Memorial
University of Newfoundland

Technical Report #2008-04

Evolvability and Acceleration in Evolutionary Computation

by

Ting Hu & Wolfgang Banzhaf

Email: tingh@cs.mun.ca; banzhaf@cs.mun.ca

Department of Computer Science
Memorial University of Newfoundland
St. John's, NL, Canada A1B 3X5

October 2008

Evolvability and Acceleration in Evolutionary Computation

Ting Hu Wolfgang Banzhaf

Abstract

Biological and artificial evolutionary systems can possess varying degrees of evolvability and different rates of evolution. Such quantities can be affected by various factors. Here, we review some evolutionary mechanisms and discuss new developments in biology that may improve evolvability or accelerate evolution in artificial systems. Biological notions are discussed to the degree they correspond to notions in evolutionary computation. We hope the findings put forward here can be used to design computational models of evolution that exhibit significant gains in evolvability and evolutionary speed.

1 Introduction

The field of *Evolutionary Computation* (EC) has seen substantial progress since it was founded in the Sixties and Seventies of the 20th century [68, 51, 128, 138, 43, 42, 84, 16], inspired by the evolution process in nature.

In EC, candidate solutions to optimization or learning problems are represented by structures similar to gene sequences and their phenotypic expressions. The ensemble of such solutions is referred to as a population. Evolutionary operators, such as mutation, recombination, and selection are applied to this population. Solutions gradually improve by repeating a variation-selection cycle in the evolution process through numerous iterations. Essentially a search method, EC often produces well-performing solutions to complex

optimization and learning problems arising from various areas [1].

The fundamental idea of EC was gleaned from biology, and more specifically, from Darwin's theory of evolution by natural selection [32] as embodied in the neo-darwinian synthesis [104, 56]. However, knowledge of natural evolution has improved profoundly in biology in the past decades. This progress has, to a large degree, not been incorporated yet in computational models of evolution, and therefore cannot be harvested in applications. We have argued that adopting new developments from areas such as molecular genetics, cell biology, developmental biology and evolutionary biology would substantially benefit evolutionary computation [14, 152].

The question then arises what are the most important and revolutionary discoveries in biology in recent times, and how can they be sufficiently abstracted to provide material for computational models. As the number of scientists working in the areas mentioned above is now higher than at any other time before, and can be estimated to be well over a million, it becomes non-trivial to select those aspects of evolution that will have most impact in computational models. A number of books have appeared in recent years that provide some guidance in this quest (see, e.g., [82, 45, 156, 121, 26, 70]), and here we are mainly interested in the concepts of evolvability and the acceleration of evolution.

1.1 Evolvability

In the process of evolution, genetic variation explores new evolutionary material, the corresponding phenotypic variation provides adaptive characteristics, and stabilization operators including recombination and selection preserve improvements over previous generations. The cooperation of these operations is what allows evolution to work. Thus, the core mechanism of evolution is to assemble the forces of these operations yielding adaptive improvements that imply the evolvability of an evolutionary system. A growing number of efforts have been dedicated to the understanding [94, 126] and evolution [33, 159] of

evolvability.

While the concept of evolvability is still very much under discussion, we want to venture to propose a definition that is equally applicable to natural and artificial systems:

Definition 1: Evolvability. The capability of a system to generate adaptive phenotypic variation and to transmit it via an evolutionary process.

Altenberg [6] describes evolvability from a viewpoint of EC as the ability of a genetic operator or representation scheme to produce offspring fitter than their parents. In biology, Kirschner and Gerhart [81] suggest that evolvability should be understood as an organism's capacity to generate heritable and selectable phenotypic variation. An explicit comparison between evolvability of biological and computational systems has been performed by Wagner and Altenberg [159]. In their view, evolvability should be considered as the ability of random variants to produce occasional improvements, which depends critically on the plasticity of the genotype-phenotype mapping. The authors emphasize "variability" determined by the genotype-phenotype mapping as the *propensity to vary*, rather than variation itself. Marrow [99] suggests that evolvability means the capability to evolve, and this characteristic should be relevant to both natural and artificial evolutionary systems. He discusses a number of important contributions on this topic in both biology and EC, and raises some open questions for further research. Recently, a growing number of evolutionary biologists and computer scientists have shown interest in this topic. In an evolutionary system, many properties of a population are considered related to evolvability, including facilitation of extra-dimensional bypass and robustness against genetic variability [31, 157], redundancy and flexibility during developmental processes [81], and mutation rate adaptation [21].

The detection and investigation of evolvability are non-trivial and intriguing problems. Phenotypic fitness is directly observable and serves as a selection criterion. However, as a *potential* to generate better fitness and a *capability* for adaptive evolution, evolvability is difficult to observe and to quantify. Although a formal methodology on measuring evolv-

ability has not yet been agreed upon in the literature, some empirical methods have been proposed. Nehaniv [108] proposes the perspective of using evolutionary system complexity to describe and measure evolvability. He specifies the *exhibited evolvability* as an observable outcome generated by evolvability, and measures evolvability by the rate of the increasing complexity of evolutionary entities in an evolutionary system. Wagner proposes to simply measure the amount of non-neutral 1-step mutation variation in a biological system of particular relevance to evolution (RNA) in order to quantify evolvability [158].

Earl and Deem [36] suggest that evolvability can be selected for by varying the environment. By observing genetic changes in protein evolution, they find that rapid or dramatic environmental change generates strong selection pressure for evolvability. Thus, high evolvability can be detected and favored by such selection pressure. Reisinger et al. [129, 130] propose an indirect encoding representation to improve the evolvability by its capacity to facilitate effective search. A gradually changing fitness function is designed to measure evolvability of such a representation and to evolve a population that is adaptive under different environments. Furthermore, as the pace of change of the fitness function increases, stronger selection pressure for evolvability is imposed.

1.2 Evolution Acceleration

Related to the theme of evolvability is the rate of evolution. Evolvability defines how likely a system can generate adaptive phenotypic variation and the rate of evolution describes how fast this evolutionary process can proceed. As an fascinating topic in evolutionary biology, the rate of evolution has caused debates already since Darwin's time. Darwin himself held the view of *phyletic gradualism*, hypothesizing that most evolution occurs uniformly, gradually molded by selective conditions. Other were of different opinion, and Eldredge and Gould [39] proposed the theory of *punctuated equilibria*. According to this idea, evolution occurs through bursts of innovation followed by long periods of stasis, a major challenge to

Darwin's orthodoxy.

In biology, the rate of evolution has different definitions and measures depending on the underlying objects examined, for instance, gene sequences, proteins, organisms, etc. In molecular biology, the rate of evolution usually describes the rate of mutants being preserved as advantageous, i.e. those that can generate phenotypic improvements. Biologists use the k_a/k_s ratio to measure the rate of gene sequences evolution [93, 107, 112, 169]. It is known that some changes to a gene sequence may lead to differences in the amino acid sequence of an encoded protein while others will not, due to the degenerate code employed for translation. Therefore, such a measure can be used to compare two homologous protein-coding gene sequences of related species. The k_a/k_s ratio resulting from a measurement of the number of nonsynonymous (amino acid) substitutions per nonsynonymous site (k_a) to the number of synonymous substitutions per synonymous site (k_s) characterizes the rate of evolution between these two sequences. Since k_s measures neutral evolution (without considering functional improvements under selection pressure), the k_a/k_s ratio reflects the rate of *adaptive* evolution against the *background* rate of variation. In case $k_a/k_s > 1$, fixation of nonsynonymous substitutions is faster than that of synonymous substitutions, which means that *positive selection* fixes amino acid changes faster than silent changes. Mostly, however, one finds $k_a/k_s < 1$, the case where deleterious substitutions are eliminated by *purifying selection* (negative selection), and the rate of fixation of amino acid changes is smaller than the background rate of variation. If $k_a = k_s$, the fixation of these two types of changes is at the same rate. To summarize, measuring a large k_a/k_s ratio suggests that adaptation has been generated and fixed at a high rate. This measurement has been widely applied in the analysis of adaptive molecular evolution, and is accepted as a general method for measuring the rate of gene sequence evolution in biology.

Other than at the molecular level, Worden has defined the concept of *genetic information in the phenotype* (GIP) in his work on the speed limit for evolution [164], GIP is meant to

be a measure of the amount of genetic information expressed in observable phenotype, and he uses the rate of increasing GIP to describe the rate of evolution. He proposes that GIP measurements can be applied in both biology and EC.

In EC systems, the goal of evolution is much more specific than in nature. That is, to find the solutions to a given problem. Therefore, the rate of evolution in EC usually refers to the speed of solving a specific problem, e.g., to the speed of fitness improvements. The ability to define explicit phenotypic fitness is one of the most distinguishing features that differentiate EC from natural evolution. In order to investigate the performance of a computation model, the rate of evolution is thus mostly measured by the speed of fitness function improvements. Other ad hoc methods are also utilized in EC, like the efficiency of algorithms and CPU time. There are, however, some methods at a deeper level than simple fitness function improvement, that can be found in the literature. Bedau and Packard [20], for instance, propose a method for visualizing evolutionary adaptation. This method is useful to identify and measure the capability of creating adaptation during evolutionary processes. It is based on calculating evolutionary activity statistics of components in an evolutionary system. During a decade of extensive development, the notion of *evolutionary activity* has been applied to various scales of genetic components, including alleles, allele tokens, phenotypic equivalence classes of alleles and whole genotypes, in both artificial evolutionary systems and the biosphere. In their later work, two aspects for evolutionary adaptation were emphasized: the *extent* and the *intensity* of evolutionary activity [19, 127]. The extent of evolutionary activity concentrates on how much of an adaptive structure is present in an evolutionary system, while the intensity concerns the capability in generating new adaptive structures. The measures of *cumulative evolutionary activity* and *mean cumulative evolutionary activity* characterize the extent of a system's evolutionary adaptation. In addition, *new activity* is a measure of the intensity of a system's evolutionary adaptation. Evolutionary activity can be quantified and visualized during evolutionary adaptation. Its

derivative is the concentration of a component's current presence, and its second derivative can be argued to reflect the rate of evolution at a particular time. Evolutionary activity is also claimed to be a straightforward method for studying evolvability [19]. The argument is that, since a system with high evolvability can create highly adaptive variation, the quantification of evolvability can be determined from different levels of extent and intensity of evolutionary activity.

A number of observations on the factors that can accelerate evolution have been made over the past few years. Simon [143] brings up the "nearly completely decomposable" property in multicellular organisms, and proposes that this is an important property that can lead to faster fitness increases. In research on yeast genes, Gu et al. [60] report rapid evolution of gene expression and regulatory divergence after gene duplication. Gene duplication in biology contributes substantially to genomic and organismal evolution, and it provides abundant material for mutation and selection events to generate new gene functions in a modular way. By studying the recent nucleotide substitutions in human evolution, Hawks et al. [65] find that, as a population becomes more adaptive to its current environment, the rate of adaptive evolution will slow down. However, growing population size can provide the potential for rapid adaptive innovation. Thus, enlarging the population size and changing environmental conditions can both promote the rate of adaptive evolution. Kashtan et al. [76] confirm in a recent report that a varying environment can speed up evolution in an artificial evolutionary system. Other properties and techniques on the acceleration of evolution have been also investigated in biology and computing.

This review discusses evolvability and methods for accelerating artificial evolution by drawing ideas from complex natural systems. Notions from biology are introduced and their potential in designing new models in EC is discussed. Section 2 starts with the characteristics of populations. Exploratory and stabilization operations are investigated in Sections 3 and Section 4, respectively. Quality differentiation is discussed in Section 5,

followed by concluding remarks in Section 6.

2 Population

The general idea of EC is to adopt mechanisms of evolution from nature. In Darwin's theory of evolution both, the notion of variation and of natural selection are based on natural populations. However, populations simulated in a computer are simplified from their natural counterparts. For example, a major difference between these two systems is that no identical individuals exist in a natural population, whereas this is allowed and often the case in a simulated population. Slight variance is considered an essential aspect of natural populations as it leads to the enormous diversity in natural evolution. Hence, more details should be taken into account also in a computational population. An essential step in EC is to determine features of the simulated population since selecting a representation for individuals in the population and a size of this population can affect the performance of a computation model.

2.1 Representation

The first step for setting up evolution with a population is to decide on the representation of evolutionary individuals. Each individual should be encoded as a candidate solution to a given problem, which subsequently determines the search space of an algorithm. Therefore, the representation strategy is important because it predicates the input to the search process that should produce a satisfactory output. Here, we highlight two biological mechanisms, a protection mechanism for robust information preservation, and a communication mechanism for information interaction between different molecules. First we review general forms of redundancy in living systems. Then we discuss molecular interaction to encourage communication among different components of an individual.

2.1.1 Redundancy

Living systems may seem wasteful and luxurious to computer scientists. The most distinguishing aspects of biology compared to other natural sciences are complexity and diversity, which are of central concern to biologists. In the face of cruel competitive circumstances, organisms show great redundancy and resilience. Redundancy exists at different levels in natural organisms, including the genomic, transcriptomic, and phenotypic levels.

In biology, the genome of an organism is defined as the information encoded in DNA sequences and inherited from generation to generation. The double helix structure of DNA sequences itself is a form of protective redundancy of genetic information. Genomes carry genes and other non-coding DNA sequences. A gene is a string of base pairs grouped by a function generating proteins and polypeptides (protein fragments). Non-coding DNA sequences, formerly called "junk DNA", are not expressed as proteins, although they might be involved in their manufacturing process. However, genes are only quite small a fraction of the entire genome [131], with more than 98% of the human genome, for instance, being non-coding DNA sequences [9]. Furthermore, even a gene sequence itself is divided into exons and introns, where exons directly participate in protein expression but introns do not. Nevertheless, these non-coding DNA sequences are not useless. Modern biological discoveries indicate that they play an important role in the regulation of gene transcription [165]. Regulation mechanisms will be discussed later in Section 5.1.1. Wren et al. [166] indicate that tandem-repeat polymorphism in genes is quite common, and that such polymorphism can enhance the ability of some genes to respond rapidly to fluctuating selection pressure. The mechanism of gene duplication will be discussed in detail in Section 3.1.2. Moreover, it is well known that the human genome has two copies of each chromosome, one copy inherited from each parent. In recent research, a great number of human DNA segments are found to have more than two copies. The *Copy Number Variations* (CNVs) in human and other mammalian genomes discovered lately accounts for a substantial amount

of genetic variation other than single necleoside polymorphisms [47, 69, 139, 153], and it is considered to have an important contribution to phenotypic variation. This phenomenon will be discussed in detail later in Section 3.1.3.

The transcriptome describes the set of all transcribed RNAs in cells. In the human transcriptome, the proportion of transcribed non-protein-coding sequences is large and shows a great complexity [48]. Substantially more DNA is transcribed than is translated, and only a small proportion of mRNAs will be translated into proteins. The remaining part is called *non-coding RNA* or ncRNA. About 98% of all transcribed output in humans are ncRNA sequences [100]. Although many of the functions of these non-coding sequences are unclear, the high complexity of the transcriptome is considered important to studying mechanisms of gene transcription in the human genome.

