JOURNAL OF COMPUTATIONAL BIOLOGY
Volume 14, Number 2, 2007

Pp. 144–155
DOI: 10.1089/cmb.2006.0127

Interpreting Chromosome Aberration Spectra

DAN LEVY,[1] CHRISTOPHER REEDER,[1] BRADFORD LOUCAS,[2] LYNN HLATKY,[2] ALLEN CHEN,[1] MICHAEL CORNFORTH,[2] and RAINER SACHS[1]

ABSTRACT

Ionizing radiation can damage cells by breaking both strands of DNA in multiple locations, essentially cutting chromosomes into pieces. The cell has enzymatic mechanisms to repair such breaks; however, these mechanisms are imperfect and, in an exchange process, may produce a large-scale rearrangement of the genome, called a chromosome aberration. Chromosome aberrations are important in killing cells, during carcinogenesis, in characterizing repair/misrepair pathways, in retrospective radiation biodosimetry, and in a number of other ways. DNA staining techniques such as mFISH (multicolor fluorescent in situ hybridization) provide a means for analyzing aberration spectra by examining observed final patterns. Unfortunately, an mFISH observed final pattern often does not uniquely determine the underlying exchange process. Further, resolution limitations in the painting protocol sometimes lead to apparently incomplete final patterns. We here describe an algorithm for systematically finding exchange processes consistent with any observed final pattern. This algorithm uses aberration multigraphs, a mathematical formalism that links the various aspects of aberration formation. By applying a measure to the space of consistent multigraphs, we will show how to generate model-specific distributions of aberration processes from mFISH experimental data. The approach is implemented by software freely available over the internet. As a sample application, we apply these algorithms to an aberration data set, obtaining a distribution of exchange cycle sizes, which serves to measure aberration complexity. Estimating complexity, in turn, helps indicate how damaging the aberrations are and may facilitate identification of radiation type in retrospective biodosimetry.

Key words: ionizing radiation, chromosome aberrations, graph theory.

1. INTRODUCTION

1.1. Chromosome aberrations

IONIZING RADIATION acting on cells during the G0/G1 phase of the cell cycle causes double strand breaks (DSBs) to occur in chromosomes. The cell is equipped with various mechanisms for repairing DSBs, but these mechanisms are imperfect and may result in large-scale rearrangement of the genome. This happens

[1]Department of Mathematics, UC Berkeley, Berkeley, California.
[2]Department of Radiation Oncology, University of Texas Medical Branch, Houston, Texas.
[3]Center for Cancer Systems Biology, Caritas St. Elizabeth's Medical Center, Tufts School of Medicine, Boston, Massachusetts.

when the two free ends of a DSB, instead of rejoining with each other, misrejoin with a free end of a different DSB (Fig. 1). This process of producing a large-scale rearrangement is called an exchange aberration, here referred to as simply a *chromosome aberration* (Savage, 1998).

There are many different kinds of chromosome aberrations, characterized by the number of chromosomes involved, the number of DSBs involved, and by subtler criteria, some of which are discussed below. Some kinds of chromosome aberrations kill cells. Aberrations are also prominent during carcinogenesis (Mitelman et al., 2004). In fact, one of the few cancers whose basic cause is well understood is chronic myeloid leukemia, produced by a very specific type of chromosome aberration (Wong and Witte, 2004). Chromosome aberrations are radiation damage endpoints that persist, sometimes for many cell generations. Consequently, they are used in retrospective biodosimetry, to see how much radiation dose an exposed individual received (Leonard et al., 2005).

Any kind of radiation makes many kinds of chromosome aberrations, with each cell typically having its own aberration pattern. However, there are differences in the spectra of aberrations produced by different radiation types. Sparsely ionizing radiations, such as x-rays or gamma rays, make DSBs scattered almost at random throughout the genome, and except at high doses, the corresponding aberrations are usually simple: involving only two DSBs and only one or two chromosomes. Densely ionizing radiations, such as alpha particles or neutrons, make DSBs in clusters along chromosomes. The resulting aberrations are often complex, involving many DSBs in a reaction that is *irreducible*: we cannot group the DSBs into disjoint sets with no exchanges among DSBs in different sets (Sachs et al., 2002, 2004). Clustering of the DSBs and complexity

