

Author's Accepted Manuscript

A computational approach to animal breeding

Tanya Y. Berger-Wolf, Cristopher Moore, Jared Saia

PII: S0022-5193(06)00395-X
DOI: doi:10.1016/j.jtbi.2006.08.028
Reference: YJTBI4464

To appear in: *Journal of Theoretical Biology*



www.elsevier.com/locate/jtbi

Cite this article as: Tanya Y. Berger-Wolf, Cristopher Moore and Jared Saia, A computational approach to animal breeding, *Journal of Theoretical Biology*, doi:10.1016/j.jtbi.2006.08.028

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

A Computational Approach to Animal Breeding

Tanya Y. Berger-Wolf², Christopher Moore¹, and Jared Saia¹

¹ Department of Computer Science, University of New Mexico, Albuquerque, NM 87131, USA
 {moore,saia}@unm.edu,

² Department of Computer Science, University of Illinois at Chicago, 851 S. Morgan St, Chicago, IL 60607, USA
 tanyabw@uic.edu

Abstract. We propose a computational model of mating strategies for controlled animal breeding programs. A mating strategy in a controlled breeding program is a heuristic with some optimization criteria as a goal. Thus, it is appropriate to use the computational tools available for analysis of optimization heuristics. In this paper we propose the first discrete model of the controlled animal breeding problem and analyze heuristics for two possible objectives: 1) breeding for maximum diversity and 2) breeding a target individual. These two goals are representative of conservation biology and agricultural livestock management respectively. We evaluate several mating strategies and provide upper and lower bounds for the expected number of matings. While the population parameters may vary and can change the actual number of matings for a particular strategy, the order of magnitude of the number of expected matings and the relative competitiveness of the mating heuristics remains the same. Thus, our simple discrete model of the animal breeding problem provides a novel viable and robust approach to designing and comparing breeding strategies in captive populations.

1 Introduction

Controlled breeding programs are common in many contexts, ranging from conservation biology and zoo animal management (Ballou and Foose, 1995; Lacy, 1994; Tonkyn, 1993; Tudge, 1992) to agricultural livestock management (CSIRO; Schneider, 1997), to laboratory animal breeding for research purposes (Lane-Petter, 1963; Amphibians, 1974). In most of these contexts heuristic mating strategies are used. These heuristics are generally based on personal experience and have not been extensively evaluated or analyzed (Lacy, 1994; Tonkyn, 1993). In this paper, we propose a discrete optimization model of the controlled breeding problem and demonstrate our approach by evaluating the effectiveness of common mating heuristics for two specific goals of breeding programs, which lie at two extremes: maximizing diversity vs. producing a particular genotype.

The first goal we consider is to produce all possible gene combinations. This is a simplified special case of the common problem in zoo breeding programs of maximizing the genetic diversity of the population in order to prevent random and directional genetic changes (Lacy, 1994). We give lower and upper bounds on the average number of matings—i.e., the number of matings that is, on average, necessary and sufficient—to achieve this goal. We also analyze the average number of matings for an analog of a strategy used in zoo breeding programs (Lacy, 1994) and show that the average number of matings needed can be significantly reduced by modifying it slightly.

The second goal we consider is breeding an animal with a specified phenotype or genotype. This is a common goal of an agricultural animal breeding program. We compare the simplified analogs of two breeding heuristics: 1) iteratively mating animals closest overall to the target and 2) breeding for one genetic trait at a time. Both strategies are commonly used in livestock management (e.g. CSIRO). We analyze the average worst case number of matings for both strategies and show that the latter, surprisingly, is more efficient.

