Combining Sea Urchin Embryo Cell Lineages by Error-Tolerant Graph Matching

J. L. Rubio-Guivernau, M. A. Luengo-Oroz, L. Duloquin, T. Savy,
N. Peyriéras, P. Bourgine and A. Santos

Abstract — **Obtaining the complete cell lineage tree of an embryo's development is a very appealing and ambitious goal, but fortunately recent developments both in optical imaging and digital image processing are bringing it closer. However, when imaging the embryos (sea urchin embryos for this work) with high enough spatial resolution and short enough time-step to make cell segmentation and tracking possible, it is currently not possible to image the specimen throughout its all embryogenesis. For this reason it is interesting to explore how cell lineage trees extracted from two different embryos of the same species and imaged for overlapping periods of time can be concatenated, resulting in a single lineage tree covering both embryos' development time frames. To achieve this we used an error-tolerant graph matching strategy by selecting a time point at which both lineage trees overlap, and representing the information about each embryo at that time point as a graph in which nodes stand for cells and edges for neighborhood relationships among cells. The expected output of the graph matching algorithm is the minimal-cost correspondence between cells of both specimens, allowing us to perform the lineage combination.**

I. INTRODUCTION

CURRENT optical imaging technologies together with the use of fluorescent proteins [1,2] for labeling cell membranes and nuclei in live embryos, make feasible the objective of building a full cell lineage [3] for species such as the sea urchin.

Sea urchin embryo is a useful animal model system because of its phylogenetic position (deuterostomes, echinoderm), its rapid development and the complete transparency of its embryo that allows microscopic observation. An attractive goal would be to get the lineage encompassing the first 30h post fertilization. At that moment, the pluteus larva consists in ~1500 cells, and metamorphosis events start.

To this aim, 3D+t images of live embryos are acquired and processed to segment and track each of the individual cells [4,5]. However, with current experimental methods photo damage and mounting artefacts impose limits to the acquisition protocol. Acquiring images with the best compromise in terms of temporal and spatial resolution to

properly achieve cell tracking and preventing the embryo from moving does not allow imaging the whole development of a single embryo.

We thus used a multi-shot approach with a temporal tiling strategy to cover the whole developmental period by imaging different embryos in overlapping imaging sequences. With this approach each embryo is imaged during the selected time frame (~3h long), and each set of 3D+t images is used to build the corresponding cell lineage tree for that specimen within that time period [6]. Then *fusing* the lineages from two different embryos is achieved by finding their overlapping developmental period and selecting one or several time points to find the correspondence between every single cell.

For the initial tests, we chose a single time point in the embryo's development in which the cell count is exactly 108. We selected this stage because it appeared quite stable (see Fig. 1). The lineage at this stage is indeed known to be largely invariant [7]. More specifically, 4 cells called *small micromeres* at the vegetal pole of the animal are easily identified. We identified these 4 cells and they were taken as prior information to help the matching algorithm.

The main piece of information used for the matching process was the neighborhood relationships of the 108 cells. This information was obtained from the raw data by using a Voronoi diagram [8].
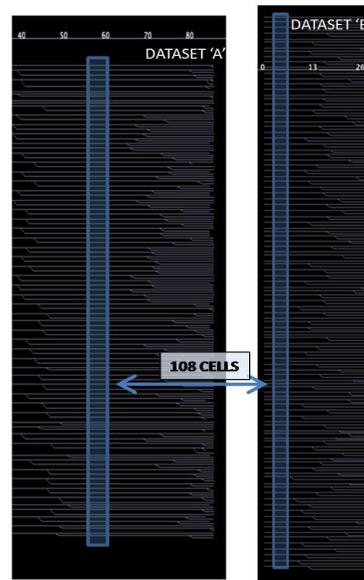

