The Emergence of Overlapping Scale-free Genetic Architecture in Digital Organisms

P. Gerlee*,**
University of Dundee

T. Lundh†
Chalmers University of Technology and Göteborg University

Abstract We have studied the evolution of genetic architecture in digital organisms and found that the gene overlap follows a scale-free distribution, which is commonly found in metabolic networks of many organisms. Our results show that the slope of the scale-free distribution depends on the mutation rate and that the gene development is driven by expansion of already existing genes, which is in direct correspondence to the preferential growth algorithm that gives rise to scale-free networks. To further validate our results we have constructed a simple model of gene development, which recapitulates the results from the evolutionary process and shows that the mutation rate affects the tendency of genes to cluster. In addition we could relate the slope of the scale-free distribution to the genetic complexity of the organisms and show that a high mutation rate gives rise to a more complex genetic architecture.

Keywords
Genetic architecture, digital evolution, Avida, scale-free networks

1 Introduction

Interaction networks are a common feature in biological systems ranging from large-scale structures like ecological networks [13] down to intracellular structures like protein interaction networks [9] and metabolic networks [10]. These structures tend to be highly complex and have received much attention in the literature, especially since the discovery that most of them tend to exhibit a scale-free architecture [4]. In order to examine the evolutionary origin of these structures we have studied the emergence and structure of genetic architecture in the ALife platform Avida [14].

Avida is an Alife platform designed to study the evolution of digital organisms. It was inspired by Tierra [16], but instead of holding all the organisms in one central memory, the organisms in Avida reside on a square lattice, each lattice point containing a CPU that executes the genome of the organism. The genome is written in an assemblerlike code, where each line of code contains a command that modifies the state of the CPU in some predefined manner. The language consists of instructions that allow for self-replication, but also other commands that can be used to perform simple computational tasks. These tasks are rewarded by giving the organism more CPU time, and thus a faster replication that results in a selective advantage. Each experiment is started with an ancestral organism, which only has the capability to replicate. But the self-replication takes place

* Corresponding author
** Division of Mathematics, University of Dundee, 23 Perth Road, DD1 4HN, Scotland. E-mail: gerlee@maths.dundee.ac.uk
† Mathematical Sciences, Chalmers University of Technology and Göteborg University, SE-41296 Göteborg, Sweden. E-mail: torbjorn.lundh@chalmers.se

under mutations, which means that genotypes with computational capabilities eventually emerge, and these will compete for the limited space on the lattice.

The genetic architecture of evolved Avida creatures turns out to be highly complex and hard to interpret [8, 12]. This is mainly due to their highly interconnected structure, where subroutines or genes that perform different tasks in the genetic code share instructions and overlap with each other. The genetic code of Avida creatures consists of assemblerlike code, where each line of code can be thought of as an amino acid. The language allows for nonlinear execution of the code; this, plus the fact that genes tend to overlap, makes the code difficult to analyze. One way to overcome this is to visualize the genetic structure by depicting the execution of a creature in a circular execution diagram. This shows the nonlinear execution, but more importantly displays where in the genome certain genes are located. A typical example of these circular diagrams can be seen in Figure 1, which exhibits a complex genetic architecture, with a nonlinear execution sequence and overlapping genes.

A natural question to ask is why overlapping genes emerge in these digital organisms. There are two main reasons: Firstly, it is much easier to develop a new gene by modifying an already existing gene; hence genes tend to cluster [12]. For such a gene development, see for example Figure 2, which shows the development of a specific gene within a lineage of creatures. The second immediate

