An Un unusually Simple HP1 Gene Set in Hymenopteran Insects

<table>
<thead>
<tr>
<th>Journal:</th>
<th>Biochemistry and Cell Biology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>bcb-2015-0046.R2</td>
</tr>
<tr>
<td>Manuscript Type</td>
<td>Article</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>17-Jun-2015</td>
</tr>
<tr>
<td>Complete List of Authors:</td>
<td>Fang, Christopher; Claremont McKenna, Pitzer, and Scripps Colleges, W. M. Keck Science Department  Schmitz, Lars; Claremont McKenna, Pitzer and Scripps Colleges, W. M. Keck Science Department  Ferree, Patrick; Claremont McKenna, Pitzer, and Scripps Colleges, W. M. Keck Science Department</td>
</tr>
<tr>
<td>Keyword:</td>
<td>Hymenoptera, HP1, chromodomain, chromoshadow domain, heterochromatin</td>
</tr>
</tbody>
</table>
Title:
An Unusually Simple HP1 Gene Set in Hymenopteran Insects

Authors:
C. Fang, L. Schmitz, and P. M. Ferree*

Author Affiliation
W. M. Keck Science Department, Claremont McKenna, Pitzer, and Scripps
Colleges, Claremont, CA, USA 91711

*Correspondence:
pferree@kecksci.claremont.edu
Tel. 909-607-8304

Important Keywords: Hymenoptera; HP1; chromodomain; chromshadow domain;
heterochromatin
Abstract

The heterochromatin protein 1 (HP1) gene family includes a set of paralogs in higher eukaryotes that serve fundamental roles in heterochromatin structure and maintenance, and other chromatin-related functions. At least ten full and sixteen partial HP1 genes exist among Drosophila species, with multiple gene gains, losses and sub-functionalizations within this insect group. An important question is whether this diverse set of HP1 genes and their dynamic evolution represent the standard rule in eukaryotic groups. Here we have begun to address this question by bioinformatically identifying the HP1 family genes in representative species of the insect order Hymenoptera, which includes all ants, bees, wasps and sawflies. Compared to Drosophila species, hymenopterans have a much simpler set of HP1 genes, including one full and two partial HP1s. All three genes appear to have been present in the common ancestor of the hymenopterans and they derive from a Drosophila HP1B-like gene. In ants, a partial HP1 gene containing only a chromoshadow domain harbors amino acid changes at highly conserved sites within the PxVxL recognition region, suggesting that this gene has undergone sub-functionalization. In the jewel wasp *Nasonia vitripennis* the full HP1 and partial chromoshadow-only HP1 are expressed in both germ line and somatic tissues. However, the partial chromodomain-only HP1 is expressed exclusively in the ovary and testis, suggesting that it may have a specialized chromatin role during gametogenesis. Our findings demonstrate that the HP1 gene family is much simpler and evolutionarily less dynamic within the hymenopterans compared to the much younger Drosophila group, a pattern that may reflect major differences in the range of chromatin-related functions present in these and perhaps other insect groups.
Introduction

As much as fifty percent of the eukaryotic genome consists of heterochromatin, a highly condensed form of chromatin that comprises the centromeres, telomeres, and their surrounding sequences (Bosco et al., 2007; Eissenberg and Reuter, 2009; Rayburn et al., 1992). A defining feature of heterochromatin is its assembly of associating proteins, a primary one being the heterochromatin protein 1 (HP1). This protein was discovered biochemically in *Drosophila melanogaster* through its tight association with DNA (James and Elgin, 1986). The functional importance of HP1 was revealed by mutational analysis in this model organism; mutations in this gene cause severe genome instability and transcriptional de-repression of genes located near heterochromatin (Eissenberg et al., 1990; Kellum and Alberts, 1995). These findings led to the suggestion that HP1 protein facilitates the condensed and transcriptionally silenced state of heterochromatin and surrounding regions (Lomberk et al., 2006). HP1 contains two functionally important domains: a chromodomain (CD) located near the N-terminus that associates with histone H3 methylated at lysine residue 9 (Jacobs and Khorasanizadeh, 2002; Lachner et al., 2001; Nakayama et al., 2001), an additional defining feature of heterochromatin. The second is a chromoshadow domain (CSD) located toward the C-terminus, which facilitates HP1 dimerization and interaction with accessory proteins such as SU(VAR)3-9, a histone H3 methylase (Schotta et al., 2002; Smothers and Henikoff, 2000). This interaction is thought to maintain both histone H3 methylation and HP1 association with this post-translational mark, together maintaining the higher order state of heterochromatin (Ebert et al., 2004; Schotta et al., 2002).