Fig. 1: Flat representation of the lineage tree for the two datasets (coming from different embryos). Time points are indicated on top of the graph. Cell divisions are seen as bifurcations in the tree. Although each of them covers a different time frame in the embryo's development, both include the 108 cells stage, which is highlighted in the figure (in blue).
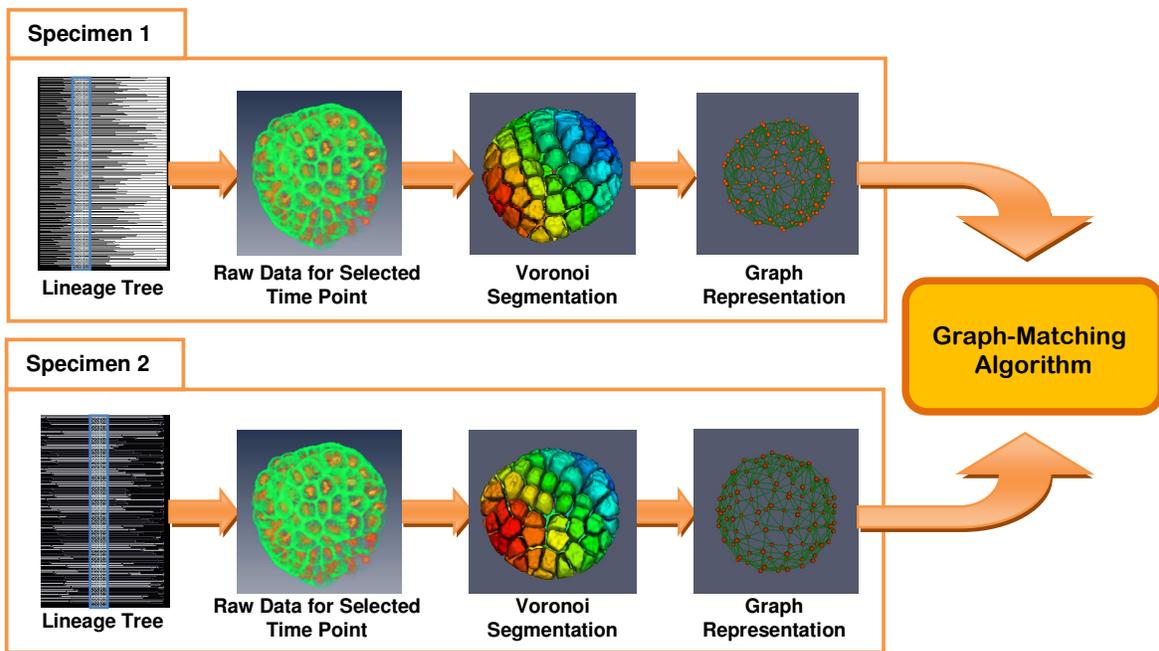
Fig. 2: General overview of the scheme proposed in this work

The whole process, from the selection of the time point in the two lineage trees to the output of the algorithm is outlined in Fig. 2.

II. DATASET DESCRIPTION

A. Image Acquisition

3D+t datasets were acquired using a Leica SP5 bi-photon laser scanning microscope equipped with a Leica 63x, 0.9NA W objective.

Embryos were injected at the one cell stage with mRNA encoding fluorescent proteins: H2B/mcherry (red fluorescent protein which is located in nuclei) and GFP-Ras (green fluorescent protein which is located in membranes).

Image resolution is 0.4805 x 0.4805 x 0.96 μm^3, which gives us images with 512x512x90 voxels (first embryo) and 512x512x101 voxels (second embryo).

The time step between 3D acquisitions was 1min 59s for the first data set, which covered from 3h45 to 6h45 post fertilization and 2min 13.774s for the second one, going from 5h45 to 9h post fertilization.

B. Data Preprocessing

After selecting the matching time point in each dataset, we took the information available in the 2-channel (one for each fluorescent protein used) 3D image and converted it into a graph form.

The most valuable piece of information to perform the matching is the cell neighborhood that has to be extracted from the raw images. Previous work [8] has shown that Voronoi diagrams provide a reasonable segmentation of the cell contours (Fig. 3), and as in the present case we do not need an extremely precise estimation of the cell shape, this approach seemed appropriate for our problem.

