Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 5483–5486, July 1989
Genetics

The homeotic gene spalt (*sal*) evolved during *Drosophila* speciation

DIRK REUTER, REINHARD SCHUH, AND HERBERT JÄCKLE

Institut für Genetik und Mikrobiologie, Universität München, Maria-Ward-Strasse 1a, D 8000 Munich 19, Federal Republic of Germany

Communicated by W. Beermann, March 27, 1989 (received for review December 21, 1988)

ABSTRACT **The region-specific homeotic gene spalt (*sal*) acts in two separate domains in the head and tail region of the *Drosophila melanogaster* embryo. Based on comparative morphology, *sal* is likely to be involved in the establishment of the head during the evolution of invertebrates and thus, it should be conserved. We have analyzed the conservation of the segmentation genes Krüppel (*Kr*) and even-skipped (*eve*) in parallel with *sal* coding sequences in several *Drosophila* species that are evolutionarily separated by up to 60 million years. To our surprise, *sal* sequences appear to be conserved in the *Sophophora* subgenus of the *Drosophila* genus but not in the *Drosophila* subgenus. On the other hand, the segmentation and other homeotic genes are conserved in the *Drosophila* subgroup as well. Our data suggest that *sal* encodes an accessory function that evolved relatively late during *Drosophila* speciation rather than playing a fundamental evolutionary role similar to that of other homeotic genes.**

Homeosis in *Drosophila* results in a diversity of phenotypic effects that have in common a transformation of segments or parts of them. Studies of homeotic mutations have revealed a limited number of selector genes that determine the unique identity of particular segments (1). Mutations in these genes result in transformations between segments or parts of them that include the conversion of antennae into legs, halteres into wings, or proboscis into antennae (2–4). In *Drosophila melanogaster* (*D.m.*), most of these mutations correspond to genes that are included in two gene complexes, the Antennapedia complex (ANT-C) and the bithorax complex (BX-C) (for review see refs. 5 and 6). Their activity is essential to form the normal segment pattern in the trunk region of the embryo (5, 6). Recently, two unique homeotic genes, spalt (*sal*) and fork head (*fkh*), have been identified (7, 8). In contrast to the genes of the ANT-C and BX-C, both *sal* and *fkh* are required in the head and in the tail. Furthermore, *sal* and *fkh* are not included in the network of hierarchical regulatory interactions of the ANT-C and BX-C genes, and they act independently of genes that regulate the BX-C and ANT-C genes (7, 8).

Several of the segmentation genes that are required to set up the spatial patterns of the ANT-C and BX-C genes, as well as the homeotic selector genes, encode a 60-amino acid DNA-binding helix-turn-helix motif, the homeodomain (9). The members of the gap class of segmentation genes encode zink fingers, which are indicative of DNA-binding properties as well (for review see ref. 6). In addition, protein motifs observed in other segmentation genes (for review see ref. 6) and genes involved in specifying the dorso-ventral axis or neurogenesis are highly conserved during insect evolution and in vertebrates (for review see ref. 6).

sal encodes a protein with previously unknown features (10). In view of the conservation of most genes required for the establishment of *Drosophila* segmentation, we expected *sal* sequences or a portion of the *sal* sequence to be conserved as well. Here we report that the protein products of the Krüppel (*Kr*) and even-skipped (*eve*) genes and their spatial distributions have been conserved during *Drosophila* evolution. By contrast, *sal* coding sequences seem to be restricted to species that are closely related to *D.m.* Our data suggest that *sal* encodes a function acquired during *Drosophila* speciation.*

MATERIALS AND METHODS

Stock Maintenance and Embryos. The different *Drosophila* stocks were maintained on standard food supplemented with fresh yeast, and embryos were collected on apple juice agar plates as described (11).

DNA Techniques. DNA preparations from flies and phages, Southern hybridization, and cloning procedures were as described (10) using lambda ZAP as a vector (Stratagene). Hybridization was carried out as described (12); low stringency hybridization was at 55°C under otherwise identical conditions. After hybridization, filters were washed at 65°C with 0.5× SSPE (1× SSPE = 0.18 M NaCl/10 mM sodium phosphate, pH 7.4/1 mM EDTA) (high stringency), at 65°C with 1× SSPE (moderate stringency), or at 65°C with 4× SSPE (low stringency) following a protocol published earlier (12). After cloning, subfragments that hybridized to the *sal* cDNA clone (10) were subcloned and sequenced using the dideoxynucleotide sequencing method as described (12).