Subsequent studies revealed that HP1 belongs to a family of paralogs within the eukaryotes. Analysis of the original *D. melanogaster* genome sequence led to the discovery of two additional HP1 genes, and the three genes were named HP1A (the original one), HP1B, and HP1C (Adams et al., 2000; Smothers and Henikoff, 2001). Like HP1A, HP1B and HP1C also contain a chromodomain and a chromoshadow domain. However, all three of the encoded proteins exhibit unique chromatin localization patterns; whereas HP1A localizes across constitutive heterochromatin (Fanti et al., 2003; James et al., 1989), HP1C associates with transcriptionally active regions of euchromatin while HP1B localizes to regions of heterochromatin and euchromatin and
may be partially redundant in function with HP1A and HP1C (Font-Burgada et al., 2008; Smothers and Henikoff, 2001). Simpler eukaryotes such as the ciliated protist Tetrahymena thermophila, the bread mold Neurospora crassa, and the mustard Arabidopsis thaliana each have a single HP1-like gene (Freitag et al., 2004; Huang et al., 1998; Mylne et al., 2006). However, higher-eukaryotes more commonly contain multiple HP1 paralogs. For example, mice and humans have HP1α, HP1β, and HP1γ, which possess a range of chromatin localization patterns and functions reminiscent of the HP1s in flies (Minc et al., 1999). Thus, it is possible that HP1 has expanded in eukaryotes with more complex genomes to perform different sub-chromatin-level functions.

This idea has been further supported by the identification of additional HP1 genes, particularly in D. melanogaster and other Drosophila species. Two genes, HP1D (also known as Rhino) and HP1E, are expressed uniquely in the female and male germ line tissues, respectively, of D. melanogaster (Vermaak et al., 2005; Volpe et al., 2001). HP1D plays a role in transposon silencing in the ovary (Klattenhoff et al., 2009), while the function of HP1E in the testis currently is unknown. A recent study uncovered a number of new HP1 genes in the genomes of multiple Drosophila species, increasing the number of HP1 paralogs within the Drosophila genus to ten (Levine et al., 2012). Additionally, sixteen partial HP1 genes, which contain either a CD or CSD, have been discovered in flies (Levine et al., 2012; Vermaak and Malik, 2009). One CSD-only HP1, Umbrea, which is most closely related to HP1B, localizes to the centromeres in somatic cells (Levine et al., 2012; Ross et al., 2013). Umbrea’s centromere association likely occurs due to evolutionary changes at conserved amino acid sites within the CSD that are responsible for recognition of the degenerate PxVxL motif in interacting proteins (Ross et al., 2013). Additionally, in the polytene nuclei of salivary glands, Umbrea co-localizes with HP1A at the heterochromatic chromocenter and also with the telomere-capping protein (Joppich et al., 2009). Phylogenetic analysis of HP1 genes across Drosophila species showed a dynamic pattern of HP1 gene duplication and loss within this group (Levine et al., 2012). Together, these patterns suggest that the HP1s both within each Drosophila genome as well as those found across the Drosophila genus are playing multiple, unique chromatin-based roles in these organisms, although many of these functions remain to be elucidated.
An important question borne from these observations is whether the dynamic pattern of HP1 diversification in the Drosophila genus is typical of other eukaryotic genomes or is instead an exception to a simpler, more common pattern. Investigating the dynamics of HP1 genes in other biological groups will be important for answering this question. In this study we have bioinformatically identified the HP1 family genes present in the genomes of representative species belonging to the prominent insect order Hymenoptera, which includes ants, bees, wasps, and sawflies. In contrast to the Drosophila species, each examined species within the four major hymenopteran groups contains one full HP1 gene and two partial HP1 genes (except for bees, which have only one partial HP1). One partial HP1 gene contains only a CSD. The second partial gene encodes two CDs near the protein’s C-terminus; to our knowledge, this is the first description of a two-CD-containing HP1 gene. Phylogenetic analysis of the HP1 genes suggests that the common ancestor of these hymenopteran species contained all three of these genes, arguing that little HP1 diversification has occurred within this group. Like Umbrea in *D. melanogaster*, the partial CSD-only genes in the two ant species differ from the other orthologs at several highly conserved amino acid positions in the PxVxL recognition region, suggesting that these genes have undergone sub-functionalization. In *N. vitripennis*, the full HP1 and partial CSD-only HP1 are expressed in both germ line and somatic tissues, whereas the partial CD-only HP1 exhibits germ-line-specific expression. These findings highlight distinct patterns of HP1 diversification between the drosophilids and the hymenopterans, and they provide a platform for future studies aimed at discerning the sub-functions of these important chromatin-associated genes.

