A THEORY OF MODULAR EVOLUTION FOR BACTERIOPHAGES *

David Botstein

Department of Biology
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

INTRODUCTION

Over the past ten years, a great deal of evidence has been accumulated suggesting that a rather large and apparently diverse group of temperate bacteriophages are related in ways not easily accounted for by standard ideas of evolution along branching trees of linear descents. In this paper a different theory of evolution which better accounts for what is known about the relatedness of temperate bacteriophages is described. It seems possible that this theory, which envisions the joint evolution of sets of functionally and genetically interchangeable elements, applies not only to the temperate bacteriophages whose properties suggested it, but also to the evolution of viruses in general.

The essentials of the modular theory of bacteriophage evolution can be stated as follows:

1. The product of evolution is not a given virus but a family of interchangeable genetic elements (modules) each of which carries out a particular biological function. Each virus encountered in nature is a favorable combination of modules (one for each viral function) selected to work optimally individually and together to fill a particular niche. Exchange of a given module for another that has the same biological function (e.g., DNA replication) occurs by recombination among a population of different viruses related only by their similar modular construction. Viruses in the same interbreeding population can differ widely in any characteristic (including morphology and host range) since these are aspects of the function of individual modules; more significantly, viruses of the same family containing virtually identical modules can be found as parasites of hosts which themselves are only distantly related.

2. Evolution acts primarily not at the level of an intact virus, but at the level of individual functional units (modules). Selection upon these modules is for: a) good execution of function; b) retention of flanking homology which ensures proper placement in the virus genome by homologous recombination and thereby guarantees proper regulation; and c) functional compatibility with the maximum number of combinations of other functional units.

THE PARTICULAR CASE OF THE TEMPERATE BACTERIOPHAGES P22 AND λ

The family of bacteriophages that best illustrates this kind of evolution (and that, incidentally, led to the development of the modular hypothesis [1, 2] is one

* Supported by the National Institutes of Health (Grants GM18973 and GM21253) and the American Cancer Society (Grant VC245).

0077–8923/80/354–484 $1.75/0 © 1980, NYAS

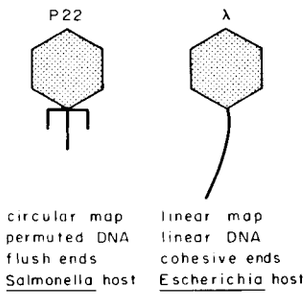

FIGURE 1. Schematic comparison of the temperate phage P22 and λ.

which contains both of the best-studied temperate bacteriophages: bacteriophage λ (whose normal host is *Escherichia coli*) and bacteriophage P22 (whose normal host is *Salmonella typhimurium*). These phages are morphologically completely unlike (FIGURE 1). Nevertheless, extensive genetic analysis of each led to the observation that their genetic maps were homologous: the order of genes (classified by their functions) is the same (FIGURE 2).[3, 4] The degree of similarity is particularly striking when one examines the arrangement of genetic sites at which transcription begins and regulation is exerted. In fact, the only

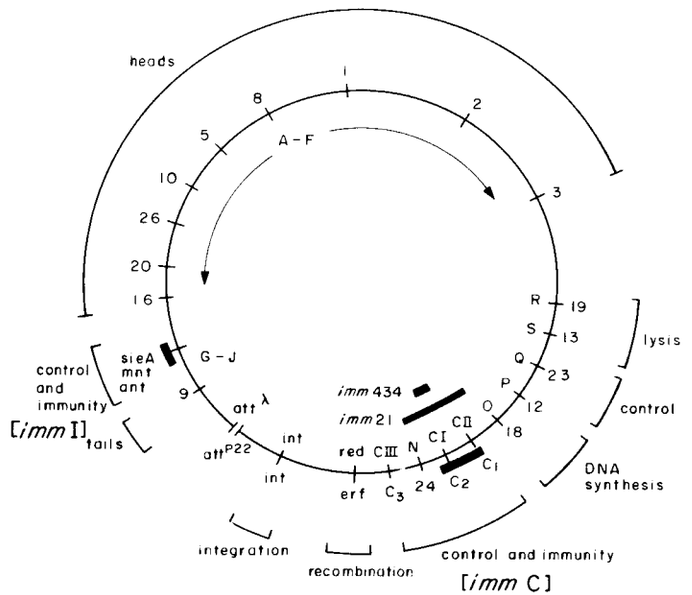

FIGURE 2. Homology between the order of functions on the genetic maps of P22 and λ. P22 genes are shown outside the circle, λ genes are shown inside. The basis for the alignment is given in Reference 3, from which the figure is adopted.

substantial difference between the two phages in genetic organization concerns the second (*immI*) immunity region, which P22 has and λ does not have.