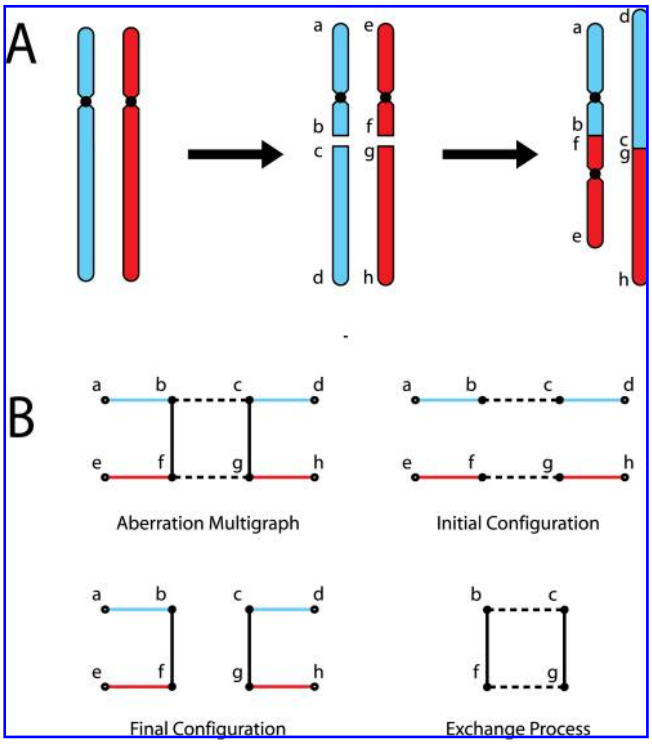

FIG . 1. **(A)** A simple aberration. The two chromosomes receive one double-strand break (DSB) each, and each break results in two free ends. In this aberration, those free ends misrejoined to produce one dicentric chromosome and one acentric chromosome. Constrictions represent centromeres (chromosome structural features used during cell division and readily visible in optical preparations of dividing cells). The ends of a chromosome (e.g., a, d, e, and h) are also special structures called telomeres. **(B)** The aberration multigraph and its subgraph components. Each free-end and telomere is a labeled vertex and the edges of the graph connect vertices that are either: (a) joined by a length of chromatin (color line), (b) have an initial partnership (dotted line), or (c) have a final partnership (solid black line). In comparative genomics, it is often assumed that all large-scale rearrangements can be reduced to a sequence of rearrangements each one of which involves only two DSBs, as in the case shown above (Bourque and Pevzner, 2002). In current radiation cytogenetics, however, it is assumed that complex rearrangements occur, involving three or more different DSBs mutually interacting, and multigraphs help keep track of the interaction process, as demonstrated in Figure 2.

of the chromosome aberrations caused by densely ionizing radiations is believed to underlie the fact that, for a given dose, densely ionizing radiations are usually much more dangerous than sparsely ionizing radiations (Nelson, 2003). Differences in aberration spectra between densely and sparsely ionizing radiations are also important in retrospective biodosimetry since they can indicate what kind of radiation an affected person received (Hande et al., 2003).

1.2. Final patterns and painting protocol

To study chromosome aberrations, methods have been developed for staining sequences of DNA with fluorescent markers. Techniques include staining a single homologue pair (one-color FISH), uniquely staining each homologue pair (mFISH), staining individual genes, or staining isochors (banding). Still other methods apply fluorescent markers to the special chromosome structures called centromeres and telomeres (Fig. 1A). By marking the DNA, the results of the DNA repair/misrepair process are typically seen as an *observed final aberration pattern* (Fig.1). These staining techniques have a large lower limit of resolution—typically locations within the genome can be specified at best to the nearest Mbp.

If we think of the staining protocol as a mapping from a chromosome aberration formation process to the final pattern observed after the process concludes, we notice that this is a many-to-one function. For instance, if two breaks occur in one chromosome arm and the intervening chromatin stretch is simply reversed, this process gives the same observed mFISH final pattern as if no aberration had occurred. So, in order to obtain information about chromosome aberrations from experimentally generated final patterns, we need a way to "invert" (get the inverse image of) this function, obtaining a set of chromosome aberrations from each observed final pattern. By applying a probability measure on the space of aberrations, we can then obtain statistics about the aberration process from a collection of observed final patterns.

In this paper, we will present systematic algorithms for analyzing aberration spectra; based on analyzing how the actual process of aberration formation is related to the observed final patterns. We will describe web-based software implementing the algorithms. To demonstrate the power of these algorithms, we analyze observed final patterns from an experiment involving the kind of densely ionizing radiation that is of particular interest for estimating risk to astronauts on extended space missions. Our approach includes the first fully systematic analysis of apparent incompleteness, where some DSB free ends apparently fail to either rejoin or misrejoin, which has confounded previous analyses, especially for densely ionizing radiation.

2. ABERRATION PROCESSES AND OBSERVED FINAL PATTERNS

2.1. Assumptions and conventions

We will provide a formalism applicable to any staining protocol that is whole-chromosome, i.e., assigns colors to entire chromosomes, rather than staining different stretches of chromosome differently. For brevity of presentation, we will confine our detailed algorithms to the mFISH case. That is, we assume that each homologous pair of chromosomes is painted its own unique color (with two additional colors for the X and Y chromosomes in the case of male cells) and that centromeres (Fig. 1A) can be recognized, but the exact original location, size, and orientation of a chromatin segment are not determined by the observed final pattern. For the time being, we also make three simplifying assumptions:

1. The aberration process is complete in the sense that every DSB free end is either rejoined to its own initial partner or misrejoined to some other free end.
2. The painting protocol has sufficient resolution that no chromatin segments are too small to be observed.
3. Telomeres (Fig. 1A) are well labeled.

We will relax these conditions later.

Further, since there are intrachromosomal and homologue-homologue exchanges which mFISH cannot reveal, there are an infinite number of aberration processes that might lead to an observed final pattern. To make the problem tractable, we will restrict the space of chromosome aberrations to those that involve a minimal number of breaks. Since, as will be discussed, the minimal number are also necessarily in any

consistent aberration process, the minimal break solutions serve as a natural starting point for generating distributions that include cryptic damage.

2.2. *Aberration multigraphs*

To proceed, we need somewhat more systematic specifications of the chromosome aberration processes, the final patterns and their interrelationship, which we discussed informally in the introduction. There are several elements to account for when describing a chromosome aberration: the double strand breaks, the stretches of chromatin, and the initial and final partnerships between break free ends. The chromatin, double strand breaks, and the initial partnerships taken together describe the *initial configuration*, the distribution of breaks on the chromosomes involved in the aberration. The chromatin and final partnerships constitute the *final configuration* that determines the observed final pattern. Lastly, the initial and final partnerships form an *exchange process*, which describes the interaction between DSBs. The *aberration multigraph*, introduced in Sachs et al. (2002), is a graph-theoretic description of a chromosome aberration that unifies the initial configuration, the final configuration, and the exchange process (Figs. 1B and 2). The vertices of the aberration multigraph are the *free-end vertices* formed by the DSBs and the *telomere vertices* of the chromosomes involved in the aberration. The multigraph edges fall into three classes: *chromatin edges*, which connect vertices joined by a stretch of chromatin; *initial edges*, which connect free-end vertices with an initial partnership; and *final edges*, which connect free-ends with a final partnership. An mFISH aberration multigraph is an aberration multigraph with a color and possibly a centromere associated to each chromatin edge.

