BIOINFORMATICS

Vol. 19 no. 7 2003, pages 851–858
DOI: 10.1093/bioinformatics/btg087

ALES: cell lineage analysis and mapping of developmental events

V. Braun[1,], R. B. R. Azevedo[2,3], M. Gumbel[1], P.-M. Agapow[2], A. M. Leroi[2] and H.-P. Meinzer[1]*

[1]Division of Medical and Biological Informatics, Deutsches Krebsforschungszentrum, Im Neuenheimer Feld 280, D-69120, Heidelberg, Germany, [2]Department of Biology, Imperial College, Silwood Park, Ascot, Berks SL5 5PY, UK and [3]Department of Biology and Biochemistry, University of Houston, Houston, TX 77204, USA

Received on July 29, 2002; revised on December 4, 2002; accepted on December 14, 2002

ABSTRACT

Motivation: Animals build their bodies by altering the fates of cells. The way in which they do so is reflected in the topology of cell lineages and the fates of terminal cells. Cell lineages should, therefore, contain information about the molecular events that determined them. Here we introduce new tools for visualizing, manipulating, and extracting the information contained in cell lineages. Our tools enable us to analyze very large cell lineages, where previously analyses have only been carried out on cell lineages no larger than a few dozen cells.

Results: ALES (A Lineage Evaluation System) allows the display, evaluation and comparison of cell lineages with the aim of identifying molecular and cellular events underlying development. ALES introduces a series of algorithms that locate putative developmental events. The distribution of these predicted events can then be compared to gene expression patterns or other cellular characteristics. In addition, artificial lineages can be generated, or existing lineages modified, according to a range of models, in order to test hypotheses about lineage evolution.

Availability: The program can run on any operating system with a compliant Java 2 environment. ALES is free for academic use and can be downloaded from http://mbi.dkfz-heidelberg.de/mbi/research/cellsim/ales.

Contact: CellO@dkfz-heidelberg.de

INTRODUCTION

The cell lineage of an organism is the sequence of cell divisions leading from a zygote to each differentiated cell, during development. The precise topology, cellular phenotypes and distribution of cell fates in a cell lineage—what Wood (1999) termed the 'cell lineage hieroglyphics'—encodes information about the sequence of molecular and cellular events that generated it (e.g. the activation and repression of particular genes, or the secretion of morphogens).

Cell lineage data has been collected from a wide range of animals, including insects, leeches, sea urchins, ascidians and frogs (reviewed in Moody, 1999). The cell lineage that has been most comprehensively described is that of the free-living nematode *Caenorhabditis elegans*, where the genealogy of every somatic cell has been completely elucidated from zygote to adult, for both hermaphrodites and males (Sulston and Horvitz, 1977; Sulston *et al.*, 1980, 1983). Although lineage analyses are routinely carried out in nematodes to elucidate developmental mechanisms and their evolution (Sternberg and Horvitz, 1981; Chalfie *et al.*, 1981; Sternberg and Horvitz, 1982; Sommer *et al.*, 1994; Fitch and Emmons, 1995; Wiegner and Schierenberg, 1998; Borgonie *et al.*, 2000), they tend to concentrate on small sublineages, comprising only a few cells and developmental events. This limit is imposed by the labour-intensive nature of the task and lack of appropriate software tools to ease this burden. Similarly, phenetic, cladistic and comparative methods were not routinely used by biologists until appropriate computer programs became available. Here we introduce tools for the mapping and analysis of developmental events in cell lineages of arbitrary size and complexity.

Our methods for evaluating cell lineages were inspired by two disciplines: algorithmic information theory and phylogenetics. A lineage can be viewed as an algorithm composed of a set of instructions carried out sequentially during development. We begin by introducing a method to identify the minimum number of instructions in a cell lineage based on computer science theory (Hopcroft and Ullman, 1979). This metric is analogous to the Kolmogorov complexity of an algorithm (Li and Vitányi, 1997).

*To whom correspondence should be addressed.

This approach of describing a lineage in terms of its 'building blocks' does not require any assumptions about developmental mechanisms, and therefore has the advantage that it can be applied to the lineage of any organism. However, this approach also has low explanatory power: by assuming that 'all instructions are equal' it does not allow us to identify which cell divisions are involved in terminal cell fate specification. We therefore introduce a more detailed method of analyzing cell lineages which depends on specific postulates concerning the nature of cell fate determination.

Parallels between evolution and development have often been noted, (Haeckel, 1866; Edelman, 1988; Raff, 1996; Frank, 1997) perhaps because they are both the outcome of branching processes. Just as extant species have arisen from the division of ancestral species, so the cells extant in an adult animal have arisen from the division of ancestral cells earlier in development. Under this analogy, the cell lineage would correspond to a phylogeny, a cell to a species, a cell fate to a character state, and cell division to speciation. Here we extend the analogy and propose that the methods used to identify character state changes in a phylogeny (for review, see Felsenstein, 1983) can be adapted to identify developmental events in a cell lineage by specifying a 'developmental' model of character evolution.

