The invariant cell lineage and cell fate of *Caenorhabditis elegans* provide a unique opportunity to decode the molecular mechanisms of animal development. To exploit this opportunity, we have developed a system for automated cell lineage tracing during C. elegans embryogenesis, based on 3D, time-lapse imaging and automated image analysis. Using ubiquitously expressed histone-GFP fusion protein to label cells/nuclei and a confocal microscope, the imaging protocol captures embryogenesis at high spatial (31 planes at 1 μm apart) and temporal (every minute) resolution without apparent effects on development. A set of image analysis algorithms then automatically recognizes cells at each time point, tracks cell movements, divisions and deaths over time and assigns cell identities based on the canonical naming scheme. Starting from the four-cell stage (or earlier), our software, named *STARryline*, can trace the lineage up to the 350-cell stage in 25 min on a desktop computer. The few errors of automated lineaging can then be corrected in a few hours with a graphic interface that allows easy navigation of the images and the reported lineage tree. The system can be used to characterize lineage phenotypes of genes and/or extended to determine gene expression patterns in a living embryo at the single-cell level. We envision that this automation will make it practical to systematically decipher the developmental genes and pathways encoded in the genome of *C. elegans*.

**embryogenesis** | **imaging** | **image analysis algorithms**

The nematode *Caenorhabditis elegans* offers a chance to understand development in molecular detail at the level of the individual cell with temporal resolution of a fraction of the cell cycle. The embryo develops from 1 to 558 cells in just 13 h via a fixed, known lineage, with a fixed relationship between the lineage history and the fate of a cell (1). The anatomy of the 959 adult somatic cells has been reconstructed with serial electron microscopy, defining the synaptic connectivity of the nervous system (2). The complete genome sequence (3) contains the recipes for the full catalog of RNA and protein molecules, along with the signals that dictate their use.

Studies exploiting these attributes have already led to the discovery of programmed cell death (4, 5), insights into organ formation (6–8), and elucidation of fundamental signaling pathways (9), including key pathways in early embryogenesis (10, 11). Microarray and serial analysis of gene expression data combined with homeotic mutants (12, 13), RNA enrichment methods (14), or FACs sorting of individual cells (15) reveal active genes within particular cells or stages of development. In situ hybridization (16) can localize mRNAs to particular stages and tissues.

However, the expression data often lack spatiotemporal resolution or are limited to a single type of cell at discrete time points. Assigning expression to individual cells from *in situ* images based on the fixed worms is difficult even for highly trained scientists, and automated cell recognition has been successful only for embryos at the eight-cell stage or earlier (17).

In contrast, GFP and other fluorescent reporters allow gene expression to be visualized continuously in the living embryo, potentially providing exquisite spatial–temporal resolution. Exploitation of the power of the system, however, relies on anatomical expertise for the interpretation of the images. The fixed lineage might provide a substitute for anatomical expertise, because knowledge of the lineage is tantamount to knowledge of the anatomy at the individual cell level. The development of 3D, time-lapse (4D) microscopy (18) greatly simplifies the task of lineage tracing, using the stored images to reconstruct essentially the whole embryonic lineage from a single embryo. Various computer programs have been developed to facilitate the analysis; the most widely used is the proprietary SIMI BIOCELL (19). An experienced user can produce one lineage in a week with sustained effort up to the point that movement starts (1).

Recently, Hamahashi et al. (20) reported an automated algorithm to identify and track nuclei by using 4D differential interference contrast imaging. Differential interference contrast image analysis relies on detection of the variation of texture between the nuclei and the cytoplasm in the image, which becomes increasingly more difficult as cells divide and become smaller. Furthermore, nuclei disappear during mitosis when the nuclear envelope dissolves, which increases the difficulty of assigning newborn cells to mother cells, especially when neighboring cells go through synchronized divisions. As a result, it can only trace up to the 24-cell stage.