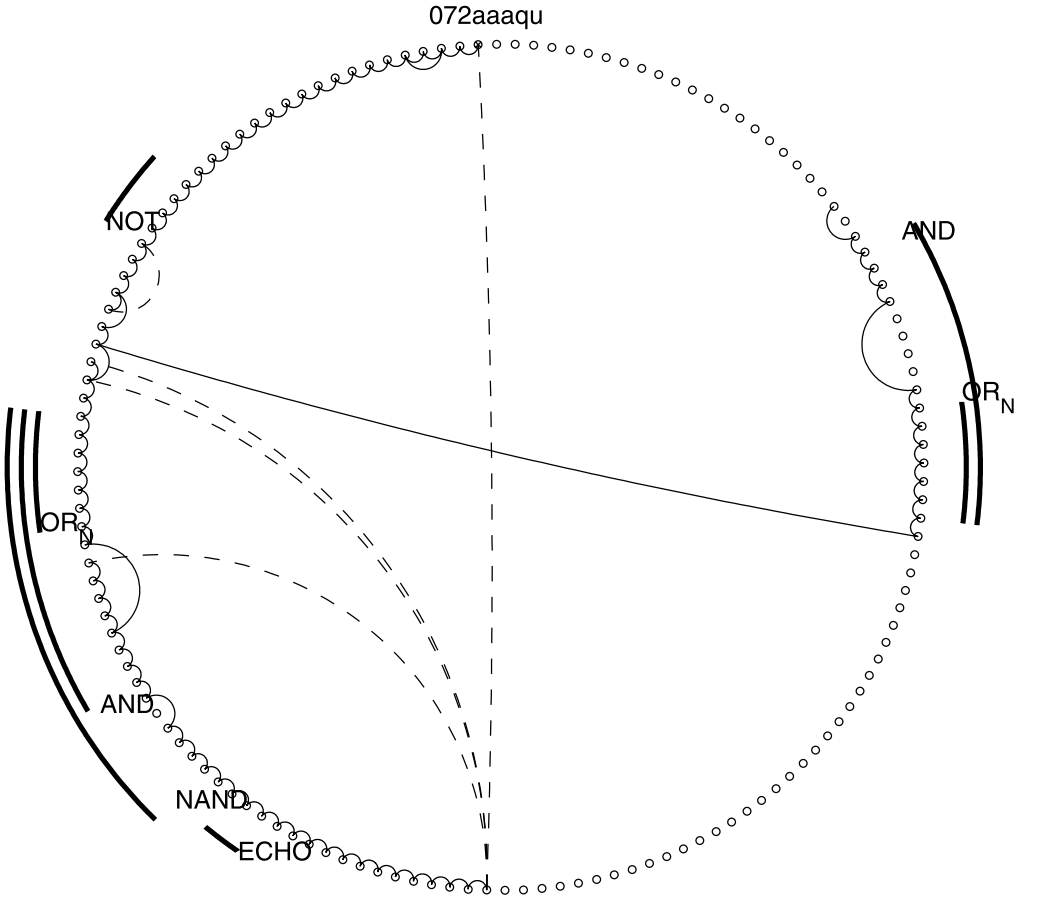

Figure 1. A useful way to depict an Avida creature is by using a circle diagram and marking the genes by solid circle segments outside the main circle, where each dot stands for a command, and where the jumps are depicted as circular arcs (hyperbolic geodesics) inside the circle. Solid arcs represent forward jumps, and dashed arcs represent backward jumps. The execution is started at 12 o'clock and proceeds in a counterclockwise fashion. Note the nonlinear execution order and that the "genes" are overlapped. This type of diagram was introduced in [7]; see also [1] for similar illustrations.

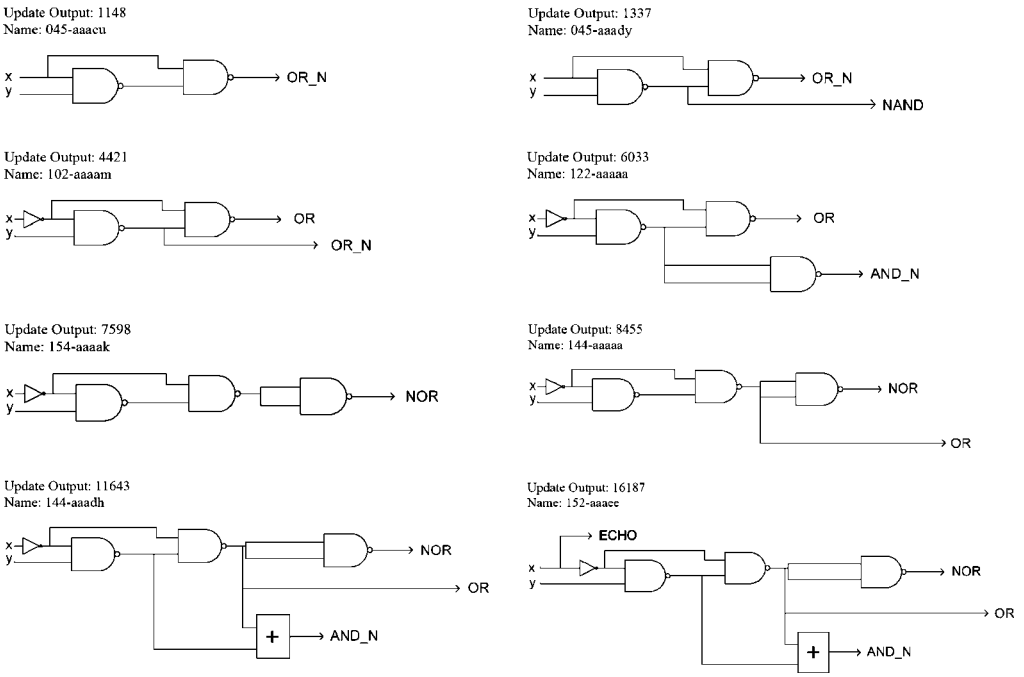

Figure 2. An illustration of how genes develop over time. We have chosen to illustrate the Boolean genes by standard electric circuits. The development of these circuits shows both the tendency to build a new gene using an old one, and also that genes are not always developing into strictly increasing higher complexity; see, for example, the jump from 6,033 to 7,598, where a less complex function is evolved. The genes are taken from a single lineage, and the time points show when the new gene appeared in the lineage.

reason is that since the Avida environment is highly competitive, it is of great value for a creature to be able to reduce the deleterious effects of random mutations [8, 6]. This is achieved by letting some instructions take part in several genes, which implies that the total number of instructions needed to encode the genes is reduced. This will also reduce the time needed for the creature to generate an offspring (i.e., the gestation time). Due to the high competition in Avida, any fitness loss will likely lead to extinction of the mutant, which implies that the simultaneous loss of several genes and the loss of only one are equally lethal for a mutant (see [7] for more details). Hence it is a good strategy to stack the genes more or less on top of each other, to make the genome less vulnerable to random point mutations.