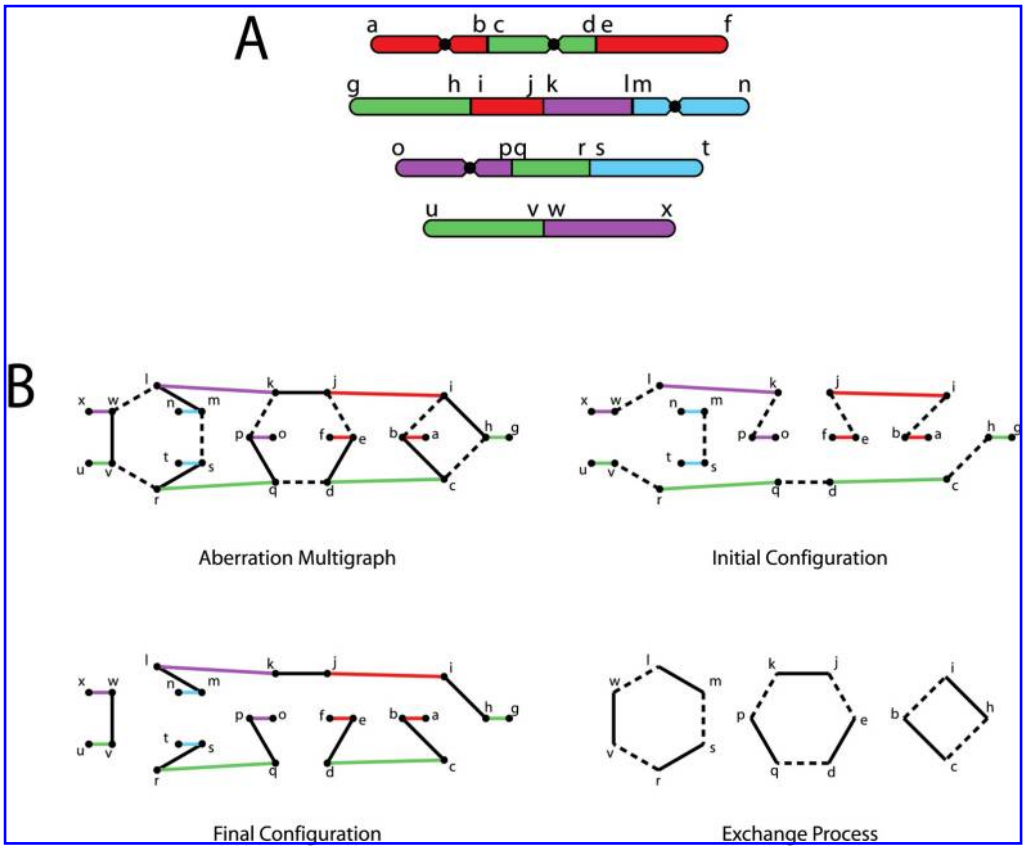

FIG . 2. **(A)** A complex aberration final pattern. **(B)** One of the 32 aberration multigraphs consistent with this final pattern, and corresponding subgraphs for the initial configuration, final configuration, and exchange process.

2.3. Initial configurations and observed final patterns

We will here assume a eukaryotic genome, where each chromosome has two telomeres and one centromere. This constrains the space of consistent mFISH multigraphs to those with an initial configuration subgraph where each connected component has two telomeres, one centromere and just one chromatin edge color.

We will consider an observed mFISH *final pattern* to be a collection of linear and ring rearranged chromosomes with lengths of DNA painted to indicate their chromosomes of origin. From an observed mFISH final pattern, one cannot determine the orientation, order, or length of the chromosome fragments. Further, if two homologous chromosomes are involved in the same aberration, mFISH cannot distinguish which fragment belongs with which chromosome. Consequently, there are often a number of mFISH aberration multigraphs consistent with an observed mFISH pattern. There are some misrejoinings that must exist in any aberration multigraph consistent with an observed mFISH final pattern and we call these *manifest misrejoinings*. Some misrejoinings will not be identifiable because of the limitations of mFISH and these we call *cryptic misrejoinings*. In reporting an observed final pattern, it is customary to exclude apparently unaffected chromosomes. However, if a color occurs in a final pattern and there is no centromeric fragment of that color, we will add an apparently unaffected chromosome of that color to the final pattern.

2.4. Manifest misrejoinings

Manifest misrejoinings occur at color junctions, between two adjacent same-color fragments both having centromeres, somewhere in a single-color ring, and between telomeres on a single-color linear rearranged chromosome. A review of each type of manifest misrejoining matched against the requisite conditions for the initial configuration demonstrates that the manifest misrejoinings must be present in any consistent aberration multigraph. We will show that if the final pattern admits an aberration multigraph, the manifest misrejoinings are not just necessary, but also sufficient.