Fluorescence microscopy provides an alternative to differential interference contrast imaging. When GFP is expressed as a histone fusion, the brightly labeled nuclei contrast strongly with dark cytoplasm. GFP–histone fusions also vividly label mitotic figures during cell divisions, providing rich timing and morphological information that can be used to match newborn cells to their mothers. Here the challenge has been to reduce excitation light exposure to a level compatible with normal development while still imaging all of embryogenesis at a sufficient frequency (W. Mohler, personal communication, and J. Waddle, personal communication).

We have developed protocols that produce 4D images of histone–GFP fusion labeled embryos with high temporal resolution and no apparent changes in development. In turn, we have developed a set of algorithms to automatically recognize nuclei and trace the lineage through 350 cells. The system should facilitate lineage tracing of mutants or RNA interference affecting embryonic development and could also be used in combination with a second fluorescent tag to trace gene expression with single-cell and minute-time resolution. In addition, the effects of RNA interference or mutation on the expression of other genes might be measured at high spatiotemporal resolution using appropriately labeled strains. We describe here the imaging protocols, the algorithms, and the system performance on test embryos. We have also developed a viewing/editing tool, ACETREE (T.B. and R.H.W., unpublished data).

**Results**

**Imaging.** Our imaging strategy minimizes the exposure of the embryo to excitation light while maintaining separation of GFP

Conflict of interest statement: No conflicts declared.

Freely available online through the PNAS open access option.

© 2006 by The National Academy of Sciences of the USA

---

*To whom correspondence should be addressed. E-mail: waterston@gs.washington.edu.*
signal and background noise, which was achieved by creating a worm strain that is brightly and ubiquitously labeled throughout embryogenesis and by carefully adjusting microscope settings, as discussed below.

We constructed RW10006, a strain that expresses histone H3.3::GFP driven by the H3.3 promoter and histone H2B::GFP driven by the germline promoter pie-1 (21). H3.3::GFP is expressed throughout embryogenesis and provides very strong GFP signals, especially in the AB, MS, and E lineages. H2B::GFP compensates for the dispersal of H3.3::GFP at mitosis for the first four rounds of embryonic cell divisions.

We found that confocal microscopy allowed the finest control of exposure, minimizing radiation damage to the embryo. We offset image quality degradation deeper in the sample by increasing the excitation intensity throughout the depth of the embryo. Similarly, we exploited the increase of GFP signal per pixel with time as the nuclei become more compact by reducing the excitation light later in development. Other parameters, such as pinhole size, scan speed, y-axis resolution, and line averaging were adjusted to improve signal-to-noise ratio with low exposure.

The combination of labeling and imaging parameters provides sufficient spatial and temporal resolution of the nuclei. At each time point, we collect images from 31 focal planes spaced 1 μm apart. Nuclear diameters range from ~11 μm in the early embryo to ~3 μm in later stages; hence, each is always represented in at least three planes. Similarly, we collect a stack of images every minute. Mitosis typically takes at least 4 min, so we have multiple time points at which to observe condensed mitotic chromosomes, an indicator of impending cell division. This temporal and spatial resolution is comparable with that used in differential interference contrast recordings (19). With these settings and parameters, gastrulation and morphogenesis take place on time, the embryo hatches on time with normal morphology and movement, and the resulting lineage is the same as the wild type (Movie 1, which is published as supporting information on the PNAS web site). Minimizing exposure levels is important; even a 2-fold increase in exposure results in hatching failure in some of the embryos.

**Image Analysis.** Automatic lineage determination in a 4D image series with labeled nuclei requires identification of the individual nuclei at each time point; tracking of the nuclei from one time point to the next; and recognition of the relative orientation of each pair of daughter cells to relate the automated lineage to the reference lineage. We have developed algorithms to perform each of these functions (see Methods for details).

