Proc. Natl. Acad. Sci. USA
Vol. 82, pp. 2814–2818, May 1985
Evolution

Silkmoth chorion gene families contain patchwork patterns of sequence homology

(multigene families/concerted evolution/unequal crossover/gene conversion)

THOMAS H. EICKBUSH AND WILLIAM D. BURKE

Department of Biology, University of Rochester, Rochester, NY 14627

Communicated by Robert K. Selander, December 21, 1984

ABSTRACT **The late chorion locus of *Bombyx mori*, containing paired members from two multigene families, has been analyzed in detail. The 15 gene pairs, irregularly spaced over 140 kilobases, exhibit an identical structure and a high overall sequence homology, while the flanking DNA (intergene regions) varies considerably. Segments of DNA of 150–300 base pairs from a reference gene pair were used as probes in a series of DNA hybridization experiments. It was found that the sequence variants within the locus are in a "patchwork" arrangement. Each gene pair contains a unique pattern of regions or "patches" that are highly homologous to the reference gene pair, interspersed with regions exhibiting less homology. We suggest that phenotypic selection is unable to account for the observed patchwork patterns; rather, sequence exchange between genes must be postulated. The nature of the DNA flanking the gene pairs would suggest that most of this sequence exchange is by means of gene conversion rather than unequal crossover events.**

When one compares the same multigene family in two species, characteristic sequence features can be found that are shared by all family members in one species but are not seen in the family members of the second species. These observations of "concerted evolution" (reviewed in refs. 1–4) imply processes, termed molecular drive, that spread and fix new sequence variants in a multigene family (3). Concerted evolution was originally identified in the highly homogeneous, tandemly repeated, ribosomal genes from a number of species. The maintenance of identical sequences within these genes can, for the most part, be explained by the expansion and contraction of the gene families by unequal crossing-over, with fixation of variants determined by a combination of selective and stochastic mechanisms (5, 6). Today, the discovery of clustered multigene families is common. Extensively studied examples include the mammalian globin loci, the immunoglobulin loci, and the major histocompatability loci. Detailed analyses have indicated concerted evolution within these families, although to a lesser extent than that in the ribosomal genes (7–11). Unequal crossing-over and gene-conversion events are not sufficiently frequent to completely homogenize sequences within these families. Thus, each member gene is able to establish and maintain its own sequence identity through the accumulation of mutations.

Another multigene system composed of clustered copies of nonidentical genes is the chorion, or eggshell, locus of the silkmoth, *Bombyx mori*. Structural studies of this system have indicated that more than 100 chorion proteins can be divided into a small number of families on the basis of their amino acid sequence and period of expression, while genetic data have shown that all chorion structural genes are localized to a specific segment of chromosome number 2 (for a comprehensive review, see ref. 12). Molecular cloning of a large region of this locus has demonstrated that members of the same multigene family are tandemly arranged (13). In particular, all members from each of two "developmentally late" families of chorion protein genes (14) are arranged in 15 pairs on a 140-kilobase (kb) chromosomal segment (13). The members of each family exhibit a high degree of sequence homology but are not identical, as determined by their differential hybridization to specific cDNA probes at an extremely high hybridization criterion (0.3 M NaCl at 85°C).

In this report we have attempted to determine the manner in which these chorion gene families are evolving. Can the high sequence homology be explained by the relatively recent amplification of the locus, with strong selective pressure to maintain identity? Or is the current arrangement of the locus best explained by assuming that the genes are undergoing concerted evolution; and if so, what are the recombinational mechanisms responsible?

MATERIALS AND METHODS

Subcloning and Mapping of Gene Regions. Genomic segments, obtained from a previously described (13) series of overlapping Charon 4 clones, were subcloned into pBR322 or pUC13 plasmid vectors. Subclones correspond to individual *Eco*RI restriction fragments containing either a single chorion gene or gene pair (see Fig. 1), except for gene pair 3, which is located on an *Sst* I fragment, and pair 4, which resides on a *Sal* I–*Eco*RI fragment. Restriction maps were constructed by simple single and double digests in the case of small subclones and by partial digestion of end-labeled fragments (15) in the case of certain larger subclones. The restriction map for gene pair 12 was obtained from its published sequence (16, 17).

