2016

Investigation of sympatric speciation as the outcome of competition for food resource by means of an individual-based modeling approach

MARYAM KARIM POUR
University of Windsor

Recommended Citation
KARIM POUR, MARYAM, "Investigation of sympatric speciation as the outcome of competition for food resource by means of an individual-based modeling approach" (2016). *Electronic Theses and Dissertations*. 5908.
https://scholar.uwindsor.ca/etd/5908

Investigation of sympatric speciation as the outcome of competition for food resource by means of an individual-based modeling approach

By

Maryam Karim Pour

A Thesis
Submitted to the Faculty of Graduate Studies
Through the Great Lakes Institute for Environmental Research
In Partial Fulfillment of the Requirements for
The Degree of Master of Science
At the University of Windsor

Windsor, Ontario, Canada
2016

Investigation of sympatric speciation as the outcome of competition for food resource by means of an individual-based modeling approach

by

Maryam Karim Pour

APPROVED BY

Dr. Boubakeur Boufama
School of Computer Science

Dr. Trevor Pitcher
Department of Biological Sciences &
Great Lakes Institute for Environmental Research

Dr. Robin Gras, Advisor
School of Computer Science

December 6, 2016

DECLARATION OF CO-AUTHORSHIP

I hereby declare that this thesis incorporates material that is result of joint research, as follows:

Some parts of the second chapter of this thesis comes from a research done by Sara Bandehbahman under the supervision of Dr. Robin Gras. Data analysis and interpretation were performed by the author.

I am aware of the University of Windsor Senate Policy on Authorship and I certify that I have properly acknowledged the contribution of other researchers to my thesis, and have obtained written permission from each of the co-author(s) to include the above material(s) in my thesis.

I certify that, with the above qualification, this thesis, and the research to which it refers, is the product of my own work.

I declare that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my thesis.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office, and that this thesis has not been submitted for a higher degree to any other University or Institution.

ABSTRACT

Sympatric speciation, the emergence of new species in the absence of geographic isolation, is one of the most controversial issues in evolutionary biology. Although today the plausibility of the occurrence of sympatric speciation is theoretically acknowledged, its underlying mechanisms are still unknown. We applied a modeling approach with three trophic levels (primary resource, prey, and predator) and supplied prey species with two different food resources (Food 1 and Food 2) to track prey lineage through evolutionary time to detect any indicators of the occurrence of sympatric speciation caused by specialized food consumption. Whereas, Food 1 was the more available resource, Food 2 had higher energy content. Initially, when there was not yet any specific food specialization, Food 1 consumption rate was significantly higher compared to Food 2. Eventually, around time step 22,000 and after the emergence of food consumption specialization, the exploitation of Food 2 was higher than Food 1 in spite of the fact that prey individuals were more frequently encountered with Food 1 than Food 2. Drawing a comparison between simulations with only one food resource and simulations with two available food resources revealed that complete reproductive isolation caused by disruptive selective pressure exerted by adaptation to different resources plays a curial role in the emergence of sympatric species. Machine learning techniques were also employed to identify the shared patterns among sympatric species. Results showed that for most lineages sympatric divergence has occurred at the beginning of the process of the emergence of specialized use. If not, these species have possessed a high spatial distribution and had to meet two conditions to be diverged sympatrically: i. high genetic diversity and ii. a large population size.

ACKNOWLEDGEMENTS

I would like to thank my supervisor Dr. Robin Gras. I also am grateful to my other committee members: Dr. Trevor Pitcher and Dr. Boubakeur Boufama. I would like to acknowledge the help of Dr. Ken Drouillard and Dr. Christina Semeniuk. Additionally, I would like to specially thank Sara Bandehbahman. Acknowledgements are also given to the faculty and staff at the Great Lakes Institute for Environmental Research (GLIER), especially Mary Lou Scratch and Christine Weisner. And finally, a special thanks is given to Nancy Barkley, Candy Donaldson, and Marina Beaudry.

TABLE OF CONTENTS

LIST OF TABLES

Chapter 2

Appendix

LIST OF FIGURES

Chapter 2

Appendix

CHAPTER 1

INTRODUCTION

1.1 Speciation and Mechanisms of Divergence

One the most essential evens in the history of life is speciation, which has happened billion times since life began 3.8 billion years ago (Herron and Freeman 2013; Stearns 1992). The origin of life has been one of the most controversial topics in biology (Bolnick and Fitzpatrick 2007). Speciation is defined as “the evolution of reproductive isolation between two populations” (Ridley 2004), and species is “the smallest evolutionarily independent unit” (Herron and Freeman 2013). Speciation is also defined as “the evolution of genetically distinct populations (clusters), maintained by reproductive isolation in the case of sexual taxa” (Bolnick and Fitzpatrick 2007). Evolution, alterations in allele frequencies across generations, is caused by evolutionary forces (selection, mutation, migration (gene flow), and genetic drift). Evolutionary independence happens when these evolutionary forces function on population separately (Herron and Freeman 2013). In other words, “species form a boundary for the spread of alleles” and as a result, each species follow its own evolutionary path (Herron and Freeman 2013). Essentially, the lack of gene flow and reproductive isolation are the central event in the process of forming a new species and the speciation process initiates when populations are genetically isolated (Herron and Freeman 2013; Ridley 2004). Therefore, species is defined as “interbreeding populations that evolve independently of other populations” (Herron and Freeman 2013). All member of one species genetically, ecologically, morphologically, and behaviorally differentiate from the members of other species. Thus, they only could interbreed with the members of their own species and the evolution of a barrier to interbreeding between two populations could cause one species to diverge into two separate species (Ridley 2004; Stearns 1992). Generally, reproductive isolation might occur through two main mechanisms: i. premating or prezygotic isolating mechanisms such as ecological or habitat isolation, seasonal or temporal isolation (different mating season), and sexual isolation that inhibit the formation of hybrid zygotes; and ii. postmating or postzygotic isolating mechanisms such as hybrid inviability and hybrid sterility that lower the viability or fertility chance of hybrid zygotes (Ridley 2004).

Speciation is conceptualized as a three-stage process: populations isolate in the first step; then divergence in traits (e.g. habitat use or mate choice) will happen through the second step; and finally populations become reproductively isolated at the third step. During the first step gene flow is disrupted and populations become isolated as a result of a barrier such as physical isolation (e.g. dispersal or vicariance) or mutation (e.g. polyploidy or chromosome changes) (Herron and Freeman 2013). The first step generates a condition for speciation; however, in order to have a continuous event the operation of genetic drift and natural selection on population is also necessary to create divergence in isolated population (Herron and Freeman 2013). Genetic drift that leads to random fixation or loss of alleles is more pronounced in small populations. Considering this fact that populations become smaller during the speciation process as a result of the stopped or reduced gene flow, genetic drift , therefore, plays a crucial role in the speciation process (Herron and Freeman 2013). Nevertheless, the role of genetic drift in the speciation process has been largely taken up with controversy and it has been asserted that genetic drift can only effectively contribute to this process if population is extremely small and remains small for a long period of time (Grant, Grant, and Deutsch 1996; Lande 1980, 1981). Whereas, natural selection is significantly recognized as the most important parameter that promotes the divergence of a new species from their ancestral population, when a portion of the original population starts inhabiting a new food resource or a new environment (Feder et al. 1997; Feder, Chilcote, and Bush 1988, 1990; Filchak, Roethele, and Feder 2000; Gras, Golestani, Andrew P Hendry, et al. 2015; Hendry and Kinnison 2001; Nosil, Crespi, and Sandoval 2002; Rundle et al. 2000). The significant role of natural selection in speciation has been illustrated by concrete empirical evidence described in a comprehensive meta-analysis done by (Funk, Nosil, and Etges 2006). They looked at hundreds of species of birds, frogs, fish, insects, and plants to investigate the occurrence of reproductive isolation throughout the evolutionary divergence from an ecological perspective. A significant correlation between ecological divergence and reproductive isolation was observed in more than 500 species pairs (Funk, Nosil, and Etges 2006).

In addition to genetic drift and natural selection, sexual selection, which works on individuals' capability to acquire their potential mates, could also facilities the process of divergence (Fisher 1958; Higashi, Takimoto, and Yamamura 1999; Lande 1981; Panhuis et al. 2001; Shaw and Lugo 2001; Uy and Borgia 2000). In summary, genetic drift through fixation or loss of specific alleles that do not function properly as heterozygote; natural selection through production of adaptation to specific conditions; and sexual selection through alteration in mating system could lead to the emergence of a new species, while produced hybrid offspring possess a remarkably reduced fitness (Herron and Freeman 2013). It means that there is a selection pressure acting against hybrid individuals (through reduced survival and fertility) and consequently, reduce their frequencies in the population. This selection force is called reinforcement. Reinforcement is the last stage of speciation, which finalizes the speciation process through the formation of a complete reproductive isolation (Coyne and Orr 1997; Herron and Freeman 2013; Higgie, Chenoweth, and Blows 2000). On the other hand, hybridization itself could results in the emergence of a new species when diverged populations produced fertile and viable hybrid offspring that unlike their parental species are able to occupy a novel habitat and indicate a higher fitness in the new habitats. As a result, eventually a distinct third species could emerge (Herron and Freeman 2013; Rieseberg et al. 1996).

Theoretically, there are three different geographic relations between a new evolving species and its ancestor. Allopatric speciation happens when a new species evolves in geographic isolation from its ancestor. In other words, in allopatric divergence or geographical speciation, new species gradually are formed from geographically isolated populations of the same ancestral gene pool (Coyne 1992; Mayr et al. 1963; Rice and Hostert 1993; Ridley 2004). Parapatric speciation occurs when "the new species evolves in a geographically contiguous population". Parapatric speciation is one of the rare forms of speciation, in which reproductive isolation happens because of temporal and behavioral reasons rather than geographic causes. Unlike allopatric speciation in which the population of one particular species is split into two separate subpopulations by a physical barrier, in parapatric speciation a subpopulation of one specific species becomes genetically isolated as a result of occupying a new niche (Bank, Bürger, and Hermisson 2012). By far the most controversial form of speciation is sympatric speciation, which happens when one single species (ancestral species) splits into two or more groups of individuals that become unable to reproduce with each other, although there is no geographical isolation or extrinsic barrier to gene flow (Berlocher and Feder 2002; Coyne 2007; Ridley 2004). In other words, sympatric speciation happens when a new species emerges "within the geographic range of its ancestor" (Ridley 2004). Basically, we can say that in the speciation process there is a continuum geographic constraint preventing interbreeding from zero in sympatric to complete in allopatric (Fitzpatrick, Fordyce, and Gavrilets 2008). In other words, from population genetic point of view sympatric speciation is considered as the most extreme lineage-splitting event that happens without presence of any physical barriers preventing gene flow (Bird et al. 2012; Gavrilets 2003; Kautt, Machado-Schiaffino, and Meyer 2016). Although there are different mechanisms of speciation, most scholars agree that the vast majority of species have been initiated through "allopatric speciation" (Coyne 1992; Mayr 1963; Rice and Hostert 1993).

1.2 Ecological speciation

Initially, Simpson (1955) argued that ecological conditions play a central role in lineage diversification; for example, when organisms switch into a new food resource or habitat (Simpson 1955). Recently, this claim has again caught the attentions of scientists and they recast it as "ecological speciation". Ecological speciation adopts a broad perspective and combines all different modes of speciation (allopatry, sympatry, and parapatry). It defines speciation as a lineage-splitting event resulting from the emergence of reproductive isolation caused by the function of divergent natural selection acting on traits between populations or subpopulation in conflicting environment (Feder and Forbes 2007; Funk, Egan, and Nosil 2011; Funk et al. 2006; Karpinski et al. 2014; Schluter 2000, 2001; Yoder et al. 2010). Biotic and abiotic factors of habitat are defined as "environment", which could be physical structure of habitat, food resources, and climate or ecological interactions between individuals such as predation and resource competition (Schluter 2001). Basically, ecological speciation happens "when barrier to gene flow (reproductive isolation) evolves between populations as a result of ecologically-based divergent selection" (Rundle and Nosil 2005). Selection acting on populations in opposite directions is considered as divergent selection (Rundle and Nosil 2005).

Traditionally, speciation modes of action are categorized based on the geographical relations of populations that are experiencing speciation event (allopatric, sympatric or parapatric). However, according to ecological speciation hypothesis, speciation “might occur in allopatry or in sympatry” (Schluter 2001). The occurrence of ecological speciation has been demonstrated through experimental observation (Rice and Hostert 1993) and also through field studies (Coyne and Orr 2004; Rundle and Nosil 2005).

The initial components required for the occurrence of ecological speciation process is “an ecological source of divergent selection, which could be differences in environment or niche, certain forms of sexual selection, and the ecological interaction of populations” (Rundle and Nosil 2005). For instance, when populations inhabit different environments or exploit different resources, they experience disruptive natural selection and eventually indicate adaptation to different environments. This could potentially lead to the evolution of barrier to gene flow among populations and therefore, the occurrence of ecological speciation (Rundle and Nosil 2005; Schluter 2001). Reproductive isolation resulted from adaptation to different environments has been observed both in sympatric (Rice and Salt 1990) and allopatric species (Rice and Hostert 1993). Generally, environmental variances is considered as one the important sources of divergent selection (Schluter 2000). Ecological interaction among living organisms is another source of divergent selection in nature, which particularly plays a central role in sympatric speciation (Rundle and Nosil 2005; Turelli, Barton, and Coyne 2001).

1.3 Sympatric Speciation

Sympatric speciation happens when one lineage is split into two new separate species without any geographical separation in ancestral species. According to the majority of models describing sympatric speciation, the initial step in sympatric divergence is polymorphism developed by natural selection and the next step is prezypotic isolation between different morphs (reinforcement process). For example, sympatric speciation happens when reinforcement process isolates two different morphs feeding on two distinct food resources that are not able to reproduce fertile and viable hybrid offspring. Unlike parapatric speciation, for the occurrence of sympatric speciation initial polymorphism does not need to be spatial polymorphism that is spread through the space within population (Ridley 2004).

It has been empirically demonstrated that there are two particular circumstances easing the occurrence of sympatric speciation as an evolutionary process in nature: genetic conditions and ecological conditions (Bolnick and Fitzpatrick 2007; Via 2001). Genotype × environment interaction in resource use and genetic variation in habitat preference are two main examples of genetic conditions facilitating sympatric speciation (Via 2001). Examples of ecological conditions leading to sympatric speciation include: i. habitat or host shift in sister species utilizing diverse habitat or host (host refers to what provides nourishment for an organism), ii. ecological opportunity for adaptive radiation in isolated environments such as small lakes or islands (Via 2001) (adaptive radiation occurs when individuals of a single population quickly branch off into several new forms as a result of a new change in the environment that provide environmental niches or new resources or new challenges (Larsen and Repcheck 2008; Schluter 2000)), and iii. imposed constraint on gene flow between populations as a result of the absence of an intermediate environment that supports hybrids (resulting in an ecological selection force against hybrids) (Via 2001).

Almost all sympatric speciation models follow a unique general outline. As such, disruptive selection in an initial random mating population leads to evolutionary changes in mating patterns in all models and this, in consequence, contributes to reproductive isolation in subpopulations of the initial population (Bolnick and Fitzpatrick 2007; Kirkpatrick and Ravigné 2002). Competition for shared resources (Bolnick and Smith 2004; Pfennig and Pfennig 2010; Roughgarden 1972), adaptation to different resources (Martin and Pfennig 2009; Wilson and Turelli 1986), and unequal distribution of resources throughout the environment (Hendry et al. 2009; Schluter and Grant 1984) are the underlying factors that could result in disruptive selective pressure. In addition to disruptive selection, other evolutionary factors play a leading role in sympatric speciation including sexual selection (van Doorn, Edelaar, and Weissing 2009; Maan and Seehausen 2011), competition, and habitat preference (Bolnick and Fitzpatrick 2007). In fact, it is believed that the sympatric speciation process stems from several fundamental causes including reproductive and behavioral strategies (Thibert-Plante and Hendry 2011). Among these, sexual selection that forces mate choice and habitat competition which leads to preferential resource use are the most popular among literature (Thibert-Plante and Hendry 2011).

Darwin (1859) successfully developed the concept that natural selection could eventually lead to species divergence. Sympatric speciation had been widely accepted by scientists until the early 1960’s when it became a divisive issue. In 1963, Mayr argued against sympatric speciation and proposed that allopatric speciation is the prevalent type of speciation. Since then many investigators such as Smith (1966) (by his simple model), and most significantly Rice (by his empirical and theoretical studies) (Rice and Salt 1990; Rice 1984, 1987) have striven to prove that disruptive selection could frequently lead to sympatric speciation. Today, thanks to a large number of empirical observations and mathematical models, it is generally acknowledged that sympatric speciation is theoretically possible and has occurred in nature. However, the underlying mechanism for it has remained unclear and controversial (Bolnick and Fitzpatrick 2007; Coyne and Orr 2004; Gavrilets 2004; Turelli et al. 2001; Via 2001). After attesting to the theoretical feasibility of sympatric speciation, its central underlying mechanism has become the main source of controversy today and much uncertainty still exists. However, exploring underlying causes of sympatric speciation by means of empirical studies is difficult (Bolnick and Fitzpatrick 2007).

In summary, there was initially a considerable uncertainty about the feasibility of the occurrence of sympatric speciation (Mayr et al. 1963; Tauber and Tauber 1989); however, today this hesitation has been fundamentally resolved and no longer exists (Bush 1994; Dieckmann and Doebeli 1999; Gavrilets and Waxman 2002; Via 2001). Necessary prerequisites for the occurrence of sympatric speciation that were primarily defined as extremely unachievable (Bush and Howard 1986; Bush 1993), today are considered as a more readily achievable set of conditions (Doebeli 1996; Gavrilets 2004; Tregenza and Butlin 1999). Today there are a substantial number of good examples (described in section 1.3.1) that indicate sympatric speciation is hypothetically more likely to happen than allopatric and parapatric speciation (Barluenga et al. 2006; Berlocher 1998; Savolainen et al. 2006; Sorenson, Sefc, and Payne 2003).

It is believed that speciation event is allopatric, if reproductive isolation is completed long before secondary contact. But it is considered as sympatric, if there is still some sort of reproductive isolation at the time of secondary contact (Rundle and Nosil 2005). However, it is very challenging to determine whether a new species has been originated through sympatric or allopatric speciation. Phylogenetic test can be applied to examine whether speciation is allopatric or sympatric. For instance, numerous species of cichlid fish have emerged in the East African lakes (Schliewen, Tautz, and Pääbo 1994), and it has been evidenced that these species have arisen through sympatric speciation (Elmer et al. 2010; Kautt, Elmer, and Meyer 2012; Malinsky et al. 2015). But how can we verify if these species have evolved allopatrically or sympatrically? Phylogenic test has revealed that African cichlids have emerged sympatrically. If the new specie occupies a geographically different habitat compared to its ancestral species (separate lake in this case), then this species has diverged allopatrically. But if the new species and its ancestor live in the same habitat (the same lake in this example), this species has arrived sympatrically (Barraclough and Vogler 2000; Berlocher 1998, 2000; Bush and Smith 1998; Via 2001). In another example of application of this method, it has been proved that whereas *pomonella* group of *Rhagoleties* has diverged rapidly through sympatric speciation, the congeneric *suavis* group, which do not indicate any signs of host shift, has slowly split through allopatric speciation (Bush and Smith 1998).

1.3.1 Well-documented Examples of Sympatric Speciation Identified Through Field and Laboratory Studies

Insect species feeding on different host plant species mainly provide a strong well documented evidence for the existence of sympatric speciation via host shift (Bush 1969; Diehl and Bush 1984; Drès and Mallet 2002; Via 2001). Moving from one host to a new one would initiate ecological speciation in the absence of geographic isolation and would eventually lead to the formation of adaptation to specific ecological niches, and thereby developing sympatric speciation (Price 1975; Soudi, Reinhold, and Engqvist 2016). Apple and hawthorn flies (maggot fly, *Rhagoleties pomonella*) are considered as the best examples to clearly illustrate the process of host shift and divergence that has happened due to the function of natural selection on preference for different food resources (Herron and Freeman 2013; Ridley 2004). *R. pomonella* are considered as the pest of apple and hawthorn fruits. Hawthorn tree and *R. pomonella* are native to North America. After introducing apple trees to this region in 1864, this species was observed for the first time exploiting apple fruits as the food resource. It seems that *R. pomonella* have moved to a new food resource (host shift) more than 200 years ago. They shifted from hawthorn fruits to apple fruits. This happened while *R. pomonella* were sharing their habitat with hawthorn flies and this shift led to reproductive isolation resulting from an incompatible mating time and habit choice. This host shift from hawthorns to apples was considered the initial step toward sympatric speciation (Bolnick and Fitzpatrick 2007; Feder and Filchak 1999; Linn et al. 2003). It has been demonstrated that *R. pomonella* exploiting different hosts are indeed different genetic races with assortative mating. They have indicated significant differences in their allele frequencies for six different enzymes. It means that host-related adaptation caused by natural selection has developed a strong divergence between *Rhagoleties pomonella* populations and as a result, they have split sympatrically by host shifts (Coyne and Orr 2004; Feder et al. 1997, 1988, 1990; Filchak et al. 2000; Herron and Freeman 2013; Hood et al. 2013; Ridley 2004).