**Nuclear identification.** Before morphogenesis, nonmitotic nuclei in *C. elegans* embryos are spherical. Although their diameters decrease upon each round of cell division, the nuclei at a given time are approximately the same size, especially at early stages. In addition, nuclei are well separated in the images for most of embryogenesis (Fig. 1). These characteristics allowed us to develop a simple but effective solution to nuclear identification. We assume that the centroids of the spherical nuclei will be maxima of local signal. The local signal associated with a given pixel is defined as the sum of intensity of all pixels within a cube centered at that pixel where the dimensions of the cube correspond to the expected nuclear size at the given time point. Candidate nuclei are then defined as the local maxima with the constraint that only one nucleus can be chosen within a spatial range based on the expected nuclear diameter.

After identifying the centroids, we generate an optimal spherical representation of nuclei. The nuclei are initially predicted to be the same size as the expected nuclear size. To optimize size, we compare the average signal intensity at the annulus of the sphere to that of the whole sphere. If the annulus intensity is below or above an empirically defined fraction of the total intensity, the sphere is shrunk or expanded, respectively. We then recalculate the centroid location as the center of gravity of the sphere using pixel intensities as weights. These two steps are iterated until both size and position converge. Besides providing more realistic representations for individual nuclei, the optimization also allows the expected nuclear diameter to be updated over time. The expected diameter is only specified for the first time point; for each later time point it is set as the average of the optimized sizes from the previous time point.

The above method may miss the closely spaced nuclei at later stages of embryogenesis. To find missed nuclei, we repeat nuclear identification with a smaller expected size and reconcile the results.

Our local maximum approach for nuclear identification is related to the watershed algorithm of object counting and is relatively insensitive to random noise in the images compared to the edge-based identification methods (22). However, although random noise does not interfere with the identification of real nuclei, it can generate local maxima that do not correspond to real nuclei; hence, lead to false positives in nuclear identification. To minimize these problems, we have developed a noise-filtering algorithm based on a low-pass filter (22) and histogram-based thresholding. The capacity to process noisy images allows us to reduce total light exposure.

**Nuclear tracking.** Tracking nuclei over time in a time-lapse image series requires matching the identified nuclei at one time point to those in the next. The biology dictates three possible kinds of match: one to one, one to two, or one to none, corresponding to movement, cell division, and cell death.

To track nuclei without relying on normal morphology or secondary labels, we image frequently and apply a nearest neighbor/minimal movement algorithm (23) to match nuclei: A nucleus at a given time is matched to the closest one in the previous time point (Fig. 2). This algorithm is highly accurate except during some mitoses when the telophase daughter nuclei are sent to the distal ends of the cell. In these cases, the newborn daughter nuclei can be closer to a neighboring nucleus than to the mother.

To accommodate these cases, we relaxed the minimal movement algorithm, when a nucleus at time *t* has a putative match to more than one nucleus at *t* + 1, indicating cell divisions. For each nucleus involved at time *t* + 1, we consider a small number of potential matches whose distances are comparable with that of the nearest match, rather than the nearest match alone.
A wild-type lineage up to the 350-cell stage produced by automated analysis followed by minor editing. The tree agrees with the canonical lineage (1).

Validation. We tested the efficiency and accuracy of the system on 20 embryos. The output lineages were visually evaluated against the images. We detected errors by first identifying and correcting obvious deviations from the wild-type lineage, such as apparent cell deaths and unusual cell division timing. To make sure that all errors were detected, we systematically verified that each cell division in the corrected lineage was supported by the images. This approach would not detect swapping of nuclei (i.e., misidentifying nucleus A as B and simultaneously B as A), unless the two have a different sublineage tree topology. We looked for such errors by visually tracking nuclei at each time point in various embryos but failed to find any. Hence, swapping errors must be extremely rare. The inspection and correction were facilitated by an interactive graphic interface, ACETREE. A resulting lineage is shown in Fig. 3.