Once we have placed all the manifest misrejoinings in our final pattern, it is easy enough to read off those elements that will be common to all consistent aberration multigraph, namely the vertices and both the final and chromatin edges. The vertices are simply the telomeres and free ends formed by the manifest breaks and chromatin edges are the lengths of DNA that stretch between them. Final edges are those free-end pairs that share the same manifest break. All that remains is to determine the initial edges.

2.5. Chromatin edge decomposition

One consequence of specifying the manifest misrejoinings is that we now have a collection of chromatin edges such that each chromatin edge has a single color, one or zero telomeres, and one or zero centromeres. Consequently, we can form a *chromatin edge decomposition*. This is a collection of chromatin edges all of the same color and divided into one of four categories:

- TC—has telomere and centromere
- TNC—has telomere and no centromere
- NTC—has no telomere but has a centromere
- NTNC—has neither telomere nor centromere

It turns out that these distinctions are the only ones necessary for determining the full range of consistent aberration multigraphs.

2.6. Initial edge algorithm

We will here provide only a brief review of the algorithm for determining all valid sets of initial edges that complete our aberration multigraph. Details may be found in Levy et al. (2004), and an overview is given in Figure 3. Since we assume no *a priori* relationship between final and initial edges, finding a valid set of initial edges requires only information about the chromatin edges. Since initial edges only connect chromatin edges of the same color, we can find all sets of valid initial edges for each color and permute over all color choices to find all consistent initial configurations.

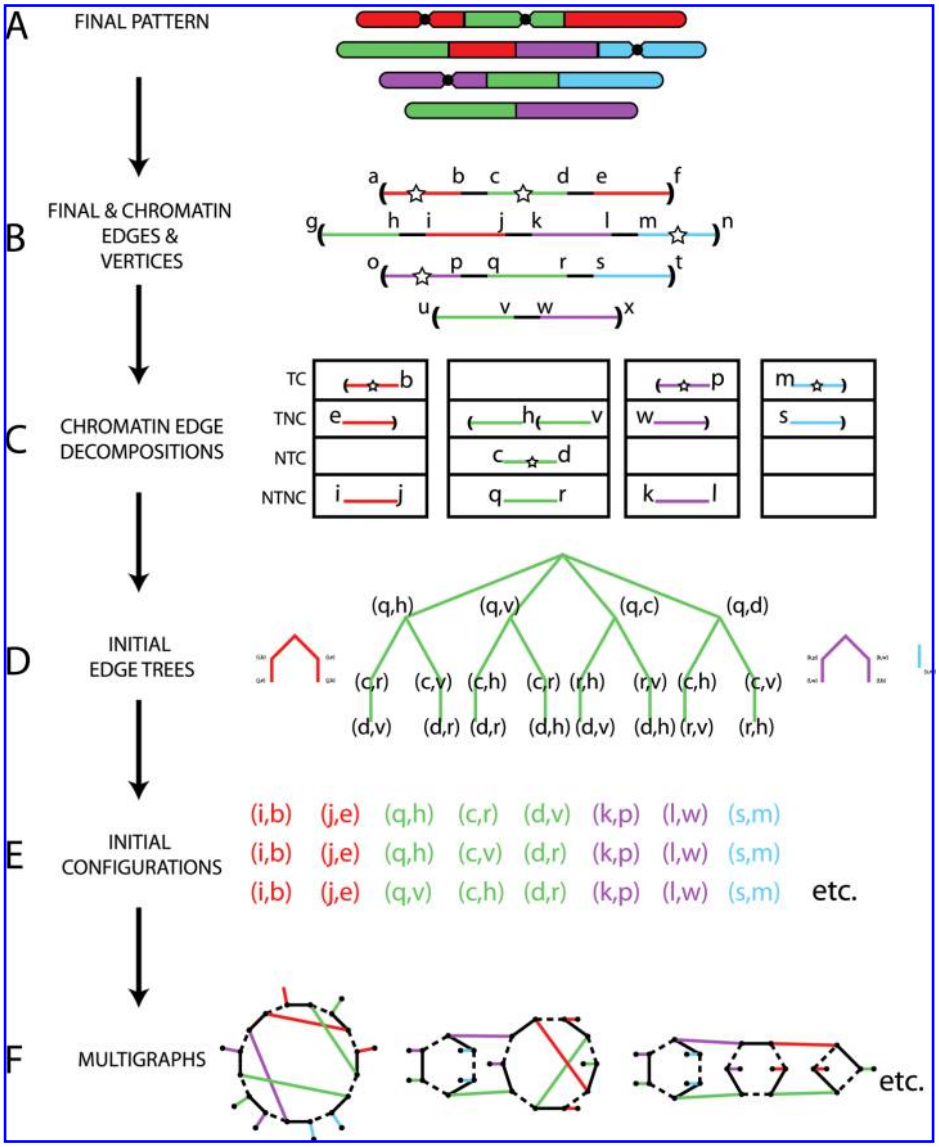

FIG. 3. A flowchart of the initial configuration algorithm, illustrated with a typical example. **(B–C)** Stars denote centromeres, and parentheses denote telomeres. From the final pattern, we can easily determine the final and chromatin edges and the vertices of our multigraph. We then obtain a chromatin edge decomposition for each color, and using a recursive algorithm, construct a tree of initial edges for each color. By taking all possible paths from root to leaf over all of the initial edge trees, we generate all possible initial edge configurations. Joining these with the final and chromatin edges, we obtain a distribution of aberration multigraphs.

