Selective Breeding analysed as a Communication Channel: Channel Capacity, and a Fundamental Limit on Adaptive Complexity

Chris Watkins
Department of Computer Science
Royal Holloway, University of London
Surrey TW20 0EX, UK
C.Watkins@cs.rhul.ac.uk

Abstract

Selective breeding is considered as a communication channel, in a novel way. The Shannon informational capacity of this channel is an upper limit on the amount of information that can be put into the genome by selection: this is a meaningful upper limit to the adaptive complexity of evolved organisms. We calculate the maximum adaptive complexity achievable for a given mutation rate for simple models of sexual and asexual reproduction. A new and surprising result is that, with sexual reproduction, the greatest adaptive complexity can be achieved with very long genomes, so long that genetic drift ensures that individual genetic elements are only weakly determined. Put another way, with sexual reproduction, the greatest adaptive complexity can in principle be obtained with genetic architectures that are, in a sense, error correcting codes. For asexual reproduction, for a given mutation rate, the achievable adaptive complexity is much less than for sexual reproduction, and depends only weakly on genome length.

A possible implication of this result for genetic algorithms is that the greatest adaptive complexity is in principle achievable when genomes are so long that mutation prevents the population coming close to convergence.

1. Introduction

Complex organisms become intricately adapted to their environments after long eons of natural selection. In some sense, natural selection creates genetic information that specifies the structure of the organism. Most of this information is encoded in the genome; in each generation, the information is degraded by mutation, and restored or increased by selection.

We will pose and answer some natural and basic questions about the amount of genetic information that can be produced by selection.

We will partially answer these questions for a simple model of evolution, variants of which have been independently studied by many people. In population genetics it is the standard model of linear selection with full linkage equilibrium, as described in classic population genetics texts such as, for example, [CK70] and [Ewe79]. Similar models have been studied in machine learning by [DH97] and as a simplified form of genetic algorithm by [BBG95]; in the genetic algorithms (GA) community they have been studied by [Bal94], [PGCP99], [HLG99], and others.

It has long been known that under certain assumptions sexual reproduction can be evolutionarily advantageous: an early study was [CK79]; [BBG95] and [Mac03] are more recent analyses. [Mac03] (chapter 20) defines a measure of the rate of acquisition of information of a species from selection, but that definition of information is different from the one developed here.

In contrast to previous studies, we quantify the advantage of sexual over asexual reproduction in terms of the maximal amount of *information* that can be maintained in the genome. We also compare the amount of information that can be maintained using both compact and highly distributed genetic codes: for sexual reproduction, it turns out that there is an enormous potential advantage in using highly distributed codes over long genomes.

It is not straightforward to define a suitable notion of the amount of information in the genomes of a population that is the result of selection rather than genetic drift. The notion we will next describe is very similar to the notion of "physical information" that was introduced by [Ada02]. We motivate the definition with two thought-experiments, and we argue that the appropriate way to consider achievable adaptive complexity is as the informational capacity of a communication channel: informational channel capacity is a standard concept in information theory, as described in numerous texts such as [CT91] and [Mac03].

2. A thought experiment: selective breeding as communication

The key question is how to define the information that is in the genomes of a species *as a result of selection*. Not all parts of the sequence of a genome are informative: many features of real genomes are determined by "random genetic drift" rather than by selection. But genetic drift is nothing but random selection: how can one distinguish the effects of random selection from "real" selection? A direct way to make the distinction to consider selective breeding as a communication channel, in the following way.

Suppose that Alice and Bob are two geneticists: Alice is to be imprisoned, and wishes to send messages to Bob from her cell. The only possible method is for Alice to capture wild Drosophila, and to breed them selectively for many generations in her cell. She must encode her message in the flies' genomes by means of selective breeding alone (we suppose that direct modification of genomes by genetic engineering is against the prison rules). The message that Alice sends corresponds to the criterion that she uses to select flies as she breeds them. On a previously appointed day, she releases her final generation of flies. The faithful Bob, waiting outside, catches one of them, sequences its genome, and decodes Alice's message according to the encoding system they agreed to use before Alice was sent to prison. How much information can Alice send, and what code should Alice and Bob agree to use? This is a concrete and practical, though fanciful, question that deserves an answer.

