

**DELETION MAPPING OF GENETIC "FINE STRUCTURE":
SUPPLEMENTING AD HOC PROBLEM SOLVING APPROACHES WITH
ALGORITHMS AND HEURISTICS**

by

John R. Jungck, Dept. of Biology, Beloit College, Beloit, WI
and

Vince Streif, Computer Center, University of Wisconsin-Eau Claire, Eau Claire, WI

INTRODUCTION

The transmission of a realistic analysis of the history of biology is exceedingly important in understanding the context of discoveries, and the personal, psychological, sociocultural, and political matrix in which ideas emerge and are propagated. Biology education typically duplicates the historically important process, which led to the discovery of the phenomena under consideration. However, as valuable as these lessons are, the extension of historically employed methods of problem solving may not be most appropriate in the education of students in contemporary laboratory practice.

We argue here that it is frequently important to re-examine tried and true methods of scientific problem solving to see if (1) more efficient and robust procedures can be constructed or borrowed from other fields, and, (2) more pedagogically sound procedures can be developed which provide greater clarity and chance of achieving success on the part of students. We believe that (1) and (2) are intimately connect-ed.

To this end, I (JRJ) have been examining (for the last 15 years) a number of problem solving techniques widely used in genetics¹ evolution, biochemistry and developmental biology (four courses which I teach frequently). The intent is to help students with techniques and processes which each year a significant number of students have considerable difficulty comprehending and using. We report here a case that has been nicely susceptible to reanalysis and has been successfully used in classrooms for nine years, namely the deletion mapping of genes. After going through the traditional ad hoc solution that is taught in basic genetics textbooks, we will describe two alternative approaches to solving such problems. We assert that both alternative approaches have several advantages compared to the traditional approach.

RELEVANCE OF NEW MATH TO BIOLOGY

Mathematics has not been stressed in the biology curriculum. The usual reasons given for this are that biology students have difficulty in studying mathematics, few biologists have & broad background in mathematics and mathematics is not fundamental to an adequate biological understanding. We counter that students are frequently not enamored with mathematics because they (1) feel mathematics is neither relevant to their own lives or biology, (2) are left out when pedagogy of the type: "It is intuitively obvious that" or "Anyone could show in a simple number of steps", are exercised on them, and (3) are selectively screened on the basis of gender and other personal characteristics as possessing mathematical prowess. All three of these subjects abound in the mathematics education literature. Less frequent in the literature is the simple observation

that (4) calculus was developed for and by physicists, not biologists and (5) that calculus is not always the most useful tool in college biology. Most of the mathematics presented in this paper has been developed in the last 20 years.

Most biology students that take first year college calculus do not use it in their first two years of college biology. In fact, many textbook authors bend over backwards to avoid the inclusion of mathematics in areas of biology where it is used routinely at the professional level. Seymour Papert has gone so far as to state:

Giving students the most powerful instruments of composition offers children an opportunity to become more like adults, indeed like advanced professionals, in their relationship to their intellectual products and to themselves. In doing so, it comes into head-on collision with many aspects of school whose effect, if not whose intention, is to infantilize "the child."

(Schwartz, 1983)

Furthermore, our personal experience is that when we have introduced alternative methods to those presented in textbooks, students respond frequently that (1) they didn't know math was and is appropriate to biology, (2) the methods become the ones of choice as the complexity of the problems increase, and (3) they are frequently surprised that (as such is the case in this paper) different kinds of mathematics can be employed on a problem and all provide equivalent results.

PEDAGOGICAL RATIONALE

Elsewhere I (Jungck, 1985) have argued that teaching frequently leads to research questions. When students have trouble understanding a concept, to what degree does our (the teachers') familiarity with the shared paradigms of our discipline blind us from seeing the lack of cohesion or the presence of intuitive leaps within our methodology. I, therefore, see learning roadblocks as an incentive to explore alternative approaches to these problems.

The development of algorithms and heuristics to solve problems has a number of other merit:

(1) Thomson and Stewart (1985) note that the development of algorithms makes each step in obtaining a solution highly explicit. Therefore, the instruction of knowledge is not hidden; it is available to all for scrutiny.

