Insect Molecular Biology (2001) **10**(5), 513–521

A molecular study of Abdominal-A in the ant *Myrmica rubra* reveals lineage dependant evolutionary rates for a developmental gene

H. Niculita, J. F. Julien and E. Petrochilo

Centre de Génétique Moléculaire, Laboratoire Evolution et Contraintes génomiques, CNRS, 1 avenue de la Terrasse, 91198 Gif sur Yvette, France

Abstract

We have characterized the *abdominal-A* locus in a Hymenopteran, the ant *Myrmica rubra*. The sequence of the homeotic domain of the Abdominal-A family of orthologous genes is known for a fairly large number of insects but the complete sequence of the Abdominal-A protein, is known only for a few. The two proteins of *Drosophila melanogaster* and *Tribolium castaneum* differ markedly outside the homeodomain. A comparison of the ant Abdominal-A protein sequences with those of these two insects shows that the ant and beetle sequences are very similar all along the length of the protein. The fruit fly has diverged considerably and equally from the other two insects. This divergence reflects different rates of evolution of the protein in different lineages.

Keywords: Hox genes, Formicidae (Hymenoptera), Abdominal-A, molecular evolution.

Introduction

The homeotic complex has been shown to play a capital role in the development of all well-studied model organisms. Initially it was thought of mainly as the determinant of regional identity along the anterior posterior body axis. This function has been confirmed in most cases, but it is becoming increasingly evident that many genes of the complex are not restricted to this purely informational role and that they are also involved in bringing about details of body morphology. Recent studies have shown that minute differences in the levels of expression of a Hox gene are directly responsible for discrete differences in appendage morphology (Stern, 1998). And, at the level of protein activity it has been shown that different regions of a Hox protein, outside of the homeodomain, are required for fulfilling its functions in different germ layers (Chauvet *et al*., 2000).

Because of their broad functional attributes, the homeotic genes have been repeatedly assigned key roles in the evolution of large-scale morphological differences, and gross alterations of the Hox complex have been proposed to explain various spectacular evolutionary events (Akam *et al*., 1994; Carroll, 1995; Gellon & McGinnis, 1998). The more recent results serve to emphasize that attention must also be paid to more discrete variations of expression and structure of the Hox proteins in evaluating their potential role in evolution.

For our part, we have been attracted to the case of the Formicidae (Insecta, Hymenoptera). In most insect orders the anterior abdominal segments are mostly monotonous repeats, distinguishable from each other only by experts and devoid of such features as wings and appendages that characterize the thoracic and the most posterior abdominal segments. This is not the case in the Hymenoptera and more particularly in the Formicidae (the ants). To begin with, in all apocrite hymenopterans, the posterior extremity of the first abdominal segment is pinched off, separating the mesosome from the gaster and functionally fusing the first abdominal segment to the thorax. All ants carry this feature a step further, as the extremity of the second abdominal segment also becomes pinched off, isolating the second abdominal segment in a tegmosal no man's land between the mesosome and the gaster, where it constitutes the petiole. Some subfamilies such as the Myrmicinae and the Pseudomyrmicinae go to further extremes, and the extremity of the third abdominal segment becomes pinched off in turn, bringing about a feature known as the post-petiole (Baroni Urbani, 1988).

It is difficult, in the presence of this feature, not to imagine that in these groups, the third abdominal segment has taken on the identity of the second, and that one is facing here a natural homeotic mutation, although alternative

Received 5 January 2001; accepted after revision 1 June 2001. Correspondence: Hélène Niculita, Centre de Génétique Moléculaire, Laboratoire Evolution et Contraintes génomiques, CNRS, 1 avenue de la Terrasse, 91198 Gif sur Yvette, France. Tel.: +33 1 69 823 208; fax: +33 1 69 824 386; e-mail: niculita@cgm.cnrs-gif.fr

explanations may spring to mind when our description of the phenotype is considered. All of these remarkable morphological events occur within the developmental realm of the homeotic gene *abdominal-A*, and mutants of this gene in both *Drosophila melanogaster* (Sanchez-Herrero, 1985) and *Tribolium castaneum* (Stuart *et al*., 1993) have been obtained that transform, very specifically in the case of *Tribolium*, the third abdominal segment into a phenocopy of the second. We became convinced that it might be worthwhile to take a molecular look at the *abdominal-A* locus in ants in order to examine whether one feature or another of this locus could be correlated with the morphological diversity exhibited by this taxon.

Keeping in mind the fact that, as emphasized above, modest differences in expression and/or structure of the protein might suffice to bring about visible phenotypic changes we have attempted a description as complete as possible of the *abdominal-A* locus in the Myrmicinae ant *Myrmica rubra* (*Mrabd-A*). We report here the complete coding sequence and the principal features of the *abdominal-A* locus.

Comparisons of this coding sequence with those of *Drosophila melanogaster* and *Tribolium castaneum*, underline the derived character of the fly's homologue and shed some light on the organization of the protein and its evolution.

Results

Isolating a complete cDNA for Myrmica rubra abdominal-A

Starting with the sequence of the honeybee *abdominal-A* homeobox (Walldorf *et al*., 1989), we synthesized a pair of non-degenerate primers capable of amplifying a 204 bp fragment. The sequence of this segment amplified from *Myrmica rubra* genomic DNA is, as expected, extremely similar to those known from other insects (G. Balavoine, personal communication).

With the help of this sequence, ant-specific primers were designed and used in a general PCR screen of a *Myrmica ruginodis* genomic library, kindly provided by Y. Bigot (Bigot *et al*., 1994). (*Myrmica ruginodis* is a sibling species of *Myrmica rubra*.) A single genomic clone was isolated in this screen. It contained the 3′ end of an intron, and a large part of the terminal exon including the entire C-terminal coding sequence and a part of the 3′ transcribed non-coding sequence. The sequence of this clone, in turn, was used to generate a set of ant *abdominal-A* specific primers with which to search for the extremities of the coding sequence by 5′ and 3′ RACE. The 5′ RACE experiment, on *Myrmica rubra* larval cDNAs, produced repeatedly a single 1.5 kb fragment that was isolated, subcloned and sequenced.