These two objectives are representative of the goals of breeding programs in conservation biology and agricultural livestock management, and thus show the wide range of applicability of our approach. While the population parameters may vary and can change the actual number of matings for a particular strategy, the order of magnitude of the number of matings on average and the relative competitiveness of the mating heuristics remain the same. Thus, our simple computational model of the animal breeding problem provides a novel, viable and robust approach to designing and comparing breeding strategies in captive populations. Our approach is different from the existing methodology for analysis of breeding strategies in that it does not follow a step-by-step process modeling the state of a population (such as gene dropping (Lacy, 1994;

MacCluer et al., 1986; Princée, 1988) or pedigree simulation (Harris et al.; Seal and Lacy, 1989)) but rather looks at any such process as an algorithm and uses algorithmic analysis techniques to evaluate and compare various strategies.

1.1 Discrete optimization model

In this section, we present the theoretical framework used for our analysis. Each animal is represented by a binary string of n bits, each bit corresponding to a genetic marker or a trait. The results in this paper are easily extendible to an alphabet of arbitrary size (other than two), allowing us to represent alleles, nucleotides, etc. Below we list the simplifying assumptions made in our breeding model. We discuss the impact of relaxing these assumptions in Section 4.

- Any two animals can mate, i.e., there is no gender. We can easily remove this assumption without changing our results significantly.
- A mating of 2 animals is a single atomic operation. This assumption can be relaxed in many ways, from fixing the number of matings per animal to the complicated concept of a generation.
- Each mating produces exactly one offspring. Again, our results are easy to generalize to multiple offspring, or to including the probability of success (fertility) of a mating.
- There are no deaths in the course of the breeding program.
- We use a simple Mendelian model of inheritance. We assume the *iid* probabilistic model on the genes (or the set of traits). That is, the outcome of a mating for each bit depends only on the values of that bit in the two parents. If a bit is the same in both parents then it is the same in the offspring. If the parents' bit values are different then the offspring has a 0 or 1 for that bit, taking its value from one parent or the other, with equal probability. We make this assumption since in most cases very little is known about the dependencies between genetic loci and the probabilities of particular outcomes (Falconer, 1981).

Given an initial population of k animals, we concentrate on two particular goals. We want to minimize the average number of matings required to achieve these goals.

1. *Maximum diversity*: generate all the binary strings that can possibly arise from the initial population. Our results for this breeding goal are described in Section 2.
2. *Target set of traits*: breed for an animal represented by a target binary string. Our results for this breeding goal are described in Section 3.

We note that our upper bounds apply to any initial population, since in our derivations we assume that the initial populations is the worst possible to achieve our goals. If the initial population is random, say, or if its distribution can be characterized in some way, then more efficient algorithms—or tighter upper bounds on the average number of matings that our algorithms use—may be possible.

1.2 Past work

There is an extensive toolkit of methods for analysis of various aspects of breeding populations, from kinship and inbreeding coefficient estimates (Ballou, 1983; Boyce, 1983; Lacy, 1994; Wright, 1921, 1969) to calculations of loss of the genetic diversity (Chesser, 1983; Crow and Kimura, 1970; Harris and Allendorf, 1989; Lande and Barrowclough, 1987; Wright, 1931, 1969) (Lacy et al. (1995) provides a good survey of the existing methodology). Presently, analysis and comparison are often done using stochastic modeling (Ballou and Lacy, 1995; Lacy, 1994; Lacy et al., 1995), that models the state of the population at each time step using the variety of the genetic analysis methods mentioned above to update the state. However, in many cases, the empirical statistical information underlying the model (how much does inbreeding affect survival rates, for instance?) is very hard to obtain. It is often collected for a different animal population (e.g. the statistics of the black rhinoceros in Kenya used for populations in other parts of Africa); or the population is too small and has large variance; or the environmental conditions have changed since the information was last collected. For these and many other biological and mathematical reasons, a time-step-based stochastic approach is either not robust or is simply impossible.

Often the information sought from the modeling process is qualitative rather than quantitative. In the case of designing and evaluating strategies for captive animal breeding, the main question we want to answer is: “Which strategy is better?”. This is inherently a question of algorithm analysis. In this paper, we answer this question by formulating the controlled breeding as a discrete optimization problem. For this reason, we do not make use of any intermediate stochastic models.