Numerous examples of host-plant shifts in insect sister species have now been traced in nature (Berlocher 1999; Claridge, Dawah, and Wilson 1997; Prowell, McMichael, and Silvain 2004; Sezer and Butlin 1998; Wood and Keese 1990). This sympatric host-shift speciation is not simply limited to insect species. Several instances among vertebrate species has been also documented (Bolnick and Fitzpatrick 2007). For instance, African indigobird of the genus (*Vidua*) act as brood parasites of different species (their hosts). Mimicking the host's courtship songs, male indigobirds manipulate their hosts into raising their offspring. It has been proven that the preparation for reproductive isolation and accordingly, the emergence of a genetically new species of indigobirds is started as soon as a new host species has been selected by indigobirds (Sorenson et al. 2003). Intermediate horseshoe bats (*Rhinolphus affinis*) and Pearson's horseshoe bats (*Rhinolphus pearsonii*) are also considered as a species having arisen from a sympatric speciation event. Investigations have illustrated that although these carnivorous bat species have an overlapped diet, they also have their own exclusive prey species. Therefore, Intermediate horseshoe bats and Pearson's horseshoe bats perfectly coexist in cave ecosystems without any competitive interactions due to their different preferential foraging specializations, thereby occupying diverse microenvironments of the cave ecosystem (Jiang et al. 2008).

Furthermore, there are two different species of three-spined sticklebacks living in small lakes of coastal British Columbia that have diverged sympatrically more than 13,000 years ago. These species possess different morphologies that are closely related to the habitat and food resources that each species exploits; one species is limnetic with a smaller mouth and the other one is a benthic species with a larger mouth. It has been demonstrated that competition for food supply has played a central role in the evolution of these sister species evolving sympatrically (Boughman, Rundle, and Schluter 2005; Rundle and Schluter 2004; Rundle et al. 2000; Schluter and McPhail 1992; Svanbäck and Schluter 2012). Different sympatric stickleback species show a significant variation in terms of their body size. This is proven to be caused by disruptive natural selection pressure exerted by exploiting diverse ecological habitats (Nagel and Schluter 1998).

Heliconiine butterflies sister species (Heliconius butterflies and their close allies (Lepidoptera: Nymphalidae: Heliconiini)) show a comprehensive geographical overlapping distribution. It has been evidenced that a substantial number of these sister species has evolved through sympatric speciation. In total, 32 to 95 percent of speciation events in this lineage has been sympatric speciation (Rosser et al. 2015). "Sister group is defined as a single species or a monophyletic group that is the closest genealogical relative of another single species or monophyletic group of species" (Wiley and Lieberman 2011). Sister species have a shared ancestral species (Wiley and Lieberman 2011), and a taxon containing two or more species including ancestral species and all of its descendants is considered as a monophyletic group (Wiley and Lieberman 2011).

Fruit Doves (genus: *Ptilinopus*, family: Columbidae) have more than 50 species that inhabit Pacific Ocean islands. Among them, there are two sister species (Red-moustached Fruit Dove, *Ptilinopus mercierii*, and the White-capped Fruit Dove, *Ptilinopus dupetithouarsii*) living in Marquesas Islands located at eastern Polynesia that have diverged sympatrically (Cibois et al. 2014).

Nesospiza buntings living in the South Atlantic Tristan da Cunha archipelago islands (Ryan et al. 2007), *Geospiza* finches in the Galápagos archipelago (Grant and Grant 2010; Huber et al. 2007), and *Oceanodroma* petrels living in in eastern Atlantic islands (Friesen et al. 2007) are other examples of the bird species that have sympatrically evolved.

Among mammals, blind subterranean mole rats (genus Spalax) living in northern Israel, which was previously categorized as allopatrically or peripatrically emerged. But recently it has been evidenced that this species has probably evolved through sympatric speciation caused by ecological adaptation to different soil types existing in their geographic distribution (Hadid et al. 2013). Subterranean Rodents foraging for underground food supplies are largely influenced by physical characteristic of the soil they burrow in to search for their food resources. This, in consequence, could lead to the emergence of adaptation to different soil types and thereby, speciation (Lövy et al. 2015).

1.3.2 Theoretical, Mathematical and Individual-based Models and Sympatric Speciation

“Sympatric speciation, the divergence of one evolutionary lineage into two in the absence of geographic isolation, has a long history”(Berlocher and Feder 2002). The process of exploration of sympatric speciation was initiated with the theoretical analysis suggested by Maynard Smith in 1966 (Smith 1966) who emphasized the critical role of disruptive selection in sympatric races (Grant and Grant 2010). Today’s hypotheses addressing sympatric speciation consider a significant role for natural selection pressure exerted by ecological factors (Rashkovetsky et al. 2015). Theoretical models that have been developed to investigate sympatric speciation mainly focus on the fundamental role of disruptive natural selection in the lineage-splitting event (Dieckmann and Doebeli 1999; Doebeli 1996; Kawecki 1997; Kondrashov and Kondrashov 1999; Tauber and Tauber 1977).

Felsenstein (1981) developed two different models, one-allele and two-alleles models, to examine how habitat differentiation could lead to the emergence of sympatric species. According to his results, unlike two-alleles models sympatric speciation happens easily in one-allele model. It means that the occurrence of sympatric divergence was more difficult in his two-alleles model (Felsenstein 1981).

Kondrashov’s “polygenic models of sympatric speciation” (1983, 1986) was another evidence of the occurrence of sympatric speciation in nature (Kondrashov and Mina 1986;

Kondrashov 1983, 1986).

Rice (1987; 1990) modeled the central role of the action of disruptive selection either on habitat choice or on other traits in sympatric speciation (Rice and Salt 1990; Rice 1987).

Diehl and Bush (1989) modeled the process of shifting into new host and the functional role of habitat choice in reproductive isolation in sympatric races (Diehl and Bush 1989).

Doebeli and Dieckmann (2000) applied a classical ecological model to indicate the importance of selective pressure acting against intermediate phenotypes, which was resulted from different ecological interactions including mutualism, competition, and predation in evolutionary branching. They showed that the integration of population genetics and mating mechanisms into ecological models could offer a pragmatic model for testing the probability of the occurrence of sympatric speciation (Doebeli and Dieckmann 2000).

Thibert-Plante and Hendry (2009) utilized an individual-based model to investigate the importance of mate choice, dispersal, gene flow, and natural selection pressure acting against migration in speciation. In order to provide a better understanding of ecological speciation and its underlying factors, in this study they measured the required time for one population to inhabit a new ecological niche (Thibert-Plant and Hendry 2009). They found that natural selection pressure acting against migration and hybrids plays a crucial role in reproductive isolation, thereby affecting speciation. Additionally, according to this investigation, mating preference also made a substantial contribution to ecological speciation. Their modeling investigation demonstrated that when a subpopulation branched from the main population and occupied a new habitat, environmental differentiation between the new and the old habitat could quickly lead to reproductive isolation wherein the subpopulation completely separates from the ancestral population. They concluded that there is a nonlinear interaction between different parameters (fluctuating environmental parameters, population size, dispersal, and mating preference) contributing to speciation (Thibert-Plant and Hendry 2009).

They also carried out another individual-based modeling investigation in 2011 to examine the potential factors (including competition, mating preference, and resource distributions) influencing sympatric speciation. In this study, male foraging ability was the main parameter exploited by females for the purpose of choosing their potential mates.

Furthermore, the capacity of individuals to utilize available resources was based on their phenotype and this capacity was used to model competition. According to the results of this study, strong mate choice is a required criterion for the occurrence of sympatric speciation; however, it is not enough. The authors found that among these three factors contributing to sympatric speciation, mate choice and resource distribution are more important factors than competition. Finally, they concluded that models involving several potential factors at the same time are more capable of modeling sympatric speciation (Thibert-Plante and Hendry 2011).

Labonne and Hendry (2010) applied an individual-based model specifically designed for guppies, *Poecilia reticulate*, to investigate how the interaction between sexual and natural selective pressures could lead to ecological speciation. They explored the evolution of male color within 20 generations under two different situations, low and high predation pressure. Their results illustrated the significant evolution of a male trait, male coloration, caused by divergent selection. This modeling study proved that the consequences of divergent natural selection could be intensely adjusted through sexual selective pressure exerted by female mating preference. They therefore concluded that estimations of ecological speciation could be changed through sexual selection (Labonne and Hendry 2010).

Gras et al. (2015) utilized an IBM approach to explore the speciation process and the primary reasons for the emergence of new genetic clusters (species) under three different scenarios. Compact and distinct clusters were clearly detectable in the first scenario, where individuals were subject to natural selection as well as spatial isolation. By contrast, clustering was weaker in the second scenario (overlapping clusters), where individuals were only subject to spatial isolation but not selection. Finally, the third scenario, where there was no natural selection and spatial isolation but genetic drift alone, did not indicate any signs of clustering (Gras, Golestani, Andrew P Hendry, et al. 2015).

Applying the same tool, Golestani, Gras, and Cristescu (2012) investigated how introducing new physical obstacles to an artificial ecosystem could influence allopatric speciation through alterations in population distribution and the patterns of gene flow between subpopulations. They found that when building up the number of existing obstacles in their virtual world, the rate of speciation increases so that there is a continuous correlation between the number of obstacles and the speed of evolution. Their results also indicated that spatial distribution of existing species in their control runs (the virtual world without any obstacles) was significantly less compact than their treatment runs (physical obstacles included) (Golestani, Gras, and Cristescu 2012).

1.4 EcoSim and Sympatric Speciation

What is the importance of investigation of sympatric speciation? Verifying the occurrence of sympatric speciation could benefit us to explain the existence of a significant proportion of lineage diversification on Earth (Berlocher and Feder 2002). Furthermore, any attempt that provide further clarification on sympatric speciation could reveal the significant role of ecology in speciation (Orr and Smith 1998; Schluter 1998). Although there is no doubt today about the possibility of the occurrence of sympatric speciation, it is not clear yet to what extent sympatric speciation contributes to our planet biodiversity (Rosser et al. 2015). Sympatric speciation and the question "how could a new species evolve without geographic isolation?" are considered as the major unknowns in evolutionary biology (Kautt, Machado-Schiaffino, and Meyer 2016; Kautt, Machado-Schiaffino, Torres-Dowdall, et al. 2016). Additionally, it is still controversial how ecological interactions could lead to sympatric speciation through creating disruptive selection (Rundle and Nosil 2005).

Direct observation and following lineage through evolutionary time are introduced as the best strategies to evidence sympatric divergence since no one could deny such a strong proof (Berlocher and Feder 2002). Although a huge number of investigations have been launched to shed light on the origin of species, sympatric speciation has not captured enough attention due to theoretical difficulties (Coyne 1992; Rice and Hostert 1993). Tracking speciation in complex organisms by means of field studies and experimental observations, which are considered as the best method to document the occurrence of sympatric speciation, is quite difficult on the grounds that speciation is a gradual genetic divergence, which requires thousands of generations to occur (Berlocher and Feder 2002; Bolnick and Fitzpatrick 2007; Coyne and Orr 2004). Therefore, it would be essential to exploit the potential abilities of new techniques such as modeling approaches to overcome such difficulties and thus obtain further insights. For instance, Individual-Based Models (IBMs), which enable us to investigate thousands of generations through a reasonable time and cost, is considered as a fully functional tool that have been widely applied to simulate ecological systems in order to offer a better understanding of speciation (Grimm and Railsback 2013). Ecological system properties evolve from adaptive behavior of individuals forming the system. Therefore, in ecology it is essential to fully comprehend the relationship between emergent system properties and adaptive traits of individuals (Levin 2007). Functionally, in the field of adaptive behavior and emergent properties IBMs are considered as a pragmatic tool to investigate the evolving characteristics of targeted system (Grimm and Railsback 2013; MacPherson and Gras 2016).

Considering the capabilities of IBMs in the area of evolutionary biology, we utilized an individual-based modeling approach to acquire further ecological insights into sympatric speciation and its underlying reasons. We employed a complex individual based evolving predator-prey ecosystem platform called “EcoSim” (Gras et al. 2009) to look at preferential resource usage causing selective pressure toward sympatric speciation. We explored the speciation process in the absence of a pre-defined fitness function (Gras, Golestani, Andrew P Hendry, et al. 2015), where the capability of individuals to cope with environmental challenges (fitness) is determined thorough their interactions with their surrounding biotic and abiotic environments (an emergent property). Three different trophic levels have been included in this model and mobile prey and predator individuals follow a non-random movement strategy influenced by many different factors including, but not limited to, search for food, search for mating partner, socialization, evasion, and exploration. One of the unique feature of the present approach is that unlike other models that simply assume that living organisms only forage and do not have any other alternative activities, the current modeling study gives careful consideration not only to performing fragging behavior but also to all other activities that influence individuals’ fitness such as predator-inflicted mortality. Each action has its corresponding sub-model and the complex interactions between these sub-models determine individual’s movement patterns at any time step. Each individual is able to perceive its environment and also its surrounding biota. Then, based on these perceptions they choose one specific activity to perform. For example, in foraging behavior or mate-searching activity, individuals perceive the closest food resources or mating partner within their vision ranges and move towards these resources. The speed of each individual determines when this individual will reach the food recourses or mating partner. Prey individuals also perform a predator avoidance behavior called the evasion, which is the movement of prey individuals in the opposite direction of the closest predators within their vision range (Gras et al. 2009).

Employing such a complex modeling approach, this study mainly focuses on crucial role of food resource acquisition in sympatric speciation. In order to achieve this goal, herbivore prey individuals were offered by two different choices of primary food resources, grass, which differentiated from each other in terms of availability and amount of energy carried by each. In other words, a dual food resource version of EcoSim was developed to create the favorable conditions for the emergence of divergent feeding behavior. We called these primary food resources: Food 1 and Food 2. While Food 1 were more available all around the world, each unit of Food 2 contained higher amount of energy that could be exploited by prey individuals feeding on this food resource. Therefore, prey individuals could choose between diverse primary food resources with different properties. This, in consequence, could lead to the evolution of food consumption specialization on specific food resource with the purpose of achieving the optimal benefits in terms of reproductive success and net energy income. In other words, by providing two food resources that differ from each other (in terms of the rate at which each food resources is encountered and energy content of each food resources), it is expected to observe the divergence of lineage into separate new species in the absence of geographic isolations. The main focus of the present study was exploring how competition for habitat and ecological specialization could contribute to sympatric speciation. More specifically, in this study we investigate preferential resource usage causing selective pressure toward sympatric speciation. Individuals from a single population may choose to feed on two different food resources while they are living in the same habitat. Under a strong force positively selecting for this, the initial population might be split into two discrete subpopulations; each specialized on their own particular food resource. Disruptive selection can exert selective pressure against hybrid individuals with an intermediate feeding behavior trait. When selection favors individuals at only the extreme ends of a feeding trait, individuals will become specialized on divergent food resources. This, in consequence, leads to reduced fitness in individuals with an intermediate expression of the trait, resulting from an inefficient exploitation of food resources (Lu and Bernatchez 1999). For instance, compared to individuals with the extreme phenotype, hybrid individuals with intermediate phenotypes may experience a higher extent of resource competition as their exploitation of available resources is inefficient (Dieckmann and Doebeli 1999; Hatfield and Schluter 1999; Wilson and Turelli 1986). Generally speaking, when selection favors extreme values of a specific trait against intermediate values of this trait and diverges the initial population into two distinct subpopulations of extremes, individuals with the intermediate value will be ultimately eliminated (Lu and Bernatchez 1999). Thus, compared to extreme morphs that tend to be the more functional phenotype, intermediate ones suffer a lower fitness (Griffith 1996). Reproductive isolation may occur either because of assortative mating (as individuals feeding on one particular food resource tend only to mate with each other) or due to a reduced probability of successful mating between individuals of two different groups feeding on distinct food resources (Lu and Bernatchez 1999). Accordingly, sympatric speciation subsequently occurs due to the restriction of gene flow between subpopulations living in the same area.

In this study we asked two main questions; first, does divergent feeding behavior promote sympatric speciation? If the answer to this first question is yes, then we would try to identify contributing behavior pattern(s) that facilitate sympatric divergence and ask the second question: what are specific pattern(s) shared between sympatric species that are primarily responsible for the occurrence of sympatric speciation?

We needed to categorize existing species based on their preference for different food resources. Two different method (FCM-Clustering Approach and Action-Perception Clustering Approach) were separately applied to categorize existing species into three different groups: species that were specialized on Food 1, species that were specialized on Food 2, and species that did not express any preferences. Then, the obtained results of both methods were compared to select the most efficient one to continue with. In order to answer the first question (detecting any evidence of the occurrence of sympatric speciation), a measuring tool was required to identify any indicators of the occurrence of sympatric speciation. Thus, we employed four different requirement criteria for sympatric speciation that were introduced by Bolnick and Fitzpatrick (2007). According to these requirements, 1. species undergoing speciation must be sister species; 2. there must be a complete reproductive isolation between sister species; 3. there must be a complete (or extensive) geographic overlap between these species; and 4. the occurrence of allopatric or parapatric speciation must be highly unlikely to be able to reject alternative hypotheses (Bolnick and Fitzpatrick 2007). The following approaches were respectively applied to test each requirement; applying phylogenetic analysis, quantifying the ratio of reproductive events leading to hybrid offspring, calculating average distance of all individuals between sister species, rejecting alternative hypotheses using the combination of phylogenetic tracking and biogeographic data.

The first question was answered and we were able to detect several runs with enough examples of sympatric species. Therefore, the second question (identifying underlying reasons leading to sympatric speciation) was perused. As such, machine learning techniques (including preparing the dataset, attribute selection, and classification) were applied to determine the shared patterns among the runs with enough examples of sympatric speciation, and thereby identifying influential conditions leading to sympatric divergence. The original version of EcoSim, which only contained one type of primary food resource, was also employed as the control treatment. Hence, single resource control simulations were compared with dual resource simulations to gain further insights into the role of the presence of different food resources in sympatric divergence. This study will be broadly discussed in the next chapter.

1.5 References

Bank, Claudia, Reinhard Bürger, and Joachim Hermisson. 2012. “The Limits to Parapatric Speciation: Dobzhansky--Muller Incompatibilities in a Continent--Island Model.” *Genetics* 191(3):845–63.

Barluenga, Marta, Kai N. Stölting, Walter Salzburger, Moritz Muschick, and Axel Meyer. 2006. “Sympatric Speciation in Nicaraguan Crater Lake Cichlid Fish.” *Nature* 439(7077):719–23.

Barraclough, Timothy G. and Alfried P. Vogler. 2000. “Detecting the Geographical Pattern of Speciation from Species-Level Phylogenies.” *The American Naturalist* 155(4):419–34.

Berlocher, Stewart H. 1998. “Can Sympatric Speciation via Host or Habitat Shift Be Proven from Phylogenetic and Biogeographic Evidence.” *Endless forms: Species and speciation* 99–113.

Berlocher, Stewart H. 1999. “Host Race or Species? Allozyme Characterization of the ‘Flowering Dogwood Fly’, a Member of the Rhagoletis Pomonella Complex.” *Heredity* 83(6):652–62.

Berlocher, Stewart H. 2000. “Radiation and Divergence in the Rhagoletis Pomonella Species Group: Inferences from Allozymes.” *Evolution* 54(2):543–57.

Berlocher, Stewart H. and Jeffrey L. Feder. 2002. “Sympatric Speciation in Phytophagous Insects: Moving beyond Controversy?” *Annual review of entomology* 47(1):773–815.

Bird, Christopher E., Iria Fernandez-Silva, Derek J. Skillings, and Robert J. Toonen. 2012. “Sympatric Speciation in the Post ‘Modern Synthesis’ Era of Evolutionary Biology.” *Evolutionary Biology* 39(2):158–80.

Bolnick, Daniel I. and Benjamin M. Fitzpatrick. 2007. “Sympatric Speciation: Models and Empirical Evidence.” *Annual Review of Ecology, Evolution, and Systematics* 459–87.

Bolnick, Daniel I. and T. Smith. 2004. “Can Intraspecific Competition Drive Disruptive Selection? An Experimental Test in Natural Populations of Sticklebacks.” *Evolution* 58(3):608–18.

Boughman, Janette Wenrick, Howard D. Rundle, and Dolph Schluter. 2005. “Parallel Evolution of Sexual Isolation in Sticklebacks.” *Evolution* 59(2):361–73.

Bush, G. L. 1993. “A Reaffirmation of Santa Rosalia, or Why Are There so Many Kinds of Small Animals.” *Evolutionary patterns and processes* 14:229–49.

Bush, Guy L. 1969. “Sympatric Host Race Formation and Speciation in Frugivorous Flies of the Genus Rhagoletis (Diptera, Tephritidae).” *Evolution* 237–51.

Bush, Guy L. 1994. “Sympatric Speciation in Animals: New Wine in Old Bottles.” *Trends in Ecology & Evolution* 9(8):285–88.

Bush, Guy L. and Daniel J. Howard. 1986. “Allopatric and Non-Allopatric Speciation: Assumptions and Evidence.” *Evolutionary Processes and Theory (S. Karlin & E. Nevo, eds.). Academic Press, New York* 411–38.

