A Graph-Theoretic Approach to Comparing and Integrating Genetic, Physical and Sequence-Based Maps

Immanuel V. Yap,* David Schneider,† Jon Kleinberg,‡ David Matthews,§ Samuel Cartinhour† and Susan R. McCouch*,[1]

**Department of Plant Breeding, Cornell University, Ithaca, New York 14853, †United States Department of Agriculture-Agricultural Research Service, Center for Agricultural Bioinformatics, Cornell Theory Center, Ithaca, New York 14853, ‡Department of Computer Science, Cornell University, Ithaca, New York 14853 and §United States Department of Agriculture-Agricultural Research Service, Cornell University, Ithaca, New York 14853*

Manuscript received March 19, 2002
Accepted for publication August 20, 2003

ABSTRACT

For many species, multiple maps are available, often constructed independently by different research groups using different sets of markers and different source material. Integration of these maps provides a higher density of markers and greater genome coverage than is possible using a single study. In this article, we describe a novel approach to comparing and integrating maps by using abstract graphs. A map is modeled as a directed graph in which nodes represent mapped markers and edges define the order of adjacent markers. Independently constructed graphs representing corresponding maps from different studies are merged on the basis of their common loci. Absence of a path between two nodes indicates that their order is undetermined. A cycle indicates inconsistency among the mapping studies with regard to the order of the loci involved. The integrated graph thus produced represents a complete picture of all of the mapping studies that comprise it, including all of the ambiguities and inconsistencies among them. The objective of this representation is to guide additional research aimed at interpreting these ambiguities and inconsistencies in locus order rather than presenting a "consensus order" that ignores these problems.

FOR any given species, multiple genetic, physical, and sequence-based maps may be available. Individual maps often have been constructed independently by different groups, for their own purposes, on the basis of diverse mapping populations or source material. Each map may contain novel markers and may segregate for novel phenotypes, providing unique and valuable genetic information. It is possible to compare and integrate these different maps, as long as a common subset of markers has been used among the different mapping studies. Map integration frequently allows for greater genome coverage and higher-order integration of data across domains than is possible using a single set of markers or one particular population. Markers and mapped genes or quantitative trait loci (QTL) that could not be mapped in one study may be placed on the basis of their relative positions in another study. Comparison of individual maps also allows one to validate or challenge marker order between different studies. Integrating these independent maps presents a challenge to geneticists because of the inherent inconsistencies and ambiguities embodied by each map.

No standard procedure for map integration has been generally agreed upon. Four main approaches to integration are currently described in the literature. Liu (1998, Chap. 11) describes three of these approaches, to which we can add a fourth approach taken by the Genome Database (GDB). The simplest approach is to visually align different maps on the basis of common markers. This visual approach was used to create a "consensus map" of the homeologous groups of wheat (Nelson *et al.* 1995a,b,c; Van Deynze *et al.* 1995; Marino *et al.* 1996). An additional step was taken by Kianian and Quiros (1992), who computed the average linkage distance from the various map studies to create a "composite map" for *Brassica oleracea*. A second approach was used by Beavis and Grant (1991), who pooled all of the marker data from different mapping populations of maize with similar size and structure and generated a "pooled map" using MAPMAKER (Lander *et al.* 1987; Lincoln *et al.* 1993). The third approach is to use software such as JoinMap (Stam 1993; Stam and Ooijen 1995), which weights for population structure and size. This technique was used, for example, by Sewell *et al.* (1999) to integrate two linkage maps for loblolly pine and by Tani *et al.* (2003) to produce a consensus map for sugi (*Cryptomeria japonica*) on the basis of two pedigrees. The GDB takes a fourth approach to constructing a "comprehensive map" for the human genome. A specific map is designated as the primary or standard map, and then additional maps are successively projected onto the standard.

The end result of a map integration exercise is typi-

[1]*Corresponding author:* Plant Breeding Department, 162 Emerson Hall, Cornell University, Ithaca, NY 14853-1901.
E-mail: SRM4@cornell.edu

cally a single map in which all of the markers used in the various mapping studies have been placed with respect to one another. This is often a subjective and time-consuming process. More importantly, because of the manner in which an integrated map is constructed or presented, information about inconsistencies and ambiguities in locus order between different map studies becomes hidden or lost. Presenting a simple integrated map may thus contribute to a false sense of accuracy about locus order and position.

The branch of mathematics that is used to model discrete objects, sets, and the relationships between them is called *graph theory*. In the biological realm, it has been applied to questions of locus order on genetic, physical, and sequence-based maps. Benzer's (1959) work to map point mutations in a gene has been cited by Golumbic and Shamir (1992) as one of the motivations for the study of a particular type of graph known as interval graphs. Interval graphs have since been used in physical mapping to solve ordering problems in yeast artificial chromosome contig assembly (Harley *et al.* 1996, 1999; Randall 1997; Fasulo *et al.* 1999), radiation hybrid mapping (Ben-Dor and Chor 1997; Slonim *et al.* 1997), and genomic sequence assembly (Idury and Waterman 1995; Myers 1995). The problem of determining the order of genetic markers along a linkage group for genetic mapping may be modeled as a special case of the traveling salesman problem, which is a classic problem in graph theory (Lander and Green 1987; Liu 1998, Chap. 9). These applications of graph theory all deal with the *de novo* ordering of markers within a single mapping population or set of experiments.

We take the next step and show how graph theory may be used to compare locus orders between multiple maps from different mapping studies. As a starting point, each mapping study is understood to represent a highest-probability locus order rather than the "correct" locus order. Maps are then modeled as abstract graphs. This emphasizes order of markers rather than the distance between them, as genetic distance is not linearly correlated with physical distance and, moreover, is poorly comparable between mapping populations. Marker order, on the other hand, is meaningful across genetic, physical, and sequence-based maps. The primary purpose of integration is to highlight similarities in marker order, derive additional inferences about marker order, and expose ambiguities and inconsistencies in marker order between mapping studies. It is possible to formulate this process algorithmically using graphs and therefore to automate the method. This is a novel approach to comparing and integrating maps in an objective, reproducible, and efficient manner. It will be of maximum utility in cases where the raw mapping data used to infer marker order are not available or when different types of maps (*i.e.*, genetic, physical, and sequence based) are being compared.

THEORY AND ANALYSIS

We make use of well-known principles of graph theory and efficient algorithms that are easy to implement. These principles may be unfamiliar to most geneticists. They are therefore described and visualized below in ways that should be intuitive to most biologists.

Modeling a map as a directed graph: A *directed graph* is a collection of *nodes* and *edges*. A node may be represented by an ellipse, while an edge is represented as an arrow that connects a pair of nodes. Note that the visual representation of a graph is not at issue—nodes may be placed anywhere and edges may be as long or as short as desired.

Conventionally, a map is represented by a line segment upon which are placed the individual loci at positions proportional to their distance from each other. Figure 1A shows corresponding maps from two different mapping studies. (The different mapping studies are color coded blue and red to more easily differentiate them. On a monochrome copy, the different studies are represented by thin and thick arrows, respectively.) Figure 1B shows how a map may be modeled as a graph; this emphasizes the order of (but not the distance between) loci along the map. Each locus is modeled as a node, represented by an ellipse, and is connected to immediately adjacent loci by arrows (*i.e.*, edges). The direction to which the arrows point is defined by some convention of order (*e.g.*, from the short to the long arm of the chromosome). Thus, if locus *A* is mapped before locus *B*, this may be represented by a graph ($A \rightarrow B$).

A locus represents a marker mapped at a distinct position along a map. Hence, if a single marker exhibits several polymorphisms, each of which maps to a different position, then that one marker denotes several distinct loci, one for each mapped position. For example, in Figure 1, the marker H was mapped to two different loci, *H1* and *H2*, in the red map. These are also modeled as distinct nodes *H1* and *H2* in the red graph. Conversely, different markers may exhibit polymorphisms that cosegregate. They will therefore appear to have the same position. However, they are still represented as distinct loci, such as *C* and *D* in the blue map. Note how these loci are represented on a graph. The nodes that represent them connect to preceding nodes and subsequent nodes, but not to each other. This represents the uncertainty in relative order among these loci. An entire map may thus be modeled in this manner. The term *map graph* is used to denote a graph that models a map.