ALES (for a screenshot see http://mbi.dkfz-heidelberg.de/mbi/research/cellsim/ales can aid the analysis of cell lineages in three ways. First, it can evaluate the lineage quantitatively, using several descriptive metrics. Second, ALES allows the visualization of cell lineages. Users can get an overview, zoom in, get information about any cell, flip the left and right daughter cells, with or without the complete sublineage, edit the progeny of a cell, save data files or lineage diagrams. Finally, the software can generate cell lineages under a variety of different null models, providing a basis for the statistical analysis of lineage metrics and the simulation of lineage evolution.

SYSTEMS AND METHODS

Data

At startup ALES reads a lineage data file (Table 1) or a Simi BioCell database (Schnabel *et al.*, 1997). The topology is reconstructed using a unique lineage identifier for each cell, consisting of a string of 0s and 1s, where 0 stands for a left and 1 for a right daughter cell. Data may be saved in this format at any time. ALES also recognizes the standard notation for *C.elegans* embryonic development (Sulston *et al.*, 1983). All planes of division are considered equivalent. The user can assign an arbitrary number of discrete cell states to a cell, such as morphological attributes or gene expression patterns (Table 1).

Table 1. ALES lineage data file for the *C.elegans* male V6L.pap sublineage (Sulston *et al.*, 1980)

LineageID	Name	Neuron	egl-5	Function
00	hyp	−	−	Hypodermis
01000	R2A	+	−	Neuron
01001	X	−	−	Death
01010	R2B	+	−	Neuron
01011	R2St	−	−	Structural
011	hyp	−	−	Hypodermis
10	hyp	−	−	Hypodermis
11000	R3A	+	+	Neuron
11001	X	−	+	Death
11010	R3B	+	+	Neuron
11011	R3St	−	+	Structural
111	hyp	−	+	Hypodermis

The first line contains meta information about the file and describes the columns of the matrix. The first columns contains lineage IDs, the second a cell descriptor, and the following strings are classification names. After the first line, each line begins with an unique cell identifier, then an arbitrary descriptor, then the class for the corresponding classification. 'Neuron' and the expression of the Hox gene *egl-5* classifications, indicated by −/+ (Ferreira *et al.*, 1999). Cell 'function' is a non-binary classification.

Visualization

The active lineage and cell classification are displayed for examination by users. To the best of our knowledge, this ability is not present in any other software package. The lineage can be displayed according to different styles. The default 'Node' (Fig. 1) style draws the lineage using small circles as inner nodes and the cell fate of all tip cells. The '*C.elegans*' (Fig. 2A–C) style omits the inner circles and draws continuous lines instead. Users can zoom in or out of the lineage image. The lineage can be reshaped by collapsing (hiding) or rotating (switching the left and right daughters) the sublineage of a cell.

Editing

A lineage can also be edited directly or generated *de novo*. Furthermore, at each cell division, the left, right or complete sublineage can be removed. Removing both daughter cells of a particular cell requires the assignment of a cell fate for the remaining (terminal) cell. A differentiated cell can be made to divide again or its fate can be changed.

Evaluation

Several metrics describing the properties of a cell lineage have been implemented. Below we describe three of these in detail.

The *reduced rules* (RR) metric measures the minimum set of 'rules' or instructions required to reconstitute the entire lineage. The algorithm begins by recording each cell division in the lineage as a unique rule. For example, the observation that cell 1 divides into cells 2 and 3, can