Slot Blots. The subclones were linearized with restriction endonucleases, heated at 65°C in 25 mM EDTA (pH 8) and stored frozen. For each experiment, equal molar amounts of the plasmid subclones (0.025–0.1 μg), along with 1 μg of calf thymus DNA, were denatured for 10 min at room temperature in 0.4 M NaOH, neutralized with 40 vol of 1 M NH_4OAc/2 mM EDTA (18), and applied to nitrocellulose paper (presoaked for 1 hr in 1 M NH_4OAc) by means of a Schleicher and Schuell slot blotter apparatus. Filters were next washed with 1 M NH_4OAc, air-dried, and baked *in vacuo* at 80°C. Equal molar concentrations of each plasmid were originally estimated from A_{260} measurements (adjusting for the size of the plasmid). Additional minor adjustments were made after scanning the autoradiograms from test hybridizations of the filter-bound DNAs to the 2.1-kb *Hae* II fragment common to both pBR322 and pUC13 plasmid vectors (19). These adjustments, based upon uniform hybridization of vector sequences, were needed because of the

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: kb, kilobase(s); Hc-A, high-cysteine A-like; Hc-B, high-cysteine B-like; bp, base pair(s).

variable degree of macromolecular contamination (predominately RNA) in the plasmid DNA preparations, which were isolated at various times over a 2-yr period. All hybridizations were conducted in 2× NaCl/Cit (1× NaCl/Cit = 0.15 M NaCl/0.015 M sodium citrate) containing 10% dextran sulfate, 100 μg of denatured calf thymus DNA per ml, 0.05 M NaH_2PO_4 (pH 6.8), 0.2% $NaDodSO_4$, and 5× Denhardt's solution (0.02% polyvinylpyrrolidone/0.02% Ficoll/0.02% bovine serum albumin) for 8 hr at 72–82°C. After hybridization the filters were washed for 20-min intervals at 70°C twice in 2× NaCl/Cit containing 0.2% $NaDodSO_4$ and twice in 0.1× NaCl/Cit containing 0.2% $NaDodSO_4$. All fragments used as probes were isolated as specific bands from low-melting-temperature agarose gels, purified by phenol extraction, and labeled by nick-translation (13).

RESULTS

General Organization of the Locus. The last two families of chorion proteins produced during *B. mori* choriogenesis are unusually rich in cysteine residues (12). By convention they have been labeled the high-cysteine A-like (Hc-A) proteins and the high-cysteine B-like (Hc-B) proteins because of their homology to previously characterized A and B chorion proteins (17, 20). The tight pairing of the genes encoding these proteins and their tandem arrangement is diagramed in Fig. 1. Each gene pair has been given a number starting at the left end of the locus. Thus, for example, Hc-A.7 corresponds to the Hc-A gene of the seventh gene pair.

Even at the low level of resolution shown in Fig. 1, several structural features of the late locus indicate that it is not simply composed of identical tandem repeats. First, four of the gene pairs (Hc-A/Hc-B.1 through Hc-A/Hc-B.4) are in an inverted orientation relative to the remainder of the gene pairs, suggesting at least one inversion event in the evolutionary history of this locus. Second, flanking certain gene pairs are DNA segments that show weak homology with either Hc-A or Hc-B probes. At least one of these segments corresponds to a transcribed pseudogene (unpublished data). Finally, the distance separating consecutive gene pairs is quite variable in length, ranging from 2 to 15 kb. Most of this flanking DNA can be classified as either middle or highly repetitive DNA of the *Bombyx* genome (unpublished results). These flanking DNA regions were used to great advantage in the original cloning of the late locus because most of this DNA is repeated infrequently within the late locus (13). Variation in the flanking DNA between consecutive A/B gene pairs has been attributed to insertion/deletion events in *Antheraea polyphemus* (21). We assume that similar mechanisms have given rise to the variable flanking DNA in the late locus of *Bombyx*.

Conserved Structure of all 15 Gene Pairs. DNA sequence information from pairs 12 and 13 indicated that in these pairs the Hc genes are divergently oriented, with approximately 250 base pairs (bp) separating their 5′ ends. Each gene is separated into two exons by an intervening sequence located at amino acid residue −4 of the leader peptide (16, 17, 22). This basic structure is identical to that reported for the A/B chorion genes of *A. polyphemus* (21). To determine how similar the remaining Hc genes are to the sequenced pairs, each of the 15 gene pairs was subcloned, and detailed restriction maps were generated. A comparison of some of the restriction sites from these maps for the approximately 4-kb region surrounding each gene pair is presented in Fig. 2. Below these restriction maps is a schematic diagram approximately locating the 5′ and 3′ borders and exon/intron regions of the divergently transcribed genes, as determined by sequence analysis of repeats 12 and 13.