Bush, Guy L. and James J. Smith. 1998. “The Genetics and Ecology of Sympatric Speciation: A Case Study.” *Researches on Population Ecology* 40(2):175–87.

Cibois, Alice et al. 2014. “Phylogeny and Biogeography of the Fruit Doves (Aves: Columbidae).” *Molecular phylogenetics and evolution* 70:442–53.

Claridge, Michael F., Hassan A. Dawah, and Michael R. Wilson. 1997. *Species: The Units of Biodiversity*. Chapman & Hall Ltd.

Coyne, Jerry A. 1992. “Genetics and Speciation.” *Nature* 355(6360):511–15.

Coyne, Jerry A. 2007. “Sympatric Speciation.” *Current biology* 17(18):R787–88.

Coyne, Jerry A. and H. Allen Orr. 1997. “‘ Patterns of Speciation in Drosophila’ Revisited.” *Evolution* 51(1):295–303.

Coyne, Jerry A. and H. Allen Orr. 2004. *Speciation*. Sinauer Associates Sunderland, MA.

Darwin, Charles. 1859. “On the Origin of Species.”

Dieckmann, Ulf and Michael Doebeli. 1999. “On the Origin of Species by Sympatric Speciation.” *Nature* 400(6742):354–57.

Diehl, S. R. and G. L. Bush. 1984. “An Evolutionary and Applied Perspective of Insect Biotypes.” *Annual review of entomology* 29(1):471–504.

Diehl, Scott R. and Guy L. Bush. 1989. “The Role of Habitat Preference in Adaptation and Speciation.” *Speciation and its Consequences* 345–65.

Doebeli, Michael. 1996. “A Quantitative Genetic Competition Model for Sympatric Speciation.” *Journal of evolutionary biology* 9(6):893–909.

Doebeli, Michael and Ulf Dieckmann. 2000. “Evolutionary Branching and Sympatric Speciation Caused by Different Types of Ecological Interactions.” *The american naturalist* 156(S4):S77–101.

Van Doorn, G. Sander, Pim Edelaar, and Franz J. Weissing. 2009. “On the Origin of Species by Natural and Sexual Selection.” *Science* 326(5960):1704–7.

Drès, Michele and James Mallet. 2002. “Host Races in Plant--Feeding Insects and Their Importance in Sympatric Speciation.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 357(1420):471–92.

Elmer, Kathryn R., Henrik Kusche, Topi K. Lehtonen, and Axel Meyer. 2010. “Local Variation and Parallel Evolution: Morphological and Genetic Diversity across a Species Complex of Neotropical Crater Lake Cichlid Fishes.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 365(1547):1763–82.

Feder, Jeffrey L., Charles A. Chilcote, and Guy L. Bush. 1988. “Genetic Differentiation between Sympatric Host Races of the Apple Maggot Fly Rhagoletis Pomonella.” *Nature* 336(6194):61–64.

Feder, Jeffrey L., Charley A. Chilcote, and Guy L. Bush. 1990. “The Geographic Pattern of Genetic Differentiation between Host Associated Populations of Rhagoletis Pomonella (Diptera: Tephritidae) in the Eastern United States and Canada.” *Evolution* 570–94.

Feder, Jeffrey L. and Kenneth E. Filchak. 1999. “It’s about Time: The Evidence for Host Plant-Mediated Selection in the Apple Maggot Fly, Rhagoletis Pomonella, and Its Implications for Fitness Trade-Offs in Phytophagous Insects.” Pp. 211–25 in *Proceedings of the 10th International Symposium on Insect-Plant Relationships*.

Feder, Jeffrey L. and Andrew A. Forbes. 2007. “Habitat Avoidance and Speciation for Phytophagous Insect Specialists.” *Functional Ecology* 21(3):585–97.

Feder, Jeffrey L., Joseph B. Roethele, Brian Wlazlo, and Stewart H. Berlocher. 1997. “Selective Maintenance of Allozyme Differences among Sympatric Host Races of the Apple Maggot Fly.” *Proceedings of the National Academy of Sciences* 94(21):11417–21.

Felsenstein, Joseph. 1981. “Skepticism towards Santa Rosalia, or Why Are There so Few Kinds of Animals?” *Evolution* 124–38.

Filchak, Kenneth E., Joseph B. Roethele, and Jeffrey L. Feder. 2000. “Natural Selection and Sympatric Divergence in the Apple Maggot Rhagoletis Pomonella.” *Nature* 407(6805):739–42.

Fisher, Ronald Aylmer. 1958. *The Genetical Theory of Natural Selection*. 2nd ed. Dover, New York. 291 pp.

Fitzpatrick, Benjamin M., J. A. Fordyce, and S. Gavrilets. 2008. “What, If Anything, Is Sympatric Speciation?” *Journal of evolutionary biology* 21(6):1452–59.

Friesen, V. L. et al. 2007. “Sympatric Speciation by Allochrony in a Seabird.” *Proceedings of the National Academy of Sciences* 104(47):18589–94.

Funk, Daniel J., Scott P. Egan, and Patrik Nosil. 2011. “Isolation by Adaptation in Neochlamisus Leaf Beetles: Host-Related Selection Promotes Neutral Genomic Divergence.” *Molecular ecology* 20(22):4671–82.

Funk, Daniel J., Patrik Nosil, and William J. Etges. 2006. “Ecological Divergence Exhibits Consistently Positive Associations with Reproductive Isolation across Disparate Taxa.” *Proceedings of the National academy of Sciences of the United States of America* 103(9):3209–13.

Gavrilets, Sergey. 2003. “Perspective: Models of Speciation: What Have We Learned in 40 Years?” *Evolution* 57(10):2197–2215.

Gavrilets, Sergey. 2004. *Fitness Landscapes and the Origin of Species (MPB-41)*. Princeton University Press Princeton, NJ.

Gavrilets, Sergey and David Waxman. 2002. “Sympatric Speciation by Sexual Conflict.” *Proceedings of the National Academy of Sciences* 99(16):10533–38.

Golestani, Abbas, Robin Gras, and Melania Cristescu. 2012. “Speciation with Gene Flow in a Heterogeneous Virtual World: Can Physical Obstacles Accelerate Speciation?” *Proceedings of the Royal Society B: Biological Sciences* 279(1740):3055–64.

Grant, Peter R. and B. Rosemary Grant. 2010. “Sympatric Speciation, Immigration and Hybridization in Island Birds.” *The Theory of Island Biogeography Revisited* 326–57.

Grant, Peter R., B. Rosemary Grant, and J. C. Deutsch. 1996. “Speciation and Hybridization in Island Birds [and Discussion].” *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 351(1341):765–72.

Gras, Robin, Didier Devaurs, Adrianna Wozniak, and Adam Aspinall. 2009. “An Individual-Based Evolving Predator-Prey Ecosystem Simulation Using a Fuzzy Cognitive Map as the Behavior Model.” *Artificial life* 15(4):423–63.

Gras, Robin, Abbas Golestani, Andrew P. Hendry, and Melania E. Cristescu. 2015. “Speciation without Pre-Defined Fitness Functions.” *Plos One* 10(9):e0137838. Retrieved (http://dx.plos.org/10.1371/journal.pone.0137838).

Gras, Robin, Abbas Golestani, Andrew P. Hendry, and Melania E. Cristescu. 2015. “Speciation without Pre-Defined Fitness Functions.” *Plos One* 10(9):e0137838. Retrieved (http://dx.plos.org/10.1371/journal.pone.0137838).

Griffith, David. 1996. “The Ecology of Resource Polymorphism in Vertebrates.” *Trends in Ecology and Evolution* 11.

Grimm, Volker and Steven F. Railsback. 2013. *Individual-Based Modeling and Ecology*. Princeton university press.

Hadid, Yarin et al. 2013. “Possible Incipient Sympatric Ecological Speciation in Blind Mole Rats (Spalax).” *Proceedings of the National Academy of Sciences* 110(7):2587–92.

Hatfield, Todd and Dolph Schluter. 1999. “Ecological Speciation in Sticklebacks: Environment-Dependent Hybrid Fitness.” *Evolution* 866–73.

Hendry, A. P. and M. T. Kinnison. 2001. “An Introduction to Microevolution: Rate, Pattern, Process.” *Genetica* 112(1):1–8.

Hendry, Andrew P., Sarah K. Huber, Luis F. De Leon, Anthony Herrel, and Jeffrey Podos. 2009. “Disruptive Selection in a Bimodal Population of Darwin’s Finches.” *Proceedings of the Royal Society of London B: Biological Sciences* 276(1657):753–59.

Herron, J. C. and S. Freeman. 2013. *Evolutionary Analysis*. Pearson Education. Retrieved (https://books.google.ca/books?id=BfKaAAAAQBAJ).

Higashi, M., G. Takimoto, and N. Yamamura. 1999. “Sympatric Speciation by Sexual Selection.” *Nature* 402(6761):523–26.

Higgie, Megan, Steve Chenoweth, and Mark W. Blows. 2000. “Natural Selection and the Reinforcement of Mate Recognition.” *Science* 290(5491):519–21.

Hood, Glen R. et al. 2013. “The Geographic Distribution of Rhagoletis Pomonella (Diptera: Tephritidae) in the Western United States: Introduced Species or Native Population?” *Annals of the Entomological Society of America* 106(1):59–65.

Huber, Sarah K., Luis Fernando De Leon, Andrew P. Hendry, Eldredge Bermingham, and Jeffrey Podos. 2007. “Reproductive Isolation of Sympatric Morphs in a Population of Darwin’s Finches.” *Proceedings of the Royal Society of London B: Biological Sciences* 274(1619):1709–14.

Jiang, Tinglei, Jiang Feng, Keping Sun, and Jing Wang. 2008. “Coexistence of Two Sympatric and Morphologically Similar Bat Species Rhinolophus Affinis and Rhinolophus Pearsoni.” *Progress in Natural Science* 18(5):523–32.

Karpinski, Anne, Sabine Haenniger, Gerhard Schöfl, David G. Heckel, and Astrid T. Groot. 2014. “Host Plant Specialization in the Generalist Moth Heliothis Virescens and the Role of Egg Imprinting.” *Evolutionary Ecology* 28(6):1075–93.

Kautt, Andreas F., Kathryn R. Elmer, and Axel Meyer. 2012. “Genomic Signatures of Divergent Selection and Speciation Patterns in a Natural Experiment, the Young

Parallel Radiations of Nicaraguan Crater Lake Cichlid Fishes." *Molecular Ecology* 21(19):4770–86.

Kautt, Andreas F., Gonzalo Machado-Schiaffino, and Axel Meyer. 2016. "Multispecies Outcomes of Sympatric Speciation after Admixture with the Source Population in Two Radiations of Nicaraguan Crater Lake Cichlids." *PLoS Genet* 12(6):e1006157.

Kautt, Andreas F., Gonzalo Machado-Schiaffino, Julian Torres-Dowdall, and Axel Meyer. 2016. "Incipient Sympatric Speciation in Midas Cichlid Fish from the Youngest and One of the Smallest Crater Lakes in Nicaragua due to Differential Use of the Benthic and Limnetic Habitats?" *Ecology and Evolution* 6(15):5342–57.

Kawecki, Tadeusz J. 1997. "Sympatric Speciation via Habitat Specialization Driven by Deleterious Mutations." *Evolution* 1751–63.

Kirkpatrick, Mark and Virginie Ravigné. 2002. "Speciation by Natural and Sexual Selection: Models and Experiments." *The American Naturalist* 159(S3):S22–35.

Kondrashov, Alexey S. 1983. "Multilocus Model of Sympatric Speciation II. Two Characters." *Theoretical Population Biology* 24(2):136–44.

Kondrashov, Alexey S. 1986. "Multilocus Model of Sympatric Speciation. III. Computer Simulations." *Theoretical population biology* 29(1):1–15.

Kondrashov, Alexey S. and Fyodor A. Kondrashov. 1999. "Interactions among Quantitative Traits in the Course of Sympatric Speciation." *Nature* 400(6742):351–54.

Kondrashov, Alexey S. and Mikhail V. Mina. 1986. "Sympatric Speciation: When Is It Possible?" *Biological Journal of the Linnean Society* 27(3):201–23.

Labonne, Jacques and Andrew P. Hendry. 2010. "Natural and Sexual Selection Giveth and Taketh Away Reproductive Barriers: Models of Population Divergence in Guppies." *The American Naturalist* 176(1):26–39.

Lande, Russell. 1980. "Genetic Variation and Phenotypic Evolution during Allopatric Speciation." *American Naturalist* 463–79.

Lande, Russell. 1981. "Models of Speciation by Sexual Selection on Polygenic Traits." *Proceedings of the National Academy of Sciences* 78(6):3721–25.

Larsen, Clark Spencer and Jack Repcheck. 2008. *Our Origins: Discovering Physical Anthropology*. WW Norton & Company.

Levin, Simon. 2007. *Fragile Dominion*. Basic Books.

Linn, Charles et al. 2003. “Fruit Odor Discrimination and Sympatric Host Race Formation in Rhagoletis.” *Proceedings of the National Academy of Sciences* 100(20):11490–93.

Lövy, Matěj et al. 2015. “Habitat and Burrow System Characteristics of the Blind Mole Rat Spalax Galili in an Area of Supposed Sympatric Speciation.” *PloS one* 10(7):e0133157.

Lu, Guoqing and Louis Bernatchez. 1999. “Correlated Trophic Specialization and Genetic Divergence in Sympatric Lake Whitefish Ecotypes (Coregonus Clupeaformis): Support for the Ecological Speciation Hypothesis.” *Evolution* 1491–1505.

Maan, Martine E. and Ole Seehausen. 2011. “Ecology, Sexual Selection and Speciation.” *Ecology letters* 14(6):591–602.

MacPherson, Brian and Robin Gras. 2016. “Individual-Based Ecological Models: Adjunctive Tools or Experimental Systems?” *Ecological Modelling* 323:106–14.

Malinsky, Milan et al. 2015. “Genomic Islands of Speciation Separate Cichlid Ecomorphs in an East African Crater Lake.” *Science* 350(6267):1493–98.

Martin, Ryan A. and David W. Pfennig. 2009. “Disruptive Selection in Natural Populations: The Roles of Ecological Specialization and Resource Competition.” *The American Naturalist* 174(2):268–81.

Mayr, Ernst. 1963. *Animal Species and Evolution*. Vol. 797. Cambridge, Massachusetts: Belknap Press of Harvard University Press.

Mayr, Ernst, Ernst Mayr, Ernst Mayr, and Ernst Mayr. 1963. *Animal Species and Evolution*. Belknap Press of Harvard University Press Cambridge, Massachusetts.

Nagel, Laura and Dolph Schluter. 1998. “Body Size, Natural Selection, and Speciation in Sticklebacks.” *Evolution* 209–18.

Nosil, Patrik, Bernard J. Crespi, and Cristina P. Sandoval. 2002. “Host-Plant Adaptation Drives the Parallel Evolution of Reproductive Isolation.” *Nature* 417(6887):440–43.

Orr, Matthew R. and Thomas B. Smith. 1998. “Ecology and Speciation.” *Trends in Ecology & Evolution* 13(12):502–6.

Panhuis, Tami M., Roger Butlin, Marlene Zuk, and Tom Tregenza. 2001. “Sexual Selection and Speciation.” *Trends in Ecology & Evolution* 16(7):364–71.

Pfennig, David W. and Karin S. Pfennig. 2010. “Character Displacement and the Origins of Diversity.” *the american naturalist* 176(Suppl 1):S26.

Price, Peter W. 1975. *Evolutionary Strategies of Parasitic Insects*. Springer.

Prowell, Dorothy Pashley, Margaret McMichael, and Jean-François Silvain. 2004. “Multilocus Genetic Analysis of Host Use, Introgression, and Speciation in Host Strains of Fall Armyworm (Lepidoptera: Noctuidae).” *Annals of the Entomological Society of America* 97(5):1034–44.

Rashkovetsky, Eugenia, Zeev Frenkel, Pawel Michalak, and Abraham Korol. 2015. “Sympatric Differentiation and Speciation: Insights from Drosophila Studies.” Pp. 107–40 in *Evolutionary Biology: Biodiversification from Genotype to Phenotype*. Springer.

Rice, William R. 1984. “Disruptive Selection on Habitat Preference and the Evolution of Reproductive Isolation: A Simulation Study.” *Evolution* 1251–60.

Rice, William R. 1987. “Speciation via Habitat Specialization: The Evolution of Reproductive Isolation as a Correlated Character.” *Evolutionary Ecology* 1(4):301–14.

Rice, William R. and Ellen E. Hostert. 1993. “Laboratory Experiments on Speciation: What Have We Learned in 40 Years?” *Evolution* 1637–53.

Rice, William R. and George W. Salt. 1990. “The Evolution of Reproductive Isolation as a Correlated Character under Sympatric Conditions: Experimental Evidence.” *Evolution* 1140–52.

Ridley, Mark. 2004. *Evolution*. 3rd ed. Oxford: Blackwell Pub.

Rieseberg, Loren H., Barry Sinervo, C. Randal Linder, Mark C. Ungerer, and Dulce M. Arias. 1996. “Role of Gene Interactions in Hybrid Speciation: Evidence from Ancient and Experimental Hybrids.” *Science* 272(5262):741.

Rosser, Neil, Krzysztof M. Kozak, Albert B. Phillimore, and James Mallet. 2015. “Extensive Range Overlap between Heliconiine Sister Species: Evidence for Sympatric Speciation in Butterflies?” *BMC evolutionary biology* 15(1):125.

Roughgarden, Jonathan. 1972. “Evolution of Niche Width.” *American Naturalist* 683–718.

Rundle, Howard D., Laura Nagel, Janette Wenrick Boughman, and Dolph Schluter. 2000. “Natural Selection and Parallel Speciation in Sympatric Sticklebacks.” *Science* 287(5451):306–8.

Rundle, Howard D. and Patrik Nosil. 2005. “Ecological Speciation.” *Ecology letters* 8(3):336–52.

Rundle, Howard D. and Dolph Schluter. 2004. “Natural Selection and Ecological Speciation in Sticklebacks.” *Adaptive speciation* 19(3):192–209.

Ryan, Peter G., Paulette Bloomer, Coleen L. Moloney, Tyron J. Grant, and Wayne Delport. 2007. “Ecological Speciation in South Atlantic Island Finches.” *science* 315(5817):1420–23.

Savolainen, Vincent et al. 2006. “Sympatric Speciation in Palms on an Oceanic Island.” *Nature* 441(7090):210–13.

Schliewen, Ulrich K., Diethard Tautz, and Svante Pääbo. 1994. “Sympatric Speciation Suggested by Monophyly of Crater Lake Cichlids.” *Nature* 368(6472):629–32.

Schluter, D. 1998. “Ecological Causes of Speciation.” *Endless forms: species and speciation* 114–29.

Schluter, Dolph. 2000. *The Ecology of Adaptive Radiation*. Oxford University Press.

Schluter, Dolph. 2001. “Ecology and the Origin of Species.” *Trends in ecology & evolution* 16(7):372–80.

Schluter, Dolph and Peter R. Grant. 1984. “Determinants of Morphological Patterns in Communities of Darwin’s Finches.” *American Naturalist* 175–96.

Schluter, Dolph and John Donald McPhail. 1992. “Ecological Character Displacement and Speciation in Sticklebacks.” *American Naturalist* 85–108.

Sezer, M. and R. K. Butlin. 1998. “The Genetic Basis of Host Plant Adaptation in the Brown Planthopper (Nilaparvata Lugens).” *Heredity* 80(4):499–508.

Shaw, Kerry L. and Ezequiel Lugo. 2001. “Mating Asymmetry and the Direction of Evolution in the Hawaiian Cricket Genus Laupala.” *Molecular Ecology* 10(3):751–59.

Simpson, George Gaylord. 1955. *Major Features of Evolution*. New York: Columbia University Press.

Smith, J. Maynard. 1966. “Sympatric Speciation.” *American Naturalist* 637–50.

Sorenson, Michael D., Kristina M. Sefc, and Robert B. Payne. 2003. “Speciation by Host Switch in Brood Parasitic Indigobirds.” *Nature* 424(6951):928–31.

Soudi, Shaghayegh, Klaus Reinhold, and Leif Engqvist. 2016. “Genetic Architecture Underlying Host Choice Differentiation in the Sympatric Host Races of Lochmaea Capreae Leaf Beetles.” *Genetica* 144(2):147–56.

Stearns, Stephen C. 1992. *The Evolution of Life Histories*. Oxford University Press Oxford.

Svanbäck, Richard and Dolph Schluter. 2012. “Niche Specialization Influences Adaptive Phenotypic Plasticity in the Threespine Stickleback.” *The American Naturalist* 180(1):50–59.

Tauber, Catherine A. and Maurice J. Tauber. 1977. “A Genetic Model for Sympatric Speciation through Habitat Diversification and Seasonal Isolation.” *Nature* 268:702–5.

Tauber, Catherine A. and Maurice J. Tauber. 1989. “Sympatric Speciation in Insects: Perception and Perspective.” *Speciation and its Consequences* 307–44.

Thibert-Plant, Xavier and A. P. Hendry. 2009. “Five Questions on Ecological Speciation Addressed with Individual-Based Simulations.” *Journal of evolutionary biology* 22(1):109–23.