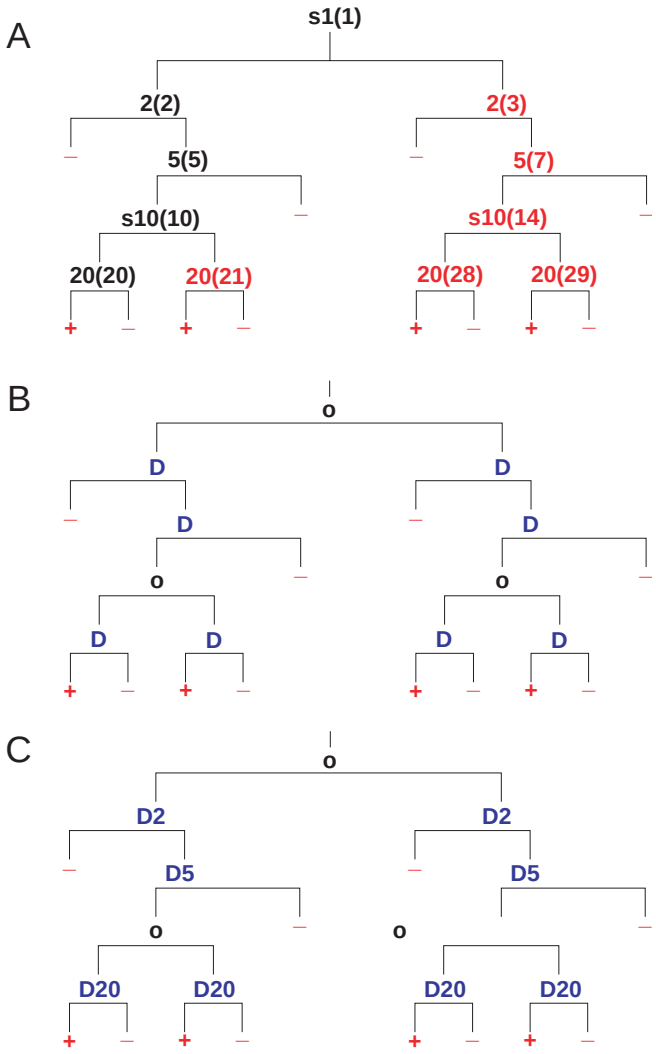

Fig. 1. Lineage of the *C.elegans* male V6L.pap cell, with neurons marked with '+' and other cells with '−'. Our analysis predicts that this lineage is generated by (**A**) 5 unique RR, marked 's1', '2', '5', 's10' and '20'. Numbers in parentheses are the RR numbers from the initial evaluation, where cell n gives rise to cells $2n$ and $2n+1$. RR marked in black have numbers that coincide with those from the initial evaluation. RR starting with 's' are symmetrical cell divisions. The four neurons in the lineage are specified by (**B**) eight DE (marked with 'D'), of which (**C**) three are UDE (marked 'D-2', 'D-5', and 'D-20'). Note that using a different classification (e.g. cell 'function', Table 1) would lead to a different evaluation.

be represented by the rule $(R1 \rightarrow R2, R3)$. If cell 2 is a differentiated cell, then $R2$ is replaced by the fate of the differentiated cell, such as $(R1 \rightarrow neuron, R3)$. A lineage containing n cells will be initially described by $n-1$ rules. Then, the algorithm identifies redundant rules, ignoring planes of cell division. For example, if $(R23 \rightarrow neuron, muscle)$ and $(R163 \rightarrow muscle, neuron)$ then we have $(R163 = R23)$. The algorithm eliminates all identical rules to generate the minimum necessary number of rules. For example, in the V6L.pap sublineage, the four terminal divisions give rise to one neuron and another cell each (Table 1) are marked as repeated occurrences of rule '20' (Fig. 1A). The number of RR for a lineage with n terminal cells ranges from a minimum of r, where r is the lowest whole number that satisfies the relationship $2^r \geqslant n$, to a maximum of $n-1$ (a stem cell lineage). We have not found an exact solution yet for the minimum number of RR for every value of n, but since the maximum number of cells that can be generated by r RR is 2^r, we have a lower bound $\mathrm{RR}_{min} \geqslant r$, where r is the lowest whole number that satisfies the relationship $2^r \geqslant n$.

In the absence of detailed knowledge about the molecular basis of cell states in a cell lineage, we postulate that cell fates are specified by the asymmetric segregation of intra-cellular fate 'determinants'. We then estimate the most parsimonious distribution of *determination events* (DE) that could account for the observed lineal origins of each phenotypic class. A DE is defined as a cell division which results in daughter cells containing different sets of cell-fate determinants, i.e. an asymmetric cell division. Each cell in the lineage can take one of a series of k fates $(f_1, f_2, \ldots, f_k)$. The k possible fates of a cell constitute a classification. A binary classification is a lineage with $k = 2$ fates. For example, in a given binary classification, each cell could be classified as either a neuron (f_1), or a non-neuron, (f_2). In a more detailed non-binary classification, cells might be divided into $k = 6$ fates: neuron (f_1), epidermis (f_2), lateral mesoderm (f_3), notochord (f_4), somite (f_5), and endoderm (f_6). A cell fate f_i is specified by a unique 'determinant' d_i. The determinant set Λ for each cell is calculated according to the following rules. If all descendants of a cell ν adopt the f_i fate (monoclone) then ν contains only determinants of type d_i. If the descendants of ν adopt more than one fate (polyclone), then ν contains the union of the determinant sets in the sublineage. Thus, the total number of DE for a lineage with n cells is given by

$$DE = \sum_{i=1}^{n} \theta(\nu_i)$$

where

$$\theta(\nu) = \begin{cases} 1, & \text{if } \Lambda(\nu_p) \cup \Lambda(\nu_q) \neq \Lambda(\nu_p) \cap \Lambda(\nu_q) \\ 0, & \text{if } \nu \text{ is a differentiated cell} \\ 0 & \text{else.} \end{cases}$$

where ν_p and ν_q are the two daughters of cell ν.

