

CoVE: A Colony Visualization System for Animal Pedigrees

Brady Cannon
Micron Technology Inc.
8000 S Federal Way,
Boise, ID 83716
bbcannon@gmail.com

Cheryl Jorcyk
Boise State University
Department of Biological
Sciences
Boise, ID 83725
cchorcyk@boisestate.edu

Minoti Hiremath
Boise State University
Department of Biological
Sciences
Boise, ID 83725
minoti.hiremath@gmail.com

Alark Joshi
University of San Francisco
2130 Fulton Street
San Francisco, CA 94403
apjoshi@usfca.edu

ABSTRACT

CoVE is a novel, scalable, interactive tool that can be used to visualize and manage large colonies of laboratory animals. Effective management of large colonies of animals with multiple individual attributes and complicated breeding schemes represents a significant data management challenge in the biological sciences. Currently available software either provides databases for record keeping or generates basic pedigrees but not both. Thus, there is a pressing need for an integrated colony management system that provides a repository for the data and addresses the visualization challenge presented by complex genealogical data. We present **CoVE**, a colony visualization tool that provides an overview of the entire colony, clusters individuals based on Gender, Litter or Genotype, and provides an individual view of any animal for detailed examination. We demonstrate that CoVE provides an efficient way to manage, generate and view complex pedigree of real world genealogical data from animal colonies, annotated with details of individual attributes. It enables interactive tracing of lineages and identification of censored subjects in tumor studies.

ACM Classification Keywords

H.5.2 Information Interfaces and Presentation: User Interfaces—*User-centered design*

General Terms

Genealogy, Pedigree, Visualization

Author Keywords

Genealogy visualization, Animal Colony Management, Interactive visualization

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

VINCI 2014, August 5–8, 2014, Sydney, Australia.
Copyright 2014 ACM 978-1-4503-2765-7...\$10.00

INTRODUCTION

Animals provide surrogates for experiments that cannot be ethically conducted in humans. Animals are used in research to model human diseases, test drugs and genetic susceptibility to diseases. Primates, dogs, cats, guinea pigs, rabbits, rats, animals, frogs and fish are all used in research. Typically, mice and rats are the most commonly used model organisms because they are easy to house and breed as compared to higher vertebrates, and can be genetically modified. Animal use in research is highly controversial because of ethical concerns and guidelines have been put in place to promote welfare of animals in research [19].

Animal colonies are often large and the size of a single colony can, in many cases, be in the high hundreds and some times even in the thousands. As compared to humans, the mating strategies used in animal experiments are fairly complex and unusual. For example, it is not uncommon to have inbreeding between siblings as well as across multiple generations. These mating strategies generate complex genealogies. For example, often more than one genetic modification is required to generate a human model of a given disease. In this case, each parent carries one genetic modification. However, due to the laws of Mendelian inheritance [8] and limited sizes of litters, a new generation may not have an animal that carries both genetic modifications. Thus the mating has to be repeated or tailored to increase the chances of generating the desired animal. Thus the process of generating animals with the desired genetic modifications can take several generations and/or complex breeding schemes.

The above-mentioned attributes pose both visualization and colony management challenges. The key visualization challenges in this domain are (1) Visualization of a large number of animals on a limited screen/paper size, (2) Representation of families and breeding schemes in an uncomplicated manner that facilitates tracing of the ancestry and (3) Simultaneous visualization of a number of attributes regarding each individual animal such as age, genotype etc. The key colony management challenges include (1) Limiting the number of animals used, to comply with institutional and federal regu-

lations and to limit costs, (2) Maintaining an accurate animal census and (3) Accurate representation of breeding schemes. Errors in colony management are frequent, in part due to inaccurate recording of the attributes or erroneous transcription. These errors lead to significant loss of time for the researcher, increased cost and loss of the lives of animals that are sacrificed due to their incorrect genotype.

We have developed *CoVE*, an animal colony visualization and management system to facilitate the exploration of large colonies at various levels. *CoVE* has a number of different views that address the visualization challenges detailed above. For example, the Colony View provides an overview of the colony and can help with identifying communities in the colony, based on defined attributes, and within generations. The Gene View facilitates the visualization of genetic information or animals at the individual, family and generation level. Animals with the same genotype are clustered and visually represented in a manner that facilitates their identification. Since each female animal gives birth to multiple animals, i.e litters, numerous individuals have the same parents. These are most easily seen in Gender View, which represents individual generations, families and the animals therein. We provide automated cues about the parents and grandparents of an individual animal by highlighting their locations in the genealogy. In a situation where the user wants to know more about a particular animal such as their paternal lineage or the number of mates, the user can click on the animal in the Gender View, which shows that animal in the Individual View. The Individual View is customized to show information pertaining to a single animal in the colony. It displays the ancestry of the “animal of interest”, the number of mates and the offspring that have been produced with each of those mates. The Statistics View provide overall statistics about the colony in terms of the distribution based on gender, genotype, alive vs dead, age of mice and so on. This provides an overall summary of the colony. In addition, we provide a “search” feature to locate any animal within the colony. Searching for an animal using the ID highlights all the instances of that animal in the colony (as a offspring, parent or grandparent).

Animal technicians and a faculty member in our group have used *CoVE* to examine their animal colonies. Based on their feedback, we find that they prefer the integration that the *CoVE* interface provides to their current system of tracking individual animals, which entails separate systems for generating genealogies and tracking animals. They find that *CoVE* simplifies the examination of individual animals and tracing the pedigree of an animal. The Gene View was found to be particularly useful since it provided users with a clear representation of the number of animals that have the same genotype, which helps in the detection of abnormal or non-Mendelian inheritance patterns. The ability to get a quick overview of the dead animals in the colony too was found to be valuable.

