L → M
				Node 15 → <i>P. humanus</i>
				A → S
				H → Q

Sites are numbered according to the *Drosophila melanogaster* sequences.