Insect Molecular Biology (1994) 3(4), 273–277

SHORT NOTE

Sequence analysis of the D1 and D2 regions of 28S rDNA in the hornet (*Vespa crabro*) (Hymenoptera, Vespinae)

J. Schmitz and R. F. A. Moritz
Technische Universität Berlin, Institut für Biologie, Berlin, Germany

Abstract

The two variable domains D1 and D2 near the 5′ end of the 28S ribosomal RNA gene (large subunit rRNA) have been sequenced for *Vespa crabro*. The sequence was aligned to corresponding rDNA regions of the wasp species *Nasonia vitripennis, Melittobia digitata* and the fruit fly *Drosophila melanogaster*. We analysed the nucleotide composition and sequence similarity for the different regions of the investigated sequences and present the inferred secondary structure of *Vespa crabro*.

Keywords: 28S rDNA sequence, secondary structure, *Vespa crabro*.

Introduction

Eukaryotic 28S rRNA consists of twelve divergent domains (D1–D12) (Hassouna *et al.*, 1984), which alternate with more or less conserved sequences. The focus of this study is the 5′ terminal region of 28S rDNA with the two variable domains D1 and D2 flanked and interrupted by conserved sequences. The rRNA contains many secondary structural elements, such as duplex, single-stranded, hairpin and bulge regions which are mainly stabilized by hydrogen bonds (Chastain & Tinoco, 1991). The secondary structure is an important requirement for interaction of the rRNA with proteins and other rRNA molecules. In this study we compare the D1 and D2 region of the European hornet *Vespa crabro* with corresponding sequence data from other insects.

Received 19 May 1994; accepted 26 July 1994. Correspondence: Dr Jürgen Schmitz, Technische Universität Berlin, FB 7-Institute für Biologie FR 1-1, Franklinstrasse 28/29, 10587 Berlin, Germany.

Results and Discussion

The amplified and sequenced 28S rDNA PCR product comprises 885 nucleotides: 160 of the D1 domain, 439 of the D2 domain and 286 from the intervening and flanking regions. The 5′ terminal sequence of the 28S rDNA was aligned with available sequences of wasp species *Nasonia vitripennis, Melittobia digitata* (Hymenoptera: Pteromalidae) (Campbell *et al.*, 1993) and *Drosophila melanogaster* (Diptera: Drosophilidae) (Tautz *et al.*, 1988) (Fig. 1). The solitary wasps were the nearest related species to the *Vespa crabro* we found in the EMBL Data Library (Heidelberg).

Nucleotide composition

The base composition of the analysed *V. crabro* sequence is 30% G, 27% C, 23% A, 20% T which is similar to the known chalcidoid wasps (Table 1). In the spacer region of *V. crabro* adenine (34%) is predominant, followed by guanine (27%), thymine (20%) and cytosine (19%). The conservation of the spacer regions may be due to functional constraint, i.e. the interaction with 5.8S rRNA (Fig. 2, underlined sites).

Sequence similarity

The pairwise percentage sequence similarity of the different regions between the studied species is given in Table 2. The aligned sequences of the spacer region show a high similarity in all compared species. The sequence of D1 is not available for the chalcidoid wasps. Hillis & Dixon (1991) argue that alignments are often ambiguous when the paired sequences differ by more than about 30%. This is the case for the *D. melanogaster* sequence. To provide nevertheless a satisfactory alignment we derived the secondary structure from *V. crabro* 28S rRNA and compared this with the secondary structure of *D. melanogaster* 28S rRNA.