PREVIOUS WORK

Hand-drawn genealogical charts have been used for centuries to trace lineages and relationships [18]. Bertin [2] described

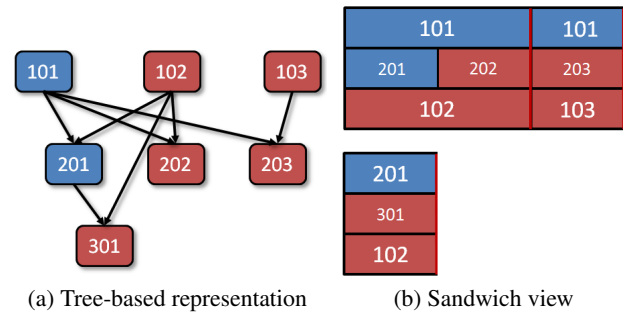


Figure 1: The left image shows a schematic of a node-link representation for a small colony with only seven mice. Note how the overlapping edges even in a small family make it hard to identify the parents and siblings of an individual. On the right is the “Sandwich View” that displays each generation as a “sandwich” with the children being represented in the middle row. Here, you can immediately identify the parents of a particular mouse as well as its siblings and cousins.

principles to represent genealogical graphs by using varying thickness of lines for males and females and connecting individual families in a node-link style graph. This approach suffers from the problem of overlapping edges for tracing pedigrees of an individual.

McGuffin et al. applied an expand-ahead metaphor to create a 2D interface for visualizing genealogical graphs [16]. The GeneaQuilts system [3] provides a matrix-based visualization of Quilts [22], adapted to bipartite layered graphs. The GeneaQuilts system was found to be less effective when visualizing animal pedigrees with very large families or relationships [12].

The PedVis system introduced by Tuttle et al [21], uses an H-tree layout to efficiently display the ancestry of an individual up to a certain number of generations. A couple of limitations of PedVis are that the amount of information visible per individual decreases as the number of generations increase and it displays only the ancestry for a single individual at a time. Draper et al. [6] developed an interactive technique for exploring large genealogies. A radial layout [24] facilitated exploration through the use of their animated fan-charts. Such representations work well for human genealogies with standard mating strategies.

Animal Genealogy Visualization

There exists significant previous work which is specifically related to the visualization of animal genealogy [23], including the genetic information of interest to biologists and geneticists. Previous approaches to visualizing animal genealogical data [1, 4, 14] represent families as Steiner diagrams. Their limitations are that they cannot handle complex mating strategies and manage only small families (around 30 animals). Other approaches have used standard visualization software such as GraphViz [7] to generate static representations of pedigree [5]. 2.5D and 3D methods too have been developed for visualizing genealogies [9]. These methods suffer from occlusion and, more importantly, depth disam-

Task Characterization		Scale					
		<i>c</i>	<i>g</i>	<i>f</i>	<i>i</i>	<i>s</i>	<i>m</i>
1	To determine if an animal is of a particular age (Eg. Breeding age = 4-5 weeks)		x	x	x	x	
2	To confirm the genotype of a given animal		x		x		
3	To identify a group of similar animals based on genotype or age		x			x	
4	To trace the pedigree of a given animal			x			
5	To determine the # of generations that the animal is removed from the founder (for backcrossing)				x		
6	To determine Mendelian inheritance or lack thereof		x	x			
7	To identify breeding females that need to be replaced (based on number of litters)	x					
8	To determine the total size of the colony (in terms of number of animals)					x	
9	To identify animals that need to be sacrificed based on genotype or age		x	x		x	
10	To earmark animals into experimental groups				x		
11	To identify a “problem” breeder that is not producing litters, or producing very small litters	x		x			
12	To identify and trace a mistyped animal by comparing its genotype to its siblings, parental genotypes		x		x		x
13	To identify the animal with the maximum number of progeny			x			
14	To specify males that are used for harem breeding			x			
15	To locate an animal within the colony						x
16	To identify which animal in the colony are dead/alive?			x	x	x	

Table 1: This table details the task characterization phase of CoVE where we identified scientific questions as well as common tasks that scientists want to perform on genealogical data. Here, *c* = colony level task, *g* = requires genotype information of an animal or a set of animals, *f* = requires family/litter level information for an animal or a set of animals, *i* = requires fine grained information at the individual level with information about its ancestors, mates and progeny, *s* = provides statistical information about the colony such as the number of males and females in the colony or the number of animals with a particular genotype, *m* = requires colony management features such as edit the information for a particular animal or search for an individual animal in the colony.

biguation issues. Most node-link diagram approaches have led to dense displays of data that fill the screen space very quickly [10]. HaploPainter discuss *duplicating nodes* in their future work to address mating schemes such as backcrossing [20], but their current system cannot handle such mating schemes or large families.

A new visualization technique termed the *sandwich view* was designed, and implemented in the Visual Pedigree Explorer (VIPER) by Graham et al [12], while considering the nature of animal pedigrees and the common existence of errors in genotype information. The design of this system focuses on the task of detecting and resolving potential genetic errors which show up as inconsistencies with respect to the Mendelian laws of inheritance. One of our views (gender view) was inspired by the ‘sandwich view’ in VIPER.

Existing visualization techniques, including node-link graphs and traditional matrix-based approaches were investigated and found to be insufficient for legibly displaying the pedigree of large animal colonies [17]. Only the GeneaQuilts [3] system was able to handle the size of the data, but lacked features for relating individuals across generations, which were considered important for detecting and resolving genotype errors. Paterson et al. [17] found their Bring& Slide navigation technique to be less effective for very large families. Histograms with sliders provide features for filtering the errors displayed in the sandwich view. A 2-stage evaluation of VIPER was performed by Paterson et al. [17], to test its effectiveness for displaying genetic errors in various data. This evaluation provided motivation for new features to separate offspring by sex, view individuals in a single family, identify

inferred genotype information and explore specific markers in detail using a new “Marker Table”.

Ghoniem et al. [11] compared node-link and matrix-based representations of graphs for readability. Based on their evaluations, they found node-link diagrams to be less readable as the graph size increased. Based on these studies and our previous prototypes, we found that node-link diagrams and matrix-based representations are limited for visualizing animal genealogies.

DATA CHARACTERIZATION

Our target users consisted of an animal technician, graduate students, postdoctoral researchers and faculty members conducting animal-based research. Based on several interviews with our target users we provide a detailed description of the type of data available, a list of tasks that are required to enhance colony management performance and scientific questions that arise when dealing with animal research.