In order to obtain the neighborhood graph we take the raw data, we segment the outer and inner shells of the

embryo, and then we compute the Voronoi diagram, yielding each cell's neighbors. It is worth remarking that the neighbors obtained by this process define the Delaunay graph (which is the dual of the Voronoi diagram) of the nuclei (see Fig. 3).
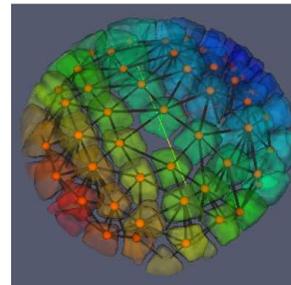

Fig. 3: Result of the Voronoi Segmentation (Delaunay graph overlaid with orange nodes and black edges).

III. ALGORITHM DESCRIPTION

A. Introduction

As explained above, we represent the information about cell neighborhood relationships in each dataset as a non-directed graph, where each node represents a single cell in the dataset, and edges between nodes stand for the contact between neighboring cells.

We started with two graphs, each one representing a specimen at the 108 cell stage. It should be noted that nodes numbering in the two specimen is a priori not related (i.e. node #1 in the first graph might represent the same cell as node #85 in the second graph). The algorithm's goal is to find the 1-to-1 correspondence providing the best match between edges of both graphs. (Fig. 4).

We did not expect to find an exact correspondence, as biological variability among different embryos gives rise

to differences in graphs connectivity. Even if we consider a single embryo, intra-specimen variability from one time point to the other as well as precision in segmentation strategies leads to 8% variability (see Table 1) in the cell neighborhood relationships throughout cell stage 108. This means that instead of an exact match, we need to search for the node correspondence preserving as many edges as possible.

As we cannot assume the same exact graph topology, instead of using the subgraph isomorphism approach (a classical NP-complete problem, for which some well-known algorithms exist [9]) we posed our problem as an *error-tolerant graph matching* [10] issue. This means that we had to define a function assigning a cost to each possible 1-to-1 correspondence between nodes of both graphs. A natural cost function definition in our case is the number of different edges remaining after re-numbering one of the graphs.

Original Graphs:

Matched Graphs:

1-to-1 Correspondence: [1,2,3,4,5,6,7,8] → [4,3,5,8,7,6,2,1]
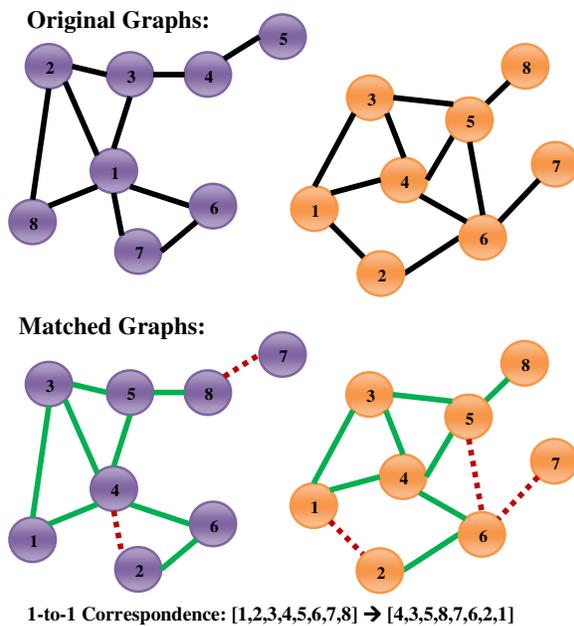
Fig. 4: Example of the graph matching algorithm. The two original graphs are shown in the top panel and the results of the matching strategy are shown in the bottom panel, with solid green lines for matched edges and dashed red lines for non-matched edges

B. Algorithm Details

It is easy to calculate how huge the search space for this problem is. Having 108 nodes on each graph, the number of possible different 1-to-1 correspondences is *factorial(108)*≈10^{174}. So the brute force approach was readily discarded.

A typical way to deal with this kind of problems is the A* algorithm [11], a best-first graph search algorithm which uses a priority queue to guide the search path, and defines priority of each partial solution as the sum of the accumulated cost and a heuristic estimation of the remaining cost. The heuristic must be optimistic, i.e. must underestimate the real cost to the goal, to maintain the optimality of the A* algorithm.

In our case, each node in the search space represents a partial (total if we already reached the goal) matching between cells from both specimens. The accumulated cost

is the number of edges that needs to be changed in order to match the already assigned cells, while a heuristic was defined which gives a lower bound on the number of edges to be changed for matching the remaining cells.