The heart of our algorithm is a function that inputs a chromatin edge decomposition (Fig. 3C) and returns a tree of initial edges (Fig. 3D) such that any path from root to leaf specifies a unique initial configuration (Fig. 3E). The only requirement for the successful termination of this function is that the chromatin edge decomposition have at least one centromere and have exactly two telomeres for every centromere. The function operates recursively as follows (Levy et al., 2004): choose an appropriate vertex from the decomposition and generate a child for every initial edge that might be incident on that vertex; by joining the two chromatin edges incident on the two partners, we obtain a reduced chromatin edge decomposition with one or two fewer chromatin edges; we then iterate until the reduced decompositions are empty. Since every path differs in at least one initial edge, every initial configuration is unique. At each iteration, we consider all possible initial edges incident on that vertex, so our algorithm is also exhaustive.

By creating an initial edge tree for each color, taking all possible unions of root to leaf paths through all the trees, and joining these with the final and chromatin edges as determined by the final pattern, we obtain all possible consistent multigraphs. It can be shown that the number of initial configurations is

$$\prod_{l \in Ab} (2B(l) - 2) \cdots (2A(l)) \cdot A(l)! = \prod_{l \in Ab} 2^{B(l)-A(l)} A(l)(B(l) - 1)!$$

where $B(l)$ is the number of double stranded breaks on chromosomes with label l and $A(l)$ is the number of chromosome arms of label l that receive at least one break. The product is taken over all labels that occur in the aberration.

3. APPARENTLY INCOMPLETE FINAL PATTERNS AND TRULY INCOMPLETE ABERRATIONS

We now consider the previously confounding case of apparent incompleteness. Reviewing, the previous sections made the following three assumptions:

1. The DSB repair/misrepair process goes to completion, with all DSB free ends either misrejoined or rejoined to their original partners.
2. The painting scheme has sufficiently high resolution that no chromatin fragments of size smaller than the resolution are involved in the aberration.
3. Telomeres are well marked.

A major problem in analyzing aberrations has been that these assumptions are not always good approximations, especially for densely ionizing radiations. This problem admits a systematic solution.

For a fixed final pattern, let $C(i)$ be the number of centromeres for color i and let $T(i)$ be the number of terminal ends of color i. Under our three assumptions, $T(i) = 2C(i)$ since there are initially two telomeres for each centromere and each terminal end in the observed mFISH final pattern is actually a telomere. We define the *incompleteness measure* for color i by $m(i) = T(i) - 2C(i)$. If the incompleteness measure for any color is non-zero, we say that the observed final pattern is *apparently incomplete*. The mFISH data set analyzed below is an example where many observed final patterns are apparently incomplete. If there are actually unrejoined free ends (i.e., if assumption 1 fails) then the aberration process, which is a more basic feature than just the observed final pattern, is called *truly incomplete*; if the painting protocol has sufficiently high resolution, a truly incomplete aberration process always generates an apparently incomplete observed final mFISH pattern because $T(i)$ is too large for one or more colors.

There are three different situations that might affect the incompleteness measure of a final pattern (Fig. 4). First, it is possible that the terminal end of a linear fragment is not a telomere, but instead a free end that failed to find a final partner. Second, it might be the case that what appears to be a telomere of color a is actually a telomere of color b, but the b fragment is smaller than the limits of resolution. Lastly, it is possible that an entire rearranged chromosome was too small to be identified. Hence, to analyze an apparently incomplete mFISH aberration, we must either turn telomeres into free ends or add small telomeric chromatin fragments. That is precisely what we will do.

3.1. Completion algorithm

If an observed mFISH pattern is apparently incomplete, it must be the case that $m(i) \neq 0$ for some i.

1. If $m(i) < 0$, we have a telomere deficit and so we add to the edge decomposition $|m(i)|$ short fragments each with one free end, one telomere, and a chromatin edge between them of color i. We will place the free ends created in this way into the set F_1.
2. If $m(i) > 0$, we have too many telomeres and so we make $m(i)$ of the $T(i)$ telomeres into free ends in all possible ways. There will be $\binom{m(i)}{T(i)}$ different ways to do this. We place these free ends into the set F_2.

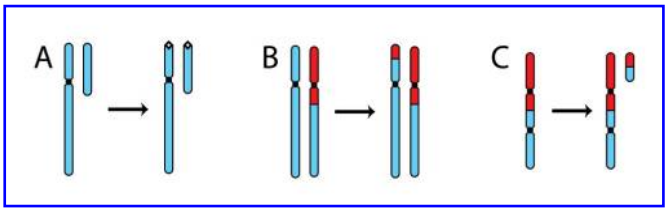

FIG . 4. Three types of incompleteness error associated with mFISH. The observed final pattern is shown on the left, and the true situation is shown on the right. **(A)** Pair of free-ends that failed to rejoin or misrejoin and were mistakenly mislabeled as telomeres. **(B)** Red telomeric segment that is too small to be scored and misidentified as a blue telomeric segment. **(C)** Entire rearranged chromosome that was too small to be identified.