Map integration via graph merger: Modeling maps as graphs allows us to use graph operations to merge them, dissect inconsistencies, and represent ambiguities. We demonstrate the procedure with a series of illustrations beginning with Figure 1. In this hypothetical scenario, the same linkage group was mapped in two different studies using different but overlapping sets of markers.

Identifying anchor nodes: Integration of maps resulting from different mapping studies is possible when

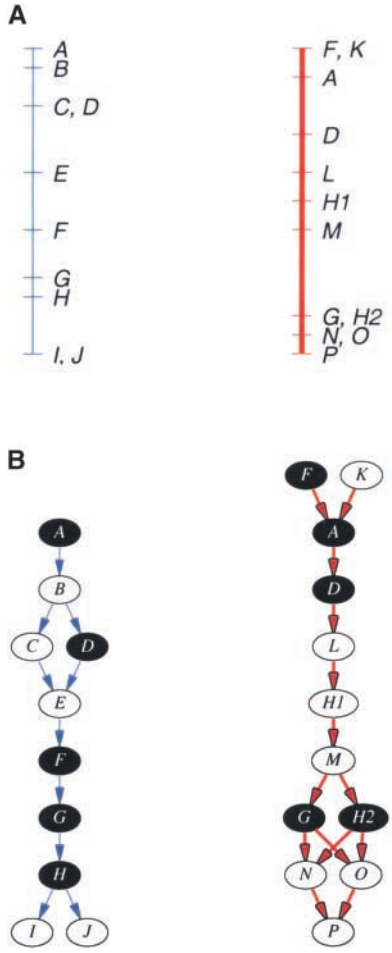

Figure 1.—Modeling maps as directed acyclic graphs and identification of anchor nodes (thin line, blue study; thick line, red study). (A) Corresponding maps from two different mapping studies. (B) The respective map graphs with anchor nodes shaded.

the different studies have used some markers in common, referred to as *anchor markers*. In a map graph, anchor markers are modeled by *anchor nodes*. These represent equivalent loci that were mapped using the same underlying marker. In Figure 2, markers A, D, F, G, and H were used in both studies and are therefore designated as anchors.

In the mapping literature, the concepts of *marker* and *locus* are often treated synonymously; indeed, locus names are usually derived from (if not identical to) the corresponding marker names. However, if a single marker yields multiple mappable polymorphisms, it will be represented as multiple loci. Typically, a letter is appended to the locus name to indicate multiply mapped polymorphisms derived from the same marker (*e.g.*, Temnykh *et al.* 2000). Care must be taken when assigning locus equivalences because locus-naming conventions are not always consistent between marker studies. For instance, a locus called *RG146*, mapped by Temnykh *et al.* (2001), was called *RG146B* by Cho *et al.* (1998); the same marker RG146 was used to detect the locus in both cases.

Figure 2A illustrates the problem of assigning locus equivalences across different map studies. Marker H, when used in the blue study, displayed only a single mappable polymorphism and therefore only a single locus *H*. Using the same marker in the red study, on the corresponding linkage group, two polymorphisms were detected and mapped and designated loci *H1* and *H2*. Matching *H* with *H1* leads to an inconsistency in the linkage graph while matching it with *H2* does not. Nodes *H* and *H2* are therefore designated as equivalent anchor nodes to be merged in the next step.

Merging anchor nodes: The key step in producing an integrated graph is merging the separate map graphs on the basis of their anchor nodes. Figure 2B shows the integrated graph resulting from the linkage graphs depicted in Figure 1B. The edges have been drawn in different colors to enable visual differentiation of the different mapping studies.

Intuitively, this can be accomplished by first drawing the linkage graph for one study and then iteratively adding in the connections for subsequent studies. Any appropriately labeled nodes that are already on the graph are used, with new nodes being added as necessary. New edges are added as in the original graph. Note that edges from the separate studies are kept distinct from each other. That is, if two studies each define (A → B), the integrated graph will merge nodes A and B and have two edges (A ⇉ B).

Mathematically, this operation is performed as a union of the node sets of the separate linkage graphs. However, to keep the edges representing different map studies distinct, only nodes are merged and not edges. Because the merger is performed as a set operation, it occurs simultaneously across all linkage graphs. There is, therefore, no bias for any particular study. At this point, the integrated graph represents a complete picture of all of the mapping studies that comprise it; unlike any of the four existing approaches outlined previously, our method has neither lost nor hidden any information about ambiguities and inconsistencies.

Identifying inconsistencies: Cycles in the integrated graph indicate an inconsistency in locus order. Intuitively, consider a map that specifies the locus order $X \rightarrow Y$ while the equivalent map from another mapping study specifies the opposite order $Y \rightarrow X$. The resulting integrated graph would show arrows that point in opposite directions $X \rightleftarrows Y$, *i.e.*, a cycle.

Examining Figure 2A, we see that the blue study specifies that *A* precedes *D*, which precedes *F*; therefore *A* precedes *F*. The red study, on the other hand, says that *F* precedes *A*. In other words, the two studies directly contradict each other. This contradiction is reflected in the graphs. If we refer to Figure 1B, we can see that there is a path $A \rightarrow F$ in the blue linkage graph. (A *path* between two nodes means that there is some sequence

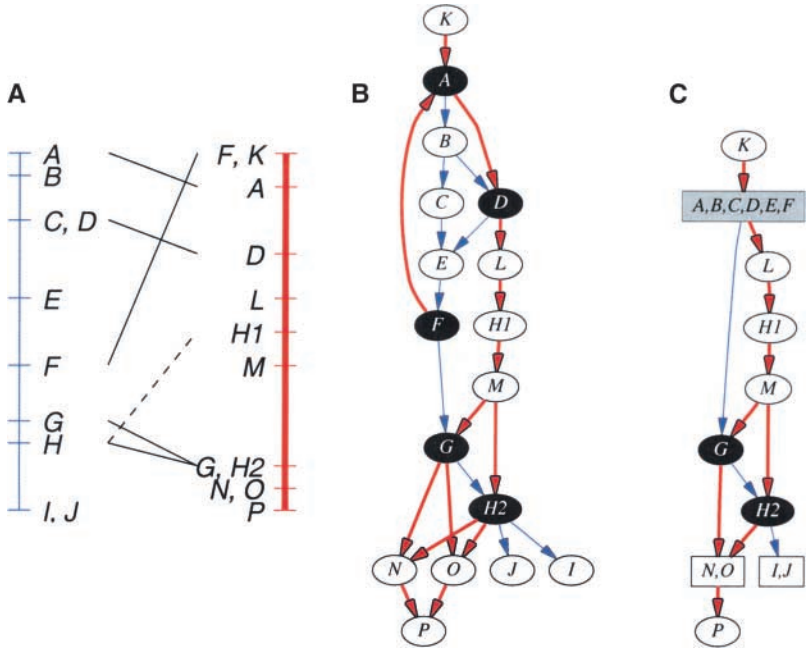

FIGURE 2.—Production of an integrated graph (thin line, blue study; thick line, red study). (A) Genetic maps from Figure 1A with common loci joined by solid black lines. The dashed line indicates that *H* could have been matched with *H1*, but a more consistent order is produced by matching it to *H2*. (B) The corresponding integrated graph generated by merging anchor nodes. (C) Integrated graph with condensed SCC and cosegregating nodes.

of nodes and edges that connects the two nodes.) In the red linkage graph, however, there is a path $F \rightarrow A$. The two graphs are inconsistent. When the two graphs are merged, this inconsistency in order is reflected as a cycle in the integrated graph. The nodes involved in the cycle {*A*, *B*, *C*, *D*, *E*, *F*} cannot be ordered relative to each other without yielding a contradiction (Figure 2B).

The concept of a cycle is generalized as a *strongly connected component* (SCC). For each pair of nodes *u* and *v* within an SCC, there is a path $u \rightarrow v$ and a return path $v \rightarrow u$. Thus an SCC defines a cycle or several interlocking cycles. A number of very efficient algorithms are able to find the SCCs in a graph (*e.g.*, CORMEN *et al.* 1990). The advantage of detecting inconsistencies as SCCs is that it is as easy to detect them in integrated graphs involving three or more map studies as it is to detect them in graphs that involve only two studies; the algorithm is affected by additional map studies only insofar as these studies add additional nodes to be processed.