As an important contributor to evolvability, genetic redundancy at both levels of genome and transcriptome has attracted increasing research interest in evolutionary biology [34], and was found to be created by a number of mechanisms. Krakauer and Plotkin [86], however, propose a new concept of *antiredundancy*. They declare that this antiredundancy emerges as well as redundancy in cells, and natural organisms can modify the redundancy properties in genotypes during evolution. Table 1 shows a summary of observed mechanisms responsible for both redundancy and antiredundancy at the cellular level. Mechanisms for redundancy mask the phenotypic effect of mutations and allow mutants to stay in sequences, while mechanisms for antiredundancy enhance the efficiency of local selection to remove damaged components.

Redundancy at the phenotypic level lies in an organism's robustness against intrinsic or environmental changes.

Definition 2: Robustness. The robustness of a biological or engineering system is its capability in functioning continuously in the face of genetic or environmental perturbations [156].

Table 1: Mechanisms responsible for creating redundancy and antiredundancy at the cellular level. (Adapted from Krakauer and Plotkin [86].)

Redundancy	*Antiredundancy*
Gene duplication	Overlapping reading frames
Neutral codon usage	Nonconservative codon bias
_	Gene silencing
Polyploidy	Haploidy
Multiple regulatory elements for n genes	Single regulatory element for n genes
Chaperone and heat shock proteins	_
Checkpoint genes promoting repair	Checkpoint genes inducing apoptosis
Telomerase induction	Loss of telomerase
Dominance	Incomplete dominance
Autophagy	_
mRNA surveillance	_
Bulk transmission	Bottlenecks in transmission
Molecular quality control	_
tRNA suppressor molecules	_
Modularity	_
Multiple organelle copies	Single organelle copies
Parallel metabolic pathways	Serial metabolic pathways
Correlated gene expression	Uncorrelated gene expression
DNA error repair	Loss of error repair

With low robustness, a species would gradually go extinct due to lethal mutations because random mutations are usually a cause of deleterious changes that destroy an individual. One may note that robustness and evolvability seem to have a contradictory relation. Since a system has high robustness, it can be tolerant to intrinsic or environmental changes that less variation will be required, thus it should be less evolvable. Wagner [157, 158] discusses this contradiction and resolves it. He distinguishes robustness and evolvability as quantities at both the genotypic and the phenotypic level. If one considers genotype, the more robust a genetic sequence is, the less innovation this sequence will produce. However, robustness and evolvability are characteristics of an entire system and if investigated at phenotypic level show a strong correlation. A system with high phenotypic robustness harbors a great number of "neutral" variations that have no functional effects. According to the notion

of *exaptation* proposed by Gould and Vrba [57], these neutral variations do not change phenotypic function during relatively static evolution periods but may be able to generate adaption later under certain genetic or environmental changes. Thus, a system with high phenotypic robustness provides a great potential for phenotypic innovation in the future, i.e., high evolvability [158].

Redundancy is wide-spread in natural organisms as an efficient protection mechanism against internal or environmental changes, whereas in EC models components that do not seem to be immediately relevant are often considered superfluous. In recent years, however, representation redundancy has arisen as a by-product of computational evolution and has attracted increasing interest from EC researchers.

Definition 3: Representation Redundancy. In genetic and evolutionary algorithms, representations are redundant if the number of genotypes exceeds the number of phenotypes [134].

Rothlauf and Goldberg [134] examine the effects of redundant representations on the performance of an EC system both theoretically and empirically, and they propose that redundant representations can increase the reliability and efficiency of EC models. Specifically in genetic programming, representation redundancy is usually identified as *introns* (or non-effective, neutral code) [16] in programs. Researchers have made great effort to investigate both the positive and negative effects of introns [91, 97, 116, 170], and a positive relation between neutral code and evolvability in genetic programming has been suggested. The important role of redundancy in evolvability has been realized though most of this redundancy comes as a by-product of evolution. We might, therefore, consider designing protective redundancy into our algorithms to make them resilient against changes while improving adaptation. Such capabilities complicates the algorithms but may be worthwhile if the resulting robustness can generate higher evolvability when applying intense changes (pressure?) to produce adaptive responses. Furthermore, evolution can be accelerated be-

cause the system has a quick and robust reply to evolution pressures. This is particularly true with the growth computational power available today.

2.1.2 Molecular Interaction

Natural living systems are remarkably diverse from simple organisms such as bacteria to highly complex creatures such as humans. This diversity is not the result of vastly different chemical constituents of creatures. In fact, many organisms carry out similar metabolic, cell division and replication processes under similar assembly principles [18]. The differences that distinguish natural species are caused by the arrangement and distribution of basic building blocks [71] and molecular interactions contribute significantly to these organizational mechanisms.

Interactions in a cell can happen between the same type of molecules, such as protein-to-protein interaction, or between different types of molecules, such as protein-to-DNA interaction. Signals can also be sent and responded to by cells in multicellular organisms. Molecular interactions can be triggered by energy supply, e.g., in *metabolic pathways*, a chain of interaction catalyzed by enzymes, or triggered by external stimuli, e.g., *signaling pathways* that enable communication through the cell membrane [28]. Proteins are not only a product enabling various organismal structures, but also work as control factors in various processes from the synthesis of cells, metabolism, gene regulation, to sexual reproduction. They accomplish complex tasks under the condition of frequent and continuous communication between molecules and cells.

Metabolism is a key process to maintain the growth and reproduction of cells. The metabolic pathway of a cell is an elaborate network of numerous chemical reactions catalyzed by enzymes. Most enzymes are proteins, and they are an important factor in cellular reactions [160]. Different types and amounts of enzymes are produced according to different energy supplies, and these enzymes will determine different metabolic pathways by their

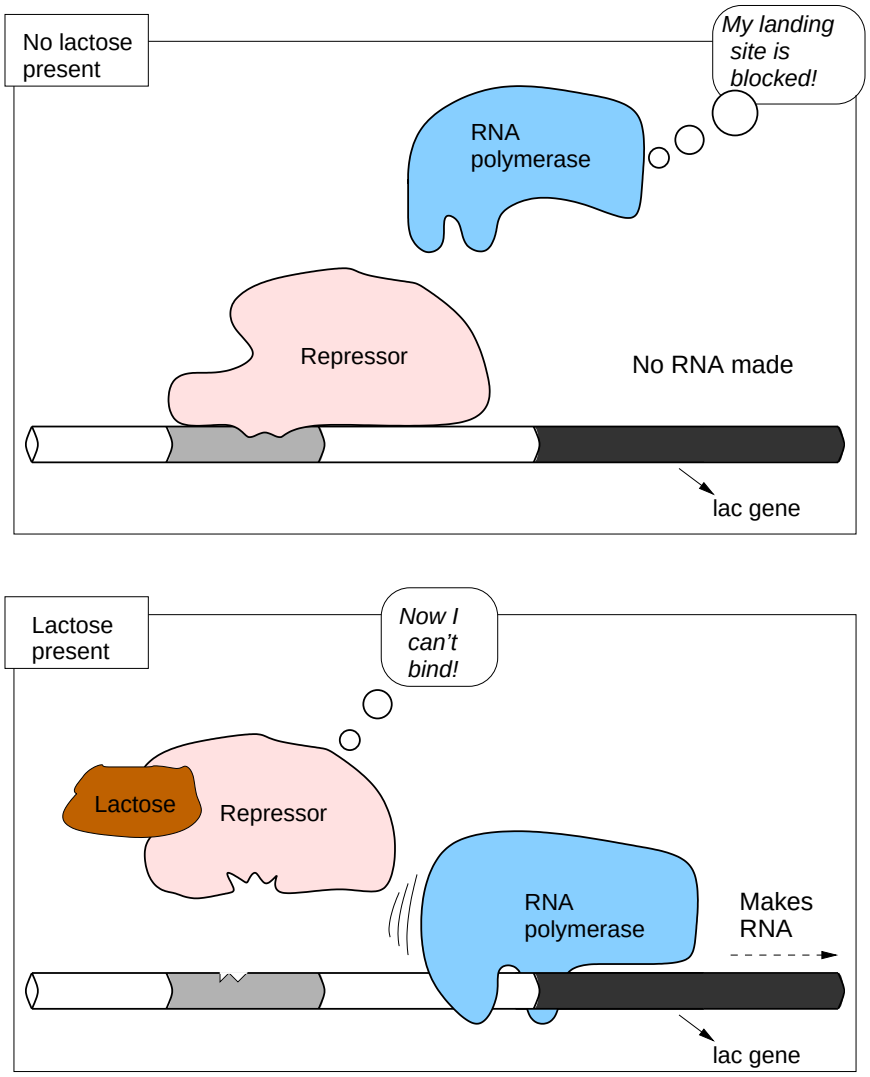

Figure 1: The genetic switch in Jacob-Monod model. A specific repressor protein acts as the switch. When it binds to a DNA site near the gene encoding beta-galactosidase, the RNA polymerase protein cannot bind nor can it synthesize RNA from the gene. The gene is turned off. When lactose is present, it binds to the repressor and keeps it from the DNA site. The gene turns on. (Adapted from Kirschner and Gerhart [82].)

catalysis. In the process of gene expression, the function achieved can be controlled by molecular interactions. For instance, the process of how a parsimonious bacterium responds to food supplies during metabolism shows a simple genetic switch mediated by molecular

interactions. Since the metabolic pathways of bacteria are much simpler than those of multicellular organisms, the regulation of gene expression is more tractable in bacteria. The phenomenon of *enzyme induction* [82] describes the adaptation of a bacterium to material supplies by producing varying amounts of enzyme. What triggers this production and how does this mechanism work? The Jacob-Monod model (shown in Figure 1) first described the regulation mechanism of inhibiting or repressing genes by an inhibitory proteins, called *repressor* in bacteria. The binding of lactose to a repressor enables the production of RNAs by removing the repressor from its binding sites on the gene sequence where RNA polymerase can bind. However, this is not a simple on-off switch model. The continuity lies in the binding duration which decides the rate of protein synthesis. Therefore, if more sugar is absorbed during metabolism, more protein is synthesized by RNA translation. This simple sugar metabolism model captures the mechanism of how a repressor affects gene functions. The enzyme here works as a trigger for the protein synthesis process under various molecular interactions. In addition, most enzyme effects are sensitive to ambient temperature [3], which is an important parameter to control metabolic interactions.

Signaling and cellular responses to signals are complex. These responses are controlled by a plethora of positive and negative feedback loops. The presence of feedback complicates the simple picture of a linear pathway, but is an essential part of the signaling process [18]. This makes signaling pathways involving molecular or cellular communication a network-like structure, with complex regulatory processes at work. The cellular infrastructure of eukaryotic organisms is only a few times larger than that of bacteria, but the complexity of signaling network control differs greatly, by orders of magnitude. The linkage between various parts of the gene expression apparatus in eukaryotic organisms is weakened by a far less-precisely defined control than that found in prokaryotic cells [81]. For instance, geometric requirements for binding sites are significantly relaxed in eukaryotic gene regulation. A repressor does not have to bind at the exact position of a target, but just in the

neighborhood. Such a weak linkage also enables potential interactions between different gene sequences by lowering constraints for cooperation. Signaling between cells is possible only after a sufficiently large number of repressors participate simultaneously. Also, a single signal may incur a much more complex response [31]. Allosteric proteins, which have multiple sites for interaction, also make gene expression much more flexible because they have different sites for different functions. The regulatory decisions on which genes are transcribed when, where, and under what circumstances makes eukaryotic cells well conserved but enormously adaptive to generate new phenotypes in changing environments [96].

Computational models used to analyze and understand multi-input/output, high-order complex signaling systems have shown recent progress in bioinformatics [114]. In contrast, current EC models are mostly limited to representing evolutionary material based on the infrastructure of natural organisms, while disregarding the vast potential of interaction mechanisms at both the molecular and cellular levels. The absence of such mechanisms in EC, however, reveals significant research opportunities in this area.

2.2 Population Size

After the encoding of an individual is determined, a population is set up. Several features of a population are tightly coupled to its evolutionary capabilities. In a computational population, the most relevant feature is population size itself.

In nature, different species have various population sizes, and this characteristic has been realized to play an important role in evolution. In the living world it is common that smaller groups constituting species evolve faster, though smaller groups sometimes become extinct, while species with larger populations evolve slower, staying unchanged for relatively long periods. However, neither a small nor a large population size is unconditionally beneficial to evolution. The relation between them should be understood in different scenarios.

The study of *population genetics* was formulated by Fisher [41], Haldane [61] and

Wright [167]. It focuses on gene frequency changes in populations under the effects of natural selection, mutation, genetic drift, and population size fluctuation. In this field, scientists have examined the role of population size in molecular evolution using mathematical analysis. The rate of molecular evolution is usually measured by the nonsynonymous to synonymous substitution ratio k_a/k_s, discussed in Section 1.2. Decades ago Kimura [78] proposed a strong dependency of the rate of molecular evolution on population size . More recently, Gillespie [49, 50] has conjectured that there is only a very weak dependency on population size. Between these two opinions, Ohta [118] finds population size to be related to the rate of evolution under different assumptions regarding mutation types. The *nearly neutral theory* of molecular evolution proposed by Kimura and Ohta [119, 120] has been regarded as one of the most important foundations for modern molecular evolution analysis. In this theory, it is predicted that there is a substantial number of nearly neutral mutations (including slightly deleterious and slightly advantageous ones) in molecular evolution, and that these contribute to evolution by providing a potential for future phenotypic innovation. Ohta [118] predicts that population size affects the rate of evolution under various mutation patterns. If most mutations are deleterious, a smaller population can evolve faster, because the chance of a slightly deleterious mutant being favored by selection is greater within a smaller population and these nearly neutral mutations bring genetic variation and may further trigger phenotypic innovation. In contrast, if mutations are mostly advantageous, the rate of evolution in a larger population is greater. If most mutations are neutral, the evolution rate is nearly independent of the population size. In general, random mutations are more deleterious than advantageous in natural systems. Thus species with a small population size usually evolve fast.

A number of studies focus on testing the relation between population size and evolution rate by using comparisons. Island endemic species usually have small population sizes because they are restricted to a limited geographical region. Woolfit and Bromham [163]

study species on islands in support of the effect of population size on the rate of molecular evolution. They compare island endemic species to closely related species on a nearby mainland and find that island endemic species have a significantly higher nonsynonymous to synonymous nucleotide substitution ratio than their counterparts on the mainland. This result indicates a decrease in the population size will lead to an increase in the rate of evolution. Wright et al. [168] study tropical species which are generally regarded to have a rapid molecular evolution rate due to several factors such as latitude and climates. It is believed that tropical organisms possess great species richness and dynamics with small but highly diverse populations [149]. However, there are also exceptions that increasing population size can accelerate evolution as well. By studying the recent rapid molecular evolution in human genomes, Hawks et al. [65] suggest that if a population is highly adapted to a current environment, evolution will become stagnant. Under these circumstances a growing population size can provide the potential for rapid adaptive innovations. Thus, enlarging the population size under chaotic environments can promote the rate of adaptive evolution.