Technically, three kinds of errors can occur in lineaging: failure to identify a nucleus (false negatives), identification of a nonexisting nucleus (false positives), and incorrect matches in tracking. The rate for each type of error is shown for different stages of development in Fig. 4. The overall error rate is low at the earlier embryonic stages. From the four-cell to 194-cell stage, or the third to the eighth round of cell division, the cumulative overall accuracy is >99%. The accuracy of cell division tracking is 97% (Fig. 4C). As the embryo develops from the 194-cell stage...
to the 350-cell stage (through the ninth or second-to-last round of embryonic cell division), errors become more frequent.

The majority of the errors are false negatives in nuclear identification, which begin increasing near the 194-cell stage. The increase of the false-negative rate has three causes. First, the C, D, and P4 lineages in RW10006 have lower expression of histone–GFP than the AB, MS, and E lineages. Under our imaging settings, some of these nuclei occasionally fall below the noise cutoff when they are at the bottom of the embryo. Second, nuclei at the top of the embryo are flattened and crowded because of mounting, to the extent that they are not always recovered by the iterative nuclear identification. This crowding becomes a significant source of error after the 150-cell stage, when nuclei first enter the very top of the embryo. Third, neighboring cells may have different levels of GFP expression. When two nuclei are close to each other, the local signal density of the weaker nucleus may blend into the edge of the high signal from the bright nucleus and, thus, no longer correspond to a local maximum.

Our determination of the axis of division sometimes is at odds with that of the canonical lineage. For example, we observe that Caap daughter nuclei align principally along the dorsal–ventral axis. In all cases examined, the dorsal daughter exhibits the distinct asymmetric sublineage of the canonical anterior daughter and on average slightly tilts to the anterior. However, in some embryos, there is no anterior–posterior separation between the daughters, and, in some extreme cases, the daughter with the canonical anterior fate is on the posterior side of her sister. In these embryos, neither of our naming methods picks the anterior–posterior axis as its choice, presumably because the canonical choice of axis is based on the initial spindle orientation instead of relative positions of newborn nuclei (J. Sulston, personal communication). Whereas nuclear positions generally serve as a good proxy, in a few instances rapid spindle rotation produces inconsistencies between the two methods. Up to the 194-cell stage, the convention-based method consistently picked the canonical choice in all test embryos for all but 27 of the 193 divisions. For these divisions, we were able to find simple alternative rules that consistently distinguish the sister cells (Table 1, which is published as supporting information on the PNAS web site). In the example of the Caap division, the rule requires that the dorsal daughter equals to the canonical anterior daughter. However, these rules are compiled based on the 20 test embryos and have not been further tested on additional ones.

The logic-based method is intended to be an alternative for analyzing mutant lineages for which both the lineage and spindle orientation could be altered from the wild type. Hence, we did not try to complement the logic-based method with special rules derived from wild-type embryos.

Discussion

We have described a system for embryonic cell lineage tracing in C. elegans using 4D GFP imaging and automated image analysis. The lineaging algorithms are not yet error-free, but the resulting lineage can be readily edited to the correct form with ACETREE. Editing requires ~2 h for the 194-cell stage, and an additional 2–8 h for the 350-cell stage, depending on slight variations in image quality. That is, with a day of work, one can reliably obtain the lineage, together with the position of each nucleus at every minute for nine of the 10 rounds of embryonic cell division.

The development of the system relied on the synergy of three components: genetic manipulation of the worm, microscopy, and algorithm development. The bright and nearly uniform nuclear GFP images in turn enabled us to develop efficient and effective image analysis algorithms that trace the lineage. Further improvement of the system will continue to rely on this synergy. For example, false negatives in nuclear identification are a major error in current automated lineaging. Improving GFP expression in the C, D, and P4 lineages will reduce the
number of false negatives in two ways. First, GFP expression increases signal-to-noise ratio for these dimmer nuclei; second, it also allows the use of a smaller pinhole size during imaging, which, at the cost of overall signal intensity, increases image resolution and enhances separation between neighboring nuclei. In parallel, these shortcomings would be addressed through improved algorithms.