2 Breeding for maximum diversity

In this section, we focus on the goal of producing all possible gene combinations. We assume that the initial population can actually produce all possible binary strings. That is, we ignore the bits that appear with a single value within the population. If the value of some bit is the same in all the strings of the initial population then there is, of course, no way to produce a string with the other value for that bit (in the absence of mutation). Thus we assume that all n bits appear with both values 0 and 1 in the initial population, so that all 2^n possible strings can be created from the initial population.

We now consider lower and upper bounds on the average number of matings to create all 2^n possible strings. A lower bound is an absolutely necessary average number of matings, no matter what particular strategy is used. An upper bound is a sufficient average number of matings. Commonly, but not always, an upper bound is an analysis of a particular strategy. (Clearly, if there exists a strategy that achieves the desired goal within some average number of matings then that number is sufficient and serves as an upper bound.) Ideally, if a particular strategy has an average number of matings equal to the lower bound then this strategy is optimal, since it uses no more matings, on average, than are absolutely necessary. In practice, it is often hard to prove a good lower bound. Thus, a strategy may be optimal even if its average number of matings is greater than the best known lower bound.

If the size of the initial population is k , an obvious and extremely optimistic lower bound for the maximum diversity goal is $2^n - k$. This bound assumes that every mating produces a new string; however, this is clearly unrealistic, since many matings produce strings already present in the population. We can show a slightly improved lower bound of $2^n - k + O(n)$, where $O(n)$ is some linear function of n . For our upper bound, we propose a GREEDY strategy that for each successive mating chooses the two animals with the highest probability of producing a new string. This strategy is in the spirit of that used in zoo breeding programs where the two “least inbred” animals are selected for each mating. However, we show that the GREEDY algorithm has at most $2.3125 \cdot 2^n - k$ matings on average, while the one used by zoos has about $.7n2^n$ —thus GREEDY is more efficient by a factor of n , the number of traits. Although the GREEDY strategy requires at most about 2.3 times as many matings as our lower bound, we believe that neither the upper and the lower bounds are tight and that the GREEDY approach is actually optimal.

2.1 Upper bound

We now present an upper bound for the maximum diversity breeding problem and compare various strategies. First we show that we can ensure that the probability of producing a new animal is always at least $1/4$. To do this, we introduce the notion of *binary hypercubes*. An n -dimensional binary hypercube is a finite n -dimensional binary lattice that consists of nodes that correspond to binary strings of n bits and edges between them. We say that two strings are a distance t apart if they differ in t bits. The nodes in a binary hypercube are connected by an edge if they are distance 1 apart. Figure 1 shows the 3-dimensional hypercube. Notice that this hypercube has six 2-dimensional hypercubes as subcubes; for example, we say that the nodes 011,010,001,000 are *spanned* by the initial population 010, 001, in the sense that this subcube is the set of all strings that can be generated by that initial population. As an extreme case, each edge is a 1-dimensional subcube, spanned by the strings at its two endpoints.

Lemma 1. *Unless all the strings have been produced, there always exists a pair with probability at least $1/4$ of producing a new string.*

Proof. Assume, to the contrary, that for every two animals in the population the probability of producing a new string is less than $1/4$. This means that the hypercube spanned by any pair of animals in the population

Fig. 1. The 3-dimensional hypercube. Each face is a 2-dimensional hypercube and each edge is a 1-dimensional hypercube.

has more than $3/4$ nodes already present in the population. That is, for any two present nodes distance t apart, the hypercube spanned by the two nodes has at most $2^{t-2} - 1$ nodes not yet present in the population. Specifically, for any two present nodes distance 2 apart the entire 2-dimensional subcube they span must be present in the population. The presence of any node adjacent to this 2-dimensional subcube necessarily means that the entire 3-dimensional subcube is spanned. This follows from the fact that both 2-dimensional faces formed by the adjacent node and the 2-dimensional subcube must be entirely present and the nodes of these faces span the rest of the 3-dimensional cube's 2-dimensional faces. In general, if each node of a t -dimensional subcube is present in the population then, by a similar inductive argument, any adjacent node forces the entire $t + 1$ -dimensional subcube to be present. Thus the nodes present in the population form a disjoint set of completely filled subcubes.