Population size is also involved in research on genome robustness. Visser et al. [34] postulate that the population size should be sufficiently large for selection to be effective to evolve the robustness of a system. Small populations have difficulty to achieve this robustness. In a different study Krakauer and Plotkin [86] find, however, that small populations will also favor evolving robustness by increasing genetic drift pressure and a buffering mechanism of hiding mutations from being diminished by selection. This hypothesis is supported by Elena et al. [40]. Among the different authors, there is an agreement that the effect of population size, either large or small, varies in different models.

In the EC community, many efforts have been dedicated to the optimization of population size [95] in algorithms, since it is believed to have a high correlation with the performance of an EC method. The challenge is that adapting population size is problem-specific

and to date it is still unclear how to estimate the relations among various EC parameters. In general, current work on this topic concentrates on two tasks: (i) initializing a proper population size prior to a run, and (ii) adjusting population size during a run. Most theoretical works on population size initialization are based on Goldberg's components decomposition approach and the notion of *Building Blocks* [52, 54]. With many other publications, these contributions propose to choose the population size according to the "hardness" of a specific problem. They reveal a general principle in setting population size: The more difficult a problem is, the more diversity is required and the larger the population should be. In later research, it has been realized that even for a specific problem the requirement for population size may differ during different stages of evolution. Therefore, in addition to setting a proper population size beforehand, many empirical methods on adjusting population size during runs have been proposed recently, such as the *Genetic Algorithm with Variable Population Size* (GAVaPS) proposed by Arabas et al. [8], the *parameter-less GA* by Harik and Lobo [62], the *Adaptive Population size Genetic Algorithm* (APGA) by Back et al. [10], and the *Population Resizing on Fitness Improvement GA* (PRoFIGA) by Eiben et al. [38]. However, mechanisms for dynamically adjusting population sizes in EC are much simpler than those found in nature, in that a fluctuating population size still has little to do with mutation and selection patterns in different evolutionary stages. This relation requires further exploration as a promising indicator of adjusting population size during a run.

3 Exploratory Operations

Evolvability is understood as the capability to generate offspring fitter than the parents [6]. Exploratory operators are mainly responsible, since generating variation is the method to search for more adaptive individuals. Due to the complex mapping process from genotypic to morphological level in biology, genotypic and phenotypic variation discussed separately.

3.1 Genotypic Variation

Genotypic variation generally means mostly changes happening to DNA sequences in both protein-coding and non-coding regions in the form of point mutation and gene rearrangement. Gene sequences are highly conserved against lethal changes that would likely lead to destructive consequences because a tiny mutant at the genetic level can cause a great change in function [82]. In contrast, changes to the regulatory or non-coding part of sequences are considered more able to increase adaptability and plasticity of a system. In this section, we will discuss the general form of mutation first and then gene duplication as the most important form of rearrangement, followed by a comparison between point mutation and gene rearrangement.

3.1.1 Mutation

Searching for the essential driving force of evolution has been a central topic in evolutionary biology. Since Darwin declared that natural selection is the main force of evolution, controversies have arisen on different aspects of this explanation. In modern biology, the two main schools of thoughts are selectionism and neutralism [111]. Some scientists argue that genotypic variation is maintained by selection, which is the central perspective of *neo-Darwinians*. Other evolutionists insist that high genotypic variation can be explained as a result of neutral mutations. In either case, mutation is accepted as a major mechanism to generate genotypic variation.

Mutation can happen at either coding or non-coding regions of DNA sequences, and may consequently cause functional or regulatory changes. The notion of neutral mutation is based on the fact that the majority of mutations have no consequent effects on protein function due to DNAs' repair abilities and the redundancy in the translation apparatus. By observing the rate of nucleotide substitution, neutralism proposes that mutations change the function of gene products barely noticeably [111]. High genotypic variation might exist, but

fatal destruction will not be able to influence species because fixation of those mutations will not happen. Therefore, innovation from random mutations is capable of generating adaptive morphology rather than random properties.

What triggers mutation and what is the relation between mutation and selection? Does selection pressure indeed generate new mutations or simply allow existing mutants to be fixed faster than before? Research on mutation under selection has received wide interest since Darwin's time. Controversies have arisen regarding the effect of selection pressure on mutation, and different models have been proposed in the meantime [133]. It is now believed that it is impossible to separate any form of mutation from the effect of selection. Roth and Andersson [132] define *adaptive mutations* as fitter mutations that arise under selective conditions to investigate "directed" mutation pathways. In subsequent work [66, 87, 145], their group propose a gene duplication-amplification model to study the mutagenesis stimulated by enhancement of selection. In addition, a recent study by Weinreich et al. [161] on the effects of Darwinian selection on random mutation argues that environmental selection can make some multi-step mutation pathways unaccessible. By studying "five point mutations" in a lactamase allele that can increase bacterial resistance to an antibiotic, several mutation pathways are in principle possible for these mutations. After calculating the different probabilities of these pathways, their experimental results show that under intramolecular interactions that increase the fitness of proteins, only a small number of pathways are really accessible. This is quite an interesting phenomenon because mutations might be supervised by some unknown fitness-increasing principles and consequently product proteins might be reproducible and even predictable. These feedback and interaction mechanisms may reduce the harm that mutations could bring to organisms. This point of view also conforms to Kirschner and Gerhart's definition of evolvability [81], which they define as "the ability to reduce the potential deleterious mutations and the ability to reduce the number of mutations needed to produce phenotypically novel traits". If mutations can be supervised, fewer

changes are needed to generate the required adaptation. Therefore, evolvability is improved by reducing the cost of mutations.

In EC, mutation is regarded as an important exploratory operator. Artificial evolutionary search should be good at both exploring suitable genetic novelty and maintaining successive improvements. Holland [68] discusses this principle as the tension between "exploration" and "exploitation". To keep this balance, a very important parameter for mutation is the mutation rate, which has already been considered as an evolvable parameter contributing to evolvability. Bedau et al. [21, 22] divide evolutionary adaptation conceptually in two stages: the novelty stage, where an evolving system enhances its adaptability against a changing environment, and the memory stage, where the evolving system is building up this adaptability through incremental improvements. By providing a simple two-dimensional model, Bedau et al. postulate that the mutation rate should increase during the novelty stage and decrease during the memory stage. This fluctuation of mutation rates is able to keep the balance between evolutionary novelty and memory, and thus to increase the evolvability of adaptive systems.

However, compared to natural evolutionary systems, genotypic variation in computation is somewhat arbitrary and not as adaptive. First, the fixation process of mutations is not simulated appropriately in EC, because all changes to individual sequences are translated into phenotypic properties. Recovery or repair mechanisms are usually not applied to individuals suffering deleterious mutations, which make those individuals unfavored during the selection process. Second, biological development of selection-driven mutation pathways is an interesting direction to explore in computational models. These topics should be considered in future research in EC to make the best use of the evolutionary power of mutation.

3.1.2 Gene Duplication

Natural selection acts mostly at the phenotypic level, but genetic changes are the major thrust of evolution. *Gene duplication* is an important mechanism creating new genes and new genetic subsystems. This mechanism has been recognized to generate abundant genetic material and contributes substantially to biological evolution [117]. A large number of duplicate genes have been discovered to exist in vertebrate genomes [110], and a repeated number of whole genome duplications is being considered as key events in evolutionary history [147]. In modern biology, gene duplication and its subsequent function-specialized divergence are widely believed to be a major reason for functional novelty.

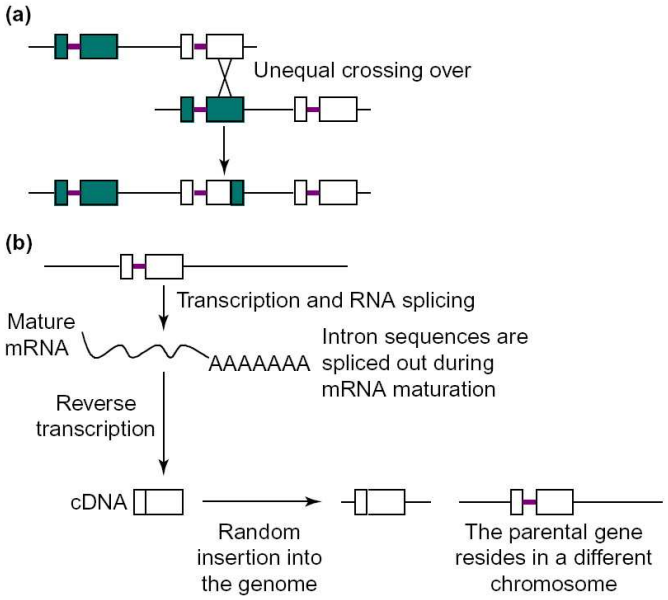

Figure 2: Two common modes of gene duplication. (a) Unequal crossing over, which results in a recombination event in which the two recombining sites lie at nonidentical locations in the two parental DNA molecules. (b) Retroposition, which occurs when a message RNA (mRNA) is retrotranscribed to complementary DNA (cDNA) and then inserted into the genome. Squares represent exons and bold lines represent introns. (Adapted from Zhang [171].)

Gene duplication is usually generated by unequal crossover or retroposition [171] (see

Figure 2). Unequal crossover is similar to but different from normal crossover that occurs when two chromosomes exchange a proportion of DNA at the same locus in base pair sequences. Unequal crossover happens if this exchange occurs in different loci, with the consequence that duplicate genes appear in one chromosome while the other turns out to contain *pseudogenes*. Retroposition happens when an mRNA is retrotranscribed into a complementary DNA (cDNA) and then inserted into the original genome. Besides such gene duplication, duplication at other scales in cells has been discovered recently [11, 147], including segmental duplication and whole-genome duplication. Here, we only consider gene duplication. The main products of gene duplication are called *paralogous genes*, a type of homologous genes. Studying homologous genes has emerged as an efficient way to search for molecular innovation and species formation. Homologous genes have two main categories, *paralogs* and *orthologs*. Paralogs are results of gene duplication and code for proteins with different functions. Orthologs are the products of speciation events and the proteins they code for serve similar functions.

Once a gene duplication has happened a complex fixation process on the duplicate genes takes place. *Purifying selection* and *gene conversion* are the main pressures affecting the survival of duplicate genes [113]. Most duplicate genes become pseudogenes after one or more mutations disable them and no promoting functions is yielded. However, multiple copies of identical genes can, after duplication, promote functional redundancy against fatal changes. The process of *pseudogenization* is reported occurring in the early stage of rapid evolution [63], with evidence of many pseudogenes found in the human genome. Other duplicate genes are changed by selection pressure and functionally diverge. Subfunctionalization and neofunctionalization are the two main channels of functional divergence [171]. In *subfunctionalization* of two gene duplicates, shown in Figure 3, each copy adopts a different aspect of the function of the original gene. Both copies will be stably maintained because both aspects of the function are indispensable. Subfunctionalization leads to functional spe-

cialization by dividing multi-functional genes once the newly emerged genes perform better. Alternatively, some relatively new function can also evolve after gene duplication [172], and this process is called *neofunctionlization*. This has been revealed earlier as the Dykhuizen-Hartl Effect [35] where a random mutation is preserved in the duplicated gene by reducing selection pressure due to functional redundancy that results from gene duplication. Such mutations may accumulate and induce a genetic function change depending on conditions of a dynamic environment. New adaptive functions may thus be generated and later preserved during evolution. By possibly creating novel functions and allowing evolution under fewer constraints, neofunctionlization is an important outcome of gene duplication.

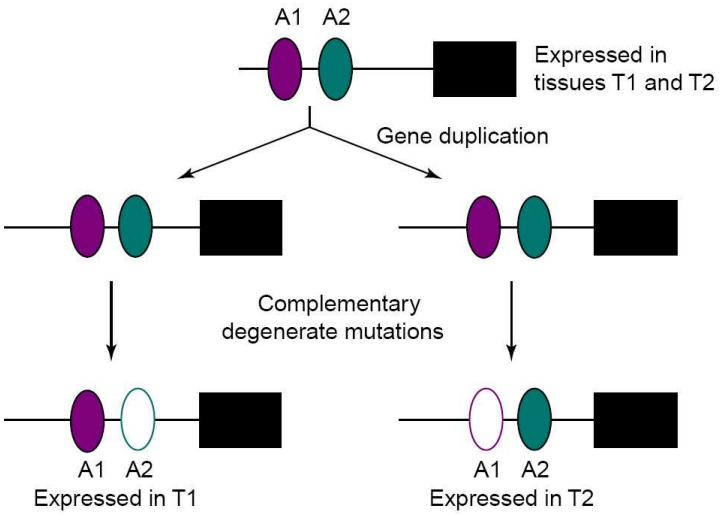

Figure 3: Division of expression after gene duplication. Squares represent genes, closed ovals represent cis-acting elements that regulate gene transcription, and open ovals represent deactivated cis-elements. Consider a gene that is expressed in tissues T1 and T2, with a cis-acting regulatory element A1 controlling the expression in T1 and A2 controlling the expression in T2. Following gene duplication, one daughter gene might lose the A1 element whereas the other gene might lose A2, so that each is expressed in only one of the two tissues. (Adapted from Zhang [171].)

In brief, gene duplication contributes substantially to genomic and organismal evolution. It provides abundant material for mutation and selection, and allows to specialize function

or generate completely new functions. The acceleration of protein sequence evolution after gene duplication has recently been reported in research on yeast genes, Gu et al. [60]. The authors use an additive expression distance between duplicate genes to measure the rate of expression divergence, and rapid evolution of gene expression and regulatory divergence after gene duplication is observed.

In summary, the mechanism of gene duplication can considerably increase evolvability of a system by reducing the cost of mutations. In EC, the idea of using gene duplication and deletion operators was proposed some time ago. Those operators are in general based on the method of variable-length genotypes, and are executed with predefined duplication or deletion probability [59, 85, 136, 137]. Unfortunately, so far only application-oriented work has appeared with different representations [25], and a common framework for this concept is missing. More details of gene duplication in biology should be taken into account to benefit computational evolution. In particular, the question of how gene duplication reduces the limitations of mutation and selection, and in the process promotes evolvability needs to be studied. Is there a way to implement functional specialization and innovation through gene duplication in EC?

3.1.3 Point Mutation vs. Gene Rearrangement

A point mutation occurs when a base on a DNA sequence is changed to another base at the same locus. Gene rearrangement is a change in the order of a DNA sequence on a chromosome. This change can be an inversion, translocation, addition, or deletion of genes. Earlier research focused mostly on *Single Nucleotide Polymorphisms* (SNPs) in genomes due to the enormous complexity of genetic sequence analysis, but gene rearrangement was still believed to contribute to evolvability more than simple point mutation does [79]. Fortunately, the recent development of technology has facilitated the shift in focus from a locus-based analysis to a genome-wide assessment of genotypic variation [47].