The algorithm iterates over the entire lineage, examining each cell division for the presence of a DE. The total number of DE is calculated and written to file. A list of

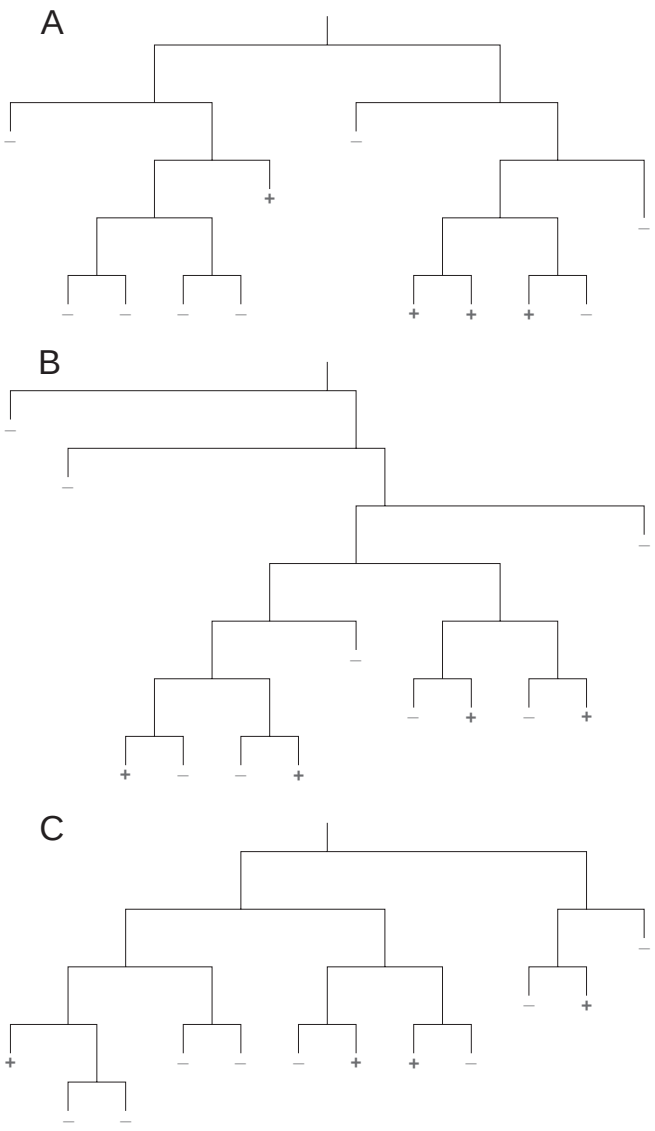

Fig. 2. Three lineages generated by different probability models, with the same cellular composition as that in Fig. 1. (**A**) The topology was retained and the cell fates were permuted (10 RR, 6 DE, 6 UDE). (**B**) Artificial lineage generated by the Markovian model (7 RR, 8 DE, 5 UDE). (**C**) Artificial lineage generated by the density-dependent model (8 RR, 6 DE, 4 UDE).

the cells with predicted DE can also be extracted. For example, this algorithm predicts eight neuronal DE in the V6L.pap sublineage (Table 1 and Fig. 1B). The number of DE found in lineages with n terminal cells and k different cell fates ranges from a minimum of $k-1$ (a lineage composed of k monoclones of each cell type), to a maximum of $n-1$ (a stem-cell lineage where the final cell division gives rise to cells of different fates).

Unique determination events (UDE) count non-redundant DE, i.e. RR that are also DE. For example, this algorithm predicts that of the eight neuronal DE, in the V6L.pap sublineage only three are UDE: 'D2' and 'D5' occur twice each, and 'D20' occurs four times (Fig. 1C). The maximum and minimum numbers of UDE for a lineage are the same as for the DE metric.

Other metrics implemented in ALES include the number of *symmetry breaking events*, defined as cell divisions that result in different sublineages, either in terms of topology or cell fate, the number and size distribution of *cell fate monoclones*.

Generation of artificial lineages

The number of possible lineages $L_{n,k}$ with n cells and k cell fates (ignoring the polarity of a given cell division) is given by the following recursion:

$$L_{n,k} = S_{n,k} - \sum_{i=1}^{k-1} \binom{n}{i} L_{n,k-i}$$

where $S_{n,k}$ is the S_n-recursion formula given by Harding (1971) with the difference that $S_1 := k$. For example, there are 451 different topologies ($k = 1$) with 12 cells like the V6L.pap sublineage (Table 1), and 495 097 different cell lineages with 12 cells and 1–11 neurons ($k = 2$). The above equation implies that the number of possible lineages increases exponentially with the number of cells (not shown). Therefore, it is not feasible to explore all conceivable cell lineages composed of more than a few cells. To interpret observed values and patterns in the lineage metrics, we need to study lineages generated under null models.