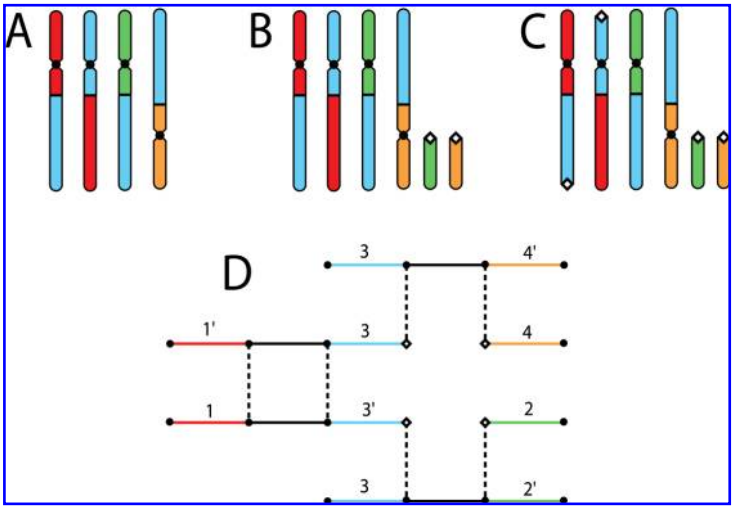

FIG . 5. The completion algorithm. Open free ends are depicted with a white diamond. **(A)** Observed final pattern. **(B)** First step in which telomeric fragments are added for those colors which have too few telomeres. **(C)** We see one of the $\binom{4}{2}$ ways in which we can select blue telomere vertices to convert to open free ends. **(D)** Aberration multigraph. We may place final edges between these open free-end vertices in 3!! = 3 ways.

We carry out this procedure for each color i to obtain

$$\prod_{\forall i | m(i) > 0} \binom{m(i)}{T(i)}$$

completions. Each of the resulting final configurations will produce a set of consistent aberration multigraphs. However, each multigraph will have $\sum_{\forall i} |m(i)| = |F_1| + |F_2|$ free ends without final partnership, where $|F_j|$ denotes the number of elements in F_j.

3.2. Joining free ends

Let $F = F_1 \cup F_2$ be the set of generated free ends that lack final partnership. A simple parity arguments shows that $\sum T(i)$ is an even integer so $\sum |m(i)|$ is also even and, consequently, $|F|$ is even. One may choose to place final edges between all, some, or none of the open free ends. Joining two F_1 vertices corresponds to an overlooked rearranged chromosome (Fig. 4C). Joining an F_1 vertex to an F_2 vertex corresponds to a telomere end below the level of resolution (Fig. 4B). And lastly, having two F_2 vertices corresponds to two free ends that were either mislabeled as disjoint (for instance, a ring misidentified as a linear fragment) or that were true free ends which failed to rejoin or misrejoin (Fig. 4A). If we place final edges between all of the free ends, the result is a truly complete aberration process. This may be done in $(|F| - 1)!!$ ways. All consistent aberration processes that are not truly complete can be obtained systematically, by omitting one or more of these final edges.

4. JAVA IMPLEMENTATION

We have implemented the initial edge algorithm, the completion algorithm, and a suite of tools for manipulating aberration multigraphs into a Java API called the Chromosome Aberration Analyzer (CAA). It was written to be interoperable with the Chromosome Aberration Simulator, a Java API for generating chromosome aberrations and final patterns. To demonstrate an application of these algorithms to experimental data, we created a set of tools for determining the probability distribution of exchange cycle structures for a given observed final pattern. The cycle structure is a measure of aberration complexity; it is a descending list of the number of breaks involved in each connected component of the exchange cycle subgraph (Levy et al., 2004). For instance, the exchange process in Figure 2 has a cycle structure of 3-2-2. A screenshot of the CAA applet (Fig. 6), available at *http:\\radiobiology.berkeley.edu\CAA*, shows the distribution of cycle structures for a complex aberration, taking into account the various possible multigraphs and computing the exchange cycle structures for each one.

5. EXPERIMENT

The CAA API also contains methods for analyzing whole sets of final patterns. We applied these methods to a set of mFISH experiments. The experiments are here treated as an example; a more detailed discussion of the experiments including methods, results and biological implications will be published elsewhere. The experiments involve densely ionizing radiation, specifically high-energy iron nuclei (i.e., fully ionized iron atoms traveling at speeds comparable to the speed of light), a type of radiation that is believed to pose a possible danger to astronauts undertaking prolonged trips outside the earth's atmosphere and magnetic field (Durante et al., 2002; Cucinotta and Durante, 2006). Such radiation results in many apparently incomplete final patterns, with, on average, 67% of final patterns apparently incomplete, regardless of dose.

Figure 7 shows the number of manifest breaks per cell, which is the most robust critereon for overall aberration frequency (Sachs et al., 2000) at various doses. It is seen that the dose dependence has the classic

