Abstract

Speciation is a lineage-splitting event that produces two or more separate species from one common ancestor. In the sympatric type of speciation, new species emerge from ancestors, while they are occupying the same geographical locations. This kind of speciation is viewed with some skepticism and is not well accepted. However, there are observations in nature, supporting this kind of speciation. To find an answer for this still unsolved problem in biology, numerous investigations are required. The only problem is that, biological and ecological problems are complicated systems. In these systems, collecting the data is a demanding task, because a very large spatial environment with a high number of individuals should be considered, which makes it almost impossible to collect tractable data. Moreover in such systems, verifying the hypothesis is extremely difficult, as manipulating a huge number of parameters in nature would not be easy, and more importantly, a long time is needed for an effect in a real ecosystem to be observed.

Agent-Based or Individual-Based Models can be considered as one solution for this problem. Individual-based modeling is a bottom-up approach to simulating ecosystems that allows for the consideration of the traits and behavior of individual organisms. These models are capable of simulating the simultaneous operations and interactions of multiple agents, in an attempt to re-create and predict the appearance of complex phenomena. These models also have the capability of the analyzing complex systems with dynamic behaviors as well as being affected by some parameters and factors. Individual-based models are as well tractable models, which allow having complete control on complicated conditions.

This survey reviews papers concerned with individual-based models, which investigated the sympatric type of speciation in the ecosystems, and categorizes them based on different approaches that are used for simulating the genome of individuals in the models.
## Contents

1. INTRODUCTION .............................................................................................................................. 3

2. USING BIT-STRINGS AS THE GENOME OF INDIVIDUALS ................................................................. 3
   2.1 Computer Simulation of Sympatric Speciation with Penna Ageing Model III, The Penna Model with Phenotype Selection ........................................................................................................... 4
   2.2 Computer Simulations of Sympatric Speciation in a Simple Food Web ........................................ 6
   2.3 Evolutionary Technology and Phenotype Plasticity: The FATINT System .................................... 7
   2.4 Cellular Automata Model of Macroevolution ............................................................................. 9
   2.5 Simple Lattice Model of Macroevolution ............................................................................ 11

3. USING A FUZZY COGNITIVE MAP AS THE GENOME OF THE INDIVIDUALS ............................... 13
   3.1 An Individual-Based Evolving Predator-Prey Ecosystem Simulation Using a Fuzzy Cognitive Map as the Behavior Model ............................................................................................................. 13

4. THE INFLUENCE OF THE GENOME ON THE EVOLUTION OF THE SPECIES AND SPECIATION CLUSTERING ALGORITHMS .......................................................................................... 15
   4.1 The Role of Genome and Gene Regulatory Network Canalization in the Evolution of Multi-Trait Polymorphisms and Sympatric Speciation .............................................................................. 16
   4.2 Identifying Species by Genetic Clustering ............................................................................. 18

5. SIMULATING EXISTING EXAMPLES IN NATURE .................................................................................. 19
   5.1 Multi-Agent Simulations of Evolution and Speciation in Cichlid Fish ....................................... 19
   5.2 Simulating Species Richness Using Agents with Evolving Niches, With an Example of Galápagos Plants ................................................................................................................................. 21

6. CONCLUDING COMMENTS .............................................................................................................. 22

ACKNOWLEDGMENTS .......................................................................................................................... 24

REFERENCES ......................................................................................................................................... 25
1. INTRODUCTION

This survey reviews papers concerned with individual-based models that focus on investigating the sympatric type of speciation in the evolution, which is still an un-answer question for biologists.

To find work in this area, Google Scholar was used as the initial source for this survey. Fifteen journal papers, two conference papers and three un-published research papers which are related to the topic of this survey were found. Ten papers were selected, including six journal papers, two conference papers and two un-published researches, which were more related to the topic of this survey, as the sources for this survey. The rest of the papers which are listed in the bibliography were either more concerned with the biological aspect of the topic or were previous works of the authors who their papers were already included in the ten selected papers.

Two kinds of modeling have been used to simulate the genome of individuals. One kind, which uses a pair of bit-strings, and the other kind, uses a Fuzzy Cognitive Map as the behavioral model for the individuals. In the first two sections of this survey, summaries are given of the papers related to each kind of the simulation. To study the sympatric speciation modeling, understanding the role of genome in the evolutionary process is very important. Furthermore it is essential to have access to suitable clustering algorithms, which allows you to split species from one common ancestor to two or more probable species. For this purpose, in the third section of this survey, two papers are summarized regarding the influence of genome on the speciation process, and some efficient clustering algorithms. Finally, in the last section, summaries are given of two papers that used existing ecosystems in the nature as the base model for their simulations, which allows the reader to have a better understanding about the concept of the biological problem that is tried to be investigated in this survey.

During this survey, it was observed that using a Fuzzy Cognitive Map as the behavioral model for the individuals, which was used for the first time by Gras et al. [2009], demonstrates the ability to represent complex internal concepts such as emotions and desires and their effects on the evolutionary process, while this kind of concepts and their effects on the evolutionary process have never been modeled by using the bit-strings as the genome of individuals.

2. USING BIT-STRINGS AS THE GENOME OF INDIVIDUALS

The first approach for interpreting the genome of each individual, or explaining their ecological properties, is to use bit-strings as the genotype of the individuals. These bit-strings define the ability of individuals to interact with other individuals as well as interacting with the ecosystem. In this section, five models are explained, which use bit-strings as the genome of individuals. Table.1 represents some information about these papers, their major contribution and their references to each other. The following subsections, provide summarized information about each paper in this section.

<table>
<thead>
<tr>
<th>Year</th>
<th>Authors</th>
<th>Title</th>
<th>Papers referred to</th>
<th>Major contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003</td>
<td>Luz-Burgoa et al</td>
<td>Computer Simulation of Sympatric Speciation with Penna Ageing Model III The Penna model with phenotype selection</td>
<td>none</td>
<td>Added a simple interpretation of one locus in the genome bit-string as identifier of species that lead to the flourishing and co-existence of two separate populations</td>
</tr>
<tr>
<td>2005</td>
<td>Luz-Burgoa et al</td>
<td>Computer simulations of sympatric speciation in a simple food web</td>
<td>Luz-Burgoa et al 2003</td>
<td>Investigated Sympatric Speciation with manipulating the resources and monitoring the changes in the population. Competition for the resources and sexual selection are assumed as the dominant forces acting on the population.</td>
</tr>
</tbody>
</table>
2007 Kampis and Gulyás
Evolutionary technology and phenotype plasticity: the FATINT system

2008 Borkowski
Cellular automata model of macroevolution

2009 Borkowski
Simple lattice model of macroevolution

Presented an speciation process, which can produce new functions and can go on indefinitely using different mechanisms for phenotype change as genetic point mutations or behaviour changes.

Divided The genotype’s bit-string, to two 8 bit masks, representing the ability to acquire resources, and the ability to defend from exploitation, respectively. The model also does not have a fix number of “species”.

The author added some more detailed study concentrated on robustness of the model’s dynamics, effects of the lattice size and autotrophs productivity to his previous model[Borkowski 2009]

<table>
<thead>
<tr>
<th>Year</th>
<th>Author(s)</th>
<th>Model Description</th>
<th>Refs</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007</td>
<td>Kampis and Gulyás</td>
<td>Evolutionary technology and phenotype plasticity: the FATINT system</td>
<td>none</td>
<td>Presented an speciation process, which can produce new functions and can go on indefinitely using different mechanisms for phenotype change as genetic point mutations or behaviour changes.</td>
</tr>
<tr>
<td>2008</td>
<td>Borkowski</td>
<td>Cellular automata model of macroevolution</td>
<td>Luz-Burgoa et al 2003, Borkowski 2009</td>
<td>Divided The genotype’s bit-string, to two 8 bit masks, representing the ability to acquire resources, and the ability to defend from exploitation, respectively. The model also does not have a fix number of “species”.</td>
</tr>
<tr>
<td>2009</td>
<td>Borkowski</td>
<td>Simple lattice model of macroevolution</td>
<td>Luz-Burgoa et al 2003</td>
<td>The author added some more detailed study concentrated on robustness of the model’s dynamics, effects of the lattice size and autotrophs productivity to his previous model[Borkowski 2009]</td>
</tr>
</tbody>
</table>

Table 1. Models that used bit-string as the genome of their individuals.