Sequence analysis of this fragment turned up a single open reading frame, beginning at position 352 and ending at position 1506. The putative protein is 394 amino acids long and contains an Abdominal-A type homeodomain, whose sequence is identical to that already determined in this species by the previously described experiments (Fig. 1).

```
MSTKFIIDSMLPAKYQQFHPQQLFSSATPGTIQCTTSPATASLESSLSAA

AVAAAAVNYAHQHNSPSPTGSSPQHSGSSASTSPAARTTSSMYPYVSAAA

AHHHHQQQQAVAAAAFGATSSMVPGFGSTAASSAALAAAAAVDATTAGDK

SCRYTASLTGNVAPASADPMVNYTLGHHHQNGATPGSLVSSASASSAVSA

ASASMAAAAQFYHQAAAASAVVGPLTSCQQPTTGQPGISDIPRYPWMSIT

DWMSPFDRVVCGPNGCPRRRGRQTYTRFQTLELEKEFHFNHYLTRRRRIE

IAHALCLTERQIKIWFQNRRMKLKKELRAVKEINEQARREREEQDMMKKQ

QAEKQAKLQQEQQSAALQHQQQHHVSGLEKTQSDLLKAVSKVPT*
```

Figure 1. The complete coding sequence of the MrAbd-A based on embryonic cDNA from *Myrmica rubra*. The homeodomain *sensu stricto* is underlined, the intron positions are indicated with arrowheads and the stop codon is indicated with an asterisk. The first methionine was putatively identified by analysing the cDNA sequence from both *Myrmica rubra* and *Tribolium castaneum*.

Analysing the Myrmica rubra abdominal-A *transcript*

Owing to a shortage of biological material, an analysis of the *Mrabd-A* transcripts by classical Northern blotting was not feasible. We therefore resorted to Reverse Transcription analysis of poly $(A)^+$ RNAs extracted from embryos and various larval stages, and 5′ and 3′ RACE experiments.

The 3′ RACE experiments failed to amplify an *abdominal-A* fragment, while the 5′ RACE experiments amplified the same fragment as described above in every extract. These results favour an extended expression of the *Mrabd-A* gene from the embryo into the early nymphal stage. The fact that we were unable to amplify more than a single *Mrabd-A* transcript, places constraints on the possible variations of the *abdominal-A* mRNA. In particular we consider it unlikely that a form of messenger initiating translation at the same position as that described for *Drosophila melanogaster*, exists in *Myrmica rubra*. Our results suggest the existence of a single form of mRNA for the *abdominal-A* locus in *Myrmica rubra*, as seen in *Tribolium castaneum*, a possibility also supported by a single Northern blotting experiment (data not shown).

Localizing the transcriptional start site

The sequence data obtained from the *Mrabd-A* cDNA were used to generate primers for Genomic Walking experiments. In this manner, 3.6 kb of upstream sequence and 1.3 kb of downstream sequence were obtained. With these sequences at hand, it is simple to verify the transcriptional start site, by comparative PCR experiments on embryonic cDNA and genomic DNA. These experiments (Fig. 2) place the transcriptional start site between –1330 and –1190 upstream of the open reading frame. Interestingly, this is considerably further upstream of the most upstream position identified

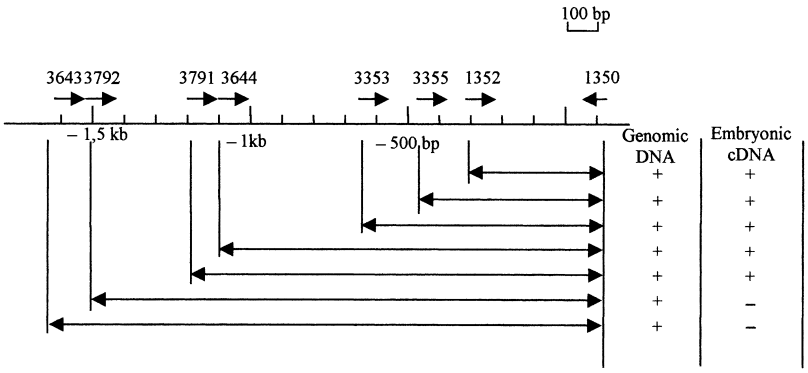

Figure 2. Localizing the transcriptional start site. Comparative PCR experiments on embryonic cDNA and genomic DNA of *Myrmica rubra* gave us the position of the transcriptionnal start site. These PCR experiments were carried out using different upstream primers (which numbers are indicated in the top of the ladder) and a single downstream primer: 1350. The 5′ region of *abd-A* locus is represented by a ladder with a unity of 100 pb. The (0) indicates the beginning of the coding region, the arrows () indicate the length of PCR product, a (+) signifies that the amplification gave the expected fragment and a (–) signifies that no amplification was observed under optimal conditions.

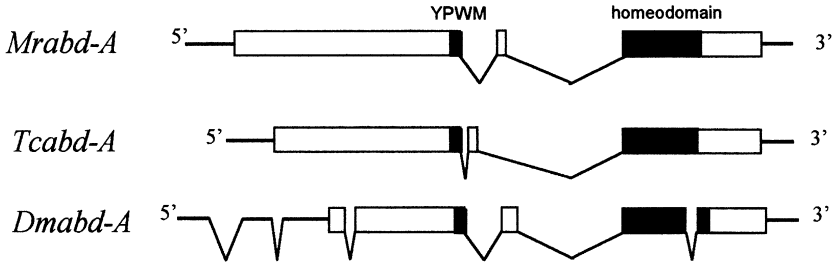

Figure 3. Schematic structure of the *abd-A* locus in different insects: *Myrmica rubra*, *Tribolium castaneum*, *Drosophila melanogaster*. The coding sequences are shown as bars, the introns as thin black lines, the 5′UTR and the 3′UTR as thick black lines. The number and position of introns are the same in the ant and in the beetle. Additional introns are present in the fly. The hexapeptide (YPWM) and homeodomain positions are indicated filled bars. *Mrabd-A*, *Myrmica rubra abdominal-A* locus; *Tcabd-A*, *Tribolium castaneum abdominal-A* locus; *Dmabd-A*, *Drosophila melanogaster abdominal-A* locus.

by the 5′RACE experiments (–316). The large untranslated 5′ leader seen for this mRNA is in keeping with those generally observed in *Drosophila*, suggesting that this is a widespread attribute of mRNAs in insects.