Our nuclear identification and nuclear tracking algorithms do not rely on knowledge of the wild-type lineage. Therefore, it can also be used to trace mutant lineages (data not shown). In addition, the system can be extended to map gene expression in living animals by using a second fluorescent protein, such as DsRed or a variant (24) in promoter or protein fusions. By tracing the lineage, gene expression would be mapped to individual cells with potentially minute-level resolution. Combined with the ability to analyze lineage phenotypes, the system could rapidly reveal the function of a gene in development. The system could also be extended to labeled proteins, following their location within cells. We expect that, with further improvement, the automated lineageing system will catalyze the systematic analysis of the developmental genes and pathways encoded in the genome of *C. elegans*.

**Methods**

**Worm Strains.** The *pie-1*::H2B-GFP strain, AZ212, was provided by the Caenorhabditis Genetics Center (University of Minnesota, Minneapolis). Briefly, for the generation of H3.3::GFP (S.L.O., J. R. Priess, and S. Henikoff, unpublished data), GFP was fused to the coding region of *his-72* at the C terminus. The *his-72::GFP* ORF was flanked by 1 kb of the 5′ and 3′ UTRs of *his-72* respectively. The entire construct was then cloned into a plasmid containing the *unc-119* marker, and the transgenic strain was generated by microparticle bombardment (21). RW10006 was derived by crossing the above two strains and selecting for a double homozygote.

**Imaging Protocol.** We used a Zeiss LSM 510 confocal microscope with a 488-nm argon laser. To enhance signal, we used a relatively large pinhole size (2 Airy units), high detector gain (1,000–1,100), and an amplifier gain of 1.1. To minimize signal loss, we used the HFT 488 filter in the light path. To reduce random noise, we averaged over two scans. To balance between noise and imaging speed, we used a scan speed of 8 and 3-gigabyte average over 10 neighboring time points.

**Nuclear Tracking Algorithms.** Upon relaxation of the minimal movement algorithm, a greedy approach is taken to prevent excessive inclusion of potential matches. First, potential matches that are more than a certain distance away (cutoff measured as a fraction of the expected nuclear size) are discarded based on the frequent imaging. Second, a nucleus at time *t* can be matched to no more than three nuclei at time *t* + 1. When violated, we retain three of the matches based on these rules: (i) nuclei at time *t* + 1 that have no other potential matches at *t* have precedence over those that do and when this condition does not apply, and (ii) a closer nucleus has precedence over one that is farther away. Our scoring system for choosing the best match includes two morphology detectors of mitotic nuclei, both of which aim to detect deviation from the stereotypical spheres of interphase nuclei with more uniform pixel intensity. The first detector measures aggregation of bright pixels (caused by chromosome condensation and aggregation) by measuring the frequency of a bright pixel occurring next to a dark pixel. Bright and dark pixels are defined as above or below the average pixel intensity within the nucleus. The second detector searches for the rod-like shape of metaphase to telophase nuclei (most of the metaphase plates during *C. elegans* embryogenesis are parallel to the *z* axis and, hence, are rod-shaped in *xy* cross sections). At the center plane of a nucleus, the circle is divided into eight equal segments, and the total intensity of the four pairs of opposing segments is calculated. If a nucleus is rod-shaped, the maximum intensity among the four would be significantly higher than the minimum.

We chose to use these detectors rather than more explicit spatial models of mitotic figures because dividing nuclei can be imaged at any stage of mitosis and are frequently captured in atypical shapes.