(2) Algorithms can be employed diagnostically. There is a rich literature on students' conceptions of science which they bring to the classroom (cf: Driver, 1983; Siegler, 1983; Brumby, 1984; Fisher, 1985; Osborne & Bell, 1983; and Kinnear, 1985). Successful teaching may involve the ability of the teacher (i) to infer the kinds of rules that students are employing to interpret their experience and to solve presented problems, in order to (ii) challenge those beliefs with empirical counterexamples.

(3) Algorithms fail when faced with new problem domains or with steps of great computational complexity (e.g., NP-complete problems). This is an asset in education. Unsolved problems should be a motivating factor to students. If turn the crank algorithms do not exist for all problems, then imagination, intuition and industriousness of scientists can contribute to science. Secondly, since many NP-complete problems do not hinder all science from proceeding, we might infer that humans are processing information, recognizing patterns, and solving problems in ways unlike contemporary computer programs in their execution.

(4) Both humans and computers can employ heuristics which, while they do not have the same guarantee of results as algorithms, do provide a "rule of thumb" for a way to proceed and usually have known cases where they are likely to fail. Thus, the use of Judgment (with aesthetic and other values involved) in decision-making is an explicit opportunity to students.

(5) If satisfactory algorithms and heuristics can be articulated for solving particular kinds of problems, computer analogs can be tested for:

- (a) Efficiency
- (b) Speed
- (c) Robustness
- (d) Reliability
- (e) Memory requirements
- (f) Iterative processes

The development of working computer "solvers" is an empirical demonstration of the methods' abilities in solving the problems in a finite number of well-described operations. Also, such programs immediately become a labor saving device in the field or laboratory because once you are sure that your data is of a given problem-type, the computer cannot only store the data, but analyze it as well. Furthermore, if ambiguities or inconsistencies appear in the data, appropriate theorem checks can be employed to detect and highlight them to the scientist. Finally, students using such programs can not only get beyond the process of solving a problem, they can use experience with many sets of data to develop a more robust intuition about subtleties of differently posed problems and complex patterns in the context of specific problems.

GRAPH THEORY IN BIOLOGY

In 1968, Joshua Lederberg stated that he hoped the applications of graph theory he found "could make it (organic chemistry) much easier for students to learn basic principles and to solve vexatious problems of classifying chemical compounds so that computers could be more readily applied to retrieve chemical information. It may be a forerunner of similar mathematical simplifications that will be applied to chemical genetics...". In a series of earlier papers, we illustrated the utility of graph theory to several problems in genetics, such as analysis of pedigrees, determination of inbreeding coefficients (Bertman and Jungck, unpublished); systematization of the properties of the genetic code (Bertman and Jungck, 1978, 1979; Jungck, in press); protein and nucleic acid sequencing (Jungck, Dick & Dick, 1982). Graph theory is a non-numerical branch of modern mathematics considered part of topology, but also closely

related to algebra and matrix theory (Ore, 1963). It has already made substantial contributions to biology in the area of taxonomy (Estabrook, 1968; Sneath and Sokal, 1973; Fitch, 1977; Penny, 1982) and ecology (Cohen, 1978) (Also see a nice high school paper on graph theory by Chartrand and Will, 1980.) Graph theory has great instructional utility because of the diversity of applications which can be learned without concomitantly learning a tremendous amount of sophisticated computational procedures.

Subsequent to our earlier effort, we learned about Gilmore and Hoffman's (1964) solution to the "Benzer problem" as it was discussed in two general texts on graph theory (Roberts, 1976; Busacker and Saaty, 1965). Benzer (1962) presented a method for the rapid mapping of point mutations, without resorting to numerous three point crosses (the classical method; for pedagogy see Mertons, 1972), by employing overlapping deletions. The construction of topological maps of deletion mutants has become a common exercise in genetic problem books (e.g., Stansfield, 1969, p. 219; Kuspira and Walker, 1973, pp. 505-508) and genetic laboratory manuals (e.g., Snustad and Dean, 1971; Hudock, 1967). The purposes of this paper are to present two alternative instructional stratagems for teaching about Benzer's 5 fine structure analysis of the gene and classical complementation mapping in a formal axiomatic fashion and, secondly, to present an algorithm for computerized determination of solutions to such problems from raw data. One benefit of the algorithm, besides allowing automatic reduction of vast amounts of data, is that it further reduces the solution to a variety of genetics problems to a series of elementary steps which can be easily understood by the neophyte geneticist.