Attributes available for each animal are the animal identifier (ID), date of birth, date of death (if applicable), gender, parent IDs, the litter number that it belongs to, the genotype and any other annotation that the animal technician may prefer to store for that animal. The genealogy that is based on the abovementioned data can be scaled into three levels. At the broadest level, information about the entire colony is crucial to the understanding of the number of generations in a colony and the number of litters produced by a given set of parents. At the intermediate level, the focus is on the individual generations and the families contained within. Here the interest is in examining families to explore common themes

such as genotypes or examine individual animals within a litter. Additionally, genetic similarity between animals in the same family is essential to determine Mendelian inheritance. At the lowest level, the focus is on an individual animal that a scientist chooses to explore further. At this level, information regarding the ancestors of the animal, its maternal/paternal lineage, its partners and their respective offspring are important.

The analysis of this data can be particularly challenging since the amount of animals present in a colony can easily be in the hundreds and even thousands in some cases. Based on interactions with our target users, we identified questions and a series of tasks that are routinely performed for colony management and can be sources of error. In addition, we defined a set of scientific questions that would be relevant to animal research. These tasks and questions are listed in Table 1.

COVE

Our design decisions were strongly influenced by our task characterization. We found that the scientists want to focus on tasks surrounding an individual family or around an individual animal and its pedigree. Representation of the data in traditional node-link diagrams results in multiple overlapping edges because of the amount and complexity of the data. Thus, tracing a single link is extremely difficult [11]. To avoid this problem, we adapted the “sandwich technique” for a single family in a generation [12]. Figure 1 shows a comparison between a small colony of seven animals being visualized as a node-link diagram and a sandwich-based visualization. The idea behind the sandwich technique is that the parents are shown on opposite sides of the offspring, in each generation and the offspring are displayed in between a row of fathers and a row of mothers in that generation. This makes it extremely easy to visualize the individual animals and identify its parents, siblings and cousins. This technique also eliminates the need for tracing complex overlapping edges and reduces the cognitive overload. A new sandwich (generation) is formed when an offspring in the current generation is used as a parent in the next generation. This means that an animal cannot be an offspring and a parent in the same generation. For example, in Figure 1 animal 301 is the progeny of animal 201 and 102. Since 201 is an offspring in the first generation, we create a new sandwich to represent the family with 201 and 102 as the parents of 301. This design results in animals being represented multiple times in the colony (animal 102 in Figure 1), but eliminates any edge crossings and keeps the genealogy segregated by generation.

Colony View

The Colony view was designed to provide an overview of the entire colony. Figure 2 shows a colony of 110 animals being visualized in CoVE. The fathers of a family are shown in blue in the top row of each generation, while the mothers are shown in red in the bottom row of that generation. Each family is made up of one or more litters and separated by vertical red lines from an adjacent family that is in the same generation. Each family consists of one or more litters, which are identified by litter numbers and separated by vertical blue lines. For example, in Figure 2 in the 1st gen-

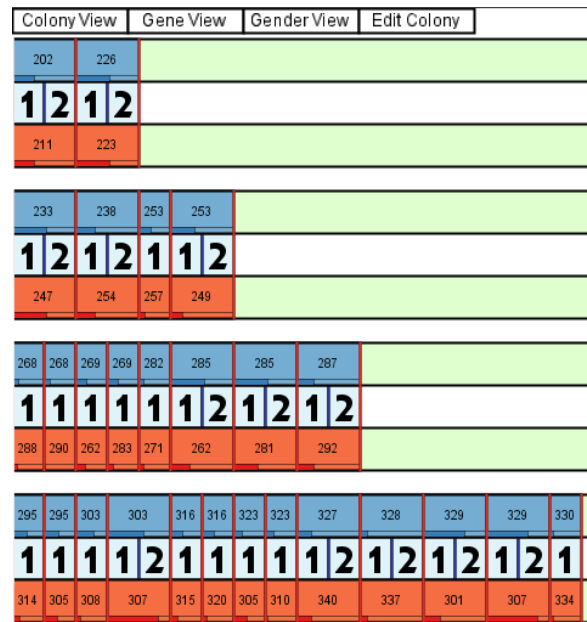


Figure 2: Colony View: In this view, a user is shown an overview of the entire colony. Fathers that comprise a generation are shown in blue in the top row of each generation, while mothers are shown in red in the bottom row. Instead of showing the details of each family, the colony view only displays the litter numbers of each family implying that the family has multiple litters.

eration animal 202 and 211 have a family with two litters: Litter ID 1 and Litter ID 2. Instead of displaying individual animals in each family, only the litter ID is displayed here to give an overview of the family and the colony. A user can select any litter ID or animal ID (father or mother) to drop down to the Gender View, where details regarding the individual animals that comprise a family are displayed.

As a users pointer hovers over a single animal, we highlight every instance of that animal in the colony as an offspring and as a parent. Figure 6 shows an example of the highlighting feature in CoVE. Additionally, we provide visual cues regarding the number of offspring that the animal has produced by scaling the size of the animal ID as shown in Figure 4. We found that merely scaling the animal ID based on the number of offspring was useful, but it did not convey whether that particular animal has the highest number of offspring in the colony. To augment this feature, we provide a small bar glyph which shows how many offspring a particular animal has produced as compared to the animal who has the highest number of offspring in the colony.

Gene View

In the Gene View all the animals within a given family that have the same genotype are clustered together. This helps identify specific genetic combinations that exist in a family and is useful for planning further mating strategies to generate animals with a specific genotype. The genotype information is represented as horizontal line segments under

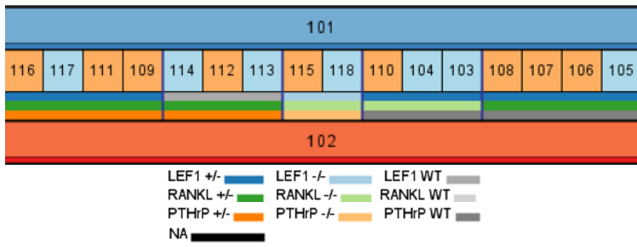


Figure 3: Gene View: In this view, mice with the same genotype are clustered together. For example, mice 116, 117, 111, and 109 mice have the same genotype (Lef1 +/-, RANKL +/- and PTHrP +/-) and therefore have longer strips below them. A legend is shown to identify individual genetic markers.