2.1 Computer Simulation of Sympatric Speciation with Penna Ageing Model III, The Penna Model with Phenotype Selection

This paper has been cited by three out of four other papers in this section. According to the authors of this paper ([Luz-Burgoa et al. 2003]), the concept of sympatric speciation is difficult to be explained by biologists, where the two populations continue to live in the same environment but nevertheless cease to mate with each other. They claim that it is easily explained if the two populations live in two separate environments, such as one on an island and the other on a continent, making the mating of males from one population with females from the other population impossible (Allopatric Speciation). To find a way to explain sympatric speciation, the authors suggest the use of simulation models.

The simulation in this paper is based on the author’s previous model which is called Penna model. To achieve the speciation event in the Penna bit-string model, the authors changed the Verhulst factors in the Penna model, and added some additional bit-strings, which regulate phenotype and mate selection to simulate the splitting of one species into two separate species via random mutations, even if both populations live together in the same environment. The Penna bit-string model for the biological ageing, simulates the mutation accumulation by storing bad mutations in a string of 32 bits in which the position of the bit corresponds to the age of the individuals, and its value corresponds to the health (zero) or to the sickness (one) while sick bits act from the age to which their positions belong up to the death of the individual, and three active sicknesses kill the animal. Moreover, all animals may be killed at each time step, with the Verhulst probability that corresponds to the current population and the limitation of food and space. For mating strategy in this model, when a female reaches the age of eight years, tries 20 times to find randomly a male aged eight or above. In the new simulation of the Penna model in this paper, each individual inherits two bit-string from mother and father, and produces gametes (single bit-strings) by random crossover between these two bit-strings and also probable random mutation. In this model, when one of the two bit-strings has zero and the other has one in a specific position, it can affect the health of that individual if this position is one for which the harmful allele is dominant and not recessive, when six out of the 32 possible positions are randomly considered as being dominant and the other 26 as being recessive.
In order to simulate the speciation, bit position 11 is defined to influence the mating, when each individual has $n = 0, 1, \text{ or } 2$ bits set at this position and a female with $n$ such bits at position 11 select only males with the same number $n$ of such speciation bits. The child does not necessarily have $n$ speciation bits set to one (because of random mutation), and the authors claim this allows the emergence of a new species out of the original one where all $n$ were zero.

To achieve speciation in this model, the aim of the authors is to start with one population and at the end have two populations coexisting with each other in a stable equilibrium without cross-mating. To achieve coexistence the Verhulst factor in the Penna model has been replaced with three separate Verhulst factors for the different populations $n = 0, 1, 2$. The authors state, it can be imagined, for example, that the original population $n = 0$ is vegetarian, and that the second population $n = 2$ is emerging out of it consists of carnivores. Both populations are limited by the amount of food, but their food sources are completely different. Thus, there is no competition between the two different populations, but the meat-eating females will not select any herbivore males for mating, and vice versa.

As another experience, the authors added some features to the standard Penna model to represent phenotypic selection. The first modification deals with the Verhulst factors for the deaths caused either by predation or because of intra-specific competition, and have a selective value, and the probability of their occurrence is dependent on the fitness of the individual to the environment. To simulate this issue in the context of the Penna model, an extra pair of bit-strings for each individual in the population has been added to the usual double strand of genes with age-dependent expression where each of these pairs will stand for a particular multi-locus phenotypic trait, such as size or color, which may have selective value. In the second modification which refers to the mating selectiveness, the authors introduced a locus into the genome that codes for this selectiveness, when if it is set to 0, the individual will not be selective in mating (panmictic mating), and it will be selective (assortative mating) if this locus is set to 1.

Figure 1 shows that, for nearly $10^8$ individuals, how the new species $N_2$ emerges, within about a hundred iterations, from the old species $N_0$. The intermediate population $N_1$ is only about one percent of the total, and the authors believe it can be assumed that there are two separate populations clearly emerging (Sympatric speciation). The authors claim that similar pseudo-speciation was observed in the asexual case or when the speciation bit was mutated with a special probability different from that of the other bits. The authors state that also on a square lattice without Verhulst factor, similar results were found.

![Figure 1](image.png)
The authors claim that by adding a simple interpretation of one locus in the genome as an identifier of the species to the original Pena model, the imposition of assortative mating from the start, and a bi-modal ecology, leads to the flourishing and co-existence of two separate populations (sympatric speciation).

This paper has been cited by Luz-Burgoa et al. [2005], Borkowski [2008], and Borkowski [2009], which are summarized in subsections 2.2, 2.4 and 2.5, respectively.

### 2.2 Computer Simulations of Sympatric Speciation in a Simple Food Web

Luz-Burgoa et al. [2005], state that using an individual-based model is one of the best possible approaches for explaining sympatric speciation, which corresponds to the division of a single local population into two or more species without any geographical barrier.

The authors refer to their own previous investigations [Luz-Burgoa et al. 2003], on their earlier individual-based model, which is called Pena and has been used as the base model for this paper’s investigations, with some modification on speciation section.

According to the authors, the original Penna model was based on Darwinian evolution and focused on problems of biological aging and applications to several different evolutionary problems. An age-structured pair of bit-strings of 32 bits each, has been used in sexual version of Penna model, in which each bit position corresponds to a period of time (year) in each individual’s life time. Deleterious mutations are represented by bits 1. If at a specific bit position, one of the two bit-strings has a bit zero, and the other has a bit one, it affects the health of that individual if this position is one for which the harmful allele bit 1 is dominant. An individual dies when it reaches a threshold number of deleterious mutations. A survived female chooses randomly a male with the same age or older for reproduction, after reaching the age R and produces b offspring, each having one of the father’s gametes and one of the mother’s gametes as its two bit-strings, followed by one random mutation. To prevent the exponential increase of the population, there is a competition for space and food given by the logistic Verhulst factor, which may kill the individual independently of its age or genome. This completely random killing factor, presented in the original Penna model, has been modified in this paper’s investigations.

In this paper, the authors try to investigate the process of sympatric speciation in a simple food web model. The authors try to examine this process using a modified version of the Penna model [Luz-Burgoa et al. 2003] assuming that competition for resource and sexual selection are the dominant forces acting on the population. To obtain sympatric speciation, it has been considered that competition for resources, changes according to the ecology. This trait is represented by a new pair of nonage structured bit-strings, added to the original structured one. The phenotypic characteristic is measured by counting, in this new pair of bit-strings, the number of recessive bit positions chosen as 16, where both bits are set to 1, plus the number of dominant positions with at least one of the two bits set, which will be a number k between 0 and 32 that is called MF the mutation probability per locus of this trait. The death probability (V) by intra-specific competition, for external phenotype, is calculated by a modified Verhulst factor, which is depended on the number of k.

For each individual in each time step a random number is generated. The individual dies if this number is smaller than V. The function of F(k,t) has been used as the basal resource species of a food chain. It may, for instance, represent plants with a given size distribution. Individuals with the extremal phenotype P₁, P₂ compete for small/large plants among the individuals with its same extremal phenotype. For mating selectiveness the authors introduced a locus into each genome that codes for the selectiveness, also
obeying the general rules of the Penna model for genetic heritage and mutation. If it is set to 0, the individual is not selective in mating (panmictic mating) and is selective (assortative mating) if this locus is set to 1. At the beginning of the simulation, all females are nonselective with the mutation probability of 0.001 in both directions. Selective female choose mating partners according to one of the following mating strategies: Mating strategy 1: A female mates with the first randomly chosen male who has its same phenotype \((K<16/K>=16)\). Mating strategy 2: In this case a female chooses, among six males, the one with the smallest difference between its phenotype \(k_F\) and the male’s phenotype \(k_M\). Mating strategy 3: The mating of a pair occurs with probability \(P=k_F-k_M/32\). The female tries to mate for at most six times, If it does not find a proper male, then it mates randomly.