Thibert-Plante, X. and a P. Hendry. 2011. “Factors Influencing Progress toward Sympatric Speciation.” *Journal of Evolutionary Biology* 24(10):2186–96.

Tregenza, Tom and Roger K. Butlin. 1999. “Speciation without Isolation.” *Nature* 400(6742):311–12.

Turelli, Michael, Nicholas H. Barton, and Jerry A. Coyne. 2001. “Theory and Speciation.” *Trends in Ecology & Evolution* 16(7):330–43.

Uy, J. Albert C. and Gerald Borgia. 2000. “Sexual Selection Drives Rapid Divergence in Bowerbird Display Traits.” *Evolution* 54(1):273–78.

Via, Sara. 2001. “Sympatric Speciation in Animals: The Ugly Duckling Grows up.” *Trends in Ecology & Evolution* 16(7):381–90.

Wiley, Edward Orlando and Bruce S. Lieberman. 2011. *Phylogenetics: Theory and Practice of Phylogenetic Systematics*. John Wiley & Sons.

Wilson, David Sloan and Michael Turelli. 1986. “Stable Underdominance and the Evolutionary Invasion of Empty Niches.” *American Naturalist* 835–50.

Wood, Thomas K. and M. C. Keese. 1990. “Host-Plant-Induced Assortative Mating in Enchenopa Treehoppers.” *Evolution* 619–28.

Yoder, J. B. et al. 2010. “Ecological Opportunity and the Origin of Adaptive Radiations.” *Journal of evolutionary biology* 23(8):1581–96.

CHAPTER 2

UTILIZING AN INDIVIDUAL-BASED SIMULATION TO INVESTIGATE THE IMPORTANCE OF SPECIALIZED FORAGING BEHAVIOR IN SYMPATRIC SPECIATION

2.1. Introduction

As pointed out in the first chapter, this modeling study was initiated to investigate the central role of foraging behavior in sympatric divergence. The adapted methods and obtained results will be thoroughly discussed in this chapter.

2.2 Methods

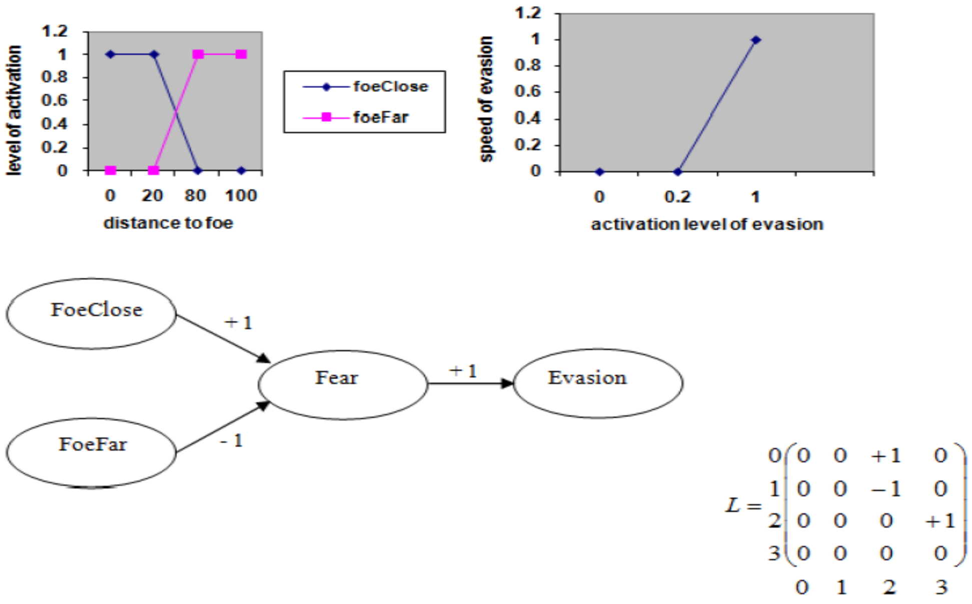
2.2.1 EcoSim

EcoSim (Gras et al. 2009) is an individual-based evolving ecosystem simulation, written in C++, and simulating a terrestrial tri-trophic dynamic food chain model of interacting organisms including: primary producer (grass), primary consumers (herbivores or prey), and secondary consumers (carnivores or top predator). This system has been used to study diverse ecological questions such as: rate of speciation (Golestani, Gras, and Cristescu 2012; Gras et al. 2015), species extinction (Mashayekhi et al. 2014), and contemporary evolution of prey in the presence of predators (Khater, Murariu, and Gras 2014).

The virtual world of EcoSim is a torus environment of 1000 × 1000 discrete cells. Each cell contains an unlimited number of prey and predator individuals, but a limited amount of primary resources. The resource amount and spatial distribution varies dynamically in time (Golestani and Gras 2011). Prey and predator individuals live in a world consisting of discrete cells. This model goes through separate time steps. During each time step, living organisms perform different actions based on their perception of their surrounding environment and of the other organisms that they are in interaction with. This, in consequence, influences the whole system. Prey and predator species coexist and they need efficient, evolvable behaviors to be able to survive and adapt to the evolving virtual world (Gras et al. 2009). The behavior of each living organism is coded in its genome and implemented via a Fuzzy Cognitive Map (FCM) (Kosko 1986). As such, individuals are able to perceive their environment using their FCM and then perform at any time step the behavior they perceive as the most beneficial. This means that at every time step, each individual will perform a unique action as determined by its behavioral model and its surrounding environment. The FCM of each agent, being coded in its genome, thus allows the evolution of the agent behavior through the simulation. As a result of utilizing such a complicated modelling approach, each individual in EcoSim can express different and divergent behavior (Gras et al. 2009).

The FCMs consist of directed graphs containing nodes that represent concepts and the edges from one concept to another, which demonstrate the influences between concepts. The influence of the concepts in an FCM with n concepts can be represented in an n×n matrix. A positive weight associated with the edge e_{ij} corresponds to an excitation of the concept c_j from the concept c_i, whereas a negative weight is related to an inhibition (a zero value indicates that there is no influence of c_i on c_j). Individuals in EcoSim have three sets of concepts: Sensitive (distance of individual from food, predator etc.), Internal (such as fear, hunger etc.), and Motor (such as evasion of predators, eating, etc.). Sensitive concepts are set by mapping a perception out of an environmental observation. At initialization, the Sensitive concepts affect Internal concepts that in turn affect Motor concepts, but evolution can add edges between any concepts allowing some complex feedback loops to emerge. A number is associated with each concept, which is called the concept’s “activation level.” Activation levels are updated at each time step, using a concept’s current activation level and the weighted sum of other concepts’ activation levels affecting that concept, transformed by a non-linear function. The activation level of a Sensitive concept is computed by performing a “fuzzification” of the information an individual perceives from the environment. For an Internal or Motor concept, the activation level is computed by applying a de-fuzzification function on the weighted sum of the current activation level of all the concepts having an edge directed toward it. Finally, the action of an individual is selected based on the maximum value of the Motor concepts' activation level. Activation levels of Motor concepts are used to determine the next action of the individual and its amplitude.

For example a simple FCM regarding two Sensitive concepts (foeClose and foeFar), one Internal concept (fear) and one Motor concept (evasion) can have three influence edges: closeness to a foe excites fear, distance to a foe inhibits fear, and fear causes evasion (Figure 2.1). Fuzzification of concepts foeClose (nearness to the predator) and foeFar (distance from predator) provide the activation of the concepts depending on the distance of prey from a predator. De-fuzzification of the evasion concept provides the speed at which preys evade. Therefore, the FCMs are weighted graphs representing the causal relationship between Sensitive, Internal, and Motor nodes. The activation levels of the concepts of an individual are never reset during its life. Hence, the previous states of an individual participate in the computation of its current state. Therefore, an individual has a memory of its own past and this will influence the individual's future states. As the action undertaken by an individual at a given time step depends on the current activation level of the motor concepts, the global behavior of the individual depends on a complex combination of the individual's perception, the current internal states, and the past states it went through during its life (Gras et al. 2009).



$$L = \begin{array}{c c}
\begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \end{array} & \begin{pmatrix} 0 & 0 & +1 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & +1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \\
 & \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array}
\end{array}$$

Figure 2.1. An FCM for detection of foe (predator) and decision to evade with its corresponding matrix (0 for ‘Foe close’, 1 for ‘Foe far’, 2 for ‘Fear’ and 3 for ‘Evasion’) and the fuzzification and defuzzification functions (Gras et al. 2009)

In EcoSim every individual possesses its own properties, which are mostly related to physical capabilities such as: age, minimum age for breeding, speed, vision distance, level of energy, and the amount of energy transmitted to the offspring. Prey individuals obtain their required energy through the consumption of the available primary producer (grass) in the environment. Throughout the world, primary resource distribution is dynamic in terms of quantity and location. Predator individuals prey on herbivores to satisfy their energy needs. As a result of performing each action (eating, reproducing, etc.), each individual loses some amount of energy based on the type of action performed and the complexity of its behavioral model (the number of existing edges in its FCM). In this evolving system the process of producing a new individual occurs when two individuals that possess essential prerequisites for mating action (being in the same cell, both choosing to express reproduction action as their first priority among other actions, having the minimum age of reproduction, having the minimum level of required energy, and being genetically close enough) perform a successful mating action. The produced offspring will inherit its parents' genome combination with some mutations.

The notion of species is also implemented in this modeling system so that species will emerge from the evolving population of agents. Accordingly, “species” is defined in this model as a set of individuals with similar genomic characteristics, and the defined genome of a given species results from the average genomic characteristics of all its individual members. Speciation events in this simulation occur as a result of the emergent properties of interactions between individuals in their spatial landscapes, where abiotic parameters are initially invariant. As a result of this speciation mechanism, a species splits if the members of the species are not genetically similar enough (based on a predefined threshold). If the genomic similarity between two individuals of a particular species becomes smaller than a predefined threshold, a speciation event occurs (Gras et al. 2009; Gras et al. 2015). Consequently, the initial species is split into two sister species using a 2-mean clustering algorithm (Aspinall and Gras 2010). The resulting sister species contains individuals that show more intraspecific genetic similarity.

2.2.2 Modeling Sympatric Speciation Using EcoSim Model

This study focuses on the relationship between the first and the second trophic level, primary food resources (grass), and prey species to model resource-based diversification. As such, a second type of food resource has been added to the model to provide more than one food resource for prey individuals to feed on (see Figure A1, a and b in the Appendix). In one single resource version (original version of EcoSim), FCM maps of prey individuals contain four Sensitive and two Motor concepts that are directly related to the prey's food consumption. These Sensitive concepts are: Food Close, Food Far, Food Local High, and Food Local Low. A Motor concept related to prey food consumption is Search For Food and Eat. Hunger, Search Partner, Curiosity, Sedentary, Satisfaction, and Nuisance are the Internal concepts in prey FCMs that are influenced by prey food consumption. In order to avoid any initial bias regarding the introduction of a new food resource to the system, the prey FCM is modified by adding four new Sensitive concepts of; Food Close 2, Food Far 2, Food Local High 2, and Food Local Low 2 as well as two new Motor concepts: Search For Food 2 and Eat 2 (Figures A2.a in the Appendix changed to Figure A2.b in the Appendix). New edges between Sensitive, Internal and Motor concepts in prey FCMs are also added. The complete FCM maps of prey individuals after adding the extra source of food is shown in Figure 2.2.

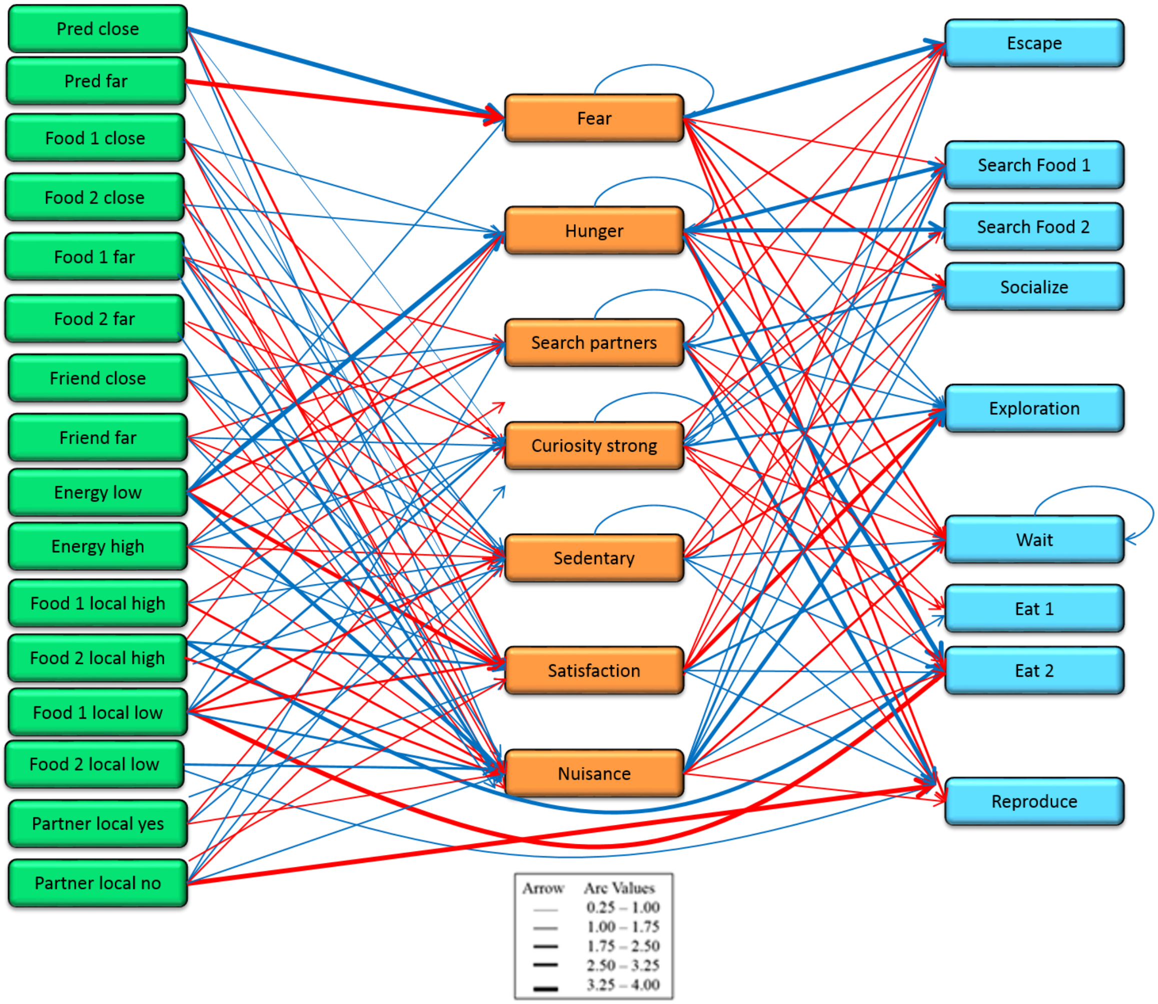


Figure 2.2. The initial Prey FCM including concepts and edges for the dual resource version of the EcoSim. The width of each edge shows its influence value. The color of an edge shows inhibitory (red) or excitatory (blue) effects (Bandehbahman 2014)

The new food resource added to the system possesses specific characteristics (Table 2.1) that we customized to create two different food resources that differ from each other in their probability of diffusion, speed of growth, and the amount of energy obtained from feeding on each food resource (the amount of energy transferred to a prey individual after eating one unit of each food resource). In general, each unit of Food 2 contains a higher amount of energy than that in one unit of Food 1. In other words, Food 2 tends to be more valuable in terms of the amount of energy transmitted to prey consumers. However, Food 1 is more accessible as it grows faster and spreads throughout the world at a higher rate than Food 2. Introducing dissimilar food resources with different levels of availability and energy content to the simulated world creates the favorable conditions for the emergence of food consumption specialization in prey individuals (either getting specialized on more available food or food with higher energy content).

Table 2.1 The main characteristics of food resources for the prey individuals (Bandehbahman 2014)

Parameter	**Description**	**Food 1**	**Food 2**	**Food (Standard EcoSim)**
Value Primary Resource	Energy value for a consumed primary resource unit	250	400	325
Max Primary Resource	Maximum number of primary resource units in a cell	4	4	8
Speed Grow Primary Resource	Speed of growing primary resource	0.3	0.2	0.3
Probability Initial Primary Resource	Initial probability of primary resource per cell	0.187	0.187	0.187
Probability Grow Primary Resource	Probability of primary resource diffusion	0.0016	0.0014	0.0016

2.2.3 Indicators of the Occurrence of Sympatric Speciation

Bolnick and Fitzpatrick (2007) defined four different basic requirements for the occurrence of sympatric speciation: firstly, species undergoing speciation must be sister species; secondly, there must be a complete reproductive isolation between these species; thirdly, there must be a complete (or extensive) geographical overlap between these species; fourthly, the occurrence of allopatric/parapatric speciation must be highly improbable to be able to reject alternative hypotheses. However, it is difficult for empirical investigations to fulfill these requirements. Computational simulations on the other hand, provide complete control over a huge number of discrete factors and facilitate the development of models addressing the complex interactions between species that give rise to sympatric speciation. Modeling simulations take advantage of computational resources, and thereby enable us to closely monitor and investigate speciation events in a reasonable time period. Additionally, these modeling approaches facilitate the process of quantitative analysis of data. Considering the pragmatic application of the modeling approach in investigating the speciation process, we employed an IBM approach and followed the suggested requirements for the occurrence of sympatric speciation as defined by Bolnick and Fitzpatrick (2007) and defined four criteria (Table 2.2) that must be fulfilled in order to consider a speciation event as a "sympatric divergence". As illustrated in Table 2.2, four different methods were employed to test each criterion. This criteria and applied methods will be further described in the following subsections. As soon as one run was complete, a large amount of information about individuals and species (e.g. their actions, their breeding information, and all the information about their behavioral FCM model), as well as a complete set of information about their surrounding environment (e.g. individual's geographic location and the food abundance distribution in the environment) were provided to analyze and evaluate the occurrence of sympatric speciation. The first filter selected the runs in which divergent eating behavior had occurred and species had expressed a significant preference for one specific type of food resource (either primary resource Type 1 or Type 2). This filter was tested following the protocol described in section 2.2.4. Observing preferential behavior for different types of food resources among different coexisting species is interpreted as the first indicator of the occurrence of sympatric speciation. The second step of the analysis process was evaluating the four selected criteria, which will be discussed in section 2.2.5.

Table 2.2. Sympatric speciation criteria and chosen strategy (Bandehbahman 2014)

Criteria (adopted from Bolnick and Fitzpatrick (2007))	**Strategy**
1. Sister Species	Phylogenetic analysis
2. Complete reproductive isolation	Ratio of reproductive events leading to hybrid offspring
3. Overlapping geographic ranges	Calculating average geographic distance between all individuals of sister species
4. Allopatric/Parapatric alternative hypotheses	Rejecting alternative hypotheses using the combination of phylogenetic tracking and biogeographic data

2.2.4 Species Categorizing Algorithm

Two different approaches (FCM-Clustering Approach and Action-Perception Clustering Approach) were applied to detect species expressing preferential behavior toward one specific food resource. Under the first approach, FCM-Clustering Approach, each species' average behavioral model was analyzed. This means that the behavioral model (FCM) of every individual of each species was averaged to obtain the average FCM for each species. This value can be used to evaluate the extent of the preference expressed by each species for each type of food resource to identify a species' category. Under the second approach, Action-Perception Clustering Approach, the action performed by individuals of each species was examined. In other words, the real actions performed by individuals and also individuals' perceptions of their surrounding resources were taken into account to evaluate whether they had any preferential behavior toward one specific resource or not. These approaches are thoroughly described in the following subsections. Employing these approaches enabled us to categorize species into three different groups based on their preferences for different food resources: Group One, the species that were more likely to choose Food1 rather than Food 2; Group Two, the species that had a preference for in consuming Food 2; and Group Three, the species without any particular preference for either food resource that simply chose the closest available food resource.

2.2.4.1 FCM-Clustering Approach

In order to determine if one species show preferential behavior toward a specific food resource or not, the weighted sum of all the edges that had influence on the Eat1 and Eat2 Motor concepts were separately calculated. Then, in order to categorize all existing species to three different groups, a threshold was defined to differentiate between the associated values of all edges coming to (influencing) the Eat1 and Eat2 Motor concepts. If the differences between the weighted sums assigned to Eat1 and Eat2 were smaller than 0.5, it was assumed that the species do not express any significant preference for either food resource and was assigned to Group Three (species with no preference). However, if the difference between the value associated to Eat1 and the value associated to Eat2 was greater than the threshold and the value of Eat1 was greater than the value of Eat2; then the species was assigned to Group One. In contrast, the species was categorized as Group Two under the opposite situation (see Figure A3 in the Appendix). After categorizing all existing species into three separate groups, the number of individuals belonging to each group was counted in each time step and (see Figure A4 in the Appendix).

2.2.4.2 Action-Perception Clustering Approach

In the second approach, instead of using the FCM behavioral model (as employed in the first approach), species' real behavior was applied as the main criteria for the classification of existing species into the three different groups (as discussed above). The rate of performed Eat 1 and Eat 2 actions by each species and the average perception for each species' local food resource availability (Food 1 and 2) were taken into consideration. Five simple logical rules were applied to evaluate these two criteria (see Table A5 in the Appendix). The thresholds were chosen to ensure that the differences in behaviors and perceptions were significantly strong (see Figure A5 in the Appendix).