Determining the position and number of introns

Comparative PCR experiments on genomic and embryonic cDNA identified an intron at a position homologous to the first intron of *Tribolium castaneum*. This is also the first intron in *Myrmica* and it has been cloned and partly sequenced, its size is approximately 3.5 kb. The sequence of the *Myrmica ruginodis* clone revealed that there is an intron–exon junction immediately upstream of the homeobox. The position of this 3′ splice site had already been identified in a previous attempt at genomic walking in *Myrmica rubra*, it is in the same position as that of the largest intron in the *abd-A* gene of *Drosophila melanogaster* and of other insects and appears to constitute a plesiomorphic trait of the locus. By genomic walking in 3′ from the first intron it was possible to identify a 5′ splice site corresponding to that of the second intron of *Tribolium* and to the 5′ splice site of the largest intron of *Drosophila*. It has not proved feasible, for the time being, to clone this intron, presumably because its too large a size. These two introns are the only ones present in the *abdominal-A* gene of *Myrmica rubra*. They are positioned at nucleotide 750 and 785 of the coding sequence; that is to say, between amino acid 250 and 251 and between the second and third position of the codon for amino acid 262, respectively (Fig. 1). All of these positions correspond exactly to those of the two introns found in the *Abdominal* gene of *Tribolium castaneum* (Shippy *et al.*, 1998).

Comparing the amino acid sequence of the putative MrAbdA protein with that of Drosophila melanogaster *and* Tribolium castaneum

A complete and annotated comparison of the Abdominal-A protein is available for only two other insects: *Drosophila melanogaster* (Karch *et al.*, 1990) and *Tribolium castaneum* (Shippy *et al.*, 1998). These two sequences have been shown to differ considerably both in the organization of the gene and in the sequence of the protein. The main conclusion of our work is that, on both accounts, the ant *abdominal-A* locus is more similar to that of the beetle than to that of the fly.

A schematic representation of the locus shown in Fig. 3 illustrates the fact that its organization is very much the same in the ant and in the beetle. In particular, the number and positions of introns are the same in these two organisms. In contrast the fly locus is more extended and contains three additional introns at the 5′ end and an intron in the homeodomain, which is absent in the other two insects. The origin of transcription of the ant gene appears to lie in the first intron of the *Drosophila* transcript, as does the origin of transcription of the Tribolium gene (Shippy *et al.*, 1998).

The protein sequences were aligned using the CLUSTALW program (Thompson *et al.*, 1994) and the alignments were hand edited. The result is shown in Fig. 4. For the purpose of this comparison the protein is split up into the following four domains:

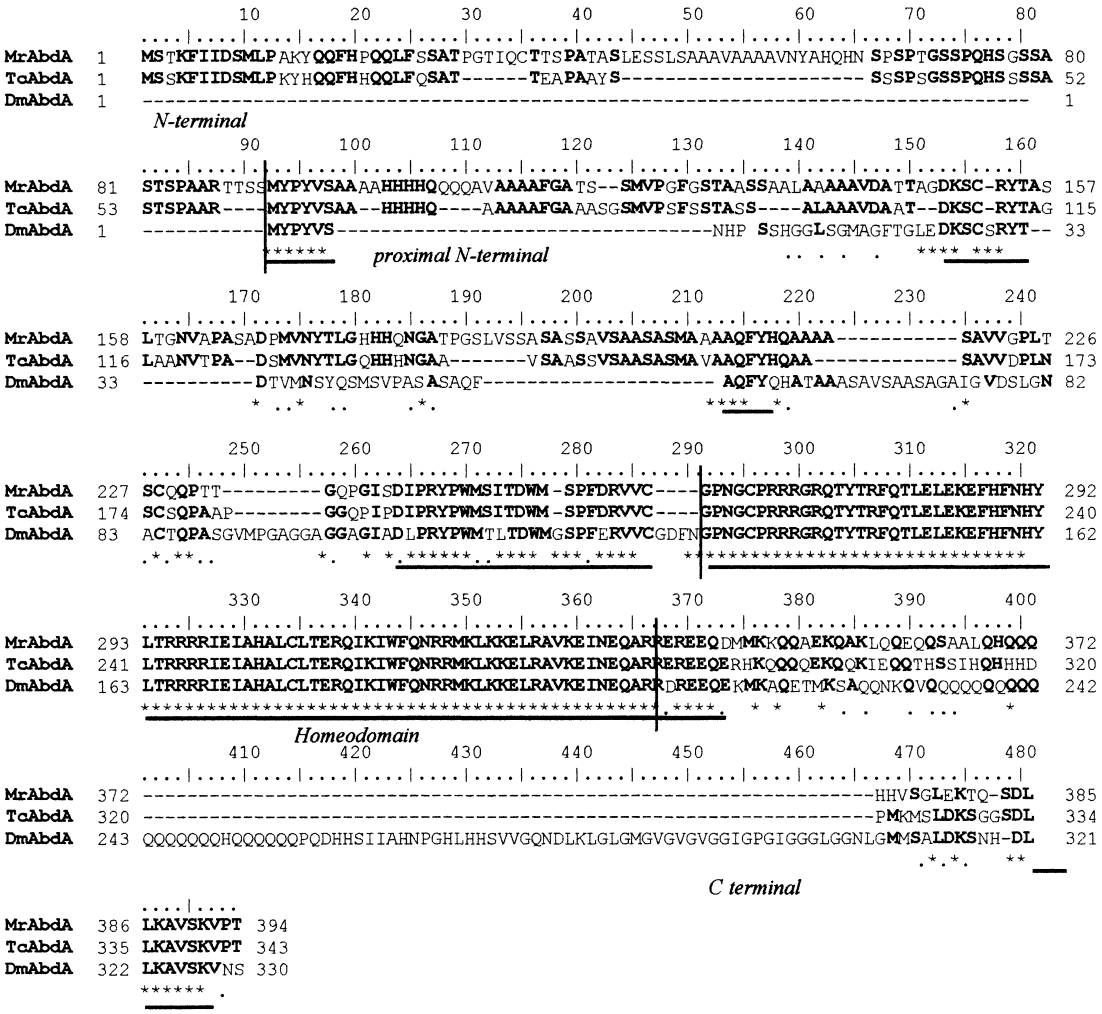

Figure 4. Sequence comparisons between the Abdominal-A proteins of *Myrmica rubra*, *Tribolium castaneum* and *Drosophila melanogaster*. These sequences were aligned using CLUSTALW (Thompson *et al.*, 1994) and the BLOSUM matrix, a gap opening penality of 10 and a gap extention penality of 0.1. Identities are indicated by an asterisk (*) and similarities are indicated by a dot (.). Each dash (–) indicates a gap of 1Aa. Bold characters indicate that the residue at this position is common between at least two sequences. Black bars below the sequences indicate conserved regions. Vertical bars delimit the four domains mentioned in the text. TcAbdA, *Tribolium castaneum* Abdominal-A protein; MrAbdA, *Myrmica rubra* Abdominal-A protein; DmAbdA, *Drosophila melanogaster* Abdominal-A protein.

1 an N-terminal region specific to the beetle and ant proteins extending from the initial Methionine to the Methionine that initiates translation in *Drosophila melanogaster* and which is part of the conserved motive MYPYV;
2 a proximal N-terminal region extending from the MYPYV motive to the homeodomain;
3 what we have called the homeodomain in fact covers all the completely invariant seventy-five residues between the sequences GPNGC ... and ... INEQA. This includes five extra residues on the N-terminal side and ten extra residues on the C-terminal side of the homeodomain *sensu stricto*;
4 the C-terminal region which includes the remaining part of the protein.

The percentage of identity is calculated for each of these regions (Table 1).

The first N-terminal domain is apparently absent from the known fly protein (see discussion) and represents the N-terminal end of the putative beetle and ant Abdominal-A proteins. The two proteins have maintained a strong identity (72%) in this region, indicating that it has some functional importance.

The proximal N-terminal part of the protein, is strikingly more conserved between the ant and beetle: 83% compared to 48% and 53%, respectively, in comparisons with the fly. In fact, in the latter comparisons, the N-terminal region is not significantly more conserved than the C-terminal, and an important part of the remaining identity results from the hexapeptide and its immediate environment.

In effect, the entire N-terminal end of the protein, right up to the hexapeptide, behaves as a unit that has diverged

Table 1. Percentage of identity for pairwise comparisons between the Abdominal-A proteins of ant, beetle and fly. The Abdominal-A proteins are split up into four domains (see table footnote). The columns represent the results obtained for each of these domains. The lines designate the corresponding pair wise comparisons. The conserved peptide motives common to all three sequences are shown in the bottom of the table in the order of their appearance along the sequence

		Abd-A protein domains			
	Full Abd-A protein	N-terminal	Proximal N-terminal	Homeodomain	C-terminal
MrAbdA/TcAbdA	80.2%	76.0%	82.8%	100%	54.6%
MrAbdA/DmAbdA	64.7%		47.9%	100%	48.0%
TcAbdA/DmAbdA	66.3%		51.4%	100%	43.80%
Conserved motifs between insects			MYPYVS, DKSCxRYT, AQFY, DxPRYPWMxxTDWMxSPFxRVVC	all	RRxREEQ, DLLKAVSKV

N-terminal = the N-terminal region specific to the beetle and ant proteins extending from the initial Methionine to the Methionine that initiates translation in the fly which is part of the conserved motive MYPYVS.
Proximal N-terminal region = the N-terminal region extending from the MYPYVS motive to the homeodomain.
Homeodomain = seventy-five residues that include the sixty residues of the homeodomain *sensu stricto*, five residues upstream and ten downstream.
C-terminal = the C-terminal region which includes the remaining part of the protein.

significantly in the fly while retaining a strong identity in the ant and beetle.

The C-terminal end of the protein is the less conserved with only 49% identity on average in the pairwise comparisons. The important point is that the three proteins appear to have diverged more or less equally from each other in this region, although the ant and beetle are slightly more identical than either is to the fly.

Despite the great similarity between the ant and beetle proteins there are a number of interesting differences between them (Fig. 4). The ant protein is fifty-one amino acids longer than the beetle's. Most of this difference in length lies in the N-terminal region of the protein which is a part from the homeodomain, the region of highest identity. A computer produced and hand edited alignment of the two protein sequences ends up with no less than fourteen gaps or insertions (indels). Thirteen gap/insertions are located in the N-terminal and only one in the C-terminal domain. There is a tendency for the material comprising the shorter insertions to consist of Alanine and Glutamine repeats, with no obvious consequence on the hydrophobicity profile of the protein. The significance of the clustering of indels in the N-terminal part of the protein, together with the fact that they appear to be mostly insertions in the ant gene (or deletions in the beetle), is unclear but underlines the differences in the evolution of the N-terminal and the C-terminal parts of the protein.

Finally, this overall analysis brings to light a number of strikingly conserved peptide motives, common to all three proteins. Their sequences are given in the bottom line of Table 1. Some of these conserved blocks harbour signals for post-translational modifications, such as phosphorylation or glycosylation, but this is not the case in others. For the latter, it has not been possible to ascribe a function to any of them although their conservation over a span of 200 million years or more definitely singles them out as significant and obvious targets for future functional studies.

Discussion

This work describes the complete Abdominal-A coding sequence and the main features of the corresponding locus in the ant *Myrmica rubra*.

A comparison of the putative Abdominal-A protein sequences with those of the beetle *Tribolium castaneum* and the fly *Drosophila melanogaster* shows that the ant and beetle sequences are very similar, retaining a high identity all along the sequence, while the fly has diverged considerably and equally from the other two insects not only in the sequence of the protein, but also in the organization of the locus, including the addition of four introns, and the use of a more posterior initiation codon, presumably in relation with the use of an entirely different transcriptional start.