Unless all strings are present in the population, there are at least two such filled subcubes. Choose the two closest nodes from different subcubes. They must be at least distance $t \geq 2$ apart (otherwise the subcubes are not disjoint) and the subcube they span is completely empty (otherwise they are not the closest nodes). Thus a mating of these two nodes produces a new string with probability $(2^t - 2)/2^t$, which is greater than $1/4$ if $t \geq 2$. This is a contradiction, so unless we have already produced every string there is some pair that produces a new animal with probability at least $1/4$.

We note that we can have very few strings present in the population, namely $n + 1$, and still have only probability $1/4$ of producing a new string. Consider an n -dimensional hypercube where the only strings present in the population are the all-zero string and strings with a single bit set to 1. Mating any two such strings produces a new string, namely one with two bits set to 1, with probability $1/4$; with probability $3/4$ we get a string already in the population.

Corollary 1. *The GREEDY strategy that always mates with with the highest probability to produce a new animal takes at most $4(2^n - k)$ matings on average.*

However, we will show that the GREEDY strategy performs better than this, bringing the factor 4 down considerably. Specifically, we now show that $2.3125 \cdot 2^n$ average number of matings is sufficient to produce all possible strings and that the GREEDY strategy requires no more than that on average. This upper bound relies on the existence of two complementary strings in the population, i.e., strings that differ on every bit. Using the techniques of Section 3, we can obtain such a pair with at most $O(n \log n)$ matings on average, which is negligible compared to 2^n . Hence, we can ignore the $O(n \log n)$ term and assume that the complementary strings indeed exist in the population already.

Theorem 1. *If two complimentary strings exist in the founding population then the expected mating sequence length is at most $2.3125 \cdot 2^n - k$.*

Proof. As we have shown earlier, there always exists a pair of animals whose probability of producing a new string is at least $1/4$. Therefore, we can mate the complementary strings until there are at most $3/4 \cdot 2^n$ animals in the population, and then mate any pair with at least $1/4$ probability of producing a new string. Assuming that we start with k distinct strings in the initial population (including the two complementary ones), the average number of matings is:

$$\begin{aligned} E(\text{number of matings}) &= \sum_{t=k+1}^{\frac{3}{4}2^n} \frac{2^n}{2^n - t} + 4 \frac{1}{4} 2^n \\ &= \sum_{t=k+1}^{\frac{3}{4}2^n} \frac{1}{1 - 2^{-n}t} + 2^n \end{aligned}$$

$$\begin{aligned}
&\leq \sum_{t=k+1}^{\frac{3}{4}2^n} 1 + 2^{1-n}t + 2^n, && \text{if } n \geq 2 \\
&= \frac{3}{4}2^n - k + 1 + 2^{1-n} \frac{(3/4 \cdot 2^n + k)(3/4 \cdot 2^n - k + 1)}{2} + 2^n \\
&= \left(\frac{3}{4}2^n - k + 1\right) \left(\frac{7}{4} + \frac{k}{2^n}\right) + 2^n \\
&= 2.3125 \cdot 2^n - k + \frac{k}{2^n}(1 - k) + 1.75
\end{aligned}$$

Since at any point the *maximum* probability of producing a new animal is at least that of the above specified algorithm, and the greedy algorithm selects the strings with the maximum probability, we can state the following corollary.