Mutations may accumulate in a different rate in single-stranded versus duplex regions (Vawter & Brown, 1993).

```
V.c.  CCCCCTGAAT TTAAGCATAT TATTAAGCGG AGGAAAAGAA ACTAACCAGG   50
N.v.  .......... .........                                    19
M.d.  .......... .........                                    19
D.m.  .......... .......... ..A.T..G.. .......... ......A...   50

V.c.  ATTTCCT-AG TAGCTGCGAG CGAAGAGGAA ATAGCCCAGC ACTGAATCCC   99  D1
D.m.  ....T..T.. ....G..... ....A..A.. .C..TT.... ...A.G..A.  100

V.c.  CCGGTACA-G CCGCAG---- GAAATGTAGT GTTCGGGAGG ATCCGCCTTA  144
D.m.  TTT..CT.TA TG...AATGT ..G...C... ..AT..-..C G..AATA..C  149

V.c.  CATCCCGAGG CGCATGACCG CGTCCAAGTC CATCTTGAAT GGGGCCATTT  194
D.m.  T.GT--AT.A GAA..T.A.. -A.TT..... .T....A... .A....----  192

V.c.  ACCCGTAGAG GGTGCCAGGC CCGTAGCGAC CGGTGCAGCG TCTTGGGAGG  244
D.m.  ---....T.A C..------- ----..AT..T TAC.A----- -------..T.  218

V.c.  ATCTCTCCTT AGAGTCGGGT TGCTTGAGAG TGCAGCTCTA AGTGGGTGGT  294
N.v.                .....T.. .......T.. .......... ...T......   57
M.d.                .....T.. .......T.. .......... ...T......   57
D.m.  ..G.T...AA .......T.. .......T.. ......A... ..........  268

V.c.  AAACTCCATC TAAGGCTAAA TATGACCACG AGACCGATAG CGAACAAGTA  344
N.v.  .......... .......... .......... .......... ..........  107
M.d.  .......... .......... .......... .......... ..........  107
D.m.  .......... ...AA..... ...A....T. .......... TA........  318

V.c.  CCGTGAGGGA AAGTTGAAAA GAACTTTGAA GAGAGAGTTC AAGAGTACGT  394
N.v.  .......... .......... .......... .......... ..........  157
M.d.  ......-... .......... .......... .......... ..........  156
D.m.  .......... .......... .....C.... T........A ..C.......  368

V.c.  GAAACCG-TT CAGGGGTAAA CCTGAGGAAC CCGAAAGATC GAATGGGGAG  443  D2
N.v.  .......-.. .......... ......A... ..A.......  ..........  206
M.d.  .......-.. .......... ......A... ..A.......  ..........  205
D.m.  .....T.C.. AGA..T...G ..C..T.... .T...TATC. .TTAT..A.A  418

V.c.  ATTCAGCGTC AGCGGCGCTG GCT-CGGCCG AGTGAGCGAT GTTGCCG--A  490
N.v.  .....T.... ....A..T.. ...T.C.TGT G.ATC..... .--....GG.  254
M.d.  .....T.... .......T.. ...T.C.TGC GACTC..... .A-....AG.  254
D.m.  .....T.A.T .---------- ---------- .AATT.TA.. A..--TAAA-  446

V.c.  CTACGGTTG- GCGACACGTT ATCACTCATG CCTTGTCCGG GGTCGTCGTC  539
N.v.  .CT.....CC .T.T....CG TC.G..-GC. GTA......A CA......G.  303
M.d.  .CT....CTC ...T....CG TC-G.C-GC. -.A.......  C.......G.  301
D.m.  ---------- ---------- -----.A..A TTA..---A. AA.A..-----  464

V.c.  GTGCACTTCT CCTCTAGTA- GAACGTCGCG ACCCGTTGGG TGCCGGTCCT  588
N.v.  .......... ..........- .G........ .......... ......-...  351
M.d.  .......... ..C.......- .G........ .......... ..T...-ATC  349
D.m.  .....T..T. T.CA..TA.G ..-.A.T.TA .T.TA..A.- CAT--A.A.C  510

V.c.  ACGGTCCGAG CGGTAGACTG TCGCGTCGC- TTCG---GCG CACGCGTCAG  634
N.v.  ....C...GT T...C..... .........G ..TACGC.T. ......G...  401
M.d.  .AT.G...GT ....C.T... .T...C.T.G ---------- -G...XG...  388
D.m.  .AA--TTT.T .A-..A.A.A .AA.T.ATAG ..TATTC--- --------..A  547
```