It is important that we require that Bob should capture only one fly. All the flies that Alice releases should therefore carry her message. This requirement corresponds to the intuition that the information to construct a member of a species is present in the genome of each member of the species. If, instead, Bob were allowed to capture a large number of flies, then Alice could encode her message in the population structure: this could be done within a single generation, rather than by producing a new variety of fly over many generations. With such a "population code", each individual fly would carry only a tiny part of the message, so that this manner of transmitting information would not be relevant to explaining the evolution of complex organisms, which is our goal. We will not, therefore, consider population codes further.

Any information that Alice can send to Bob in this way must be the result of Alice's selections of which flies to breed. The capacity of this communication channel, therefore, is a conservative measure of the amount of information that Alice can put into her flies' genomes by selective breeding.

One possible method for Alice and Bob to use would be to have a code-book of distinct varieties of fly, each of which Alice could reliably produce by selective breeding. Each time Alice set out to breed a particular variety, she would produce a detectably different final population – an achievable variety would need to be defined sufficiently broadly to ensure that Alice could produce it by following a prespecified selection policy. The amount of information that could be sent in this way would be the log of the number of distinct achievable varieties in the code-book. Although the notion of a code-book of distinct, achievable varieties is concrete and intuitively attractive, the notion of channel capacity is formally more convenient, and will be used below.

From the point of view of the flies, the channel capacity measures the variety and precision of the fly population's possible responses to selection. The structural adaptation of an individual fly to its environment is limited by the amount of information from selection that is stored in its genome. In principle, the greater the number of distinct achievable varieties, the more precise and well-specified the fly can be, both in body and innate behaviour, and the greater the range of possible responses to environmental challenge.

2.1. A formal framework for describing selective breeding

In a more formal model, we view the breeding population as a collection of genomes, and consider selection to be performed directly on known genome sequences. We will first set up a general framework that can be applied to many computational models of evolution, and we will then consider two specific models.

Let the (finite) set of all possible genomes be $\mathcal{G}$. A breeding population of genomes, which will be called a *collection* of genomes, is denoted as $\mathbf{c} = (\mathbf{x}^1, \ldots, \mathbf{x}^n)$. Three operations are defined on collections: selection, breeding, and mutation.

A *selection rule* s assigns a weight to each genome in a collection: that is, $s(\mathbf{c}) = \mathbf{w}$, where $\mathbf{w} = (w_1, \ldots, w_n)$ and $w_i \geq 0$ and $\sum_i w_i = 1$. We suppose that there is a set $\mathcal{S}$ of possible selection rules that a breeder can apply.

A *breeding system* b is a stochastic function that constructs a new collection from an existing weighted collection: $C = b(\mathbf{c}, \mathbf{w})$, where C is a random variable ranging over the set of possible collections.

A *mutation function* m modifies the genomes in a collection by incorporating mutations. m is also a stochastic function, in the sense that, for a collection $\mathbf{c}$, $m(\mathbf{c})$ is a random variable ranging over all possible collections of the same size as $\mathbf{c}$.

Given a selection rule s and a starting collection C^0, we may construct a sequence of collections $C^1, C^2, \ldots, C^t, \ldots$, such that for $t = 1, 2, \ldots, T$, $C^{t+1} = m(b(C^t, \mathbf{w}^t))$, where $\mathbf{w}^t = s(C^t)$. Note that the same selection rule s is used for all T generations. We also define an associated sequence $X^0, X^1, \ldots, X^T$ such that

$X^t \sim (C^t, \mathbf{w}^t)$: that is, X^t is sampled from C^t according to the probabilities $\mathbf{w}^t$. C^t is a Markov sequence, but X^t is in general not Markov.[1]

We define an *evolutionary system* (ES) $\langle \mathcal{G}, \mathcal{S}, b, m, C^0, n, T \rangle$, consisting of a set $\mathcal{G}$ of possible genomes, a set $\mathcal{S}$ of possible selection rules, a breeding system b, a mutation function m, a starting population C^0 (which may be a random variable), population size n, and a stopping time T.