Antibody Staining. Antibodies directed against the *sal* protein were prepared as described for the *Kr* protein (11). Specificity of the antibodies was shown by the lack of the *sal* protein band on Western blots loaded with protein extracts from homozygous *sal* embryos (which carry a lack of function allele) that were allowed to react with the affinity-purified antibodies (11). Whole-mount staining and Western blotting were carried out as described (11).

RESULTS AND DISCUSSION

Conservation of Proteins Controlling Embryonic Pattern Formation in Different *Drosophila* Species. Embryogenesis of the different *Drosophila* species (Fig. 1) varies in duration, but no significant morphological differences can be observed between the embryos of different members of the *Sophophora* and *Drosophila* subgenera except for their final size. The lack of morphological differences suggests that the genetic basis for pattern formation of *Drosophila* is highly conserved, which implies that the underlying molecular mechanism are conserved as well.

Drosophila hydei (*D.h.*) and *Drosophila virilis* (*D.v.*) represent two species groups belonging to the subgenus *Drosophila*. Both species are separated from *D.m.* by 60 million

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: *D.m.*, *D.h.*, *D.v.*, *D.s.*, *D.o.*, and *D.p.*, *Drosophila melanogaster*, *D. hydei*, *D. virilis*, *D. simulans*, *D. orena*, and *D. pseudobscura*, respectively; ANT-C, Antennapedia complex; BX-C, bithorax complex.

*The sequences reported in this paper have been deposited in the GenBank data base (accession nos. S00262 for *D. melanogaster*, M21227 for *D. simulans*, and M21579 for *D. orena*).

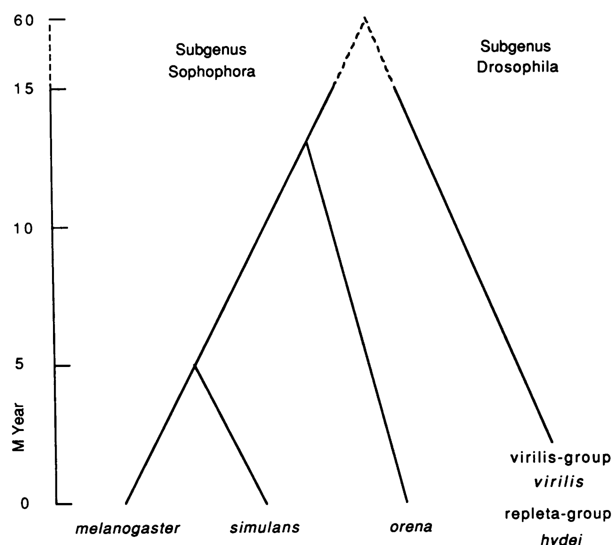

FIG. 1. Evolutionary separation of *Drosophila* species: The evolutionary distances of different *Drosophila* species that are separated into the *Sophophora* and *Drosophila* subgroups. The distances refer to million years of separation (ordinate). For details see ref. 13.

years of evolution (Fig. 1). The homeotic gene Ultrabithorax (*Ubx*) and the segment polarity gene engrailed (*en*) of *D.m.* are conserved in *D.v.* The degree of amino acid conservation is more than 70% and 85%, respectively (13, 14). To examine whether other segmentation gene products and their spatial domains of action are conserved in *Drosophila* evolution as well, we used antibodies directed against other segmentation gene products, the proteins encoded by the gap gene Krüppel (*Kr*) and by the pair rule gene even-skipped (*eve*) of *D.m.* In *D.h.* embryos (Fig. 2 *A* and *B*), the staining patterns of *Kr* and *eve* antibodies were similar, if not identical, to those of *D.m.* (6, 11). Thus, the protein products of these two genes, and their domains of action, have been conserved during the 60 million years of evolution that separate the two species (Fig. 1). Similarly, antibodies against the *D.m.* gap gene hunchback (*hb*) and the pair rule gene fushi tarazu (*ftz*) recognize the corresponding proteins in *D.v.* They also show appropriate spatial regulation in *D.v.* (15).