Genetic rearrangements, rather than point mutations, can maintain the connective information internal to gene sequences. Because genes form networks of functional control, rearrangement is better able to preserve internal structures. Genetic changes are highly constrained to gene sequences and gene rearrangements occur far more frequently than point mutations.

The ubiquity of *Copy Number Variations* (CNVs) has been realized recently in mammalian genomes by different groups of biologists, such as Iafrate et al. [69], Sebat et al. [139], and Tuzun et al. [153]. CNV is regarded as a predominant type of genotypic variation leading to vast phenotypic diversity in mammalia. CNVs show that large segments of DNA, with sizes from thousand to millions of base pairs, can vary in copy number of genes. This variation can lead to protein dosage differentiation in the expression of genes, and CNV is therefore regarded as being responsible for a significant proportion of phenotypic variation [47]. The mechanisms that create CNV have not yet been clearly understood, but some hypotheses have been proposed in the literature. Fredman et al. [46] and Shaw and Lupski [140] propose that CNV might be the result of large segmental gene duplications or non-homologous recombination events.

Recent bioinformatics research uses statistical and computational tools to analyze chromosomal evolution by a comparison of genome-rearrangements between sequences of related species [135]. Although the biochemical mechanisms of gene rearrangement are still far from being fully understood, it is time to start simulating such rearrangement operations in computational models in EC. Particularly, the recent discovery of CNVs requires attention by computer scientists, in order to achieve similar benefits in EC.

3.2 Phenotypic Variation

As mentioned in Section 2.1.2, despite their vast phenotypic differences, metabolic processes and cell structures in bacteria and humans are quite similar [82]. What, then, makes

humans so different morphologically from other organisms? It is phenotypic variation. Unfortunately, the relation between genotypic variation and phenotypic variation is still not fully clarified in current biological research. Since selection acts on phenotypes rather than on genotypes, phenotypic variation should be used to explain the immense diversity among organisms. Here, we discuss several aspects of phenotypic variation. We leave the discussion of the mapping process between genotype and phenotype that controls the direction of phenotypic changes resulting from genotypic variation to Section 5.1.

3.2.1 Conservation and Relaxation

According to Kirschner and Gerhart, evolution possesses two important features: conservation at the molecular level and relaxation at the anatomical and physiological level [82]. By conservation it is meant that the genetic components of organisms tend to maintain relatively stable structures; relaxation refers to the less constrained phenotypic diversification of organisms. The authors state that conservation on the genotypic level reduces the constraints on the phenotypic level.

In Darwin's evolutionary theory, all organisms have evolved from the same ancestor. After primal initialization and evolution, genetic structures of organisms are highly conserved during the course of billions of years [160]. This can well explain why the number of human genes is only a few times that of bacterial genomes but significant anatomical and behavioral differences exist between them. A surprisingly small number of genes in humans and other complex organisms shows that the great diversity and complexity at the anatomical and physiological levels have to rely on and organize / reuse limited genetic material. When certain organisms need to improve their adaptivity in order to survive in a new environment, the regulation system only has to recombine existing mechanisms for the generation of adaptive functions, which requires little or no new genetic material [81]. Not only gene sequences are highly conserved, the *core processes* of coordination of the genetic

material are also well conserved since the time they initially emerged [82]. These conserved core processes are used repeatedly for different purposes and functions under different circumstances, at different times, with different genetic material. The Baldwin Effect [144] explains that phenotypic variation is not generated out of the blue but through regulation of existing components in organisms: Mutation simply happens to stabilize and to extend what has already existed to improve somatic adaptability towards external stimulations.

This conservation mechanism can efficiently prevent lethal changes in genotypic variation and is an economic method to increase the adaptability of organisms. New material is not needed to adapt to changing environments, but a few modifications will suffice.

Functional innovation is heavily constrained due to molecular interactions among various genetic components that are involved to produce a specific trait. If the participation of more genetic components is needed, it becomes harder for functions to change. In fact, relatively little genetic material is required to generate all proteins of organisms. Under selection pressure from a changing environment, organisms have to yield adaptive phenotypic traits to survive, however, and the highly conserved core processes mentioned above are used repeatedly to generate new cooperation among the conserved genetic material, bringing about fitter function and behavior. Relying more on the combinatorics of components is equivalent to a relaxation on phenotypic variation.

The relaxation on phenotypic variation has been highlighted as the notion of "deconstraint" in Kirschner and Gerhart's [81] research on evolvability which studies the mapping from genotype to phenotype. Enhancing phenotypic variability under changing environmental conditions allows nature more evolvability. Not only can deleterious changes be avoided, but nonlethal genetic and phenotypic variation is indeed the material from which innovation can be generated.

What is the role of conservation and relaxation in EC? First, an economic use of genomes or building blocks can help to conserve genetic information. Second, by reducing the con-

straints on making changes to a phenotype we can try to enhance the exploratory capability of a computational system to find better solutions. How that can be implemented in actual systems is presently unknown, but a worthwhile line of inquiry.

3.2.2 Modularity

Modularity is a widespread structural property of complex systems. It has attracted considerable interest from studies of both natural and artificial evolutionary systems, and is regarded as strongly related to evolvability [159] and the acceleration of evolution [143].

Modularity exists at various levels, .e.g., at the level of gene expression or embryonic development. Here we adopt the definition of modularity proposed by Simon [142] in his research on hierarchies in complex systems.

Definition 4: Modularity. In a complex system, modularity refers to the property that a loose horizontal coupling exits between the entities at the same level of this system. [142].

Simon [141, 143] further defined that "a system is *nearly decomposable* if it consists of a hierarchy of components, such that, at any level of the hierarchy, the rate of interaction within components at that level is much higher than the rates of interactions between different components". Although this "Near Decomposability (ND)" is attributed to a *vertical* separation while modularity describes the separable property of components at the same level, they seem closely related that they both describe how a complex system is decomposed into sub-systems.

The modularity property of genotype-phenotype mappings has been extensively studied in gene expression. It reduces harmful pleiotropic effects of gene expression and can lead to adaptive phenotypic variation. *Pleiotropy* is a general property of genotypic variation, expressing the fact that one change at the genetic level can cause a multitude of functional changes at the phenotypic level. Pleiotropy can generate both advantageous and disad-

vantageous results. Pleiotropy can sometimes generate unexpectedly improved functions, but can also be harmful or even fatal to evolutionary systems [7]. Since a gene can affect multiple functions, optimizing one particular function at the phenotypic level inevitably incurs side-effects on other functions. Bonner [24] proposes the notion of "gene nets" by grouping gene actions and their products into discrete units during evolution. In general, for a given organism, the mapping from genotype to phenotype can be divided into modules such that the sets of genes in one module only affect the functions in that same module. The mapping is therefore decomposed into groups of independent "sub-mappings". Bonner finds that the phenomenon of gene nets becomes increasingly prevalent as organisms become more complex. Wagner and Altenberg [159] investigate modularity in genotype-phenotype mapping in both biology and EC views. They interpret modularity as a means for dividing phenotypic traits into different "compartments" to reduce interference among different optimization modules. With such modularity, optimization of a function in one module has no effect on functions in other modules. As a result, pleiotropy can be confined to a known set of functions during evolution. Figure 4 shows a simple example of this idea of modular separation.

Wagner and Altenberg [159] further propose that modularity results from evolutionary modifications in natural organisms. In their view, the evolution of modularity follows two mechanisms, dissociation and integration. Dissociation is the suppression of pleiotropic effects by disconnecting interactions between different modules, while integration is realized by strengthening of pleiotropic connections among traits in the same modules. Both mechanisms are driven by the selection pressure.

Thus, modularity can be conceptualized as an evolutionary mechanism to promote evolvability. It reduces the interdependence of disjoint components and consequently reduces the chance of pleiotropic damage by mutation [81]. It allows genotypic variation and selection to affect separate features in a complex system and to evolve various functions without

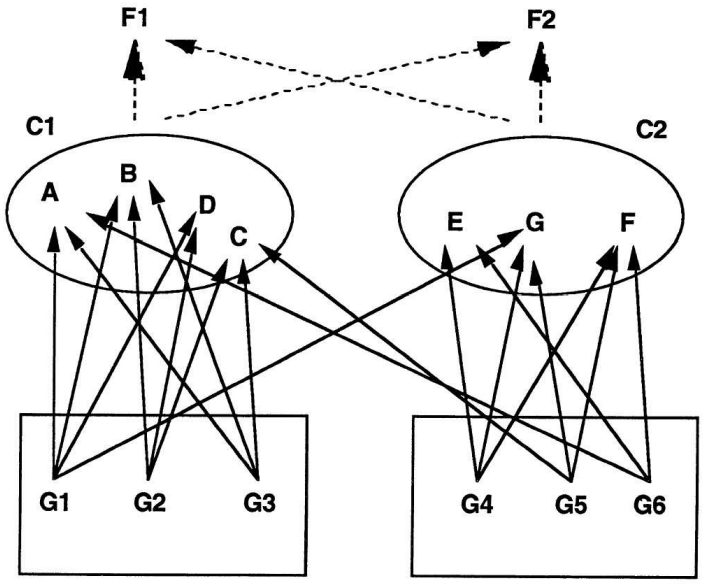

Figure 4: Example of a modular representation. Complexes C1 ={A, B, C, D} and C2 ={E, F, G} serve to functions F1 and F2. Each character complex has a primary function, F1 for C1 and F2 for C2. Only weak influences exist of C1 on F2 and vice versa. The genetic representation is modular because the pleiotropic effects of the genes M1 ={G1, G2, G3} have primarily pleiotropic effects on the characters in C1 and M2 ={G4, G5, G6} on the characters in complex C2. There are more pleiotropic effects on the characters within each complex than between them. (Adapted from Wagner and Altenberg [159].)

interference [92]. Sub-systems as part of an entire system can evolve faster to optimize their local sub-functions individually, by decreasing crosstalk between genetic changes. In a study of encoding schemes in EC by Kazadi et al. [77], a *compartment* is defined similar to a module in the genotype-phenotype mapping, and such compartmentalization at different levels is claimed to contribute to the acceleration of evolution. In RNA research, Manrubia and Briones [98] propose that the increase of molecule length and subsequent increase in function complexity could be mediated by modular evolution. They find that short replicating RNA sequences with a small population size can be assembled in a modular way and create complex multi-functional molecules faster than conventional evolution of complex individuals toward multiple optima.

Modularity in general has been widely used in computer science and engineering by subdividing complex entities into smaller components to yield higher computational efficiency, and we expect it to play a major role in EC. In current EC models, phenotypic variation is mostly generated from genotypic variation with mappings that are not very complex. It is our opinion that considering genotypic and phenotypic variation separately but connected with a number of complex and sophisticated evolutionary mechanisms like modularity, will allow EC to benefit substantially.

3.2.3 Facilitated Variation

Kirschner and Gerhart [82] emphasize that variation is much less random at the phenotypic level of organisms than at the genotypic level, where genetic mutations show considerable randomness. Since phenotypic variation should be favored by selection via modifying existing evolutionary components, this variation is *facilitated*.

Kirschner and Gerhart summarize three principles of facilitated variation. It serves (i) to reduce lethal pleiotropic effects, (ii) to increase phenotypic variation in light of a given number of genetic changes, and (iii) to improve genetic diversity in evolutionary populations (by reducing lethality). Evolution is not so much affected by the content of genetic and protein structures but by regulation capabilities to organize and reuse these functional parts and to decide the targets of such regulation. The core processes instead are conserved being built in a special way, only to be linked together under new circumstances like time, place, and the number of genetic material that may participate in generating new phenotypic variation. It is clear that only adaptive phenotypic variation can be maintained during evolution, and the relevant product proteins mostly will have multiple functions for various adaptive requirements under selection.

We must realize that EC systems use more randomness in searching for improvements than natural evolution. Despite the limitations in recognizing these phenomena in biology,

we should explore methods to reduce randomness in computation models to make evolving processes more "intelligent" and facilitate their finding of good solutions.

3.3 Epigenetic Mechanism

Epigenetics has become a new research direction in evolutionary biology [70]. Literally, "epi"-genetic control lies in the regulation of gene expression *without* changing the DNA sequence itself, so it is "beyond the conventional genetic" control. Epigenetic regulation arises during the processes of organism development and cell proliferation, triggered by intrinsic signals or environmental stimulations [72]. Epigenetic changes are heritable in the short term from cell generation to cell generation, and these stable alterations do not involve mutations on DNA sequences. Epigenetic regulation of DNA expression lies at the heart of many complex and long-term human diseases [17].

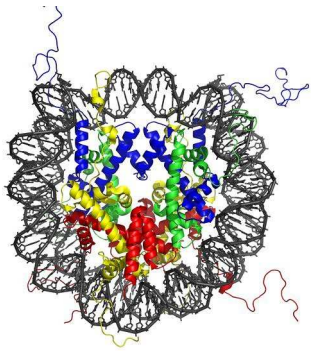

Figure 5: Nucleosome structure. A DNA sequence winds with four classes of histones marked by four colors here. Histone tails in a nucleosome can be attached by different groups, which will lead to different chromatin compositions. (Adapted from Vidal [2].)

Previous research in genetics mostly focused on the sequential information carried by DNA. However, DNA sequences are coiled up in cells in intimate complexes with the help of so called histone proteins. Figure 5 depicts a DNA sequence wrapped with histones that comprise a *nucleosome*. *Chromatin* is the complex of nucleosomes in the nucleus of cells

which participates in the control process of gene expression. The chromatin composition varies according to cell type and response to internal and external signals. The different composition of chromatin may affect expression and thus change the produced proteins even in the absences [5].

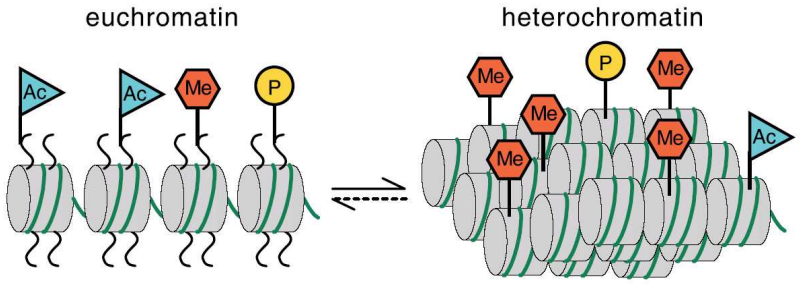

Figure 6: Euchromatin and Heterochromatin. Histone tails have three types of modification including acetylation (Ac), phosphorylation (p), and methylation (Me). Euchromatin (left) is the loosely packed state that most histone tails are attached by acetyl groups. Heterochromatin (right) is the tightly packed state that most histone tails are attached by methyl groups. (Adapted from Jenuwein and Allis [73].)