A variety of restriction cleavage sites is conserved in every or nearly every Hc gene pair, which unambiguously positions the genes. These include *Hin*dIII, *Pst* I, and *Hin*cII sites in the 5′ flanking region, a *Bgl* I site in the small exon of Hc-B genes, an *Ava* I site in the large exons of all Hc-B genes, and a *Sau*96I site within the large exon of all Hc-A genes. Most of the remaining restriction sites shown in Fig. 2 are conserved within only a fraction of the gene pairs. For example, variable combinations of *Hin*dIII, *Bgl* II, and *Eco*RI sites are found in the intron of the Hc-A genes, while *Bgl* II, *Hin*cII, or *Sst* I sites are conserved immediately 3′ of certain Hc-A and Hc-B genes. Differences in the relative position of these 3′ flanking sites, especially in the Hc-B genes, correspond to differences in the length of the carboxyl-terminal arms of Hc proteins, which are composed of simple tandem repeats (17, 22). This size difference (indicated by the multiple arrowheads on the exon/intron diagram of Fig. 2) is probably more extreme in the case of the Hc-B proteins, as judged by the greater shifts in position of the conserved *Hin*cII and *Sst* I sites flanking Hc-B genes than shifts in the *Bgl* II site flanking Hc-A genes. The only major structural variation in any of the 15 gene pairs is the presence of a 900-bp insertion within Hc-A.3 (jagged lines in Fig. 2). Additional mapping (data not shown) indicated that this insert is located within the Hc-A intron.

While the exact spacing of conserved restriction sites between repeats indicates that the 5′ flanking and most of the coding regions are highly conserved between the Hc gene pairs, the regions immediately 3′ are more variable. No conserved restriction sites were found beyond the *Bgl* II site at the 3′ border of the Hc-A genes, except for two interesting cases where the maps are completely identical: Hc-A.6 and Hc-A.7, Hc-A.13 and Hc-A.15. This identity terminates abruptly at the end of the map in Fig. 2 in the case of Hc-A.6 and -.7 and extends for up to another kilobase in the case of Hc-A.13 and -.15 (data not shown). For the Hc-B genes, there are nine instances of an *Eco*RI site 600–900 bp from the 3′ end of the gene. These sites would suggest a somewhat

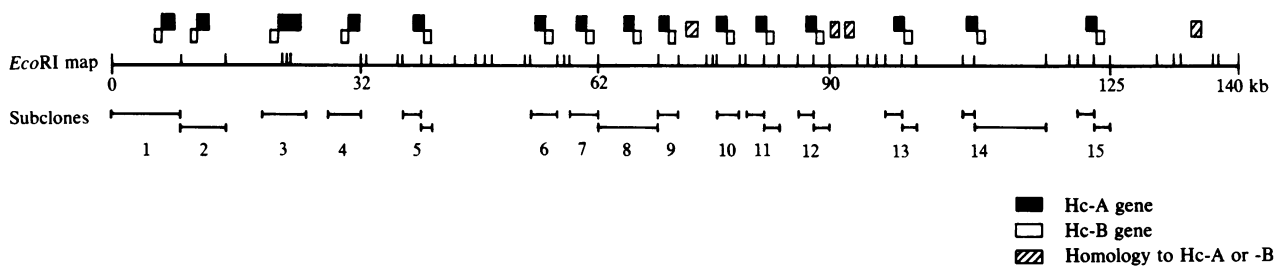

FIG. 1. Chromosomal *Eco*RI map of the Hc chorion locus. The *Eco*RI site at the left end of the locus has arbitrarily been made a reference point, and the distances from it of other approximately equally spaced *Eco*RI sites (vertical cross bars) are indicated. The location of pairs of chorion genes belonging to the two Hc gene families are indicated by the boxes above the map and are numbered sequentially from the left. The location of genes or pseudogenes that exhibit limited sequence homologies to these families are also indicated but not numbered. Regions of the locus subcloned for the purposes of restriction mapping and hybridization experiments are drawn below the *Eco*RI map.