In contrast to these antibodies, the antibodies directed against the *sal* protein of *D.m.* fail to detect the *sal* protein in whole-mount embryos (R.S., unpublished observation). However, the antibodies recognize a single protein band of 17 kDa (molecular mass) on Western blots prepared from denatured proteins of *D.m.* embryos (Fig. 2*C*). This is in good agreement with the molecular mass of the *sal* protein deduced from the DNA sequence (10). As shown in Fig. 2*C*, the *sal* antibodies also recognize the corresponding protein band of *D.s.* and *D.o.*: all of them members of the *Sophophora* subgenus, but they fail to detect *sal* protein in *D.h.*, a member of the *Drosophila* subgenus (Fig. 1). This observation suggests that *sal* epitopes are only conserved in *Sophophora*. Alternatively, *sal* antibodies may recognize only epitopes that are subject to evolutionary changes. To address this issue, we examined the conservation of *sal* coding sequences in different species of the *Drosophila* genus.

Absence of *sal* Homologous DNA in *D.h.* We hybridized the *sal* cDNA clone of *D.m.* to Southern blots loaded with genomic DNA of various *Drosophila* species under different stringency conditions (Fig. 3). Under moderate stringency, *sal* cDNA hybridized to a single DNA fragment of *D.o.* and *D.s.*, the two *Sophophora* species analyzed (Fig. 3). In contrast, no hybridization signal could be observed with the more distantly related species *D.v.* and *D.h.* (example shown in Fig. 3). At low stringency hybridization conditions a series of weakly hybridizing DNA fragments could be observed with all five *Drosophila* species examined, including *D.h.* (Fig. 3). However, DNA sequence analysis showed that the hybridization signals obtained under the low stringency conditions derived from short stretches of sequence identity of 20–30 bp in length (data not shown). Thus, significant sequence homology between the *sal* cDNA and *D.h.* DNA could not be observed. This observation is consistent with the results obtained by antibody staining (see above), and it

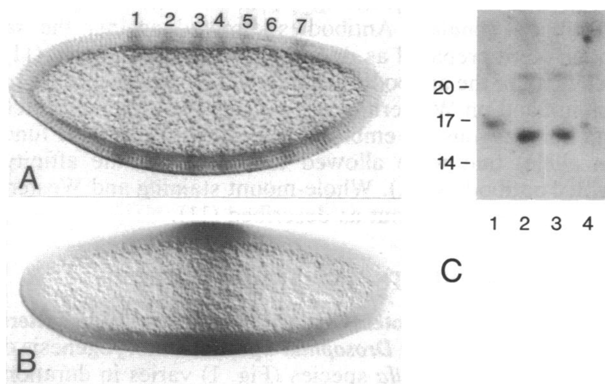

FIG. 2. Conservation of *D.m.* segmentation gene products in other *Drosophila* species. (*A*) *D.h.* embryo at blastoderm stage stained with *eve* antibodies prepared from the *D.m.* protein. Numbers at the dorsal side (posterior is left) refer to the stripes 1–7. Note that the width and position of stripes are identical to those observed in *D.m.* embryos at the same stage of development. (*B*) *D.h.* embryo at early blastoderm stage stained with *Kr* antibodies prepared against the *D.m.* protein (orientation of the embryo as in *A*). Note that the *Kr* domain is identical to the corresponding domain in *D.m.* embryos at the same stage. (*C*) Western blot stained with *sal*-specific antibodies. Note a 17-kDa protein band in *D.m.* (lane 1), a 16.5-kDa band in *Drosophila simulans* (*D.s.*) (lane 2) and in *Drosophila pseudobscura* (*D.p.*) (lane 3), but no band in *D.h.* (lane 4). The differences in molecular mass between the *sal* proteins *D.m.*, *D.s.*, and *D.p.* may be due to differences in protein loading and/or migration artifacts caused by the predominant yolk material. Numbers refer to the molecular masses of standard proteins (given in kDa) used as markers.

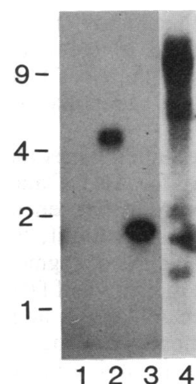

FIG. 3. Southern blots hybridized with the *sal* cDNA probe under high and low stringency conditions. Numbers to the left refer to size markers (given in kb). Note the lack of hybridization signals in *D.h.* DNA (lane 1) (except for a smear after about 10-fold overexposure; data not shown) and two distinct bands in *Drosophila orena* (*D.o.*) (lane 2) and *D.s.* (lane 3). Lane 4 shows *D.h.* DNA hybridized and washed under low stringency conditions. Note a number of bands not seen under higher stringency (lane 1). The corresponding DNA fragments were cloned and sequenced. Sequence identity with *sal* cDNA was 30 base pairs (bp) maximum.

implies that the *sal* corresponding gene of *D.h.* is not coded by *sal* similar DNA sequences.