ALES can generate cell lineages with random-topologies, permuted-fates and mutated-fates. *Random-topology* lineages with n terminal cells can be produced according to three different null models. Under the *Markovian* model (Fig. 2B) lineages are formed such that, at each round of division, one cell is chosen at random from the existing cells and made to divide (Harding, 1971). After r rounds of cell division, the lineage contains $r+1$ terminal cells. The probability of adding two daughters at a terminal cell v is $p_v = 1/(r+1)$. The *density-dependent* model (Fig. 2C) allows each terminal cell of the same generation to divide with the same probability. At every round, a terminal cell has a probability of dividing equal to $p_v = (1/2)^g$, where g is the generation number of that cell such that the founder cell has $g = 0$, the first two daughter cells $g = 1$, and so on.

Permuted-fate lineages for a given cell lineage topology and cell fate composition are generated by reshuffling the cells to which the fates are assigned (Fig. 2A; (Maddison and Slatkin, 1991)).

Mutated-fate lineages are generated by 'mutating' the fate of each cell in turn with a given probability.

For example, the V6L.pap sublineage in Figure 1 might be submitted to fate-mutation with the following probabilities: $P(neuron \rightarrow neuron) = 0.9$, $P(neuron \rightarrow other) = 0.1$, $P(other \rightarrow other) = 0.8$, and $P(other \rightarrow neuron) = 0.2$. ALES can be used in batch mode. Any desired number of artificial cell lineages can be created and evaluated without user interaction.

Output

The lineage data can be saved in a simple textual format for editing, analysis and re-loading. A lineage drawing can be saved in .jpg format. The results of any analysis, e.g. the number of DE per lineage, can be saved as tab-delimited text files for further analysis. The cellular distribution of an evaluation metric can also be saved to file.

Cell list comparison

Cell lists can be compared using ALES. One use for this is for comparing a distribution of DE or RR with gene expression patterns from the literature. For example, the *C.elegans* database WormBase contains data on the expression patterns of hundreds of genes in individual cells or groups of cells (WormBase, release WS91, 23/11/2002 http://www.wormbase.org/db/searches/expr_search). In another application, the distributions of terminal fates or DE in the lineages of different species may be compared and used directly in comparative analyses.

Movie creation

Movies of lineage growth can be generated if data on cell positions and timing of cell divisions are available. ALES can use data sets generated with the Simi Biocell software (Schnabel *et al.*, 1997). Any group of cells (e.g. those predicted to undergo DE) can be highlighted in the movie. 2D and 3D movies can be recorded directly using ALES (Fig. 3). Any time resolution can be set. This provides a powerful tool for use in lineaging studies.

IMPLEMENTATION

ALES is implemented in *Java* 2 to ensure platform independence and has been tested on the MS Windows, Mac OS X and Linux platforms. Movie creation is based on the CellO-Framework (Gumbel *et al.*, 2000) and the *Java Media Framework API* http://java.sun.com/products/java-media/jmf. ALES can be extended with further evaluation algorithms or null models for artificial lineages using a plugin-mechanism.

Current developments include the true scaling of the time-axis of the lineage, if information on the timing of cell divisions is provided in the data set, and the creation of new lineage evaluation metrics based on deterministic finite automata and algorithms for the generation of artificial 'mutated-topology' lineages. As lineage data accumulates, connecting ALES to a DBMS should be considered.

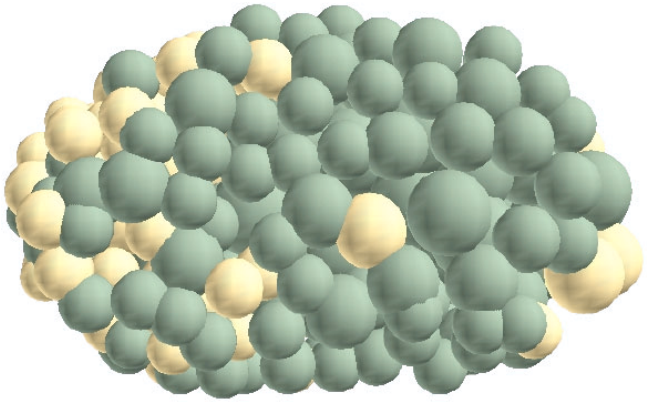

Fig. 3. *C.elegans* embryo at $t = 202$ min with predicted neuronal DE highlighted (frontal view). Most DE occur in the anterior part of the worm, where the pharynx and nerve ring will form.

RESULTS

Our analysis predicts that the neurons in the V6L.pap sublineage are specified by eight DE, of which three are unique (UDE). In addition to the three UDE, our analysis also predicts that there are two other unique RR. Do these inferred developmental events correlate with known developmental mechanisms? The second and third UDE, (marked 'D-5' and 'D-20', respectively in Fig. 1C) require the function of the basic-helix—loop-helix transcription factor *lin-32* and the E/*daughterless* ortholog *hlh-2* (Zhao and Emmons, 1995; Portman and Emmons, 2000). LIN-32 and HLH-2 appear to act as a heterodimeric complex that activates different sets of target genes at multiple steps in the ray sublineage (Portman and Emmons, 2000). This example, suggests that the algorithms introduced here successfully capture a fundamental aspect of the programming of the *C.elegans* lineage.