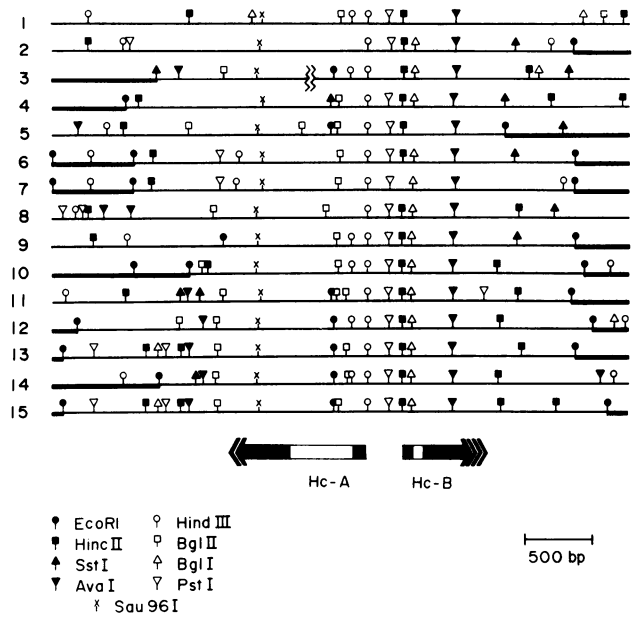

FIG. 2. Comparison of the restriction endonuclease maps for each of the 15 Hc gene pairs. Gene pair numbers at the left are those used in Fig. 1. The maps for pairs 1–4 have been inverted relative to their orientation in Fig. 1 in order to compare all 15 gene pairs directly. Thicker line width at the ends of some maps correspond to DNA regions outside the regions subcloned into plasmids. The restriction maps for these regions are derived from the original charon 4 clones (13) and does not include *Hinc*II, *Ava* I, or *Pst* I sites. The single *Sau*96I cleavage site located in each gene pair is diagnostic of Hc-A genes and was mapped by either partial digestion of end-labeled DNA (15) or by Southern analysis using as probe a *Sau*96I–*Hin*fI site from repeat 13 (probe d in Fig. 3). Additional *Sau*96I sites are present throughout the gene regions but in most cases have not been precisely mapped. The orientation and approximate boundaries of the Hc-A and Hc-B genes, derived from the sequence information for gene pairs 12 and 13 (16, 17, 22), are indicated by the arrows below the maps. Solid regions of the arrows correspond to exons while open regions correspond to the introns. Multiple arrowheads correspond to the diversity in the DNA region coding for the carboxyl-terminal arm of the mature protein.

greater conservation of the 3′ flanking DNA of Hc-B genes, while the variable distance of these sites from the 3′ border of the genes is consistent with the occurrence of small insertion/deletion events (21). As with the Hc-A genes, cases of identical 3′ restriction maps exist for the Hc-B genes. Additional restriction mapping (not shown) indicates that Hc-B.6 and Hc-B.9, and Hc-B.11 and Hc-B.13, have very similar 3′ flanking regions, but once again these similarities end abruptly only 2–3 kb beyond the 3′ end of the Hc-B genes.

Patchwork Patterns of DNA Homologies. In a previous report comparing DNA sequences of Hc-B.12 and Hc-B.13 (22), a gene conversion event was postulated to explain a segment of decreased sequence homology in a part of the genes coding for the central conserved region of Hc protein. If many such localized regions (patches) of decreased or increased homology are present in the locus, they should be revealed by hybridization experiments utilizing short fragments of genes as probes. To directly test for such a patchwork of segmental homology, short DNA restriction fragments (150–300 bp) from gene pair 13 were hybridized to equal amounts of DNA from the 15 Hc gene pairs. The DNAs used in this experiment were the same subcloned genomic fragments shown in Fig. 1 and were bound to nitrocellulose filters by means of a slot blotter. The locations of the various restriction sites used to divide gene pair 13 into 10 fragments are shown in Fig. 3 *Upper*.