Global ordering of nodes: Despite the presence of local inconsistencies in the maps being integrated, it is still possible to resolve the global order of markers that are not involved in the inconsistency. This is accomplished in the integrated graph by "condensing" the subgraph that comprises an SCC into a single node. This *condensed node* can be treated as a single discrete unit within the context of the encompassing integrated graph.

From Figure 2B, nodes {*A*, *B*, *C*, *D*, *E*, *F*} form an SCC that is condensed to form the resulting graph shown in Figure 2C. The condensed node can be ordered after node *K* and before node *L*. Additionally, the order of nodes {*N*, *O*} cannot be resolved relative to each other in the red study (*i.e.*, they are cosegregating). These nodes both have similar topology, in that they all come after the same particular node (or set of nodes) and they all come before another particular node. Hence, they may also be represented by a single condensed node. Similarly, nodes {*I*, *J*} cosegregate in the blue study and can therefore be represented by a single condensed node. However, *P* can be ordered after {*N*, *O*}. We can therefore deduce a global ordering of markers as

$$(K \rightarrow \{A, B, C, D, E, F\} \rightarrow L \rightarrow H1 \rightarrow M \rightarrow G \rightarrow H2 \rightarrow \{\{I, J\}, (\{N, O\} \rightarrow P)\}).$$

Note that this string representation is similar to the notation used by MAPMAKER. Arrows define order of elements. If elements are unordered, they are separated by commas. Parentheses are used to group together ordered elements; curly braces are used to group elements with undefined order. Although suitable within text for discussing relatively simple graphs or sections of graphs, this rendering quickly becomes cumbersome for more complex graphs.

Dissecting inconsistencies: Even within an SCC, it may still be possible to resolve the order of certain elements, even though the SCC as a whole indicates an inconsistency. One way to do this is by determining which set of edges would, if removed, eliminate the cycle. Finding the smallest such set may not always have a biological explanation, but it would be the most parsimonious solution. Consider the SCC discovered previously (Figure 3A). If we removed the edge ($F \rightarrow A$), the resulting graph would be acyclic (Figure 3B). We could also remove ($E \rightarrow F$) to break the cycle (Figure 3C). Either edge would be an acceptable solution.

The smallest set of edges that may be removed to leave an acyclic graph is called the *minimum feedback edge set.* The problem of finding such a set is NP-complete and, hence, no algorithms are known that can solve this efficiently (GAREY and JOHNSON 1979). However, integrated graphs that represent genetic maps tend to be

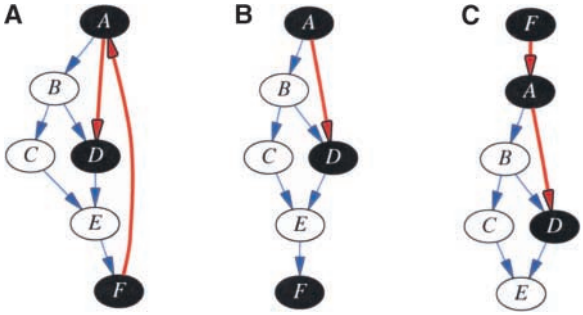

FIGURE 3.—Feedback edge sets (thin line, blue study; thick line, red study). (A) The SCC from Figure 2B. (B) Resulting graph when ($F \rightarrow A$) is removed. (C) Resulting graph when ($E \rightarrow F$) is removed.

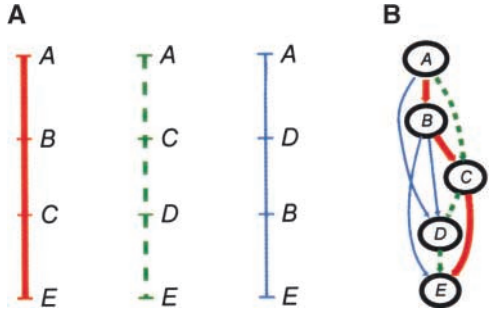

FIGURE 4.—Simultaneous comparison of (A) three maps by graph integration reveals a (B) three-way inconsistency involving *B*, *C*, and *D* (thin line, blue study; thick line, red study; dotted line, green study).

sparse, where each node is generally connected to a small set of other nodes—those representing the preceding loci and those representing subsequent loci. Further, an SCC generally occurs only in a limited part of the entire integrated graph. Hence we may be able to simply use the brute-force approach of testing every edge in an SCC to see whether its removal would leave an acyclic graph. This will find all feedback sets with a single edge. If no set has been found, then—given a sufficiently small SCC, or enough time and computing power—this brute-force algorithm could be extended to find all pairs, then sets of three, and so on.

Note that there may be more than one minimum feedback edge set. In Figure 3, two sets (each composed of a single edge) were found. Larger or more complex SCCs may well have many feedback edge sets.

Identifying ambiguities: In addition to inconsistencies, the graph approach may also highlight ambiguities in locus order between different map studies. This is a more subtle problem caused by insufficient information, such as when different studies use different subsets of markers or when markers cosegregate on a single genetic map. Returning to the blue study in Figure 1A, due perhaps to lack of recombination between *C* and *D*, the two loci map *genetically* to the same position. This does not mean, however, that they are at the same *physical* location on the chromosome. Hence, a more precise statement would be that both *C* and *D* are known to come after *B* and before *E*, but their order relative to each other is unknown. This is reflected in the blue map graph in Figure 1B, which shows that nodes *C* and *D* are both connected to the previous node *B* and the subsequent node *E*, but *C* and *D* are not connected to each other. The lack of a connection or path between *C* and *D* indicates the ambiguity in their order. The same ambiguity in ordering is true for {*I*, *J*} in the blue study and for {*F*, *K*}, {*G*, *H*2}, and {*N*, *O*} in the red study.

In an integrated graph, the same lack of a path between nodes can indicate ambiguity in locus order *between* different mapping studies. Examining Figure 2B, we can see that *I* and *J*, which were mapped only in the blue study, cannot be ordered relative to *N* and *O*, which were mapped only in the red study. Furthermore, although *P* can be definitively ordered after {*N*, *O*}, its order relative to {*I*, *J*} is still ambiguous.

Higher-order comparisons: The examples thus far have used only two maps for integration. Multiple maps may be integrated as easily as two using the graph approach. Indeed, certain inconsistencies may become evident only when integrating three or more maps. Consider Figure 4, where three different studies mapped some combination of four out of five loci. Simple pairwise comparisons of red to green, red to blue, and green to blue are not inconsistent. Only if we perform a simultaneous comparison of the three maps do we find that there is indeed an inconsistency involving all three studies. Note that none of the existing approaches to map integration would highlight this kind of inconsistency.

Graph representation: As a data structure, the integrated graph is the most accurate representation of the marker order specified by the various map studies that comprise it. Note that it is not necessary to visualize the graph to analyze it. Indeed, a major difficulty with graph visualization is that an integrated graph (indeed, any kind of abstract graph) may be rendered in a variety of ways. Two different graph representations that appear to be visually distinct may actually describe the same set of relationships. Hence, it may be difficult to compare visually two distinct graph representations to see whether they describe the same graph or different graphs. Graphs that are equivalent in this manner are referred to as *isomorphic*. Rather than simply comparing where nodes are positioned, one must examine the edges to determine whether they describe the same set of relationships.