An ES may be viewed as a communication channel in the following sense. The message sender chooses a selector $s \in \mathcal{S}$: s is the "message" that is "sent". Starting with a collection distributed as C^0, a sequence of collections $C^1, \ldots, C^T$, each of size n, is generated, such that $C^{k+1} = m(b(C^k, s(C^k)))$. Finally the sample $X^T \sim (C^T, \mathbf{w}^T)$ is the message that is "received". The receiver of the message may then infer some information about s by examining X^T. All characteristics of the system, including the stopping time T are known to the receiver: the receiver is ignorant only of the sender's choice of s.

The channel capacity I of an ES is defined using a "sending" probability distribution Q over $\mathcal{S}$, so that the selection rule used is a random variable $S \in \mathcal{S}$ and such that $S \sim Q$. The channel capacity I is defined in the standard way as:

$$I = \max_Q \left\{ H(S) - H(S|X^T) \right\} = \max_Q \left\{ H(X^T) - H(X^T|S) \right\} \tag{1}$$

where H is the entropy function and conditional entropy is defined in the standard way.

3. A Simplified GA Model

We use a model identical to that of [BBG95]: very similar models have been used by others such as [DH97], and by [CK70] for the case of truncation selection with reversible mutation in full linkage equilibrium.

The set of possible genomes $\mathcal{G} = \{0,1\}^L$ for some chosen genome length L (we will consider the effect of varying the choice of L).

3.1. Selection Rules

We consider selection rules based on the Hamming distance from an "ideal genome" $\mathbf{z}$: the set $\mathcal{S}$ consists of one selection rule based on each of the 2^L possible values of $\mathbf{z}$. Define the level of agreement between a genome $\mathbf{x}$ and and "ideal" genome $\mathbf{z}$ as $f(\mathbf{x}, \mathbf{z}) := \frac{1}{L}\#\{i : x_i = z_i\}$, which is the fraction of indices at which they agree. Given a collection $C = (\mathbf{x}^1, \ldots, \mathbf{x}^n)$, let us define $f_i = f(\mathbf{x}^i, \mathbf{z})$. Let σ be a permutation of $(1, .., n)$ such that $f_{\sigma(1)} \geq \cdots \geq f_{\sigma(n)}$. Let $s_{\mathbf{z}} \in \mathcal{S}$ be the selection rule based on ideal genome $\mathbf{z}$, and, assuming n to be even, define

$$w_{\sigma(i)} := \begin{cases} \frac{2}{n} & \text{if } 1 \leq i \leq \frac{n}{2} \\ 0 & \text{if } \frac{n}{2} < i \leq n \end{cases} \tag{2}$$

In other words, we select for breeding the 50% of genomes in a collection that agree best with the "ideal genome" $\mathbf{z}$.

By symmetry, $H(X^T|s_{\mathbf{z}})$ will be equal for all $s_{\mathbf{z}} \in \mathcal{S}$, so to compute the channel capacity we need consider only the case where $\mathbf{z} = (1, 1, \ldots, 1)$ and let us define $f(\mathbf{x}) := \frac{1}{L}\sum_{i=1}^{L} x_i$.

3.2. Model of Sexual Reproduction

We model sexual reproduction as follows. A child genome $\mathbf{x}'$ is constructed by selecting each element x'_i from the corresponding element of a genome $\mathbf{x}$ of the parent population, where $\mathbf{x} \sim (C, \mathbf{w})$, and $\mathbf{x}$ is sampled afresh and independently for each x'_i. Each element of the child genome $\mathbf{x}'$ may be drawn from a different parent genome, therefore. This might be termed "hypersexual reproduction", since each child is a mixture of the alleles of the entire parent population, instead of having just two parents. This model of reproduction ensures that the elements of a child genome are independent Bernoulli variates.

This "hypersexual" model is simple to analyse but not fully biologically realistic. It is equivalent to the simplifying assumption of linkage equilibrium, often used in population genetics.

3.3. Model of Mutation

Finally, mutation is modelled by inverting each element of each genome with probability u, independently of other elements and other genomes. The parameter u is the *mutation rate*, and is typically small.

3.4. Channel capacity of sexual breeding

We will estimate the channel capacity of the evolutionary system defined in the previous section, when the stopping time T is large enough for the collections to have reached mutation-selection equilibrium. We write $X := X^T$ as a genome observed when the process has reached equilibrium.