Several points about the hybridization conditions used in this experiment should be emphasized. All fragments were first hybridized to the blots at relatively low criteria (65°C in 0.6 M NaCl) to insure that no gene pairs contain a deletion of gene sequences. These hybridizations resulted in little or no discrimination between gene pairs, consistent with the extremely high renaturation criteria previously shown to be required (13, 23). An initial estimate of suitably stringent criteria for each fragment, which would allow discrimination between gene pairs, was based upon its length and G+C content (the G+C content of the noncoding regions of Hc gene pairs 12 and 13 is 30–40% and as high as 60–70% within the coding region). The results from these initial hybridizations were then used to adjust the criteria upward or downward to permit maximum discrimination between pair 13 and the other gene pairs. Criteria adjustments were made by keeping all conditions identical except the renaturation temperature (range 72–82°C). The exact temperature used for each fragment is indicated in the legend to Fig. 3. It is important to keep in mind that the hybridization criterion of each fragment in the figure was adapted to maximize the difference between gene pairs. Thus, this figure can be used to directly compare relative homologies within a column but not between columns. Finally, a similar series of hybridization experiments was performed in the presence of 50% formamide. These experiments consistently gave less hybridization discrimination between the different Hc gene pairs and, thus, are not presented in this report.

The patchwork structure of the Hc gene pairs can be seen clearly in Fig. 3. Each gene segment from Hc-B.13 exhibited a unique pattern of hybridization with the other 14 gene pairs. For example; fragment c, corresponding to the COOH arm of Hc-A genes, hybridized most intensely to pairs 11 to 15; fragment e, corresponding to the central region of the Hc-A intron, hybridized most intensly to pairs 1, 4, 6, 7, and 8; and fragment g, corresponding to the 5′ flanking regions of the gene pair, hybridized most intensively to pairs 7, 8, 9, 11, and 12. In agreement with the location of conserved restriction sites (Fig. 2), the 3′ region of Hc-A.13 (fragment b) is homologous to pairs 11–15, while the 3′ region of Hc-B.13 (fragments j and k) is homologous to pair 11. These unique patterns of segmental homology suggest that the Hc genes have exchanged sequence information within the family.

DISCUSSION

We have presented data indicating that the Hc chorion gene pairs have an identical organization and a high overall sequence homology. Sequence variants in the locus are in a "patchwork" arrangement in that each gene pair contains a unique pattern of regions or "patches" that are highly homologous to a reference gene pair (pair 13), interspersed with regions exhibiting less homology. What role might phenotypic selection play in the evolution of such a patchwork structure? Two possible selective mechanisms can be considered. First, the patches of homology between the various family members and pair 13 may represent regions that have not diverged from an ancestral sequence. This hypothesis requires that pair 13 has remained similar to the ancestral sequence while all other gene pairs have diverged in most regions. The second possible selection mechanism assumes that pair 13 did diverge from the ancestral sequence, but that in most instances there has been a similar evolution of several other gene pairs. This mechanism would require selective pressures that severely constrain the types of variation possible, while placing few if any constraints on which of these variants are combined in a gene. While it is certain that selective pressures serve to limit variation in the chorion locus, and probably do so more severely in some