Two different scenarios have been considered to analyze the model. The first scenario uses a two species food web with a basal resource and a consumer (herbivore) that feeds on this resource, which has genetic properties and evolves for 250 generations with a constant food distribution. Suppose for instance, that during a given season, this food distribution consists of plants which sizes favor the individuals of the second species presenting medium phenotypes.

The second scenario has three species food web in which the consumer (a predator) feeds solely on the herbivore and has genetic properties. The herbivore also has genetic properties without mating preference, and evolves for 250 generations with the same initial basal resource distribution of the first food web scenario, when the basal resource distribution suddenly changes, the phenotype distribution of the herbivores, consequently changes.

An interesting result is that for the mating strategies 1 and 2 the second species phenotypic distributions are the same since for strategy 1 the female knows the drift direction of the ecological change, and it is easy to understand why the final population presents two substantially different phenotypes and how reproductive isolation between them has driven the elimination of all intermediate phenotypes while females with mating strategy 2 do not know this direction and, even so, the ecological change drives their preferences in the same way as with choice 1. For the mating strategy 3, there is no correlation between ecological changes and female’s preference, however the authors believe that this strategy is more realistic than strategy 2, since the female’s preference is subject to the male’s availability.

The authors claim that a real process of speciation of three lineages of the Darwin’s finches, which is called three finch fits into their model. This group consisting of six finches, all of them are insect eaters except one vegetarian. Vegetarian finch diverged from the ancestral stock before the divergence from the rest of the tree finch group. They also state that in the first scenario (only two species), the sympatric speciation for the herbivores (top species) is always obtained, however in the second scenario, speciation is not always obtained for the predators (top species). Comparing the results for two and three species food webs, the authors conclude that higher-level consumers take longer to be speciated when the distribution of the basal resource changes abruptly.

There is no citation to this paper by other papers in this survey.

### 2.3 Evolutionary Technology and Phenotype Plasticity: The FATINT System

According to Kampis and Gulyás [2007], the open-ended evolution is widely recognized as a difficult and unsolved problem in biology. They believe that currently, there is no accepted general evolutionary theory for the origin of complexity or the maintenance of the evolutionary changes. The authors also state that simulating the biological populations for understanding the Evolutionary Technology can be considered
as an attempt to realize open-ended evolution showing complexification and to harness it for the technological needs.

Linking together the Evolutionary Technology (i.e. open-ended evolution) with the species concept, niche construction and plastic phenotypes, the authors advanced a general Phenotype Hypothesis, PH, which predicts that no evolution model and no evolutionary technology will be able to produce a sustained evolutionary process or support the increase of functional complexity without admitting a changing phenotype space as a natural part of the interaction potential of the organism. They considered emergent species that occupy dynamically created niches based on phenotype interactions and examined two Hypotheses:

Hypothesis 1: An evolutionary engine with phenotype-based emergent species and without phenotype plasticity, remains stable and non-productive of species. Hypothesis 2: The introduction of phenotype changes in the above systems leads to a sustained production of new emergent species.

To address the above crucial issues of phenotype representation and to study their effects in maintaining open ended progress, the authors performed their study using digital organisms. Their system which is called FATINT, was written in an encapsulated agent-based approach, using the Java based REPASt environment, and it offers a graphical user interface which the authors claim allows for a wide selection of models and a broad range of interactive experiments by different parameter settings. A high number of automatic observers, including a metric distance plot, various histograms, population plots (such as number and age distributions, and programmable 2D projections), species graphs and others are implemented in this system. In this system, each agent’s phenotype is represented as a vector $v_i$ of integers from the interval $[V_{min}, V_{max}]$, where the length of the vector is $L$, and it is always identical with all agents but it is variable in time. In the FATINT model, sexually reproducing gender-less (i.e. ‘snail-like’) agents, represent the individuals and reproductive success depends on a phenotype similarity metric. In the FATINT system, a species emerges as a dynamically maintained cluster of interbreeding individuals, around a dynamically defined center characterized by the predominant types, and also any individual can become the ‘center’ of a species. An evolution engine that consists of a population of agents, who consume energy, and some evolutionary operators for crossing over, mutation, resource uptake and ageing or death, is used at the base level of the system.

For reproduction policy in this system, an active agent picks a random mate from the list of potential partners, which are similar individuals and their similarity is measured by the Euclidean distance between the agents’ phenotype vectors.

The authors demonstrated three different automatic species observers to study the reality of species found. Fig. 2 shows convergence to a single cluster in phenotype space, using a standard hierarchical clustering method which is characterized with a clustering constant $d_{cluster}$. The $d_{cluster}$ expressed here as a ratio to $M_{limit}$, where at $d_{cluster} = 1$ they ask about clusters of size $M_{limit}$. The authors state that in order to test the effects of the phenotype changes, in FATINT’s phenotype change mode, at every time step a phenotype change can be initiated by hand or, at every reproduction event, a new phenotype dimension can be introduced with a variable probability $P_{change}$ per offspring. According to the authors, when such an event occurs, a new component is added to the agents’ phenotype vector with a value depending on the used method.
Fig. 2 shows the clusters against time. It is the summary of 60 runs, where 10 for 6 random seeds each, from $d_{\text{cluster}} = 0.9$ to 1.4. In this figure, time goes left to right and $d_{\text{cluster}}$ increases right to left and the vertical axis shows the number of the clusters.

To understand the process of phenotype change and the speciation transition more closely, the authors produced a sample run, and its comparisons can be seen in the Table 2. In this run, stable species situations before speciation are characterized by a Rand index $R = 1$ for all pair-wise comparisons of the species metrics. Then, around the time of the introduction of a new phenotype trait ($t = 450$), clusters formed using phenetic (similarity-based) and cladistic (lineage-based) classifications strongly diverge from each other, which according to the authors shows that phenotype properties are more different than are lineages.

The authors claim that the FATINT evolution engine also permits further studies such as drift, equilibrium, density-dependent effects and other phenomena, which are well understood at population level models but still require further study in the present individual-based framework. They also claim that the typical process though the introduction of a new phenotype trait is the appearance and amplification of phenotype heterogeneity, which often leads to the appearance of new emergent species centers which are strong enough to facilitate the separation of the formerly stable species into new species, which shows the ability of having different functionalities. They also showed that changing phenotype-to-phenotype interactions can repeatedly split species by the production of new selection constraints and sustained ecological evolution always proceeds by similar mechanisms. The authors also claim that, their presented speciation process, can produce new functions (i.e. new sexual selection processes) and can go on indefinitely using different mechanisms for phenotype change as genetic point mutations or behavior change.

There is no citation to this paper, by other papers in this survey.

2.4 Cellular Automata Model of Macroevolution

Borkowski [2008], states that as in the context of many ecosystems, the macro level of evolution typically acts on a very long time scale, thus mathematical models and numerical simulations would be extremely important for the study of the macro level of evolution as they may be the only way to understand a variety of observable facts, repeatable and recurrent processes and patterns. In this paper, the author tried to study emergent properties of macroevolution using a cellular automaton model of a multi-species ecosystem.

The author refers to previous work of Luz-Burgoa et al. [2003], and his future work, which was unpublished at that time [Borkowski 2009].