TRADITIONAL AD HOC SOLUTION

Rather than discussing Benzer's problem abstractly, let us examine a concrete example and then compare the solutions of Benzer (1959) and Gilmore and Hoffman (1964). Readers unfamiliar with the genetic mapping experiments are referred to Benzer (1961). Generally, we can explain these experiments by indicating that mutations in a chromosome which overlap will not be capable of recombining to form a prototroph while nonoverlapping deletion mutants can recombine to form prototrophic recombinants (+) or the absence of prototrophic recombinants (0). Stansfield (1969) presents the following data (p.213) from such an experiment:

	1	2	3	4	5	6	7
1	0	+	0	0	+	0	0
2		0	0	0	+	+	0
3			0	0	+	+	0
4				0	+	0	0
5					0	0	0
6						0	0
7							0

Table I. Data from a "Benzer experiment"

The matrix elements could be completely filled in; however, because of symmetry around the diagonal, this process would only provide redundant information. Secondly, the diagonal is unnecessary information because no deletion mutant can recombine with itself to form a

prototroph.

Traditionally, to begin to solve this problem, Row One is analyzed for overlaps because it contains the most information of any of the rows. Thus since Deletion Number 1 overlaps with 3, 4, 6 and 7 (because they all have zeroes in the matrix) and does not overlap with Deletion Numbers 2 or 5, we can draw the following map of the deletions (Figure 1).

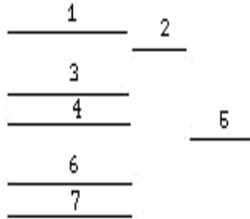


Fig 1. Topograph of row 1, Table 1.

In each successive step, this topographic map is modified by shortening or lengthening each of the lines (as required) representing a deletion or, in some cases, literally transposing a deletion line to the right or left.

If we proceed to the next steps, we see several such modifications. Thus, in analyzing Row 2, Deletion 5 has to be moved completely and Deletions 3, 4 and 7 have to be elongated (Figure 2).

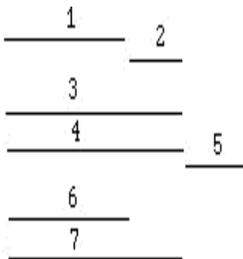


Fig. 2. Topograph modified by row (Table 1) data.

Conversely, shortening of Deletions 6 and 3 is necessary after examining Row 3 (Figure 3).

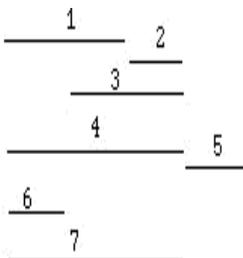


Fig. 3. Topograph modified by row 3 (Table 1) data.

Luckily, Row 4 does not require any modification of our topograph. On the other hand, the data in Row 5 requires substantial revision; we must now transpose Deletion 5 all the way to the left and extend both Deletions 6 and 7 to the left accordingly to now overlap Deletion 5.

Similar to the case of Row 4, neither the data in Row 6 or Row 7 (a diagonal element only) necessitates modification of our topograph. It should be noted in passing that the absolute length of any deletion line is arbitrary and that we do not have any reason for assigning the overall left-right orientation. This traditional ad hoc solution of "Benzer's problem" did not allow a student to process a given piece of information only once. In order to solve a problem this way, the student: must conscientiously retrace all her previous steps at each successive step. Thus, as above, this frequently involves multiple lengthening, erasing, and transposing different deletion lines. In addition, this usually requires a fabulous memory of what you've done before and/or constant rechecking. Furthermore, the arbitrary lengths of lines yield solutions that are not parsimonious. More intervals of the linear chromosome may be illustrated in a solution than are defensible based solely on the original matrix of data (Figure 4).

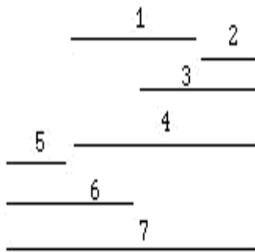


Fig. 4. Topograph modified by row 5 (Table 1) data.

A GRAPH THEORY ALGORITHM

Roberts' (1976) graph theory solution of the "Benzer's problem" avoids these potential sources of error. Let us begin by representing each mutant by a vertex in a graph (Figure 5a). Then we can process all the data in Table I row by row in a continuous series of steps, without any erasures or backtracking, by simply connecting any two vertices representing overlapping deletions with an edge; i.e., we construct the intersection graph equivalent to Table I (Figure 5b). Second, we construct the complement of the intersection graph by placing edges between all the vertices not connected in the intersection graph (Figure 5c).