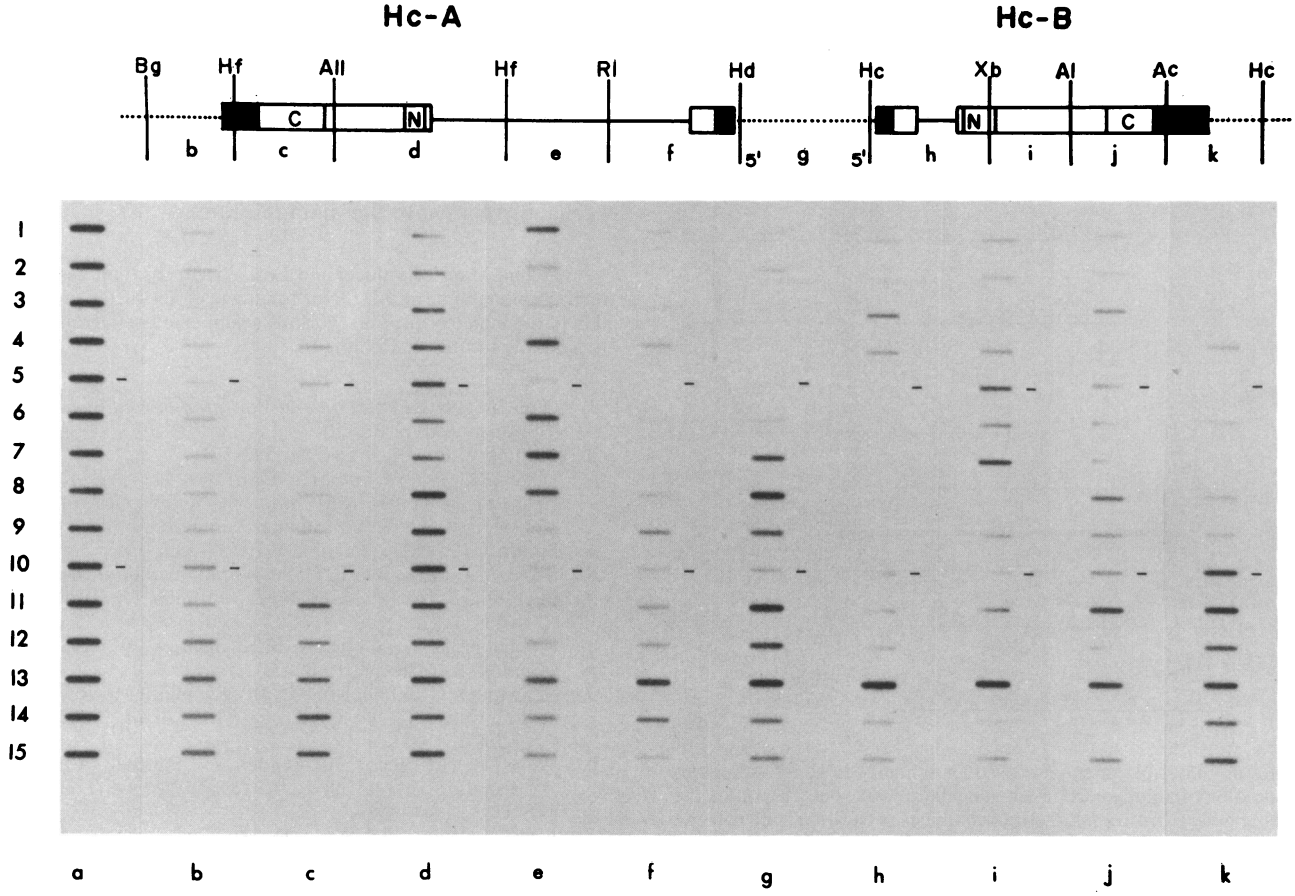

FIG. 3. Hybridization study demonstrating that short segments of gene pair 13 exhibit unique patches of very high sequence homology with other gene pairs. (*Upper*) Schematic diagram of gene pair 13. Dotted lines show the 5′ and 3′ flanking regions; solid lines, the introns; boxed regions, the exons; filled regions of the boxes, the untranslated regions of the transcript; N, the region encoding the amino-terminal arm of the mature protein; and C, the region encoding the carboxyl-terminal arm of the protein. Restriction endonuclease sites (vertical lines) used to isolate specific fragments were: A1, *Ava* I; Bg, *Bgl* II; Hf, *Hin*fI; All, *Ava* II; R1, *Eco*RI; Hd, *Hin*dIII; Hc, *Hin*cII; Xb, *Xba* I; and Ac, *Acc* I. Equal molar amounts of denatured plasmid DNA representing each gene pair (numbered at left) were applied to nitrocellulose paper by means of a slot blotter. The letter below each blot corresponds to the fragment of pair 13 used as probe (as shown in the diagram). The size of each fragment and the hybridization temperature used are as follows: a, (plasmid DNA control) 2.1 kb (82°C); b, 150 bp (72°C); c, 200 bp (82°C); d, 300 bp (82°C); e, 210 bp (72°C); f, 250 bp (78°C); g, 260 bp (78°C); h, 230 bp (82°C); i, 150 bp (82°C); j, 170 bp (82°C); and k, 180 bp (82°C).

regions of the genes than others, we are unable to account for the observed patchwork variation by means of selection.