Condensed nodes: Despite its limitations, visualizing a graph may still aid in understanding how nodes (*i.e.*, loci) are ordered relative to each other. If the maps being integrated have more than a moderate number of loci, their integrated graph may become visually complex. One way to reduce complexity without loss of information is to condense certain logically related sets of nodes. The condensed node itself forms a discrete unit

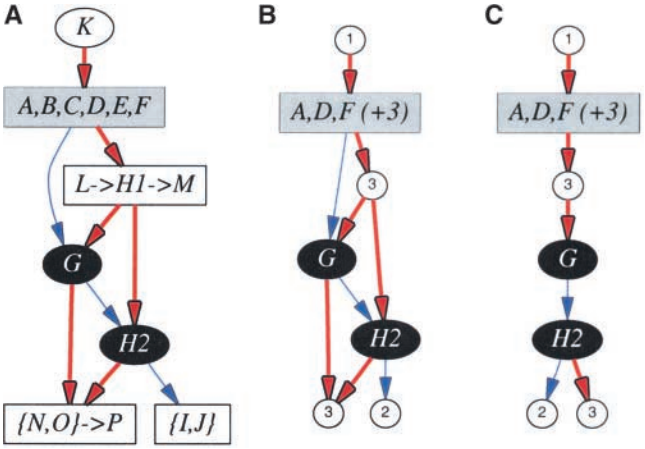

FIGURE 5.—Condensed nodes (thin line, blue study; thick line, red study). (A) Integrated graph with condensed nonanchor nodes. (B) Fully condensed integrated graph. (C) Graph with redundant edges removed.

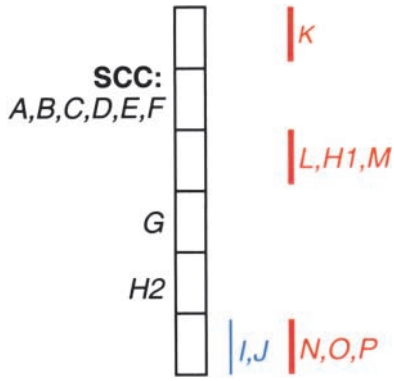

FIGURE 6.—Interval representation of the integrated graph from Figure 2B.

that may be treated like any other node in the graph. We have already shown in Figure 2C how an SCC as well as cosegregating markers may be condensed into a single discrete node. Certain other sets of nodes may also be condensed without altering the overall topology of the integrated graph.

A series of *adjacent* nodes (*i.e.*, a contiguous series of nodes) that are not anchors may also be condensed. These represent loci that were mapped in only one study and therefore add no additional information to locus ordering between studies. In Figure 5A, the sequential nonanchor nodes ($L \rightarrow H1 \rightarrow M$) are represented by a single condensed node. Further, node P can also be condensed into the previously condensed $\{N, O\}$ to form a single condensed node containing ($\{N, O\} \rightarrow P$).

Figure 5B shows an even more compact graph representation, in which each condensed and nonanchor node has been reduced in size and labeled with the number of loci represented by that node. This rendition has a particular advantage in that it hides the visual complexity caused by markers that were mapped in only a single study, including multiple cosegregating markers. It also emphasizes the order of common markers and highlights the inconsistent portion of the graph. Figure 5C shows a further simplification, in which redundant edges have been hidden. These are edges that can be safely removed because they contribute no additional information in terms of node ordering.

Interval representation: An integrated graph carries complete information about the relative locus ordering of the map studies that comprise it. However, this representation may become quite visually complex, making it difficult to follow the ordering of loci. We have explored the construction of a linear representation to reduce visual complexity. Nodes are represented as intervals, which represent the uncertainty in their position.

This approach cannot represent inconsistencies, *i.e.*, if an integrated graph has an SCC, then it is obviously not possible to portray the graph in linear form. Therefore, an SCC must be treated as a single condensed node, as in Figure 2C. To further simplify the problem, we can also condense nonanchor nodes and remove redundant edges, as in the condensed graph shown in Figure 5C.

The condensed graph intuitively suggests an interval representation like that in Figure 6. The anchor nodes (including the SCC) are placed as intervals, in the order specified by the condensed graph (SCC $\rightarrow G \rightarrow H2$). Intervals representing condensed nonanchor nodes can then be placed relative to the anchor intervals. From the blue study, node $\{I, J\}$ was ordered after $H2$, while B, C, and E are already incorporated into the SCC. From the red study, K lies before the SCC, $\{L, H1, M\}$ can be found between the SCC and $H2$, and $\{N, O, P\}$ can be found after G.

Ambiguous orders and misleading intervals: The problem in constructing an interval representation is to accurately portray the ambiguities present in the integrated graph. If such an interval representation exists, then the graph is known as an *interval order* (FISHBURN 1985). GOLUMBIC and SHAMIR (1992) indicate that determining whether a graph is an interval order is solvable in polynomial time; however, the graph may not be representable at all (*i.e.*, an integrated graph is not necessarily an interval order).

Consider the two maps shown in Figure 7A. The integrated graph, shown in Figure 7B, tells us that an interval representation would need to satisfy all of these constraints:

1. A precedes B precedes C precedes D.
2. W precedes X.
3. Y precedes Z.
4. X and Y both fall between B and C.
5. X overlaps Y (*i.e.*, the relative order of X and Y is unknown because there is no path between X and Y).
6. W overlaps B.
7. Z overlaps C.

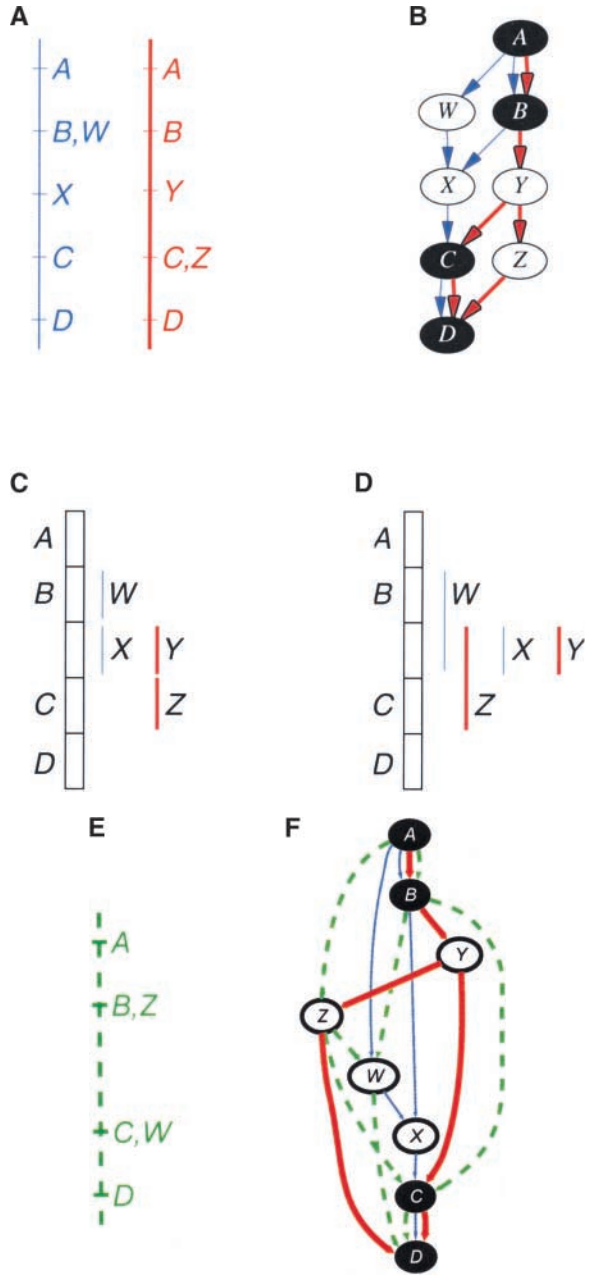

FIGURE 7.—Inaccurate interval representation (see text for details).

8. *W* overlaps *Y*.
9. *X* overlaps *Z*.
10. *W* overlaps *Z*.

These constraints break down into two types: whether an interval is known to precede another interval or whether two intervals overlap.

Figure 7C shows a reasonable interval representation for this scenario that is similar to the original maps. However, this fails to satisfy constraints 8, 9, and 10. If the interval representation is adjusted so that it follows these three constraints, as in Figure 7D, the intervals so drawn would contradict constraints 2 and 3. It may seem counterintuitive that *Z* could fall before *W*, but consider that a third study that specifies the order $Z \rightarrow W$ (Figure 7E) may be integrated without introducing an inconsistency (Figure 7F). In fact, there is no interval representation that can accurately portray all of these constraints. Only the integrated graph can represent ambiguities and inconsistencies accurately in all cases. Despite this, interval representation is sometimes useful in gaining a simplified picture of the order of nodes.

