METHODS & TECHNIQUES

SAPling: a Scan-Add-Print barcoding database system to label and track asexual organisms

Michael A. Thomas and Eva-Maria Schötz*
170 Carl Icahn Laboratory, Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ 08544, USA
*Author for correspondence (eschoetz@princeton.edu)

Accepted 15 August 2011

SUMMARY

We have developed a ‘Scan-Add-Print’ database system, SAPling, to track and monitor asexually reproducing organisms. Using barcodes to uniquely identify each animal, we can record information on the life of the individual in a computerized database containing its entire family tree. SAPling has enabled us to carry out large-scale population dynamics experiments with thousands of planarians and keep track of each individual. The database stores information such as family connections, birth date, division date and generation. We show that SAPling can be easily adapted to other asexually reproducing organisms and has a strong potential for use in large-scale and/or long-term population and senescence studies as well as studies of clonal diversity. The software is platform-independent, designed for reliability and ease of use, and provided open source from our webpage to allow project-specific customization.

Supplementary material available online at http://jeb.biologists.org/cgi/content/full/214/21/3518/DC1

Key words: barcoding, population study, aging, diversity, flatworm, asexual, regeneration.

INTRODUCTION

In recent years, asexual organisms have become an important model for the study of questions about the emergence of diversity within a population and the relationship between aging and reproduction (e.g. Martinez and Levinton, 1992; Yoshida et al., 2006; Verhoeven et al., 2010; Richards, 2011). Existing studies, however, are often inconclusive or controversial, which may be due in part to the difficulty of tracking many individuals over an extended period of time, and in particular across multiple generations. To overcome this limitation, we have developed SAPling, a ‘Scan-Add-Print’ barcoding database system that can be utilized to record data on the reproductive histories of thousands of individuals within an asexual population.

The SAPling system can generate and visualize family trees and aid in the statistical analysis of the reproductive behavior of clonal individuals. Combined with molecular and genetic tools, this platform enables large-scale studies of genetic diversity and other heritable variation in populations of asexual organisms, which will also serve to deepen our understanding of sexual reproduction and evolution. It has been hypothesized that exclusively asexual populations suffer slower adaptive evolution because of the lack of genetic recombination (reviewed in Browne and Hoopes, 1990), but despite various studies on the genetic diversity accumulated by mutations in clonal populations (Lynch, 1985; Lynch et al., 1989; Browne and Hoopes, 1990; Meirmans and Tienderen, 2004), this question is still subject to debate. Additionally, it has recently been found that epigenetic effects are also passed between parents and offspring (Wilson et al., 2003; Jablonka and Raz, 2009; Verhoeven et al., 2010) and thus provide another source of adaptive potential. Because asexual organisms largely avoid the problem of disentangling epigenetic from genetic variability (Monteuxois et al., 2008; Johannes et al., 2009; Richards et al., 2010; Verhoeven et al., 2010; Richards, 2011), they are particularly well suited for studies of epigenetic variation.

Asexual organisms also raise interesting questions in the field of aging research, as studies in hydra (Bell, 1984; Martinez, 1998; Yoshida et al., 2006; Estep, 2010), flatworms (Child, 1911; Sonneborn, 1930; Haranghy and Balázs, 1964; Martinez and Levinton, 1992), jellyfish (Ojimi et al., 2009) and annelids (Bell, 1984; Martinez and Levinton, 1992; Martinez, 1996) conflict on whether these regenerative animals undergo senescence in their asexual state or whether a correlation exists between aging and sexual reproduction. It has, however, been shown that senescence does occur in bacteria and yeast using a large-scale approach involving hundreds of specimens over many generations (Barker and Walmsley, 1999; Lai et al., 2002; Stewart et al., 2005; Ackermann et al., 2007). By enabling large-scale and long-term studies incorporating many reproductive cycles, SAPling has the potential to help reach a definitive answer to this debate.

To our knowledge, SAPling is the only existing freely available database system for tracking asexual organisms. A number of systems are available for managing sexual animals, primarily mice (Boulukos and Pogonuce, 2001; Hopley and Zimmer, 2001; Pryor et al., 2001; Ching et al., 2006; Milisavljevic et al., 2010), but these are largely based on commercial software and would be difficult to customize for asexual organisms. SAPling, in contrast, is a ready-to-use database system that can be utilized to record data on each individual in populations of asexual organisms and produce family trees mapping the entire population. It is designed to print a barcoded label for each individual to provide easy and error-free tracking and database input, and provides data structures for handling computational analysis. Here, we illustrate the capabilities of SAPling using a two-year-long study of over 5000 non-interacting asexual freshwater planarians and discuss how the system could be applied to other organisms.
MATERIALS AND METHODS