The interconnected genetic structure of Avida creatures makes them comparable to metabolic networks of many organisms, which are known to be highly complex and interconnected and to often exhibit a so-called scale-free structure [10]. Viewing the genetic architecture as a network structure makes it possible to illustrate its connectivity properties and to compare it with real metabolic networks. By doing this we can get another perspective on the genetic architecture in the Avida codes and point out analogies with other evolutionary driven systems.

For metabolic networks the scale-free structure means usually that the probability $P(k)$ of finding a metabolic substrate that takes part in k separate reactions scales as $k^{-\gamma}$. If one represents the metabolic network as a graph where the nodes correspond to substrates and the edges to chemical reactions, then this means that the probability of finding a node with connectivity k scales as $k^{-\gamma}$. This relation has also been shown to be true for other complex networks like the WWW, social networks, and power grids [4]. It should be noted that there exists a degree of controversy regarding the scale-free properties of these networks; see for example [18, 17].

Scale-free networks can be generated by a growth algorithm that has two basic ingredients: growth and preferential attachment [4]. The former means that number of nodes in the network

does not remain constant, but instead new nodes are added over time; the latter, that when new links are being added to the network, they are more likely to be connected to nodes that already have a large number of links.

The connectivity of scale-free networks tends to be dominated by a few highly connected hubs. This structure leads to a high robustness against random errors, meaning that the removal of a random node in the network is unlikely to affect the overall performance of the network [3]. For further reading on scale-free networks consider [5] and [2].

It has been suggested [15] that scale-free networks are likely to appear in an evolutionary setting, as these networks are tolerant to mutations, but still have the capability to evolve through additions of new nodes. Motivated by this, we propose to study the emergence of scale-free network architecture by studying the genetic structure of digital organisms in the Alife platform Avida. This is of course a simplified system compared to the evolution of real genetic or metabolic networks, but the simplicity also allows for a greater understanding of the dynamics. First of all, there is no representation of gene products in Avida; instead the genes are used for computations, which are meant to represent beneficial chemical reactions. This implies that the interactions can only occur on the level of the genome, which manifests itself as gene overlap or different genes sharing the same instructions. This means that the Avidian creatures do not have any metabolic networks, but there still exist interactions between different genes that constitute their digital metabolism, and these interactions will be the subject of our study.

2 Genetic Architecture of Digital Organisms

The genetic architecture of digital organisms has been investigated in a large number of publications. Much of the work has focused on the effects of epistasis and the robustness of the genetic architecture. In a study of simple and complex digital organisms (organisms that could only replicate and organisms with computational abilities), Lenski et al. [11] showed that simple organisms in general are more fragile than complex ones. They also showed a difference in how mutations affect the fitness of the organisms, in that complex organisms display a higher degree of epistatic effect of mutations, that is, the genes of the complex organisms have a higher degree of interdependence than their simple counterparts. In another study on gene epistasis by Wilke and Adami [19], they investigated the average effect of mutations and the strength of directional epistasis. They show that there exists a tradeoff between the average fitness loss of a single mutation and epistatic effects, and that the two cannot be optimized separately in the evolutionary process. The robustness of digital organisms has been investigated by Edlund and Adami [6]. By evolving digital organisms in a environment with a low mutation rate up to a certain complexity and then inserting them into a high-mutation environment, they could show that average deleterious effect of mutations is decreased. This effect is achieved by breaking up links between genes and thus reducing the epistatic effects of mutations.

The above studies have all investigated the average effect of mutations without considering the detailed structure of the genetic architecture. This was considered by Lenski et al. [12] in a study on the evolution of complex features in digital organisms. Their results show that complex genes often emerge from simpler existing genes and that evolution of complex traits indeed is possible as long as the building blocks of the complex function are also rewarded. Further, they demonstrated that the organisms with complex traits exhibited a high degree of interaction between different genes. This was done by analyzing functional genomic arrays (FGAs) or knockout matrices, a technique that has been widely used [12, 8, 6]. The FGA gives detailed information about the genetic architecture of Avida creatures, and we intend to use this technique to investigate the structure of the genetic architecture.

2.1 Functional Genomic Array

The FGA is an $N \times M$ binary matrix, where N is the length of the genome and M is the number of possible tasks the creature might perform, including replication. The entry at position (i, j) is 1

if instruction number i is involved in the calculation of task number j and 0 otherwise. An example of a typical FGA can be found in Figure 3. The FGA not only serves as a visual aid, but can also be used to gain quantitative knowledge about the genetic structure. One example is the gene correlation [8], which is calculated from the FGA. This is a measure of how clustered the genes are and has been shown to depend on evolutionary settings like the mutation probability.