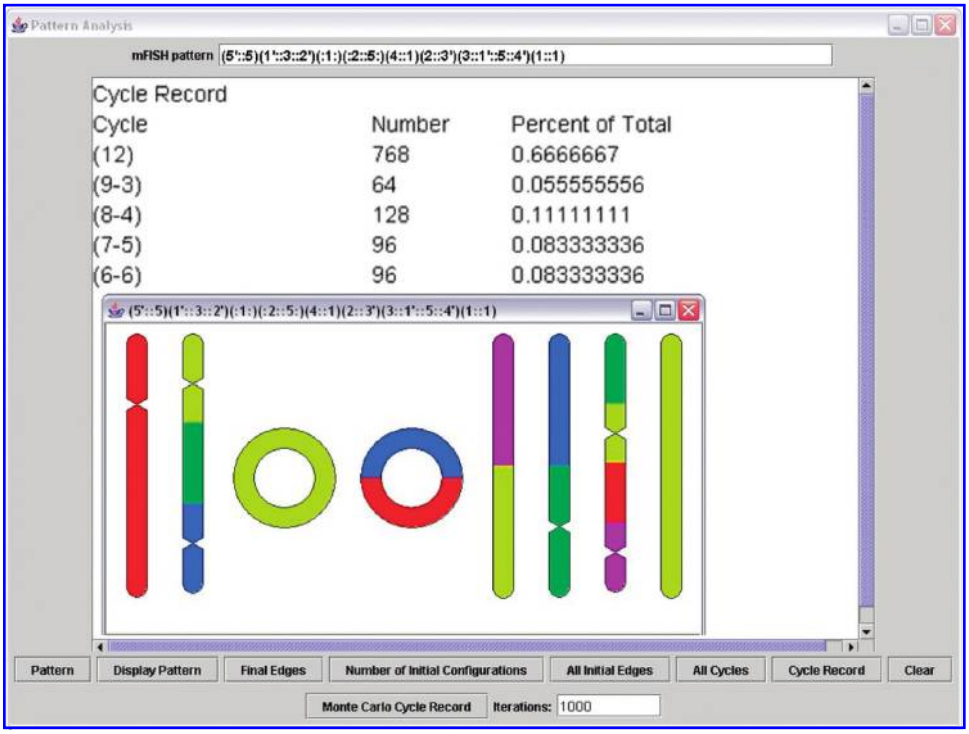

FIG . 6. A screenshot of the CAA java applet showing the distribution of cycle structures for a complex aberration.

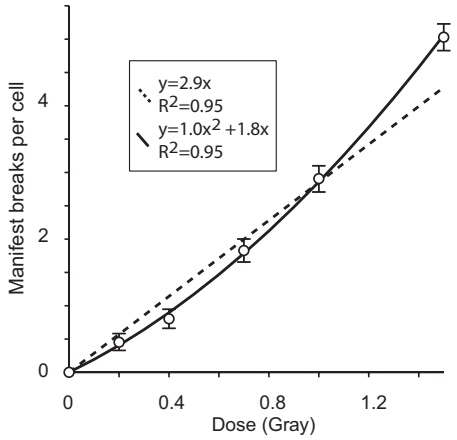

FIG . 7. Manifest breaks per cell as a function of dose. Error bars show standard deviations computed from Poisson statistics. Variance-weighted least squares linear and linear-quadratic fits are also shown with their equations and adjusted R values.

linear-quadratic form; the linear term is indicative of 1-track action (misrejoinings involving breaks made by a single radiation track) and the quadratic term is indicative of multitrack action (Hlatky et al., 2002).

Assuming a minimum number of breaks, an equal likelihood of free-end misrejoining, and an equal likelihood of final edge pairings for open free ends, we computed the number of exchange events involving n breaks. Subtracting out the zero Gray data, we can find the number of n-cycles caused by each experimental dose and from this, we determined the proportional representation of simple and higher order cycles (Fig. 8). For all doses, there is a strong admixture of cycles with order higher than two, in contrast to results for sparsely ionizing radiation, where significant fractions of such complex aberrations appear only at higher doses (Vazquez et al., 2002). For the smaller doses, where there are a few breaks per cell, there are fluctuations in cycle number, presumably due to noise. For doses at or above 0.7 Gy, the results are consistent with Figure 7: the proportion of each kind of cycle is nearly constant, as expected for one-track action, but the proportion of higher order breaks increases somewhat with dose, corresponding to multi-track action.

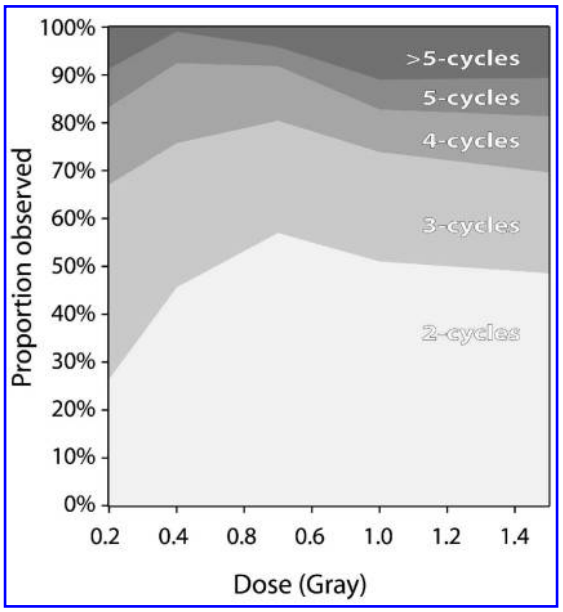

FIG . 8. Proportional representation of n-cycles as a function of dose. We determined the average number of n-cycles per cell for each dose, and by subtracting the values for zero gray, we determine the excess n-cycles at each dose. From these values, we determine the proportional representation of each n-cycle for each experimental dose.

6. LIMITATIONS AND EXTENSIONS

6.1. Cryptic damage

The algorithms presented here use only the manifest misrejoinings and, in the apparently incomplete case, those misrejoinings that are required for completion. Hence, they do not take into account the effects of cryptic damage. Cryptic damage is easy enough to simulate by distributing additional breaks into the final configuration before analysis. This is a costly method: depending on the number of cryptic breaks added and the number of cryptic break scenarios explored. However, the addition of cryptic breaks does not alter the functionality of the completion or multigraph algorithms. Simply note, that if you use the multigraph algorithm in this way, it will also consider the restitution of cryptic damage, where free ends are rejoined with their own initial partners, among the exchange scenarios.

6.2. Multiple counting of completions

The completion algorithm accurately determines the distribution of completion scenarios assuming an equal likelihood of misrejoining for any pair of free ends; however; there is some duplication of effort in the process. F_1 free ends of the same color are indistinguishable and swapping any two in the final configuration changes neither the configuration nor the completion. We could eliminate this extra work by refining the algorithm to treat F_1 fragments of the same color as identical. However, it is worth noting that the current algorithm does not alter the outcome of the distribution and may be preferable in methods that do not exhaustively search the solution space — e.g., Monte Carlo or simulated annealing.