2.2.5 Verifying Required Criteria of Sympatric Speciation

2.2.5.1 First Criterion: Sister Species

The first criterion was identifying the sets of sister species that were specialized on different food resources. More precisely, it was necessary to consider any set of two species and determine whether they are sister species (each other closest relative) or not. This assessment had to be applied for all couples of species. Since every single individual of the prey and predator species were trackable through evolutionary time in our simulation study, we could simply construct the exact phylogenetic trees to determine the precise time of the occurrence of speciation. Thus, the phylogenetic trees were made to distinguish species with preference for one specific food resource. Consequently, this made it possible to categorize species on their phylogenetic tree in terms of their expressed preference for specific food resources. Based on the first criterion, three different categories obtained from the previous step were used to find a set of sister species, one specialized on Food 1 and the other specialized on Food 2.

Phylogenetic trees were computed using information such as parent species ID, offspring species ID, and the time steps where speciation events occurred. The life span of each species was applied to determine that species' associated branch length in their phylogenetic tree. It should be noted that this program also had the related information obtained from the previous step regarding the species categorizing algorithm and their food preferences.

We needed to apply a method to visualize species that belonged to different categories (either specialized on Food 1 or Food 2), so that we could easily detect sister species with different food resource specializations. Therefore, a graphical editor for phylogenetic trees called TreeGraph (Stöver and Müller 2010) was applied. A truncated phylogenetic tree, rooting on a speciation event occurring at time step 17400, is presented in Figure 2.3. This represents a good example of a set of sister species that has met the first criterion. This set of sister species has lived for more than 400 time steps, that is why the length of their branches is so long. All other lines in this figure (shown in light blue and light red) indicate other species with shorter life spans.

Since the phylogenetic tree of each run was huge, containing thousands of species and all the speciation events, it was impossible to manually trace a set of sister species with different food preferences. As such, an algorithm was developed to distinguish all sets of sister species that met both of the following conditions: i. one of them was specialized on Food 1, while the other one showed preference for Food 2; and ii. both were able to survive for at least 100 time steps.

The speciation event in EcoSim is determined by a two-means clustering method. Therefore, at any speciation event only two sister species emerge from a parent species. This means that in cases where there is potential for the emergence of more than two sister species, it is possible to observe two consecutive speciation events within a very short period of time. In such cases, these species with such sequential speciation events are still considered as sister species as long as the difference between their originating time steps is less than five.

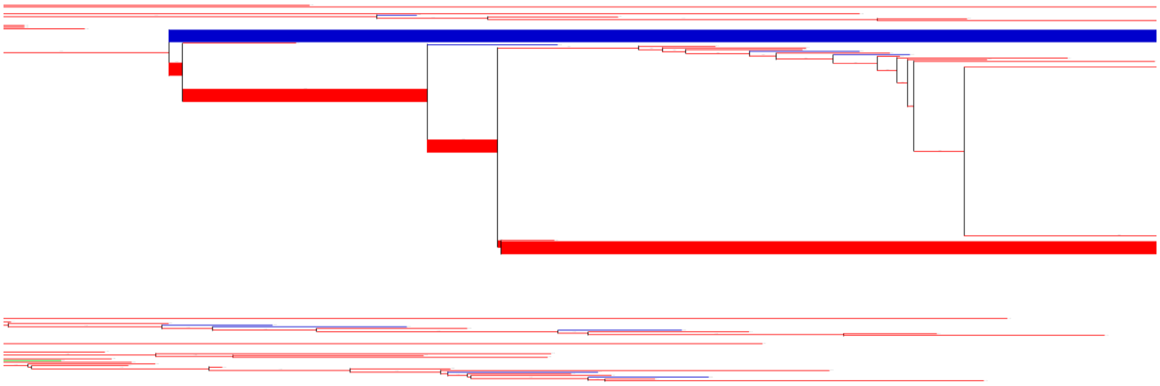


Figure 2.3. A truncated phylogenetic tree centered on one species splitting in two sister species each expressing a preferential resource behavior for Food 1(blue bold) or Food 2 (red bold). The length of the branches are proportional to the number of time steps. A set of sister species presented in bold color (red and blue) is detectable in this figure, each indicating different food preferences (in each color). The branch in the color of bold blue belongs to Group 1 with preference for Food 1, whereas its sister species, the branch in the color of bold red, belongs to Group 2 specialized on Food 2 (Bandehbahman 2014).

2.2.5.2 Second Criterion: Complete Reproductive Isolation

The second criterion was to ensure that there was a complete reproductive isolation between sister species that had already passed the first criterion. This criterion quantified the extent of reproductive isolation between sister species. The level of reproductive isolation between two sister species could be determined by measuring the number of occurrences of hybridization events. In other words, reproductive isolation level would be low if sister species frequently mate with each other and reproduced hybrid offspring. Therefore, measuring hybridization events was used to determine the level of reproductive isolation between sister species. The hybridization events were calculated as a ratio of all reproductive events that had occurred between all individual members of two sister species through evolutionary time. This ratio, then, measured intra- and inter-specific reproduction events. As the parents of each single individual were trackable in our simulation study, we only needed to go through all individual members of each sister species (that had already passed the first criterion) and calculate the ratio of intra-specific reproduction versus inter-specific reproduction occurring at each time step. The calculated hybridization ratio indicated that there were no occurrences of hybridization events between identified sister species from the first criterion.

2.2.5.3 Third Criterion: Overlapping Geographic Ranges

Spatial distribution of species was also examined to ensure that the two sister species occupied the same geographic habitat. In nature, dispersal ability of all individuals of one particular species determines the spatial extent of the habitat occupied by that species (Bolnick and Fitzpatrick 2007). In studies focusing on resource distribution or host-plant mediated interactions, what matters is the dispersal ability of every single individual rather than the average of the population’s dispersal ability as a whole. To validate our third criteria, it was necessary to verify that speciation events occurred among individuals sharing the same geographical range. Thus, for all individuals belonging to either of two sister species (that had passed the first and the second criteria), the average distance was measured in number of cells for the first 200 time steps after the occurrence of a speciation event. Using this information we were able to calculate the minimum distance between the two closest individuals, the average distance of the 200 closest individuals, and the average distance between all the individuals in either sister species to determine the level of geographical closeness of species (Figure A6 in the Appendix).

Furthermore, to get an idea about the distance between the set of candidate sister species (that had already passed the first and the second criteria) compared to the distance between all other sister species in the simulation (that had not arisen through sympatric speciation), the above parameters (the minimum distance, the average distance of 200 closest, and the total average distance) were also measured between all other sister species. The measurement of the minimum and the average distance between all other sister species provided an estimation about the level of cohabitation. Thereby, we could define a threshold for the highest acceptable minimum and maximum distances between the individuals of the candidate sister species. These thresholds could be, ultimately, used to examine the third criterion. In other words, it is crucial to know: i. what is the highest acceptable minimum distance, and ii. the maximum acceptable average total distance between the individuals of the sets of candidate sister species. According to the obtained results, the average geographic distance between individuals of the candidate sister species was significantly less than the average distances between all other sister species. Furthermore, in order to make sure that this important criterion (shared geographic habitat) was met, the statistical significance of the distances between every set of candidate sister species and all other sister species were also calculated through a T-Test. The result of this T-Test demonstrated that the distances between the candidate species (species that had already passed the first and the second criteria of sympatric speciation) were significantly differentiated from the distances between all other sister species. More importantly, the thresholds were estimated; i. the minimum distance between the individuals of the sets of candidate sister species and the average distance between their closest 200 individuals must be zero (less than 0.01) during the first 50 time steps after the speciation. Also, ii. the total average distance between the sister species populations must be less than 13 during the same time (the first 50 time steps after the speciation).

In summary, as the third criterion, the distances between individuals of the candidate sister species (all couples of sister species, which had already successfully passed the two previous required sympatric speciation criteria) were measured.

When the distances for the individuals of the candidate sister species were equal or below the thresholds, this couple of sister species were considered to have passed the third required criterion, which implies that this particular couple of sister species occupies the same geographical habitat. More precisely, if the minimum distance between individuals of a couple of sister species and the average distance between their closest 200 individuals was 0 during the first 50 time steps after the speciation event, and also at the same time the total average distance between their populations was less than 13, this couple of sister species, then, fulfilled the third criterion.

2.2.5.4 Fourth Criterion: Rejecting Alternative Hypothesis (Allopatric/Parapatric Speciation)

In evolutionary modeling studies, it has been proven that sufficient evidence of the biogeography and evolutionary history of a sister species couple is required to validate the emergence of a new species through sympatric divergence, and reject the possibility of their resulting from allopatric or parapatric processes (Bolnick and Fitzpatrick 2007; Coyne and Orr 2004). In this study, the biogeography of the two species in relation to one another was taken into account under the third criterion and the species' phylogenetic lineage was examined through the first criterion.

The last required criterion was to reject the alternative hypothesis of allopatric and parapatric speciation, to attest that the two species supposing to have arisen through sympatric speciation have not undergone any geographic isolation. One of the advantages of this study is that it was possible to track all the phylogenetic and biogeographic information of every single individual within the populations. As a result of such a population tracking capability, sampling errors that are intrinsically unavoidable in experimental investigations were eliminated from this modeling study. This study enables us to follow the complete biogeographic and phylogenetic history of all species through evolutionary time. Furthermore, there were no physical barriers in EcoSim that could restrict individuals' dispersal and movement to isolate the populations. As such, as soon as the first three criteria are met, the fourth criterion is also automatically met, and consequently, the possibility of the contribution of alternative hypothesis (allopatric/parapatric divergence) is contradicted.

2.2.6 Experimental Conditions

In order to detect the implications of resource partitioning on sympatric speciation, more than 50 runs of the dual resource version of the EcoSim with different initializations in terms of the foods' specifications were executed on SHARCNET[1]. Each run was executed for about three months and provided 25000 time steps, which was long enough to observe the evolutionary behavior of the species. The process of evaluation of simulations for monitoring speciation phenomena was started at time steps 15,000 – 20,000, when populations had enough time to stabilize. All necessary data was stored individually for each simulation. Furthermore, 10 runs of the classic version of the EcoSim with only one food resource were also submitted as the control. The initial number of prey and predator in each run was 12000 and 4900 respectively.

2.3. Results and Discussion

The Action-Perception Clustering Approach (which categorized species into three groups based on the actual behavior of the individuals) provided a significantly higher number of sister species fulfilling the sympatric speciation requirements compared to the FCM-Clustering Approach (which categorized species into three groups based on their FCM behavioral model). Under the FCM-Clustering Approach, only between 1 and 4 instances of sympatric speciation were tracked in each run. However, under the Action-Perception Clustering Approach, the number of observed instances of sympatric speciation in each run was between 11 and 53 (Table 2.3). The reason behind such a difference is that the FCM-Clustering Approach did not differentiate between the importance of the concepts influencing the Eat 1 and Eat 2 actions. For instance, some genes may be associated with an important concept in an FCM, which has a positive influence on eating action. At the

---

1. This work was made possible by the facilities of the Shared Hierarchical Academic Research Computing Network (SHARCNET): www.sharcnet.ca and Compute/Calcul Canada.

same time, other genes may be related to a less important concept, which could then negatively influence on eating action. However, the importance of the concepts corresponding to genes influencing the Eat 1 and Eat 2 actions was not included in calculating the weighted sum of the genes. Therefore, the weight effect of an important positive gene was not compensated for by a negative less important gene with the same absolute weight of influence on the eating action. In consequence, some species specializing on one specific food resource may not have been found by simply examining their FCM through the FCM-Clustering algorithm. In general, five runs with more than ten candidate instances of sympatric speciation, seven runs with one or two candidate instances, and eight runs with no candidate instance of sympatric speciation (the total number of submitted test runs was twenty) were observed in total.

The three criteria were implemented on the five runs with the highest number of observed instances of sympatric speciation. Table 2.3 summarizes how speciation events were filtered step by step. As it can be observed, most of the speciation events have been filtered out after applying the first criterion, leaving the sister species that were specialized on different food resources and that had a life span greater than 100 time steps. Interestingly, all sets of sister species that passed the first criterion also successfully met the second required criterion (they were also found to be reproductively isolated). In some runs, a small number of sister species that had passed the first and the second criteria, failed to fulfill the third criterion since they lived too far from each other (Table 2.3).

The results of these five runs were used to create a dataset to investigate the probability of the occurrence of sympatric speciation. Although we observed very promising results in all runs, presenting all the results obtained from these five runs is beyond the scope of this study. The results presented here focus on run 4 since this run had the highest number of the occurrences of sympatric speciation. However, we obtained similar results for the other four runs as well.

Table 2.3. Initial number of sister species and the number of sister species that successfully met each of the required criterion in five runs with the most promising results of the occurrences of sympatric speciation (Bandehbahman 2014)

	Sets of sister species						
Runs	Initial number of sister species	After applying the first criterion		After applying the second criterion		After applying the third criterion	
		FCM-Clustering	Action-Perception Clustering	FCM-Clustering	Action-Perception Clustering	FCM-Clustering	Action-Perception Clustering
1	8449	2	12	1	12	1	12
2	9106	1	13	1	13	1	11
3	10173	1	19	1	19	1	17
4	10880	4	53	4	53	3	47
5	9770	2	15	2	15	2	15

2.3.1 Obtained Results From Run 4

The total abundance of the different food resources is shown in Figure 2.4. As mentioned earlier, Food 1 had a relatively higher probability of diffusion and grew faster than Food 2. On the other hand, Food 2 was less available but was a more valuable resource regarding the amount of energy transmitted to prey.

As described earlier, we were able to track the rate of any successful or failed action performed by prey individuals. The rate of successful or failed searching action for Food 1 and Food 2, (as a ratio to all performed actions by all prey individuals at each time step), is represented in Figure 2.5 for the two food resources. The very low level of a failed searching for food action shows that prey individuals in this run did not face any difficulties in finding either of the food resources.

Another important action that was investigated in this study was the eating action performed by prey individuals, feeding on two different food resources. Figure 2.6 indicates the ratio of the number of successful or failed eating actions performed to the total number of all actions performed by all prey individuals at every time step.

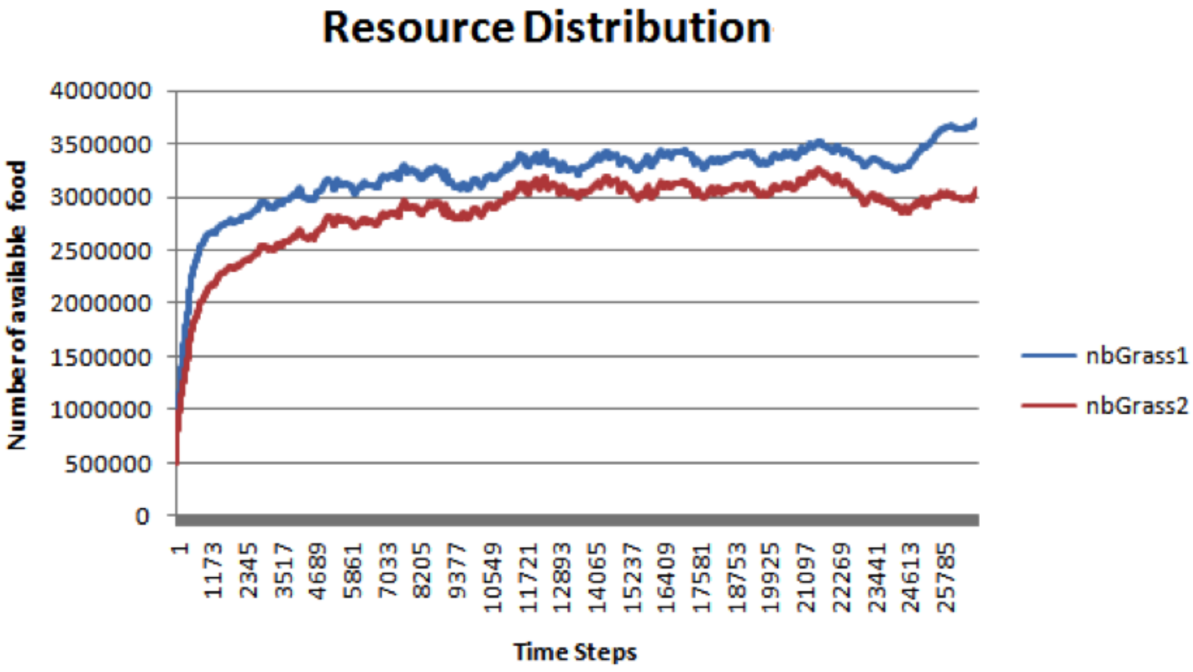


Figure 2.4. The total resource abundance of Food 1 (blue) and Food 2(red) in different time steps (Bandehbahman 2014)

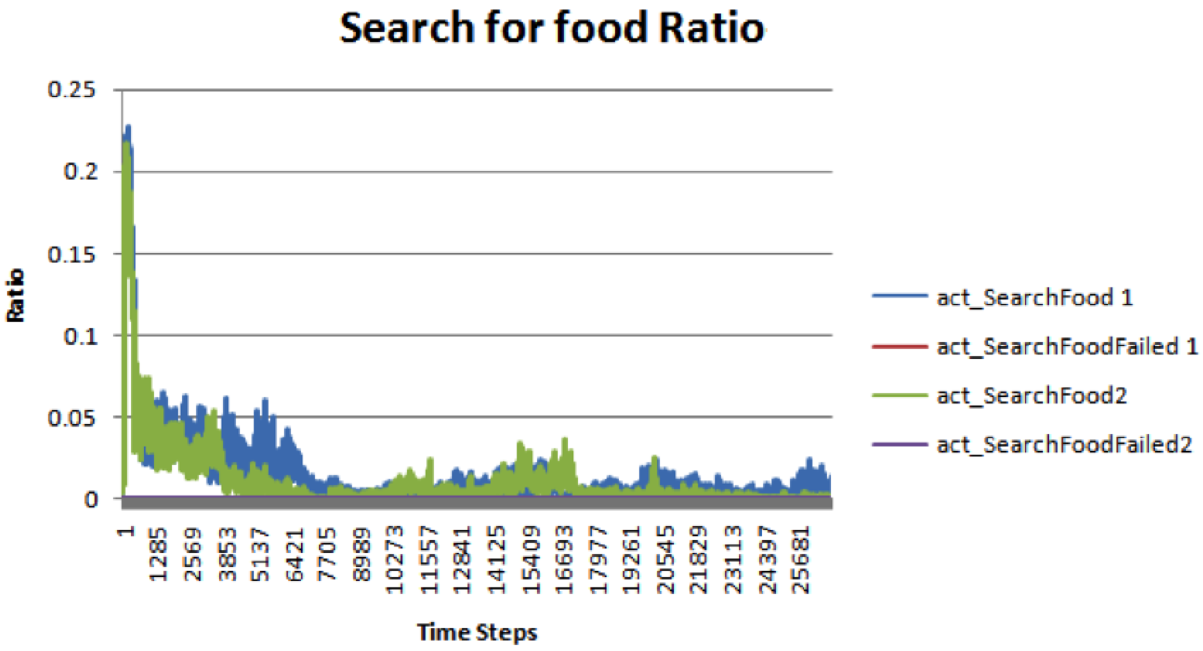


Figure 2.5. The success or failure of searching for each food resource as a ratio to all actions performed by all prey individuals at every time step (Bandehbahman 2014)

Initially, the rate of eating Food 1 is significantly higher than the rate of eating Food 2 (Figure 2.6). This is because at the beginning of the simulation, prey individuals were not specialized on any specific food resource and they simply fed on the most available food resource (Food 1) (Figure 2.4). Starting approximately from time step 20,000 the rate of eating Food 2 suddenly built up (an increasing trend for Eat 2 action; Figure 2.6), and at the same time, an evident decreasing trend for the Eat 1 action occurred. As such, the ratio of these actions (Eat 1 and Eat 2) crossed each other near time step 22,000. Accordingly, from time step 22,000 the rate of the Eat 2 action was clearly higher than the rate of Eat 1 (Figure 2.6). Initially, there was no food consumption specialization and the majority of individuals consumed the more abundant food. However, after the occurrence of food specialization at time step 22,000, the consumption rate of Food 2 was greater than that of Food 1, although Food 2 was less available than Food 1. This means that, although there were higher costs and risks associated with the exploitation of Food 2 (such as “longer search time, vulnerability to variation in habitat abundance, etc.” (Kawecki 1997)), specialization evolved nevertheless.

Resource preference distribution for Food 1 (blue), Food 2 (red), and for both food resources (green) is illustrated in Figure 2.7. Starting near time step 22,000, a large proportion of the prey population specialized on Food 2 despite a higher availability of Food 1 (Figure 2.7). This explains the observed increase in the Eat 2 action after time step 22,000 (Figure 2.6). Looking back to Figure 2.4, it is demonstrated that while the difference between the availability levels of Food 1 and Food 2 follows a steady trend, starting from time step 22,000 this difference begins to increase, which reflects the effect of the preference for Food 2.