The author argues unlike most of the previous models that investigate rather different aspects of predator-prey or predator-herbivore-plant systems, a rare number of them such as [Luz-Burgoa et al. 2003] raise questions more related to macroevolution, like speciation. The author also mentions that some
more detailed study concentrated on robustness of his model dynamics, effects of the lattice size and autotrophs productivity will be published in a more elaborative paper [Borkowski 2009].

The author designed an individual-based model of macroevolution as similar as possible to a cellular automaton, using basic concepts, such as simple entities in partially occupied rectangular lattice, governed by local rules of interaction in Monte-Carlo dynamics, which their attributes are based on the central principles of Ecology, that are:

1. Using a network of energy flow from producers (autotrophs) to a number of consumers (heterotrophs) as a community of species.
2. Using a very large and multidimensional space which allows for simulation of many “ways of life”, adaptations for acquiring energy and for defending from exploitations and different levels of specialization.
3. Searching the space of niches by a “random walk” process, in which newly emerged species takes a niche adjacent to the niche of its ancestral species. Populations need to obtain enough energy from a biotic source or from coexisting populations, or they would vanish later or sooner (rather).
4. All populations are continuously evolving on a longer time scale, when in changing conditions a population fails in the balancing task, it becomes extinct and at the same time changes the environment for other interconnected species, which may lead to vanish of other populations. Author names the model CO-EVO to underline the essential role of multi-species co-evolution in long time evolutionary dynamics.

The author implemented the principles as follows: one simple entity which is called “agent” is characterized by a 16-bit string to define its ability to interact with other agents and two additional attributes of energy and age. Each agent belongs to a number of possible classes and in simulation, the position and states of the agents change in the Monte-Carlo dynamics, where agents are randomly chosen for activity and the lattice site for their action are randomly selected from the so-called “Moore neighborhood”. Based on its level of energy and the state of adjacent site, each agent can choose its action (moving, setting offspring or eating the current owner of the site), which increases the agent’s age (each agent dies after exceeding a particular limit). Reproduction is asexual; therefore, the offspring copies the ecological characteristics of the parent with probability of “mutation” of 1 out of 100, by changing 1 bit of its bit-string. In this model, if a newly emerged agent can survive and reach a threshold number of entities can be considered as a new “species.”

The author states that in this model, each agent has a 16 bit-string which is divided to two 8 bit masks (acquiring and defense mask). When an active agent has at least one bit set at the same position as the defense mask of the attacked agent, exploitation would be possible. The acquiring mask with all bits equal to 1 allows the agent to be an autotroph, but a 0 at any position causes the agent to be a heterotrophy. The prey always dies, and the aggressor receives only a fraction of its energy (E_f), and the rest of the energy disappears.

Borkowski claims that this model may produce analogous results, based on different techniques and assumptions. On a short time scale, the simulation dynamics would be similar to ones reported from individual based predator-prey models. On medium time scales, evolutionary cycles would be observable and finally in long time scale repeats the same dynamics, composed from few successive stages.

Borkowski also claims that in a broad range of parameters, the CO-EVO model repeats qualitatively the same dynamics, composed from few successive stages. Exponential colonization of lattice, first
dynamic equilibrium of the ecosystem based on relatively weakly armored autotrophs, “great extinction” caused by a series of “key innovations” in autotrophs defense mask and a second equilibrium with the ecosystem based on heavy armored autotrophs. The author mentions, in any equilibrium state the system has a constant rate of species turn-over, and during the exponential growth at the beginning of the simulation course and just after the avalanche mass extinction, new species are produced much faster, that resembles the predictions of “Punctuated Equilibrium” theory.

The author claims comparing mean number or species, autotrophic species and ratio of autotrophic species taken from last one thousand M-C steps of the second equilibrium stage, it is clearly visible that fragmentation may increase the number of species, but within some range, and differently for autotrophic and heterotrophic species. He argues that such a result agrees more or less with basic biological knowledge - naturally fragmented ecosystems are usually richer than homogeneous ones, but additional anthropogenic fragmentation impoverishes them significantly.

Borkowski claims that the main concept which distinguishes his model from other evolutionary simulations (e.g. [Luz-Burgoa et al. 2003]), is the interpretation of the “genotype” bit-strings of an agent as whole ecological properties of it (the bit-string is divided to two 8 bit masks, the first one represents the ability to acquire resources, and the second defines the ability to defend from exploitation). The author also mentions that unlike most versions of the predator-prey and classical models, his model does not have a fix number of “species”, which allows him to examine relations between global characteristics of an environment and ecosystem complexity, simply measured by the number of inhabiting agents and their classes. There is no citation to this paper by other papers in this section.

2.5 Simple Lattice Model of Macroevolution

Borkowski [2009], states that, the study of the macro level of evolution is a very difficult task for biologists, because they have to consider a huge body of observable facts, repeatable and recurrent processes and patterns, and disputable consequences of such phenomena like adaptive radiations, emergence or vanishing of natural barriers, climate change, local and global catastrophes, spontaneous or induced mass extinctions. The author explains that mathematical models and numerical simulations make it easier for the biologists to study the macroevolution. According to the author, these models can be considered as an approach to answer a very important question for evolutionary biology and paleontology: How much is macroevolution determined by the environmental restrictions, and how much does it depend on the random events?

Borkowski refers to Luz-Burgoa et al. [2003] as previous work in this field of investigation. According to the author, a rare number of contemporary works such as [Luz-Burgoa et al. 2003] raise questions much closer to macroevolution like sympatric speciation or the causes of mass extinctions, while most of them explore rather different aspects of the dynamics of predator–prey or predator–herbivore–plant systems, or optionally consider also environmental changes.

In this paper, Borkowski presents a minimalistic individual-based model of macroevolution, which is very similar to cellular automata, and combines the basic concepts used typically in classical kinds of micro-simulation, with the most important aspects designed from a biological point of view. The author believes that considering a community of species living in a particular ecosystem as a system of energy (or biomass) that flows from every population of producers (autotrophs) to a number of interconnected populations of consumers (heterotrophs) is a quite basic idea for modern ecology. The author claims that, all populations in his ecosystem are constantly evolving, pressed by their enemies, competitors, and the
smartest prey and if a population fails in such a task, it becomes extinct, and could no longer play this evolutionary game. The author named his model CO–EVO to underline the important role of multi-species co-evolution.

The environment of the simulation is a rectangular lattice which its side proportions are 2:1, because author believes square environments are rare in the nature. Each cell of the lattice can be empty or occupied by an agent. Agents are randomly selected for action in a randomly selected site in the neighborhood. A 16 bit “genotype” and two additional attributes, including energy and age, define the ability of agent to interact with other agents and the ecosystem. The action of the active agents, depend on many factors such as their abilities and their energy level and the state of the selected adjacent site. Based on these factors the action of an agent can be just moving in a site, setting his offspring in a site, attacking and eating the owner of the site or just switching the sites with the neighbor agent. In this model, reproduction is asexual, and a copy of one parent’s bit-string with a probability of mutation (1 out of 10 times) in one bit is considered as the genome of offspring. The author states that mutations in this model can be considered as the speciation events. In this model, the agent’s genotype bit-string is divided into two 8-bit masks, which first one represents a strategy for acquiring resources, and second one represents a strategy for defending from other heterotrophic agents. In the attack process, the prey always dies and a part of its energy, based on the energy reserve of the prey, the value of the defense mask of the prey and the value of the acquiring mask of the aggressor, will be given to the predator, and the rest of its energy disappears. According to the author, at the beginning of the simulation, an autotrophic agent, which has no defending strategy, harvests only natural resources. Then probable mutations can affect the agents’ strategy which may lead to the appearance of new strategies for taking resources from other agents. The author claims in this model, the mutations can produce armored classes of agents, and a complete food network can emerge that naturally goes to its own equilibrium state.