If we sum the FGA along the rows, we obtain a vector T, which we name the *gene distribution*. This vector is of length N, and T_i is the number of tasks (genes) that depend on instruction i. From this it is obvious that $0 \leq T_i \leq M$, where $T_i = 0$ means that instruction i is not involved in any genes, and $T_i = M$ means that instruction i is used by all genes in the digital organisms. From another perspective we can view T_i as the number of genes that are connected to instruction i in the genome. An instruction that is connected to a large number of genes can be viewed as a hub in the genetic structure, and a mutation to this instruction would most probably have a disastrous effect on the fitness of the creature. On the other hand, a mutation to an instruction with no connection would most likely be neutral for the creature. With this in mind, we can observe that although the instructions do not form a network, they still share features with the metabolic networks found in real organisms. It is therefore interesting to measure how the vector T behaves on average, to see if it shares the scale-free structure of the metabolic networks. What we specifically ask is if we select a random instruction i, what is the probability $P(T_i = k)$ that it will be used in k different genes. To measure the probability distribution $P(k)$, we evolved populations of digital organisms to a high complexity (i.e., with a large number of genes) and for each individual created an FGA and calculated the vector T. In order to also measure the effect of environmental settings, the above was performed for two different values of the mutation probability, a parameter known to affect the genetic architecture of digital organisms [8].

3 Materials and Methods

In these experiments we used the default settings in Avida,[1] and the only parameter that was changed was the (divide) mutation probability.[2] We considered two different values of the mutation probability: low ($p = 0.001$) and high ($p = 0.005$). In this type of experiment it would be optimal to compare creatures from the same phenotype (i.e., that can perform the same tasks), but as evolution in Avida proceeds at a different pace in each run, one has no guarantee that a certain phenotype has evolved after a fixed number of updates. One way to accomplish this would be to reward only those functions that we want the phenotype to perform and use a large number of fixed updates, but this approach also has its drawbacks. If the required phenotype appears early in evolution, then the code is optimized during the rest of the run, as no new functions are rewarded. If a creature from the above run is compared with a creature from a run where the required phenotype appeared just before the maximal number of updates, their genetic architecture will certainly be found to differ, as one creature has optimized its coding while the other has not. Instead we decided to compare creatures with approximately the same complexity. This was done by extracting a creature from the dominant genotype after 40,000 updates in each run; if it had reached a certain degree of complexity (it managed at least five tasks including replication), it was kept for analysis. For each value of the mutation probability, 50 creatures were collected and analyzed. That is, for each creature the FGA was calculated using a replica of the CPU in Avida, and from the FGA the vector T was calculated. The probability distributions of the entries in T were then averaged over all creatures from the same setting.

4 Results

The creatures that had evolved at the low mutation rate ($p = 0.001$) on average managed 6.3 ± 1.3 tasks. The probability distribution $P(k)$ can be seen in Figure 4, where it is shown in a log-log plot.

1 Avida version 1.3 was used for all experiments reported in this article.

2 Supplementary material can be found at http://www.math.chalmers.se/~torbjrn/scale-free.

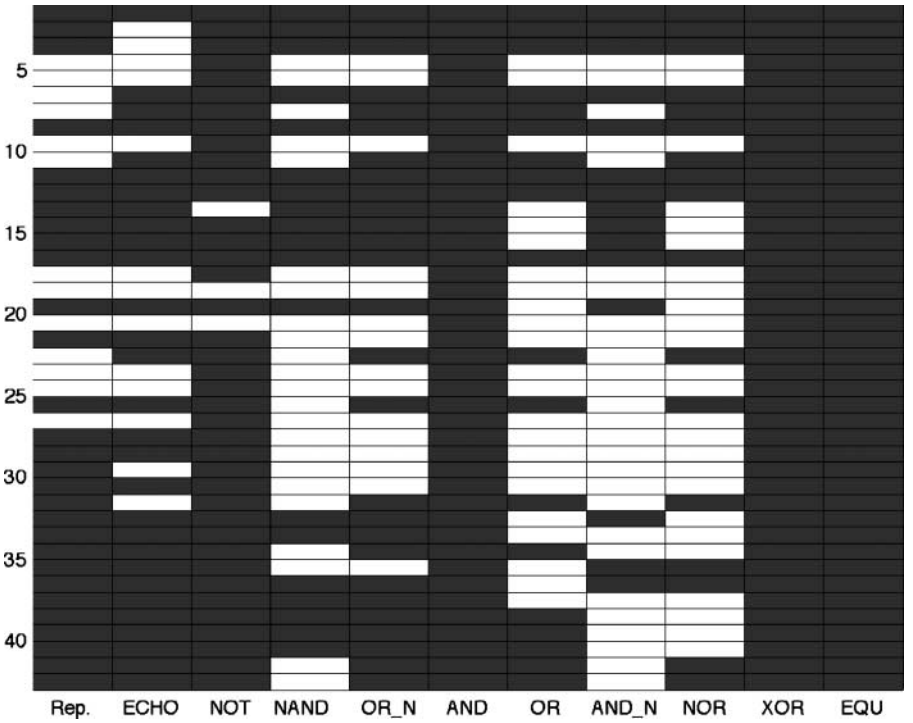

Figure 3. Typical example of a FGA. Each row represents an instruction, and each column a task (gene). An instruction is colored white in a task column if the task depends on that instruction, and gray if not.