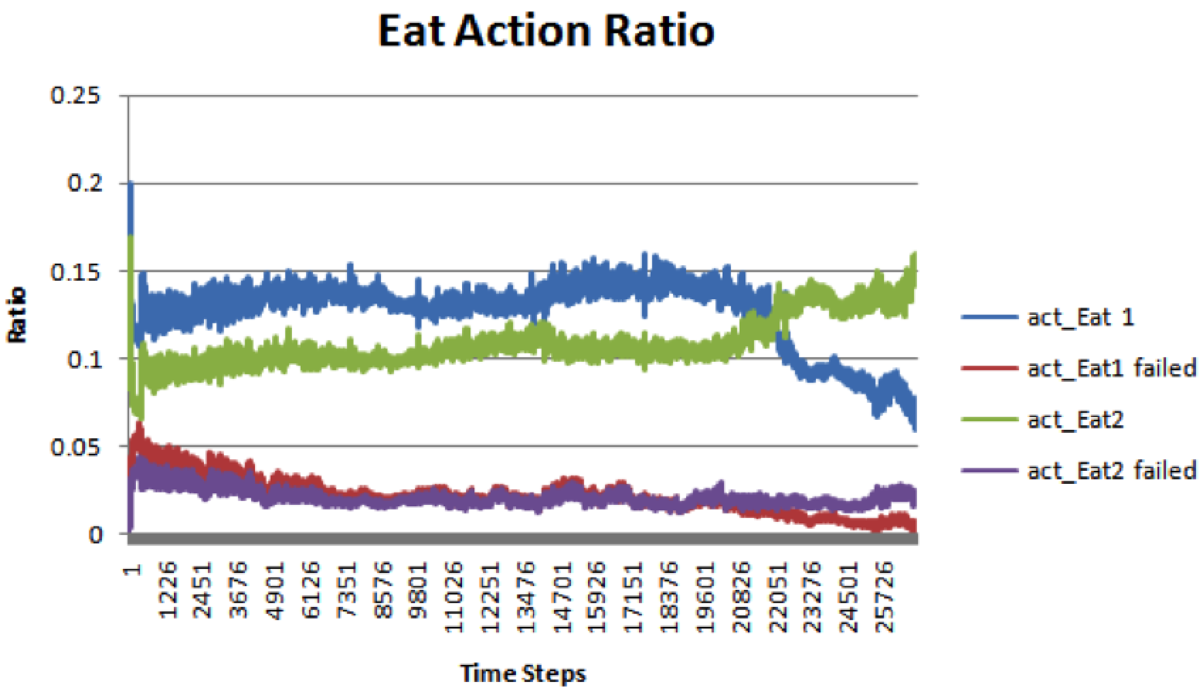


Figure 2.6. The successful and failed eating action on each type of food resource as a ratio to all actions performed by all prey individuals at every time step of the simulation (Bandehbahman 2014)

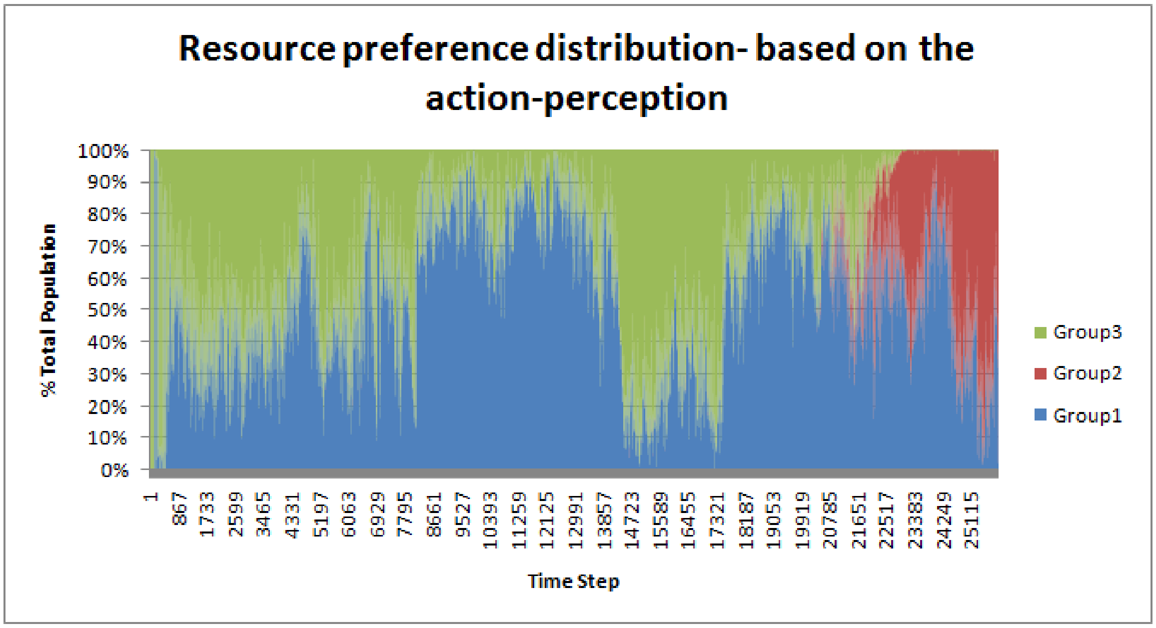


Figure 2.7. Resource preference distribution for Food1 (blue), Food2 (red) or both resources (green). Each individual's preference from the total prey population is calculated at every time step for the duration of the simulation (Bandehbahman 2014)

2.3.2 Comparing Sympatric Sister Species with Non-sympatric Sister Species

All sets of sympatric sister species (that had passed sympatric speciation requirements) were compared with all other sets of sister species (called non-sympatric sister species) in the simulations that failed to meet at least one sympatric speciation requirements. In other words, all sets of sister species (either sympatric or non-sympatric) with a minimum lifespan of 100 time steps in the duel resource version were compared with each other in terms of the hybridization ratio (between sister species) and the average geographical distance (between sister species) following the application of the same methods employed for testing the second and the third required criteria for the occurrence of sympatric speciation. Obtained results enabled us to draw a comparison between sympatric and non-sympatric sets of sister species in terms of the reproductive isolation level and the amount of geographical overlapping. This potentially illustrates the importance of required conditions for sympatric divergence.

As it is indicated in Table 2.3, there were five runs that each contained more than 10 candidates for the occurrence of sympatric speciation. These runs were used to calculate the hybridization ratio between the individual members of the sister species as well as the average geographical distance between their individuals. These distances were calculated for all sets of sister species with a minimum lifespan of 100 time steps (Figure 2.8, a and b).

Figure 2.8 represents the scatter plot (a) and logarithmic plot (b) of the hybridization ratio and the average geographical distance between all individuals of all sister species in the dual resource simulations. Each red circle represents a candidate couple of sympatric sister species showing the level of the hybridization ratio between their populations and the average geographical distance between their individuals. Each green circle indicates a couple of sister species which failed to meet at least one of the required criteria and are thus categorized as non-sympatric sister species. According to this figure, the sympatric sister species (red circles) are strongly clustered in the lower left part of the graph, whereas the non-sympatric sister species (green circles) are distributed along the two axes. This means that the non-sympatric sister species were either not completely reproductively isolated or that they lived in a non-overlapping area. The differences between sympatric and non-sympatric sister species are even stronger when presented in the form of logarithmic plot

(Figure 2.8, b). For the occurrence of sympatric speciation, divergent species are required to inhabit the same habitat and share the same geographical range as their common ancestral species (Berlocher and Feder 2002; Coyne 2007; Ridley 2004). Therefore, it was expected to observe sympatric species exclusively in overlapping geographical habitats, in the absence of geographical isolation. More importantly, this study demonstrated how exploiting different resources could exert disruptive selective pressure. This process thereby induces the formation of barriers to gene flow (reproductive isolation) and consequently, the emergence of new sympatric species (Kautt, Machado-Schiaffino, and Meyer 2016; Rice and Salt 1990; Rundle and Nosil 2005; Schluter 2001). This modeling study therefore indicates that sympatric speciation could result from assortative mating driven by differential resource use as a divergent selective pressure. In this study, we showed that ecological divergence in the form of diverse feeding preferences and differential foraging behavior could lead to reproductive isolation and thereby, the emergence of sympatric species. Similar results have been observed in natural populations. For instance, 11 different cichlid species all share the same ancestral species (tilapiines cichlid). Each of these species specializes on a particular food resource. Schliewen, Tautz, and Pääbo (1994) proved that these species have sympatrically diverged from their common ancestor species (Schliewen, Tautz, and Pääbo 1994). Furthermore, two species of three-spined sticklebacks have been verified to have arisen through sympatric speciation as a result of becoming specialized on different food resources (Boughman, Rundle, and Schluter 2005; Rundle and Schluter 2004; Rundle et al. 2000; Schluter and McPhail 1992; Svanbäck and Schluter 2012).

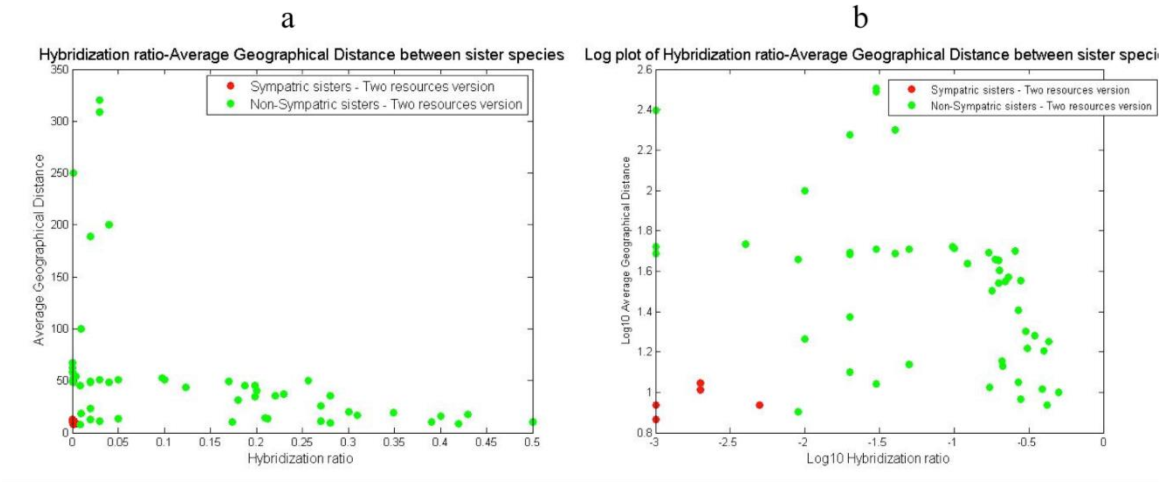


Figure 2.8. The scatter plot (a) and logarithmic plot (b) of the hybridization ratio and the average geographical distance between all individuals of sister species in the dual resource version of EcoSim. Red circles represent sympatric sister species, while green circles shows non-sympatric sister species (Bandehbahman 2014).

Moreover, all sets of sister species in the dual resource simulation that had a minimum lifespan of 100 time steps were compared with all sets of sister species with a lifespan of more than 100 time steps in the single resource simulation (control simulations). In order to illustrate the importance of the presence of two different food resources in sympatric divergence, a comparison was made between all sets of sister species from the dual and single resource simulations (Figure 2.9, a and b).

Figure 2.9 shows the scatter plot (a) and logarithmic plot (b) of the hybridization ratio and the average geographical distance of the sister species for both the dual and single resource simulation versions of the EcoSim. Again, the differences are easier to spot in the logarithmic plot (Figure 2.9, b). The blue circles show all sister species in five classic runs of the single resource version of EcoSim, while the red and green circles correspond to the dual resource version of EcoSim. Again, the red circles indicate sympatric sister species and the green circles signify non-sympatric sister species. According to this figure, there are no examples of sister species fulfilling the required criteria in the single resource runs, and species that met the required criteria are all from the dual resource simulations. Therefore, this model demonstrated that divergent foraging behavior could potentially result in reproductive isolation between sister species and eventually lead to sympatric speciation.

This study indicates how environmental variation in the case of diverse resource acquisition could play a very fundamental role as the main driver of divergent selection leading to the evolution of sympatric races. This observation supports previous claims regarding the crucial role of “ecologically-based divergent selection” (Rundle and Nosil 2005) and divergent selection caused by environmental variances (Schluter 2000) in the evolution of sympatric species.

When one population is offered different choices of food resources, a proportion of the population may begin exclusively exploiting one particular resource, and this could initiate a barrier to gene flow between this part of the population and the main population. That is why natural selection is considered the most central factor in the emergence of new species (Feder et al. 1997; Feder, Chilcote, and Bush 1988; Filchak, Roethele, and Feder 2000; Hendry and Kinnison 2001; Nosil, Crespi, and Sandoval 2002; Rundle et al. 2000). Our observation is also consistent with studies that consider ecological interactions to have an extremely important role among living organisms as a source of divergent selection in sympatric speciation (Rundle and Nosil 2005; Turelli, Barton, and Coyne 2001).

These results therefore support the main hypothesis of this modeling investigation regarding the importance of the presence of multiple resources in sympatric divergence. It has been established that different local environments could result in the evolution of distinct characteristics, and consequently lead to the emergence of sympatric species (Schemske and Bradshaw 1999). In fact, specialization on different food resources exerts varying extents of ecological forces that lead to the emergence of prezygotic isolation through natural selection (Turelli et al. 2001). African Finches (*Pyrenestesostrinus*), Salamander (*Ambystomatigrinum*), and Arctic Charr (*Salvelinusalpinus*) are typical examples of vertebrate species that have indicated discrete intraspecific morphs, varying in food and habitat preference, and have evolved to exploit diverse resources (Skulason and Thomas B Smith 1995). Indo-pacific goby and its sister species are another example that could clearly illustrate the fundamental role of foraging behavior in a lineage-splitting event. Scientists have identified a brand new species of goby within the range of the Indo-pacific goby species' habitat that is in fact its sister species and is exclusively specialized on a distinct coral host (Munday, van Herwerden, and Dudgeon 2004).

Reproductive isolation or the emergence of barrier to gene flow might occur either before or after the formation of a hybrid zygote (respectively called the prezygotic or postzygotic isolating mechanisms) (Ridley 2004). It is believed that compared to postzygotic (e.g. hybrid sterility), a prezygotic isolation (e.g. behavioral mating preference), which is considered an "earlier-evolving barrier to gene flow", plays a more significant role in the speciation process (Network 2012).

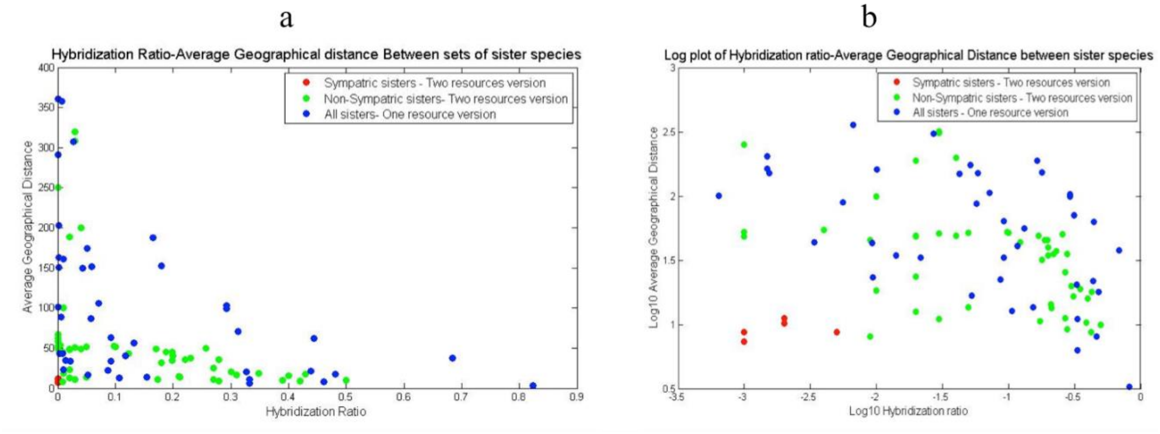


Figure 2.9. The scatter plot (a) and logarithmic plot (b) of the hybridization ratio and the average geographical distance of the sister species for the dual and single resource versions of the EcoSim. The blue circles represent all sister species in five classic runs of the single resource version. The red and the green circles indicate sympatric and non-sympatric sister species respectively (Bandehbahman 2014).

2.3.3 Do Sympatric Species Share Some Common Patterns?

We applied machine learning techniques to find the shared patterns among sympatric species in the five runs with more than 10 instances of sympatric speciation events. As such, three steps were followed (preparing the dataset, attribute selection, and classification), to analyze the results of these runs for further detailed information concerning the specific conditions leading to sympatric speciation.

2.3.3.1 Preparing and Preprocessing the Dataset

The results obtained from the five runs that had a high number of occurrences of sympatric speciation were used as the main dataset for applying the machine learning methods. In this dataset, sympatric species were labeled as positive instances, while other sister species at the same period of time were marked as negative instances. Initially, we included all attributes describing the species and their environment to create the initial dataset. These 81 attributes covered a broad range of information including general species information (such as population size of each species, their interbreeding ratio, and the amount of their energy transferred to an offspring), and behavioral specifications (such as the frequency of each action, and an individual's perception of their environment).

Accordingly, five initial datasets were created from the five different runs. However, four of them were imbalanced, meaning that the number of positive samples was only one third of the number of negative samples. This can negatively affect the machine learning method's ability to discover significant rules. One main approach to solve the imbalanced dataset problem is to either oversample the minority class or undersample the majority class (He and Garcia 2009). Therefore, for those four imbalanced datasets, we applied the smote algorithm (Chawla and Bowyer 2002) to resample the minority class, which corresponded to our sympatric species (positive samples). After balancing the datasets, each had approximately 6000 to 7000 instances, where each instance contained the values of all the attributes describing one species (either in the positive or negative class).

2.3.3.2 Attribute Selection

Each attribute describes one particular characteristic about a species, but not all attributes impact sympatric speciation. Thus, the most influential attributes were identified to classify the datasets in a way that will generate the most accurate results. Consequently, different attribute selection methods were used and their results were combined to select the attribute subset that most significantly discriminates between the two classes. We used the Info Gain Attribute Evaluator implemented in Weka (Hall et al. 2009), combined with the Ranker search method and Cfs subset Evaluator in three different search methods (including Best First, Greedy Stepwise, and Genetic Search) (Hall et al. 2009). Subsequently, all attributes were sorted by their corresponding scores, returned from the Ranker plus Info Gain Attribute Evaluator. The Ranker, combined with the Info Gain Attribute Evaluator, assigned a score to each attribute based on their relative importance for the learning process. The lower the rank of an attribute, the higher the importance. The Best First search method combined with the Cfs subset Evaluator only selected 8 attributes, corresponding to attributes already having a high importance based on the Ranker and Info Gain Evaluator combination. The Greedy Stepwise method combined with the Cfs subset Evaluator also returned a rank for the first 20 most important attributes. The Genetic Search method combined with the Cfs subset Evaluator was applied on a 10-fold cross-validation attribute selection basis. If an attribute was selected by evaluation of all 10 folds, a score of 100% was assigned to that attribute. Similarly, if an attribute was not selected by the evaluation of any fold, a score of 0% was assigned to that attribute. Accordingly, the attributes with the lowest score from all the attribute selection methods were removed. For this purpose, we removed attributes with a score of less than 30% from the Genetic Search and Cfs subset Evaluator, or with a rank higher than 40 on the Ranker and InfoGain attribute Evaluator. Since the removed attributes also had a low score in the GreedyStepwise+Csf method, they were not selected by the BestFirst+Csf method. As a result, the number of the attributes was reduced to 29. The list of these attributes is provided in the Appendix (Table A6 and A7).

2.3.3.3 Classification

2.3.3.3.1 Specific Rules Associated to Each Run

The J48 classifier in Weka (Hall et al. 2009), the CRF combined rule extraction and feature elimination method in supervised Random Forest classification (Liu et al. 2012), and the Random Forest classification combined with feature selection using hill climbing method (Mashayekhi and Gras 2015) were applied to each dataset individually to find a fit method for classification.

First, each dataset was tested separately to extract the rules on each run. Then, all datasets were combined to identify the shared patterns among all runs. The J48 classification method returned a lower number of rules than the Random Forest methods. However, the Random Forest method provided the highest level of accuracy of classification, whereas the accuracy obtained with J48 was still reasonably high. Hence, we decided to use the J48 classifier to classify each dataset separately since it returned the lowest number of rules with a high accuracy.

J48 classifier was used with different attribute selection methods to find the minimum number of attributes, the minimum number of rules, and the highest accuracy. The classification started with the 29 attributes, selected using the attribute selection method (section 2.3.3.2). We pruned the decision tree by increasing the minimum number of instances per leave as this technique helped us to decrease the number of rules, which facilitated an explanation of the rules related to each class. A small part of each dataset was put aside to be utilized as a validation set. Hence, each dataset went through each step (pruning and removing attributes) separately. Starting with 29 attributes and 17 rules, it was possible to reduce down to 5 attributes and 11 rules. Consequently, the total accuracy declined from 96.26% to 86.79%, with the advantage of obtaining a reasonable number of short rules for interpretation. However, an accuracy greater than 86% is sufficient to capture the main properties and to provide a primary analysis of the conditions leading to sympatric speciation. As an example, the decision tree generated for Run #2 is presented in Figure 2.10 (the trees obtained from the other runs are given in the Appendix, Figure A7-A11). As is noticeable in this example, sympatric speciation has occurred at low values of disEvol (the average genetic distance between the initial reference genome and the current genomes). The evolutionary distance (disEvol) is always increasing with time; therefore, a low value of disEvol represents the beginning of the data measurement near time step 20,000 when the food specialization process was about to begin. This means that sympatric speciation has occurred at the beginning of the food specialization process, when an initial specialization on different food resources was developing (Figure 2.10).