A division is scored as follows. If the mother nucleus is considered mitotic, a score of 1 is given; otherwise, the score is 0. For each daughter considered mitotic or significantly smaller than the mother (a difference in diameter of >3 pixels), +1 is added to the score; otherwise, the score is −1. If a daughter’s total intensity is at least 30% less than the mother’s, the score is
+1; otherwise, the score is −1. If the difference in diameter between the two daughters is <3 pixels, the score is +1; otherwise, the score is −1. If the relative difference in total intensity is <10%, the score is +1; otherwise, the score is −1. A nondividing match is scored as follows. Starting with a score of 0, if the difference in diameter is <3 pixels, +1 is added; otherwise, the score is −1. If the relative difference in total intensity is <30%, the score is +1; otherwise, the score is −1. If neither nucleus is considered mitotic, the score is +3. A one to zero match (apparent cell death) is scored −3.

In terms of temporal information, if a nucleus is considered mitotic but no division occurs within 5 min (before or after), a score of −3 is given. If a nucleus divides twice within the allowed minimal cell cycle length (10 min), the one with a lower morphology score (see above) is given a score of −1, and the other is scored +1.

The tracking algorithm also includes a component to correct sporadic errors in nuclear identification. If a nucleus at time t does not have a match at time t + 1 (indicating cell death or a false negative), a nucleus is tentatively added at the same position at time t + 1. If this tentative nucleus is matched to a nucleus at t + 2, we assume a nucleus was missed, and the tentative nucleus is permanently added. Otherwise, the tentative nucleus is removed, and a cell death is declared. Similarly, if a nucleus exists for only one time point (division followed immediately by death), it is considered a false positive and removed. The tracking algorithm runs iteratively until no adjustment to nuclear identification is needed.

Cell Naming Algorithms. During C. elegans embryogenesis, the embryonic axes are first defined at the four-cell stage. The embryo subsequently goes through two rotations along the anterior–posterior axis, first during gastrulation and again during morphogenesis. The direction of the first rotation is stereotypical: Viewed from the anterior end of the embryo, the rotation is clockwise. The direction of the second rotation appears random (1).

Following this scheme, we have chosen first to determine the axes at the four- to eight-cell stages (see above) and then follow the axes through the embryo rotations. Because the vast majority of the cells are named based on the anterior–posterior axis, which do not change over time, we have opted for a simple implementation to track the other two axes: We rotate these axes 90° once at the 51-cell stage soon after gastrulation begins. We have not implemented the second rotation during morphogenesis, because the number of cells to be named in the left–right or dorsal–ventral directions in the wild-type embryo is below the error rate of our current automated analysis at that stage.

In addition to using the canonical choice of axis to distinguish sister cells, we have developed a standardized, logic-based rule to analyze non-wild-type lineages. If the distance between two centroids along the anterior–posterior axis is greater than one-eighth of the sum of the two nuclear diameters and greater than 5 pixels, the sister cells are named by the anterior–posterior axis; otherwise, the left–right or dorsal–ventral axis is used. To choose between these two axes, the one that is not parallel to the z axis of the image has precedence, provided the distance along that axis is greater than the above cutoff. The cutoffs and precedence are set empirically to best approximate the canonical choices when applied to wild-type embryos.

When the embryonic axes cannot be inferred, we give arbitrary names to cells at the first time point (“Nuc” followed by a sequential number) and name subsequent cells using the image axes, giving precedence to the x axis over the y axis and the y axis over the z axis. Because of various rules in the tracking algorithms, a cell at a later time point may not be matched to any cell in the previous time point. The unmatched cell is also given an arbitrary name.

Software implementation. STARRYNITE is implemented in C and is freely available as a supporting TAR file that is published as supporting information on the PNAS web site. STARRYNITE takes two inputs: the 3D, time-lapse image series and a text file containing the tunable runtime parameters. The output is a set of text files, one for each time point containing information about the identified cells, including their positions and sizes, which cells they are matched to in the previous and next time points, and their names. Our lineaging algorithm is very efficient. It takes ~25 min to lineage through the 350-cell stage on a desktop computer with a Pentium 4 central processing unit (2.8 GHz) and 1 gigabyte of memory, or 6 s per time point.