The existence of the longer protein in *Drosophila melanogaster* has been proposed (Martin *et al*., 1995) but not substantiated, while the presence of the shorter form is well documented and mutants are known that can only affect the latter (Karch *et al*., 1990; Macias *et al*., 1990). Moreover, another complete sequence of an insect Abdominal A orthologue, that of *Anopheles gambiae*, is present in the GenBank database. The proposed coding sequence for this orthologue begins with the same 'MYPYV' motive as the identified protein of *Drosophila melanogaster*, suggesting that the use of this translational initiation site is a synapomorphy of the Dipterae. Therefore, our working hypothesis is that the shorter form of the protein is the functional entity in *Drosophila melanogaster*.

In order to analyse the relationships between the sequences of the Abdominal-A protein in insects, we chose the spider *Cupiennus salei* (Damen *et al*., 1998) as an outgroup. The multiple alignments were produced by the CLUSTALW program and hand edited to minimize the number of gaps.

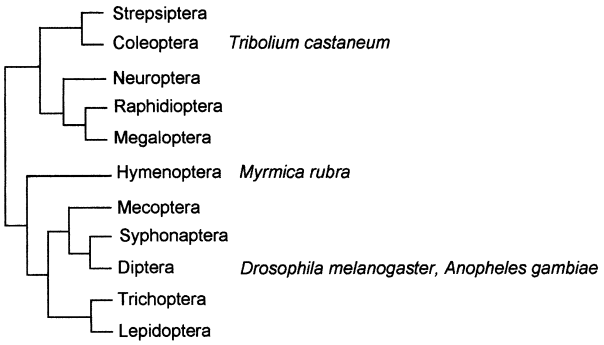

Figure 5. The phylogenetic relationships between the Holometabolous insects (according to Brusca & Brusca, 1990). The species mentioned in the present paper are shown to the right of the corresponding Order.

The consensus concerning the evolutionary relationships between the holometabolous insects is depicted in Fig. 5 (Brusca & Brusca, 1990). The Coleopterans are believed to have emerged first, approximately 250 million years ago, followed in relatively quick succession by the Hymenopterans and the Dipterans. The latter were already present 200 million years ago, when it is believed that the lineage leading to *Drosophila* was already distinct from that leading to *Anopheles*. In the most comprehensive study of Insect molecular phylogeny to date, Whiting *et al.* (1997) also favour this relationship. Furthermore, they are careful to point out that although the relative position of the Coleoptera and Hymenoptera is still questionable, the latter emergence and monophyletic nature of the Diptera is strongly supported.

We have used Tajima's relative rate test (Tajima, 1993) in the reverse mode between the three taxa under study to find out which one performs best as an outgroup to the two others, under the molecular clock hypothesis. The test was performed on both nucleic acid and amino acid sequences and for the entire protein, the N-terminal domain and the C-terminal domain. The results are shown in Table 2. It is quite clear that according to this test the only satisfactory outgroup would be *Drosophila*. The idea that the lineage leading to *Drosophila* stems from the base of the holometabolous insects contradicts all available evidence. On the other hand, this result is also compatible with a specific increase in the rate of evolution of the *Drosophila* lineage, leading to an increase in the number of differences between this lineage and all others. Interestingly, it is the N-terminal domain that is mainly responsible for the behaviour of the *Drosophila* sequence, because it has accumulated no less than thirty-three specific differences as opposed to four in *Myrmica* and two in *Tribolium*, while the C-terminal domain has ten specific sites as opposed to six and five for *Tribolium* and *Myrmica*, respectively. Considering that a significant part of this domain has been lost entirely in *Drosophila* and that this same part shows a high degree of conservation between *Myrmica* and *Tribolium* one might be tempted to believe that some function of Abdominal-A has been forfeited altogether in *Drosophila.* Alternately, the function of the Abdominal-A protein might have changed in this lineage, in which case the sequence divergence would reflect this change and not a simple relief from selective pressure.

Two recent publications have confirmed the sequence of the *Abdominal-A* gene of *Anopheles gambiae* (Devenport *et al.*, 2000; Powers *et al.*, 2000) that was present in the

Table 2. Results of Tajima's relative rate test as applied to the three possible permutations of *Myrmica*, *Tribolium* and *Drosophila* AbdominalA coding sequences. Combinations followed by [A.A.] concern amino acid sequences. Each line represents a test of the taxons listed in the first column. The last taxon sited is taken as the outgroup. The first six columns describe the data on which the test is run. Divergent sites are sites that contain different residue in each sequence. The last two columns give the result of the test in the form of a Chi squared value for one df and the pertaining probability

Taxons A, B and C Taxons are rooted on C	No. sites	Site category					χ^2(1 df)	Probability
		Identical	Divergent	A specific	B specific	C specific		
Dmabd-A, Mradb-A, Tcadb-A	681	365	74	104	78	60	3.71	0.054
DmAbd-A, MrAbd-A, TcAbd-A [A.A.]	226	142	24	43	9	8	22.23	< 0.001
Dmabd-A, Mradb-A, Tcadb-A N-terminal	282	136	34	59	34	19	6.72	0.01
DmAbd-A, MrAbd-A, TcAbd-A N-terminal [A.A.]	94	45	10	33	4	2	22.73	< 0.001
Dmabd-A, Mradb-A, Tcadb-A C-terminal	168	75	29	25	18	21	1.14	0.286
DmAbd-A, MrAbd-A, TcAbd-A C-terminal [A.A.]	55	20	14	10	5	6	1.67	0.197
Tcabd-A, Dmabd-A, Mrabd-A	681	365	74	60	104	78	11.8	0.001
TcAbd-A, DmAbd-A, MrAbd-A [A.A.]	226	142	24	8	43	9	24.02	< 0.001
Tcabd-A, Dmabd-A, Mrabd-A N-terminal	282	136	34	19	59	34	20.51	< 0.001
TcAbd-A, DmAbd-A, MrAbd-A N-terminal [A.A.]	94	45	10	2	33	4	27.46	< 0.001
Tcabd-A, Dmabd-A, Mrabd-A C-terminal	168	75	29	21	25	18	0.35	0.555
TcAbd-A, DmAbd-A, MrAbd-A C-terminal [A.A.]	55	20	14	6	10	5	1	0.317
Mrabd-A, Tcabd-A, Dmabd-A	681	365	74	60	78	104	2.35	0.125
MrAbd-A, TcAbd-A, DmAbd-A [A.A.]	226	142	24	8	9	43	0.06	0.808
Mrabd-A, Tcabd-A, Dmabd-A N-terminal	282	136	34	19	34	59	4.25	0.039
MrAbd-A, TcAbd-A, DmAbd-A N-terminal [A.A.]	94	45	10	2	4	33	0.67	0.414
Mrabd-A, Tcabd-A, Dmabd-A C-terminal	168	75	29	21	18	25	0.23	0.631
MrAbd-A, TcAbd-A, DmAbd-A C-terminal [A.A.]	55	20	14	6	5	10	0.09	0.763