The same pattern was observed in all other generated decision trees (Figure A7-A11).

The rules generated by the decision tree for this run (Run #2) demonstrated that sympatric speciation had mostly occurred at the beginning of the food specialization (disEvol low,) except when the species' spatial distribution was large (diversitySpatialRatio high). Under this circumstance, two different reinforcements were needed for the occurrence of sympatric speciation. First, sister species required a high number of genes in their genomes (nArc high). This is intuitive since more genetic diversity results in a higher mutations rate and thereby, drives a faster genetic divergence. Kawecki (1996, 1997) illustrated the importance of the accumulation of beneficial or deleterious mutations corresponding to habitat and resource exploitation. His research showed that disruptive selection through habitat-specific deleterious or beneficial mutations could result in sympatric speciation (Kawecki 1996, 1997). It has been proven that the expression of a habitat preference behavior could be spread among the gene pool of an initially random dispersing population via beneficial (Diehl and Bush 1989; Kawecki 1996; Rice 1987; Smith 1966) or deleterious (Kawecki 1997) mutations, when selective pressure favors habitat preference over generalism. This eventually leads to the evolution of polymorphism and sympatric divergence.

The second condition occurred when species contained a large number of individuals (individualRatio high). This means that species with a larger population size (compared to the whole populations of all species living in the simulation's world) had a higher chance of experiencing sympatric speciation. This observation supports the claim that the extent of genetic diversity builds up with an increasing effective population size (Frankham 1996). Additionally, as mentioned above, such increased genetic variability leads to a more diverse ancestral gene pool and thereby, increases the chance that sympatric speciation will occur (Kawecki 1996, 1997).

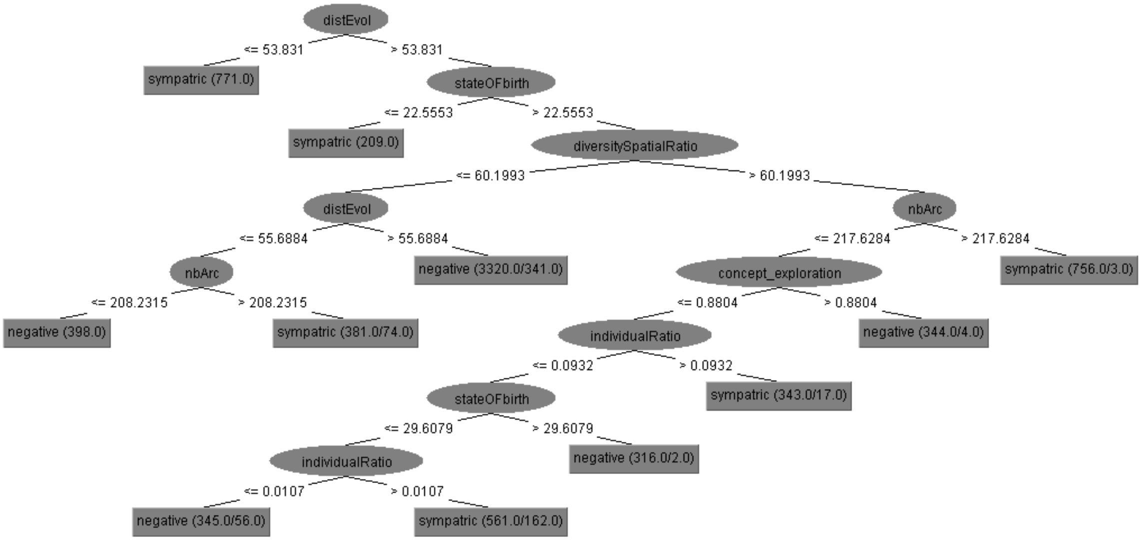


Figure 2.10. Decision tree corresponding to Run #2 with 11 rules (Bandehbahman 2014)

2.3.3.3.2 Generic Rules Valid for All Runs

The results of all the five runs were united to create a dataset to identify the shared patterns between all their sympatric species. The validation set consisted of 30% of the dataset put aside. Two methods of feature selection (the Info Gain Attribute Evaluator implemented at Weka (Hall et al. 2009) with the Ranker Search method, and the Cfs subset Evaluator with the Genetic Search method) were employed. Initially, 81 attributes were present in the dataset. First, the attributes were removed with scores less than 30% in the Cfs subset Evaluator with Genetic Search method or those with a rank higher than 30 in the Info Gain Attribute Evaluator with the Ranker search method were removed. As a result of the first step of feature selection, 25 attributes remained. Although a high number of attributes were removed, the total accuracy only dropped by approximately 1%, (from 97.25% [with 81 attributes] to 96.34% [with 25 attributes]). Accordingly, the number of rules decreased from 69 (with 81 attributes) to 56 (with 25 attributes).

In a second step, the J48 classification method was applied to the dataset with the remaining set of attributes. The tree pruning method was also utilized by increasing the minimum number of objects per leaf, which led to a decrease in the number of leaves and thereby, a decrease in the number of rules per class. The amount of pruning was chosen to significantly decrease the number of rules when keeping the total accuracy at a reasonable level.

The total accuracy marginally declined to 94.95% and the number of rules dropped to 42. These steps were repeated three more times and 13, 11, and 9 attributes were selected respectively after each step. The decision tree returned by the J48 classifier on all datasets combined together with 11 attributes and 20 rules is shown in the Appendix, Figure A11. In order to estimate how generic the discovered rules were, the classification process was repeated five more times. Each time the results of four out of the five datasets were united to use as the training set, while the results of the fifth dataset were exploited as the validation set. The attributes with the lowest score (as previously discussed) were removed step-by-step by applying the Info Gain Attribute Evaluator implemented in Weka (Hall et al. 2009) with the Ranker search method, and the Cfs subset Evaluator with Genetic Search method leading to the selection of 10 attributes. The J48 decision tree and Random Forest classification methods were also used in each experiment.

As was expected, the total accuracy of the validation set in this experiment was much lower than the total accuracy of the 10-fold cross validation on the training set. This was due to the validation set having been created from the results of a different run. We observed that the Random Forest method strongly outperformed the J48 algorithm on the validation set and had a consistently higher accuracy on the training set.

The averages of the classification results of the five experiments are summarized in Table 2.4, giving the TP; the True Positive rate and the AUC; the Area Under the Receiver Operating Characteristics (ROC) Curve. ROC curves are curves that are exploited in machine learning and data mining investigations with the purpose of both organizing classifiers and obtaining a clear visualization of their performance (Fawcett 2006). “An ROC curve is a two-dimensional depiction of classifier performance” (Fawcett 2006). In order to draw a comparison between classifiers, ROC performance needs to be decreased to “one single scaler representing expected performance” (Fawcett 2006). The AUC method is frequently used to measure the area under the ROC curve (Bradley 1997; Hanley and McNeil 1982). The AUC varies between 0 and 1, but a realistic classifier should not have an AUC less than 0.5 (Fawcett 2006). Applying the Random Forest method we can predict the occurrence of sympatric speciation on the training set with an average accuracy of 99.97%. Furthermore, the unseen validation sets from different runs obtained an average accuracy of 82.22%, which is considered a high accuracy, indicating that our method was able to discover very generic rules that have the potential to reflect some meaningful properties of sympatric speciation.

Table 2.4. The average results of five experiments of classification using J48 and Random Forest classification methods. For each experiment four out of five datasets were used as the training set, while the fifth dataset was used as the validation set (Bandehbahman 2014)

#Features	RandomForest----Training Set – 10 fold C.V.					Validation set				
	Total accuracy	TP Rate		AUC		Total accuracy	TP Rate		AUC	
10	99.97%	0.99	1	1	1	82.22%	0.63	0.95	0.89	0.89

2.4 Conclusion

There is still long-standing controversy surrounding sympatric speciation. Despite a general agreement on the theoretical plausibility of the incidence of sympatric divergence in nature, the extent that sympatric speciation may contribute to biodiversity and its root causes are still unknown today. It is believed that strong disruptive selective pressure exerted by both competition for and specialization on resources could play a significant role in sympatric divergence (Bolnick and Fitzpatrick 2007; Bolnick and Smith 2004; Kirkpatrick and Ravigné 2002; Martin and Pfennig 2009; Pfennig and Pfennig 2010; Roughgarden 1972; Thibert-Plante and Hendry 2011; Wilson and Turelli 1986). However, the importance of ecological interactions and consequent disruptive selection in sympatric speciation still needs further investigation.

In order to obtain a better understanding of the evolutionary impact of the arrival of a new species, and to investigate speciation and lineage-splitting events, it is necessary to have access to a species' complete evolutionary history including thousands of generations leading to a speciation event (Berlocher and Feder 2002; Bolnick and Fitzpatrick 2007; Coyne and Orr 2004). Achieving such insight is challenging by means of experimental and field investigations due to the unreasonable time investment for such a field study. Therefore, in this study we utilized the ability of an individual-based modeling approach in tracking the evolutionary paths of species (Grimm and Railsback 2013).

According to the results of this investigation, prey individuals mainly fed on the more abundant resource (Food 1) at the beginning of the simulations, before they had adapted to efficiently exploiting each specific resource. However, after the evolution of specialization around time step 22,000, consumption of Food 2 exceeded that of Food 1 in spite of the fact that Food 1 was more available and prey individuals encountered this resource more frequently. The main focus of this study was to investigate whether and under which circumstances the selective pressures acting on foraging behaviors could sympatrically diverge lineages. Four different criteria suggested by Bolnick and Fitzpatrick (2007) were employed, and we detected an indicator of the occurrence of sympatric speciation in 12 of our runs out of 20. After testing these four required criteria to identify sympatric speciation in the dual resource simulations, sympatric and non-sympatric sister species with a minimum lifespan of 100 time steps were compared in terms of their level of reproductive isolation and amount of geographical overlapping (between individual members of the sister species). This was employed to obtain a better understanding of the underlying causes of sympatric divergence. As it was expected, the instances of sympatric species were exclusively observed among sister species that shared the same geographical ranges. Moreover, this comparison revealed the significant role of reproductive isolation and assortative mating caused by disruptive selection pressure exerted by the exploitation of different resources in sympatric speciation (Kautt et al. 2016; Rice and Salt 1990; Rundle and Nosil 2005; Schluter 2001). Comparing the results obtained from the dual resource simulations with the single resource control simulations highlighted the importance of divergent foraging behavior and consequent reproductive isolation in sympatric divergence. This is because there were no incidences of sympatric speciation in the single resource control simulations. This result is consistent with previous observations regarding the role of ecologically-based divergent selection and ecological interactions among living organisms in sympatric speciation (Feder et al. 1997, 1988; Filchak et al. 2000; Hendry and Kinnison 2001; Nosil et al. 2002; Rundle and Nosil 2005; Rundle et al. 2000; Schluter 2000; Turelli et al. 2001). The results of this study support the theoretical claim that reproductive isolation caused by assortative mating as a result of divergent selection pressures inflicted by resource differentiation could potentially lead to sympatric speciation (Coyne and Orr 2004; Forbes et al. 2009; Nosil 2013).

Our unique modeling approach does not simply assume that individuals are involved in foraging and mating activities; it also comprises all other possible considerations, which might play an important role from evolutionary perspective. Applying this complex modeling approach we highlighted significant indicators of behavioral modifications caused by preferential resource use. Finally, when employing the several machine learning techniques, explicit rules were extracted to gain more information regarding the most essential patterns that lead to sympatric speciation. According to our acquired rules, the majority of incidences of sympatric divergence occurred at the beginning of the process of resource specialization. However, if species had a high spatial distribution, they needed to fulfill two different conditions to diverge sympatrically: i. high genetic diversity, and ii. large population size. This means that the probability of sympatric divergence was higher if a population had a more diverse gene pool and also a higher number of individual members. It has been empirically verified that genetic conditions and ecological conditions are the key components that facilitate the occurrence of sympatric speciation (Bolnick and Fitzpatrick 2007; Via 2001). In the case of specialized resource use, genotype × environment interaction is the leading contributor to sympatric divergence (Via 2001). Our modeling study indicated the crucial role of these factors in the occurrence of sympatric speciation and stressed the importance of genetic diversity and population size.

One of the difficulties of empirical investigations of sympatric speciation is that it is almost impossible to reach a solid conclusion about ancestor species, as it is difficult to gain access to the genetic conditions of the initial population (ancestral species) prior to a divergence event (Barluenga et al. 2006; Schliewen et al. 2006). In most empirical studies addressing speciation, a speciation event has either completed and species have completely diverged, or it is currently happening. On the other hand, it has been claimed that the most accurate estimations about the initial conditions leading to sympatric speciation could be obtained from lineages that are beginning the divergence process (Martin 2012). Since modeling approaches provide us with an ideal opportunity to monitor speciation events at early stages of divergence, these tools are considered one of the strongest candidate approaches to achieve an accurate prediction of the initial requirements for speciation (Martin 2012). Our modeling investigation strongly supports this claim and illustrates the importance of an early stage of resource specialization in the occurrence of sympatric speciation. This modeling study provided us a golden opportunity to follow the speciation process since its initiation stage, something that is impossible in nature. The valuable obtained results of this study shed light on the central role of sympatric speciation in evolutionary ecology.

From a biological point of view, however, this modeling study has some limitations in spite of its major contributions to investigathis field of study. EcoSim is intrinsically designed to address broad ecological and biological questions and it is not able to exclusively model a specific ecological system or a distinct species with high specificity.

Furthermore, the extent of the complexity of interactions and behavioral patterns among components of a real ecosystem is much greater than that modeled in this simulated ecological system. More importantly, it has been demonstrated that phenotypic adaptation is a key in sympatric speciation caused by specialized resource use. For example, it has been verified that a population with different food preferences than its original population gradually evolves a dissimilar phenotypic structure (Ratciliffe and Grant 1983; Schluter 1996). This is evident in the evolution of phenotypic polymorphisms in amphibians, in terms of teeth length and mouth size in order to increase foraging efficiency (Skulason and Thomas B. Smith 1995); the evolution of different morphs in benthic and limnetic species of stickleback fish (Boughman et al. 2005; Rundle and Schluter 2004; Rundle et al. 2000; Schluter and McPhail 1992; Svanbäck and Schluter 2012); and the evolution of divergent shape and size in Darwin's finches to exploit different resources (Ratciliffe and Grant 1983; Schluter 1996). Because this modeling study did not include any particular phenotypic traits, we did not evaluate the role phenotypic adaptation and adaptive radiation may have in sympatric divergence caused by resource specialization. This would therefore be a beneficial area for future work. In the current study, we simply evaluated runs with sympatric species; however, it would be valuable in future work to also involve runs without sympatric divergence. This may allow us to obtain a greater understanding of the environmental and behavioral differences between these runs that might have led to sympatric speciation.

2.5 References

Aspinall, Adam and Robin Gras. 2010. “K-Means Clustering as a Speciation Mechanism within an Individual-Based Evolving Predator-Prey Ecosystem Simulation.” Pp. 318–29 in *Active Media Technology*. Springer.

Bandehbahman, Sara. 2014. “Investigating The Effects of Food Chain on Sympatric Speciation Using ECOSIM.”

Barluenga, Marta, Kai N. Stölting, Walter Salzburger, Moritz Muschick, and Axel Meyer. 2006. “Sympatric Speciation in Nicaraguan Crater Lake Cichlid Fish.” *Nature* 439(7077):719–23.

Berlocher, Stewart H. and Jeffrey L. Feder. 2002. “Sympatric Speciation in Phytophagous Insects: Moving beyond Controversy?” *Annual review of entomology* 47(1):773–815.

Bolnick, Daniel I. and Benjamin M. Fitzpatrick. 2007. “Sympatric Speciation: Models and Empirical Evidence.” *Annual Review of Ecology, Evolution, and Systematics* 459–87.

Bolnick, Daniel I. and T. Smith. 2004. “Can Intraspecific Competition Drive Disruptive Selection? An Experimental Test in Natural Populations of Sticklebacks.” *Evolution* 58(3):608–18.

Boughman, Janette Wenrick, Howard D. Rundle, and Dolph Schluter. 2005. “Parallel Evolution of Sexual Isolation in Sticklebacks.” *Evolution* 59(2):361–73.

Bradley, Andrew P. 1997. “The Use of the Area under the ROC Curve in the Evaluation of Machine Learning Algorithms.” *Pattern recognition* 30(7):1145–59.

Chawla, NV and KW Bowyer. 2002. “SMOTE: Synthetic Minority Over-Sampling Technique Nitesh.” *Journal of Artificial Intelligence Research* 16(1):321–57. Retrieved (http://arxiv.org/abs/1106.1813).

Coyne, Jerry A. 2007. “Sympatric Speciation.” *Current biology* 17(18):R787–88.

Coyne, Jerry A. and H. Allen Orr. 2004. *Speciation*. Sinauer Associates Sunderland, MA.

Diehl, Scott R. and Guy L. Bush. 1989. “The Role of Habitat Preference in Adaptation and Speciation.” *Speciation and its Consequences* 345–65.

Fawcett, Tom. 2006. “An Introduction to ROC Analysis.” *Pattern recognition letters* 27(8):861–74.

Feder, Jeffrey L., Charles A. Chilcote, and Guy L. Bush. 1988. “Genetic Differentiation between Sympatric Host Races of the Apple Maggot Fly Rhagoletis Pomonella.” *Nature* 336(6194):61–64.

Feder, Jeffrey L., Joseph B. Roethele, Brian Wlazlo, and Stewart H. Berlocher. 1997. "Selective Maintenance of Allozyme Differences among Sympatric Host Races of the Apple Maggot Fly." *Proceedings of the National Academy of Sciences* 94(21):11417–21.

Filchak, Kenneth E., Joseph B. Roethele, and Jeffrey L. Feder. 2000. "Natural Selection and Sympatric Divergence in the Apple Maggot Rhagoletis Pomonella." *Nature* 407(6805):739–42.

Forbes, Andrew A., Thomas H. Q. Powell, Lukasz L. Stelinski, James J. Smith, and Jeffrey L. Feder. 2009. "Sequential Sympatric Speciation across Trophic Levels." *Science* 323(5915):776–79.

Frankham, Richard. 1996. "Relationship of Genetic Variation to Population Size in Wildlife." *Conservation Biology* 10(6):1500–1508.

Golestani, Abbas and Robin Gras. 2011. "Multifractal Phenomena in EcoSim, a Large Scale Individual-Based Ecosystem Simulation." Pp. 991–99 in *Int. Conf. Artificial Intelligence, Las Vegas*.

Golestani, Abbas, Robin Gras, and Melania Cristescu. 2012. "Speciation with Gene Flow in a Heterogeneous Virtual World: Can Physical Obstacles Accelerate Speciation?" *Proceedings of the Royal Society B: Biological Sciences* 279(1740):3055–64.

Gras, Robin, Didier Devaurs, Adrianna Wozniak, and Adam Aspinall. 2009. "An Individual-Based Evolving Predator-Prey Ecosystem Simulation Using a Fuzzy Cognitive Map as the Behavior Model." *Artificial life* 15(4):423–63.

Gras, Robin, Abbas Golestani, Andrew P. Hendry, and Melania E. Cristescu. 2015. "Speciation without Pre-Defined Fitness Functions." *Plos One* 10(9):e0137838. Retrieved (http://dx.plos.org/10.1371/journal.pone.0137838).

Grimm, Volker and Steven F. Railsback. 2013. *Individual-Based Modeling and Ecology*. Princeton university press.

Hall, Mark et al. 2009. "The WEKA Data Mining Software: An Update." *SIGKDD Explorations* 11(1).

Hanley, James A. and Barbara J. McNeil. 1982. "The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve." *Radiology* 143(1):29–36.

He, Haibo and E. A. Garcia. 2009. "Learning from Imbalanced Data." *Knowledge and Data Engineering, IEEE Transactions on* 21(9):1263–84.

Hendry, A. P. and M. T. Kinnison. 2001. "An Introduction to Microevolution: Rate, Pattern, Process." *Genetica* 112(1):1–8.

Kautt, Andreas F., Gonzalo Machado-Schiaffino, and Axel Meyer. 2016. "Multispecies Outcomes of Sympatric Speciation after Admixture with the Source Population in Two Radiations of Nicaraguan Crater Lake Cichlids." *PLoS Genet* 12(6):e1006157.

Kawecki, Tadeusz J. 1996. "Sympatric Speciation Driven by Beneficial Mutations." *Proceedings of the Royal Society of London B: Biological Sciences* 263(1376):1515–20.

Kawecki, Tadeusz J. 1997. "Sympatric Speciation via Habitat Specialization Driven by Deleterious Mutations." *Evolution* 1751–63.