Borkowski states that, at the beginning of the simulation, a short number of heterotrophic mutants may accidentally appear in the short-lived clones, until the population of producers achieves sufficient density at some area. After that, the number of clones begins to grow exponentially following the number of agents to the moment of saturation. Author believes that it is a starting point for a real competition which clears out the poorly adapted clones, and the system arrives at the first dynamic equilibrium state (on Fig. 3. after 1010-th step). He adds, during several thousands of M-C steps the numerous heterotrophic clones exploit few populations of weakly armored autotrophs (Fig. 4.I). Some particular clones are replaced at statistically constant speed, but the general organization of the ecosystem remains broadly the same.
According to Borkowski, starting from the size of 150×100 cells, the system always reaches the second equilibrium. Nevertheless, for fragmented lattices, the amplitude of the agent fluctuations in general is smaller.

Borkowski claims that there are very important differences between his model and other micro-interactive models. According to the author what mainly differentiates his model from others is the organization of the agent genotype and its consequence, which is divided into two 8-bit masks. He states that in his model, the number of species, both producers and consumers are not fixed, all populations could potentially interact with each other and by speciation, they are able to take every ecological niches possible in the particular moment which allows the model to show properties and to achieve results similar to both stochastic models of macroevolution and for “predator–prey” based micro-simulations. The author also claims that the CO–EVO model, although is based on very different techniques and assumptions and seems to be able to produce analogous results to other better-known models.

This paper has been cited by Borkowski [2008] as a parallel project, which was unpublished at that time.

3. USING A FUZZY COGNITIVE MAP AS THE GENOME OF THE INDIVIDUALS

The second approach for simulating an individuals’ behavior, which for the first time was used by Gras et al [2009] in a large scale individual-based evolutionary process, is to use a Fuzzy Cognitive Map as the behavioral model for the individuals. The FCM enables individuals to achieve a perception from the environment and choose their action based on the achieved perception. In this section, an explanation is given of the only individual-based model which used this approach.

3.1 An Individual-Based Evolving Predator-Prey Ecosystem Simulation Using a Fuzzy Cognitive Map as the Behavior Model

As natural ecosystems are very complex (in terms of the number of species, and of ecological interactions) the authors believe that using ecosystem models, which typically simplify the systems, which they are representing, to a limited number of components and allows for the development of computer-aided ecosystem simulations that are tractable would be beneficial in macroevolution investigation because ecosystem simulations offer a global view of the evolution of the system, which is difficult to be observed in the nature.

As all the previous works on ecosystem-modeling which are mentioned by Gras et al [2009], does not concentrate specifically on sympatric speciation, which is the topic of this survey, they have not been explained here.

Gras et al. [2009] mention that to their knowledge there have been just two models embodying mechanisms of speciation before them, those in which species were migrating and becoming distinct but
did not originate, and those in which the number of species was a predefined parameter. They do not provide any detailed information about previous works in sympatric speciation modeling.

Gras et al [2009] developed an ecological modeling to characterize the major dynamics of ecosystems, in order to synthesize the understanding of such systems, and to allow predictions of their behavior. As authors claim this study “presents an individual-based predator-prey model with, for the first time, each agent’s behavior being modeled by a fuzzy cognitive map (FCM), allowing the evolution of the agent behavior through the epochs of the simulation”. The FCM enables the agent to evaluate its environment (e.g., distance to predator or prey, distance to potential breeding partner, distance to food, energy level) and its internal states (e.g., fear, hunger, curiosity), and to choose several possible actions such as evasion, eating, or breeding. The FCM of each individual is unique and is the result of the evolutionary process. The notion of species is also implemented in such a way that species emerge from the evolving population of agents. The authors claim that their system is the only one that allows the modeling of links between behavior patterns and speciation. According to the authors, a notable amount of data can be produced by the model, including the number of individuals, level of energy by individual, choice of action, age of the individuals, and average FCM associated with each species which allows for numerous investigations on macro-evolutionary processes.

The FCM is used to model the agent behaviors (structure of the graph) and to compute the next action of the agent (dynamics of the map). A map contains three kinds of concepts: sensitive, internal, and motor. The activation level of a sensitive concept is computed by a “fuzzification” of the information coming from the environment. The activation level of the motor concept is used to determine what the next action of the agent will be, and a “defuzzification” of its value can be used to determine the amplitude of the action. Finally, the internal concepts’ activation levels correspond to the levels of intensity of the internal states of the agent and affect the computation of the dynamic of the map.

Gras et al states that in this model, each agent has its FCM, which represents its genome. The FCM contains, sensitive concepts (e.g. foeClose, foeFar, foodClose, foodFar, mateClose, mateFar, etc.), internal concepts (e.g. hunting, fear, hunger, sexual needs, curiosity, sedentarity, satisfaction, annoyance), and motor concepts (e.g. evasion, searchForPreys, searchForFood, socialization, eating, breeding, etc.). It also contains links and weights among different concepts representing the mutual influence of these concepts. The authors claim that, such a behavioral model allows to representation of very complex phenomena. For example, the concept of evasion is excited by the concepts of fear and annoyance and inhibited by the concepts of hunger, sexual needs, curiosity, and satisfaction. These concepts in turn are excited or inhibited by all the sensitive concepts that means, the activation level of the motor concept of evasion depends on a complex and nonlinear combination of all the sensitive concepts and of six internal concepts. This is true for all motor concepts. Another important thing to notice is that the activation levels of the concepts of an agent are never reset during its life. As the previous time step activation level of a concept is involved in the computation of its next activation level, this means that all previous states of an agent during its life participate in the computation of its current state. It means therefore, that an agent has a memory of its own past that will influence its future states. As the action undertaken by an agent at a given time step depends on the current activation level of its motor concepts, the global behavior of an agent dynamically depends on a complex combination of the information it currently receives from its environment, its current internal states, and the past states it went through during its life.

The authors explain that, in this model, a species is a set of individuals associated with the average genetic characteristic of its members. The average map is computed on the basis of the FCM matrices of all individuals in that species. An individual is considered to be belonged to a species if the difference
between its matrix and the average matrix of the species is below a speciation threshold. Interbreeding can take place if the distance between individual matrices is below the reproduce threshold. When a new born appears, the distance between its matrix and the average matrices of all species is computed, if its distance from the most similar species is greater than the speciation threshold, then the individual forms a new species.

This simulation takes place in a toric virtual world, composed of 1000 cells in both dimension and each cell contains a limited number of resources (grass and meat), and can contain an unlimited number of individuals, which the authors claim that, this will allow for completion for resources between individuals to occur.

According to Gras et al [2009], the evolution of species is derived directly from the evolution of individuals which occurs when there is an interbreeding event, when an offspring is conceived by two parent agents. The offspring inherits a combination of the genomic information of its parents with possible mutations, which is all the information that is transmitted from the parents to the child and submitted to possible mutations.

The longest run of the simulation at that time, which was corresponded to 7112 time steps, lasted for two and a half months and the authors used its results for this paper’s analysis. Gras et al [2009] claim that, having access to details of all parameters of all components at any time of the process makes it useful for a biological interpretation. They also mention that, the results allow having access to the speciation events and then construct the exact phylogeny of the evolving predator and prey species. In order to illustrate the total behavior of the simulation and to find if its properties are similar with those who are known in nature, several of these parameters were extracted by the authors, and the cross correlation between them has been computed.

Gras et al claim that even though the simulation is in a very large and adaptive system, its whole behavior is fairly stable and presents interesting correlations. They state that, the agents make decisions of action, the evolutionary system generates behavioral models with new innovative notions and there is a correlation between the number of species and the number of individuals in accord with the existing correlation.

Gras et al also claim that an FCM has never been used in large-scale individual-based modeling and has never been used in an evolutionary contest. They also argue that using an FCM as a behavioral model for agents, led to very promising results that demonstrate the ability to represent complex internal concepts such as emotions and desires, and to build agents, who are able to perceive, make decisions, and act. They also claim that this is the only simulation modeling the fact that individual behaviors affect evolution and speciation.