We thank Drs. James Priess and James Thomas and their groups, Drs. James Waddell, William Mohler, and John Sulston for their discussions and Adrienne Waterston for advice on image display. This work was supported in part by the National Institutes of Health. Z.B. and S.L.O. are Damon Runyon Fellows supported by Damon Runyon Cancer Research Foundation Fellowships DRG-1813-04 and DRG-1818-04, respectively. J.I.M. is a fellow of The Jane Coffin Childs Memorial Fund for Medical Research.
### Table 1. Special naming rules

<table>
<thead>
<tr>
<th>Mother Cell</th>
<th>Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABal</td>
<td>i0ya</td>
</tr>
<tr>
<td>ABala</td>
<td>a0zp</td>
</tr>
<tr>
<td>ABalaaa</td>
<td>i0zp</td>
</tr>
<tr>
<td>ABalapaa</td>
<td>i0ya</td>
</tr>
<tr>
<td>ABalppaa</td>
<td>a0ya</td>
</tr>
<tr>
<td>ABalpp</td>
<td>a0ya</td>
</tr>
<tr>
<td>ABar</td>
<td>a0zp</td>
</tr>
<tr>
<td>ABaraaaa</td>
<td>a0xa</td>
</tr>
<tr>
<td>ABrp</td>
<td>a0ya</td>
</tr>
<tr>
<td>ABpraaa</td>
<td>a0zp</td>
</tr>
<tr>
<td>ABprappp</td>
<td>i0ya</td>
</tr>
<tr>
<td>ABprapppp</td>
<td>a0ya</td>
</tr>
<tr>
<td>ABprpap</td>
<td>a0ya</td>
</tr>
<tr>
<td>ABpp</td>
<td>i0ya</td>
</tr>
<tr>
<td>ABpappp</td>
<td>a0ya</td>
</tr>
<tr>
<td>ABpapppp</td>
<td>i0ya</td>
</tr>
<tr>
<td>ABpappppp</td>
<td>i0ya</td>
</tr>
<tr>
<td>Caap</td>
<td>i0za</td>
</tr>
<tr>
<td>Cpa</td>
<td>a0ya</td>
</tr>
<tr>
<td>Cpap</td>
<td>i0za</td>
</tr>
<tr>
<td>Cpp</td>
<td>a0yp</td>
</tr>
<tr>
<td>CPP</td>
<td>i0ya</td>
</tr>
<tr>
<td>Cppp</td>
<td>a0ya</td>
</tr>
<tr>
<td>D</td>
<td>i0za</td>
</tr>
<tr>
<td>Daa</td>
<td>a0za</td>
</tr>
<tr>
<td>Ea</td>
<td>i0yr</td>
</tr>
<tr>
<td>Ep</td>
<td>i0yr</td>
</tr>
</tbody>
</table>

How to read the rules: (i) The rules assume a standard orientation of the embryo as shown in Fig. 5. Images of other orientations would need proper transformation. (ii) \( x, y, \) and \( z \) are the Cartesian axes of the images, whereas anterior–posterior (a–p), left–right (l–r), and dorsal–ventral (d–v) are the embryonic axes. (iii) The rule a0zp for ABala means that when ABala divides, first look at the distance between the two daughters along the canonical choice of a–p.
axis (the first “a”). If the distance falls below the standard cutoff (see Methods), look for their relative position along the z axis (the third digit, “z”), and the one with the smaller z coordinate should be called the “p” daughter of ABal (the last digit, “p”). The rule i0ya for ABal means that when ABal divides, ignore (the first digit “i”) the canonical choice of axis. The daughter with the smaller y coordinate (the third digit, “y”) should be called the “a” daughter (the last digit “a”). (iv) The rules use a combination of the embryonic and the image axes because, first, the a–p axis can be easily recognized and does not change over time, and, second, the d–v and l–r axes change during gastrulation, and it is not clear how they are defined at various intermediate stages. However, at any given stage of the change, the orientation of embryo relative to the image axes is similar from embryo to embryo.