The homology between the maps of phages P22 and λ, in spite of the gross dissimilarities between these phages in morphology, host range, DNA structure and replication, morphogenesis, and regulation (see Susskind and Botstein [4] for review), stimulated ultimately successful attempts to obtain hybrids between them.[5, 2] Extensive analysis of such hybrids [2, 6, 7] showed, surprisingly, that many different possible hybrids are possible, suggesting that homologous crossovers between them at virtually any position have the result of producing viable functional viruses.

Both phages P22 and λ had already been known to be members of families of bacteriophages within which most recombinants result in viable hybrids. Many viable hybrids of λ with other members of the "lambdoid" family defined by morphology, DNA structure (particularly the cohesive ends), and regulation had been studied in detail, particularly by DNA heteroduplex mapping.[8] On the basis of these studies, Hershey [1] proposed that the apparently modular construction of all of the phages in the lambdoid family was the agency of their evolution. The viability of many different hybrids between P22 and λ thus meant that the two families were in fact one, and suggested that modules from either subfamily were interchangeable with ones of similar function in the other.

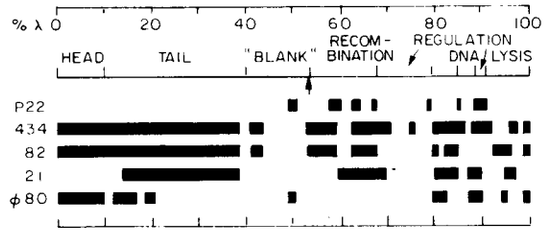

FIGURE 3. Summary of homology among lambdoid phages as determined by DNA heteroduplex analysis. The bars indicate position and extent of homology between the particular phage and λ itself. The figure is redrawn from Hershey,[1] with P22 data [2, 6, 9] added. The scale is in % physical length on λ.

This view of the relationship among the lambdoid coliphages (now taken to include the *Salmonella* phage P22 and its relatives) can best be visualized in the diagram shown in FIGURE 3 which is an adaptation of Hershey's [1] summary of the DNA heteroduplex analysis of the lambdoid bacteriophages. Regions of nonhomology are interspersed with regions (often quite short) of regions of homology. The homologous segments are frequently found in regions of no known function: it is as if the regions of function are relatively free to diverge but somehow are constrained to remain bounded by homologous segments. The relationship of phage P22 to λ is similar, reinforcing the idea that P22 is a member of the lambdoid group.

The notion that P22 and λ are members of the same interbreeding family suggests that identical modules might be found in both the P22 and λ subfamilies, despite the fact that the two subfamilies parasitize different species of bacteria. This prediction was fulfilled in the analysis of P22/λ hybrids, which

showed that the P22 *immC* region (which contains a repressor gene and its operators) was functionally identical in specificity as well as totally homologous (by DNA heteroduplex criteria) to the analogous region of the lambdoid phage 21.[2] This conclusion has since been further confirmed by direct nucleotide sequence analysis (A. Poteete, personal communication). Similarly, by both functional specificity and homology, the late regulator gene *23* of P22 was found to be identical to the late regulator gene λ of phage λ itself.[9]

A more striking example is the recent discovery by Friedman and colleagues (personal communication) that a segment of DNA which is homologous and functionally indistinguishable from P22s DNA replication module exists as a totally silent genetic element in *Escherichia coli* K12. In this case, the module seems not to be associated even with a complete virus. Similar rescue of functional modules from the bacterial genome had previously been observed [10] but these had never been related to modules also found in a naturally occurring lambdoid phage, let alone a phage of another bacterial species.

In summary, genetic and heteroduplex analysis of the P22 and λ subfamilies of bacteriophages suggested that these phages are modular in their construction. Similar modules exist in different species of bacteria; a module can be exchanged for one of similar function from any phage in the extended family.

THE NATURE OF THE MODULE

The basic requirements of a functional module are two: first, that it carry out its biological function effectively; second, that it retain its interchangeability, both in terms of its ability to be placed into other genomes and in terms of its functional compatibility with a variety of different combinations of modules carrying out the other essential biological functions.

In principle, the theory places no constraint upon the way in which the biological function is carried out. In fact, a strong prediction of the modular idea is that several different ways of carrying out each function will be found among the interbreeding family. For example, there are many ways in which one might imagine that the cell wall of an infected cell might be dissolved during lytic growth. In fact, the lambdoid phages have two known methods: phages 80 and P22 have a true lysozyme (which breaks a glycosidic bond) whereas phage λ uses an endopeptidase. The genes specifying these enzymes are completely nonhomologous and the proteins are in no way similar except in their biological function. Yet these genes are found in the same position on the genetic maps of the phages where they first were found and are all interchangeable with each other in hybrids.