The main mechanisms of epigenetic control are DNA methylation and histone modification [72]. Modifications to chromatin, either on the DNA sequence itself (*DNA methylation*) or on its surrounding proteins (*histone modification*), affect gene expression and can be inherited from cell generation to cell generation during cell division. DNA methylation is a chemical addition to DNA sequences. Genes with methyl marks are repressed in expression, despite unchanged DNA content [162]. In histone modification, the tails of histone proteins are modified by different molecular attachments, e.g., acetyl, phosphoryl, and methyl groups (see Figure 6). If acetyl groups are attached to the histone tails of a chromatin, it will be loosely packed, a state called *euchromatin*. In euchromatin, DNA is readable and can be expressed in RNA and later translated into proteins. In contrast, if methyl groups are attached to histone tails, chromatin is tightly compressed, a state called *heterochromatin*. In the heterochromatin state, genes are inaccessible to the transcriptional machinery such as RNA polymerase or to transcription factors, and genes are prevented from being transcribed [58].

Other recognized mechanisms responsible for epigenetic regulation of gene expression include chromatin remodeling, histone variant composition, and non-coding RNA regulation. A discussion of these mechanisms can be found in Allis et al. [4].

The key feature of epigenetic mechanisms is their ability to coordinate internal and environmental signals can collaborate for modification of protein production [72]. The underlying interactions involve various molecules, such as DNA, RNA, and protein, but the extensive feedback between these molecules is still beyond our current understanding.

Epigenetics indeed opens up a new field in evolvability studies both in biology and EC. Sophisticated epigenetic feedback networks suggest a new structure for EC compared to the linear computing flow usually employed in the literature. For instance, in some dynamic optimization problems, not all genes responsible for different subfunctions need to be expressed all the time. We anticipate that a "controller switch" can be integrated into the genotype allowing short-term changes, where fragments of the genome can be turned on and off in response to feedback from outside. Such a mechanism for repression of expression has barely been used in computation. Similar multi-layer adaptive encoding schemes have been proposed, e.g., the *messy Genetic Algorithm* (mGA) [55] that combines short building blocks to form variable-length chromosomes to increasingly cover all features of a problem, or *diploid Genetic Algorithm*, e.g., [154] using a two-chromosome representation to adapt phenotypic variation in dynamic environments. However, existing work has not embedded the organizational epigenetic control in algorithms that would allow significant flexibility in changing environments. We anticipate that epigenetic mechanisms will play a crucial role in increasing the evolvability of EC algorithms.

4 Stabilization Operations

There are two main stabilization operations in evolution: recombination and selection, which will be reviewed in this section.

4.1 Recombination

Recombination is a process that generates combinations of existing genetic material in contrast to mutation which creates new alleles. Recombination is regarded as an important force shaping genomes and phenotypes. Since some highly efficient and accurate computational methods can be used in biology, analysis of gene recombination has made much progress by way of comparing aligned genome sequences. These comparisons facilitate a better understanding of several aspects of genetic and evolutionary biology, notably genotypic and phenotypic variation and genome structures [125].

Recombination exchanges genetic material between two DNA sequences swapping strands between one or multiple crossover points. Recombination can occur on homologous or non-homologous sequences. The former is more prominent in research because it is more common and efficient in generating adaptation in nature. Generally, research on recombination focuses on prevalent eukaryotic organisms rather than prokaryotes, which do not have sex. Unequal crossover is fairly rare and may lead to duplication or loss of some genes (discussed in Section 3.1.2) and other results [146]. Combination events can take place between different gene sequences, as in *intergenic recombination*, or between alleles on the same gene sequence, as in *intragenic recombination* [125]. Despite various forms of recombination, their outcome is crossover at one or multiple points and a swapping of fragments of genetic sequences.

The rate of recombination can significantly affect the rate of adaptation. It is usually higher than the rate of mutation, which implies that recombination introduces much less lethality to an evolutionary population than mutation. Instead, it advances evolution remarkably by stabilizing adaptive traits from parents to offspring. By drawing a recombination map of the human genome, Kong et al. [83] discovered that recombination rates vary in different regions of the genome. This variation is due to such functional features as gene density, other gene properties, and frequency of sequence repetitions. Recombination

rates are also different in autosomes between different sexes. Recombination contributes to producing both genotypic and phenotypic variation, and is able to repair DNA double strand breaks. *Sexual reproduction* is an important outcome of recombination.

In EC, recombination operations are considered an essential search strategy. Chromosome coding is much more flexible in computation than in nature, and thus, various recombination techniques have been proposed and studied, including double-parent and multi-parent crossover [37], fixed-length chromosome and variable-length chromosome crossover [55, 64], and homologous and non-homologous crossover [90, 115, 124]. High recombination rates are usually also adopted in computation because of its perceived efficiency in generating beneficial genetic and phenotypic variation. Elsewhere, adaptive recombination rates are proposed to strike a balance between exploration and exploitation [148]. In most of these adaptive recombination rate schemes, modification of recombination rates is based on fitness value. Different from natural recombination mechanisms, most adaptive recombination rate proposals simply react to the current search status, in order to escape from local optima. However, rate adaptation in biology is much more complex and suggests other models for computation. For instance, the rate may vary among different individuals or in different modules serving sub-functions in the genome. Such function-specific recombination rates could also consider the method of "compartmentalization" for modularity (Section 3.2.2). The notion of *epistatic clustering* in contributing to evolution of evolvability has recently been studied [122]. Genetic linkage patterns between different loci are claimed to affect recombination rates, and the simultaneous optimization of different recombination rates on different traits would be realized by a method called epistatic clustering. Evolvability would be improved through co-evolution of trait clustering and recombination mechanisms.

4.2 Selection

Although Darwin's theory of evolution being directed primarily by natural selection has been the subject of much argument, selection is an extremely important operation to stabilize the functional traits already generated by some exploratory operations [111]. Selection mechanisms are divided into two types by their effects on different stages of evolution. First, positive selection, i.e., Darwinian selection, enhances the fixation of advantageous alleles thus improving the diversity in early stages of evolution [171, 172]. Second, negative selection, also known as stabilizing selection or purifying selection, occurs at later stages of evolution when genetic diversity decreases when such selection eliminates deleterious alleles and only stabilizes specific traits [113]. The balance between selection and diversity of an evolutionary population has been a critical problem, and the dynamic pressure and some consequences of selection are still under active investigation. In general, selection pressure is produced by two factors, the environment and the mating competition, both of which will be discussed next.

4.2.1 Environmental Selection

Environmental selection comes from external surroundings enforcing the adaptivity of organisms to survive. Since Darwin environmental selection has received extensive research in evolutionary biology. Natural selection is an extremely important driving force for adaptive evolution in natural populations [67].

The first response of a living organisms to a changing environment is *somatic adaptation*. A simple example of somatic adaptation is human temperature compensation [82]. When the external temperature increases above normal, humans will sweat to adapt to this new environment. Shivering will happen if temperature falls below a normal value. Somatic adaptation happens directly as an organismal reaction to a changing environment, and is not fixed in morphological structures as an evolutionary change unless some deeper adapta-

tions caused by somatic changes can increase the survivability of an organism. Organisms have plenty of latent traits within their somatic adaptability, so they have a fairly high tolerance to changes in their environment [82]. Somatic adaptation mechanisms can only adjust existing functions to external changes. However, if evolution acts for a long time under environmental selection, changes can be stabilized by mutations in the germ-line after somatic adaption has tested them through promotion of survivability of the organism.

Selection can act at different levels depending on its targets [81]. These might be individual selection, individual-and-clade selection, or clade selection. At the individual level, the selection process has the fewest constraints since it directly affects phenotypic function fitness, and fewer mutation changes are required for a new adaptive trait. An individual can also interact with others in a clade, such as through recombination, and survive under selective pressure as a member in this clade. At the highest level, selection can happen on the level of an entire clade given large environmental impact, and the entire clade can, as a whole, escape from extinction. Some small groups of the lineage might go extinct, but the entire will be able to survive.

Interactions between different species may also cause environmental changes. Phillips and Shine [123] report an interesting phenomenon on species invasion. Toxic cane toads induced morphological changes among a species of snakes in Australia. Generally, native natural ecosystems can be devastated by the invasion of new species. At the beginning of the arrival of such an invasive species, the number of native organisms may decrease. As these native organisms adapt towards the invaders, the impact of the invasion declines and a new balance is achieved. Morphological changes are fixed subsequently. Complex natural ecosystems possess communities with highly frequent and dense interactions between species, as well as between species-specific functional traits within a species.

Although environmental selection is widely accepted as contributing significantly to natural evolution, selection has not been the mainstream of studies in evolvability. As a

potential to generate adaptation, evolvability is difficult to observe and to select for. However, there is increasing research arguing that evolvability is selectable and environmental selection can improve the evolution of evolvability. In the real world, the environment is changing constantly and fixes beneficial mutations, and there is a growing acceptance that a changing environment is a key ingredient to studying evolvability. Selection pressure is a critical operator to control an evolutionary process. Earl and Deem [36] suggest that selection pressure is increasingly strong when the environment becomes uncertain. Dramatic environmental changes lead to selection for better evolvability. They consider evolvability as a selectable trait, and facilitating environmental changes can be a method to accelerate evolution. A recent simulation by Kashtan et al. [76] in a biologically realistic setting also suggests that varying environments may accelerate natural evolution. In their work, different scenarios of temporarily changing optima were used. Kashtan and Alon [75] report that a goal that varies in a modular way can speed up evolution. Other work [30] takes into account the effect of the rate of environmental change. By observing the dynamics of *adaptive walks* under varying speed scenarios, they find that environments with varying rates of change have noticeably different effects on the fixation of beneficial mutations, the substitution time required, and the final phenotypic variation.

In EC, selection strategies are considered affecting search capability significantly during an evolutionary process. Different selection strategies have been proposed and the dynamics of selection pressure has been studied extensively [23, 53]. Since the effects of environmental selection on the evolution of evolvability have been recognized, further research on the dynamics of selection is required. Moreover, somatic adaptation should be considered when applying selection. Group-based selection methods should also be studied for varying selection pressure, so that a balance between the development of a minority and of the entire population can be dynamically achieved.

4.2.2 Sexual Selection

Sexual selection was proposed by Darwin as the pressure away from the possibility of mating failure. Two forms of sexual selection pressure are met by mature high-level animals: the battle between male individuals who fight, and the competition through mating choice made by females. Fisher [41] proposed a *runaway process*, where a male trait and female preferences for it can both evolve dramatically over time until finally checked by severe counter-selection. In modern biology, scientists pay much attention to these sex-based competitions that can generate and evolve several kinds of traits in high-level organisms. For instance, Kirkpatrick and Ravigne [80] find that some secondary sexual characteristics among individuals of the same sex can trigger rapid speciation.

Sexual selection happens at the inter-species level and affects reproductive fitness of individuals. Reproductive fitness is the probability of successfully generating offspring. Sexual selection has two main forms: intra-sexual selection and inter-sexual selection. Intra-sexual selection is known as the combat between competitive male individuals, and usually occurs in the form of fight. Inter-sexual selection is based on the choice made by the opposite sex. Male secondary sexual characteristics and female mating preferences can affect each other and evolve cooperatively [89]. This joint selection pressure, combined with natural selection, is a powerful force for rapid evolution.

Recent research in biology has connected sexual selection to the acceleration of evolution. Colegrave [29] finds that the rate of adaptation can be increased by *sex mechanisms* because sexual selection allows a rapid adaptive response under changing conditions by fixing beneficial mutations. Swanson and Vacquier [151] observe that rapid evolution emerges in reproductive proteins. This rapid evolution is forced by three main selective factors: sperm competition, sexual selection, and sexual conflict. Sperm competition is quite fierce in that each sperm will compete with billions of others to fuse with the only egg, and this competition exists in multiple steps for the sperm. Sexual selection happens when different

eggs have variant affinities for a special allele of a sperm-surface protein, and only the egg with the highest affinity is most likely to bind to this sperm. Sexual conflict means only one egg can be fused with the sperm to avoid polyspermy such that only one embryo is fertilized. These types of mechanisms add considerable selection pressure to reproductive proteins, and thus trigger rapid evolution in certain regions of these proteins.

The concept of *mating choice* was already applied in EC decades ago by Miller [106, 105]. Some co-evolutionary algorithms have been proposed to simulate mechanisms from sexual selection by constructing subgroups which can affect each other cooperatively to evolve in parallel. As more and more knowledge has been accumulated by biologists on the complex process of sexual selection, especially on the advantages that sex mechanisms contribute to the acceleration of speciation and evolution, this should be better incorporated in EC.

5 Quality Differentiation

As discussed previously, exploration and stabilization processes generate variation and adaptation at different levels. In contrast, the process of quality differentiation allows to quantify this adaptation and to distinguish between individuals. Two aspects are involved in quality differentiation, genotype-phenotype mapping and fitness evaluation. Genotype-phenotype mapping translates genetic information into visible functional phenotypes, and fitness evaluation measures the adaptation of variant individuals based on its ability to survive.

5.1 Genotype-Phenotype Mapping

In EC, mapping from genotype to phenotype is an encoding process, especially in evolutionary algorithms and evolutionary strategies, where the mapping mechanism is simply to calculate a fitness function of each individual. However, in nature, the mapping process is much more complex, typically from the highly conserved genotypic information to greatly divergent polymorphism in phenotypes. The fundamental process in biological genotype-

phenotype mapping is gene expression, and the most important mechanism in this process is regulation of gene expression, which will be discussed first here. Since research on transcriptional regulation has discovered increasing evidence that RNA plays an important role in gene expression, the transcriptome, i.e., the set of all transcribed RNAs, will be reviewed next.

5.1.1 Regulation of Gene Expression

In biology, the core processes (Section 3.2.1) in organisms are responsible for generating anatomy and behavior using genetic and cell materials. These core processes include metabolism, gene expression, and interaction among molecules and cells [82], which are well conserved but still under exploration. Regulation of gene expression is the most important mechanism among the core processes to facilitate organismal novelties in evolution. Kirschner and Gerhart highlight the characteristics of "conservation" and "economy" in regulatory core processes in [82].

Scientists have been trying to understand the process of gene expression for decades. In 1956, Crick proposed the *Central Dogma* of molecular biology, as shown in Figure 7, which describes the transmission of genetic information from DNA to protein. The circular arrow around DNA means that a DNA is a template for self-replication. The arrow from DNA to RNA indicates that an RNA is transcribed on a DNA template, and the arrow from RNA to protein signifies that a protein is translated on an RNA template.

However, later biological research revealed that the process of gene expression is much more complex than such a linear flow, and involves a considerable number of complex regulation operations. The Central Dogma was challenged by discoveries of proteins playing an important role in regulation of gene expression, and most recently, the non-coding RNA control of chromosome architecture proposed by Mattick [101]. In this section, we concentrate on gene expression regulation by proteins and will discuss RNA effects in next

section.