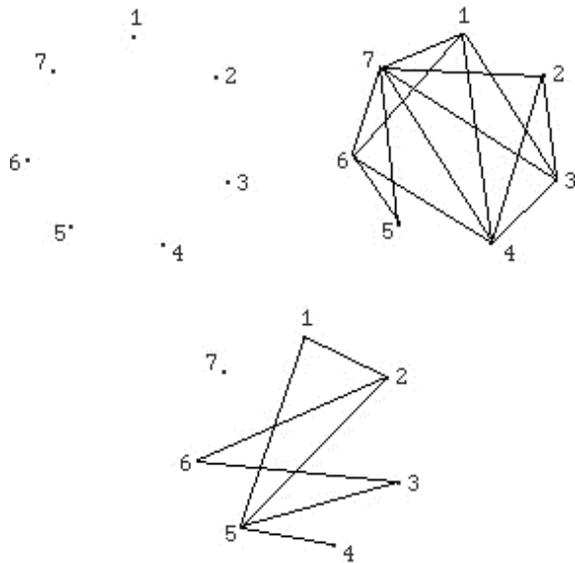


Fig. 5. (a) Seven vertices which represent the seven mutants employed in Table I. (b) intersection graph of Table I data on the vertices shown in (a). (c) Complementary graph of the intersection graph shown in (b).

Now we are at a point in a solution where we can check to see if the original intersection graph is, in fact, consistent with the expectations of a linear genetic map. First, the intersection graph should contain no Z4's (Figure 6) which are four vertices connected together into open squares with no diagonals.

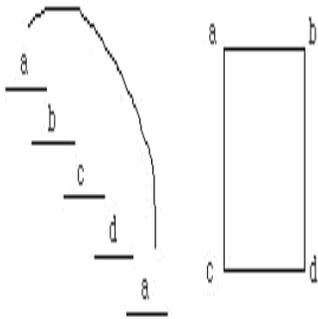


Fig. 6. The successive overlapping of the four fragments on the left are represented by the intersection graph, a Z4, on the right. Such representations imply a circular structure.

On checking Figure 5 (b), we see that it does not contain any Z4's and, therefore, Figure 5 (b) is a legitimate graph of an overlapping deletion experiment according to this criterion.

Secondly, the complement of the intersection graph (Figure 5c) should be capable of being made transitive. Most readers are familiar with the transitive relationship: if $a < b$ and $b < c$, then $a < c$. In graphs, a closed loop is transitive if $a \rightarrow b$ and $b \rightarrow c$, then a must $\rightarrow c$ also (Figure 7).

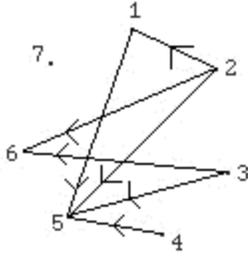


Fig. 7. Transitive orientation of the complementary graph illustrated in Fig 5c.

Since both criteria are shown to hold for the data in this example, we are thus assured it is worth proceeding to solve the problem. The next step is to find all the maximal cliques (synonyms: complete subgraphs, universal graphs, cliques) of the original intersection graph (Figure 5b). Cliques originally referred to exactly what the reader might first construe them to mean; namely, a group of n persons who all speak to one another (Festinger, 1949). Thus, a maximal clique is the largest subgraph in a graph in which all the vertices are connected by edges. There are four maximal cliques in Figure 5b (see Figure 8).

Thus, in this example there are four maximal cliques and each vertex was contained in at least one maximal clique. Next we order these maximal cliques in the same order as the direction between noncommon vertices of two maximal cliques in the transitive complement of the intersection graph (refer back to Figure 7). Thus, for example, Vertices 1 and 2 are not common to maximal cliques A and B and are connected $2 \rightarrow 1$ in Figure 7; therefore, we say the order of the maximal cliques is $B \rightarrow A$. By applying similar reasoning to all the maximal cliques, Figure 9 (a) can be obtained.