To demonstrate how ALES may be used to interpret the evolution of the developmental program predicted by the evaluation algorithms, we generated 1000 random-topology lineages with the same cellular composition as the V6L.pap neuronal lineage and calculated the different metrics for each run. Figure 4 shows the resulting frequency distributions for each metric. The analysis shows that the numbers of DE and UDE observed in the V6L.pap neuronal lineage (eight and three, respectively) are not significantly different from the numbers obtained in random lineages (2.5 percentiles: four and three, respectively). However, the observed numbers of RR (5) is significantly lower than the number obtained in random lineages (2.5 percentile: seven). This result suggests

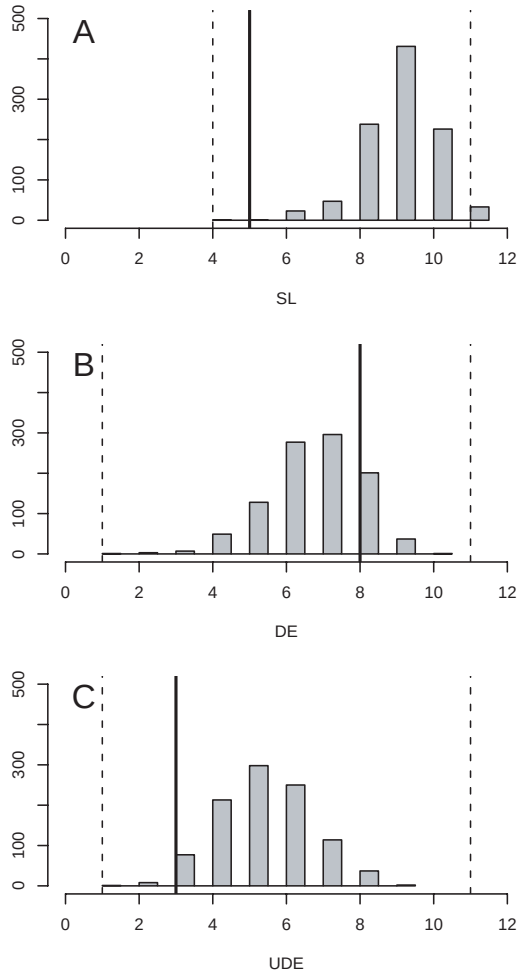

Fig. 4. Frequency distribution of the different lineage metrics in 1000 Markovian lineages with the same cellular composition as the cell lineage in Figure. 1. The solid lines mark the numbers of (**A**) RR, (**B**) DE, and (**C**) UDE calculated in the corresponding sections of Figure. 1. The dashed lines mark the maxima and minima possible of each metric for a lineage of the same cellular composition.

that the specification of neuronal fates in the V6L.pap sublineage requires fewer 'instructions' than are expected if the lineage had evolved at random.

The V6L.pap sublineage shows high DE but low RR and UDE. Does this mean that lineages with simultaneously low values for all the metrics are impossible? A correlation analysis of the metrics calculated in the 1000 random lineages suggests that such a constraint may exist. The numbers of DE and UDE are positively correlated (Pearson correlation coefficient, $r = 0.55$), as are the numbers of UDE and RR ($r = 0.65$), but the numbers of DE and RR are negatively correlated ($r = -0.10$). The same pattern of correlations was

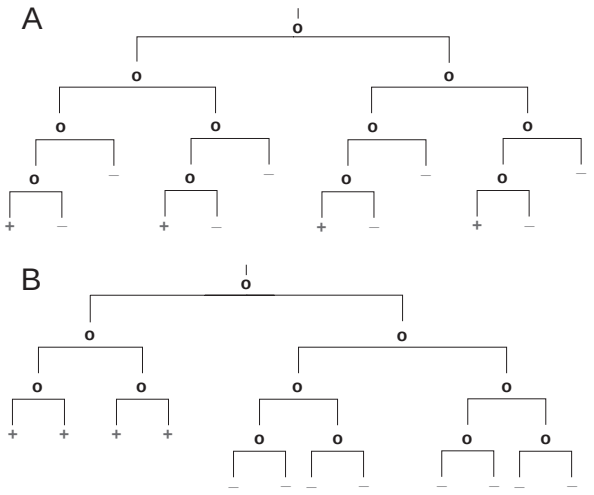

Fig. 5. Lineages with the same cellular composition as V6L.pap (Fig. 1) that minimize (**A**) RR (RR=4, DE=8, UDE=2), and that minimize (**B**) DE and UDE (RR=6, DE=1, UDE=1).

observed in 1000 random-topology density-dependent lineages ($r = 0.39, 0.66, -0.21$, respectively) and 1000 Permuted-fate lineages ($r = 0.37, 0.71, -0.29$, respectively). Furthermore, none of the 3000 random or permuted lineages showed $DE < 9 - RR$. Figure 5 shows examples of lineages found to minimize different metrics.