***sal* Sequence Diversity in Sibling Species.** The *sal* coding sequences of *D.m.* are conserved in *D.s.* and *D.o.*, two members of the *Sophophora* subgenus (Fig. 4). Cloning of the *sal* homologous genes followed by sequence analysis revealed that 92% of the nucleotide sequence and 94% of the amino acids are conserved in *D.s.* (Fig. 5). *D.o. sal* sequences are conserved to 87% both at the nucleotide and the amino acid levels. In all three species, the position of the small (57 bp) *sal* intron after the fourth amino acid is conserved (Fig. 4). The nucleotide replacements within the intron are more pronounced than in the coding region—i.e., 23% and 46% of the nucleotides in the intron of *D.s.* and *D.o.* are different from the *D.m.* intron sequence.

For *Drosophila*, the rate of silent nucleotide replacements ("synonymous replacements") in the coding region is 1×10^{-8} nucleotides per site per year of separation (16). Based on this value, the comparison of *sal* sequences from *D.m.*, *D.s.*, and *D.o.* revealed an evolutionary distance of 9.8 million years for *D.m.* and *D.s.* and 21 million years for *D.m.* and *D.o.* These values are more than three times the value obtained with the sequences of the alcohol dehydrogenase gene and a heat shock gene. The latter values of 3.1 and 13 million years are in agreement with the evolutionary distance calculated from protein polymorphisms (17, 18). This indicates that the *sal* sequences diverge more rapidly during *Drosophila* evolution than the other genes or gene products analyzed so far.

CONCLUSIONS

Homeotic mutations have been observed in other insects, including *Musca*, *Aedes*, *Anopheles*, and *Blattella* (for review see ref. 19). In addition, there is genetic evidence that the homeotic genes that specify the trunk region are conserved among the insects. A homeotic gene cluster in the flour beetle *Tribolium* contains the genetic elements of both gene complexes of *D.m.* and shows sequence similarity with the *Drosophila Antennapedia* homeobox (19). A tightly linked series of homeotic mutations affecting thorax and abdominal segments has been described in the silkworm *Bombyx mori* (20). Furthermore, all of the segmentation and homeotic genes of *D.m.* that have been analyzed so far—with the exception of *sal*—appear to be conserved at the molecular level. These observations are consistent with the hypothesis that the molecular principles that establish the *Drosophila* body are conserved during insect evolution.

Genetic analysis of the *sal* function in *D.m.* revealed that the *sal* gene product is required for head development as opposed to trunk development (7). The other homeotic selector genes are thought to be involved in the evolution of insects from myriapod-like ancestors (1–3). By analogy, *sal* would fit the hypothesis that insects arose from annelid-like ancestors that reached the organizational level of myriapods by integrating the anterior trunk region into a primitive head (7). This hypothesis, advanced by comparative morphologists, would imply that the *sal* gene or a *sal*-related gene is evolutionarily older than the BX-C and the ANT-C genes.