The penultimate step in the solution is to find the Hamiltonian path connecting all the maximal cliques. A Hamiltonian path is a path connecting all of the vertices, but traverses through any one vertex only once. In Figure 9 (a), it is very easy to see that B has all outgoing edges and thus must be a beginning point of the Hamiltonian path. Also, since D has all incoming edges, it must be the end of the Hamiltonian path. Therefore, the Hamiltonian path shown in Figure 9 (b) emerges easily.

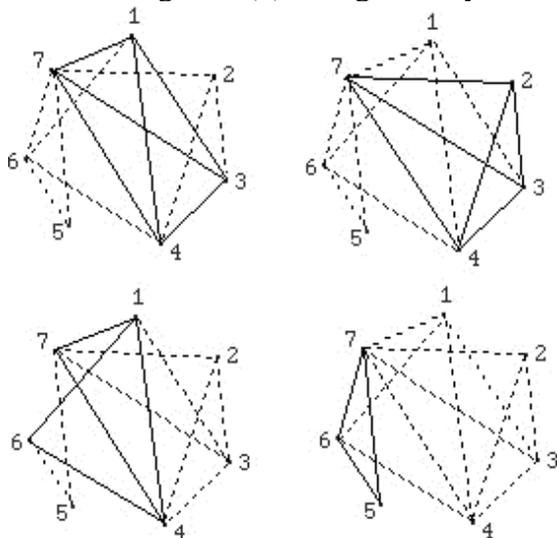


Fig. 8 The four maximal cliques contained in Fig. 5b. Each maximal clique is outlined in solid lines. (a) Maximal clique A contains vertices 1, 3, 4 and 7. (b) Maximal clique B contains vertices 2, 3, 4 and 7. (c) Maximal clique C contains vertices 1,4,6 and 7. (d) Maximal clique D contains vertices 5,6 and 7.

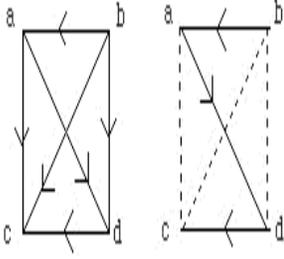


Fig. 9 (a) The order relationships between the maximal cliques shown in Fig. 8. (b) The Hamiltonian path in Fig. 9 (a) is shown in solid lines ($B \rightarrow A \rightarrow C \rightarrow D$).

Finally, we now are able to construct the interval graph equivalent to that shown earlier in Figure 4. We simply construct a line with the maximal cliques ordered in the same relative sequence as their Hamiltonian path and then each deletion will overlap those maximal cliques of which it is a vertex. Thus, Figure 10, produced in this way, is topologically equivalent to the topographic map of the deletions, seen in Figure 4, produced by the traditional solution method.

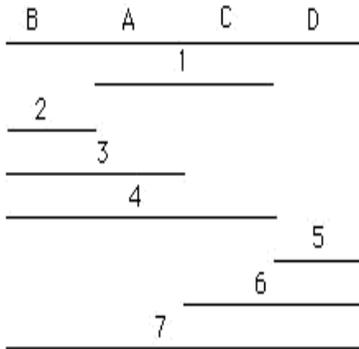


Fig. 10. Interval graph of the deletions (1 through 7) overlapping the four maximal cliques (A through D) ordered on a Hamiltonian path.

Inspection shows that the only difference between Figures 4 and 10 is the left-right orientation, which was arbitrary anyway, and the ambiguity of the lengths of lines in Figure 4. Figure 10 has four intervals of deleted regions of the chromosome and is a parsimonious solution. Although the graph theory solution to "Benzer's problem" seems longer than the classical solution, it has the three distinct advantages of (1) depending on a formal series of logical steps which can be axiomatized, of (2) being adaptable to automatic processing on a computer, and of (3) producing a parsimonious solution. To reiterate the steps involved in the graph theory solution of "Benzer's problem", for easy reference, the steps are laid out in Table II.

Table II. Steps in the graph theory solution of "Benzer's problem"

- I. Convert each deletion mutant into a vertex.
 2. Construct the intersection graph by placing an edge between each pair of vertices which represent overlapping deletions.
 3. Construct the complement of the intersection graph.
 4. Check for absence of Z4's in the intersection graph.
 5. Determine whether the complement of the intersection graph can be made transitive.
 6. Find all the maximal cliques in the intersection graph.
 7. Order these maximal cliques in the same way as in the transitive complementary graph.
 8. Find the Hamiltonian path of all the ordered maximal cliques.
 9. Construct the interval graph by assigning deletions to each interval of the line, which sequentially orders the maximal cliques, for all the cliques to which the deletion vertex belongs.