TABLE 1 lists the modules one can simply define in a temperate bacteriophage of the lambdoid family, along with the number of easily distinguishable mechanisms used by at least one of the modules in the family. However, mechanism is not the only (and probably not the most important) variability in function of modules. Many of the functions of the phage exhibit a specificity (for example for the host surface, or host DNA replication proteins) that can vary among modules. As shown in TABLE 1, the number of distinguishable specificities is very large for some modules (such as the tail, which specifies host range). It may well be that the most important advantage of modular evolution (as opposed to linear descent) is the easy and continual access to a large variety of different specificities.

The Role of Homologous Recombination

The requirement that modules retain their means of interchange must be met by a method that will assure that the whole assembly of modules (i.e. the virus) be properly regulated regardless of which particular modules are present for each biological function. This assurance is intrinsic to the way in which these phages are regulated and by the method of interchange, which is homologous recombination. The regulation of the lambdoid bacteriophages (including, of course P22) is at the level of transcription, through a cascade of regulators, each of which controls transcription at sites immediately adjacent to itself. Thus the phage repressor regulates transcription in both directions at operators immediately flanking itself. The repressor and its operators are a module that can be interchanged since all the specificity-bearing elements are linked. The transcription leftward includes the positive regulator N (or its many other modular analogues, such as gene 24 of P22) which allows transcription to proceed into the DNA replication, recombination, and late control modules. The late control gene, when expressed, allows all the late modules to be transcribed, as a single operon. Thus a module interchange simply required that homology between (or at the termini of) functional segments be retained. Since recombination will exert specificity for homology, each module (provided it has the homology at its ends) automatically is placed in the correct position and orientation to be transcribed in proper sequence. Thus proper regulation is assured no matter which combinations of modules are present, and thus it is easy to account for the observation that virtually all hybrids among the temperate bacteriophages (even using modules that exist as cryptic

Table 1

Numbers of Specificities and Mechanisms Associated With Lambdoid Phage Modules

Function (Module)	λ Genes	Specificities	Mechanisms
DNA encapsulation	A–F	2	2
Tails	G–J	10+	2
Antigen conversion	b2	10+	2+
Integration	*att,int,xis*	10+	1
Recombination	*red,gam*	(3)	
Early control	N	5	1
Immunity	*cI*	10+	2
DNA replication	O,P,*ori*	3	2
Late control	Q	6	1
Lysis	S,R	—	2

The meaning of differences in specificity and function is given in the text. Estimates of the numbers of specificities and mechanisms are based on the literature and upon experiments in the author's laboratory. For example, the large number of tail specificities is based on the observation of very many host ranges; the two mechanisms represent the short tail (P22) and the long tail (λ) which are obviously mechanistically different.

host genes or that come from distantly related hosts) are viable and normally regulated.

The role of homology also accounts for the observation that there are relatively many instances of homology between phages at positions not known to code for any function (Figure 3). The conservation of these homologies is a natural consequence of the modular theory: these regions allow correct interchange of modules of like function, and are thus selected since a module that has lost its ability to be placed in the proper sequence has lost a crucial selective advantage.

In summary, the individual module is selected for retention of some common sequences at its ends which ensure ability to be recombined into the proper position in a variety of different phages, as well as for good execution of its biological function. An extension of this idea, of course, is the notion that modules also will be selected for functional compatibility with any other modules where joint function is required. This requirement is to some degree vitiated by the observation that interacting functions are, in these viruses, usually adjacent genetically and can, under some circumstances, be joined into a single module. This is clearly the case, for example, in the head assembly system: the P22 and λ systems, which are very different from each other, contain submodules [11] with respect to their own subfamilies; but with respect to each other the vast difference in mechanism requires that the entire head region be exchanged as a unit. As might be anticipated, no homology between the two head assembly regions has been observed (Botstein, unpublished observation).

Modular Evolution Accounts for the P22 Antirepressor System

One of the great puzzles in P22 biology has been the observation that the secondary immunity system (*immI*) [12–14] is entirely deletable with no obvious negative consequences to the normal lytic or lysogenic life cycles of the phage. If one accepts the modular theory, however, the advantage of association with an antirepressor is explicable. Susskind and Botstein [15] showed that P22 antirepressor inactivates the repressors of not only P22, but also of λ and some of its relatives. It is, in other words, very broad in its specificity for temperate phage repressors. Therefore a phage that carries an antirepressor and infects a cell in which one of its own modules fails to work has a selective advantage, because it will induce the resident prophages and elicit thereby the function of all the modules present in that cell. Induction of the prophages will also frequently result in replication of the prophage DNAs, increasing the probability of a module exchange, and thereby the creation of a new phage containing many of the old modules, which is now a fitter host for growth on this particular host.