Let $p := \mathbb{E}[f(X)]$, the expected fraction of 1s in the selected population at equilibrium. The maximum entropy distribution of X for a given value of p would be the factorial distribution in which the elements of X^T are independent Bernoulli variables, such that $\mathbb{P}(X_i^T = 1) = p$ for all i. The entropy of this distribution is $Lh(p)$ where

[1] In fact $X^1, \ldots, X^T$ is a sequence of observations from a hidden Markov model, but standard tools of HMM estimation turn out not to be needed for the analysis given below.

$h(p) := -p\log_2 p - (1-p)\log_2(1-p)$. The actual entropy of X must be less than equal to this.

As $H(X|s_{\mathbf{z}})$ is the same for all $s_{\mathbf{z}} \in \mathcal{S}$, the maximal channel capacity is achieved when the sending distribution Q is the uniform distribution over all $s_{\mathbf{z}}$, in which case, without the conditioning on S, X is uniformly distributed over $\mathcal{G}$, so that $H(X)$ attains its maximal possible value of L bits. The channel capacity is therefore

$$\begin{aligned} I &= H(X) - H(X|S) && (3)\\ &\geq L(1-h(p)) && (4)\\ &= \frac{2}{\ln 2}L(p-\tfrac{1}{2})^2 + O((p-\tfrac{1}{2})^4) && (5)\end{aligned}$$

using the Taylor series for $h(p)$ (measured in bits) expanded at $\frac{1}{2}$. To compute the channel capacity, it remains to estimate the equilibrium value of p.

3.5. Channel Capacity with Small Genomes

For small L, the equilibrium selected population may consist of identical genomes. The mean number of elements of a genome that are inverted by mutation is Lu; using a Poisson approximation of the number of elements of a genome that are inverted, the probability that no element of a genome is inverted — that is, the probability that a genome is unchanged by mutation — is approximately e^{-Lu}.

If more than half of the genomes are unchanged by mutation — that is, if $e^{-Lu} > \frac{1}{2}$ — then in equilibrium 50% truncation selection can maintain a selected collection that with high probability consists of identical genomes. Each selection rule $s_{\mathbf{z}}$ can maintain a selected collection of genomes that are with high probability all equal to $\mathbf{z}$. This implies that $H(X|S) \approx 0$, and therefore the channel capacity $I = L$ bits, provided that $L < \frac{\log 2}{u}$. Hence

$$I \leq \frac{\log 2}{u} \qquad (6)$$

for a regime in which the selected collection consists of identical copies of a single genome.

The result of equation (6) was derived by [ES79], in their investigation of the possible origins of life. Eigen et al. argued that an early replicator would have been inaccurate, with some relatively large mutation rate u, and that the mutation rate would set a limit on the possible length of the genetic sequence of such a proto-organism. If such an organism required, for its specification, an accurate genetic sequence longer than approximately $\frac{1}{u}$, it would suffer an "error catastrophe", as more errors accumulated in its sequence than could be feasibly eliminated by selection.

3.6. Channel Capacity with Large Genomes

In contrast to short genomes, we now consider the case of long genomes, so long that in equilibrium, the selected collection will have a value of p only slightly larger than $\frac{1}{2}$. Clearly, such genomes will be underdetermined by the selection rule, and they will have a low density of information — but on the other hand, the genomes are long, and so may contain a large amount of information at low density.

To determine the channel capacity, we estimate p at equilibrium as follows. At equilibrium, the fraction of 1s introduced by selection in each generation will equal the fraction of 1s removed by mutation. Mutation changes the fraction of 1s by $-2u(p-\frac{1}{2})$.

The increase in the fraction of 1s as a result of selection depends on the intra-collection variance of $v := \mathbb{E}_{C^T}\mathrm{Var}(f_1,\ldots,f_N)$. Assuming that the f_i are approximately normally distributed, the expected fraction of 1s in selected half of the population is $p+\sqrt{\frac{2}{\pi}}\sqrt{v}$ since the mean deviation of a normal variate from the mean is $\sqrt{\frac{2}{\pi}}\sigma$. As stated above, the exact method of selection is not important: let us suppose that the effect of selection is to increase the fraction of ones by an amount $\alpha\sqrt{v}$, where α is $\sqrt{\frac{2}{\pi}}$ for 50% truncation selection. The equilibrium equation is therefore

$$2u(p-\tfrac{1}{2}) = \alpha\sqrt{v} \qquad (7)$$

It remains to estimate the expected variance v of the fractions of 1s $\mathrm{Var}(f_1,\ldots,f_n)$ of genomes in the collection.