EXAMPLE

To illustrate the utility of the graph-theoretic approach, we use it to analyze maps of chromosome 1 of rice (*Oryza sativa* L.). Several dense molecular genetic maps have been developed by different groups from diverse populations. Among these are the interspecific *O. sativa* × *O. longistaminata* BS125/2/BS125/WLO2 population (CAUSSE *et al.* 1994; WILSON *et al.* 1999), the intersubspecific *indica* × *japonica* IR64/Azucena doubled-haploid population (TEMNYKH *et al.* 2001), and the *japonica* × *indica* Nipponbare/Kasalath population (HARUSHIMA *et al.* 1998). The first two maps were developed at Cornell University, mostly using simple sequence repeat and restriction fragment length polymorphism (RFLP) markers. They are referred to in this article as the SL01 and DH01 maps, respectively. The third was developed by the Japanese Rice Genome Project, using an independently developed set of RFLP markers. In this article, it is referred to as the JP98 map. The release of chromosome 1 sequence by SASAKI *et al.* (2002) affords us the opportunity to test certain inferences generated by comparing these three genetic maps.

The total numbers of mapped markers on the SL01, DH01, and JP98 maps were 129, 91, and 289, respectively. Certain markers have been mapped in common between the three maps, which provides a foundation for genetic map alignment and integration. The number of common (anchor) markers is relatively low, however. There are 17 common markers between the SL01 and DH01 maps, 15 between the SL01 and JP98 maps, and 2 between the DH01 and JP98 maps. Among all three maps, only 2 markers were mapped in common. Figure 8 shows the three maps aligned on the basis of anchor markers.

Inconsistencies: The integrated graph of chromosome 1 showed that there were two inconsistent intervals (Figure 9A). The first involved {*AMY1B*, *RG146*, *RZ276*} on DH01 and SL01 at 3.5 and 1.05 cM, respectively. These markers have been sequenced and are available as GenBank accessions M59350 (*AMY1B*), AQ074233 (*RG146*), AI978355 (*RZ276.F*), and AI978356 (*RZ276.R*). None of these have been used to anchor genomic sequence or physical map, but a BLAST search reveals significant hits on bacterial artificial chromosomes (BACs) or P1-derived artificial chromosomes (PACs) AP003275, AP00 3854, AP003444,

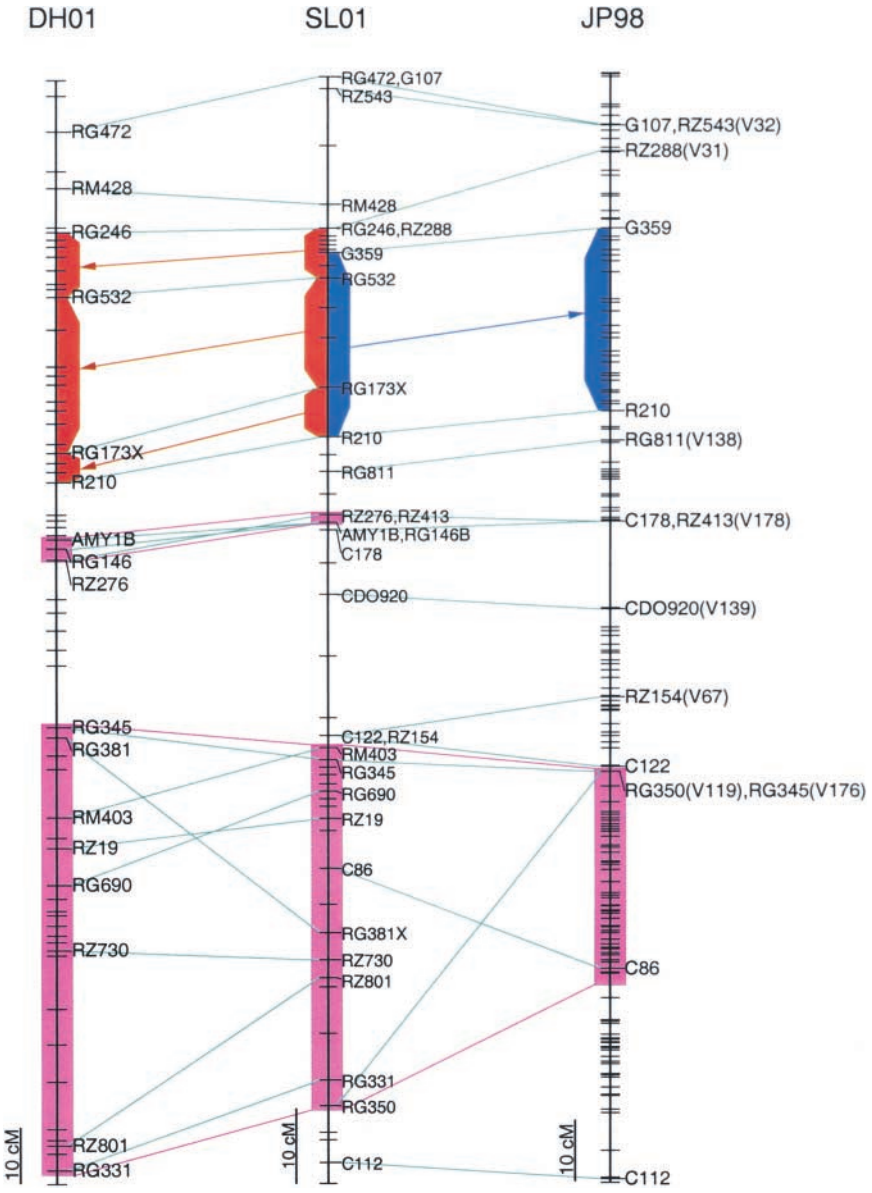

FIGURE 8.—Three genetic maps of chromosome 1 of rice aligned on the basis of anchor markers. Anchor markers are connected by cyan lines. Inconsistencies in locus order between maps are indicated by red-shaded regions. Ambiguous intervals between SL01 and DH01 are indicated by a magenta-shaded region in SL01 connected by an arrow to the corresponding region on DH01. Ambiguous intervals between SL01 and JP98 are similarly indicated by blue-shaded regions. Ambiguity between DH01 and JP98 is not indicated.

and AP003275, respectively; note that *AMY1B* and *RG276.R* hit the same PAC while *RG146* was not identified on that PAC (Table 1).

According to the chromosome 1 sequence published by the International Rice Genome Sequencing Project, these clones are in the order AP003275 → AP003444 → AP003854 (SASAKI *et al.* 2002), indicating that the markers should be ordered *AMY1B* → *RZ276* → *RG146*. This is a different order from that specified by either DH01 or SL01. Both *AMY1B* and *RG146* were mapped at a low LOD on SL01; thus this inconsistency was likely due to statistical error inherent in the genetic mapping process. Alternatively, there may be true differences in the genomes of the parental lines.