In fact, the antirepressor advantage nicely illustrates the overall advantage of the modular system in the rapidity and flexibility of response to new environments. Clearly the availability of substitute modules of different specificity and/or mechanism but which have been evolved individually for good function will be an advantage even when a system of antirepressors is not present to increase the frequency of exchange. The module exchanges that have been observed in the laboratory have occurred at frequencies comparable to or

higher than the mutation rate. At such frequencies, module exchange seems to offer greater speed and flexibility than evolution by linear descent.

Conclusion

The modular theory of virus evolution has clear experimental support among the temperate bacteriophages of the enteric bacteria. However, there is also similar genetic and DNA heteroduplex evidence for such evolution among other families of bacteriophages: the virulent bacteriophages of the enterics comprise several families: the T-even group, the T3-T7 group (which has many members among different species of bacteria, including bacteria as widely divergent as *E. coli* and *Caulobacter crescentus*.[16] It nicely explains the diffusion of very similar homologous bacteriophages into hosts whose own DNAs have diverged very greatly from each other in nucleotide sequence. It also accounts for the rigorous maintenance of regulatory schemes while units of function (including regions coding for proteins) diverge more rapidly.

It should also be noted that the considerations that make modular evolution seem advantageous for bacteriophages apply equally well to viruses of higher organisms. Furthermore, the kinds of heteroduplex similarity observed among animal viruses are reminiscent of what is found for bacteriophages. Viruses found in widely divergent hosts show much greater similarity than would be expected; quite possibly animal viruses also evolve as a population of interchangeable modules.

References

1. Hershey, S. 1971. Carnegie Inst. Washington Yearb. **1970:**3–7.
2. Botstein, D. & I. Herskowitz. 1974. Properties of hybrids between the *Salmonella* phage P22 and coliphage. Nature **251:**584–589.
3. Botstein, D., R. K. Chan & C. H. Waddell. 1972. Genetics of bacteriophage P22. II. Gene order and gene function. Virology **49:**268–282.
4. Susskind, M. M. & D. Botstein. 1978. Molecular genetics of bacteriophage P22. Microbiol. Rev. **42:**385–413.
5. Gemski, P., Jr., L. S. Baron & N. Yamamoto. 1972. Formation of hybrids between coliphage λ and *Salmonella* phage P22 with a *Salmonella typhimurium* hybrid sensitive to these phages. Proc. Natl. Acad. Sci. USA **69:**3110–3114.
6. Hilliker, S. & D. Botstein. 1976. Specificity of genetic elements controlling regulation of early functions in temperature bacteriophages. J. Mol. Biol. **106:** 537–566.
7. Friedman, D. I. & R. Ponce-Campos. 1975. Differential effect of phage regulator functions on transcription from various promoters: evidence that the P22 gene *24* and the λ gene *N* products distinguish three classes of promoters. J. Mol. Biol. **98:**537–549.
8. Davidson, N. & W. Szybalski. 1971. Physical and chemical characteristics of lambda DNA. *In* The Bacteriophage Lambda. A. D. Hershey, Ed.: 45–82. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
9. Hilliker, S. 1974. Specificity of regulatory elements in temperature bacteriophages. Ph.D. Thesis. Massachusetts Institute of Technology. Cambridge, Mass.
10. Zissler, J., E. Signer & F. Schaefer. 1971. The role of recombination in growth of bacteriophage lambda. II. Inhibition of growth by prophage P2. *In* The Bacteriophage Lambda. A. D. Hershey, Ed.: 469–475. Cold Spring Harbor Laboratory. Cold Spring Harbor, N.Y.
11. Casjens, S. & R. Hendrix. 1974. Comments on the arrangement of the morphogenetic genes of bacteriophage lambda. J. Mol. Biol. **90:**20–23.
12. Botstein, D., K. K. Lew, V. M. Jarvik & C. A. Swanson, Jr. 1975. Role of antirepressor in the bipartite control of repression and immunity by bacteriophage P22. J. Mol. Biol. **91:**439–462.
13. Levine, M., S. Truesdell, T. Ramakrishnam & M. M. Bronson. 1975. Dual control of lysogeny by bacteriophage P22: an antirepressor locus and its controlling elements. J. Mol. Biol. **91:**421–438.
14. Botstein, D. & M. M. Susskind. 1974. Regulation of lysogens and the evolution of temperature bacterial viruses. *In* Mechanisms of Virus Disease. W. S. Robinson & C. F. Fox, Eds.: 363–384. W. A. Benjamin, Inc. New York, N.Y.
15. Susskind, M. M. & D. Botstein. 1975. Mechanism of action of *Salmonella* phage P22 antirepressor. J. Mol. Biol. **98:**413–424.
16. West, D., C. Langenaur & N. Agabian. 1976. Isolation and Characterization of *Caulobacter crescentus* bacteriophage/CD1. J. Virol. **17:**568–575.