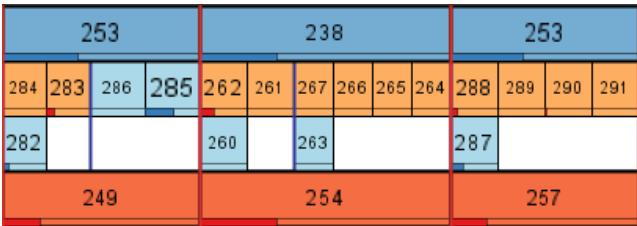


Figure 4: Gender View: The mice in a family can be examined in this view. The mice within each litter are sorted according to their gender with the list of females being displayed first followed by males. The text label of each mouse ID is scaled according to the number of children it has produced. A small glyph below each mouse indicates how many children it has produced as compared to the mouse with the maximum number of children in the entire colony.

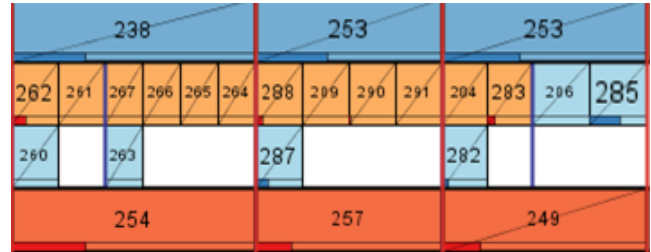
each animal with a genotype legend shown in the status bar of the screen. Additionally, tool tips for each horizontal strip convey individual genetic markers for each animal. Figure 3 shows a single generation of animals clustered according to their genotype. In the family of 101 and 102, animals 116, 117, 111 and 109 have the same genotype. Similarly, animals 108, 107, 106 and 105 also have the same genotype. The genotype can be easily distinguished based on the allele legend. Each allele is represented by a different color and the WT alleles for all genes are represented by shades of gray for easy identification. The colors for the legend were obtained from ColorBrewer [13]. Therefore, a given combination of alleles represents a distinct genotype. For example, 109 is Lef1 +/-; RANKL +/-; PTHrP +/- where as 106 is Lef1 +/-; RANKL +/-; PTHrP WT. The color black indicates to the user that information for that allele is not available (NA) in the database.

Gender View

The Gender View provides the ability to examine the animals in a family and understand the family structure. The animal IDs of individual animals can be seen. As the users pointer hovers over individual animals, the status bar at the bottom displays the individual attributes of that animal such as the date of birth, its gender, the ID of its father (Father ID), the ID of its mother (Mother ID), the litter number that



(a) Litter ID overlay



(b) Depicting dead mice in CoVE

Figure 5: This set of figures shows other features of Gender View. The generation shown here is the same as in Figure 4.

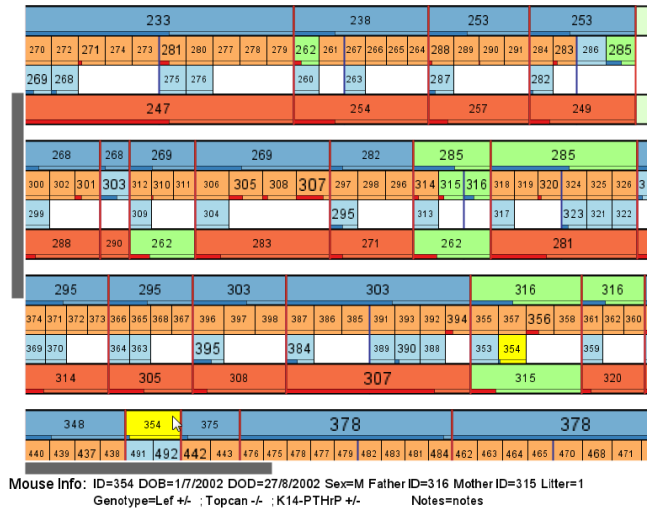


Figure 6: Gender View: As a user's mouse pointer hovers over a mouse, the status bar below shows all the relevant information about a mouse in the colony. Additionally, to help trace its ancestors, we highlight the parents and grandparents of that mouse.

the animal belongs to, date of death (if applicable), its genotype followed by any specific notes a user may have entered for that animal. Further, we provide information to the user by highlighting the parents and grandparents of that animal in the colony. We also highlight other instances of that animal in the colony, where it is a parent of another family. The ability to examine individual details in a quick manner helps the user get an understanding of the distribution of animals in the family and individual litter. The label of each animal in the family is *scaled* according to the *number of offspring that it has produced*. It helps draw a users attention to specific (potentially important) animals in a family. For exam-

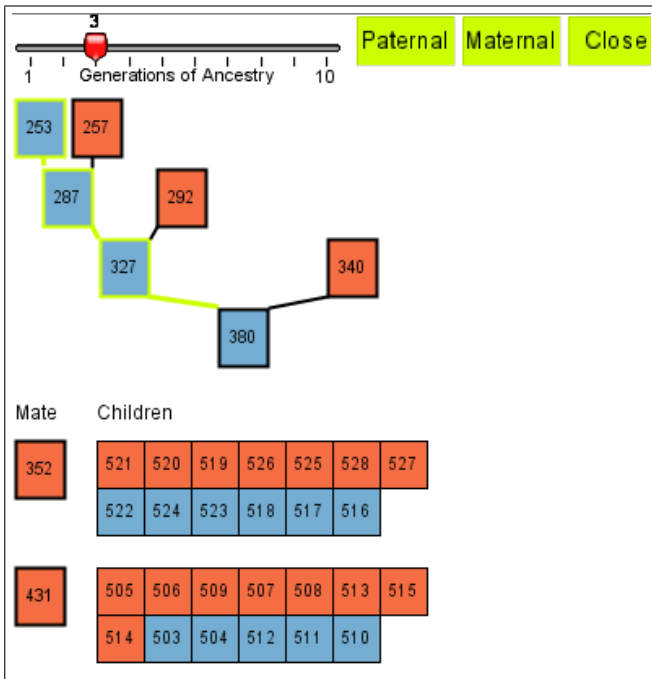


Figure 7: Individual View: Animal ID=380 being examined in the Individual View. Its ancestors are shown above it with the paternal lineage highlighted.

ple, in Figure 4 the animal 285 has a larger font than its siblings, which implies that it has produced more offspring than them. Additionally, we also display a small bar glyph under each animal ID to allow comparison between the animal being examined and the animal with the maximum number of offspring in the colony. To differentiate litters in a family, a vertical blue line is used. Animals 286 and 285 (whose parents are 253 and 249) are the only animals in that litter and are separated from their siblings (in a different litter) by a vertical blue line.