Corollary 2. *The expected number of matings used by the greedy algorithm is at most $2.3125 \cdot 2^n$.*

Note that the actual GREEDY algorithm discussed here is the following

Algorithm Greedy

1. If no complimentary animals exist in the population, breed an animal complimentary to one of the animals in the population
2. Mate two animals with the highest probability of producing an animal not yet in the population

The first step does not require any computations in addition to the $O(n \log n)$ matings. The second step requires at most $2.312 \cdot 2^n$ matings. However, each mating requires additional computations to find the two animals with the highest probability of producing a new animal. One possible way to achieve this is with a *priority queue* data structure (Cormen et al., 2001) of all the animal pairs in the population. For each pair (a, b) the data structure maintains the probability of (a, b) to produce a new animal: $pr(a, b)$. The priority queue automatically returns the pair with the maximum probability. In addition, for each pair we maintain pointers to all the animals that are in the span of this pair: $span(a, b)$. When a new animal c is produced we update the data structures as follows:

Update Probability (c)

```

FOR each pair  $(a, b)$ ,  $a, b \neq c$ 
  IF for all bits  $i$  such that  $a_i = b_i$  the bit  $c_i = a_i$  THEN
    add  $c$  to  $span(a, b)$ 
     $pr(a, b) \leftarrow pr(a, b) - \frac{1}{2^{H(a, b)}}$  where  $H(a, b)$  is the hamming distance between  $a$  and  $b$ 
  FOR all  $a$ 
    FOR all  $b > a$ 
      FOR all  $d \in span(a, b)$ 
        IF for all bits  $i$  such that  $a_i = c_i$  the bit  $d_i = a_i$  THEN
          add  $d$  to  $span(a, c)$ 
        ELSE add  $d$  to  $span(b, c)$ 
       $pr(b, c) = 1 - \frac{|span(b, c)|}{2^{H(b, c)}}$ 
       $pr(a, c) = 1 - \frac{|span(a, c)|}{2^{H(a, c)}}$ 

```

Thus, the update of the probability value when a new animal is produced takes $O(2^{2n})$ operations (since there are at most 2^{2n} pairs) and the priority queue maintenance is constant for all practical purposes with the current best implementations. Some time may be saved by eliminating pairs that have $pr(a, b) = 0$ but the overall asymptotic worst case complexity remains the same.

2.2 Comparison with a currently used strategy

The GREEDY strategy always mates the two animals in the population that have the highest probability of producing an animal not yet in the population. However, the strategy currently used by the conservation breeding programs is to breed the two animals whose pedigrees differ by as much as possible; in our model, the two strings farthest apart from each other in the hypercube. We will call this the DISSIMILAR strategy. In particular, this strategy will breed any two complementary strings if they exist. As we discussed earlier, we can assume that two complementary strings indeed exist in the population since the number of matings needed to ensure this is negligible compared to the total number of matings needed to produce all the strings.

Theorem 2. *The minimum expected number of matings used by the DISSIMILAR algorithm is at least $n \ln 2 \cdot 2^n$.*

Proof. Mating two complementary strings will create a string which is chosen uniformly at random from the set of all possible strings. We can then ask how many matings of these complementary strings are required before all strings have been created at least once. This is known as the *coupon collector's problem* (Cormen et al., 2001), and the answer for $X = 2^n$ different strings is $X \ln X = n \ln 2 \times 2^n$.

Even this lower bound for the DISSIMILAR strategy is greater than the upper bound on the GREEDY strategy, which is a constant times 2^n , roughly by a factor of n . For example, if we wish to create a population of 1,024 genetically distinct animals with all possible combinations of $n = 10$ traits, the DISSIMILAR strategy requires at least 7,168 matings on average while the GREEDY strategy requires at most 2,367 matings on average. The advantage of GREEDY becomes even greater as n increases.

Thus, GREEDY provably performs significantly better than DISSIMILAR in terms of the average number of matings in our model. However, GREEDY requires more computational effort than DISSIMILAR as discussed above. However, the time it takes to do this computation is surely negligible compared to the time and effort required to actually breed two animals!