The second inconsistency involved almost the entire long arm of the chromosome and contained anchors {*RG690*, *RG331*, *RG350*, *C86*, *RZ801*, *RG381*, *RG345*, *RZ730*, *RZ19*, *RM403*}. Twelve minimum feedback edge sets (MFES) were found, each involving removal of three edges:

C86 → *RG381*;	*RG345* → *RG381*;	*RG690* → *RZ19*
C86 → *RG381*;	*RG345* → *RG381*;	*RZ19* → *RG690*
C86 → *RG381*;	*RG381* → *RM403*;	*RG690* → *RZ19*
C86 → *RG381*;	*RG381* → *RM403*;	*RZ19* → *RG690*
C86 → *RG381*;	*RM403* → *RG345*;	*RG690* → *RZ19*
C86 → *RG381*;	*RM403* → *RG345*;	*RZ19* → *RG690*
RG331 → *RG350*;	*RG381* → *RM403*;	*RG690* → *RZ19*
RG331 → *RG350*;	*RG381* → *RM403*;	*RZ19* → *RG690*
RG350 → *C86*;	*RG381* → *RM403*;	*RG690* → *RZ19*
RG350 → *C86*;	*RG381* → *RM403*;	*RZ19* → *RG690*
RG381 → *RM403*;	*RG381* → *RZ730*;	*RG690* → *RZ19*
RG381 → *RM403*;	*RG381* → *RZ730*;	*RZ19* → *RG690*

These sets were used to focus attention on intervals within the larger inconsistency. Half of these sets in-

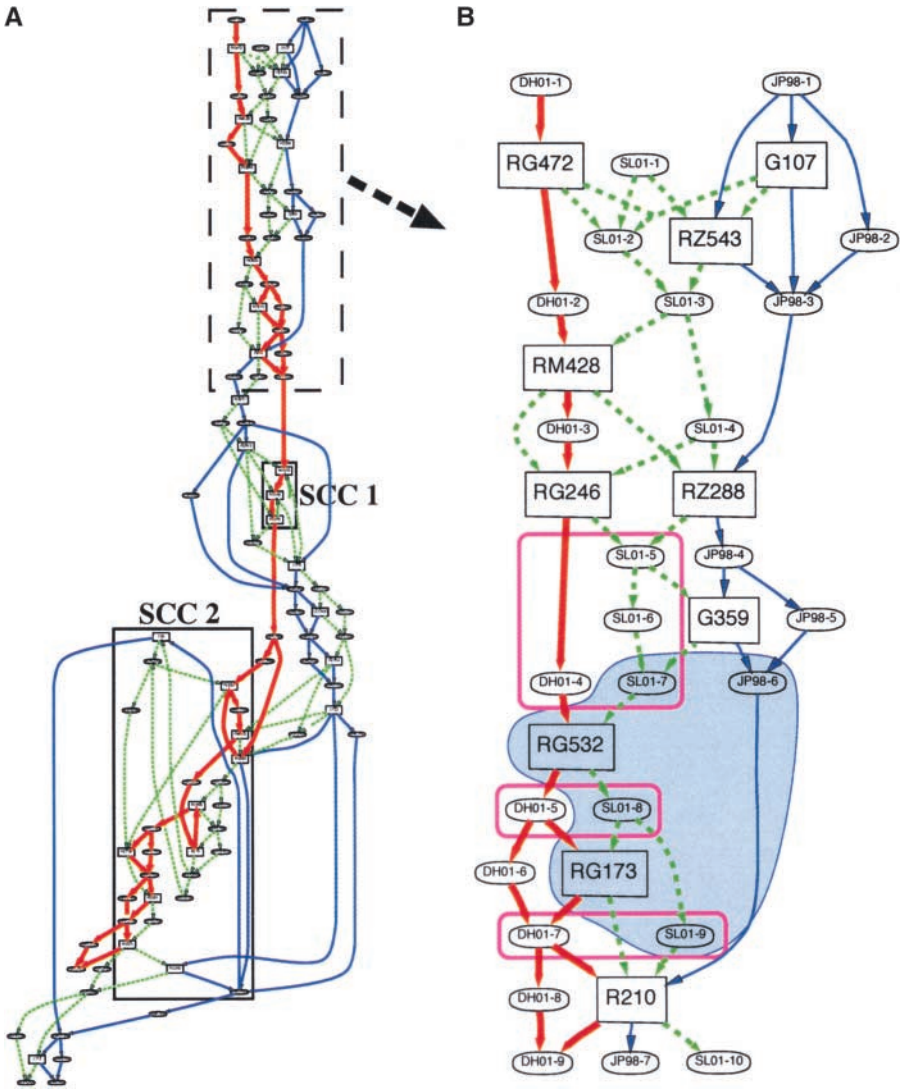

FIGURE 9.—Integrated graph of chromosome 1. Locus orders specified by DH01, SL01, and JP98 are indicated by boldface red, dashed green, and blue arrows, respectively. (A) Complete integrated graph of chromosome 1 showing two SCCs. (B) Top portion of the integrated graph. This rendering is isomorphic to the corresponding subgraph in A. The graph has been redrawn to better illustrate certain relationships. An ambiguous interval between SL01 and JP98 is indicated by a blue-shaded box. This overlaps or encompasses certain ambiguous intervals between SL01 and DH01, indicated by magenta boxes with rounded corners.

cluded the edge *RZ19* → *RG690*, which is the order specified on the DH01 map with a distance of 6.5 cM between these two loci. The other half included the opposing edge *RG690* → *RZ19*, which is the order on the SL01 map at 3.7 cM. Sequence for *RZ19* and *RG690* can be found in GenBank with accessions G73632 and AQ074147, which can be mapped to genomic BAC/PACs AP002972 and AP003377, respectively. Since the BAC/PACs were ordered AP002972 → AP003377, the loci are therefore ordered *RZ19* → *RG690* on the sequence-based map.

To resolve the rest of the region, note that the 12 MFES specified only nine edges and nine markers:

RG381 → *RM403*

RZ19 → *RG690*

RG690 → *RZ19*

C86 → *RG381*

RG350 → *C86*

RM403 → *RG345*

RG381 → *RZ730*

RG331 → *RG350*

RG345 → *RG381*

Of these nine edges, two have already been accounted for. Hence only seven edges with seven markers need to be examined more closely. First, the two markers *RG331* and *RG350* appear together on the same map only on SL01. Their order, *RG331* → *RG350*, is confirmed by comparison to chromosome 1 sequence (Table 1). Next, observe that the distance between *RM403* and *RG345* is only 1.6 cM on the SL01 map while the same two markers mapped in reverse order at a distance of 16 cM on the DH01 map. This suggests that the order of these two markers may have been reversed on the SL01 map due to a low frequency of recombination in this region, while ordering was much clearer on the DH01 population in which recombination was high, providing

TABLE 1

Loci along chromosome 1 that showed inconsistent order on genetic maps ordered on the basis of available genomic sequence

Cycle	Order along chr 1	Marker	GenBank accession	BAC/PAC
1	1	*AMY1B*	M59350	AP003275
	2	*RZ276*	AI978355	AP003854
			AI978356	AP003275
	3	*RG146*	AQ074233	AP003444
2	4	*RG345*	G73768	AP003222
	5	*RM403*	(AQ051193)	AP003273
	6	*RZ19*	G73632	AP002972
	7	*RG690*	AQ074147	AP003377
	8	*RZ730*	AI978485	AP003446
			AI978486	
	9	*C86*	D15115	AP003286
				AP003235
	10	*RG331*	G73765	AP003687
				AP004073
		RG350	G73729	AP004326
				AP003238

The corresponding GenBank accession(s) for each marker are given along with the significant hit for that sequence to genomic BAC/PAC sequence. Chr, chromosome.

less chance of statistical error. Comparison to genomic sequence confirmed that the order should indeed be *RG345* → *RM403*. Similarly, the sequence-based map confirms the locus order *C86* → *RG350* as given by the SL01 map, rather than the reverse order shown on JP98.

The remaining edges all involved *RG381*. Unfortunately, sequence for this marker is not available. However, note that marker *RG381* was mapped on SL01 as locus *RG381X*. This indicates that this marker was observed to have multiple copies on the basis of hybridization signal on Southern blots, only one of which was mapped (Causse *et al.* 1994; Wilson *et al.* 1999). Reexamination of the existing hybridization filters indicates that marker *RG381* does indeed exhibit multiple bands and is likely to map to more than one position in the rice genome (S. Harrington, personal communication). It appears, therefore, that one copy of *RG381* may have been mapped to position 112.35 cM on SL01 and another copy to position 117.1 cM on DH01.

Ambiguities: Examination of the integrated graph for pairs of nodes that have *no* path between them detects ambiguity. Figure 9B shows the integrated graph for the top portion of chromosome 1. Note that within the blue-shaded region is a node with the label "JP98-6"; this represents the loci that were mapped only on JP98 within the interval delineated by *G359* and *R210*. Note further that no path exists from JP98-6 to any of the other nodes (SL01-7, *RG532*, SL01-8, *RG173*, and SL01-9) within the blue region, all of which were mapped on SL01. The blue region hence indicates ambiguity between SL01 and JP98 within the interval *G359–R210*. This ambiguity is indicated on the corresponding map in Figure 8, in which a blue-shaded interval from *G359* to *R210* on SL01 is connected by an arrow to the corresponding blue-shaded interval on JP98.