Let $\theta = (\theta_1,\ldots,\theta_L)$ be the marginal frequencies of 1s in C. That is, let $\theta_i = \frac{1}{n}\sum_{k=1}^{n} x_i^k$. The breeding system ensures that the values of each element of each genome are statistically independent, so:

$$v = \frac{1}{L^2}\mathbb{E}_C \sum_{i=1}^{L} \theta_i(1-\theta_i) \qquad (8)$$

By symmetry, for $1 \leq i \leq L$

$$\begin{aligned} \mathbb{E}_C\theta_i &= p - 2u(p-\tfrac{1}{2}) && (9)\\ &= p - 2up + u && (10)\end{aligned}$$

For sufficiently large collection size n, each θ_i will with high probability be close to its expected value $p-2u(p-\frac{1}{2})$. It follows that for large collections

$$\begin{aligned} v &= \frac{1}{L}\left(p(1-p) + 4(p-\tfrac{1}{2})^2(u-u^2)\right) && (11)\\ &\geq \frac{1}{L}p(1-p) && (12)\end{aligned}$$

We are most interested in the case where u is small, so that $v \approx \frac{1}{L}p(1-p)$, and we will proceed using this approximation.

The equilibrium equation is

$$2u(p-\tfrac{1}{2}) = \alpha\sqrt{\frac{p(1-p)}{L}} \tag{13}$$

which implies

$$p-\tfrac{1}{2} = \frac{1}{2\sqrt{\frac{4}{\alpha^2}Lu^2+1}} \tag{14}$$

For $L \gg \frac{\alpha^2}{16u^2}$ the channel capacity is

$$I = \frac{\alpha^2}{8\ln 2}\cdot\frac{1}{u^2} + O\left(\frac{1}{Lu^4}\right) \tag{15}$$

hence for large L and n,

$$I \propto \frac{1}{u^2} \tag{16}$$

Equation (16) is a remarkable result. The mutation rate u is typically small, so that the maximal channel capacity is achieved with large, ill-determined genomes, and is much larger than the channel capacity with small genomes of size $O(\frac{1}{u})$.

3.7. How large a population is needed?

An important question is how large the population size n needs to be to approach this channel capacity. A standard result of population genetics (given in [CK70]) is that the expected variance of the fraction of 1s with symmetric mutation and near-neutrality at individual loci, which is achieved with large L, is given by

$$v = \frac{1}{L}\cdot\frac{Nu}{4Nu+1} \tag{17}$$

This implies that $I \propto \frac{1}{u^2}$ for N greater than approximately $\frac{1}{4u}$.

4. A model of selective breeding for asexual organisms

To model asexual reproduction, we need alter only the breeding system b. To produce a new collection C' asexually from an existing weighted collection $(C, \mathbf{w})$, we sample each element $\mathbf{x}'^i$ of C' independently from $(C, \mathbf{w})$. In this model of asexual reproduction, genomes do not recombine, so that each element of C' is a copy of some element of C. Mutation is the only source of new genetic variety.

4.1. Channel capacity of asexual breeding: strong selection

Equation (7) remains valid for asexual breeding, but v is not easy to estimate for truncation selection as the distribution of $(f_1, \ldots, f_N)$ is no longer normal.

Instead of seeking to compute the channel capacity for truncation selection, we will bound the capacity of asexual breeding with a population of size n for *any* type of selection.

In asexual breeding, the "children" are cloned from the parent, and differ only in the mutations they accumulate. The expected fitness of a child, therefore, increases monotonically with the fitness of the parent. It follows that the fittest possible child population is obtained by breeding the entire child generation C' from fittest genome in C. That is, the selection rule — that we term *strong selection* — which gives the highest expected fitness of the child population is to set w_i to 1 where $\mathbf{x}_i$ is a maximally fit genome in C, and to set the weights of the rest of the genomes in C to zero.