3 Breeding a target animal

We now focus on the problem of breeding an animal with a specified target set of gene variations or traits. For simplicity we assume that the target string is the all-ones string $111 \dots 1$. We also assume that this string is a possible descendant of the initial population, i.e., in the subcube spanned by the initial population; this simply means that for every i , there is some initial string whose i 'th bit is 1. Again, we wish to find lower and upper bounds on the average number of matings required to produce this target string.

In the worst case, the founder population consists of n strings, x_1, x_2, \dots, x_n , where the i 'th bit of x_i is 1 and its others are 0. This is indeed the worst case since if there is an founder individual y whose i 'th bit is 1 and it has additional ones in its string, then for the purposes of producing the all-ones string we can only gain by using y in lieu of x_i . In this case, an obvious lower bound on the number of matings needed to get the all-ones string is $n - 1$, since every x_i must appear at least once as one of its ancestors.

In the remainder of this section, we show an upper bound of $2n \log_2 n$. First, we consider the following subproblem. We are given a string x with t ones and a string y with a single one in a position where x does not have a one. Our goal is to produce a string with a 1 in every position where either x or y has a 1.

Lemma 2. *Given a string x with i ones and a string y with a single 1 that are Hamming distance $i + 1$ apart, it takes at most $2 \cdot \lg i$ expected number of matings to produce the string with $i + 1$ ones.*

Proof. This is achieved by the following algorithm ADDONETRAIT:

Algorithm AddOneTrait

```

WHILE  $y$  has less than  $t + 1$  ones DO:
  1. Mate  $x$  and  $y$  repeatedly until we produce a string  $y'$  with the following properties:
    (i)  $y'$  has 1s in all the bit positions where  $y$  has a 1; and (ii)  $y'$  has 1s in at least half
    the bit positions that are 1s in  $x$  but not in  $y$ .
  2.  $y \leftarrow y'$ 
RETURN  $y$ 

```

Without loss of generality, let x have ones in the first t bits and y have a 1 in the bit $t + 1$. After the first two matings of the original x and y we expect one of the offspring to have a 1 in the $(t + 1)$ 'st bit and ones in half of the first t bits. We choose this to be our new partner for x . In any mating of this string with x , all the common 1-bits will, of course, be retained. One of the two matings of this new string with x is expected to produce a string with a 1 in the $(t + 1)$ 'st bit and ones in about half of the bits which are 1 only in the string x . Thus, after two matings, we expect to have a string with a 1 in $(t + 1)$ st bit and $3/4 \cdot t$ ones in the first t bits. Continuing in this manner (by induction), after $\log_2 t$ iterations on average we produce a string with $t + 1$ ones. Since each iteration takes two matings on average, the total expected number of matings is $2 \cdot \log_2 t$.

Theorem 3. *For any set of animals in the founding population the expected length of the mating sequence that produces the string of all ones is at most $2n \log_2 n$ (if the target string is in the span of the founding population).*

Proof. This is achieved by the algorithm BREEDTARGET. We assume that the initial population consists of the x_i defined above, since this is the worst case. The target string is the all-ones string, so our goal is to reach $t = n$.

Algorithm BreedTarget

```

 $x \leftarrow x_1$ 
FOR  $t = 2$  to  $n$  DO:
  1. Call AddOneTrait on  $x$  and  $x_t$  to produce a string,  $y$ , which has 1s in positions 1 through  $t$ .
  2.  $x \leftarrow y$ 
RETURN  $x$ 

```

It is straightforward to see that the total expected number of matings for BREEDTARGET is at most $2n \log_2 n$. The t 'th call to ADDONETRAIT requires $2 \log_2 t$ matings in expectation. Therefore the total expected number of matings is

$$\sum_{t=1}^{n-1} 2 \log_2 t \leq 2n \log_2 n.$$

Biologically, this result can be interpreted as follows. Our algorithm repeatedly adds one new trait to the population and then uses back-breeding to breed an animal with both this new trait and all the previously added traits. The downside of this strategy is that new offspring are bred back with one of the parents several times, producing a highly inbred population. However, such practices are not uncommon in agricultural animal breeding programs so we do not consider it unrealistic.