Thus, the algorithm is capable of processing the original recombination matrix data through each of these nine steps.

A HEURISTIC SOLUTION

However, let us reconsider the original problem (Table I) in a different way. In 1965, Shkurba developed a matrix manipulation to a canonical form which allowed an analysis of Gershenzon's "hypothesis that supplementary nutrition of *Drosophila* larvae with preparations of DNA caused mutations which affect whole sections of the chromosome." We can easily see that this biological problem is exactly equivalent to the "Benzer" problem. Therefore, let us consider Shkurba's solution because we contend that by considering the 3 kinds (ad hoc, algorithmic, and heuristic) of solutions to topological mapping of genes, we will offer students multiple ways of understanding problem solving. The first step in Shkurba's solution is to convert the matrix as displayed in Table I to the fully symmetrical form and to use blocked out cells in the matrix rather than zeroes and ones because it is easier to visualize when one has reached an appropriate canonical form of the matrix which he refers to as possessing the "basic property".

At this point, we will alter from our previous process by presenting Shkurba's answer first and then returning to an analysis of the process whereby he arrived at his conclusion. We will use the same example as before. First, Table I is easily completed to make a fully symmetrical matrix as seen in Figure 11.

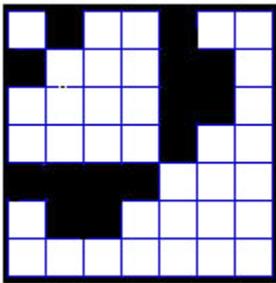


Figure 11. Symmetrical matrix for the data in Table 1.

Second, try to gather all the black squares about the diagonal in order (1) to minimize the moment of the matrix and (2) to make sure that no squares are unattached to other squares by less than a full side. Furthermore, Shkurba says that (3) all arrowheads should be pointing toward the same corner of the matrix. These three conditions, if they can be satisfied, represent the conditions for constructing an interval graph of the deletion mutants. Thus, if we rearrange rows and columns in Figure 11, we see that the conditions are approached if we interchange the order of 5 and 6 mutants (Figure 12).

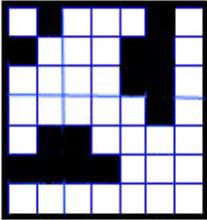


Figure 12. A rearrangement of Figure 11 which has mutants 5 and 6 interchanged in the sequence of their rows and columns.

Only one additional third step is required in this case to draw all the black squares together and satisfy the conditions for the “basic property”; this is simply achieved by inserting mutant 1 between mutants 4 and 6 in Figure 12 (see Figure 13).

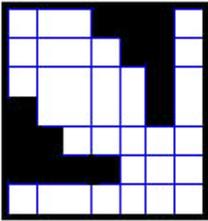


Figure 13. Matrix form of Table I which has Shkurba’s “basic property”; achieved by inserting mutant 1 between mutants 4 and 6 in Figure 12.

We can easily see that the diagonal can be represented as four overlapping blocks (Figure 14).

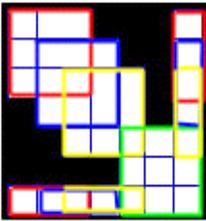


Figure 14. The four overlapping blocks contained in matrix form depicted in Figure 16.

Now it is a trivial problem to assign a mutant to an interval corresponding to the block (and only block or blocks) to which it belongs (see Figure 15).

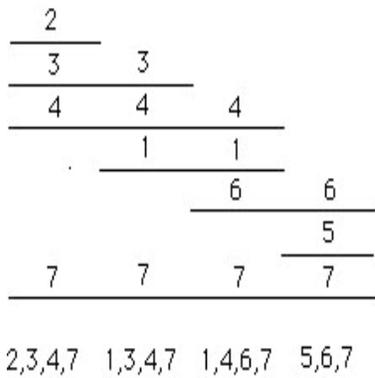


Figure 15. The interval graph determined by the Shkurba (1965) method on the same data employed throughout this paper. Each block in Figure 14 is given one interval on a line and the mutants are represented by lines extending over each interval (block) of which they are a member.