A more likely explanation for the patchwork homologies involves the exchange of sequence information between genes. Since the locus is composed of highly similar gene pairs, tandemly repeated, two recombinational mechanisms could be responsible for the patchwork homologies: unequal crossover and gene conversion. The first mechanism would involve consecutive crossovers at different positions within the gene pairs, producing expansions and contractions of the locus. One characteristic outcome of these unequal crossovers would be the concerted evolution of the large amounts of DNA flanking the 3′ borders of the gene pairs. However, because these intergene regions appear to be rapidly diverging by the insertion/deletion of DNA segments (21), only the most recent crossovers leading to expansion of the locus might be identifiable as gene pairs with identical 3′ flanking regions. This identity would include one or more gene pairs depending upon the number of pairs offset between the chromosomes (see Fig. 4, the first crossover in *A* and *B*). Four examples of highly similar 3′ flanking restriction maps are present in the Hc locus (see Fig. 2). Significantly, in each of these cases the homology does not extend throughout the 3′ region to include the next gene pair, as would be expected from a simple expansion event. To explain the 3′ region identities detected in the locus as resulting from unequal crossovers, each of the most recent expansion events must be correlated with a contraction event—as depicted in Fig. 4, the second crossover in *A* and *B*. Obviously, alternative pathways to those presented in the figure can be envisaged. For example, a crossover between gene pairs could occur first, followed by one or more crossovers with the gene pairs. The important features in any proposed crossover pathway for the 3′ homologies is that the last event in all four cases is a contraction of the locus; and in each series, at least one crossover must occur between the gene pairs in regions of limited homology.

The second possible mechanism to account for the patchwork structure of the Hc locus, gene conversion, has been used frequently to explain patches of increased or decreased sequence homology in a variety of multigene systems (3, 7–11). Because gene conversion does not require the expansion and contraction of the locus, it can explain conveniently the occasional 3′ homologies as resulting from rare conversion events that extend beyond the 3′ border of the genes into regions of limited homology. By analogy to recent data on gene conversions in yeast (24, 25), the conversions in the Hc locus are probably mitotic or intrachromosomal meiotic exchanges, as they do not appear to involve (unequal) crossovers of flanking sequences at a significant frequency. If intrachromosomal, simple reciprocal exchanges (crossovers) would spread these variants between homologues. As

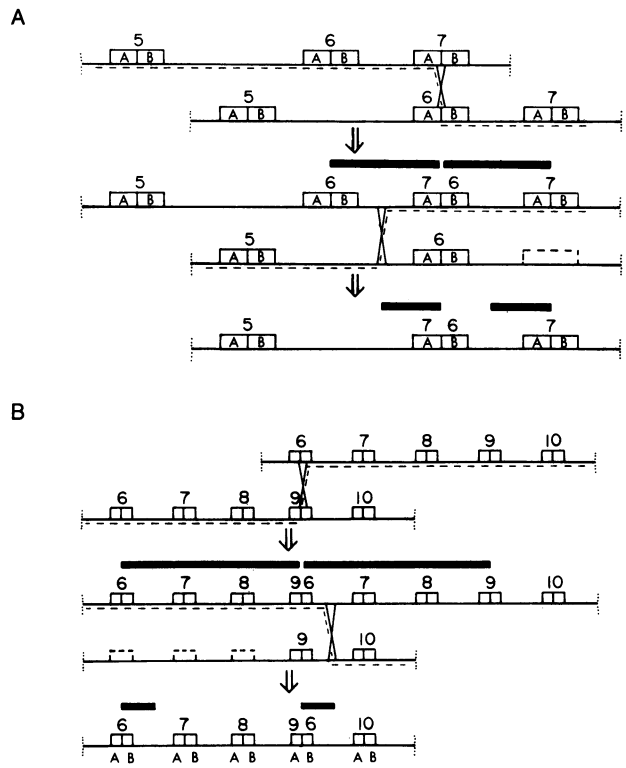

FIG. 4. Possible unequal crossover schemes to generate regions of homology flanking certain Hc gene pairs. Boxes corresponding to specific gene pairs are numbered. Distances separating gene pairs do not reflect the variable length of the DNA separating these gene pairs in the actual locus. Dotted lines correspond to the chromosome to be followed after each crossover. Solid bars correspond to duplicated regions created by the first crossover (expansion) and what is left of the duplication after the second crossover (contraction). (A) Scheme for the 3'-flanking homology found between Hc-A.6 and Hc-A.7. (B) Scheme for the 3'-flanking homology between Hc-B.6 and Hc-B.9.

in other higher eukaryotic multigene systems, the inability to conduct tetrad analysis of *Bombyx* means it is not possible to distinguish between gene conversion and closely spaced double-reciprocal crossovers among nonhomologous genes. However, we find very little evidence for single (unequal) crossover events and would expect that double crossovers are even less frequent.