**Results**

*Identification of one full and two partial HP1 genes in the hymenopterans*

We sought to identify the HP1 genes present in a representative group of hymenopteran insects with sequenced genomes. We included in our analyses two ants (*Solenopsis invicta* and *Linepithema humile*), two bees (*Apis mellifera* and *Bombus*).
terrestris), and two sawflies (Orussus abietinus and Athalia rosae). For wasps, we included Nasonia vitripennis and its two sister taxa, N. giraulti and N. longicornis, as well as the fig wasp Ceratosolen solmsi marchali. Based on previous phylogenetic studies, the sawflies are ancestral to the group that includes ants, bees and wasps (Figure 1A) and the common hymenopteran ancestor of these taxa dates back to 250-350 mya (Johnson et al., 2013; Ronquist et al., 2012b). In order to identify the hymenopteran HP1 genes, we BLAST-searched the CDs and CSDs from the five full-length D. melanogaster HP1 genes against the genomes of these representative hymenopteran species (see Methods for details). To be consistent with previous studies, we defined any top hits containing both a CD and a CSD as a full HP1 gene (Levine et al., 2012). We also considered genes with only a CSD as partial HP1s because CSDs are found only in HP1 genes (Aasland and Stewart, 1995; Levine et al., 2012; Smothers and Henikoff, 2001). However, because a number of non-HP1 genes that perform chromatin-related functions, such as polycomb and chd-1, also contain CDs (Messmer et al., 1992; Woodage et al., 1997), we considered any new CD-only genes as partial HP1s only if they best matched CDs from known HP1 genes and not CDs from other chromatin genes.

By using these criteria and approach, we uncovered one full HP1 gene (referred to from here as HP1-full) in each of the hymenopteran genomes examined (Figure 1B; Supplemental Table 1). The HP1-full genes encode proteins with a hinge region between the CD and CSD, as well as small regions of flanking sequence on the N- and C-terminal sides of these domains (Figure 1B), similar to HP1s in other eukaryotes (Vermaak and Malik, 2009). We also found one partial CSD-only gene (referred to from here as HP1-csd) in each of the ant, wasp, and sawfly species (Figure 1B; Supplemental Table 2). However, this gene is absent from both bee genomes. Because sawflies are ancestral to the other hymenopterans (Ronquist et al., 2012b), it is likely that the HP1-csd gene was present in the common hymenopteran ancestor but was lost in the bee lineage.

In addition, each hymenopteran genome contains one partial CD-only gene. Interestingly, in each genome this gene contains two CDs toward the C-terminus of the encoded protein (thus we refer to this gene from here onward as HP1-cdx2) (Supplemental Table 3). The presence of two CDs in this gene is in contrast to all other
known HP1 genes, which contain only one CD. Certain chd-1 paralogs are known to contain two CDs in addition to a DEAH-box helicase domain and an ATP-binding domain (Hall and Georgel, 2007). However, the HP1-cdx2 gene does not contain these or any other identifiable domains and, moreover, both CDs best-BLAST to the *N. vitripennis* HP1-full gene and not to *N. vitripennis* chd-1 (Supplemental Table 3). Reciprocal best-BLASTing to the *D. melanogaster* and hymenopteran genomes revealed that all hymenopteran HP1 genes – the full and two partials from each hymenopteran species – are more similar to the fly HP1B gene than to any other HP1 paralog (Supplemental Tables 1-3). The only exception to this pattern is the more C-terminal CD of the HP1-cdx2 genes, which better matches the CD of HP1C (Supplemental Table 3). The presence of a full HP1 gene and either one or two partial HP1 genes in each examined hymenopteran species is in striking contrast to the larger set of paralogs present in the previously sequenced Drosophila species genomes (Clark et al., 2007; Levine et al., 2012; Vermaak and Malik, 2009).