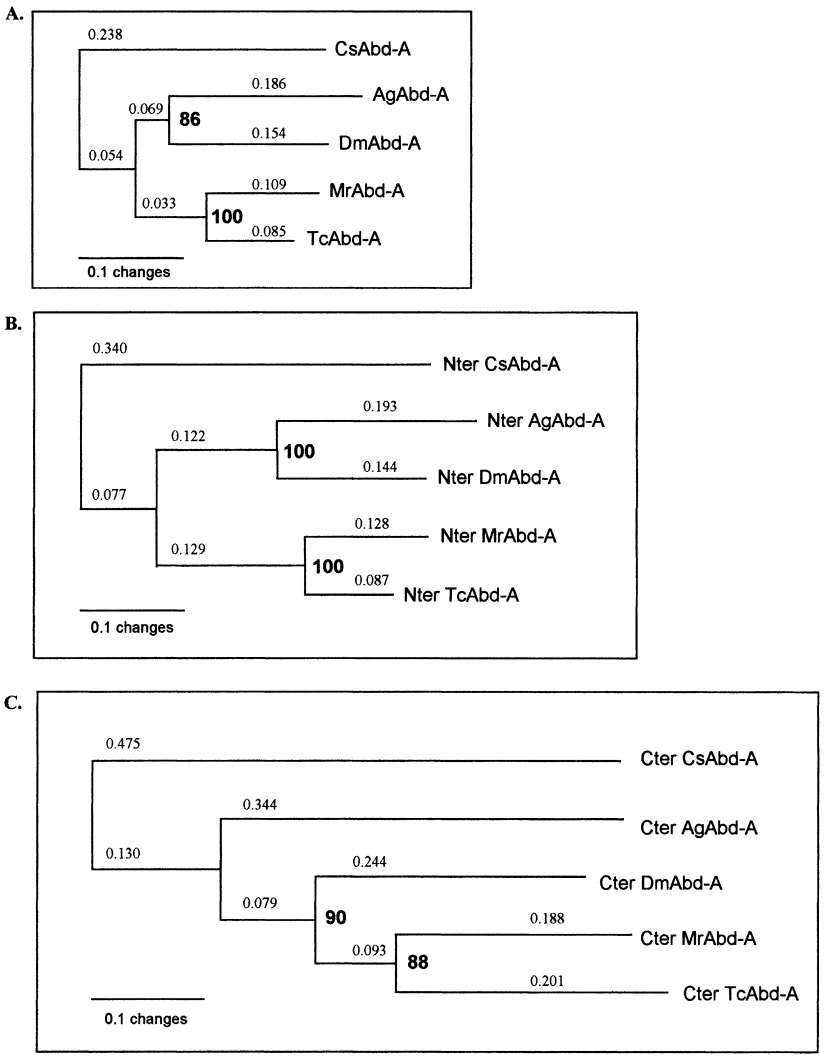

Figure 6. These phylograms concern the entire AbdA proteins (A) and the two variable parts of the Abdominal-A protein: the entire N-terminal region (B) and the C-terminal region (C) of *Cupiennius salei* (accession number AJ007436), *Tribolium castaneum* (accession number AF017415), *Myrmica rubra, Anopheles gambiae* (accession number AF080566) and *Drosophila melanogaster* (accession number X54453). The multiple alignements were produced by the CLUSTALW program and hand edited to minimize the number of gaps. The phylogenetic trees were constructed by the Neighbour Joining method using the observed distances and the pair wise gap removal option. Nter, N-terminal region of each protein; Cter, C-terminal region of each corresponding protein. The bootstrap values are in bold digits and the branch lengths in small digits.

GenBank database. As mentioned previously, the N-terminal part of the *Anopheles* protein shows the same modifications as that of *Drosophila*. If these modifications were correlated with a loss of function of the Abdominal-A protein, one might expect a considerable amount of divergence between the two Dipterans. On the other hand, if the function of the protein had been altered and if the nature of the alteration were common to the Dipterans, one would expect a clear relationship between these sequences to emerge. The phylograms that include *Anopheles gambiae* are presented in Fig. 6. It is clear that the Dipteran sequences cluster unambiguously, yet they do show more divergence between themselves than do the ant and beetle. Could it be the case that the retained identity that allows the clustering of these sequences simply has not had the time to be dispersed? Two aspects of the data refute this explanation.

1 On the one hand, the clustering of the Dipteran sequences is increased by using only the N-terminal domains to draw the phylogram. This should not be the case if this domain had simply lost its function. Furthermore, the observed clustering cannot be the result of long branch attraction on the N-terminal tree because the length of the branches are shorter than on the C-terminal tree.

2 On the other hand, the Dipteran sequences do not cluster but instead emerge independently. This domain serves as an internal control that loss of clustering might well have occurred, had it not been for some selective pressure that has maintained the identity of the Dipteran proteins mainly in the N-terminal region.