Figure 1

```
V.c.  ACCC-TCGGT CGCCCGGCCG GCTGCCCGGC GGTC--GAAC -TTATAA---  677
N.v.  ...-AC.... T......... ......... ...ATA..T. GC.....AAC  450
M.d.  ...GA..... TCAA..T... .....T.... ...ATA.... GC.TA.TCAC  438
D.m.  TTAAA.T.C. T..AT-TTTA A.------A. A.AATAA.TG T.AT...TTT  590

V.c.  GGTATCTTGC CGCAATTATC GAACTGCGT- CGGGCCCGC- -GCAGGCGCG  724
N.v.  .....TGG.. ..G.--.--T A.G-.....T .T...T..TC G......T..  495
M.d.  ....ATGG.. .....C.--. ...-.....T .A...T..TC G......T..  485
D.m.  .A..AAG... T.------..A ..TT.ATA.- --.ATT---- -A...-T...  626

V.c.  GTCAGTGTCT CGGAGGTACG GACCCGGTGC CGTCCCCGAG --CCTGAC-C  771
N.v.  TCTC.---.. ....TT.... ....TA.... ...TG...G. CG.T..G.GT  542
M.d.  TCTC.---.. ....TT.... ....TA.C.. ...TG...G. CG.T..G.GT  532
D.m.  -.T.A.T.T. ....AT..TA ----TAA--- ---------- -----..G.AT  654

V.c.  AGCTGTTGGC AGGCGGAGTC CTCTGACGGG CAAT-ACAC- --AATAA---  814
N.v.  G..C...A.T CAA..AT... ...G...T.. .TC.T...TT CG..CGGATT  592
M.d.  G......A.T C.A..AT... ...G...T.. .TC.------ ----------  566
D.m.  .AT.A.CA-T T.AT-----.T T.G..TTTAT T.TATG...T T.T..G.--T  697

V.c.  TACCGGTCTG CGACGCTACT GC-TTT-GGG -TACTCTCAG GACCCGTCTT  861
N.v.  ........G. .......... ..-...-... -....T.... ..........  639
M.d.  .......... .......... ..-...-... -....T.... ..........  613
D.m.  ..-------A .A.T..G.AA .--A..CA.. A...CT..G. ..........  738

V.c.  GAAACACGGA CCAAGGAGTC TAAC                               885
N.v.  .......... ....                                         653
M.d.  .......... ....                                         627
D.m.  .......... .......... ....                              762
```

Figure 1. Primary structure of 5′ terminal 28S rDNA compared with other eucaryotic homologues. The underlined sequences represent the two variable domains D1 and D2 (partition as in Hassouna *et al.*, 1984).

We found compensatory mutations in duplex regions, where one fixed mutation positively selects another complementary substitution. In single-stranded sequences, on the other hand, mutations accumulated independently, except in the binding sites for other molecules. Compared to *D. melanogaster*, we detected that more than 80% of mutations in duplex regions of the investigated sequence are compensatory mutations.

We calculated the sequence similarity from alignments of duplex (including the 5.8S RNA binding site) and single-stranded regions (Table 3). The similarity values of the two separate regions within the wasps are almost identical to the similarity of the whole region. This would allow for using of the entire DNA stretch in phylogenetic studies of wasps (Hillis & Dixon, 1991). For the distantly related fruit fly the difference between single-stranded and duplex regions is higher, which would make a separate treatment of the two regions useful for phylogenetic problems.

Our analysis does not pretend to provide new insights into insect phylogeny. However, it demonstrates the feasibility of the selected 28S rDNA sequence for phylogenetic studies after considering the secondary structure for sequence alignment of more distantly related organisms. 28S rRNA is a good candidate for further unravelling the phylogeny of wasps and may allow for a comprehensive understanding of wasp evolution.

Table 1. Nucleotide composition in the conserved regions (spacer) and the variable regions (D1, D2) of *Vespa crabro*, *Nasonia vitripennis*, *Melittobia digitata* and *Drosophila melanogaster*. Nuc = Number of nucleotides in the different regions. A, C, G, T = nucleotides and their percentage of occurrence. — = sequence not available.