**Phylogenetic analysis of hymenopteran HP1 genes**

We sought to address the evolutionary relationships of the hymenopteran HP1 genes. In order to achieve this goal we performed phylogenetic analyses with the CDs and the CSDs of these genes (Figure 2A,B). Because each species in our study contains one HP1-full, one HP1-cdx2, and one HP1-csd (with the exception of the bees which are devoid of this latter gene), we hypothesized that each of these genes belongs to a set of orthologs across the hymenopteran insects, and that the common ancestor of the hymenopteran species in this study already contained these three genes. If this were true, one would expect to find a tree with three well-supported clades of HP1 genes. An alternative scenario is that some of the HP1 genes arose subsequently to the divergence of the hymenopteran common ancestor. In this latter case, multiple duplication events would be required to explain the observed pattern of all HP1s across all hymenopterans. Given that the duplication events would have occurred more recently and the orthologs would have had less time available to accrue differences in amino acid sites, it is expected that the more recently produced HP1 partials would display similarities with their closest HP1 gene relatives, whether they be full or partial HP1s. Hence, the resulting tree would group the HP1 partials of a given organism with the HP1-fulls of that same organism.
Our phylogenetic analyses uncovered several specific trends. First, when either the CDs or CSDs were used for tree building, the HP1-full genes formed a well-supported monophyletic group (Figure 2A,B). In some clades the relationships among these domains do not reflect the patterns existing in the species phylogeny (Figure 1A), a discrepancy that may result from non-neutral evolutionary influences such as functional constraints on certain amino acid sites within these domains. In addition, these genes formed a close sister group with *D. melanogaster* HP1B (Figure 2A,B), a pattern that is consistent with our BLAST results indicating the similarity of each hymenopteran HP1 gene to *D. melanogaster* HP1B. Second, each of the two CDs of the HP1-cdx2 genes formed strongly supported monophyletic groups (Figure 2A). In the CD tree, the non-HP1 chromatin remodeling genes, polycomb and chd-1, from *N. vitripennis* and *D. melanogaster* clustered together in a group that falls outside of either HP1-full or HP1-cdx2 groups (Figure 2A). This finding supports our previous conclusion that the HP1-cdx2 genes are indeed members of the HP1 family and not of chromatin remodeling gene families despite their unusual character of having two CDs. Third, the HP1-csd genes form a monophyletic (*i.e.*, completely unified) group, although the support for this clustering is not as strong as for other groups (Figure 2B). This effect may result in part from the higher divergence levels of these genes as indicated by longer branch lengths, compared to branch lengths of the HP1-full genes (see also below). In essence, we found no evidence that any hymenopteran HP1 genes were more closely related to HP1 genes outside of their respective group. Taken together, these patterns strongly argue that the common ancestor of the hymenopterans contained all three of these HP1 genes, and that minimal diversification within this gene family has occurred since the appearance of the major groups within the Hymenoptera. Theoretically, the resulting tree may also be the result of gene convergence, a mechanism that may be more widespread than previously thought (Parker et al., 2013). However, the number of required evolutionary events required for convergence makes this scenario very unlikely.

The clustering of the three hymenopteran HP1 genes into discrete monophyletic groups suggests that the genes within each group are orthologous to one another. To further address this possibility, we tested for chromosomal synteny of each HP1 gene by looking for similarity of neighboring genes across the hymenopterans. Because several
of the genomes in our study were not assembled, we focused on the assembled genomes of the ant *S. invicta*, the honeybee *A. mellifera*, and the jewel wasp *N. vitripennis*. In all three of these species, the HP1-full gene is flanked on the 3’ side by the same gene, a CCAAT enhancer binding protein zeta, in each genome (Figure 3). In the ant and bee genomes, a common gene encoding isocitrate dehydrogenase resides on the 5’ side of HP1-full (Figure 3). However, a different gene is present in this position within the wasp genome (Figure 3). For the HP1-cdx2 gene, we found one common gene, RPL38, on the 3’ side in both ant and bee but not wasp (Figure 3). No similarity was observed among the genes present on the 5’ side of the HP1-cdx2 gene (Figure 3) or on either side of the HP1-csd gene (not shown) across all three species. These results and those from our phylogenetic analyses together suggest that the HP1-full, HP1-cdx2, and HP1-csd genes represent orthologs across hymenopterans, although little synteny exists for the two partial HP1 genes, likely due to chromosomal rearrangements over evolutionary time.