The magenta boxes in Figure 9B represent ambiguity between SL01 and DH01 in the intervals *RG246–RG532*, *RG532–RG173*, and *RG173–R210*. Likewise, in Figure 8, magenta-shaded intervals on SL01 point to corresponding intervals on DH01. Note how the magenta- and blue-shaded intervals overlap each other. This shows how ambiguities may overlap, depending on the pair of comparisons being made. These ambiguous intervals may be considered to define a set of bins within which locus order is only partially known, but the boundaries of the bins depend on which set of comparisons is being made.

In Figure 8, lines connecting anchor loci may themselves indicate ambiguity, if loci cosegregate to that position. For example, on chromosome 1, *RG246* and *RZ288* cosegregate in SL01, but DH01 mapped only *RG246* and JP98 mapped only *RZ288*. This is reflected in the integrated graph, in which there is no path between the nodes representing *RG246* and *RZ288*. Hence, no conclusion can be drawn as to the relative order of these two loci since there is no evidence (due to lack of recombination, small population size, double crossover, missing data, etc.) to prefer one order over the other.

DISCUSSION

An implicit assumption behind the notion of an integrated map is the view that, for a species as a whole, there is one correct order of markers. Under this assumption, the data from individual mapping studies—using different crosses and different sets of markers—represent different samplings of the species map. Thus, the objective of map integration is to combine individual primary maps into a single integrated species map. The four approaches to map integration currently in use have as their goal to generate a simplified resolution of the inconsistencies and ambiguities that exist in the data. Presentation of a single integrated map tends to obscure these anomalies; our graph approach, on the other hand, highlights them. It is of interest to compare the way these different approaches handle inconsistencies and ambiguities.

One conceptualization of consensus mapping is that marker positions may be averaged across the different mapping studies being integrated. This is the view taken directly by the various *ad hoc* visual approaches to map integration. The visual approach is arguably the most flexible of the four approaches, since access to the raw marker segregation data is not required. This was the method of choice for the creation of the wheat consensus map, for instance, where the problem involved bringing together maps from the three different genomes that reside together in the nucleus of this allo-

hexaploid species (Nelson *et al.* 1995a,b,c; Van Deynze *et al.* 1995; Marino *et al.* 1996). Since construction of the consensus map involved integration across three homeologous groups of chromosomes that do not recombine with each other, this precluded the use of existing mapping software. The disadvantage of visual interpolation is that it is highly subjective and hence not wholly reproducible. It is also difficult to visually integrate more than two or three maps at a time.

The GDB (1999) approach prescribes a heuristic algorithm, in which a specific map is designated as the standard map. The coordinate system of a second map is then projected onto the standard map. The resulting "comprehensive" map is then used as the basis for incorporating the next map. This process is iterated until all maps have been incorporated into the comprehensive map. It is more objective than the visual approach and can be used to integrate multiple maps of different types. Although access to the raw data is not necessary, it requires that the map distances between markers be known. Most importantly, the final comprehensive map is dependent on the order in which the primary maps are incorporated.

Despite progress in physical and sequence-based mapping, genetic linkage maps are often the only type of map available for many species. Genetic maps continue to be a valuable tool for genomic research. They are often used to guide the assembly of physical and sequence-based maps. Genetic mapping monitors the recombinatorial process, allowing researchers to determine how genes are inherited in relationship to other genes in the genome. This is useful, for instance, when studying coinheritance or coregulation of linked genes, investigating the evolutionary significance of conserved chromosomal segments (homeologous regions) between species, or designing a plant or animal breeding strategy.

When all the maps to be integrated are genetic maps and the raw segregation data are available, it becomes possible to plug these data into standard mapping software such as MAPMAKER. MAPMAKER first estimates the most likely order of loci on the basis of two-point, three-point, and multipoint recombination values. Then it computes a genetic map on the basis of the estimated locus order. With complete data, one can simply count the number of meiotic recombinations that occurred between loci, adjust that data to take account of the probability of double-crossover events, and compute recombination fractions (and hence genetic distance). To handle missing data, MAPMAKER uses an iterative procedure to estimate recombination fractions. From an initial guess for the recombination fractions, the expected number of recombinant and nonrecombinant meioses in each interval for the complete data is computed as if the guess were actually correct. From this expected value, the maximum-likelihood estimate for the recombination fractions is computed. These new estimates for recombination fractions are then used to compute new expected values and the process is repeated until the likelihood converges into a maximum. MAPMAKER uses a hidden Markov chain model described by Lander and Green (1987) to simultaneously compute the expected number of recombinant meioses and the maximum-likelihood recombination fractions.

The models to estimate the recombination fractions and maximum likelihood are dependent on population size and structure. Because of this, one cannot use MAPMAKER or similar software to create an integrated map from an arbitrary set of mapping populations. However, map integration is possible if the populations are of similar size and identical structure. The pooled map for maize developed by Beavis and Grant (1991) used data from four different maize populations with similar numbers of individuals. Essentially, the individual populations were treated as different replications of a single map study. For the most part, however, an arbitrary set of mapping studies will differ in population size, structure, and resolution (*i.e.*, number of markers), rendering invalid the assumptions of standard mapping software such as MAPMAKER.

As an alternative to software such as MAPMAKER, JoinMap was designed to allow integration of genetic maps from disparate populations. Pairwise recombination frequencies are calculated as well as LOD scores for those pairs that are available in the entire data set. Beginning with the most informative pair of markers, JoinMap iteratively builds up the map by adding a new marker on the basis of LOD score. The best-fitting position of this new marker is determined without changing the order of markers that were placed earlier. Periodically, the locus orders are reshuffled so that the best-fit order for the entire map can be found. JoinMap adjusts for the differences in population size and structure by assigning different weights when estimating map distances (Stam 1993). However, this still does not account for the fact that different sets of parentals have differentially nonlinear recombination rates along the length of the chromosome. Further, crossovers occur nonrandomly in the genome. Observed recombination frequency not only varies with distance from the centromere, but also varies due to crossover interference, the presence of recombinatorial "hot spots" and "cold spots," *e.g.*, genes that change the rate of recombination either *in cis* or *in trans* (Muller 1916), and nonrandom gamete elimination resulting from both genetic and environmental causes.

Different maps produced by different studies are assumed to sample the same underlying physical order of markers. However, there will invariably be differences in marker order among the different studies. Resolving these inconsistencies prior to the construction of the consensus map generally requires an investment in additional laboratory work and can be very labor intensive and subjective. Combining multiple maps by current

approaches tends to obscure the ambiguities and inconsistencies among the different maps. While many of the inconsistencies in marker order may be attributable to statistical or experimental error, there may, in fact, be real biological differences in genetic colinearity even among individuals of the same species due to duplications, translocations, inversions, or movement of transposable elements as has been recently demonstrated by Fu and Dooner (2002).

CONCLUSION

We have taken a novel approach to map integration by modeling maps as graphs. Map integration and analysis are performed on these graphs and include the following steps:

1. Identifying and merging anchor nodes
2. identifying inconsistencies
3. global ordering of nodes
4. dissecting inconsistencies
5. identifying ambiguities.

Although an integrated graph may not be amenable to direct inspection, it can be used as a data structure to drive analysis in an objective and reproducible manner. The algorithms described in this article are very efficient and relatively easy to implement.

This approach to map integration allows for comparisons purely on the basis of marker order and does not require access to the raw mapping data or information about distances between markers. Furthermore, it can be used to integrate maps of different types, such as genetic, physical, or sequence based. Each linkage graph, in essence, is a statement about the order of markers derived for that linkage group from that particular mapping study. Map integration using these directed graphs allows us to reason directly about marker order and more easily determine both similarities and differences between distinct maps from different studies. When used directly, the resulting integrated graph is a faithful representation of the marker orders of its component maps, including all of their ambiguities and inconsistencies. Knowledge of where these irregularities occur can help to motivate additional research into investigating the biological reasons, if any, behind the inconsistencies or to better choose markers to resolve ambiguities.