To enable correlation between the Colony View and the Gender View, we also provide the ability to overlay litter IDs on individual families. Figure 5a shows a view of the same generation (as shown in Figure 4) but with litter ID overlay. A user can click on a specific litter in the Colony View, which would then bring up the Gender View with the litter IDs overlaid. Dead animals in a colony are represented with a diagonal line that goes across the rectangle. Figure 5b shows the dead animals in that generation. Only the two mothers (254 and 257) are alive from this generation.

Figure 6 shows the interaction capability of CoVE when examining an animal in Gender View but a similar functionality is available in Gene View. As a user's pointer hovers over a particular animal ID, its information is displayed in the status bar below. To facilitate easy tracing of its lineage, we highlight all the instances of that animal (animal 354 in this case) as well as its parents (315 and 316) and grandparents. To examine more than 2 levels of ancestry, the user can click on the individual animal to display it in the Individual View, where up to 10 levels of ancestry can be displayed.

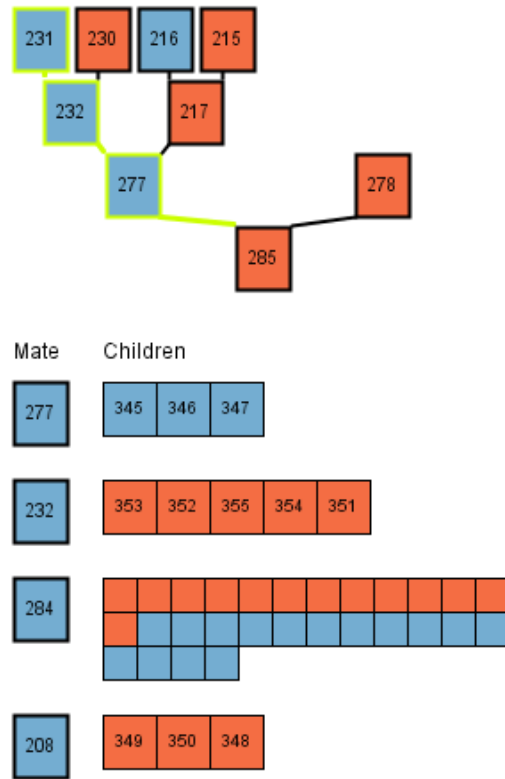


Figure 9: When displaying the children for a large family, the size of each square representing a child is scaled down to eliminate the label.

Individual View

The Individual View provides a focused view for a single animal in the colony. When a user clicks on an animal in Gender or Gene view, it is displayed in the Individual view. The animal of interest is depicted with its ancestors being drawn as a tree. The mating partners are shown below and the offspring from that mate are shown to the right of the mate. Figure 7 shows a mouse (mouse ID 380) being examined in the Individual View. Its ancestors are shown above it with the paternal lineage highlighted. The “Generations of Ancestry” slider can be used to change the number of generations of ancestors being displayed. The Mates for the mouse of interest are shown below it with a list of all the children displayed next to it. Here, mouse 380 has mice 352 and 431 as its mates and their respective children are shown next to them. Additionally, we continue to show the detailed mouse information in the status bar as shown in Figure 6 to learn more about its date of birth, genotype, and so on.

Figure 8 shows five levels of ancestors for a particular animal being displayed. At any point the user can click on any animal in this view to make it the “focus” of the individual view. The view is then updated to display its ancestors, its mates and their offspring. Thus, the individual view can be used to traverse the family tree and get a better understanding of relationships in the colony. The paternal and maternal lineages of an individual animal are of paramount importance. We provide buttons at the top of the individual view to highlight the paternal or maternal lineage of the

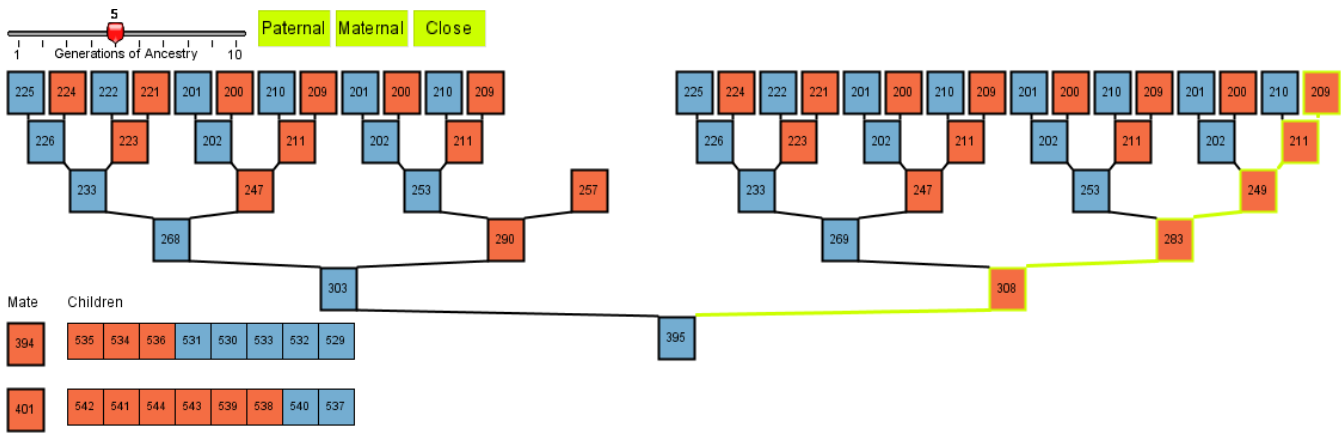


Figure 8: Multiple levels of ancestry can be visualized for that mouse by varying the ancestor slider. Here the user is being shown five levels of ancestors for a particular mouse of interest. Additionally, the maternal lineage is being highlighted to examine the pedigree of the mouse.

animal. For example, Figure 8 shows the maternal lineage being highlighted. Such lineage is crucial in determining the pedigree of a specific animal as well as to identify any errors in mating that may have inadvertently occurred. At times, the animal in “focus” has many mates and many children from those mates.