This paper is cited by Murdock and Yaeger [2011], which is summarized in the subsection 4.2, of this survey.

4. THE INFLUENCE OF THE GENOME ON THE EVOLUTION OF THE SPECIES AND SPECIATION CLUSTERING ALGORITHMS

In the first paper of this section, [Ten, Tusscher and Hogeweg 2009], the role of a flexible, evolving genome and gene regulatory network architecture in the evolutionary process and speciation has been
investigated. The second paper in this section, [Murdock and Yaeger 2011], explains how a clustering algorithm can be used to identify the newly emerged species, which are the result of the evolutionary process.

4.1 The Role of Genome and Gene Regulatory Network Canalization in the Evolution of Multi-Trait Polymorphisms and Sympatric Speciation

Ten Tusscher and Hogeweg [2009] state that in the biology field, the sexual reproduction has classically been considered as a barrier to the buildup of discrete phenotypic differentiation, and this notion has been confirmed by models of sympatric speciation in which a fixed genetic architecture and a linear genotype phenotype mapping were assumed, but it has never been studied for the nonlinear genetic architecture. According to the authors, simulating the influence of a flexible genetic architecture and non-linear genotype phenotype map on differentiation under sexual reproduction can be a suitable way for investigation about this biological problem.

In this study, the authors explicitly investigate “the influence of a flexible, evolving genome and gene regulatory network architecture and the resulting non-linear genotype phenotype mapping on the evolution of phenotypic divergence under sexual reproduction.” They built an individual-based model in which organisms compete in a multi-niche environment to induce disruptive selection. In their model, organisms have a genome consisting of genes and transcription factor binding sites, where mutations affect individual genes or binding sites or stretches of genome. In this model, the genome codes for a gene regulatory network, that determines the gene expression dynamics and the phenotype of the individual. According to the authors, this model contains a complex, non-linear, many-to-one genotype phenotype mapping. For the comparison purposes, they used a null model in which individuals have a simple genome only containing genes, and in which mutations only affect whether a gene is expressed or not.

The proposed individual-based model is consisting of a population of organisms, living on a two-dimensional grid world (typically 60 × 140 that each position of it can be either empty or occupied by a single individual).

At the start of the simulation, the population is initialized with 1500 identical individuals who are placed randomly in the middle region of the grid. Organisms within a certain radius of an empty position compete to reproduce into that position using probabilistic fitness-based selection. The authors state that, in this model, the population is allowed to evolve for 500000 time steps, and organisms have a linear genome that contains genes, and their upstream regulatory regions composed of transcription factor binding sites (TFBSs). This model includes two types of genes, including the transcription factors (TF), which influence the expression of genes, and phenotype genes, which determine the phenotype of the organism. The simulation is started by randomly generating, the genome of the first 1500 identical individuals and randomly placing an average of 2 TFBSs upstream of each gene. To keep their model computationally tractable, the authors used a Boolean gene network in which genes are either expressed (1 or on) or not expressed (0 or off), where the next state of a gene “i”, depends on its current state and the occupancy of its TFBSs.

To impose frequency-dependent disruptive selection the authors used a multiple-niche environment and local, frequency-dependent competition with 20 different ecological niches to which individuals can adapt. They presented all niches at each grid position, to be able to study the sympatric speciation and also be able to interpret it as ecological roles rather than physical habitats. At the start of the simulation, they randomly picked 20 different niches from the space of $2^{20}$ possible niches. The strength of the
competition between individuals depends on the similarity between the competing individuals and their number.

A general overview of the model is presented in the figure 5. In the figure 6, the evolution of phenotypic diversity under random mating in the full and the null model has been represented. “Part A, shows the evolution of niche-fitness in the full and the null model. Part B, shows the evolution of the difference in niche-fitness between hybrids and non-hybrids in the full and the null model. Part C shows the evolution of morph heterogeneity in the full and the null model” [Ten Tusscher and Hogeweg 2009]. In part D, the evolution of morph cohesion in the full and the null model is represented. Parts E and F represent. “Randomly 10% of individuals are selected from the population to study the distribution of phenotypes presented in the population for the full (E) and the null (F) model, where the black colors show the not expressed genes, and whites represent expressed genes” [Ten Tusscher and Hogeweg 2009]. The distribution of niches on which the population has specialized at time 500000, for the full and null model are represented in G and H, respectively, where a niche is characterized by the phenotype gene expression pattern needed to be perfectly adapted to that niche, and black color shows the genes that should be off in this niche, while white color shows the genes that should be on in this niche.

Ten Tusscher and Hogeweg claim that their model can be assumed as a generic outcome of the evolution of discrete clusters of organisms, which are capable of adapt to different niches, despite random mating. They also claim that in case of a flexible genetic architecture and a non-linear genotype phenotype mapping, sexual reproduction does not constrain phenotypic differentiation, but instead constrains the genotypic differences underlying it. They hypothesize that, as genotypic canalization enables differentiation despite random mating, and increases the effectiveness of assortative mating, sympatric speciation is more likely than is commonly suggested.
This paper is not cited by other papers in this survey.

4.2 Identifying Species by Genetic Clustering

Murdock and Yaeger [2001], believe that, understanding speciation is one of the key problems in biology. They state that, artificial life simulations exhibit complex agent-based behaviors, which persist and evolve through genetic recombination and mutation. They argue that, where explicit speciation is built into the simulation, identifying emergent species in these simulations would be difficult, both theoretically and practically and as a result a technique for identifying subpopulations of agents using a clustering algorithm to identify groups of agents with shared genetic attributes would be essential. They claim that, the resulting clusters might reasonably be considered distinct species, and allows identifying some of the different adaptation mechanisms adopted in the simulation.

Murdock and Yaeger refer to previous works of Gras et al. [2009], which is explained in subsection 3.1 of this survey. According to the authors, use of gene clustering for speciation has been explored in computational ecosystems by Gras et al. [2009] where their predator-prey simulator has some traits in common with the authors’ model, but defines two distinct agent classes, which do not interbreed, and the clustering analysis is performed during the simulation and allowed to control reproductive success, thus allowing it to drive the speciation process.

The authors’ simulation model name is Polyworld, that is a computational ecology with a long history, in which populations of haploid agents evolve, each possessing a suite of primitive behaviors (move, turn, eat, mate, attack, light, focus) under continuous control of an Artificial Neural Network (ANN) employing discrete-time. Input to the ANN consists of pixels from a rendering of the scene from each agent’s point of view. Output from the ANN consists of the aforementioned primitive behaviors. All Agents replenish their energy levels by seeking out and consuming food or by killing and eating other agents. Reproduction in this model, occurs when two collocated agents simultaneously express their mating behaviors.

The authors state that, the structural architecture of the ANNs and the activation of every neuron at every time step for every agent are recorded during the progress of the simulation, allowing investigation into evolutionary trends in network structure and function. Agent genomes also are recorded, and are used as the basis for the clustering analysis.

The clustering task in this system, includes two sub-problems: a distance function for measuring object similarity by use of entropy-weighted Euclidean distance over each agent’s genome, and a clustering algorithm to partition objects, which uses a variation of the QT-Clust algorithms and adds a new algorithmic improvement to allow for multiple cluster selection on each pass and a pre-calculation of point-wise distances for greater efficiency. According to the authors, effective clusters should maximize inter-cluster distances, while minimizing intra-cluster distances (cluster diameter). The authors’ clustering algorithm uses nearest-neighbor clustering approach by fixing the cluster diameter (ε), rather than the number of clusters which allows for data discovery problems, such as gene analysis.