We started with an initial search node that matched only the 4 *small micromeres* from both embryos, and the algorithm followed the steps:

1. We add the initial search node to the priority queue
2. The first node in the queue is extracted
3. We check if the extracted node is the goal (i.e. if it matched all the cells from both embryos). In that case, we are done
4. A set of possible candidates is generated from the state of the extracted node
5. Some feasibility conditions are checked for each of the candidates, and those which pass are added to the priority queue
6. Back to step 2

For step 4, i.e. the selection of candidate matches for continuing a given search node, we follow a semi-local strategy. The unassigned cell from the first graph having the highest number of edges connecting it to the already matched cells is selected, and all the cells from the second graph that are connected with the set of matched cells in one or two steps are proposed to be paired with it. The motivation behind this strategy is to softly guide the algorithm while preserving good tolerance to biological variability between specimens.

Even with the use of the A* algorithm, the search space remains huge, and as it is well known in this situation the A* algorithm is usually limited by the space requirements of maintaining all the candidate solutions opened. In order to overcome this issue, we have implemented a number of heuristic feasibility conditions (step 5) that must be fulfilled before a candidate solution is added to the priority queue. These conditions take advantage of the extra information that we have with respect to a general graph, coming from the fact that our graphs are representing embryos with similar morphological structure. For instance we don't expect a cell in the vegetal pole of the first embryo to match a cell in the animal pole of the second embryo, so we imposed some limits in the distance between the already matched cells and the next pair of cells to be matched.

IV. RESULTS

As a first test for the algorithm, we matched two different time points coming from the same specimen, both with 108 cells (as already mentioned, cell stage 108 goes on for several time points). The algorithm was able to find the right correspondence between both graphs, and we found that about 8% (32 out of 410, see Table 1) of edges had changed. This justifies the use of the *error-correcting graph matching* approach.

	# Edges in Graph 1	# Edges in Graph 2	# Matched Edges
Intra-Specimen	408	410	378 (92%)
Inter-Specimen	408	376	279 (71%)

Table 1: Summary of intra and inter-specimen results

When trying to match graphs coming from two different specimens, to date best result gives a 71% of matched edges. Fig. 5 shows some images of this partial matching.

Naturally, the fact that we use the previously mentioned heuristic conditions for pruning the search space means that this result is not guaranteed to be optimal with respect to the metric used, as another path leading to a higher score matching might get cut at some point. Relaxing those conditions could help here, but the combinatorial nature of the search space makes aggressive pruning necessary to avoid an explosion of the search space.

Checking the matching by eye is actually hard, as inter-specimen variability is high enough to make difficult the choice of equivalent cells between two different embryos. However, visual comparison of the result (Fig. 5) shows reasonable coherence between the *large micromeres* (8 cells around the *small micromeres*) as well as between the *macromere* rings in both embryos.
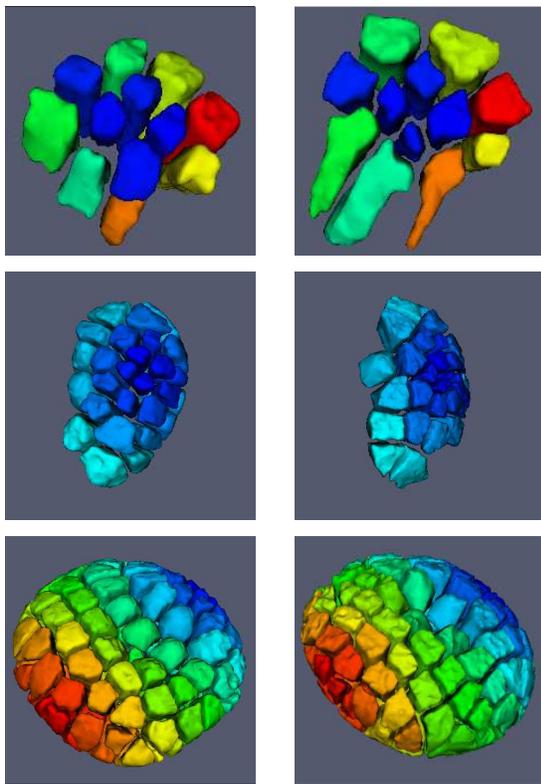