These results confirm that the N-terminal domain of the Dipteran protein has retained significant functional importance, as established by the work of Chauvet *et al.* (Chauvet *et al.*, 2000). Nevertheless, the Dipteran sequences have diverged considerably more from the ant and beetle sequences than would be predicted on the basis of the accepted phylogeny. Obviously, one might question the systematic position of the Dipterans, but we feel the evidence from all fields (palaeontological, physiological and developmental) to be overwhelmingly in favour of the accepted relationships among these groups. The alternative

explanation for the observed discrepancy is that somewhere between the emergence of the Hymenopterans and that of the Dipterans, the function of the *abdominal-A* gene has changed, either because it has acquired a new function or because it has relinquished in part some of the functions of the ancestral form. The case for the latter hypothesis lies with the fact that the divergence between the two Dipterans is significantly more important than it is between the ant and the beetle, suggesting that, in its more recent role, the protein is less constrained than it is in its ancestral functions. This point is also supported by preliminary results we have obtained on part of the N-terminal sequence (from residue 81 to residue 213) in eight other ant species, covering a large range of ant subfamilies (Niculita *et al.*, in preparation). All ant sequences are essentially identical at the protein level (unpublished results). Of the 132 residues examined, only 10 (7.5%) show a variation, and, in pairwise comparisons, the amount of divergence never exceeds 3%, despite the fact that these species have diverged over a timespan of more than 100 million years. This very strong conservation in ants and the limited divergence observed between *Myrmica rubra* and *Tribolium castaneum*, emphasize that the rate of evolution of the N-terminal domain of the protein is lineage dependant, while no such effect is detected on the evolution of the C-terminal domain.

The strong conservation of sequences outside the homeodomain over a timespan of more than 200 million years was unexpected. Numerous reports (Bachiller *et al.*, 1994; Zhao *et al.*, 1993) of partial, yet spectacular, complementations between vertebrate homeotic genes and their fly orthologues, in cases when outside of the homeodomain there was little sequence identity, have created the impression that the variable parts of these proteins were not essential for their function, and, in our opinion at least, the less numerous reports (Chauvet *et al.*, 2000; Grenier & Carroll, 2000; Peltenburg & Murre, 1996) that point to the functional importance of the extra-homeotic domains have not yet turned the tides. Be that as it may, our data point at once to the functional importance of the N-terminal region of the Abdominal-A protein and to an evolutionary shift of this function between the Dipterans and the Hymenopterans. Obviously future efforts will have to concentrate on obtaining experimental evidence for this prediction.

Experimental procedures

Ant care and handling

The colonies of *Myrmica rubra* were a kind gift from Alain Lenoir (Tours University). They were maintained in the laboratory at 24 °C and under a 12 h light/12 h dark photoperiod. The colonies are harboured in plastic boxes filled with plaster of Paris moulded to define connected chambers. This nest is connected to a foraging area, from which the ants are prevented from escaping by the use of Fluon. They are regularly provided with water and fed with live *Tenebrio* larvae and sugar. Embryogenesis lasts about one week. The embryos were collected directly from the nest after the colonies had been anaesthesized by CO_2 treatment.

Screening phage libraries by PCR

The library is divided in ninety-four subsets whose complexity is approximately 1/30 that of the complete bank. Aliquots of phage lysates obtained under these conditions of infection, in microplate cultures (200 µl each) are tested by PCR for the amplification of the *abdominal-A* homeodomain fragment. The positive subsets are selected and a treated in the same way. The amplification factor per round is about 30 so that after four rounds the complexity of the positive sub banks is reduced to the point were it is feasible to isolate the desired clone by testing individual phage plaques.

Isolation of MrabdA cDNA

Total RNA was extracted from fresh embryos aged from a few hours to a week, according to the 'SV total RNA Isolation System Kit' protocol (Promega). Poly (A) + RNA was isolated with the help of the Dynabeads mRNA purification Kit (Dynal). Double Stranded cDNA synthesis is performed, and the products are ligated to the proprietary adaptors provided by the Marathon cDNA amplification kit (Clontech). 5′ RACE is carried out using the following *Mrabd-A* specific primer located near the 3′ end of the coding sequence: 5′-AGATCGCTTTGCGTTTTTTCCAGGCCGC-3′, and the primers provided by the kit. Amplification of the cDNA requires 1 M GC melt and the following reaction conditions: 94 °C, 1 min; then 94 °C, 25 s; 68 °C, 3 min for 40 cycles; finally 68 °C, 3 min. The GenBank accession number of *Mrabd-A* cDNA is AF332515.

Isolation of abdominal-A genomic regions

A genomic bank of *Myrmica ruginodis* (a sibling species of *Myrmica rubra*) was screened by PCR, as described above, with a *Mrabd-A* specific set of primers. A single clone was obtained that contained the 3′ end of the gene. Genomic DNA was extracted from freshly killed adult workers with the help of a QIAamp Tissue Kit (Qiagen). Amplification of specific *Mrabd-A* genomic fragments was performed either straightforwardly as in the case of the first intron with primers based on the cDNA sequence, or with the help of Clontech's Genome Walker Kit. Sequences isolated in this manner include 3.8 kb of material 5′ to the translational start (accession number AF332516) and 1.3 kb of material 3′ to the translational stop (accession number AF332517).

Localization of the transcriptional start

Comparative PCR experiments on embryonic cDNA and genomic DNA were carried out using the following upstream primers:
3643: 5′-ATTCGAAAAAGTGCTGCCTTGACGAGT-3′
3792: 5′-GAGGAGGACAGAGGCAGAATAAGAACGAG-3′
3791: 5′-CCTAAAACAGGACAGAAAATACCGTGTGAG-3′
3644: 5′-GGGTGTAAGGCAACGTGGCGAGCA-3′
3353: 5′-AACACCTGACAGGGGCCACGCTCTT-3′
3355: 5′-GCGCAGGATTCAAATTCAAACGCGA-3′
1350: 5′-GATAATGTGCTGAGTTGCGTAAGTGGC-3′
and the single downstream primer:
1352: 5′-CCGGACTCGTCGTGCACTGAATGGTAC-3′

Advantage 2 Polymerase Mix (Clontech, UK) was used in all these amplifications following Clontech instructions.

Cloning and sequencing

PCR products were cloned using the TOPO TA cloning system from Invitrogen. Clones were sequenced by the Applied Biosystems dye terminator method and run on an ABI automated sequencer. Sequences were analysed using the Wisconsin™ Package v.9.1, Genetics Computer Group (GCG), Madison, WI.