We use the asexual strain of the species Schmidtea mediterranea Benazzi, Baguña, Ballester, Puccinelli & Del Papa 1975 for our planarian population studies (Dunkel et al., 2011; Quinodoz et al., 2011). Individual worms are kept in separate 100 mm Petri dishes, and the lid of each dish is tagged with an adhesive label listing its family number, name and date of birth, as well as a barcode encoding this information (Fig. 1). When a worm undergoes asexual reproduction it divides into a head piece and a tail piece. Occasionally, planarians can drop off multiple pieces in succession (Quinodoz et al., 2011); these are referred to as single ‘fragmentation events’ but are treated for the purposes of data entry as consecutive divisions. Each division is recorded and the pieces are separated into new dishes and given new labels. A full description of the other aspects of planarian maintenance is available in Quinodoz et al. (Quinodoz et al., 2011).

The SAPling database system was written in Java (Oracle, Redwood Shores, CA, USA) using NetBeans 6.9.1 (Oracle) with JDK 6 (Oracle). The user interface layout was created with the NetBeans GUI Builder. Extensions to the system for data analysis for the planarian project were written both in Java and in MATLAB (version R2009b, MathWorks, Natick, MA, USA) using the MATLAB Java interface. When a division is added to SAPling, the offsprings’ labels are automatically generated and printed on a Zebra TLP2842 model thermal desktop printer (catalog no. 2824-11100-0001, BarcodeSource, Inc., Chicago, IL, USA). Labels are printed on Zebra Z-Ultimate 4000T white, removable, waterproof adhesive labels (catalog no. 10002629, BarcodeSource, Inc.) using Zebra 5095 Resin Ribbon (catalog no. 800132-202, BarcodeSource, Inc.) and read with a hand-held barcode scanner (catalog no. IDA-SC5USB-D, IDAutomation, Tampa, FL, USA). An HP Pavilion MS225 All-in-One Desktop PC is used to run the SAPling database and interface with the scanner and printer.

RESULTS AND DISCUSSION

In the SAPling system, asexual organisms are separated into ‘families’, each corresponding to one of the individuals in the original population. Within a family, each individual is associated with a binary name specifying its reproductive history. The key to using SAPling is applying this binary naming convention to the asexual organism at hand. We describe the naming scheme for the planarians in our population study and then show how binary names can be assigned to other organisms.

When a planarian reproduces asexually, it splits into a larger head and a smaller tail piece. After each division, a ‘1’ is appended to the parent worm’s name for the head piece’s new name, and a ‘0’ for the tail (Fig. 2A). For human readability, the binary strings are converted into a form such as T1-H3-T2, representing a worm that was the tail piece of the first division, then the head piece of the next three divisions, and finally the tail piece of the next two divisions, a.k.a. the worm 011100.

In this nomenclature, both head and tail receive new names after a division. Even in cases where it may be preferable to think of the head as the ‘parent’ and the tail as the ‘offspring’, for example in studies of reproductive strategies (Quinodoz et al., 2011), this allows each reproductive event to be uniquely identified. This schema of parent and offspring applies to a wide range of asexually reproducing organisms. For instance, hydra typically reproduce by budding so we can append a ‘1’ for the mother hydra at each division and a ‘0’ for the bud (Fig. 2B). To convert the binary names for this case

![Fig. 1. Planarian in a 100 mm diameter Petri dish, identified with a label and barcode containing its reproductive history.](image)

![Fig. 2. Illustration of the naming convention using unique binary codes for (A) dividing planarians, (B) budding hydras and (C) parthenogenetic lizards.](image)
Fertilizing hermaphrodites, such as Caenorhabditis elegans (Cuellar, 1981; Cuellar, 1984; Badaeva et al., 2008) and self-fertilizing plants (Wilson et al., 2003; Kanbe and Akimoto, 2009), certain lizards (Martínez and Levinton, 1992; Martínez, 1996; Bely, 1999; Bely, 2006), acoels (Sikes and Bely, 2008; Sikes and Bely, 2010), starfish (Ereskovsky and Tokina, 2007), annelids (Berrill, 1952; Gottlieb, 1991; Stibor and Tokle, 2003; Ojimi et al., 2009), sea anemones (Hand and Uhlinger, 1995; Burton and Finnerty, 2009), sponges (Ereskovsky and Tokina, 2007), and also leaves two unused Code 39 characters that could be used for distinguishing special barcodes, if necessary. For example, because of the size of our labels, we can create barcodes for planarians up to at most generation 63. If some worms reached this limit, we could use the reserved characters to create a barcode table for these special cases.