The distribution shows a clear tendency toward scale-free behavior, with a negative slope of $\gamma = 2.5$ (correlation coefficient $\rho = -0.99$). This means that the distribution of connections for an instruction scales as $P(k) \sim k^{-\gamma}$.

For the high mutation rate ($p = 0.005$) the mean number of tasks was 7.1 ± 1.3. The probability distribution $P(k)$ can be seen in Figure 5, where it is shown in a log-log plot. Again the distribution shows a clear tendency toward scale-free behavior, this time with a negative slope $\gamma = 2.2$ (correlation coefficient $\rho = -0.99$). The data was also fitted with exponential curves ($P(k) \sim e^{-\beta k}$), but this gave fits with correlation coefficient $\rho = -0.95$ in both cases, which suggests that a power law describes the data better.

5 Dynamic Model of Gene Clustering

In random networks the scale-free behavior can be obtained from a growth algorithm that only contains two basic components: growth and preferential attachment. Can the scale-free behavior of the gene distribution observed in the previous experiments be explained from a similarly simple growth algorithm?

In order to answer this question we have constructed a simple dynamic (Monte Carlo) model that mimics the complex process of gene clustering in Avida creatures. The basic idea behind the model is that new genes are more likely to appear in connection with existing genes [12]. This preferential attachment of new genes is controlled by a parameter r, the gene attraction. To a first approximation r can be interpreted as the probability that a mutation transforms an existing gene into one that is capable of performing multiple tasks. This depends on the instruction set that is being used, and also on the type of functions that are rewarded. If functions that are *computationally* close are rewarded, then the gene attraction will be high. But part of the gene attraction can also be

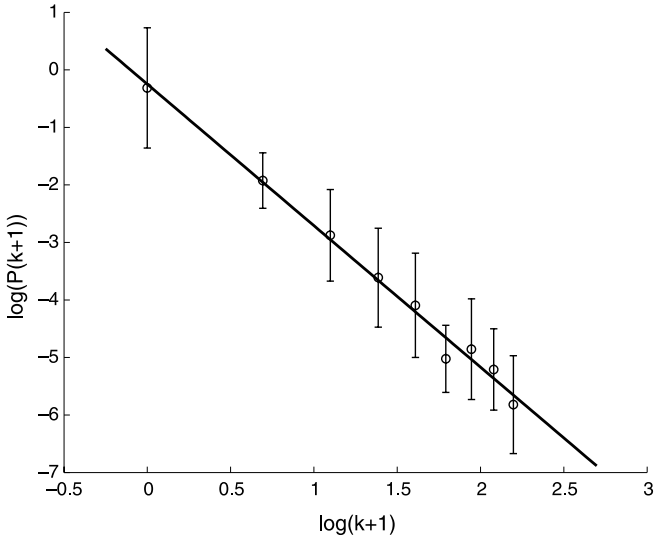

Figure 4. The gene distribution $P(k)$ for the low mutation rate $p = 0.001$ plotted in a log-log diagram. Note that graph is shifted to the left [i.e., $\log(k + 1)$ is plotted] to allow for a logarithmic plot. The gene distribution shows a scale-free tendency with $\gamma = 2.5$. The error bars correspond to 1 *SD*.

given an evolutionary interpretation. From our experiments with Avida we have observed that creatures that have evolved under a high mutation rate on average have a higher clustering of genes (0.8 ± 0.5 genes per instruction for the low mutation rate, compared to 1.1 ± 0.5 for the high mutation rate). As the instruction set and rewarded functions are the same for both the high and the low mutation setting, this implies that the mutation rate influences the probability of genes to cluster. In fact the gene attraction is a priori (i.e., explicitly) independent of the mutation rate, but the environment in which the population evolves selects for genotypes with a high gene clustering [8],