Figure 15 is isomorphic with Figure 10. The overlapping blocks in Figures 13 and 14 correspond exactly to "maximal cliques." Shkurba is able to present an algorithm employing the theory of partially ordered sets; however, we want to use his "basic form" to illustrate the power of heuristics. While heuristics are thought to be weaker methods than algorithms, for simple problems such as occur in undergraduate text-books, our students usually preferred the Shkurba method over the other two because by a little trial and error they could quickly convert a data matrix into the "basic form." Students have quite powerful pattern recognition experience and they find it easy to check their progress towards a "basic form" solution to such problems. However, if the matrix is extremely large, most students prefer to use the graph theoretic algorithm.

CONCLUSION

Biological problem solving can be taught in three different ways: (1) Historically based; i.e., by simply presenting the way a famous scientist solved a particular type of problem for the first time; (2) Algorithmically; i.e., by employing a mathematical technique which explicitly lays out each step to take in a solution path; and (3) Heuristically; i.e., by a rule of thumb or weak method which is likely to lead to a solution. We have illustrated each of these three approaches to a specific problem; namely, the deletion mapping of genes.

We believe that students have much to gain by learning the two new approaches described. First, they will learn powerful techniques which can easily be used to solve similar problems. For each, interval graphs can be used to sequence proteins (Jungck, Dick, and Dick, 1982), nucleic acids, complementation groups in genetics, food webs and niches, archaeological layers, preferences in psychological series or preferences in tastes, etc. By moving from a line to a circle, circular genetic maps, restriction maps, and boxity in food webs can be understood.

Second, by replacing an ad hoc technique with more general mathematical techniques, biologists will be better able to communicate with mathematicians. Especially with the availability of computers, it is easier to employ these powerful methods in laboratory analysis of

data. If data can be rapidly analyzed, particular genetic crosses can be identified which may violate the linear chromosome hypothesis if a Z4 arises. Also, when extremely large data sets are obtained, the computer program can be employed to check results.

Third, the conditions stated explicitly in an algorithm can be used to discuss what happens when they are violated. Thus, circular genetic maps, overlapping genes, and when one gene's intron is another gene's exon all violate the assumptions of the algorithm presented herein. Students can see that such violations are illustrative of new genetic phenomena and are not simply the fault of poor experimental technique.

We have developed two computer programs in BASIC on a Tektronix 4052 related to this paper. First, one program will generate "Benzer" problems with as many deletion mutants and intervals as the user specifies. This program offers many more opportunities for students to develop their skill than exist in any textbook or book of problems. Second, another program will draw the intersection graph and the complementary graph on a Tektronix 4662 plotter for any intersection matrix which is input. Both program lists are available upon request.

ACKNOWLEDGMENTS:

The teaching of Martha O' Kennan (previously Bertman) and the collaboration of B. Dennis Sustare were essential to the development of the ideas in this paper. Figures 1 through 10 were previously published in Jungck, Dick and Dick, 1982. Gregg and Amy Dick developed the Hamiltonian path technique described herein. Discussions with James Stewart and Angelo Collins at the Department of Curriculum and Instruction, University of Wisconsin-Madison, were important in developing the articulation of the ideas. Also, many, many students in genetics classes at Clarkson University and Beloit College were exceedingly helpful in pushing us to develop these ideas in a more understandable fashion. All of this help is greatly appreciated.

BIBLIOGRAPHY

Benzer, S., 1959, On the topology of genetic fine structure. Proc. Natl. Sci. USA 45, 1607-1620.

Benzer, S., 1961, On the topography of the genetic fine structure. Proc. Natl. Acad. Sci., USA 47, 403-426.

Benzer, S., 1962, The fine structure of the gene. Sci. Am. 206, 70-84.

Bertman, M. O. and Jungck, J. R. 1978, Some unresolved mathematical problems in genetic coding. Notices of the American Mathematical Society 25, A-174.

Bertman, M. O. and Jungck, J. R., 1979, Group graph of the genetic code. J. Heredity 70, 379-384.

Bertman, M. O. and Jungck, J. R., unpublished, Graph theory applications to genetics.

Brumby, M., 1979, Problems in learning the concept of natural selection, *J. Biological Education* 13, 119-122.

Busacker, and Saaty, 1965.

Chartrand, G. and Wall, C. E., 1980, Graph theory and the High School student. *School Science and Mathematics*, 479-485.