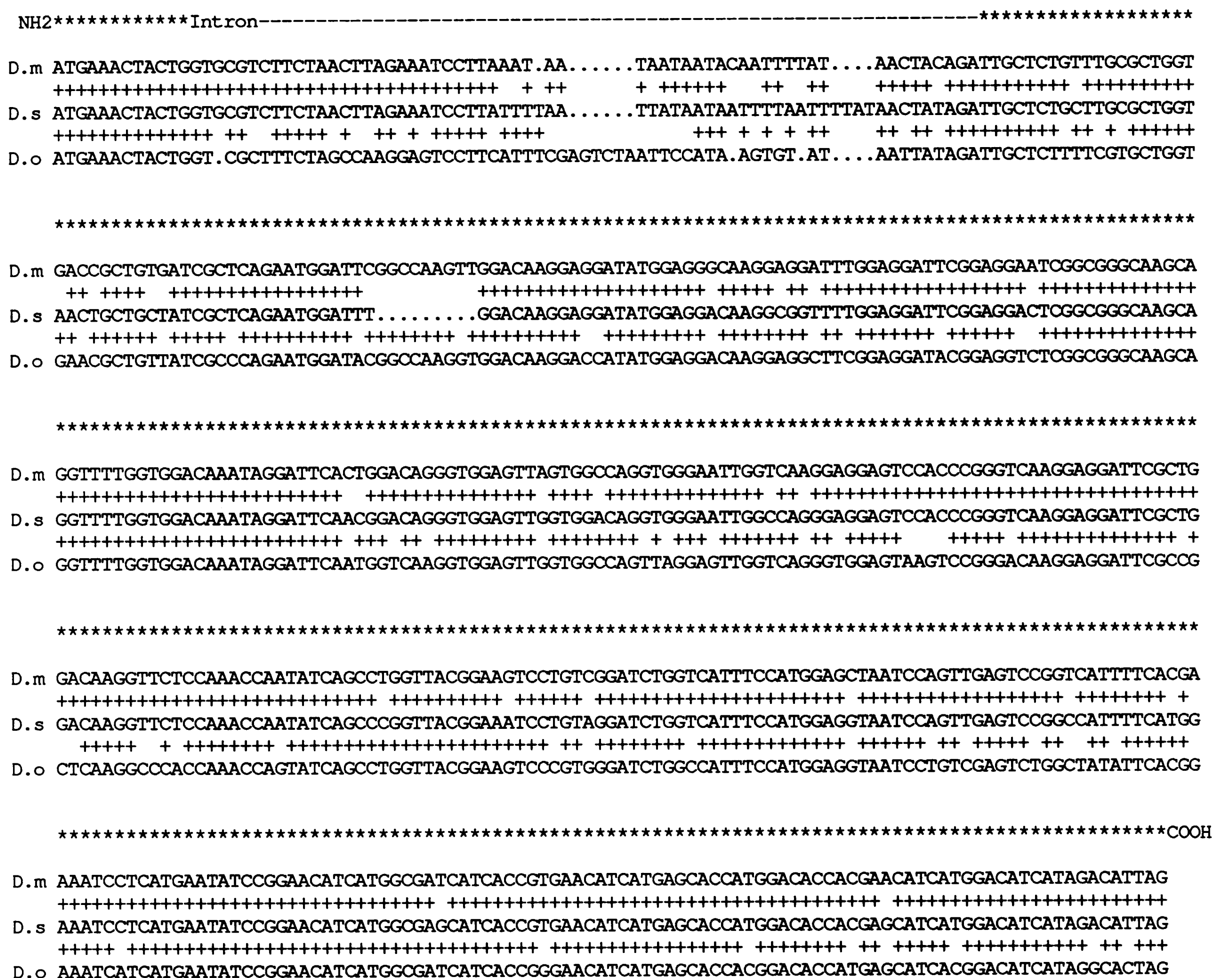

FIG. 4. Comparison of the genomic *sal* DNA sequences of *D.m.*, *D.s.*, and *D.o.* Sequences include the coding region (asterisks) and the intron, starting at nucleotide position 733 of the published *sal* sequence (10). Identity of nucleotides is indicated by crosses. For the corresponding amino acid sequence see Fig. 5.

```
    1           5              10             15
D.m MetLysLeuLeuIleAlaLeuPheAlaLeuValThrAlaValIle
D.s  *  *  *  *  *  *  * Leu *  *  *  *  * Ala * 
D.o  *  *  *  *  *  *  *  * Val *  * Asn *  *  * 

    16          20             25             30
D.m AlaGlnAsnGlyPheGlyGlnValGlyGlnGlyGlyTyrGlyGly
D.s  *  *  *  *  *  -  -  -  *  *  *  *  *  *  * 
D.o  *  *  *  * Tyr *  * Gly *  *  * Pro *  *  * 

    31          35             40             45
D.m GlnGlyGlyPheGlyGlyPheGlyGlyIleGlyGlyGlnAlaGly
D.s  *  *  *  *  *  *  *  *  * Leu *  *  *  *  * 
D.o  *  *  *  *  *  * Tyr *  * Leu *  *  *  *  * 

    46          50             55             60
D.m PheGlyGlyGlnIleGlyPheThrGlyGlnGlyGlyValSerGly
D.s  *  *  *  *  *  *  * Asn *  *  *  *  * Gly * 
D.o  *  *  *  *  *  *  * Asn *  *  *  *  * Gly * 

    61          65             70             75
D.m GlnValGlyIleGlyGlnGlyGlyValHisProGlyGlnGlyGly
D.s  *  *  *  *  *  *  *  *  *  *  *  *  *  *  * 
D.o  * Leu * Val *  *  *  *  * Ser *  *  *  *  * 

    76          80             85             90
D.m PheAlaGlyGlnGlySerProAsnGlnTyrGlnProGlyTyrGly
D.s  *  *  *  *  *  *  *  *  *  *  *  *  *  *  * 
D.o  *  * Ala *  * Pro *  *  *  *  *  *  *  *  * 

    91          95             100            105
D.m SerProValGlySerGlyHisPheHisGlyAlaAsnProValGlu
D.s Asn *  *  *  *  *  *  *  *  * Gly *  *  *  * 
D.o  *  *  *  *  *  *  *  *  *  * Gly *  *  *  * 

    106         110            115            120
D.m SerGlyHisPheHisGluAsnProHisGluTyrProGluHisHis
D.s  *  *  *  *  * Gly *  *  *  *  *  *  *  *  * 
D.o  *  * TyrIle * Gly * His *  *  *  *  *  *  * 

    121         125            130            135
D.m GlyAspHisHisArgGluHisHisGluHisHisGlyHisHisGlu
D.s  * Glu *  *  *  *  *  *  *  *  *  *  *  *  * 
D.o  *  *  *  *  *  *  *  *  *  *  *  *  *  *  * 

    136         140
D.m HisHisGlyHisHisArgHisEND
D.s  *  *  *  *  *  *  * 
D.o  *  *  *  *  *  *  * 
```