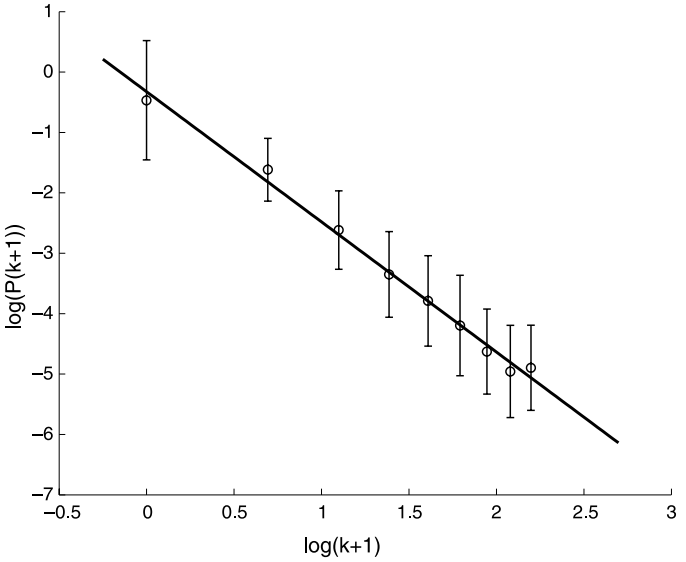

Figure 5. The gene distribution $P(k)$ plotted in a log-log diagram. The mutation rate is $p = 0.005$. Note that the graph is shifted to the left to allow for a logarithmic plot. The gene distribution shows a scale-free tendency with slope $\gamma = 2.2$. As above, the error bars correspond to 1 *SD*.

which can be interpreted as a dependence of the mutation rate on the gene attraction. A high mutation rate would then correspond to a higher value of the gene attraction.

5.1 Model Description

Each Monte Carlo simulation is started with a genome of length L, which does not contain any genes. For simplicity we will consider a fixed genome length, although this is not the case in the evolution of Avida creatures (under the default settings). We will for simplicity consider genes of a fixed length Δ. The genes will further be assumed to be constructed of a set of simply connected instructions, which is not strictly true in Avida. At each time step of the simulation a new gene is added to the genome. The position of the new gene is determined by the current gene distribution, and the probability $Q(i)$ that the center of the new gene appears at instruction i is given by

$$Q(i) = \frac{rk_i + 1}{\sum_{j=1}^{L}(rk_j + 1)} \tag{1}$$

where k_i is the number of genes that already are connected to that instruction. After the site of the new gene has been chosen all instructions that the gene covers have their k_i's incremented by one. This model contains the two components that shape the growth of scale-free networks namely (i) growth (new genes are constantly added to the genome and thus creating new connections), and (ii) preferential attachment (new genes are more likely to appear where the gene distribution already is high).

Figure 6 shows the result of two typical simulations with differing values of the gene attraction r. The left plot shows the gene distribution for a low gene attraction ($r = 1$), while the right is for a higher attraction ($r = 10$). The other parameters were fixed at $L = 100$, $\Delta = 15$, and the number of genes $N = 8$. From this plot it can be seen that a higher gene attraction gives rise to highly overlapping genes, while the low r case gives a more uniform gene distribution.

The comparison between the dynamic model and the results from Avida creatures was performed by simulating a large number of gene distributions (1,000 different realizations of the process), calculating the probability distribution $P(k)$, and comparing it to the one obtained from Avida. All parameters for the model were estimated from the Avida organisms, except the gene attraction, which was used to fit the model to the data obtained from the Avida experiment. Figure 7 shows the gene distribution $P(k)$, with parameters estimated from the high mutation rate ($p = 0.005$). The genome length was set to the average value among those creatures, $L = 130$; the gene length, to the

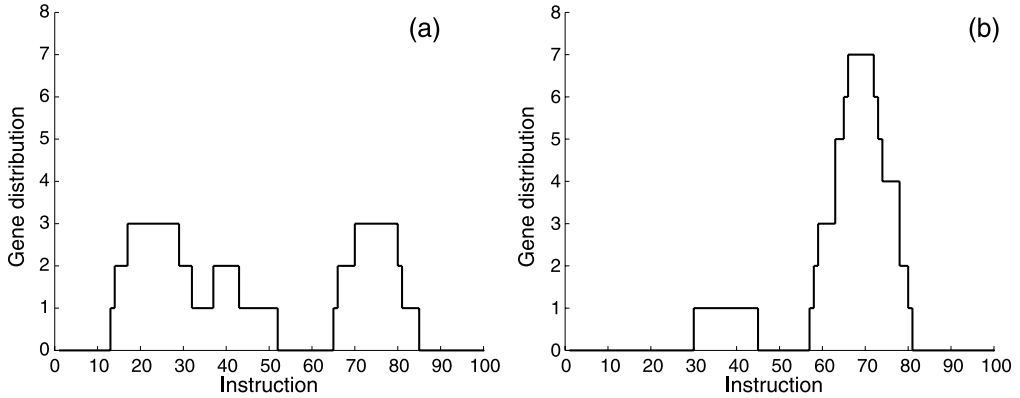

Figure 6. The gene distribution at the end of two simulations with different values of r, all other parameters being fixed at L = 100, Δ = 15, and the number of genes N = 8: (a) r = 1, and (b) r = 10. The higher gene attraction gives rise to clustered genes; the lower, to a more uniform gene distribution.