In equilibrium, the expected fraction of 1s in the best child should be equal to the fraction of 1s in the parent. For large L, the distribution of the fraction of 1s in the children will be approximately normal, so that

$$v = \frac{u(1-u)}{L} \tag{18}$$

$$\approx \frac{u}{L} \quad \text{when } u \text{ is small} \tag{19}$$

The mean fraction of 1s in the children will be $p-2u(p-\frac{1}{2})$. The maximum of N samples from a normal distribution is $O(\sigma\sqrt{\log N})$ above the mean. It follows that

$$2u(p-\tfrac{1}{2}) = O\left(\sqrt{\frac{u\log N}{L}}\right) \tag{20}$$

so that

$$I = O\left(\frac{\log N}{u}\right) \tag{21}$$

Hence the channel capacity for asexual breeding is much lower than for sexual breeding for low mutation rates. The difference is large: to achieve a channel capacity comparable to sexual breeding with a population size of $\frac{1}{4u}$, the size of an asexual population needs to be $\exp(O(\frac{1}{u}))$, which is infeasible for small u.

5. Simulations

Figure 1 shows the channel capacity for sexual and asexual breeding, for genomes of various lengths, and with various mutation rates. The value of collection size N was 500 for all experiments. Three different mutation rates — 0.1,

0.01, and 0.001 — were used. The genome sizes ranged from 10 to 1,000,000. The lines labelled "S" are for sexual breeding, and those labelled "A" show the results for asexual breeding. As expected, for low mutation rates there is a large difference in channel capacity for sexual and asexual breeding. In this experiment, sexual rather than hypersexual breeding was used: each child was created from two "parent" genomes, with each element of the child being independently selected with equal probability from either parent. This breeding system can be implemented very efficiently, so that large simulations may readily be done. The channel capacity of this form of sexual breeding is somewhat less than that of hypersexual breeding.

6. Discussion

Although the analysis has been quite abstract, the informational advantage of distributed encodings for sexual breeding is in principle so large that it seems plausible that such encodings occur in nature. Well known aspects of the genetics of eukaryotic organisms may make sense from this point of view. The (sexual) eukaryotes usually have large genomes, consisting mostly of "junk", while the genomes of (asexual) prokaryotes are generally smaller with a higher proportion of genes. Although the classical triplet genetic code is the same in both kingdoms, the genome encodes much information other than protein sequences. The encoding of regulatory and developmental information would be informationally efficient if it were diffuse. "Junk" DNA is produced by many processes: once it exists it will accumulate genetic variety as mutations occur, and this genetic variety provides potential channel capacity. Where channel capacity exists, it is likely to be used.

Eigen et al in [ES79] argue that a primitive genome must be limited in length to $O(\frac{1}{u})$ because otherwise errors would accumulate that would prevent the genome from replicating properly, so that there would be an "error catastrophe" for genomes of excessive length. We take a different view. There are no doubt parts of the genome that must be accurate for an organism to be viable, and for basic components of cells to function properly. The length of these parts of the genome is indeed limited, by Eigen et al's argument, to $O(\frac{1}{u})$. However, for complex organisms, the mechanisms of regulation of gene expression, and the processes of development, could conceivably be influenced by very many loci, and might be robust enough to ce able to interpret a code of low information density. The adaptive structural complexity of an organism is necessarily limited by the amount of genetic information available to the organism's developmental processes, and the largest amount of information may be supplied by diffuse, low-density codes.

For genetic algorithms, these results imply that more interesting behaviour and better adaptation may occur if the algorithms use diffuse encodings on long genomes, long enough so that "convergence" of the population to a single genome never occurs because of mutation and drift. We do not yet know how to devise suitable diffuse encodings, or under what circumstances such encodings may spontaneously arise.

The next unanswered question is whether highly distributed genetic codes that enable high channel capacity tend to evolve spontaneously. The channel capacity is the maximum possible maintainable information in the genome under an encoding that enables the most favourable type of fitness function. We do not know whether such stable highly distributed and favourable codes can evolve and themselves be stable under natural selection.

Acknowledgments

Anna Fukshansky read earlier versions of this manuscript. Thanks to Peter Dayan, Vincent Jansen, and John Shawe-Taylor for useful discussions. Quaid Morris made a most helpful comment.

References

[1] C. Adami. What is complexity? *BioEssays*, 24:1085–1094, 2002.

[2] S. Baluja. Population-based incremental learning: A method for integrating genetic search based function optimization and competitive learning,. Technical Report CMU-CS-94-163, Pittsburgh, PA, 1994.