FIG. 5. Comparison of putative *sal* protein sequences of *D.m.*, *D.s.*, and *D.o.* Amino acid identities are indicated by asterisks. Note the lack of three amino acids (dashes) in *D.s.* protein.

However, *sal* sequences, in contrast to those of other segmentation and homeotic genes analyzed (see above), appear to be restricted to the melanogaster group of the subgenus *Sophophora*. This suggests that the formation of normal head and tail structures is independent of *sal* function in the *Drosophila* subgenus. Therefore, *sal* is likely to serve an accessory or modulating function that acts in combination with another older gene that may have the evolutionary traits to myriapod-like ancestors. This proposed gene could have escaped the mutant screens in *D.m.* that were designed to identify segmentation genes or genes that act in the trunk region.

We thank our colleagues in the laboratory for various contributions and Dr. Ulrike Gaul for providing the *eve* antibodies. D.R. was supported by a fellowship of the Studienstiftung des deutschen Volkes. The work was supported by a grant of the Deutsche Forschungsgemeinschaft (Leibniz-Programm).

1. Garcia-Bellido, A. (1977) *Am. Zool.* **17,** 613–629.
2. Kaufmann, T. C. (1983) in *Time, Space and Pattern in Embryonic Development,* eds. Jeffrey, W. R. & Raff, R. A. (Liss, New York), pp. 365–383.
3. Lewis, E.B. (1978) *Nature (London)* **276,** 565–570.
4. Kaufmann, T.C., Lewis R. & Wakimoto, B. (1980) *Genetics* **94,** 115–133.
5. Akam, M. (1987) *Development* **101,** 1–22.
6. Ingham, P.W. (1988) *Nature (London)* **335,** 25–34.
7. Jürgens, G. (1988) *EMBO J.* **7,** 189–196.
8. Jürgens, G. & Weigel, D. (1988) *Wilhelm Roux' Arch. Dev. Biol.* **197,** 345–354.
9. Levine, M. & Hoey, K. (1988) *Cell* **55,** 537–540.
10. Frei, E., Schuh, R., Baumgartner, S., Burri, M., Noll, M., Jürgens, G., Seifert, E., Nauber, U. & Jäckle, H. (1988) *EMBO J.* **7,** 197–204.
11. Gaul, U. & Jäckle, H. (1987) *Cell* **51,** 549–555.
12. Schuh, R., Aicher, W., Gaul, U., Cote, S., Preiss, A., Maier, D., Seifert, E., Nauber, U., Schröder, C., Kemler, R. & Jäckle, H. (1986) *Cell* **47,** 1025–1032.
13. Wild, C.D. & Akam, M. (1987) *EMBO J.* **6,** 1393–1401.
14. Kassis, J.A., Poole, S.J., Wright, D.K. & O'Farrell, P.H. (1986) *EMBO J.* **5,** 3583–3589.
15. Treier, M., Pfeifle, C. & Tautz, D. (1989) *EMBO J.* **8,** 1517–1525.
16. Hayashida, H. & Miyata, T. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 2671–2675.
17. Bodmer, M. & Ashburner, M. (1984) *Nature (London)* **309,** 425–430.
18. Blackman, R. & Meselson, M. (1986) *J. Mol. Biol.* **188,** 499–515.
19. Beeman, R. (1987) *Nature (London)* **327,** 247–249.
20. Ouweneel, W.J. (1976) *Adv. Genet.* **18,** 179–248.