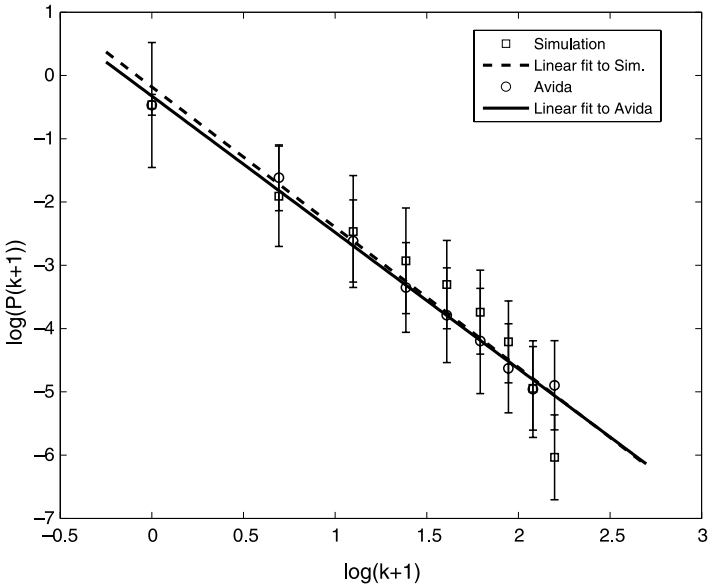

Figure 7. The gene distribution from the simple Monte Carlo model of gene attachment plotted together with the actual gene distribution from creatures that have evolved under the high mutation rate. The plot shows that the model recovers the scale-free behavior and shows good agreement between the model and the data when the gene attraction $r = 6.5$. The error bars correspond to 1 *SD*.

average $\Delta = 15$; and the number of genes, to $N = 7$, the average in the population. The gene attraction r was fitted to the data using a least squares method and was found to be $r = 6.5$. The plot shows that the model also gives rise to scale-free behavior and that it matches the results from Avida quite well.

In order to further investigate the behavior of the model, we measured how the exponent γ in the scale-free distribution depends on the gene attraction r. The results can be found in Figure 8, and

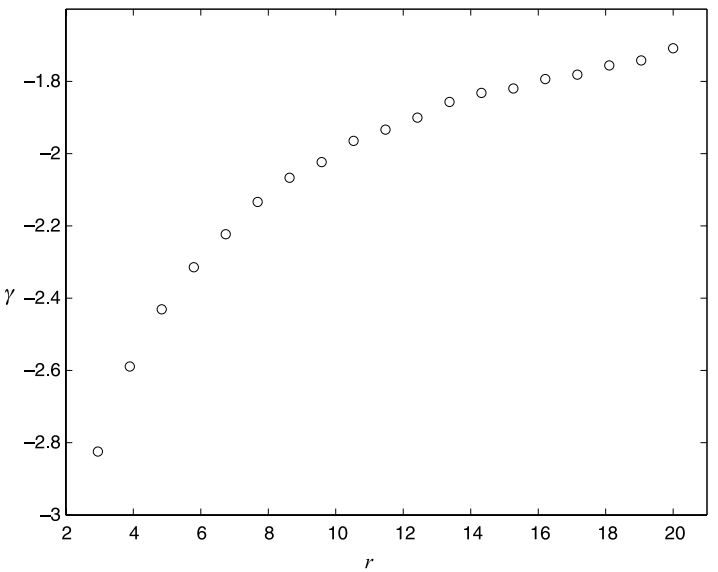

Figure 8. Using the dynamic model of gene clustering, the fitted exponent γ in the scale-free distribution is plotted as a function of the gene attraction r.

show that γ is an increasing function of the gene attraction. This result is intuitive in that a larger r gives rise to more clustered genes, and thus to a slower decay in $P(k) \sim k^{-\gamma}$.

6 Discussion

The results from our experiments with Avida show that for both low and high mutation rates the genetic architecture of Avida creatures tends to a scale-free structure. Although the genome does not form a network in a strict sense, the overlap of genes can be interpreted as interactions between different genes, which in turn can be seen as an interaction network. It is known that scale-free networks can be generated using a simple growth algorithm that relies on growth of the network and preferential attachment. The emergence of a scale-free architecture in the genetic architecture therefore suggests that it is preferential attachment or gene clustering that is the dominating force behind gene development. It should be noted that this preferential attachment is dependent on the mutation probability, which implies that although the a priori probability for an existing gene to mutate into one that performs more functions is independent of the environment, the mutation rate affects the gene clustering because it causes selection to favor more robust coding. This was also observed in our experiments where creatures from the low mutation probability had a distribution that tailed off faster ($\gamma = 2.5$) than that of the creatures from the high-mutation environment ($\gamma = 2.2$), implying that the low-mutation creatures had less clustered genes.

The scale-free structure was recovered from a simple Monte Carlo model of gene clustering, which suggests that the genetic architecture indeed is driven by preferential attachment to existing genes. For creatures that had evolved under the mutation probability $p = .001$ we observed a slope of $\gamma = 2.5$ in the gene distribution, which in the dynamic model corresponds to a gene attraction of $r \approx 5$. On the other hand, for creatures that had evolved under $p = .005$ we measured $\gamma = 2.2$, which corresponds to $r = 6.5$. This confirms our hypothesis that the mutation rate affects the gene attraction, and that a high mutation rate selects for creatures with a higher degree of overlap among the genes.