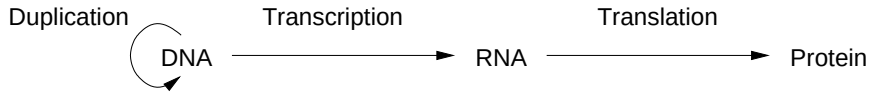

Figure 7: Central Dogma. The Central Dogma of biology by Crick holds that genetic information normally flows from DNA to RNA to protein, which involves the mechanisms of gene replication, transcription and translate.

Recall the discussion of genome redundancy in Section 2.1.1. Coding regions on genetic sequences that can be expressed into proteins only occupy a small portion of the entire genome in eukaryotic cells. This discovery indicated that a huge number of regulatory elements exist in genomes that participate in generating adaptation in evolution according to changes in environments. Although living systems have evolved for billions of years, regulatory core processes in various organisms have remained mostly unchanged despite species divergence. By comparison of related species from the same ancestors, such as humans and chimpanzees, at both the molecular and organismal levels King and Wilson [79] had already found in 1975 that genetic structures in these two species are almost the same; while at the organismal level, the anatomy, physiology, behavior and ecology of these two species are significantly different. This suggested to them that the complex adaptive evolution lies is produced by a combination and multiple utilization of similar, highly conserved genetic components under the control of regulatory systems.

A key step in the regulation of gene expression is transcription. Studies there are concentrated on two primary components: *promotors* and *transcription factors*. Promotors, also known as cis-regulatory sequences, are responsible for regulatory transcription. Cis-regulatory sequences are a part of non-coding DNA sequences, and they can determine the target genes and the length of the loci that will be transcribed under which conditions. Transcription factors are proteins interacting with these cis-regulatory sequences by binding to certain sites on DNA sequences. Interested readers are referred to Wray et al. [165]

for more details. Transcription factors act either as activators or as repressors of gene expression. For example, if a transcription factor A binds to a site on a DNA sequence that is responsible for generating protein B, then this factor A is regarded as a repressor to protein B. In addition, as a protein itself, factor A also has its template gene sequence. If another transcription factor C can bind to this site and represses the generation of protein A, C acts as a repressor to A but in turn as an activator to the expression of protein B. These activators and repressors can work together as a network of logic control. Promotors usually contain a number of binding sites for transcription factors, where each site can only be occupied by one factor at a time. These binding sites occupy, however, only a small fraction of sequences, and are distributed unevenly. Some binding sites of different functions can overlap. Furthermore, binding affinities of different materials are important for regulation as well. On the other hand, most transcription factors have numerous target genes and use priorities in binding with any of them [165]. This sophisticated network endows the regulation system with high robustness and plasticity necessary for evolution of capabilities of organisms.

Evolution of cis-regulatory sequences as non-coding sequences is considerably different from that of protein-coding sequences, and is less understood. King and Wilson [79] suggested that protein-coding sequences are highly conserved during evolution since they were synthesized. It is mutations on promoters that causes most morphological variation. Research on the evolution of transcriptional regulation has become mainstream in molecular biology in recent years [165]. In particular, Roderiguez-trelles et al. [131] find that significant substitution rate differences exist among different promotors, and even some neighboring cis-regulatory promotors involved in the same regulatory network can have different evolution rates. Moreover, Stone and Wray [150] propose that local point mutations on binding sites can lead to rapid evolution in gene expression, which indicates their potential of accelerating evolution. Wagner [155] points out that other simple changes such as gene

duplication and deletion of promotors can also result in rapid evolution in gene regulatory networks. By comparing genomes, Fondon and Garner [44] discovered that gene-associated tandem repeat expansions and contractions exist and give rise to rapid morphological evolution. In their experimental research, a tandem repeat mutation shows both elevated purity and intensive length polymorphism among different dog breeds. Mutations on non-coding sequences can modify regulation of the target genes, the length of coding loci to transcribe, and the occurrence conditions. Furthermore, they also result in morphological variation and accelerated phenotypic evolution.

Since the mechanisms of regulation of gene expression can well explain many phenomena in evolvability and rapid evolution in living systems, research on artificial regulatory networks has also arisen in computer science. Several models of artificial evolution regulatory networks have been proposed such as Banzhaf et al. [12, 13, 15, 88], Chavoya and Duthen [27], Mattiussi and Floreano [103], Nehaniv [109], etc. These artificial models intend to generate regulatory behavior akin to that of natural systems. However, these research efforts are still in the early stage, and more work on evolvability and dynamics in artificial regulatory networks is necessary.

5.1.2 The Transcriptome

The transcriptome, or collection of transcripts, refers to all RNAs produced in a single or a group of cells, working as an intermediate component of gene expression. In high-level eukaryotes such as humans, most regions of the transcriptome are not translated into protein. What necessitates the existence of such a large number of RNAs in the transcriptome of high-level eukaryotes? Regulatory functions is one answer to this question. Although regulation of gene expression starts with the transcription step, these transcribed but non translated sequences or non-coding RNA sequences act as regulators for translation in gene expression, and are currently attracting increased interest in biological research [48,

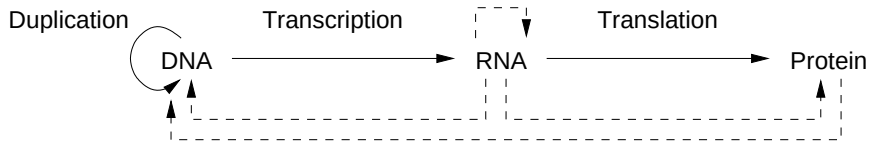

Figure 8: Eukaryotic genetic system. Expression of genes is not with an irreversible linear flow in eukaryotes, but involves frequent feedback and interactions among different molecules including DNA, RNA, and protein, as the dotted lines shown here.

102].

An RNA is not just a temporary medium between genes and proteins as described in the Central Dogma. In high-level eukaryotes, the information transmission from DNA to protein is not a one-way process, but involves many functionalities of the transcriptome. The new perspective of gene expression proposed by Mattick [101, 102] can be described in Figure 8. Compared to a prokaryotic genetic system, an eukaryotic system has a parallel control mechanism with multiple outputs and information transfers. Rather than a simple medium of gene expression, RNA metabolism and interaction have been discovered playing an important role in gene expression regulation.

Mattick [100] proposes that non-coding RNAs participate extensively in gene expression regulation, being present in about 98% of all transcriptional outputs in eukaryotes. In the research on human transcriptome, Frith et al. [48] found that non-coding RNAs play an important role in generating phenotypic variation. Non-coding RNAs are classified into two categories: introns and other non-coding RNAs.

Regulation of the transcriptome shows contributions to evolvability and rapid evolution. Introns, an important category of non-coding RNAs, are found more susceptible to mutations than their neighboring protein-coding exons. Rather than having no function, as thought previously, the fewer constraints on introns offer flexibility to generate new functions and rapid protein sequence evolution during the process of regulation, especially in connection with alternative splicing. The evolution of RNA communication networks may also accelerate the evolution of gene expression, as observed by Mattick [100]. These RNA

communication networks, which describe interaction among different layers of RNA signaling, provide a sophisticated regulatory architecture, enabling DNA-DNA, DNA-RNA, or RNA-RNA communication, DNA methylation, chromatin generation, and RNA translation.

Compared to natural systems, the genotype-phenotype mapping in EC is rather primitive still and a transcriptome is mostly missing in algorithms. The complex RNA parallel information transfer framework inspires various applications. Based on what computational models have already achieved with artificial regulatory networks, more mechanisms should be implemented, especially the newly discovered powerful mechanisms of transcriptome regulation.

5.2 Fitness Evaluation

Fitness evaluation measures behavior or function of individuals or species. In nature, fitness of an individual or species is implicit and subject to natural selection, whereas in EC, fitness is mostly based on numerical values of an individual as solution to a given problem, and this fitness is explicit.

In nature, adaptable species survive through different challenges, and less fit species may become extinct during evolution. Adaptability lies not only in the currently existing adaptivity to the environment, but also in the capability to generate more adapted offspring, also known as fertility. Such adaptability enables organisms to survive despite the pressure of environmental selection. Some factors are included in fertility characteristics, such as age and sexuality [82]. The age indicates how long an individual has been in an evolutionary system and, usually, a greater age means relatively higher adaptability but lower fertility. An individual dies at a certain age in a natural evolution population, and this is a way to keep diversity and to make an evolutionary system evolve continually. Sexuality varies under sexual selection, as discussed in Section 4.2.2, and mostly depends on mating preferences

of female individuals. An individual with stronger secondary sexual characteristics is more likely to be chosen by the opposite sex to generate offspring.

The above implicit fitness in natural organisms emphasizes evolvability under intricate pressures from interactions among evolutionary components, internal or external to these organisms during a long, continuing evolutionary process [122]. In reality, the fitness of individuals in a system can vary a great deal. Moreover, a large-scale quality differentiation exists in almost every natural evolutionary system, and these vastly diverse evolution systems exhibit great evolvability. Since selection and evaluation act directly on observable phenotypic functions but evolvability only provides the potential for better functions, selection and evaluation for evolvability are not observable directly. Reisinger and Miikkulainen [129] propose an evolvable representation and an evaluation strategy to exert indirect selection pressure on evolvability. In their work, a systematically changing fitness function is adopted according to a special evolvable representation that can reflect efficiently how genetic changes restructure phenotypic variation. Thus, evolvability can be evaluated through how such a systematical structure can expand in phenotypes.

As the method of EC has been widely applied in many areas of industry and academia, fitness evaluation arises as a difficult problem because it is usually very CPU-intensive. In the current literature, two main methods of fitness evaluation are employed, absolute fitness and relative fitness. Absolute fitness of each individual usually refers to its value of a specified fitness function. Relative fitness compares different individuals and gives a rank to each individual to produce a record of winners. This methods is good at suppressing exceptionally good individuals, thus, helping an evolutionary system to escape from premature convergence. In fact, evaluating the fitness of each individual is usually difficult for many optimization problems in the real world because explicit fitness can be hard to define and expensive to calculate. As a result, a *fitness approximation* has been proposed with differing levels of approximation, including "problem approximation", "functional ap-

proximation" and "evolutionary approximation". Jin [74] has surveyed these approaches. They are sensitive to training data and to varying constraints of different models, so a common framework would be required. Nevertheless, it is a good starting point to simulate the implicit adaptive fitness evaluation from nature, a method that has good prospects for detecting or evolvability in EC.

6 Conclusion

Since Darwin proposed his theory of natural evolution based on heritable variation and natural selection, numerous research efforts have been dedicated on this subject. In modern biology, many details about mechanisms of evolution and factors that can affect evolution have been revealed. Besides understanding the history of evolution, biologists are currently paying more attention to the capability of organisms to evolve and to the evolution of such capability in the open-ended natural evolutionary process. Varying evolution rates among different species or different regions of genetic material in an organism attract research interest for acceleration of evolution. Meanwhile, in research of artificial evolutionary systems, one is also working on improving the capability of computational models of evolution by studying more intelligent and adaptive mechanisms.

Evolvability, as the capability to generate adaptation by producing fitter offspring via evolutionary operations, has received considerable interest in recent research in both biology and EC. Substantial work has been done on this topic in both disciplines, and many factors are found to contribute to evolvability. After some phenomena of rapid evolution have been reported, acceleration of evolution has also become an important research topic. Evolvability indicates the *potential* of a system to evolve, and it can well be a long-term method to accelerate evolution. However, in the short term, concentrating on evolvability may slow down the rate of evolution, since some individuals or groups which are more capable to evolve do not show sufficient fitness improvement. Consequently, rather than

improving immediate fitness, evolvability concentrates on significantly longer, or even open-ended evolution, especially under changing environments. Evolution should be accelerated in the long term with such capability. Therefore, evolvability and acceleration of evolution are interrelated and crucial aspects in biology and EC.

In this review, we started from notions and new discoveries in biology, including aspects that have been recognized to improve evolvability or to accelerate the evolution of natural systems. The order in the presentation of these ideas conforms to the flow of an evolutionary algorithm. In each part of the flow, we first reviewed relevant results in biology. Next, a brief survey was provided to describe current research status in EC, followed by an outlook for further research. Our goal was to describe and present new research outcomes to computer scientists, especially in the field of EC. Since it is accepted that artificial evolutionary systems are much less evolvable than natural systems, we hope these ideas can inspire new methods and applications in EC.

References

[1] Annual "HUMIES" Awards for Human-Competitive Results Produced by Genetic and Evolutionary Computation held by GECCO: Genetic and Evolutionary Computation Conference since 2004. http://www.human-competitive.org. ACM, New York, NY, USA.

[2] Christopher Vidal personal website. Accessed on 2008-08-30. http://www.christopher-vidal.com/thegenome.htm.

[3] B. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts, and J. D. Watson. *Molecular Biology of the Cell (4th edition)*. Garland, New York, NY, USA, 2002.

[4] C. D. Allis, T. Jenuwein, and D. Reinberg, editors. *Epigenetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA, 2007.

[5] C. D. Allis, T. Jenuwein, and D. Reinberg. Overview and concepts. In *Epigenetics*, pages 23–61. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA, 2007.

[6] L. Altenberg. The evolution of evolvability in genetic programming. In *Advances in Genetic Programming*, pages 47–74. MIT Press, Cambridge, MA, USA, 1994.

[7] L. Altenberg. The schema theorem and price's theorem. In *Fundations of Genetic Algorithms 3*, pages 23–49. MIT Press, 1995.

[8] J. Arabas, Z. Michalewicz, and J. Mulawka. GAVaPS – a genetic algorithm with varying population size. In *Proceedings of the 1st IEEE Conference on Evolutionary Computation*, pages 73–78. IEEE Press, 1994.

[9] S. Asthana, W. S. Noble, G. Kryukov, C. E. Grant, S. Sunyaev, and J. A. Stamatoyannopoulos. Widely distributed noncoding purifying selection in the human genome. *Proceedings of the National Academy of Sciences*, 104(30):12410–12415, 2007.

[10] T. Back, A. E. Eiben, and N. A. L. van der Vaart. An empirical study on gas "without parameters". In *Proceedings of the 6th International Conference on Parallel Problem Solving from Nature (PPSN VI)*, volume 1917 of *Lecture Notes in Computer Science*, pages 315–324. Springer, 2000.

[11] J. A. Bailey, Z. Gu, R. A. Clark, K. Reinert, R. V. Samonte, S. Schwartz, M. D. Adams, E. W. Myers, P. W. Li, and E. E. Eichler. Recent segmental duplications in the human genome. *Science*, 297(5583):1003–1007, 2002.

[12] W. Banzhaf. Artificial regulatory networks and genetic programming. In *Genetic Programming Theory and Practice*, pages 43–62. Kluwer Academic, Norwell, MA, USA, 2003.