Sequence analysis

Alignments were produced using the CLUSTALW program and hand edited to minimize the number of gaps and improve the identities. Phylograms were produced with the Phylowin package (Galtier *et al.*, 1996) using the Neighbour Joining Algorithm (Saitou & Nei, 1987), an uncorrected observed distance matrix and the pair wise gap removal option. Tajima's relative rate test was performed with the help of the program package MEGA v.2.0 (Kumar *et al.*, 2001).

Acknowledgements

We thank Alain Lenoir for introducing us to ant biology and breeding and for providing us with specimens of *Myrmica rubra*. Work in our laboratory was supported by the Centre National de la Recherche Scientifique, Université Pierre et Marie Curie, and CNRS 'Programme Génome'. Financial support from the Ministère de la Recherche Scientifique to H.N. is gratefully acknowledged.

References

Akam, M., Averof, M., Castelli-Gair, J., Dawes, R., Falciani, F. and Ferrier, D. (1994) The evolving role of Hox genes in arthropods. *Dev Suppl*: 209–215.

Bachiller, D., Macias, A., Duboule, D. and Morata, G. (1994) Conservation of a functional hierarchy between mammalian and insect Hox/HOM genes. *Embo J* **13**: 1930–1941.

Baroni Urbani, C. (1988) Phylogeny and behavioural evolution in ants, with a discussion of the role of behaviour in evolutionary processes. *Ethol Ecol Evol* **1**: 137–168.

Bigot, Y., Hamelin, M.H., Capy, P. and Periquet, G. (1994) Mariner-like elements in hymenopteran species: Insertion site and distribution. *Proc Natl Acad Sci USA* **91**: 3408–3412.

Brusca, R.C. and Brusca, G.J. (1990) *Invertebrates.* Sinauer Associates, Sunderland, MA.

Carroll, S.B. (1995) Homeotic genes and the evolution of arthropods and chordates. *Nature* **376**: 479–485.

Chauvet, S., Merabet, S., Bilder, D., Scott, M.P., Pradel, J. and Graba, Y. (2000) Distinct hox protein sequences determine specificity in different tissues. *Proc Natl Acad Sci USA* **97**: 4064–4069.

Damen, W.G., Hausdorf, M., Seyfarth, E.A. and Tautz, D. (1998) A conserved mode of head segmentation in arthropods revealed by the expression pattern of Hox genes in a spider. *Proc Natl Acad Sci USA* **95**: 10665–10670.

Devenport, M.P., Blass, C. and Eggleston, P. (2000) Characterization of the Hox gene cluster in the malaria vector mosquito, Anopheles gambiae. *Evol Dev* **2**: 326–339.

Galtier, N., Gouy, M. and Gautier, C. (1996) Seaview and Phylowin, two graphic tools for sequence alignment and molecular phylogeny. *Comput Appl Biosci* **12**: 543–548.

Gellon, G. and McGinnis, W. (1998) Shaping animal body plans in development and evolution by modulation of Hox expression patterns. *Bioessays* **20**: 116–125.

Grenier, J.K. and Carroll, S.B. (2000) Functional evolution of the Ultrabithorax protein. *Proc Natl Acad Sci USA* **97**: 704–709.

Karch, F., Bender, W. and Weiffenbach, B. (1990) abdA expression in Drosophila embryos. *Genes Dev* **4**: 1573–1587.

Kumar, S., Tamura, K., Jakobsen, I.B. and Nei, M. (2001) *MEGA2: Molecular Evolutionary Genetics Analysis Software*. Arizona State University, Tempe, Arizona, USA.

Macias, A., Casanova, J. and Morata, G. (1990) Expression and regulation of the abd-A gene of Drosophila. *Development* **110**: 1197–1207.

Martin, C.H., Mayeda, C.A., Davis, C.A., Ericsson, C.L., Knafels, J.D., Mathog, D.R., Celniker, S.E., Lewis, E.B. and Palazzolo, M.J. (1995) Complete sequence of the bithorax complex of Drosophila. *Proc Natl Acad Sci USA* **92**: 8398–8402.

Peltenburg, L.T. and Murre, C. (1996) Engrailed and Hox homeodomain proteins contain a related Pbx interaction motif that recognizes a common structure present in Pbx. *Embo J* **15**: 3385–3393.

Powers, T.P., Hogan, J., Ke, Z., Dymbrowski, K., Wang, X., Collins, F.H. and Kaufman, T.C. (2000) Characterization of the Hox cluster from the mosquito *Anopheles gambiae* (Diptera: Culicidae). *Evol Dev* **2**: 311–325.

Saitou, N. and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**: 406–425.

Sanchez-Herrero, E., Vernos, I., Marco, R. and Morata, G. (1985) Genetic organisation of the Drosophila Bithorax complex. *Nature* **313**: 108–113.

Shippy, T.D., Brown, S.J. and Denell, R.E. (1998) Molecular characterization of the Tribolium abdominal-A ortholog and implications for the products of the Drosophila gene. *Dev Genes Evol* **207**: 446–452.

Stern, D.L. (1998) A role of Ultrabithorax in morphological differences between Drosophila species. *Nature* **396**: 463–466.

Stuart, J.J., Brown, S.J., Beeman, R.W. and Denell, R.E. (1993) The Tribolium homeotic gene Abdominal is homologous to abdominal-A of the Drosophila bithorax complex. *Development* **117**: 233–243.

Tajima, F. (1993) Simple methods for testing molecular clock hypothesis. *Genetics* **135**: 599–607.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl Acids Res* **22**: 4673–4680.

Walldorf, U., Richard, F. and Gehring, W.J. (1989) Comparison of homeobox-containing genes of the honeybee and Drosophila. *Proc Natl Acad Sci USA* **86**: 9971–9975.

Whiting, M.F., Carpenter, J.C., Wheeler, Q.D. and Wheeler, W.C. (1997) The Strepsiptera problem: phylogeny of the holometabolous insect orders inferred from 18S and 28S ribosomal DNA sequences and morphology. *Systematic Biol* **46**: 1–68.

Zhao, J.J., Lazzarini, R.A. and Pick, L. (1993) The mouse Hox-1.3 gene is functionally equivalent to the Drosophila Sex combs reduced gene. *Genes Dev* **7**: 343–354.