**Tissue expression patterns of HP1 genes in *N. vitripennis***

Given that several HP1 genes in *D. melanogaster* exhibit sex- and tissue-specific expression patterns (Levine et al., 2012; Vermaak et al., 2005; Volpe et al., 2001), we addressed if any of the hymenopteran HP1 genes show a similar pattern. In particular, we focused on HP1 expression in *N. vitripennis*, a rising model system for hymenopteran insects (Akbari et al., 2013; Werren and Loehlin, 2009). By using RT-PCR, we found that the HP1-full and HP1-csd genes are expressed in the germ line and adult somatic tissues of both sexes, suggesting that these genes perform functions common between these tissue types. In contrast, the HP1-cdx2 gene is expressed exclusively in the ovary and testis (Figure 4). Additionally, PCR amplification of cDNAs from the HP1-cdx2 gene produced two bands that differ by ~200bp, suggesting that the pre-mRNA of this gene may be differentially spliced. The germ line-specific expression of the HP1-cdx2 gene offers the intriguing possibility that in *N. vitripennis* this gene may serve unique chromatin-related roles during gametogenesis.

**Divergence of the HP1-csd gene in the ant lineage**
The branch lengths of several of the HP1-csd genes, including those of the two ant species, appear longer than branches of the HP1-full and HP1-cdx2 genes (Figure 2A,B), demonstrating that the CSDs in the HP1-csd genes have undergone greater sequence change over evolutionary time. We therefore carefully examined the encoded CSD sequences of the hymenopteran HP1-csd proteins in order to identify the specific amino acid changes in these domains. Sites comprising the chromoshadow fold, which are important for HP1 protein dimerization, are largely unchanged among all hymenopteran HP1-csd genes (Figure 5). For the wasps and sawflies, the PxVxL recognition region of the HP1-csd does not vary in sequence relative to the HP1-full gene of *N. vitripennis* and HP1B of *D. melanogaster*. In contrast, two sites within this region in the ant *Linepithema humile* have undergone non-synonymous amino acid changes (Figure 5). Additionally, four sites within this region have changed in the ant species *Solenopsis invicta* (Figure 5). Interestingly, two of these altered amino acids (positions 15 and 57) are the same sites that have changed in the *D. melanogaster* Umbrea protein (Ross et al., 2013), although the specific amino acid changes are not the same as in Umbrea (Figure 5). Nevertheless, this finding raises the possibility that, like Umbrea, the ant HP1-csd genes may have undergone sub-functionalization from their counterparts in the other hymenopteran species.

**Discussion**

In this study we found that in comparison to mammals and Drosophila species, the hymenopteran insects have substantially fewer HP1 genes: one full and two partials (only one partial for the bee species). Additionally, the HP1 gene family lacks the dynamic pattern of diversification in hymenopterans compared to Drosophila species, with phylogenetic evidence pointing towards a deep origin of the three HP1 genes within hymenopterans or even earlier. It is possible that the smaller repertoire of HP1 genes present in hymenopteran species signifies a less complex set of chromatin-based functional requirements, compared to organisms with more HP1 genes. Genome size *per se* does not seem to drive this pattern. For example, *N. vitripennis* contains 250 mega-bases in its genome, approximately twice that found in *D. melanogaster* (Adams et al., 2000; Bosco et al., 2007; Desjardins et al., 2013; Werren et al., 2010). However,
some organisms may have a wider array of chromatin sub-types, particularly within heterochromatin, that could necessitate more HP1 genes. For example, Drosophila and mammals have sex chromosomes, which contain large amounts of simple and complex satellite sequences in their heterochromatin (Bonaccorsi and Lohe, 1991; Hoskins et al., 2002; Lohe et al., 1993). The presence of sex chromosomes in these organisms may add to the complexity of their genomes through multiple unique chromatin-related processes. In contrast, hymenopteran insects, as well as many non-hymenopteran insects including mealybugs, do not have sex chromosomes and could therefore have less complex genomes regarding sub-chromatin/heterochromatin types. If true, then one or just a few HP1 paralogs in these organisms may be needed to perform a smaller number of chromatin-related functions.