[13] W. Banzhaf. On the dynamics of an artificial regulatory networks. In *Advances in Artificial Life – Proceedings of the 7th European Conference on Artificial Life*, volume 2801 of *Lecture Notes in Artificial Intelligence*, pages 217–227. Springer-Verlag, 2003.

[14] W. Banzhaf, G. Beslon, S. Christensen, J. A. Foster, F. Kepes, V. Lefort, J. F. Miller, M. Radman, and J. J. Ramsden. From artificial evolution to computational evolution: A research agenda. *Nature Reviews Genetics*, 7(9):729–735, 2006.

[15] W. Banzhaf and P. Kuo. Network motifs in natural and artificial transcriptional regulatory networks. *Journal of Biological Physics and Chemistry*, 4(2):85–92, 2004.

[16] W. Banzhaf, P. Nordin, R. E. Keller, and F. D. Francone. *Genetic Programming – An Introduction: On the Automatic Evolution of Computer Programs and its Applications*. Morgan Kaufmann, San Francisco, CA, USA, 1998.

[17] S. B. Baylin. DNA hypermethylation in tumorigenesis. *Trends in Genetics*, 16(4):168–174, 2000.

[18] M. Beckerman. *Molecular and Cellular Signaling*. Springer, New York, NY, USA, 2005.

[19] M. A. Bedau. Quantifying the extent and intensity of adaptive evolution. In *GECCO'99: Proceedings of the 1999 Genetic and Evolution Computation Conference Workshop Program*, pages 34–37. Morgan Kaufmann, 1999.

[20] M. A. Bedau and N. H. Packard. Measurement of evolutionary activity, teleology, and life. In *Artificial Life II*, pages 431–461. Addison-Wesley, Redwood City, CA, USA, 1992.

[21] M. A. Bedau and N. H. Packard. Evolution of evolvability via adaptation of mutation rates. *BioSystems*, 69(2):143–162, 2003.

[22] M. A. Bedau and R. Seymour. Adaptation of mutation rates in a simple model of evolution. In *Complex Systems: Mechanism of Adaptation*, pages 179–183. IOS Press, Amsterdam, The Netherlands, 1994.

[23] T. Blickle and L. Thiele. A comparison of selection schemes used in evolutionary algorithms. *Evolutionary Computation*, 4(4):361–394, 1997.

[24] J. T. Bonner. *The evolution of complexity by means of natural selection*. Princeton University Press, Princeton, NJ, USA, 1988.

[25] R. Calabretta, S. Nolfi, and G. P. Wagner. Duplication of modules facilitates the evolution of functional specialization. *Artificial Life*, 6(1):69–84, 2000.

[26] S. Carroll, J. Grenier, and S. Weatherbee. *From DNA to Diversity*. Blackwell Publishing, 2005.

[27] A. Chavoya and Y. Duthen. Evolving an artificial regulatory network for 2d cell patterning. In *ALIFE'07: IEEE Symposium on Artificial Life 2007*, pages 47–53. IEEE Press, 2007.

[28] J. Cohen. Bioinformatics-an introduction for computer scientists. *ACM Computing Surveys*, 36(2):122–158, 2004.

[29] N. Colegrave. Sex releases the speed limit on evolution. *Nature*, 420(6916):664–666, 2002.

[30] S. Collins, J. de Meaux, and C. Acquisti. Adaptive walks toward a moving optimum. *Genetics*, 176(2):1089–1099, 2007.

[31] M. Conrad. The geometry of evolution. *BioSystems*, 24(11):61–81, 1990.

[32] C. Darwin. *On the Origin of Species by means of natural selection.* John Murray, London, 1859.

[33] R. Dawkins. The evolution of evolvability. In *Artificial Life: The Quest for a New Creation*, pages 201–220. Addison-Wesley, Reading, MA, USA, 1989.

[34] J. A. G. M. de Visser, J. Hermisson, G. P. Wagner, L. A. Meyers, H. Bagheri-Chaichian, J. L. Blanchard, L. Chao, J. M. Cheverud, S. F. Elena, W. Fontana, G. Gibson, T. F. Hansen, D. Krakauer, R. C. Lewontin, C. Ofria, S. H. Rice, G. von Dassow, A. Wagner, and M. C. Whitlock. Perspective: Evolution and detection of genetic robustness. *Evolution*, 57(9):1959–1972, 2003.

[35] D. Dykhuizen and D. L. Hartl. Selective neutrality of 6pgd allozymes in e. coli and the effects of genetic background. *Genetics*, 96(4):801–817, 1980.

[36] D. J. Earl and M. W. Deem. Evolvability is a selectable trait. *Proceedings of the National Academy of Sciences*, 101(32):11531–11536, 2004.

[37] A. E. Eiben. Multiparent recombination in evolutionary computing. In *Advances in evolutionary computing: theory and applications*, pages 175–192. Springer-Verlag New York, Inc., New York, NY, USA, 2003.

[38] A. E. Eiben, E. Marchiori, and V. A. Valkó. Evolutionary algorithms with on-the-fly population size adjustment. In *Proceedings of the 8th International Conference on Parallel Problem Solving from Nature (PPSN VIII)*, volume 3242 of *Lecture Notes in Computer Science*, pages 41–50. Springer, 2004.

[39] N. Eldredge and S. J. Gould. Punctuated Equilibria: An alternative to phyletic gradualism. In *Models in Paleobiology*, pages 82–115. Freeman Cooper and Co., San Francisco, CA, USA, 1972.

[40] S. F. Elena, C. O. Wilke, C. Ofria, and R. E. Lenski. Effects of population size and mutation rate on the evolution of mutational robustness. *Evolution*, 61(3):666–674, 2007.

[41] R. A. Fisher. *Genetical Theory of Natural Selection*. Clarendon, Oxford, UK, 1930.

[42] D. B. Fogel. *System Identification through Simulated Evolution: A Machine Learning Approach to Modeling*. Ginn Press, Needham Heights, MA, USA, 1991.

[43] L. J. Fogel, A. J. Owens, and M. J. Walsh. *Artificial Intelligence through Simulated Evolution*. John Wiley and Sons, New York, NY, USA, 1966.

[44] J. W. Fondon III and H. R. Garner. Molecular origins of rapid and continuous morphological evolution. *Proceedings of the National Academy of Sciences*, 101(52):18058–18063, 2004.

[45] G. Forgacs and S. Newman. *Biological Physics of the Developing Embryo*. Cambridge University Press, 2005.

[46] D. Fredman, S. J. White, S. Potter, E. E. Eichler, J. T. D. Dunnen, and A. J. Brookes. Complex snp-related sequence variation in segmental genome duplications. *Nature Genetics*, 36(8):861–866, 2004.

[47] J. L. Freeman, G. H. Perry, L. Feuk, R. Redon, S. A. McCarroll, D. M. Altshuler, H. Aburatani, K. W. Jones, C. Tyler-Smith, M. E. Hurles, N. P. Carter, S. W. Scherer, and C. Lee. Copy number variation: New insights in genome diversity. *Genome Research*, 16(8):949–961, 2006.

[48] M. C. Frith, M. Pheasant, and J. S. Mattick. The amazing complexity of the human transcriptome. *European Journal of Human Genetics*, 13(8):894–897, 2005.

[49] J. H. Gillespie. The role of population size in molecular evolution. *Theoretical Population Biology*, 55(2):145–156, 1999.

[50] J. H. Gillespie. Is the population size of a species relevant to its evolution? *Evolution*, 55(11):2161–2169, 2001.

[51] D. E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley, Boston, MA, USA, 1989.

[52] D. E. Goldberg. Sizing populations for serial and parallel genetic algorithms. In *Proceedings of the 3rd International Conference on Genetic Algorithms*, pages 70–79. Morgan Kaufmann, 1989.

[53] D. E. Goldberg and K. Deb. A comparative analysis of selection schemes used in genetic algorithms. In *Foundations of Genetic Algorithms 1*, pages 69–93. Morgan Kaufmann, 1991.

[54] D. E. Goldberg, K. Deb, and J. H. Clark. Genetic algorithms, noise, and the sizing of populations. *Complex Systems*, 6(4):333–362, 1992.

[55] D. E. Goldberg, B. Korb, and K. Deb. Messy genetic algorithms: Motivation, analysis, and first results. *Complex Systems*, 3(5):493–530, 1989.

[56] S. Gould. *The Structure of Evolutionary Theory*. Harvard University Press, 2002.

[57] S. Gould and E. Vrba. Exaptation: A missing term in the science of form. *Paleobiology*, 8(1):4–15, 1982.

[58] S. I. S. Grewal and D. Moazed. Heterochromatin and epigenetic control of gene expression. *Science*, 301(5634):798–802, 2003.

[59] F. Gruau. Genetic micro programming of neural networks. In *Advances in genetic programming*, pages 495–518. MIT Press, Cambridge, MA, USA, 1994.

[60] X. Gu, Z. Zhang, and W. Huang. Rapid evolution of expression and regulatory divergences after yeast gene duplication. *Proceedings of the National Academy of Sciences*, 102(3):707–712, 2005.

[61] J. B. S. Haldane. *The Causes of Evolution*. Harper and Row, New York, NY, USA, 1932.

[62] G. R. Harik and F. G. Lobo. A parameter-less genetic algorithm. In *GECCO'99: Proceedings of the Genetic and Evolutionary Computation Conference*, pages 258–267. Morgan Kaufmann, 1999.

[63] P. M. Harrison, H. Hegyi, S. Balasubramanian, N. M. Luscombe, P. Bertone, N. Echols, T. Johnson, and M. Gerstein. Molecular fossils in the human genome: Identification and analysis of the pseudogenes in chromosomes 21 and 22. *Genome Research*, 12(2):272–280, 2002.

[64] I. Harvey and I. Harvey. The saga cross: the mechanics of recombination for species with variablelength genotypes. In *Proceedings of the 2ed International Conference*

on Parallel Problem Solving from Nature (PPSN II), pages 269–278. North-Holland, 1992.

[65] J. Hawks, E. T. Wang, G. M. Cochran, H. C. Harpending, and R. K. Moyzis. Recent acceleration of human adaptive evolution. *Proceedings of the National Academy of Sciences*, 104(52):20753–20758, 2007.

[66] H. Hendrickson, E. S. Slechta, U. Bergthorsson, D. I. Andersson, and J. R. Roth. Amplification-mutagenesis: Evidence that "directed" adaptive mutation and general hypermutability result from growth with a selected gene amplification. *Proceedings of the National Academy of Sciences*, 99(4):2164–2169, 2002.

[67] A. P. Hendry. The power of natural selection. *Nature*, 433(7027):694–695, 2005.

[68] J. H. Holland. *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor, MI, USA, 1975.

[69] A. J. Iafrate, L. Feuk, M. N. Rivera, M. L. Listewnik, P. K. Donahoe, Y. Qi, and S. W. S. C. Lee. Detection of large-scale variation in the human genome. *Nature Genetics*, 36(9):949–951, 2004.

[70] E. Jablonka and M. J. Lamb. *Evolution in Four Dimensions: Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life*. MIT Press, Cambridge, MA, USA, 2005.

[71] F. Jacob. Evolution and tinkering. *Science*, 196(4295):1161–1166, 1977.

[72] R. Jaenisch and A. Bird. Epigenetic regulation of gene expression: How the genome integrates intrinsic and environmental signals. *Nature Genetics*, 33:245–254, 2003.

[73] T. Jenuwein and C. D. Allis. Translating the histone code. *Science*, 293(5532):1074–1080, 2001.

[74] Y. Jin. A comprehensive survey of fitness approximation in evolutionary computation. *Soft Computing - A Fusion of Foundations, Methodologies and Applications*, 9(1):3–12, 2005.

[75] N. Kashtan and U. Alon. Spontaneous evolution of modularity and network motifs. *Proceedings of the National Academy of Sciences*, 102(39):13773–13778, 2005.

[76] N. Kashtan, E. Noor, and U. Alon. Varying environments can speed up evolution. *Proceedings of the National Academy of Sciences*, 104(34):13713–13716, 2007.

[77] S. T. Kazadi, D. Lee, R. Modi, J. Sy, and W. Lue. Levels of compartmentalization in artificial evolution. In *GECCO'00: Proceedings of the Genetic and Evolutionary Computation Conference*, pages 841–848. Morgan Kaufmann, 2000.

[78] M. Kimura. Evolutionary rate at the molecular level. *Nature*, 217(5129):624–626, 1968.

[79] M. C. King and A. C. Wilson. Evolution at two levels in humans and chimpanzees. *Science*, 188(4184):107–116, 1975.

[80] M. Kirkpatrick and V. Ravigne. Speciation by natural and sexual selection: Models and experiments. *The American Naturalist*, 159(3):S22–S35, 2002.

[81] M. W. Kirschner and J. C. Gerhart. Evolvability. *Proceedings of the National Academy of Sciences*, 95(15):8420–8427, 1998.

[82] M. W. Kirschner and J. C. Gerhart. *The Plausibility of Life*. Yale University Press, New Haven, CT, USA, 2005.

[83] A. Kong, D. F. Gudbjartsson, J. Sainz, G. M. Jonsdottir, S. A. Gudjonsson, B. Richardsson, S. Sigurdardottir, J. Barnard, B. Hallbeck, G. Masson, A. Shlien,

S. T. Palsson, M. L. Frigge, T. E. Thorgeirsson, J. R. Gulcher, and K. Stefansson. A high-resolution recombination map of the human genome. *Nature Genetics*, 31(3):241–247, 2002.

[84] J. R. Koza. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. MIT Press, Cambridge, MA, USA, 1992.

[85] J. R. Koza. Gene duplication to enable genetic programming to concurrently evolve both the architecture and work-performing steps of a computer program. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, pages 695–717. Morgan Kaufmann, 1995.

[86] D. C. Krakauer and J. B. Plotkin. Redundancy, antiredundancy, and the robustness of genomes. *Proceedings of the National Academy of Sciences*, 99(3):1405–1409, 2002.

[87] E. Kugelberg, E. Kofoid, A. B. Reams, D. I. Andersson, and J. R. Roth. Multiple pathways of selected gene amplification during adaptive mutation. *Proceedings of the National Academy of Sciences*, 103(46):17319–17324, 2006.

[88] P. Kuo, W. Banzhaf, and A. Leier. Network topology and the evolution of dynamics in an artificial genetic regulatory network model created by whole genome duplication and divergence. *Biosystems*, 85(3):177–200, 2006.

[89] R. Lande. Models of speciation by sexual selection on polygenic traits. *Proceedings of the National Academy of Sciences*, 78(6):3721–3725, 1981.

[90] W. B. Langdon. Size fair and homologous tree genetic programming crossovers. *Genetic Programming and Evolvable Machines*, 1(1):95–119, 2000.

[91] W. B. Langdon, T. Soule, R. Poli, and J. A. Foster. The evolution of size and shape. In *Advances in Genetic Programming: Volume 3*, pages 163–190. MIT Press, Cambridge, MA, USA, 1999.