We want the individual view to be able to provide a comprehensive overview of all aspects of an individual mouse. This means that if a mouse has many children with one of its mates, then the individual view should still be able to represent them. Since each mouse can have families with multiple mates, we show the mates and their respective children on subsequent rows, as shown in Figure 8. Additionally, since the individual view window is resized depending on the number of ancestors being shown, the amount of space available to display the children is dependent on the size of the window. If there are more than a certain number of children, we draw them on a new row (as shown in Figure 7).

In situations, where a family may have a large number of children, we further drop down the representation of an individual mouse to a pixel-based representation [15]. We reduce the size of each individual square to display all the children in the view as shown in Figure 9. For larger families, the size of the square decreases further. There is an allowed horizontal and vertical space to draw the children in. The children are assumed to be square, and the side length is reduced from the size of the mate button until all children fit in the space. If the text can fit into that size (minus the width of the lines on left/right sides) and is still legible, it is drawn. These squares are clustered and displayed according to the gender of the mice in the family. All the females in the family are shown first followed by the males in the family. Even though the mouse ID’s of the individual children are not visible, a user can still hover their mouse pointer over the squares to see individual mouse information in the status bar at the bottom of the screen as shown in Figure 6.

Statistics View

Total Population: 268			Families: 13			Close
Dead (%): 215 (80.2)			Alive (%): 53 (19.8)			
Males (%): 72 (26.9)		Females (%): 188 (70.1)		Gender NA (%): 8 (3.0)		
Markers: LEF1 (+/-) RANKL (+/-) PTHrP (+/-)	Markers: LEF1 (NA) RANKL (NA) PTHrP (+/-)	Markers: LEF1 (+/-) RANKL (NA) PTHrP (NA)	Markers: LEF1 (WT) RANKL (+/-) PTHrP (+/-)	Markers: LEF1 (WT) RANKL (NA) PTHrP (+/-)	Markers: LEF1 (WT) RANKL (NA) PTHrP (NA)	
#Mice: 2 (0.7%) #Male: 2 (100.0%) #Female: 0 (0.0%)	#Mice: 21 (7.8%) #Male: 4 (19.0%) #Female: 12 (57.1%)	#Mice: 25 (9.3%) #Male: 3 (12.0%) #Female: 22 (88.0%)	#Mice: 4 (1.5%) #Male: 3 (75.0%) #Female: 1 (25.0%)	#Mice: 7 (2.6%) #Male: 2 (28.6%) #Female: 5 (71.4%)	#Mice: 99 (36.9%) #Male: 31 (31.3%) #Female: 67 (67.7%)	
Living Individuals:						
Age:1 days						
1 LEF1 +/- RANKL +/- PTHrP +/-						
6 LEF1 +/- RANKL +/- PTHrP NA						
8 LEF1 +/- RANKL NA PTHrP NA						
2 LEF1 NA RANKL +/- PTHrP NA						
1 LEF1 NA RANKL NA PTHrP +/-						
3 LEF1 WT RANKL +/- PTHrP +/-						
16 LEF1 WT RANKL NA PTHrP NA						
Age:30 days						
1 LEF1 WT RANKL NA PTHrP NA						
Age:35 days						
1 LEF1 NA RANKL NA PTHrP NA						
2 LEF1 WT RANKL NA PTHrP NA						
Age:61 days						
1 LEF1 +/- RANKL +/- PTHrP +/-						
4 LEF1 +/- RANKL NA PTHrP NA						
1 LEF1 WT RANKL +/- PTHrP +/-						

Figure 10: The Statistics View provides an overall statistical summary of the colony.

Animal researchers need to know a summary of the mice in their colony for housekeeping purposes. To aid in this

Figure 11: Mouse colony management: New mice can be added to the colony using the Add Mice user interface shown here. Information about individual mice can be edited as well as new mice can be added to the colony using this interface. As soon as a user enters a mouse ID into the Mouse ID text field, the relevant mouse information is displayed and can be edited.

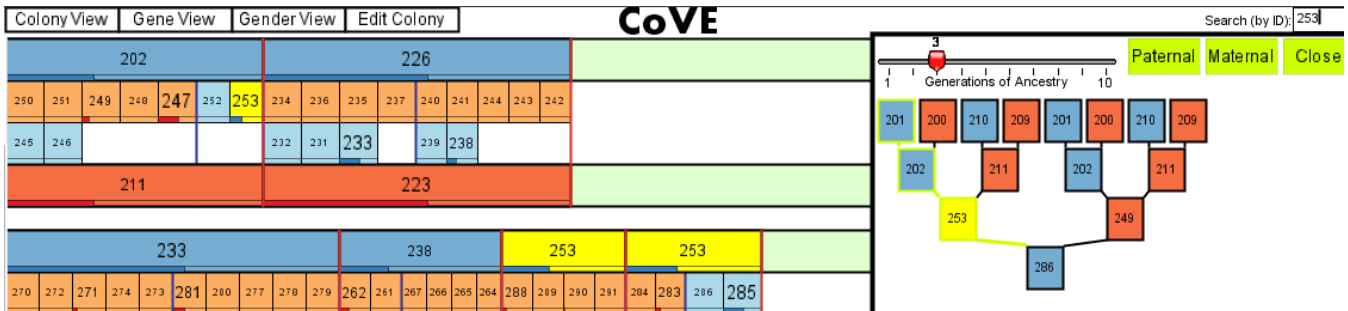


Figure 12: Search: A mouse can be searched for in the Search box. Here mouse 253 is found and highlighted in the individual view as well as both the generations in the Gender View.

process, we provide the Statistics View. The Statistics View provides a detailed summary of the colony in terms of the total population of the colony, the total number of families in the colony, the number of animals that are dead, the number that are alive, the number and percentage of males and females in the colony. For individual genotypes, we also show the total number of mice that have the same genotype and further divide that into males and females. Additionally, we also provide summary statistics in terms of their age. This can help with answering the question #3 in Table 1 - Identify a group of similar animals based on genotype or age. Figure 10 shows the Statistics View for a colony of 268 mice. The number of mice in the colony, the number of families in the colony and a distribution of the mice based on whether they are dead or alive, their gender, their genotype, and their age is presented to the viewer.