Alternatively, one or a few HP1 genes could perform multiple different chromatin functions in hymenopteran genomes through association with other chromatin proteins. This idea is supported by the fact that in \textit{D. melanogaster}, HP1A interacts with several different proteins including SU(VAR)3-9, SU(VAR)3-7, Lethal Hybrid Rescue (LHR), and HP2 (Brideau et al., 2006; Delattre et al., 2000; Schotta et al., 2002; Shaffer et al., 2002; Smothers and Henikoff, 2001). It is conceivable that in hymenopterans, a range of interacting partners could give the single HP1-full protein the ability to change specificity for different sub-functions. This effect also could be achieved by different phosphorylation states during development and in different tissue types. HP1A becomes differentially phosphorylated across multiple sites in the early \textit{D. melanogaster} embryo (Eissenberg et al., 1994), and a subset of these protein phosphorylation states seems to play a role in DNA/chromatin binding (Zhao et al., 2001). Likewise the hymenopteran HP1-full protein harbors a number of potential phosphorylation sites (data not shown), making differential phosphorylation and/or association with different binding proteins an option for performing multiple chromatin functions.

An intriguing question is what functional roles the partial HP1s in particular play in hymenopterans. Despite the fact that many partial HP1s exist in Drosophila species (Levine et al., 2012; Vermaak and Malik, 2009), very little is known about their functions. In fact, the only detailed analysis of a partial HP1 focused on the \textit{D. melanogaster} gene Umbrea (Levine et al., 2012). The encoded protein of this gene localizes specifically to the centromeres in culture cells (Ross et al., 2013) and with HP1A at the chromocenter...
and the telomere-capping protein HOAP in polytene nuclei of salivary glands (Joppich et al., 2009). Removal of Umbrea through RNAi causes telomere fusions in the polytene nuclei of salivary glands, demonstrating that Umbrea protein is essential for telomere maintenance (Joppich et al., 2009). Interestingly, the PxVxL recognition sequence within the CSD of Umbrea has changed at three highly conserved sites; conversion of these sites to those found in the PxVxL consensus sequence causes loss of Umbrea from the centromeres (Ross et al., 2013), demonstrating that these evolutionary changes at least partly dictate the ability of Umbrea protein to undertake its centromere localization. These previous studies of Umbrea function and our findings presented here provide a basis for thinking about possible functions of the hymenopteran HP1-csd orthologs. In particular, because the ant HP1-csd orthologs have diverged at conserved PxVxL sites, it is possible that they, like Umbrea, perform centromere-related roles. In contrast, the HP1-csd orthologs of the non-ant species in our study share very high sequence identity at PxVxL recognition and chromoshadow-fold sites with the HP1-full of hymenoperans as well as D. melanogaster HP1s (Figure 5). Thus, these genes may have chromatin-related functions that are more similar to the HP1-full genes.

The hymenopteran HP1-cdx2 orthologs are interesting for multiple reasons. First, they contain two CDs, in contrast to all other previously described HP1-family genes, which have only one CD. The two chromodomains of the chromatin remodeling enzyme CHD-1 appear to be important for its activity and less so for its recognition of chromatin and localization; mutations in the CDs of CHD-1 do not affect its association with RNA Pol II or methylated histone H3 at lysine residue 4, but instead, they disrupt the ability of this enzyme to remodel chromatin in vitro and properly stimulate transcription (Morettini et al., 2011). Unlike the chd-1 orthologs in flies and wasps, the hymenopteran HP1-cdx2 orthologs do not contain an ATP-binding motif, a DEAH-box helicase domain, or any other recognizable domains. It is possible, therefore, that the two CDs in the encoded protein function solely in chromatin recognition unlike in CHD-1. Alternatively, the encoded HP1-cdx2 protein may associate with other proteins with enzymatic activity for function. Second, the HP1-cdx2 gene in N. vitripennis is expressed in the germ line but not somatic tissues, arguing that the encoded protein plays a chromatin role in gametogenesis. And third, the transcripts of this gene appear to be present in two forms, which may represent differentially spliced variants. Molecular studies will be required to further investigate this possibility and test for functional differences between any protein.
isoforms resulting from alternative splicing.