6.3. Weighted final edge pairings

The assumption of an equal likelihood of final edge pairing is only a coarse assumption (Sachs et al., 2004), and the completion algorithm is structured to allow the application of a distribution when assigning final partnership for the generated free-ends. As a test case, we recomputed the results shown in Figures 7 and 8 assuming a strong bias for observed one-color acentric fragments to be acentric rings with the two ends misrejoining with each other rather than failing to rejoin or misrejoin elsewhere. The results do not differ significantly from those shown in the figures. In general, the parameters for the final partnership distribution could be determined by an experiment that combines mFISH with telomere probes. The probability of an open free-end that fails to find final partnership may be determined from the number of linear rearranged chromosomes with one or no telomere signal. The probability of misreading a telomere color end may be determined from the number of times the telomeres signals for a particular color exceed the true number. In this way, we can determine an empirical distribution for the completion scenarios.

6.4. Complexity of algorithms

Lastly, our algorithm details a complete exploration of the solution space and hence is factorial in the number of breaks per color. However, that assessment is in the worst case and a desktop computer is currently capable of analyzing aberrations exceeding 30 breaks. Further, we have determined that these techniques are amenable to a Monte Carlo approach, sampling the solution space to obtain an approximate distribution.

7. CONCLUSION

Chromosome aberration patterns provide a wealth of information about DNA geometry, radiation repair/misrepair processes, doses occurring in, for example, accidental, occupational, or medical exposures. However, to access this information requires a method that infers the aberration formation process from the observed final pattern. It is usually possible to reconstruct a few likely scenarios ad hoc by visual inspection and, so far, this is the technique most often used. However, with increasingly powerful techniques revealing increasingly complex aberrations, more systematic methods are needed.

We have provided fully systematic ways to determine the aberration processes consistent with a given observed final pattern and to find all completions or partial completions of apparently incomplete patterns. As shown by an example, these algorithms, even considered as a first approximation to more detailed scenarios, provide powerful means for analyzing data on chromosome aberration spectra while making explicit the conditions of the underlying model.

ACKNOWLEDGMENTS

D.L. was supported by the NIH (grant R01-GM68423). B.L. and M.C. were supported by the Office of Science (BER), U.S. Department of Energy (grant DE-FG03-02ER63442), and the National Aeronautical and Space Administration Office of Biological and Physical Research (NASA/OBPR). L.H. was supported by NASA (grant NNJ04HF42G), A.C. by DOE (grant DE-FG02-03ER63668), and R.S. by NASA NSCOR (grant NNJ04HJ12G).

REFERENCES

Bourque, G., and Pevzner, P.A. 2002. Genome-scale evolution: reconstructing gene orders in the ancestral species. *Genome Res.* 12:26–36.

Cucinotta, F.A., and Durante, M. 2006. Cancer risk from exposure to galactic cosmic rays: implications for space exploration by human beings. *Lancet Oncol*. 7:431–435.

Durante, M., George, K., Wu, H., et al. 2002. Karyotupes of human lymphocytes exposed to high-energy iron ions. *Radiation Res.* 158:581–590.

Hande, M.P., Azizova, T.V., Geard, C.R., et al. 2003. Past exposure to densely ionizing radiation leaves a unique permanent signature in the genome. *Am. J. Human Genet.* 72:1162–1170.

Hlatky, L., Sachs, R.K., Vazquez, et al. 2002. Radiation-induced chromosome aberrations: insights gained from biophysical modeling. *Bioessays*. 24:714–723.

Leonard, A., Rueff, J., Gerber, G.B., et al. 2005. Usefulness and limits of biological dosimetry based on cytogenetic methods. *Radiat. Prot. Dosimetry* 115:448–454.

Levy, D., Vazquez, M., Cornforth, M., et al. 2004. Comparing dna damage-processing pathways by computer analysis of chromosome painting data. *J. Comput. Biol.* 11:626–641.

Mitelman, F., Johansson, B., and Mertens, F. 2004. Fusion genes and rearranged genes as a linear function of chromosome aberrations in cancer. *Nat. Genet.* 36:331–334.

Nelson, G.A. (2003). Fundamental space radiobiology. *Gravit. Space Biol. Bull.* 16:29–36.

Sachs, R., Levy, D., Hahnfeldt, P., et al. 2004. Quantitative analysis of radiation-induced chromosome aberrations. *Cytogenet. Genome Res.* 104:142–148.

Sachs, R.K., Arsuaga, J., Vazquez, M., et al. 2002. Using graph theory to describe and model chromosome aberrations. *Radiat. Res.* 158:556–567.

Sachs, R.K., Rogoff, A., Chen, A.M., et al. 2000. Underprediction of visibly complex chromosome aberrations by a recombinational-repair ("one-hit") model. *Int. J. Radiat. Biol.* 76:129–148.

Savage, J.R.K. (1998). A brief survey of aberration origin theories. *Mutat. Res.* 404:139–147.

Vazquez, M., Greulich-Bode, K.M., Arsuaga, J., et al. 2002. Computer analysis of mfish chromosome aberration data uncovers an excess of very complicated metaphases. *Int. J. Radiat. Biol.* 78:1103–1115.

Wong, S., and Witte, O.N. (2004). The bcr-abl story: bench to bedside and back. *Annu. Rev. Immunol.* 22:247–306.

Address reprint requests to:
Dr. Rainer Sachs
970 Evans Hall
Berkeley, OA 94720

E-mail: sachs@math.berkeley.edu