Colony Management

To explore and make changes in the mouse colony, it is imperative to be able to edit the data easily. Moreover, since CoVE is an integrated visualization and management system, we provide the functionality to add new animals into the colony. Figure 11 shows the add/edit animal information interface. The “add animals” user interface allows a user to enter the information for a particular animal and add the animal to the colony. In addition to specifying the information for a new animal in the colony, a user can specify how many offspring have the same gender and same genotype. The animal IDs are then serially incremented and the animals are added to the colony. This makes the addition of new ani-

mals to the colony fast, intuitive and less error prone since the gene combinations are pre-specified and a user only has to select the correct combination. When adding new animals, specifying the animal IDs of the parents is optional, since there are situations when an animal is “bought” for a breeding and its parents, who are not in the colony being maintained. CoVE automatically calculates the age of the animal in days, based on the date of birth of the animal and the current date. This helps with identifying whether the animal is ready for mating (breeding age = 4-5 weeks) or for experimentation.

To edit the information for an existing animal, the user types the ID into the Animal ID text field at which point all of the relevant fields are populated automatically. The user can then make any changes to the animal information and hit the Apply button. The animal information in the colony is updated immediately. Additionally, a user can search for a specific animal in the entire colony. If a match is found, all the instances of that animal are highlighted in the Gender View (and Individual View if applicable). Figure 12 shows a scenario where a user searches for animal 253. Since animal 253 is being visualized in the Individual View, it is highlighted there as well as in the Gender View in both the generations where 253 occurs.

CASE STUDY

CoVE is currently being used to examine and manage a few small colonies of 30/40 mice and a larger 300 mouse colony. Here we describe the experience of a co-author exploring the

family of 300 mice in CoVE. One of the first observations made was about the ability to visualize the entire colony and how easy it was to see all the mice in the Colony View along with their litter ID's.

Mendelian Inheritance

She found it is easy to spot whether Mendelian Inheritance was being followed in the litters as well as families of the colony. For example in Figure 13, she saw that the first litter mostly follows Mendelian inheritance with an almost equal balance between males and females in that litter. But in the rest of the litters (separated by vertical blue lines), the mice are all females. Other such interesting patterns were found and explored to examine Mendelian inheritance.

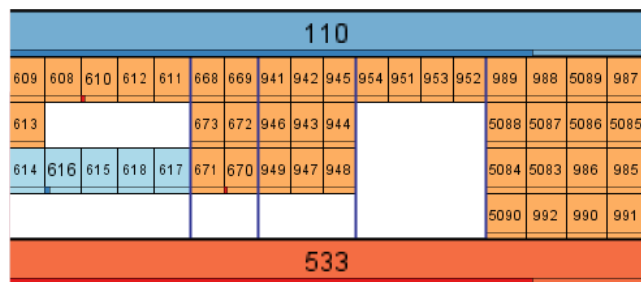


Figure 13: Children in the first litter follow Mendelian inheritance and have a balance between the males and females in that litter. The rest of the litters do not exhibit such a balance and are all females.

Identifying Genotype

One of the biggest sources of error in a mouse colony is from incorrect genotype information for a mouse. Part of the problem is that the genotype information can easily be misread. She wanted to find a particular mouse in the colony, that she bred and examine its genotype. Figure 14 shows a generation with that mouse in it. Here, the combination that she wanted in the “ideal” mouse was *Lef1* *-/-*, *RANKL* *+/-*, *PTHrP* *+/-*. Now, they had had trouble obtaining such a mouse in previous matings and so to increase their odds they mated two sets of opposite gender pairs with the same genotypes. Mouse 616 (male) with the genotype *Lef1* *+/-*, *RANKL* *+/-*, *PTHrP* WT was mated with mouse 801 (female) with the genotype *Lef1* *+/-*, *RANKL* WT, *PTHrP* *+/-*. Similarly, 817 (male) with the genotype *Lef1* *+/-*, *RANKL* WT, *PTHrP* *+/-* was mated with mouse 610 (female) with the genotype *Lef1* *+/-*, *RANKL* *+/-*, *PTHrP* WT.

These mice gave birth to a number of mice with a wide range of genotypes, as shown in Figure 14. Fortunately, they found a single mouse (302) that had the genotype that they wanted, which was *Lef1* *-/-*, *RANKL* *+/-*, *PTHrP* *+/-*. She was able to identify this mouse very quickly in the generation without remembering the mouse ID.

Dead mice

Identifying dead mice in a colony is crucial to make sure you do not have any mice that you are not using for experiments but is still inadvertently alive. Using the Gender View, she

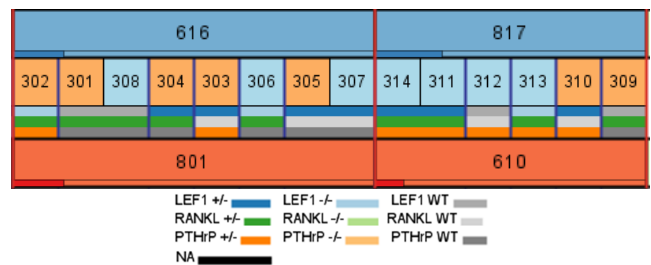


Figure 14: Identifying a mouse with a specific genotype was found to be easy using the Gene View. Mouse 302 with the desired genotype *Lef1* *-/-*, *RANKL* *+/-*, *PTHrP* *+/-* was quickly identified.

was able to quickly examine the dead mice in the colony and was surprised to find a specific mouse to be alive, even though it should have been sacrificed with its siblings. Figure 15 shows the example from the same generation where mouse 302 is alive for experimental purposes and further mating but mouse 313 has been inadvertently kept alive and is not being used for any further experiments.