We acknowledge Golan Yona, Pankaj Jaiswal, and two anonymous reviewers for their invaluable comments and suggestions. We also thank Lois Swales for her help in formatting and submitting. The software package Graphviz, used to render the graphs in this article, is available for download at http://www.graphviz.org. This work was funded in part by the U.S. Department of Agriculture-Agricultural Research Service (USDA-ARS) specific cooperative agreements 58-3655-7-101 and 58-1907-0-041, USDA-ARS CRIS 1907-21000-008, and USDA-ARS IFAFS 00-52100-9622.

LITERATURE CITED

Beavis, W. D., and D. Grant, 1991 A linkage map based on information from four F_2 populations of maize (*Zea mays* L.). Theor. Appl. Genet. **82:** 636–644.

Ben-Dor, A., and B. Chor, 1997 On constructing radiation hybrid maps. J. Comput. Biol. **4:** 517–533.

Benzer, S., 1959 On the topology of the genetic fine structure. Proc. Natl. Acad. Sci. USA **45:** 1607–1620.

Causse, M. A., T. M. Fulton, Y. G. Cho, S. N. Ahn, J. Chunwonse *et al.*, 1994 Saturated molecular map of the rice genome based on an interspecific backcross population. Genetics **138:** 1251–1274.

Cho, Y. G., S. R. Mccouch, M. Kuiper, M.-R. Kang, J. Pot *et al.*, 1998 Integrated map of AFLP, SSLP and RFLP markers using a recombinant inbred population of rice (*Oryza sativa* L.). Theor. Appl. Genet. **97:** 370–380.

Cormen, T. H., C. E. Leiserson and R. L. Rivest, 1990 *Introduction to Algorithms.* MIT Press, Cambridge, MA.

Fasulo, D., T. Jiang, R. M. Karp, R. J. Settergren and E. Thayer, 1999 An algorithmic approach to multiple complete digest mapping. J. Comput. Biol. **6:** 187–207.

Fishburn, P. C., 1985 *Interval Orders and Interval Graphs: A Study of Partially Ordered Sets.* John Wiley & Sons, New York.

Fu, H., and H. K. Dooner, 2002 Intraspecific violation of genetic colinearity and its implications in maize. Proc. Natl. Acad. Sci. USA **99:** 9573–9578.

Garey, M. R., and D. S. Johnson, 1979 *Computers and Intractability: A Guide to the Theory of NP-Completeness.* W. H. Freeman, New York.

GDB, 1999 GDB comprehensive map construction (http://www.gdb.org/).

Golumbic, M. C., and R. Shamir, 1992 Complexity and algorithms for reasoning about time: a graph-theoretic approach. J. ACM **40:** 1108–1133.

Harley, E., A. J. Bonner and N. Goodman, 1996 Good maps are straight, pp. 88–97 in *Proceedings, Fourth International Conference on Intelligent Systems for Molecular Biology (ISMB),* edited by D. J. States, P. Agarwal, T. Gaasterland, L. Hunter and R. Smith. AAAI Press, Washington University, St. Louis. (ftp://ftp.cs.toronto.edu/%2Fcs/ftp/pub/bonner/papers/genome.mapping/).

Harley, E., A. Bonner and N. Goodman, 1999 Revealing hidden interval graph structure in STS-content data. Bioinformatics **15:** 278–285.

Harushima, Y., M. Yano, A. Shomura, M. Sato, T. Shimano *et al.*, 1998 A high-density rice genetic linkage map with 2275 markers using a single F_2 population. Genetics **148:** 479–494.

Idury, R. M., and M. S. Waterman, 1995 A new algorithm for DNA sequence assembly. J. Comput. Biol. **2:** 291–306.

Kianian, S. F., and C. F. Quiros, 1992 Generation of a *Brassica oleracea* composite RFLP map: linkage arrangements among various populations and evolutionary implications. Theor. Appl. Genet. **84:** 544–554.

Lander, E. S., and P. Green, 1987 Construction of multilocus genetic linkage maps in humans. Proc. Natl. Acad. Sci. USA **84:** 2363–2367.

Lander, E. S., P. Green, J. Abrahamson, A. Barlow, M. J. Daly *et al.*, 1987 MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics **1:** 174–181.

Lincoln, S. E., M. J. Daly and E. S. Lander, 1993 *Constructing Genetic Linkage Maps with MAPMAKER/EXP Version 3.0: A Tutorial and Reference Manual.* Whitehead Institute, Cambridge, MA (http://www.broad.mit.edu/ftp/distribution/software/mapmaker3/).

Liu, B. H., 1998 *Statistical Genomics: Linkage, Mapping, and QTL Analysis.* CRC Press, Boca Raton, FL.

Marino, C. L., J. C. Nelson, Y. H. Lu, M. E. Sorrells, P. Leroy *et al.*, 1996 Molecular genetic maps of the group 6 chromosomes of hexaploid wheat (*Triticum aestivum* L. em. Thell.). Genome **39:** 359–366.

Muller, J., 1916 The mechanism of crossing over. Am. Nat. **50:** 193–207.

Myers, E. W., 1995 Toward simplifying and accurately formulating fragment assembly. J. Comput. Biol. **2:** 275–290.

Nelson, J. C., M. E. Sorrells, A. E. Van Deynze, Y. H. Lu, M. Atkison *et al.*, 1995a Molecular mapping of wheat: major genes

and rearrangements in homoeologous groups 4, 5, and 7. Genetics **141:** 721–731.

Nelson, J. C., A. E. Van Deynze, E. Autrique, M. E. Sorrells, Y. H. Lu *et al.*, 1995b Molecular mapping of wheat: homoeologous group 2. Genome **38:** 516–524.

Nelson, J. C., A. E. Van Deynze, E. Autrique, M. E. Sorrells, Y. H. Lu *et al.*, 1995c Molecular mapping of wheat: homoeologous group 3. Genome **38:** 525–533.

Randall, J. R., 1997 Using interval graphs for solving map assembly problems. Master's Thesis, Department of Computer Science, University of Toronto, Toronto, Ontario, Canada (ftp://ftp.cs.toronto.edu/pub/bonner/papers/genome.mapping/).

Sasaki, T., T. Matsumoto, K. Yamamoto, K. Sakata, T. Baba *et al.*, 2002 The genome sequence and structure of rice chromosome 1. Nature **420:** 312–316.

Sewell, M. M., B. K. Sherman and D. B. Neale, 1999 A consensus map for loblolly pine (*Pinus taeda* L.) I: construction and integration of individual linkage maps from two outbred three-generation progenies. Genetics **151:** 321–330.

Slonim, D., L. Kruglyak, L. Stein and E. Lander, 1997 Building human genome maps with radiation hybrids. J. Comput. Biol. **4:** 487–504.

Stam, P., 1993 Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. Plant J. **3:** 739–744.

Stam, P., and J. W. V. Ooijen, 1995 *JoinMap Version 2.0: Software for the Calculation of Genetic Linkage Maps.* CPRO-DLO, Wageningen, The Netherlands.

Tani, N., T. Takahashi, H. Iwata, Y. Mukai, T. Ujino-Ihara *et al.*, 2003 A consensus linkage map for sugi (*Cryptomeria japonica*) from two pedigrees, based on microsatellites and expressed sequence tags. Genetics **165:** 1551–1568.

Temnykh, S., W. D. Park, N. Ayres, S. Cartinhour, N. Hauck *et al.*, 2000 Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.). Theor. Appl. Genet. **100:** 697–712.

Temnykh, S., G. Declerck, A. Lukashova, L. Lipovich, S. Cartinhour *et al.*, 2001 Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Res. **11:** 1441–1452.

Van Deynze, A. E., J. Dubcovsky, K. S. Gill, J. C. Nelson, M. E. Sorrells *et al.*, 1995 Molecular-genetic maps for group 1 chromosomes of Triticeae species and their relation to chromosomes in rice and oat. Genome **38:** 45–59.

Wilson, W. A., S. E. Harrington, W. L. Woodman, M. Lee, M. E. Sorrells *et al.*, 1999 Inferences on the genome structure of progenitor maize through comparative analysis of rice, maize and the domesticated Panicoids. Genetics **153:** 453–473.

Communicating editor: G. A. Churchill