DISCUSSION

We believe that ALES fills a need in developmental biology. Although the example discussed above is relatively simple, it is representative of our understanding of cell fate specification in *C.elegans*: we know most about relatively small, terminal sublineages. We believe ALES provides a useful tool for identifying cell fate determinants for cell types which, like neurons, are specified by multiple determination steps (see Fig. 3). For example, (Kalb *et al.*, 1998) used a similar procedure to the DE algorithm to identify pharyngeal precursor cells in the *C.elegans* embryonic lineage. They then predicted which cells should express a pharyngeal determinant, and finally tested whether the expression of the transcription factor *pha-4* matched that expectation. This type of analysis can be simplified and expanded with ALES.

Another application of ALES is in the investigation of cell lineages from species closely related to model organisms such as *C.elegans* or the ascidian *Halocynthia roretzi* (Nishida, 1987), particularly in what we might term 'lineage annotation'. Such lineages are likely to be generated in the near future given current improvements in microscopy and lineage reconstruction techniques (Schnabel *et al.*, 1997; Yasuda *et al.*, 1999; Mochizuki, 1999; Bor-

gonie *et al.*, 2000; Heid *et al.*, 2002).

Although we have dealt primarily with examples from nematode lineages, this merely reflects the current quality and availability of data sets. ALES can be used to analyze any embryonic or post-embryonic lineages from any species. The only limit on the number of cells is available memory: cell lineages with up to 20 000 terminal cells have been generated successfully. The applicability of the DE metric to the lineage of a given species will depend on the degree to which it meets the assumptions of the 'determinants' model, the correctness of which is still an open question, even in *C.elegans* (Schnabel *et al.*, 1997; Labouesse and Mango, 1999; Moody, 1999). In any event, ALES will enable developmental biologists to generate testable predictions of the 'determinants' model, and hopefully will stimulate the invention of metrics incorporating other biological assumptions.

Each metric describes a different aspect of the lineage program. The RR metric is related to the number of 'sublineages', independent patterns of cell division and/or differentiation (Sulston and Horvitz, 1977; Sulston *et al.*, 1980; Chalfie *et al.*, 1981; Sternberg and Horvitz, 1981, 1982). The observation that the neurons of the V6L.pap sublineage are specified with fewer rules than random lineages with the same composition highlights the repeated use of 'modular' sublineages in *C.elegans* development (Sulston and Horvitz, 1977; Sulston *et al.*, 1980; Chalfie *et al.*, 1981; Sternberg and Horvitz, 1981, 1982). For example, the ray sublineage is executed independently by V6L.papap, V6L.pappp (Fig. 1) and 16 other ray precursor cells (Sulston *et al.*, 1980). Furthermore, the analysis of artificial lineages, suggested that there is an evolutionary trade-off between the numbers of RR and DE. This might be because the number of RR measures the number of *different* instructions required to generate a lineage, whereas the total number of DE is composed of iterations of particular RR (UDE). Biologically, the RR metric might be related to the complexity of the gene network underlying a lineage, whereas the DE metric might be related to the overall level of transcriptional activity associated with terminal cell fate specification.

It is unknown how powerful such metrics will turn out to be in deciphering the lineage 'hieroglyphics' of different species, or how closely the lineages generated under different null models resemble the actual products of evolution. As lineage data accumulates, the answers to these questions will reveal much about the development and evolution of organisms.

ACKNOWLEDGEMENTS

We thank Douglas Portman for helpful comments on the manuscript. RA was supported by a fellowship from the Foundation for Science and Technology (Portugal). P-MA was supported by Natural Environment Research Council (UK) grant GR3/11526 and Biotechnology and Biological Sciences Research Council (UK) grant 45/G09600.

REFERENCES

Borgonie,G., Jacobsen,K. and Coomans,A. (2000) Embryonic lineage evolution in nematodes. *Nematology*, **2**, 65–69.

Chalfie,M., Horvitz,H.R. and Sulston,J.E. (1981) Mutations that lead to reiterations in the cell lineages of *C.celegans*. *Cell*, **24**, 59–69.

Edelman,G.M. (1988) *Topobiology: An Introduction to Molecular Embryology*, Basic Books, New York.

Felsenstein,K. (1983) Parsimony in systematics: biological and statistical issues. *Ann. Rev. Ecol. Syst.*, **14**, 313–333.

Ferreira,H.B., Zhang,Y., Zhao,C. and Emmons,S.W. (1999) Patterning of *Caenorhabditis elegans* posterior structures by the *Abdominal-B* homolog, *egl-5*. *Dev. Biol.*, **207**, 215–228.

Fitch,D.H. and Emmons,S.W. (1995) Variable cell positions and cell contacts underlie morphological evolution of the rays in the male tails of nematodes related to *Caenorhabditis elegans*. *Dev. Biol.*, **170**, 564–582.

Frank,S.A. (1997) The design of adaptive systems: optimal parameters for variation and selection in learning and development. *J. Theor. Biol.*, **184**, 31–39.