For analysis, the authors configured the world with two barriers, running 90% of the depth of the world but left open for the remaining 10% of the world, to enable the populations to mix relatively easily, but not with complete freedom. The clustering algorithm is performed on Polyworld simulation data containing 29,564 agents (distributed over 30,000 time steps), contained in 1.9GB of genomic data with
the purpose of teasing apart the dynamics of a particular simulation, and to obtain the degree to which cluster analysis and a species/sub-population perspective can inform the understanding of those dynamics.

Table 3 shows the number of clusters identified for varying levels of $\varepsilon$. Figure 7 shows the population of each cluster over time for $\varepsilon = 2.25$. The progression from a large diameter to a smaller diameter shows each cluster splintering. The authors claim that the results are in general agreement with previous simulations, showing strong selection for complexity in early populations during the period in which they are evolving to adopt an Ideal Free Distribution of agents to the heterogeneous resources of the simulated environment.

The authors claim that, clusters can be considered as species within their artificial simulation. They argue since the simulation does not explicitly prevent interbreeding between clusters or base reproductive success on genetic distance, they should be considered proto-species, but the fall and rise of sub-populations, with significantly different genetic makeup from the dominant population, suggests a degree of specificity and persistence of species identity. The authors claim that both forms of speciation are to be found in their simulations as in the movies showing cluster membership over time it can be seen that clusters emerge and persist alongside existing clusters in a fully sympatric fashion, but also the evidence of allopatric speciation can be seen when new clusters emerging in and coming to dominate one food patch before spreading to the other.

This paper is not cited by other papers in this survey.

5. SIMULATING EXISTING EXAMPLES IN NATURE

In this section, two individual-based simulating models, which used real example in the nature for their simulations, are explained. The first paper, [Clement 2003a], models speciation in the cichlid fish in the Great Lake of Africa. The evolution of the plants species richness patterns on 22 Galapagos Island, Ecuador, is modeled in the second paper [Boone 2010]. These examples can be beneficiary for giving a better understanding about the biological problems that can be investigated by individual-based modeling approaches.

5.1 Multi-Agent Simulations of Evolution and Speciation in Cichlid Fish

Clement [2003b] claims that this paper is beneficial as it contributes to solve open problems in cichlid biology, and provides a library of field research to drive the design and parameterization of the simulation. The investigation of conditions under which sympatric speciation would, and would not, occur is one of the open problems that according to the author has been addressed in this paper.
The author refers to his own side projects [Clement 2003b], and [Clement 2003c] as references for this paper. He mentions that previous works in the speciation simulation field generally attempt to find general rules for speciation, by modeling highly abstract ecologies including trophic preference, adaptation (preferred foods), some sexual signaling phenotype (e.g. color), and a sexual preference (e.g. the preferred mate color of a female). The author does not provide more detailed information regarding other authors’ previous work. A side project of the author examined the possible role of social learning behavior in partially dividing populations as a precursor to speciation [Clement 2003c]. Work on detailed visualization of the process (rather than just the result) of speciation (using fuzzy sets to represent species) is being undertaken in parallel to the main simulation system, and has been reported in [Clement 2003a].

In order to investigate an open problem in cichlid fish speciation in the African Great Lakes, the author describes the use of a purpose-built multi-agent simulation system (TDLP – “The Digital Lake Project”). An individual-based platform has been implemented in Java in order to allow the maximum flexibility during the development and evolution of the project. TDLP combines both ecological modeling and a varying genetic model. It simulates an ecosystem in discrete time steps, roughly approximating one week each time step. The basic living entity in the simulation is a fish which is modeled by a number of parameters defined by a 'lineage'. The lineage has some parameters such as a lineage name (basic type such as cichlid, catfish, or cypriniform), sublineage (a species name), size at birth, maximum size, growth increment, typical lifespan, etc. These identical properties control the ecological properties of the Fish during the simulation, with generic values chosen for each. According to the author, the lineage also defines the plug-in models that control the three phenotypes (trophic adaptation, sexual phenotype (male color), and (female) sexual preferences). A random function with equal probabilities for male and female offspring has been used to assign the gender of the fish. At each time step, the population agent calls methods of the fish agents to simulate various life processes and events. If a female fish survives and riches sexual maturity, and has enough energy, at the start of a new breeding period, it chooses a mate according to a probability function. Food sources are represented by an interface and supply a fixed amount of energy to fish in the population every time interval. Each food source is assigned a location (a N-dimensional point) in the 'trophic space' and it can be evaluated how suitably adapted fish are for harvesting the food source. The fish compete for the energy available from food sources, and that fishes who are better adapted to the food source obtain proportionally more energy from the food source.

Two different types of experiment were conducted by the author, on this simulation. The goal of the experiments was to verify the changes occurring in the simulated population are really speciation. The system was run for 4000 time steps (around 770 years). Phenotype of the final population was extracted from the simulation trace, and visualized by conversion to an xfig format file. In order to confirm speciation, it was necessary to show that individual fish are not randomly assorted by the groups, but that there is a strong covariance (or high mutual information) between members in the groups. The goal of the second experiment was to examine a competitive exclusion principle, which states that two species competing for the same resources cannot co-exist. The author states that, eventually one of the species will go extinct leaving the other species to dominate the environment.

Arbitrarily labeling groups for the three phenotypes (sexual, preference, and trophic): As, Bs, Ap, Bp, and At, Bt, and counting the membership of all eight combinations of groups given in the Table 4. The author claims that, all the results show clear separation of each of the phenotypes into two groups (of fish) consistent with speciation for the first experiment. He also claims that, the results of the second experiment show the possibility that species that have lost the ability to adapt (due to inbreeding among small populations) may take

<table>
<thead>
<tr>
<th></th>
<th>ApAs</th>
<th>ApBs</th>
<th>BpAs</th>
<th>BpBs</th>
</tr>
</thead>
<tbody>
<tr>
<td>At</td>
<td>96</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Bt</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>136</td>
</tr>
</tbody>
</table>

Table 4 [Clement 2003b], page 7.
much longer to exclude other species from an environment.

The author claims that, the (virtual) fish simulated in TDLP clearly undergo a process that closely parallels speciation in the real world, producing species groups that are differentiated in terms of diet, and have developed a method (emergent covariance between color and preference) of maintaining different species within a single location. The author also mentions that the investigation of the effect on adaptability on competitive exclusion suggests that the ability to adapt increases the likelihood of competitive exclusion occurring. Adaptive organisms are mentioned to be capable to evolve, towards a global optima for harvesting a number of food sources, forcing out other species that might also gravitate towards the same optima. However, the author mentions that, if the species (fish) are initially specialists in different food sources, and cannot adapt, then competitive exclusion will not occur. The author claims that using a real natural system as the basis for TDLP has many advantages, e.g. the number of parameters and model types that need to be considered can be reduced to only those that apply to the natural system in the question.

This paper is cited by Clement [2003a] and Clement [2003c].

5.2 Simulating Species Richness Using Agents with Evolving Niches, With an Example of Galápagos Plants
Bonne [2010], believes that, in order to improve our understanding of ecosystems’ structure and investigate the probable changes in niches and speciation, it is essential to use simulation models, which are flexible and can be easily made and also allow biologists to analyze the results which are difficult to obtain in natural ecosystems.

The author refers to Sadedin [2005], as a previous work on speciation simulation. The author states that for decades, field work and analytical modeling have been the only sources available for investigation of community structure and change, while recently, agent-based simulations have been used to emulate changes in niches and speciation such as the model introduced by Sadedin [2005]. The author does not represent more details about the Sadedin [2005].