Cohen, J. E., 1978, *Food webs and niche space* (Princeton University Press, Princeton, New Jersey).

Driver, Rosalind, 1983, *The Pupil as Scientist?* The Open University Press, Milton Keynes, England.

Estabrook, C. F., 1968, A general Solution in partial orders for the Camin-Sokal model in phylogeny. *J. Theoretical Biology* 21, 421-438.

Festinger, L., 1949, The analysis of Sociograms using matrix algebra. *Hum. Relat.* 2.

Fisher, K. M., 1985, A misconception in biology: Amino acids and translation. *J. of Research in Science Teaching* 22, 53-62.

Fitch, W., 1977, On the problem of discovering the most parsimonious tree. *American Naturalist* 11, 223-257.

Gilmore, P. C. and Hoffman, A. J., 1964, A characterization of comparability graphs and of interval graphs. *Can. J. Math.* XVI, 539-548.

Hudock, G. A., 1967, *Experiments in Modern Genetics*. New York: Wiley.

Jungck, J. R., Dick, G. and Dick, A. G., 1982, Computer-assisted sequencing, interval graphs, and molecular evolution. *BioSystems* 15, 259-273.

Jungck, J. R., 1985, Mendel, Mendeleev and Me: Teaching's impact upon research. *American Biology Teacher* 47, 197-201.

Jungck, J. R., in press, Circular concatenated genetic code: impact of selection on overlapping codons. *BioSystems*.

Kinnear, J., 1983, Identification of misconceptions in genetics and the use of computer simulations in their correction. In H. Helm and J. Novak, eds., *Proceedings of the International Seminar, Misconceptions in Science and Mathematics*, Ithaca, NY: Department of Education, Cornell University, 84-92.

Kuspira, J. and Walker, G. W., 1973, *Genetics: Questions and Problems*. New York: McGraw-Hill Book Company.

- Lederberg, J., 1968, Topology of molecules, in: The mathematical sciences: A collection of essays, edited by the National Research Council's Committee on Support of Research in the Mathematical Sciences (COSRIMS) (M.I.T. Press, Cambridge, Massachusetts) pp. 37-51.
- Mertons, T., 1972, Investigation of three point linkage. *American Biology Teacher*.
- Ore, O., 1963, *Graphs and their uses*. New York: Random House.
- Osborne, Roger J. and Bell, Beverley F., 1983, Science teaching and children's views of the world. *European Journal of Science Education* 5 (1) : 1-17.
- Penny, D., 1982, Graph theory, evolutionary trees and classification. *Zoological J. of the Classification Society* 74, 277-292.
- Roberts, F. S., 1976, *Discrete mathematical models with applications to social, biological, and environmental problems* (Prentice-Hall, Engelwood Cliffs, NJ) pp. 111-140.
- Schwartz, L., 1983, Teaching writing in the age of the word processor and personal computers. *Educational Technology* (June), 33-35.
- Shkurba, V. V., 1965, Mathematical processing of a class of biochemical experiments. *Kibernetika* 1, 62-67.
- Siegler, Robert S., 1983, How knowledge influences learning. *American Scientist* 71 : 631-638 (November-December).
- Sneath, P. H. A. and Sokal, R. R., 1973, *Numerical Taxonomy*. San Francisco: W.H. Freeman and Company.
- Snustad, D. P. and Dean, D. S., 1971, *Genetics experiments with bacterial viruses*. San Francisco: W. H. Freeman and Company.
- Stansfield, W. D., 1969, *Schaum's outline of theory and problems of genetics*. New York: Schaum's Outline Series, McGraw-Hill Book Company. (In his 1983 edition, the problem is on pages 347-348; seven intervals exist in his solution instead of the parsimonious four intervals shown herein.)
- Thomson, N. and Stewart, J., 1985, Secondary school genetics instruction: making problem solving explicit and meaningful. *J. of Biological Education* 19, 53-62.

Drawing inspiration from natural selection, genetic algorithms (GA) are a fascinating approach to solving search and optimization problems. While much has been written about GA (see: [here](#) and [here](#)) Building our genetic algorithm. While each part of our GA is built from scratch, we'll use a few standard packages to make things easier: `import numpy as np, random, operator, pandas as pd, matplotlib.pyplot as plt.`