We now compare our algorithm with a greedy GREEDYTARGET strategy, that always mates the pair of individuals that are currently closest to the target animal. Again, we assume the worst case founder population of n strings with a single 1 bit each.

Theorem 4. *For the founder population of $\{x_i\}$ strings GREEDYTARGET algorithm requires at least $n^2/2$ expected number of matings to produce an all-ones string.*

Proof. The final mating that produces the all-ones string is, in expectation, a mating of two strings with $n - 1$ ones each (otherwise it takes more than a constant expected number of matings to produce an all-ones string in that last step). There would need to be at least three strings with $n - 2$ ones to produce these $n - 1$ ones' strings. Continuing by induction, such greedy strategy would need a total number of strings of

$$\sum_{t=1}^n t = \frac{n(n+1)}{2} > n^2/2 .$$

Thus, the greedy strategy requires asymptotically more (when $n \geq 15$) than the $2n \log_2 n$ matings on average required by our algorithm.

4 Conclusions and extensions

To the best of our knowledge, we have formulated the first combinatorial computational model of the controlled breeding problem, allowing us to design and evaluate breeding strategies in the context of discrete optimization. We have used this model to analyze common mating heuristics for two specific goals of a breeding program: 1) breeding for maximum diversity and 2) breeding a target animal. These two goals are representative of two extreme examples of breeding programs, namely conservation biology and agricultural livestock management. We have proved upper and lower bounds for the average number of matings to achieve these breeding goals. Overall, we have demonstrated the viability and robustness of our mathematical approach to analyzing controlled breeding problems.

In formulating our discrete optimization model of the breeding problem, we made many simplifying assumptions. However, most of these are easy to relax: for example, introducing the concept of gender rather than allowing any pair of animals to mate at most doubles the average number of matings for any breeding strategy. Similarly, assuming that each mating can produce up to a fixed number of offspring reduces the number of matings by no more than a constant factor. One non-trivial assumption we made is that only two animals can mate at any step, i.e., that breeding takes place serially rather than in parallel. It would be interesting to investigate various strategies that minimize the number of *generations* needed to achieve a breeding goal. There are several ways to define a generation in mathematical terms, each one providing a challenging optimization problem. Finally, the assumption of the *iid* probabilistic distribution model of genes is the most challenging from a biological point of view, since little is known about genetic transfer. As biologists learn more about this area, we can develop more realistic models of controlled breeding.

Acknowledgments

We would like to thank Robert Lacy and Onnie Byers of the Conservation Breeding Specialists Group and Sarah Long from the Chicago's Lincoln Park Zoo for providing information and pointing out the relevant literature. We would also like to thank Alex Russell, Bernard Moret, Alla Sheffer, and the anonymous reviewers for many useful comments. This work was supported by the National Science Foundation postdoctoral fellowship grant EIA 02-03584 (Tanya Berger-Wolf), by the National Science Foundation grant CCR-0313160 (Jared Saia), and by the Sandia University Research Program Grant No. 191445 (Jared Saia).