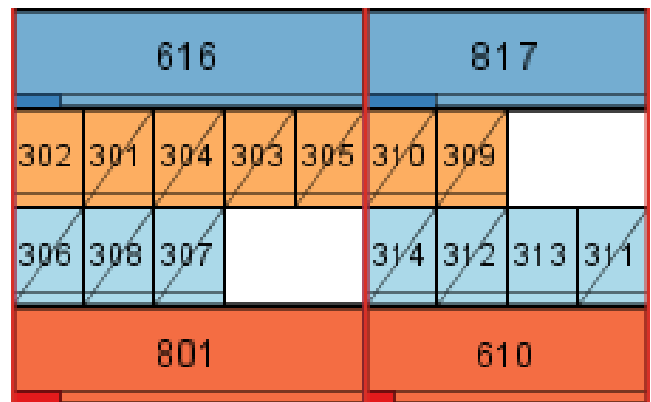


Figure 15: Identifying dead mice quickly in a colony can save time and expense in mice that are not required for experiments being kept alive.

Overall the co-author found CoVE to be very easy to use and has been using it to maintain her colony. The ability to add mice was found to be very useful. She suggested adding text message alerts when mice are of a certain age (for e.g. 21 days).

CONCLUSIONS AND FUTURE WORK

We have presented CoVE: a novel genealogy visualization system for management of large and complex animal colonies. CoVE provides the ability to examine individual animals in a huge colony and facilitates planning of matings to generate “desirable” animals with minimal impact on cost, time and lives of the animals involved. We plan on adding an “alert” feature that would send an email, text message or an in-program alert to the user according to the user-specified rules. For example, alerts would be sent if an animal is of a certain age and has a specific genotype, or if an animal has not been mated for a certain number of days. We also plan

to provide an automated genotype selection feature for new animals, based on the genetic combination of their parents. This feature will significantly reduce errors in transcription. Conduct a longer term evaluation to identify specific features and also conduct usability testing with a large group possibly at multiple sites.

REFERENCES

1. R. Agarwala, L. Biesecker, K. Hopkins, C. Francomano, and A. Schaffer. Software for constructing and verifying pedigrees within large genealogies and an application to the old order amish of lancaster county. *Genome research*, 8(3):211–221, 1998.
2. J. Bertin. Sémiologie graphique. les diagrammes, les réseaux, les cartes. *The Geographical Journal*, 135:144–146, 1967.
3. A. Bezerianos, P. Dragicevic, J. Fekete, J. Bae, and B. Watson. Geneaquils: A system for exploring large genealogies. *Visualization and Computer Graphics, IEEE Transactions on*, 16(6):1073–1081, 2010.
4. L. Brun-Samarq, S. Gallina, A. Philippi, F. Demenais, G. Vaysseix, and E. Barillot. Cope: a collaborative pedigree drawing environment. *Bioinformatics*, 15(4):345–346, 1999.
5. J. Cole. Pypedal: A computer program for pedigree analysis. *Computers and electronics in agriculture*, 57(1):107–113, 2007.
6. G. M. Draper and R. F. Riesenfeld. Interactive fan charts: A space-saving technique for genealogical graph exploration. In *In Proceedings of the 8th annual Family History Technology Workshop, Brigham Young University*, 2008.
7. J. Ellson, E. Gansner, L. Koutsofios, S. North, and G. Woodhull. Graphvizopen source graph drawing tools. In *Graph Drawing*, pages 594–597. Springer, 2002.
8. W. Ewens. *Mathematical population genetics: theoretical introduction*, volume 1. Springer Verlag, 2004.
9. C. Fuchsberger, M. Falchi, L. Forer, and P. Pramstaller. Pedvizapi: a java api for the interactive, visual analysis of extended pedigrees. *Bioinformatics*, 24(2):279–281, 2008.
10. J. Garbe and Y. Da. Pedigraph user manual. version 2.4. *Department of Animal Science, University of Minnesota*, 2008.
11. M. Ghoniem, J.-D. Fekete, and P. Castagliola. A comparison of the readability of graphs using node-link and matrix-based representations. In *Proceedings of the IEEE Symposium on Information Visualization*, pages 17–24, Washington, DC, USA, 2004. IEEE Computer Society.
12. M. Graham, J. Kennedy, T. Paterson, and A. Law. Visualising errors in animal pedigree genotype data. In *Computer Graphics Forum*, volume 30, pages 1011–1020. Wiley Online Library, 2011.
13. M. Harrower and C. Brewer. Colorbrewer.org: an online tool for selecting colour schemes for maps. *Cartographic Journal, The*, 40(1):27–37, 2003.
14. M. He and W. Li. Pedidraw: A web-based tool for drawing a pedigree in genetic counseling. *BMC medical genetics*, 8(1):31, 2007.
15. D. Keim. Pixel-oriented database visualizations. *ACM SIGMOD Record*, 25(4):35–39, 1996.
16. M. J. McGuffin and R. Balakrishnan. Interactive visualization of genealogical graphs. In *Proceedings of the Proceedings of the 2005 IEEE Symposium on Information Visualization*, pages 3–, Washington, DC, USA, 2005. IEEE Computer Society.
17. T. Paterson, M. Graham, J. Kennedy, and A. Law. Evaluating the viper pedigree visualisation: detecting inheritance inconsistencies in genotyped pedigrees. In *Proc. of IEEE BioVis 2011*. IEEE Computer Society Press, Providence, RI, USA, 2011. Paper presented at IEEE Symposium on Biological Data Visualization, Providence, RI, USA Oct 23-24 2011.
18. J. Priestley. A description of a chart of biography, 1764.
19. W. Russell, R. Burch, and C. Hume. The principles of humane experimental technique. 1959.
20. H. Thiele and P. Nürnberg. Haplopainter: a tool for drawing pedigrees with complex haplotypes. *Bioinformatics*, 21(8):1730–1732, 2005.
21. C. Tuttle, L. Nonato, and C. Silva. Pedvis: A structured, space-efficient technique for pedigree visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 16(6):1063–1072, 2010.
22. B. Watson, D. Brink, M. Stallmann, R. Devarajan, M. Rakow, T. Rhyne, and H. Patel. Visualizing very large layered graphs with quilts. In *IEEE Information Visualization Conference Poster*, 2007.
23. L. Wong. Visualization and manipulation of pedigree diagrams. *Genome Informatics*, 11:63–72, 2000.
24. K. Yee, D. Fisher, R. Dhamija, and M. Hearst. Animated exploration of dynamic graphs with radial layout. In *Presented at IEEE Symposium on Information Visualization*, 2001.