Methods

Bioinformatical analyses

We identified HP1 orthologs by BLAST-searching concatenated amino acid sequences of the chromodomains (CDs) and the chromoshadow domains (CSDs) of the five HP1 gene family members (HP1A, HP1B, HP1C, HP1D, and HP1E) from *Drosophila melanogaster* against the nucleotide genome databases of the following species: *Nasonia vitripennis, Nasonia giraulti, Nasonia longicornis, Ceratosolen solmsi marchali, Solenopsis invicta, Linepithema humile, Apis mellifera, Bombus Terrestris, Orussus abietinus, and Athalia rosae*. The default settings were used under tBLASTn for these searches. This approach is very similar to that used in a recent genomic study of HP1 family genes across drosophilid species (Levine et al., 2012). From this point we excluded all hits with an e-value higher than 0.1. The domain of each gene with an e-value of 0.1 or lower was then BLAST-searched against the *D. melanogaster* genome. A hymenopteran gene was considered to be an HP1 if it (i) best matched a *D. melanogaster* HP1 gene and (ii) retrieved the same hymenopteran gene used for the original query when re-BLAST-searched against the respective hymenopteran genome. The CDs and CSDs from newly identified hymenopteran HP1 genes were then added to our concatenated query sequences to further BLAST-searched against the hymenopteran genomes in order to best insure that we uncovered all likely matches. This iterative BLAST searching was repeated until no new HP1 genes were found in the queried genomes. We confirmed the CDs and/or CSDs in the newly discovered HP1 genes by using Pfam (Punta et al., 2012). This search strategy assumes completeness of each queried genome in order to identify all HP1-family genes, an assumption that may not be met due to variability in genome quality across species queried. Additionally, our search did not include querying with the less conserved “hinge” and “tail” regions that lie outside the CD and CSD domains. Thus, our study could have missed some HP1 family genes in the queried hymenopteran genomes.

*Nomenclature of new HP1 family genes*
None of the hymenopteran HP1 genes uncovered in our study were previously annotated as HP1 genes in the genome databases. To be consistent with previous studies (Levine et al., 2012), we considered any gene containing both a CD and a CSD to be a full HP1 gene. In order to distinguish among HP1 orthologs, we use the first letter of the genus name capitalized followed by the first three letters of the species name in lowercase to precede the HP1 name (e.g., Nvit HP1-full). Orthologs for the partial HP1 genes were named based on whether they contained a CD or CSD. Additionally, all partial CD-containing genes contained two CDs; therefore, we named these genes by using the genus/species distinction (as above) followed by HP1 and either cdx2 or csd (e.g., Nvit HP1-cdx2 or Nvit HP1-csd, respectively).

**Phylogenetic analyses of hymenopteran HP1 genes**

Phylogenetic trees were generated by using the MCMC package BEAST v1.8.1. Hand-aligned amino acid sequences for the entire CDs and CSDs (Supplemental Tables 1, 2 and 3) were used as data sets for building trees. An uncorrelated log normal relaxed clock and the Blosum62 substitution model were used for these analyses (Henikoff and Henikoff, 1992). The MCMC Chains were run for 100,000,000 generations and the Tracer program was used to assess acceptability of the traces and effective sample size (ESS) of the parameters (http://tree.bio.ed.ac.uk/software/tracer) as well as the stability of the trace pattern. The first 25% of each MCMC run was discarded as ‘burn-in.’ Maximum clade credibility trees were generated from posterior tree samples. Each tree was repeated by running five independent runs in order to ensure consistency of the patterns.

**Gene expression analyses of HP1 genes in Nasonia vitripennis**

RT-PCR was used to assess gene expression in the germ line and somatic tissues of *N. vitripennis*. Ovaries and testes were dissected in 1x phosphate-buffered saline (PBS). RNA was extracted from tissues by using Trizol Reagent (Ambion) based on the manufacturer’s protocol. Extracted RNA samples were treated with DNase and subsequently with DNase Inactivator Reagent (Ambion). Complementary DNAs (cDNA) were made by using the iScript cDNA Synthesis Kit (BioRad). PCR amplification was
performed with the following primer sequences: Nvit HP1-full, 5’-GTC GAC CGA GGG AGA AAC GG-3’ and 5’-CCG CCT TCA TCG TCG TGA AC-3’; Nvit HP1-cdx2, 5’-CCA GAG GAA AGC AAA GAA AGT CAG-3’ and 5’-GCG CCG AGA T TG TCT CTT ATC TG-3; Nvit HP1-csd, 5’-GGA GAA AAA GAT TGA CAA TAA AGC TGC C-3’ and 5’-GGT TGA CTT GGC AGT CTG CC-3’. All reactions were conducted by using a 54°C annealing temperature and 35 cycles.

Acknowledgements

We thank Mia Levine and Bryan Thines for critical input during the development of this study. An NSF CAREER award (NSF-1451839) made to Patrick M. Ferree supported this work.