Gumbel,M., Schnabel,R. and Meinzer,H.P. (2000) Analysis of cell migrations in *C.elegans* using computer simulations. In van Landeghem,R. (ed.), *Proceedings of the 14th European Simulation Multiconference*. SCS Publications, pp. 605–611.

Haeckel,E. (1866) *Generelle Morphologie der Organismen*. Reimer, Berlin.

Harding,E.F. (1971) The probabilities of rooted tree-shapes generated by random bifurcation. *Adv. Appl. Prob.*, **3**, 44–77.

Heid,P.J., Voss,E. and Soll,D.R. (2002) 3D-DIASemb: a computer-assisted system for reconstructing and motion analyzing in 4D every cell and nucleus in a developing embryo. *Dev. Biol.*, **245**, 329–347.

Hopcroft,J. and Ullman,J. (1979) *Introduction to Automata Theory, Languages and Computation*. Addison-Wesley, Reading, MA.

Kalb,J.M., Lau,K.K., Goszczynski,B., Fukushige,T., Moons,D., Okkema,P.G. and McGhee,J.D. (1998) *pha-4* is *Ce-fkh-1*, a *fork head* hnf-3 α, β, γ homolog that functions in organogenesis of the *C.elegans* pharynx. *Development*, **125**, 2171–2180.

Labouesse,M. and Mango,S.E. (1999) Patterning the *C.elegans* embryo: moving beyond the lineage. *Tr. Genet.*, **15**, 307–313.

Li,M. and Vitányi,P. (1997) *An Introduction to Kolmogorov Complexity and Its Applications*, 2nd edition, Springer, New York.

Maddison,W.P. and Slatkin,M. (1991) Null models for the number of evolutionary steps in a character on a phylogenetic tree. *Evolution*, **45**, 1184–1197.

Mochizuki,A. (1999) Estimating cell lineage from distributions of randomly introduced markers. *J. Theor. Biol.*, **197**, 227–254.

Moody,S.A. (eds) (1999) *Cell Lineage and Fate Determination*. Academic Press, San Diego.

Nishida,H. (1987) Cell lineage analysis in ascidian embryos by intracellular injection of a tracer enzyme. 111. Up to the tissue restricted stage. *Deb. Biol.*, **121**, 526–541.

Portman,D.S. and Emmons,S.W. (2000) The basic helix-loop-helix transcription factors LIN-32 and HLH-2 function together in multiple steps of a *C.elegans* neuronal sublineage. *Development*, **127**, 5415–5426.

Raff,R.A. (1996) *The Shape of Life: Genes, Development and the Evolution of Animal Form*. Chicago University Press, Chicago.

Schnabel,R., Hutter,H., Moerman,D. and Schnabel,H. (1997) Assessing normal embryogenesis in *Caenorhabditis elegans* using a 4D microscope: variability of development and regional specification. *Dev. Biol.*, **184**, 234–265.

Sommer,R.J., Carta,L.K. and Sternberg,P.W. (1994) The evolution of cell lineage in nematodes. *Development Suppl.*, 85–95.

Sternberg,P.W. and Horvitz,H.R. (1981) Gonadal cell lineages of the nematode *Panagrellus redivivus* and implications for evolution by the modification of cell lineage. *Dev. Biol.*, **88**, 147–166.

Sternberg,P.W. and Horvitz,H.R. (1982) Postembryonic nongonadal cell lineages of the nematode P*anagrellusl redivivus*: description and comparison with those of *Caenorhabditis elegans*. *Dev. Biol.*, **93**, 181–205.

Sulston,J.E., Albertson,D. and Thomson,J. (1980) The *Caenorhabditis elegans* male: postembryonic development of nongonadal structures. *Dev. Biol.*, **78**, 542–576.

Sulston,J.E. and Horvitz,H.R. (1977) Post-embryonic cell lineages of the nematode *Caenorhabditis elegans*. *Dev. Biol.*, **56**, 110–156.

Sulston,J.E., Schierenberg,E., White,J.G. and Thomson,J.N. (1983) The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Dev. Biol.*, **100**, 64–119.

Wiegner,O. and Schierenberg,E. (1998) Specification of gut cell fate differs significantly between the nematodes *Acrobeloides nanus* and *Caenorhabditis elegans*. *Dev. Biol.*, **204**, 3–14.

Wood,W.B. (1999) Cell lineages in *Caenorhabditis elegans* development. In Moody,S.A. (ed.), *Cell Lineage and Fate Determination*. Academic Press, San Diego, pp. 77–95.

Yasuda,T., Bannai,H., Onami,S., Miyano,S. and Kitano,S. (1999) Towards automatic construction of cell-lineage of *C.elegans* from nomarski dic microscope images. *Genome Informatics*, 144–154.

Zhao,C. and Emmons,S.W. (1995) A transcription factor controlling development of peripheral sense organs in *C.elegans*. *Nature*, **373**, 74–78.