In this paper, the author tries to review potential uses of an agent-based representation of evolving niches on 22 Galápagos Islands, Ecuador, as a test-bed for an exploration of the utility of evolutionary computation in biogeography research. The author programmed an agent-based model in FORTRAN, and conducted simulations on a local Linux multiprocessor cluster. The simulation is claimed to be spatially explicit, where agents are plant-mono-cultures defined by three niche dimensions, lava (yes or no), elevation, and slope. According to the author, niches are represented as standard normal curves subjected to selection pressure, where neighboring plants breed if their niches overlap sufficiently, are considered the same species, otherwise they are different species. Plants that breed produce seeds with mutated niches. Seeds disperse locally and longer distances, and establish if the habitat is appropriate given the seed’s niche. From a single species colonizing a random location, hundreds of species evolve to fill the islands. The Galápagos Islands and ocean of the archipelago are represented as cells of the model. The plants are assumed as agents and represented by three niche dimensions: two are unimodal based on elevation and slope, and a third is whether the land is soil or unsuitable lava. A single species is modeled to colonize a randomly selected patch of soil, and in a repeating cycle, neighboring plants can breed if their niches overlap sufficiently and produce seeds with niches that include slight mutations. Seeds also are modeled to disperse in a random direction with a declining probability with distance, and if the destination habitat is suitable, a seedling is established. With a one-third chance, a mutation value would be added to the niche dimension, and with an equal chance the value would be subtracted. In rare cases (1
in 100), niches are mutated by larger values to speed the adaptation of plants and increase the likelihood that unique habitats would be colonized and may be thought to include hybridization, an important source of variation in island populations.

The author used a digital elevation model produced from the Shuttle Radar Topography Mission of 2000 to characterize elevation and slope on the islands which was composed of cells representing $92.5 \times 92.5\text{m}$ of land or ocean to simulate the model’s cells consisting of 3499 columns and 3884 rows. The attributes of each plant location and its niche dimensions are stored in a file, which is processed by a Visual Basic program and is used for confirming simulation behavior.

Figure 8, shows a schematic demonstrating plant species niche spaces, showing plants in which with niches that overlap sufficiently in niche dimension are considered the same species and are able to breed. Plants with niches that do not overlap sufficiently are separate species. Figure 9, represents modeled native plant species richness plotted against observed richness for 22 islands of the Galápagos Archipelago. A regression line provides reference ($r = 0.957$, 60 simulations).

Boone [2010], claims that the evolved plant species richness agrees very well with observed plant species richness in the nature except on Isabela and Santa Cruz Islands where their simulation richness did not agree as well with observed richness as for the remaining islands. In the simulations, plants spread quickly from the single pair that colonized the Galápagos Islands, and secondary colonies appeared on neighboring islands through direct and stepping-stone dispersal. All the available habitats on the islands filled with plants within the generations modeled. The author claims the results show increasing the overlap parameters for elevation and slope, increased the number of species modeled and reduced the agreement of simulated and observed richness and reducing the habitat suitability coefficient, reduced the number of seedlings being established and ultimately, the number of species evolved, and the agreement between observed and simulated richness declined.

This paper is not cited by other papers in this survey.

6. CONCLUDING COMMENTS

This survey reviewed ten papers, which were simulation studies of sympatric speciation using the individual-based models, including six journal papers, two conference papers and two un-published researches. The major contribution of these papers and their relations are presented in the Table 5.
<table>
<thead>
<tr>
<th>Year</th>
<th>Authors</th>
<th>Title</th>
<th>Papers referred to</th>
<th>Major contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003</td>
<td>Luz-Burgoa et al</td>
<td>Computer Simulation of Sympatric Speciation with Penna Ageing Model III The Penna model with phenotype selection</td>
<td>none</td>
<td>Added a simple interpretation of one locus in the genome bit-string as an identifier of species that led to the flourishing and co-existence of two separate populations.</td>
</tr>
<tr>
<td>2005</td>
<td>Luz-Burgoa et al</td>
<td>Computer simulations of sympatric speciation in a simple food web</td>
<td>Luz-Burgoa et al 2003</td>
<td>Investigated Sympatric Speciation with manipulating the resources and monitoring the changes in the population. Competition for the resources and sexual selection are assumed as the dominant forces acting on the population.</td>
</tr>
<tr>
<td>2007</td>
<td>Kampis and Gulyás</td>
<td>Evolutionary technology and phenotype plasticity: the FATINT system.</td>
<td>none</td>
<td>The presented speciation process can produce new functions and can go on indefinitely using different mechanisms for phenotype change as genetic point mutations or behavior change.</td>
</tr>
<tr>
<td>2008</td>
<td>Borkowski</td>
<td>Cellular automata model of macroevolution</td>
<td>Luz-Burgoa et al 2003, Borkowski 2009</td>
<td>The genotype’s bit-string is divided to two 8 bit masks, representing the ability to acquire resources, and the ability to defend from exploitation, respectively. The model also does not have a fix number of “species”</td>
</tr>
<tr>
<td>2009</td>
<td>Borkowski</td>
<td>Simple lattice model of macroevolution</td>
<td>Luz-Burgoa et al 2003</td>
<td>The author added some more detailed study concentrated on robustness of his model dynamics, effects of the lattice size and autotrophs productivity to his previous model[Borkowski 2009].</td>
</tr>
<tr>
<td>2009</td>
<td>Gras et al</td>
<td>An individual-based evolving predator-prey ecosystem simulation using a fuzzy cognitive map as the behaviour model.</td>
<td>none</td>
<td>Used Fuzzy Cognitive Map, for the first time, as the behavioural model of the individuals.</td>
</tr>
<tr>
<td>2009</td>
<td>Ten Tusscher and Hogeweg</td>
<td>The role of genome and gene regulatory network canalization in the evolution of multi-trait polymorphisms and sympatric speciation.</td>
<td>none</td>
<td>For the first time, investigated the effects of nonlinear genetic architecture on the evolutionary process.</td>
</tr>
<tr>
<td>2010</td>
<td>Boone</td>
<td>Simulating species richness using agents with evolving niches, with an example of galápagos plants.</td>
<td>none</td>
<td>Modeled the evolution of the plants species richness patterns on 22 Galapagos Island, Ecuador.</td>
</tr>
<tr>
<td>2011</td>
<td>Murdock and Yaeger</td>
<td>Identifying species by genetic clustering.</td>
<td>Gras et al 2009</td>
<td>Described how a standard clustering algorithm is used to discover and characterize distinct subpopulations in a common environment.</td>
</tr>
</tbody>
</table>

Table 5: The papers reviewed in this survey.

During the survey, it was observed that, using a Fuzzy Cognitive Map can be the best approach for implementing the genome of the individuals, as it allows the simulation of the internal concepts as well as other physical concepts, where the role of internal concepts in the evolutionary process, is a subject that could not be ignored.

Some future work are suggested by some of the authors of the papers. Clement [2003] states that the future work of his investigations, in general, can be improving the biological accuracy of the simulation, improving the visualization of the results, and also continuing the investigating conjectures concerning cichlid biology. Kampis and Gulyás [2007], claim that their presented works signify first steps about the relationship to the distant goal of evolutionary technology, and their results imply that fat phenotypes may
play a key role in pushing speciation forward under specific conditions. They state that a next step in their investigation would be to extend these results to functional differentiation and to the study of body plans using evolvable genotype-phenotype maps. In Borkowski [2009]’s model, when the availability of free space becomes a main constrain, only the mean number of agents can grow slightly more. The author believes, such an effect is inherited from the cellular automata paradigm and is an important limit of this version of his model. Borkowski [2009] suggests a more flexible solution to be invented, for the next iteration of the CO–EVO model, which may improve conformity of the model with real data and allow dealing with more actual questions. Murdock and Yaeger [2011] state that, as a future direction they can apply their clustering methods to simulations with clearly differentiated niches that are geographically either overlapping or isolated, to distinguish and quantify the relative effects of niche specialization versus geographic isolation.

ACKNOWLEDGMENTS
I would like to gratefully thank Dr. Richard Frost for his help and detailed comments on organizing the structure of the survey, and Dr. Robin Gras for his support and advice and also the faculty of computer science of University of Windsor for preparing suitable environment to have access to academic resources and scientific papers.
REFERENCES