	Spacer					D1					D2				
Species	Nuc	A	C	G	T	Nuc	A	C	G	T	Nuc	A	C	G	T
V.c.	199	34	19	27	20	160	19	30	31	20	439	16	31	33	20
N.v.	199	34	20	25	21	—	—	—	—	—	454	15	29	33	23
M.d.	198	34	19	25	22	—	—	—	—	—	429	15	30	32	23
D.m.	200	37	19	22	22	133	32	14	20	34	341	36	11	14	39

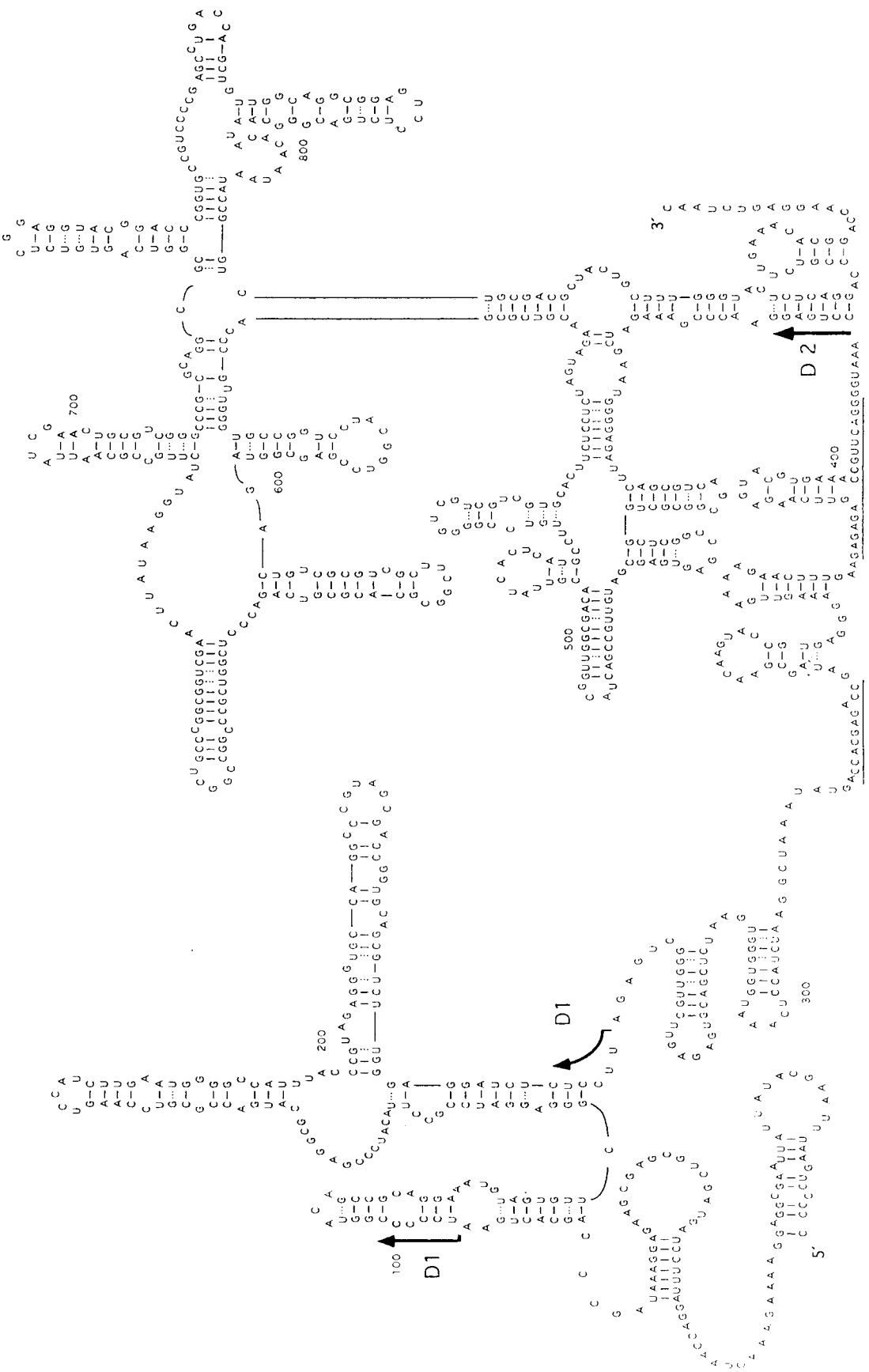

Figure 2. Putative secondary structure of the investigated sequence of *Vespa crabro*. Variable domains (D1, D2) are marked by arrows. Underlined sequences represent the binding sites for the 5.8S rRNA.

Table 2. Percentage of sequence similarity in the different regions in pairwise comparisons. Gaps were treated as one mutation event without consideration of gap length. Spa – spacer regions (sequence positions 1–19, 257–413, 853–875) D1 – sequence positions 96–255 and D2 sequence positions 414–852. — = sequence not available.