To generate the binary string, the family ID number, binary name and date of birth. It uses the Code 39 barcode format, which is self-checking and has 43 distinct patterns representing the characters A–Z, 0–9 and several punctuation symbols (Pavlidis et al., 1990). To conserve space on the labels, we developed a method for converting a binary string into a Code 39 barcode, which SAPling uses to encode the information. The binary string is broken up by converting each 16 bits into three Code 39 characters, which can be done with little loss because 216 is just slightly less than 413. This is computationally simpler than converting the entire string at once and also leaves two unused Code 39 characters that could be used for distinguishing special barcodes, if necessary. For example, because of the size of our labels, we can create barcodes for planarians up to at most generation 63. If some worms reached this limit, we could use the reserved characters to create a barcode table for these special cases.

To generate the binary string, the family number and date of birth are first converted into binary representations and concatenated. The entire string must be a multiple of 16 bits long, so a ‘1’ is placed in front of the worm’s binary name and then zeros in front of that
to fill up to the next multiple of 16 bits. The ‘1’ is necessary to separate the meaningless padding zeros in front from any zeros at the beginning of the binary name. The algorithm described above is then used to convert this binary string to a barcode. The number of bits allotted to the family number and date of birth can be customized as necessary for the specific project (supplementary material Fig. S1).

Fig. 3 illustrates the usage of the SAPling system for data logging and processing. When the database is opened, the user is prompted to enter the current date. This will be the date associated with recorded divisions and other events. It can be changed at any time and the new date will apply to all subsequently entered data, allowing events from multiple dates to be logged at the same time.

The main part of the Graphical User Interface has three tabs: Add, Lookup Worm and Lookup Date. The Add tab allows the user to add events to the database including reproduction, sampling individuals by fixing or freezing, and deaths from either natural or other causes. The type of event can be selected from a drop-down list. For reproductive events, there is a second menu for specifying a number of offspring. The event is recorded, the Barcode, Name and Birthdate fields are automatically cleared, and (where applicable) the SAPling system automatically generates and prints labels for the offspring (Fig. 4A and supplementary material Movie 1). The system also keeps log files detailing all data entered in each database session.

If an individual with the specified name and birthdate is not found or is found to have already divided, died, etc., the database displays an error message. The user can see more details on the conflicting individual or date on the Lookup Worm or Lookup Date tabs to try to resolve the discrepancies immediately, or can press the Log Worm button to make a note in the session log files for later reference. If the database is only used with the automatically printed barcodes, these errors should not occur. Nonetheless, these features can be useful when transitioning to the SAPling system from previously handwritten labels and manually kept records. The Lookup Worm tab allows the user to retrieve family history information on a specific individual (Fig. 4B). On the Lookup Date tab, the user can see a list of all worms that were born or that reproduced on a specific date (supplementary material Fig. S2). The interface labels are customizable and the families can be assigned specific names as in our study or simply be numbered.

The SAPling database is built on top of a tree data structure that represents a single family of organisms. It stores objects representing each individual with links between parents and offspring as nodes in a binary tree. The code for reading SAPling’s data files and building and visualizing family trees (Fig. 5) is provided, so computational analysis routines can take advantage of this intuitive tree structure instead of having to handle the raw data. In addition, the system provides a list data structure and a filtering system for building lists of individuals with specified characteristics (e.g. all individuals born on a certain date). The code for these data structures is written in Java, so it is easiest to write computational analysis in the same, but they can also be used in analysis routines written in MATLAB using the MATLAB Java interface or routines written in Python using the Jython implementation, which allows integration of Java and Python code.

**CONCLUSIONS**

By storing data on the complete reproductive history of large family trees, SAPling enables large-scale and/or long-term...
population experiments of asexual organisms. Experiments tracking individuals across many generations are important for studying questions of diversity in clonal populations and whether senescence occurs in regenerative organisms, and we hope SAPling will help to resolve open questions in these fields. The hardware for the system is inexpensive ($1500, including the cost of the computer) and space efficient, and the software is provided open source and free of charge at: http://www.genomics.princeton.edu/schoetzlab/software.html.

ACKNOWLEDGEMENTS

The authors thank S. Quinodoz for assistance with the supplemental movie and for being the primary database tester/user, R. Sedgewick and K. Wayne for their Java I/O Libraries, C. Liu for designing the SAPling logo, and J. Cebra-Thomas for comments on the manuscript.

FUNDING

E.-M.S. is funded by the Lewis-Sigler Fellowship and the Burroughs Wellcome Fund. M.A.T. was supported by the National Institutes of Health – National Institute of General Medical Sciences [grant P50 GM071508 to the Center for Quantitative Biology]. Deposited in PMC for release after 12 months.

REFERENCES


THE JOURNAL OF EXPERIMENTAL BIOLOGY


<table>
<thead>
<tr>
<th>Family</th>
<th>Birthdate</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peter</td>
<td>12/10</td>
<td>T1-H3-T2</td>
</tr>
<tr>
<td>5</td>
<td>344</td>
<td>011100</td>
</tr>
</tbody>
</table>

- Family: Peter
- Birthdate: 12/10
- Name: T1-H3-T2

O N H
O 2 A