Fig. 5: **Left hand:** First embryo,　　　**Right Hand:** Second embryo
Top panel: Comparison of the 12 first cells, the 4 *small micromeres* (prior information) are shaded in blue, while the 8 *large micromeres* appear in other colors (same color on both sides for matched cells)
Middle: First 28 match cells in both embryos. For the first embryo the colors change according to the distance to the 4 *small micromeres.* The second embryo is colored according to the cell matching.
Bottom: Full embryo matching, observed from the animal pole. Same color map as images in the middle panel

V. DISCUSSION AND FUTURE LINES

In this work we have implemented an A* search algorithm applied to finding the best matching between cells from two different sea urchin embryos observed at the same developmental stage. This strategy allows the combination of cell lineage trees obtained from different specimens.

Results obtained here are promising, but there are some

aspects that might be improved in order to enhance both the algorithm's performance and its general usefulness.

For instance, the overlap between lineages from different specimens is longer than one single time-step (Fig. 1), so using the information from several time steps in parallel could help the algorithm finding the best match.

We imposed in the present study a 1-to-1 correspondence that seemed reasonable with the overall low variability observed at the 108 cell stage. However, for a broader range of application and for example, to use the same strategy for vertebrate embryos, it would be useful to perform a *region to region* matching to deal with non-deterministic lineages in later stages of development. In other words, we could specify the matching between cell populations in the two embryos and the algorithm would restrict the search to obey this constraint.

Validation of the algorithm's results might also be improved by again taking advantage of the fact that we have several 3D volumes of each specimen having the same number of cells, which allows us to perform several matching and cross-check the results.

We conclude that the chosen approach is well-suited for our problem and can be used to build a full lineage from several 3D+t images acquired in a multi-shot approach.

REFERENCES

[1] B.N.G. Giepmans, S.R. Adams, M.H. Ellisman, and R.Y. Tsien, "The Fluorescent Toolbox for Assessing Protein Location and Function," *Science*, vol. 312, Apr. 2006, pp. 217-224.

[2] S.G. Megason and S.E. Fraser, "Digitizing life at the level of the cell: high-performance laser-scanning microscopy and image analysis for in toto imaging of development," *Mechanisms of Development*, vol. 120, 2003, pp. 1407–1420.

[3] Z. Bao, J.I. Murray, T. Boyle, S.L. Ooi, M.J. Sandel, and R.H. Waterston, "Automated cell lineage tracing in Caenorhabditis elegans," *Proceedings of the National Academy of Sciences*, vol. 103, 2006, pp. 2707–2712.

[4] C. Melani, N. Peyrieras, K. Mikula, C. Zanella, M. Campana, B. Rizzi, F. Veronesi, A. Sarti, B. Lombardot, and P. Bourgine, "Cells tracking in a live zebrafish embryo," *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, 2007, pp. 1631–1634.

[5] P. Frolkovic, K. Mikula, N. Peyrieras, and A. Sarti, "Counting Number of Cells and Cell Segmentation Using Advection-Diffusion Equations," *Kybernetika*, vol. 43, 2007, p. 817.

[6] L. Duloquin, N. Peyrieras, et al. *Unpublished Data.*

[7] R.A. Cameron, B.R. Hough-Evans, R.J. Britten and E.H. Davidson, "Lineage and fate of each blastomere of the eight-cell sea urchin embryo," *Genes & Development*, vol. 1, Mar. 1987, pp. 75-85.

[8] M. Luengo-Oroz, L. Duloquin, C. Castro, T. Savy, E. Faure, B. Lombardot, R. Bourgine, N. Peyrieras, and A. Santos, "Can voronoi diagram model cell geometries in early sea-urchin embryogenesis?," *Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on*, 2008, pp. 504-507.

[9] J.R. Ullmann, "An Algorithm for Subgraph Isomorphism," *J. ACM*, vol. 23, 1976, pp. 31-42.

[10] B. Messmer and H. Bunke, "A new algorithm for error-tolerant subgraph isomorphism detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 20, 1998, pp. 493-504.

[11] E.P.N.J. Nilsson, *Principles of Artificial Intelligence.*