Figure Legends

Figure 1. Hymenopteran species harbor three HP1 genes. (A) A tree depicts the general relationships among some of the major groups within the Hymenopteran insect order, based on several recent studies (Johnson et al., 2013; Ronquist et al., 2012a). The tree does not represent branch length distances to scale. (B) A schematic symbolizes the three hymenopteran HP1 genes. Green boxes represent chromodomains; red boxes symbolize chromoshadow domains. Black lines indicate protein sequences outside of these domains. The amino acid residue lengths of the Nasonia vitripennis proteins are shown in parentheses.

Figure 2. Phylogenies depict the evolutionary relationships among the hymenopteran HP1 genes. The trees were constructed by using the amino acid sequences of the (A) chromodomains or (B) chromoshadow domains of the HP1 genes from each species examined in the study. For clarity, only the posterior probability values for the major nodes are shown. Asterisks mark the unified CD and CSD clades across all of the species in each phylogeny, indicating the monophyly of these groups. The root was
arbitrarily assigned to the branch joining the HP1-full gene groups and the partial gene groups. Scale bars indicate the number of expected amino acid substitutions per site.

**Figure 3.** A schematic symbolizing the limited level of synteny among the (A) HP1-full and (B) HP1-cdx2 orthologs for *N. vitripennis* (wasp), *Apis mellifera* (bee), and *Solenopsis invicta* (ant). Genes are depicted as thick arrows, with the direction of the arrow indicating the 5’ to 3’ orientation of each gene. The HP1 genes are shown as red arrows, while green and blue arrows represent likely orthologous genes to the left and right of HP1 genes, respectively. Black arrows depict non-homologous genes. Numbers in parentheses correspond to genome LOC identifiers.

**Figure 4.** Expression patterns for the HP1 genes in the germ line and somatic tissues of the jewel wasp *Nasonia vitripennis*. Non-quantitative reverse transcription PCR was used to assay gene expression of each HP1 gene. Single bands are seen for products of the HP1-full and HP1-csd genes. A pair of bands appears from amplification of the HP1-cdx2 gene’s cDNA, which may indicate alternative splicing of the pre-mRNA. The first lane in each gel shows the size ladder, with arrows indicating the approximate size of the amplified products.

**Figure 5.** A schematic depicting a sequence alignment of the chromoshadow domains of the hymenopteran HP1-csd orthologs. The PxVxL sites are shaded in green, while the chromofold sites are shaded in gray. Numbers mark every tenth amino acid within the part of the CSD sequence shown. HP1B from *D. melanogaster* and HP1-full from *N. vitripennis* are shown at the top as references. Amino acid sites that are altered from the conserved sequence are highlighted in red.
References


<table>
<thead>
<tr>
<th>Entity</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dmel HP1B</td>
<td>GFERGIASGLGATDSSGHLMPFAMKWDGDHADLVPARLANTRCQPOVVTQFYEEERLTW</td>
</tr>
<tr>
<td>Nvlt HP1-full</td>
<td>GFDRNLDPERRIGATDSSGELMPLMKWGDGDADLVPARANEMCPQIVIRFYEEERLTW</td>
</tr>
<tr>
<td>Dmel Umbrea</td>
<td>GFDLGLEPLRLGLGNWLSGLTPMOKGQCEAYLVPAEVLRNQVMVSFFYEEERIVF</td>
</tr>
<tr>
<td>Nvlt HP1-csd</td>
<td>GFKRQDEADKIGATESGSLMFLMKWKGTEETLDVAEEANVLEQVYVYQFYEDRLKV</td>
</tr>
<tr>
<td>Csol HP1-csd</td>
<td>GFKRQDEADKIGATESGSLMFLMKWKGTEETLDVAEEANVLEQVYVYQFYEDRLNV</td>
</tr>
<tr>
<td>Lhum HP1-csd</td>
<td>GFKRELEAEIRILGADDMLMLVQMOKGCTMVMLTVEAKENKCPDPIVFYEDRLAV</td>
</tr>
<tr>
<td>Cinv Hri-csd</td>
<td>GFRNCLEADDKIGADDMLMLVQGFKCSDAVLVCAKEAKAKQQIVIRFYEDRLTV</td>
</tr>
<tr>
<td>Aros HP1-csd</td>
<td>GFKRALGPDKIIGATDSSGELMFLMKWKGTEETLDVAEEANVLEQVYVYQFYEEERLTW</td>
</tr>
<tr>
<td>Gabi HP1-csd</td>
<td>GFKRALQPDKIIGATDSSGELMFLMKWKGTEETLDVAEEANVLEQVYVYQFYEEERLTW</td>
</tr>
</tbody>
</table>

203x60mm (300 x 300 DPI)