Bibliography

- Amphibians (1974) , *Amphibians: guidelines for the breeding, care, and management of laboratory animals*, Institute of Laboratory Animal Resources (U.S.). Subcommittee on Amphibian Standards
- Ballou, J. D. (1983) , In *Genetics and Conservation: A Reference for Managing Wild Animal and Plant Populations*. (C. M. Schonewald-Cox, S. M. Chambers, B. MacBryde, and W. L. Thomas eds.), pp. 509–520, Menlo Park: Benjamin/Cummings
- Ballou, J. D. and Foose, T. J. (1995) , In *Wild Animals in Captivity*. (D. G. Kleinman, S. Lumkin, M. Allen, H. Harris, and K. Thompson eds.), pp. 263–283, University of Chicago Press, Chicago, USA
- Ballou, J. D. and Lacy, R. C. (1995) , In *Population Management for Survival and Recovery*. (J. D. Ballou, M. Gilpin, and T. J. Foose eds.), Methods and Cases in Conservation Biology, pp. 76–111, Columbia University Press, New York, USA
- Boyce, A. J. (1983) , *Journal of Heredity* **74**, 400
- Chesser, R. K. (1983) , In *Genetics and Conservation: A Reference for Managing Wild Animal and Plant Populations*. (C. M. Schonewald-Cox, S. M. Chambers, B. MacBryde, and W. L. Thomas eds.), pp. 66–77, Menlo Park: Benjamin/Cummings
- Cormen, T. H., Leiserson, C. E., Rivest, R. L., and Stein, C. (2001) , *Introduction to Algorithms*, Chapt. 5.4.2, pp. 109–110, The MIT Press, Cambridge, MA, McGraw-Hill Book Company, 2nd edition
- Crow, J. F. and Kimura, M. (1970) , *An Introduction to Population Genetics Theory*, Harper and Row, New York
- CSIRO, *Breeding for worm resistance: A component of sustainable worm control*, Commonwealth Scientific and Industrial Research Organisation Livestock Industries, CSIRO Livestock Industries, Armidale NSW 2350, Australia, <http://www.csiro.au/proprietaryDocuments/CSIROLIBreedingForWormResistance.pdf>
- Falconer, D. S. (1981) , *Introduction to Quantitative Genetics*, New York: Longman, 2nd edition
- Harris, R. B. and Allendorf, F. W. (1989) , *Conservation Biology* **7**, 289
- Harris, R. B., Metzgar, L. H., and Bevins, C. D., *GAPPS (Generalized Animal Population Projection System) User's Manual*, Montana Cooperative Wildlife Research Unit, University of Montana, Missoula, Montana
- Lacy, R. C. (1994) , In *Restoration of Endangered Species: Conceptual Issues, Planning, and Implementation*. (M. L. Bowles and C. J. Whelan eds.), pp. 63–89, Cambridge University Press
- Lacy, R. C., Ballou, J. D., Princée, F., Starfield, A., and Thompson, E. A. (1995) , In *Population Management for Survival and Recovery*. (J. D. Ballou, M. Gilpin, and T. J. Foose eds.), Methods and Cases in Conservation Biology, pp. 76–111, Columbia University Press, New York, USA
- Lande, R. and Barrowclough, G. F. (1987) , In *Viable Populations for Conservation*. (M. E. Soulé ed.), pp. 87–123, Cambridge University Press, Cambridge
- Lane-Petter, W. (1963) , *Animals for research; principles of breeding and management*, London, New York, Academic Press
- MacCluer, J. W., VandeBerg, J. L., REad, B., and ryder, O. A. (1986) , *Zoo Biology* **6**, 147
- Princée, F. P. G. (1988) , *Zoo Biology* **7**, 219
- Schneider, K. (1997) , *Computers and information technologies in agricultural production and management. Part 2 : January 1994 - June 1997*, Vol. 10 of *Quick bibliography series*, Beltsville, Md. : USDA, ARS, National Agricultural Library
- Seal, U. S. and Lacy, R. C. (1989) , *Florida Panther Population Viability Analysis*, Captive Breeding Specialists Group, Species Survival Commission, IUCN, Apple Valey, Minnesota, Report to the U. S. Fish and Wildlife Service
- Tonkyn, D. W. (1993) , *Endangered Species Update* **10(8)**, 1
- Tudge, C. (1992) , *Last animals at the zoo : how mass extinction can be stopped*, Washington, D.C. : Island Press
- Wright, S. (1921) , *Genetics* **6**, 111
- Wright, S. (1931) , *Genetics* **16**, 97
- Wright, S. (1969) , *Evolution and the Genetics of Populations. Vol. II: The Theory of Gene Frequencies.*, Chicago. University of Chicago Press