The exponent γ can also be given another meaning by considering the genetic architectures that give rise to low and high values of γ. For example, consider an Avidian creature where all genes are coded sequentially in the genome (without overlap). This architecture will give rise to a connectivity distribution for which $P(k) > 0$ only for $k = 0, 1$ (redundant instructions and instructions involved in one gene) and $P(k) = 0$ for all $k > 1$. This distribution can be viewed as the limiting case of a scale-free distribution where γ is tending to infinity. On the other hand consider a creature which has a fully connected architecture. This creature will have a uniform distribution of connectivity over all instructions, and thus $P(k) = c$ for some $c > 0$. This can be interpreted as the limiting case of a scale-free distribution where γ is tending to zero. From this point of view we can interpret γ as a measure of genetic complexity, as the former example has a very simple genetic architecture, while the latter would be considered more complex. This would then allow us to determine the complexity of a genetic architecture by studying the exponent in the scale-free degree distribution.

In conclusion, we have shown that a scale-free genetic architecture spontaneously emerges in Avida creatures and that the exponent γ in the distribution depends on the mutation rate. This suggests that the scale-free architecture is a natural outcome of an evolutionary process and that the genetic information self-organizes into a state that is robust with respect to mutations, but that still can evolve toward higher complexity. In the future we aim to test this hypothesis by analyzing the genetic architecture in other Alife platforms, as for example Tierra [16]. Other possible extensions of the results presented here would be to investigate how the genetic architecture evolves over time, and also to make systematic measurements of how the slope γ of the scale-free distribution depends on the mutation rate.

Acknowledgments

The authors would like to thank the referees for many valuable comments and suggestions. The work of P. Gerlee was funded by the National Cancer Institute, Grant 1 P50 CA113007-01.

References

1. Adami, C. (2006). Digital genetics: Unravelling the genetic basis of evolution. *Nature Review Genetics*, *7*, 109–118.
2. Albert, R. (2005). Scale-free networks in cell biology. *Journal of Cell Science*, *118*, 4947–4957.
3. Albert, R., Jeong, H., & Barabasi, A.-L. (2000). Error and attack tolerance of complex networks. *Nature*, *406*, 378–382.
4. Barabasi, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*, 509–512.
5. Barabasi, A.-L., & Oltvai, Z. (2004). Network biology: Understanding the cell's functional organization. *Nature Review Genetics*, *5*, 101–113.
6. Edlund, J., & Adami, C. (2004). Evolution of robustness in digital organisms. *Artificial Life*, *10*, 167–179.
7. Gerlee, P. (2005). *The genetic coding style of digital organisms*. Master's thesis, Chalmers University of Technology.
8. Gerlee, P., & Lundh, T. (2005). The genetic coding style of digital organisms. In M. Capcarrere, A. Freitas, P. Bentley, C. Johnson, & J. Timmis (Eds.), *Advances in artificial life* (pp. 854–863). Berlin: Springer.
9. Jeong, H., Mason, S., Barabsi, A.-L., & Oltvai, Z. (2001). Lethality and centrality in protein networks. *Nature*, *411*, 41–42.
10. Jeong, H., Tombor, B., Albert, R., Oltvai, Z., & Barabasi, A.-L. (2000). The large-scale organization of metabolic networks. *Nature*, *407*, 651–654.
11. Lenski, R., Ofria, C., Collier, T., & Adami, C. (1999). Genome complexity, robustness and genetic interactions in digital organisms. *Nature*, *400*, 661–664.
12. Lenski, R., Ofria, C., Pennock, R., & Adami, C. (2003). The evolutionary origin of complex features. *Nature*, *423*, 139–144.
13. Montoya, J., Pimm, S., & Sole, R. (2006). Ecological networks and their fragility. *Nature*, *442*, 259–264.
14. Ofria, C., & Wilke, C. (2004). Avida: A software platform for research in computational evolutionary biology. *Artificial Life*, *10*, 191–229.
15. Oikonomou, P., & Cluzel, P. (2006). Effects of topology on network evolution. *Nature Physics*, *2*, 532–536.
16. Ray, T. (1992). An approach to the synthesis of life. In C. Langton, C. Taylor, J. Farmer, & S. Rasmussen (Eds.), *Artificial life II*, pp. 371–408. Reading, MA: Addison-Wesley.
17. Stumpf, M. P., Wiuf, C., & May, R. M. (2005). Subnets of scale-free networks are not scale-free: Sampling properties of networks. *Proceedings of the National Academy of Sciences of the USA*, *102*, 4221–4224.
18. Tanaka, R., Yi, T., & Doyle, J. (2005). Some protein interaction data do not exhibit power law statistics. *FEBS Letters*, *579*, 514.
19. Wilke, C., & Adami, C. (2001). Interaction between directional epistasis and average mutational effects. *Proceedings of the Royal Society B: Biological Sciences*, *268*, 1469.