[3] E. B. Baum, D. Boneh, and C. Garrett. Where genetic algorithms excel. Technical report, NEC Research Institute, 4 Independece Way, Princeton, NJ 08540, 1995.

[4] T. Cover and J. Thomas. *Elements of Information Theory*. Wiley-Interscience, New York, 1991.

[5] J. Crow and M. Kimura. *An Introduction to Population Genetics Theory*. New York: Harper and Row, 1970.

[6] J. Crow and M. Kimura. Efficiency of truncation selection. *Proc. Natl. Acad. Sci. USA*, 76:396–9, 1979.

[7] P. Dayan and G. E. Hinton. Using EM for Reinforcement Learning. *Neural Computation*, 9:271–278, 1997.

[8] M. Eigen and P. Schuster. *The Hypercycle: a principle of Natural Self-Organization*. Springer-Verlag, 1979.

[9] W. J. Ewens. *Mathematical Population Genetics*. Springer-Verlag, 1979.

[10] G. R. Harik, F. G. Lobo, and D. E. Goldberg. The compact genetic algorithm. *IEEE-EC*, 3(4):287, November 1999.

[11] D. J. MacKay. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.

[12] M. Pelikan, D. E. Goldberg, and E. Cantú-Paz. BOA: The Bayesian optimization algorithm. In W. Banzhaf, J. Daida, A. E. Eiben, M. H. Garzon, V. Honavar, M. Jakiela, and R. E. Smith, editors, *Proceedings of the Genetic and Evolutionary Computation Conference GECCO-99*, volume I, pages 525–532, Orlando, FL, 13-17 1999. Morgan Kaufmann Publishers, San Fransisco, CA.

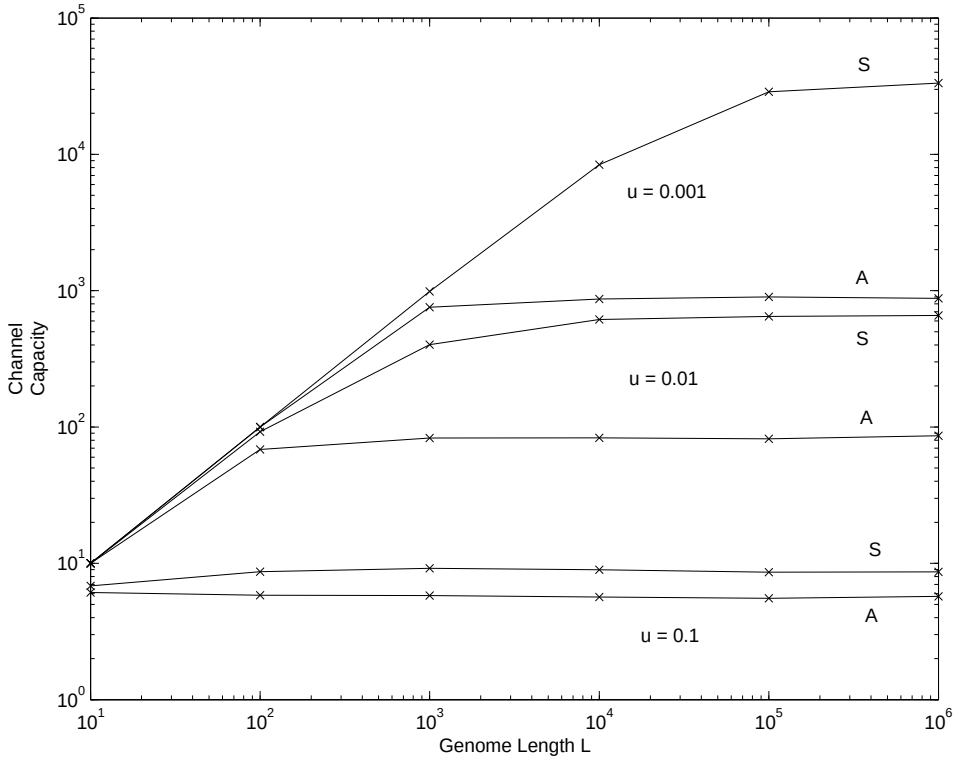

Figure 1. The channel capacity as a function of genome size, for sexual and asexual breeding, for three mutation rates.