[92] J. E. G. Leigh. Levels of selection, potential conflicts, and their resolution: The role of the "common good". In *Levels of Selection in Evolution*, pages 15–30. Princeton University Press, Princeton, NJ, USA, 1999.

[93] W.-H. Li, C.-I. Wu, and C.-C. Luo. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Molecular Biology and Evolution*, 2(2):150–174, 1985.

[94] K. F. Liem. Key evolutioanry innovations, differential diversity, and symecomorphosis. In *Evolutioanry Innovations*, pages 147–170. University of Chicago Press, Chicago, IL, USA, 1990.

[95] F. G. Lobo and C. F. Lima. A review of adaptive population sizing schemes in genetic algorithms. In *GECCO'05: Proceedings of the 2005 workshops on Genetic and Evolutionary Computation*, pages 228–234. ACM, 2005.

[96] M. Z. Ludwig, N. H. Patel, and M. Kreitman. Functional analysis of eve stripe 2 enhancer evolution in Drosophila: Rules governing conservation and change. *Development*, 125(5):949–958, 1998.

[97] S. Luke. Code growth is not caused by introns. In *GECCO'00: Late Breaking Papers at the 2000 Genetic and Evolutionary Computation Conference*, pages 228–235. Morgan Kaufmann, 2000.

[98] S. C. Manrubia and C. Briones. Modular evolution and increase of functional complexity in replicating RNA molecules. *RNA*, 13(1):97–107, 2007.

[99] P. Marrow, M. Heath, and I. I. Re. Evolvability: Evolution, Computation, and Biology. In *GECCO'99: Proceedings of the 1999 Genetic and Evolutionary Computation Conference Workshop on Evolvability*, pages 30–33. Morgan Kaufmann, 1999.

[100] J. S. Mattick. Non-coding RNAs: the architects of eukaryotic complexity. *EMBO Reports*, 2(11):986–991, 2001.

[101] J. S. Mattick. Challenging the dogma: The hidden layer of non-protein-coding RNAs in complex organisms. *BioEssays*, 25(10):930–939, 2003.

[102] J. S. Mattick. RNA regulation: a new genetics? *Nature Reviews Genetics*, 5(4):316–323, 2004.

[103] C. Mattiussi and D. Floreano. Analog genetic encoding for the evolution of circuits and networks. *IEEE Transactions on Evolutionary Computation*, 11(5):596–607, 2007.

[104] E. Mayr and W. Provine, editors. *The Evolutionary Synthesis*. Harvard University Press, 1998, 2nd edition.

[105] G. F. Miller. Exploiting mate choice in evolutionary computation: Sexual selection as a process of search, optimization, and diversification. In *Selected Papers from AISB Workshop on Evolutionary Computing*, pages 65–79. Springer-Verlag, 1994.

[106] G. F. Miller. *The Mating Mind: How Sexual Choice Shaped Human Nature*. Anchor, Port Moody, BC, Canada, 2001.

[107] T. Miyata and T. Yasunaga. Molecular evolution of mRNA: A method for estimating evolutionary rates of synonymous and amino acid substitutions from homologous nucleotide sequences and its application. *Journal of Molecular Evolution*, 16(1):23–36, 1980.

[108] C. L. Nehaniv. Measuring evolvability as the rate of complexity increase. In *Proceedings of the Artificial Life VII Workshop*, pages 55–57. MIT Press, 2000.

[109] C. L. Nehaniv. Self-replication, evolvability and asynchronicity in stochastic worlds. In *Stochastic Algorithms: Foundations and Applications*, pages 126–169. Springer, New York, NY, USA, 2005.

[110] M. Nei. Gene duplication and nucleotide substitution in evolution. *Nature*, 221(5175):40–42, 1969.

[111] M. Nei. Selectionism and neutralism in molecular evolution. *Molecular Biology and Evolution*, 22(12):2318–2342, 2005.

[112] M. Nei and T. Gojobori. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution*, 3(5):418–426, 1986.

[113] M. Nei and A. P. Rooney. Concerted and birth-and-death evolution of multigene families. *Annual Review of Genetics*, 22(15):121–152, 2005.

[114] S. R. Neves and R. Iyengar. Modeling of signaling networks. *BioEssays*, 24(12):1110–1117, 2002.

[115] P. Nordin, W. Banzhaf, and F. Francone. Efficient evolution of machine code for cisc architectures using blocks and homologous crossover. In *Advances in Genetic Programming: Volume 3*, pages 275–299. MIT Press, Cambridge, MA, USA, 1999.

[116] P. Nordin, F. Francone, and W. Banzhaf. Explicitly defined introns and destructive crossover in genetic programming. In *Advances in Genetic Programming: Volume 2*, pages 111–134. MIT Press, Cambridge, MA, USA, 1996.

[117] S. Ohno. *Evolution by Gene Duplication*. Springer-Verlag, New York, NY, USA, 1970.

[118] T. Ohta. Population size and rate of evolution. *Journal of Molecular Evolution*, 1(4):305–314, 1972.

[119] T. Ohta. The nearly neutral theory of molecular evolution. *Annual Review of Ecology and Systematics*, 23:263–286, 1992.

[120] T. Ohta and M. Kimura. On the constancy of the evolutionary rate of cistrons. *Journal of Molecular Evolution*, 1(1):18–25, 1971.

[121] S. Palumbi. *The Evolution Explosion*. Norton, New York, 2002.

[122] J. W. Pepper. The evolution of evolvability in genetic linkage patterns. *BioSystems*, 69(2):115–126, 2003.

[123] B. L. Phillips and R. Shine. Adapting to an invasive species: Toxic cane toads induce morphological change in australian snakes. *Proceedings of the National Academy of Sciences*, 101(49):17150–17155, 2004.

[124] R. Poli and N. F. Mcphee. Exact schema theory for gp and variable-length gas with homologous crossover. In *GECCO'01: Proceedings of the Genetic and Evolutionary Computation Conference*, pages 104–111. Morgan Kaufmann, 2001.

[125] D. Posada, K. A. Crandall, and E. C. Holmes. Recombination in evolutionary genomics. *Annual Review of Genetics*, 36:75–97, 2002.

[126] R. A. Raff. *The Shape of Life: Genes, Development, and the Evolution of Animal Form*. University of Chicago Press, Chicago, IL, USA, 1996.

[127] M. J. Raven and M. A. Bedau. General framework for evolutionary activity. In *Proceedings of the Seventh European Conference on Artificial Life*, pages 676–685. Springer-Verlag, 2003.

[128] I. Rechenberg. *Evolution Strategy*. Frommann-Holzboog, Stuttgart, Germany, 1973.

[129] J. Reisinger and R. Miikkulainen. Selecting for evolvable representations. In *GECCO'06: Proceedings of the 2006 Genetic and Evolutionary Computation Conference*, pages 1297–1304. ACM, 2006.

[130] J. Reisinger, K. O. Stanley, and R. Miikkulainen. Towards an empirical measure of evolvability. In *GECCO'05: Proceedings of the 2005 Genetic and Evolutionary Computation Conference Workshop Program*, pages 257–264. ACM, 2005.

[131] F. Roderiguez-trelles, R. Tarrio, and F. J. Ayala. Evolution of cis-regulatory regions versus codifying regions. *International Journal of Developmental Biology*, 47(7):665–673, 2003.

[132] J. R. Roth and D. I. Andersson. Adaptive mutation: How growth under selection stimulates lac+ reversion by increasing target copy number. *Journal of Bacteriology*, 186(15):4855–4860, 2004.

[133] J. R. Roth, E. Kugelberg, A. B. Reams, E. Kofoid, and D. I. Andersson. Origin of mutations under selection: the adaptive mutation controversy. *Annual Review of Microbiol*, 60:477–501, 2006.

[134] F. Rothlauf and D. E. Goldberg. Redundant representations in evolutionary computation. *Evolutionary Computation*, 11(4):381–415, 2003.

[135] D. Sankoff. Rearrangements and chromosomal evolution. *Current Opinion in Genetic and Development*, 13(6):583–587, 2003.

[136] K. Schmitt. Using gene duplication and gene deletion in evolution strategies. Technical report, Collaborative Research Center 531, University of Dortmund, 2004.

[137] M. Schutz. Other operations: gene duplication and deletion. In *Handbook of Evolutionary Computation*, pages 8–15. Oxford University Press, Oxford, UK, 1997.

[138] H.-P. Schwefel. *Evolution and Optimum Seeking: The Sixth Generation*. John Wiley and Sons, New York, NY, USA, 1995.

[139] J. Sebat, B. Lakshmi, J. Troge, J. Alexander, J. Y. P. Lundin, S. Manr, H. Massa, M. Walker, M. Chi, N. Navin, R. Lucito, J. Healy, J. Hicks, K. Ye, A. Reiner, T. C. Gilliam, B. Trask, N. Patterson, A. Zetterberg, and M. Wigler. Large-scale copy number polymorphism in the human genome. *Science*, 305(5683):525–528, 2004.

[140] C. J. Shaw and J. R. Lupski. Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease. *Human Molecular Genetics*, 13(1):R57–R64, 2004.

[141] H. A. Simon. The architecture of complexity. *Proceedings of the American Philosophical Society*, 106:467–482, 1962.

[142] H. A. Simon. The organization of complex systems. In *Hierarchy Theory: The Challenge of Complex Systems*, pages 1–27. George Braziller, New York, NY, USA, 1973.

[143] H. A. Simon. Near decomposability and the speed of evolution. *Industrial and Corporate Change*, 11(3):587–599, 2002.

[144] G. G. Simpson. The Baldwin Effect. *Evolution*, 7(2):110–117, 1953.

[145] E. S. Slechta, K. L. Bunny, E. Kugelberg, E. Kofoid, D. I. Andersson, and J. R. Roth. Adaptive mutation: General mutagenesis is not a programmed response to stress but results from rare coamplification of dinb with lac. *Proceedings of the National Academy of Sciences*, 100(22):12847–12852, 2003.

[146] G. P. Smith. Evolution of repeated dna sequences by unequal crossover. *Science*, 191(4227):528 – 535, 1976.

[147] J. Spring. Genome duplication strikes back. *Nature Genetics*, 31(2):128–129, 2002.

[148] M. Srinivas and L. M. Patnaik. Adaptive probability of crossover and mutation in genetic algorithms. *IEEE Transactions on Systems*, 24(4):656–667, 1994.

[149] G. C. Stevens. The latitudinal gradient in geographical range: How so many species coexist in the tropics. *The American Naturalists*, 133(2):240–256, 1989.

[150] J. R. Stone and G. A. Wray. Rapid evolution of cis-regulatory sequence via local point mutations. *Molecular Biology and Evolution*, 18(9):1764–1770, 2001.

[151] W. J. Swanson and V. D. Vacquier. The rapid evolution of reproductive proteins. *Nature Reviews Genetics*, 3(2):137–144, 2002.

[152] J. Timmis, M. Amos, W. Banzhaf, and A. Tyrrell. Going back to our roots: Second generation biocomputing. *International Journal of Unconventional Computing*, 2(4):349–378, 2006.

[153] E. Tuzun, A. J. Sharp, J. A. Bailey, R. Kaul, V. A. Morrison, L. M. Pertz, E. Haugen, H. Hayden, D. Albertson, D. Pinkel, M. V. Olson, and E. E. Eichler. Fine-scale structural variation of the human genome. *Nature Genetics*, 37(7):727–732, 2005.

[154] A. S. Uyar and A. E. Harmanci. A new population based adaptive domination change mechanism for diploid genetic algorithms in dynamic environments. *Soft Computing*, 9(11):803–814, 2005.

[155] A. Wagner. Evolution of gene networks by gene duplications: A mathematical model and its implications on genome organization. *Proceedings of the National Academy of Sciences*, 91(10):4387–4391, 1994.

[156] A. Wagner. *Robustness and Evolvability in Living Systems*. Princeton University Press, Princeton, NJ, USA, 2005.

[157] A. Wagner. Robustness, evolvability, and neutrality. *Federation of European Biochemical Societies Letters*, 579(8):1772–1778, 2005.

[158] A. Wagner. Robustness and evolvability: A paradox resolved. *Proceedings of The Royal Society B*, 275(1630):91–100, 2008.

[159] G. P. Wagner and L. Altenberg. Complex adaptations and the evolution of evolvability. *Evolution*, 50(3):967–976, 1996.

[160] J. D. Watson, N. H. Hopkins, J. W. Roberts, J. A. Steitz, and A. M. Weiner. *Molecular Biology of the Gene: 5th edition*. Benjamin-Cummings, Menlo Park, CA, USA, 2004.

[161] D. M. Weinreich, N. F. Delaney, M. A. Depristo, and D. L. Hartl. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science*, 312(5770):111–114, 2006.

[162] A. P. Wolffe and M. A. Matzke. Epigenetics: Regulation through repression. *Science*, 286(5439):481–486, 1999.

[163] M. Woolfit and L. Bromham. Population size and molecular evolution on islands. *Proceedings of The Royal Society B*, 272(1578):2277–2282, 2005.

[164] R. P. Worden. A speed limit for evolution. *Journal of Theoretical Biology*, 176(1):137–152, 1995.

[165] G. A. Wray, M. W. Hahn, E. Abouheif, J. Balhoff, M. Pizer, M. Rockman, and L. Romano. The evolution of transcriptional regulation in eukaryotes. *Molecular Biology and Evolution*, 20(9):1377–1419, 2003.

[166] J. D. Wren, E. Forgacs, J. W. Fondon III, A. Pertsemlidis, S. Y. Cheng, T. Gallardo, R. S. Williams, R. V. Shohet, J. D. Minna, and H. R. Garner. Repeat polymorphisms

within gene regions: Phenotypic and evolutionary implications. *American Journal of Human Genetics*, 67(2):345–356, 2000.

[167] S. Wright. Evolution in mendelian populations. *Genetics*, 16(2):97–159, 1931.

[168] S. Wright, J. Keeling, and L. Gillman. The road from santa rosalia: A faster tempo of evolution in tropical climates. *Proceedings of the National Academy of Sciences*, 103(20):7718–7722, 2006.

[169] Z. Yang and J. P. Bielawski. Statistical methods for detecting molecular adaptation. *Trends in Ecology and Evolution*, 15(12):496–503, 2000.

[170] T. Yu and J. Miller. Neutrality and the evolvability of boolean function landscape. In *Proceedings of the Forth European Conference on Genetic Programming (EuroGP 2001)*, volume 2038 of *Lecture Notes in Computer Science*, pages 204–217. Springer, 2001.

[171] J. Zhang. Evolution by gene duplication: An update. *Trends in Ecology and Evolution*, 18(6):292–298, 2003.

[172] J. Zhang, H. F. Rosenberg, and M. Nei. Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proceedings of the National Academy of Sciences*, 95(7):3708–3713, 1998.