	N.v.			*M.d.*			*D.m.*		
	Spa	D1	D2	Spa	D1	D2	Spa	D1	D2
V.c.	98	—	75	98	—	72	91	57	46
N.v.				99	—	87	91	—	48
M.d.							91	—	48

Table 3. Percentage of sequence similarity in pairwise comparison in regions selected by the secondary structure. Complete region = C; duplex regions = D; single-stranded regions = S. Gaps were treated as one mutation event without consideration of gap length.

	N.v.			*M.d.*			*D.m.*		
	C	D	S	C	D	S	C	D	S
V.c.	97	96	99	96	95	98	78	74	83
N.v.				99	99	99	78	73	84
M.d.							78	74	83
D.m.									

Experimental procedures

Larval material of *V. crabro* was collected from one nest in Berlin (Germany). The nuclear DNA was prepared by standard protocols (Sambrook *et al.*, 1989).

Amplification and sequencing

We chose two primers in the 5′ terminal part of the 28S rRNA for amplifying, using PCR (Polymerase Chain Reaction) as described in Beye & Moritz (1994). The first primer is located in the highly conserved 5′ end of 28S rDNA and the second in the conserved intersperse between D2 and D3, corresponding to Vossbrinck & Friedman (1989).

Primer 1: Mo 6 5′-CCC CCT GAA TTT AAG CAT AT-3′
Primer 2: Mo 7 5′-GTT AGA CTC CTT GGT CCG TG-3′

The PCR product was purified by gel elution and Geneclean II Kit (BIO 101, Inc.), cloned in pUC 19 and sequenced by the Sanger dideoxy-chain termination technique with the Sequenase kit (Sequenase Version 2, US Biochemical). The α^{35}S-dATP labelled sequence reactions were electrophoresed in 7% acrylamide, 7 M urea gels and visualized by autoradiography. Corresponding sequences were found by screening the EMBL Data Library (Heidelberg). Clustal V (Higgins & Sharp, 1989) was used for multiple sequence alignment.

Acknowledgements

We thank Dr Andreas Pahl (TU Berlin, Institut für Biochemie) for assistance with DNA sequencing and Erik Schmolz (FU Berlin, Institut für Zoologie) for providing hornets. The Deutsche Forschungsgemeinschaft granted financial support (DFG Mo 373/4-3).

References

Beye, M. and Moritz, R.F.A. (1994) Lack of variability in the D1/D2 region of the 28S rDNA region of honeybees (*Apis mellifera*). *Apidologie* (in press).

Campbell, B.C., Steffen-Campbell, J.D. and Werren, J.H. (1993) Phylogeny of the *Nasonia* species complex (Hymenoptera: Pteromalidae) inferred from an internal transcribed space (ITS2) and 28S rDNA sequences. *Insect Molec Biol* **2**: 225–237.

Chastain, M. and Tinoco, I. (1991) Structural elements in RNA. *Prog Nucl Acid Res Molec Biol* **41**: 131–177.

Hassouna, N., Michot, B. and Bachellerie, J.-P. (1984) The complete nucleotide sequence of mouse 28S rRNA gene. Implications for the process of size increase of the large subunit rRNA in higher eukaryotes. *Nucl Acids Res* **12**: 3563–3583.

Higgins, D.G. and Sharp, P. M. (1989) Fast and sensitive multiple sequence alignments on a microcomputer. *CABIOS* **5**: 151–153.

Hillis, D.M. and Dixon, M.T. (1991) Ribosomal DNA: molecular evolution and phylogenetic inference. *Quart Rev Biol* **66**: 411–453.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory Press.

Tautz, D., Hancock, J.M., Webb, D.A., Tautz, C. and Dover, G.A. (1988) Complete sequences of the rRNA genes of *Drosophila melanogaster*. *Molec Biol Evol* **5**: 366–375.

Vawter, L. and Brown, W.M. (1993) Rates and patterns of base change in the small subunit ribosomal RNA gene. *Genetics* **134**: 597–608.

Vossbrinck, C.R. and Friedman, S. (1989) A 28S ribosomal RNA phylogeny of certain cyclorrhaphous Diptera based upon a hypervariable region. *Syst Entomol* **14**: 417–431.