We believe the Hc chorion locus can be viewed as an interesting intermediate between the ribosomal RNA loci, where sequence exchange homogenizes the family (5, 6), and the immunoglobulin or multiple histocompatibility loci, where sequence exchange may homogenize certain regions of some genes (subfamilies) but does not prevent extensive diversity within the larger family (3, 9–11). While the Hc locus must have originally expanded by unequal crossover events, the present structure of the locus is best explained by exchanges of sequence information, predominately by gene conversion mechanisms, resulting in the concerted evolution of the gene families. We plan to continue our analysis of the intergene regions, in an attempt to determine how the locus has expanded, and to sequence a number of additional gene pairs from throughout the locus to determine both the sizes of the converted patches and the frequency of sequence exchange relative to the mutation rate.

We thank Maurice Palucci and Lori Wright for their assistance in restriction mapping and Uzi Nur for suggestions during the preparation of the manuscript. This work was supported by a grant from the National Institutes of Health.

1. Ohta, T. (1980) *Evolution and Variation of Multigene Families* (Springer-Verlag, Berlin).
2. Baltimore, D. (1981) *Cell* **24,** 592–594.
3. Dover, G. (1982) *Nature (London)* **299,** 111–117.
4. Arnheim, N. (1983) in *Evolution of Genes and Proteins*, eds. Nei, M. & Kohn, R. K. (Sinauer, Sunderland, MA), pp. 38–61.
5. Smith, G. P. (1976) *Science* **191,** 528–535.
6. Szostak, J. W. & Wu, R. (1980) *Nature (London)* **284,** 426–430.
7. Slightom, J. L., Bechl, A. E. & Smithies, O. (1980) *Cell* **21,** 627–638.
8. Liebhaber, S. A., Goosens, M. & Kan, Y. W. (1981) *Nature (London)* **290,** 26–29.
9. Krawinkel, U., Zoebelein, G., Bruggemann, M., Radbruch, A. & Rajewsky, K. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 4997–5001.
10. Weiss, E. H., Mellor, A., Golden, L., Fahrner, A., Simpson, E., Hurst, J. & Flavell, R. A. (1983) *Nature (London)* **301,** 671–674.
11. Pease, L. R., Schulze, D. H., Pfaffenbach, G. M. & Nathenson, S. G. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 242–246.
12. Goldsmith, M. R. & Kafatos, F. C. (1984) *Annu. Rev. Genet.* **18,** 443–487.
13. Eickbush, T. H. & Kafatos, F. C. (1982) *Cell* **29,** 633–643.
14. Iatrou, K., Tsitilou, S. G. & Kafatos, F. C. (1982) *J. Mol. Biol.* **157,** 417–434.
15. Smith, H. O. & Brinstiel, M. L. (1976) *Nucleic Acids Res.* **3,** 2387–2398.
16. Iatrou, K. & Tsitilou, S. G. (1983) *EMBO J.* **2,** 1431–1440.
17. Iatrou, K., Tsitilou, S. G. & Kafatos, F. C. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 4452–4456.
18. Kafatos, F. C., Jones, C. W. & Efstratiadis, A. (1979) *Nucleic Acids Res.* **7,** 1541–1552.
19. Vieira, J. & Messing, J. (1982) *Gene* **19,** 259–268.
20. Rodakis, G. C. & Kafatos, F. C. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 3551–3555.
21. Jones, C. W. & Kafatos, F. C. (1980) *Cell* **22,** 855–867.
22. Rodakis, G. C., Lecanidou, R. & Eickbush, T. H. (1984) *J. Mol. Evol.* **20,** 265–273.
23. Beltz, G. A., Jacobs, K. A., Eickbush, T. H., Cherbas, P. T. & Kafatos, F. C. (1983) *Methods Enzymol.* **100,** 266–285.
24. Jackson, J. A. & Fink, G. R. (1981) *Nature (London)* **292,** 306–311.
25. Klein, H. L. (1984) *Nature (London)* **310,** 748–753.