Khater, Marwa, Dorian Murariu, and Robin Gras. 2014. "Contemporary Evolution and Genetic Change of Prey as a Response to Predator Removal." *Ecological Informatics* 22:13–22.

Kirkpatrick, Mark and Virginie Ravigné. 2002. "Speciation by Natural and Sexual Selection: Models and Experiments." *The American Naturalist* 159(S3):S22–35.

Kosko, Bart. 1986. "Fuzzy Cognitive Maps." *International journal of man-machine studies* 24(1):65–75.

Liu, Sheng et al. 2012. "Combined Rule Extraction and Feature Elimination in Supervised Classification." *IEEE transactions on nanobioscience* 11(3):228–36.

Martin, Christopher H. 2012. "Weak Disruptive Selection and Incomplete Phenotypic Divergence in Two Classic Examples of Sympatric Speciation: Cameroon Crater Lake Cichlids." *The American Naturalist* 180(4):E90–109.

Martin, Ryan A. and David W. Pfennig. 2009. "Disruptive Selection in Natural Populations: The Roles of Ecological Specialization and Resource Competition." *The American Naturalist* 174(2):268–81.

Mashayekhi, Morteza, Abbas Golestani, Yasaman Majdabadi Farahani Farahani, and Robin Gras. 2014. "An Enhanced Artificial Ecosystem: Investigating Emergence of Ecological Niches." *ALIFE 14: The Fourteenth Conference on the Synthesis and Simulation of Living Systems* 14:693–700.

Mashayekhi, Morteza and Robin Gras. 2015. "Rule Extraction from Random Forest: The RF+ HC Methods." Pp. 223–37 in *Advances in Artificial Intelligence*. Springer.

Munday, Philip L., Lynne van Herwerden, and Christine L. Dudgeon. 2004. "Evidence for Sympatric Speciation by Host Shift in the Sea." *Current Biology* 14(16):1498–1504.

Network, The Marie Curie SPECIATION. 2012. "What Do We Need to Know about Speciation?" *Trends in Ecology & Evolution* 27(1):27–39.

Nosil, Patrik. 2013. “Degree of Sympatry Affects Reinforcement in Drosophila.” *Evolution* 67(3):868–72.

Nosil, Patrik, Bernard J. Crespi, and Cristina P. Sandoval. 2002. “Host-Plant Adaptation Drives the Parallel Evolution of Reproductive Isolation.” *Nature* 417(6887):440–43.

Pfennig, David W. and Karin S. Pfennig. 2010. “Character Displacement and the Origins of Diversity.” *the american naturalist* 176(Suppl 1):S26.

Ratciliffe, Laurene M. and Peter R. Grant. 1983. “Species Recognition in Darwin’s Finches (Geospiza, Gould) I. Discrimination by Morphological Cues.” *Animal Behaviour* 31(4):1139–53.

Rice, William R. 1987. “Speciation via Habitat Specialization: The Evolution of Reproductive Isolation as a Correlated Character.” *Evolutionary Ecology* 1(4):301–14.

Rice, William R. and George W. Salt. 1990. “The Evolution of Reproductive Isolation as a Correlated Character under Sympatric Conditions: Experimental Evidence.” *Evolution* 1140–52.

Ridley, Mark. 2004. *Evolution*. 3rd ed. Oxford: Blackwell Pub.

Roughgarden, Jonathan. 1972. “Evolution of Niche Width.” *American Naturalist* 683–718.

Rundle, Howard D., Laura Nagel, Janette Wenrick Boughman, and Dolph Schluter. 2000. “Natural Selection and Parallel Speciation in Sympatric Sticklebacks.” *Science* 287(5451):306–8.

Rundle, Howard D. and Patrik Nosil. 2005. “Ecological Speciation.” *Ecology letters* 8(3):336–52.

Rundle, Howard D. and Dolph Schluter. 2004. “Natural Selection and Ecological Speciation in Sticklebacks.” *Adaptive speciation* 19(3):192–209.

Schemske, Douglas W. and H. D. Bradshaw. 1999. “Pollinator Preference and the Evolution of Floral Traits in Monkeyflowers (Mimulus).” *Proceedings of the National Academy of Sciences* 96(21):11910–15.

Schliewen, U. K., T. D. Kocher, K. R. McKaye, O. Seehausen, and Diethard Tautz. 2006. “Evolutionary Biology: Evidence for Sympatric Speciation?” *Nature* 444(7120):E12–13.

Schliewen, Ulrich K., Diethard Tautz, and Svante Pääbo. 1994. “Sympatric Speciation Suggested by Monophyly of Crater Lake Cichlids.” *Nature* 368(6472):629–32.

Schluter, Dolph. 1996. “Ecological Causes of Adaptive Radiation.” *American Naturalist* S40–64.

Schluter, Dolph. 2000. *The Ecology of Adaptive Radiation*. Oxford University Press.

Schluter, Dolph. 2001. “Ecology and the Origin of Species.” *Trends in ecology & evolution* 16(7):372–80.

Schluter, Dolph and John Donald McPhail. 1992. “Ecological Character Displacement and Speciation in Sticklebacks.” *American Naturalist* 85–108.

Skulason, Skuli and Thomas B. Smith. 1995. “Resource Polymorphisms in Vertebrates.” *Trends in ecology & evolution* 10(9):366–70.

Skulason, Skuli and Thomas B. Smith. 1995. “Resource Polymorphisms in Vertebrates.” *Trends in Ecology & Evolution* 10(9):366–70.

Smith, J. Maynard. 1966. “Sympatric Speciation.” *American Naturalist* 637–50.

Stöver, Ben C. and Kai F. Müller. 2010. “TreeGraph 2: Combining and Visualizing Evidence from Different Phylogenetic Analyses.” *BMC Bioinformatics* 11(7).

Svanbäck, Richard and Dolph Schluter. 2012. “Niche Specialization Influences Adaptive Phenotypic Plasticity in the Threespine Stickleback.” *The American Naturalist* 180(1):50–59.

Thibert-Plante, X. and a P. Hendry. 2011. “Factors Influencing Progress toward Sympatric Speciation.” *Journal of Evolutionary Biology* 24(10):2186–96.

Turelli, Michael, Nicholas H. Barton, and Jerry A. Coyne. 2001. “Theory and Speciation.” *Trends in Ecology & Evolution* 16(7):330–43.

Via, Sara. 2001. “Sympatric Speciation in Animals: The Ugly Duckling Grows up.” *Trends in Ecology & Evolution* 16(7):381–90.

Wilson, David Sloan and Michael Turelli. 1986. “Stable Underdominance and the Evolutionary Invasion of Empty Niches.” *American Naturalist* 835–50.

CHAPTER 3

CONCLUSION

Ecological specialization plays a significant role in speciation. One of the remaining questions in the field of ecological socialization is “what are the main circumstances that cause a specialized population to diverge to the point that it becomes a new separate species?” (Caillaud and Via 2000). Individuals of each species are expected to efficiently exploit their available resources to reach their optimal capacities in terms of energy gain and reproduction success. Different parameters such as resource availability and resource value could potentially influence individual’s acquired benefits (Bolnick et al. 2003). Ultimately, the interactions between these parameters in addition to individual’s phenotype determine the extent of expressed preference by each individual for any alternative food resources. This preference is then regulated by some other factors such as environmental heterogeneity and social interactions to determine individual’s actual resource use (Bolnick et al. 2003). Comprising all of these influential factors along with their interactions in empirical and experimental investigations is quite challenging and problematic, whereas these elements and their relations could simply be involved and controlled in modeling investigations. It is also possible to analyze these components in modeling studies and determine their relative contribution to the context of specialized habitat or resource use. The modeling approach employed in this study has given careful consideration to such prominent components including prey-predator interactions, intra- and inter- competition, and social interactions. Thus, we were able to evaluate the leading role of each parameter in adaptive specialization.

Although individuals of each species are unique and differentiate from each other, empirical and theoretical investigations looking at resource use and its ecological and evolutionary consequences usually consider all members of species ecologically equivalent (Bolnick et al. 2003). However, the novelty of the modeling approach in this investigation is that every single individual in this investigation is treated as a unique agent and all of individuals are fully participated in obtained results.

It is also believed that behavior plays a central role in initiation of exploiting a new resource and then gradually becoming specialized and diverging as a new species (Futuyma and Moreno 1988). The strength of our tool is that living organisms in this investigation are genetically and behaviorally distinct from each other. Additionally, the behavioral complexity of each individual is also a contributing factor to prey/predator's energy expenditure. Not only foraging activity, but also all other fitness-determining behavior such as resting, predator avoidance behavior, socialization and reproduction were also put into consideration. These aspects make the employed modeling approach an appropriate tool for performing this investigation.

One of the central aspects in investigation of the divergence a new separate species from its ancestral population as a result of adaptive specialization through sympatric speciation is to illustrate whether the evolution of ecological specialization and assortative mating have been caused by the same traits and also if there is a synergic interaction between them or not (Caillaud and Via 2000). The number of empirical examples with identified contributing traits to both subjects is extremely limited (Caillaud and Via 2000; Schluter 1996). Empirical observations on Darwin's finches are one of these rare examples that illustrated that divergent morphology (in the form of body size and also shape and size of the beak) simultaneously makes contribution to both assortative mating and efficient resource acquisition in each environment (Ratciliffe and Grant 1983; Schluter 1996). Although such empirical studies offer a concrete evidence, it is usually difficult to monitor and investigate natural populations for generations. Therefore, it is crucial to utilize new techniques to obtain further ecological insights into this subject. Individual-based modeling approaches are capable of providing a clear and understandable insight into the correlations between adaptive behavior and emergent system properties (Grimm and Railsback 2013); as a result, IBMs could potentially offer a functional tool with the ability to track down any important features involved in ecological specialization and reproductive mating. In the current study we utilized an IBM to model a dual resource system and examine the evidence of sympatric species caused by specialized resource use. The obtained results supports previous findings in regard to the synergic interaction between adaptive specialization and assortative mating (Caillaud and Via 2000; Doebeli 1996; Kondrashov and Kondrashov 1999; Rice and Hostert 1993; Rice 1987). However, in this study we didn't incorporate any particular phenotypic traits and therefore, this would be a potential alternative for future works to look at phenotypic adaptation and adaptive radiation caused by specialized resource use. Furthermore, it will be well worthwhile to perform a more comprehensive analysis by the means of machine learning techniques in future to reveal more detailed information regarding the major contributing factors to sympatric divergence caused by specialized resource use. In this study we merely focused on runs, which had indicated evidence of the sympatric emergence. However, it will be particularly beneficial to also look at runs without any incidences of sympatric speciation to acquire a better understanding of their environmental and behavioral differences.

Despite considerable advantages, IBMs also have some major disadvantages when they are applied in the field of evolutionary biology. For instance, the applied IBM in this study does not exclusively target any specific ecosystems, aquatic or terrestrial ecosystems. Moreover, this system does not model any particular species. Living organisms in this model do not have any particular mating season. They also do not have any specific phenotypic traits; therefore, we were not able to look at the emergence of phenotypic adaptation and plasticity in this modeling study.

Regardless of these limitations, this modeling investigation enabled us to investigate thousands of generations with their complete evolutionary history. Thanks to the following characteristics, EcoSim provided a pragmatic approach to investigate the evolution of individual behavior and thereby the whole ecosystem over generations: i. every individual possesses its own genomic information, ii. this specific genetic information influences individual behavior and thereby individual’s fitness, iii. each individual inherits its genome from its parents with some possible modifications, iv. the model makes it possible to have a large number of time steps and consequently, a huge amount of generations during a reasonable period of time, v. at each time step there is a substantial number of individuals, which coexist and coevolve with each other while they are involved in a complex interaction, vi. finally, this model make it possible to monitor each individual using its ID and its parent’s ID; therefore, it facilities the investigation of speciation. These aspects make EcoSim a functional tool for this study. Consequently, it provided a valuable insight into the major role of disruptive selection pressure exerted by specialized recourse use in reproductive isolation and thereby, sympatric speciation. More importantly, the current study showed the importance of tracking lineage since the early stage of divergence.

As pointed out in the first chapter, there are limited examples of utilizing IBMs in the field of sympatric speciation. To our knowledge, there are only three investigations that have employed IBMs to study speciation including: (Labonne and Hendry 2010; Thibert-Plant and Hendry 2009; Thibert-Plante and Hendry 2011). Among them, only one single study (Thibert-Plante and Hendry 2011) has specifically focused on sympatric speciation. The current study is the second modeling study that has adapted IBMs to shed light on sympatric speciation, which is considered as a rare form of speciation occurred in nature. Therefore, this study paves the way for the future investigations to develop further insights into sympatric speciation. These are the main contributions of this modeling investigation to the literature of this research area.

3.1 References

Bolnick, Daniel I. et al. 2003. “The Ecology of Individuals: Incidence and Implications of Individual Specialization.” *The American Naturalist* 161(1):1–28.

Caillaud, Marina C. and Sara Via. 2000. “Specialized Feeding Behavior Influences Both Ecological Specialization and Assortative Mating in Sympatric Host Races of Pea Aphids.” *The American Naturalist* 156(6):606–21.

Doebeli, Michael. 1996. “A Quantitative Genetic Competition Model for Sympatric Speciation.” *Journal of evolutionary biology* 9(6):893–909.

Futuyma, Douglas J. and Gabriel Moreno. 1988. “The Evolution of Ecological Specialization.” *Annual Review of Ecology and Systematics* 207–33.

Grimm, Volker and Steven F. Railsback. 2013. *Individual-Based Modeling and Ecology*. Princeton university press.

Kondrashov, Alexey S. and Fyodor A. Kondrashov. 1999. “Interactions among Quantitative Traits in the Course of Sympatric Speciation.” *Nature* 400(6742):351–54.

Labonne, Jacques and Andrew P. Hendry. 2010. “Natural and Sexual Selection Giveth and Taketh Away Reproductive Barriers: Models of Population Divergence in Guppies.” *The American Naturalist* 176(1):26–39.

Ratciliffe, Laurene M. and Peter R. Grant. 1983. “Species Recognition in Darwin’s Finches (Geospiza, Gould) I. Discrimination by Morphological Cues.” *Animal Behaviour* 31(4):1139–53.

Rice, William R. 1987. “Speciation via Habitat Specialization: The Evolution of Reproductive Isolation as a Correlated Character.” *Evolutionary Ecology* 1(4):301–14.

Rice, William R. and Ellen E. Hostert. 1993. “Laboratory Experiments on Speciation: What Have We Learned in 40 Years?” *Evolution* 1637–53.

Schluter, Dolph. 1996. “Ecological Causes of Adaptive Radiation.” *American Naturalist* S40–64.

Thibert-Plant, Xavier and A. P. Hendry. 2009. “Five Questions on Ecological Speciation Addressed with Individual-Based Simulations.” *Journal of evolutionary biology* 22(1):109–23.

Thibert-Plante, X. and a P. Hendry. 2011. “Factors Influencing Progress toward Sympatric Speciation.” *Journal of Evolutionary Biology* 24(10):2186–96.

APPENDIX

A1. ODD Description of EcoSim, an Individual-based Predator-Prey Model without Predefined Fitness Function

In this section a detailed description of EcoSim will be provided using the updated 7-points Overview, Design concepts, and Details (ODD) standard protocol (Grimm et al. 2006) for explaining individual-based models. It should be noted that not all of the presented materials in this section have been published in (Gras et al. 2015).

A1.1 Purpose

With the purpose of investigating biological and ecological theories, this model was initiated to simulate individual behavior in an evolving dynamic ecosystem to create a complex adaptive system leading to a generic ecosystem platform, which possesses properties similar to real ecosystems (Gras et al. 2009; Mashayekhi et al. 2014). The Fuzzy Cognitive Map (FCM) (Kosko 1986) is applied in this approach to model individual behavior.

Since the FCM is coded in the genome, behavior can evolve during the simulation. Essentially, the fitness of a given set of behaviours is not set in advance. Instead, fitness emerges from interactions between simulated organisms and their abiotic and biotic environments. For instance, according to the observed results of our simulations, reproduction action was given priority over foraging action by a prey with a high fitness behavioral model when there was enough food resource available to prey individuals, whereas in the absence of food resources foraging was prioritized over reproduction.

A1.2 Entities, State Variables, and Scales Individuals

In general, simulated living organisms in EcoSim correspond to two main categories, prey and predator. Every individual in this model has a set of life-history characteristics, such as age, minimum age for breeding, speed, vision distance, level of energy, and amount of energy transmitted to the offspring. Energy is provided to individuals by consumption of available food resources in their environment. Prey feed on primary resources: grass, which are dynamic in quantity and location. Predators hunt for prey or scavenge for dead prey (in the following called ‘meat’). Each individual performs one unique action during a given time step, based on its perception of the environment. Each individual possesses its own FCM coded in its genome, and its behaviors are determined by the interaction between the FCM and the environment. FCMs are weighted graphs representing the causal relationship between sensitive, internal, and motor nodes. Consuming one unit of primary resources provides 250 units of energy to prey individuals, whereas eating one prey or one each unit of meat provides 500 units of energy to predator individuals. At any given time step, each individual performs a unique action based on its perception of the environment, which leads to the consumption of some unit of energy. Energy consumption for each individual is a matter of the type of selected action (e.g. eating, waiting, escaping) and the complexity of its behavioral model (number of existing edges in its FCM). On average, performing a movement action (such as escape or exploration) requires 50 units of energy, reproduction action requires 110 units of energy, and no action at all (basal metabolic rate) contributes to a small expenditure (18 units of energy).

Cells and virtual world: The smallest unit of the environment, the cell, represents a large space containing an unlimited number of individuals and/or some amount of food (of course, the actual number will be limited by the food). The virtual world consists of a 1000 × 1000 matrix of cells that wraps around in a torus to remove any spatial bias.

Time step, Population, and Species: Each time step involves the time needed for each individual to perceive its environment, make a decision, and perform one action. Furthermore, during each time step species memberships, including speciation events, are updated and all relevant variables are recorded (e.g., quantity of available food). The term generation refers to the total number of required time steps for an individual to reach the age of reproduction (6 for prey and 8 for predators). In general, the speed of each simulation per time step is proportional to the number of existing individuals in that simulation. There are approximately 250,000 individuals (as members of one or several species) in the world in each time step. A species is a set of individuals with a similar genome relative to a threshold. This concept will be clarified in the following section.

A1.3 Process Overview and Scheduling

Before using their behavioral model to choose a single action, each prey individual first perceives its surrounding environment (all the surrounding cells in their vision range).
Then it may choose among different possible actions, which include: i. for prey individuals: evasion (escape from predator), search for food (if not enough grass is available in the current cell, prey can move to a nearby cell to search for grass), socialization (moving to the closest prey in the vicinity), exploration (random movement), resting (to save energy), eating, and breeding; ii. for predator individuals: searching for food, hunting (catching and eating prey), scavenging (eating dead prey = ‘meat’), socialization, exploration, resting, and breeding. After prey and predator individuals perform each action, each individual’s level of energy is adjusted and its age is incremented by one unit. If the current level of energy of one individual ~~is~~ becomes less than or equal to zero, the individual will die. After all individuals complete their actions, the amount of grass and meat (dead prey) in each cell is adjusted, and the value of the state variables of individuals and cells are updated (see section A1.6 Sub-models).

A1.4 Design Concepts

A1.4.1 Basic Principles

In order to observe the evolution of individual behaviour through generations, several features were implemented in the model: i. every individual possesses genomic information; ii. this information influences individual behavior and, consequently, fitness; iii. the inheritance of genetic material allows for modification (i.e., mutation); iv. the number of individuals is sufficiently high to allow for complex interactions and spatial configurations to emerge; v. species are identified based on a measure of genomic similarity; and vi. a large number of time steps is required. These complex conditions lead to computational challenges so that models must combine the compactness and ease of computation with a high potential for complex representation.

Each individual possesses an FCM (Figure 2.1). As described in chapter two (see section 2.2.1), the FCM is the basis for the computation of an individual's behavior. Therefore, the action performed by each individual at all time steps arises from that individual's FCM. The FCM is integrally coded in the genome and, therefore, is heritable, mutable and subject to evolution. When a new offspring is created, it receives a genome that combines the genomes of its parents with some possible mutations.

A1.4.2 Emergence

In each FCM, three kinds of nodes are defined: sensitive (such as distance to enemy or food, amount of energy, etc.), Internal (fear, hunger, curiosity, satisfaction, etc.), and Motor (evasion, socialization, exploration, breeding, etc.). The activation level of a sensitive node is computed by performing a fuzzification of the information the individual perceives in the environment (changing its real scalar value into a fuzzy value, i.e., transforming the input value by a non-linear function). For an Internal or Motor node, C, the activation level is computed from the weighted sum of the current activation level of all input nodes by applying a de-fuzzification function (another non-linear function transforming the fuzzy input value into the final 'real' value). These fuzzification/de-fuzzification mechanisms allow for non-linear transformations of the perception signal, which may represent, for example, a saturation of information. Finally, the action of an individual is selected based on the activity node with the highest activation level. This concept is comprehensively clarified in chapter two (section 2.2.1). At the initiation of the simulation, prey and predator individuals are scattered randomly throughout the virtual world. As the simulation proceeds, ~~i~~ndividual's distribution pattern is gradually formed. Many factors (prey escaping from predators, individuals socializing and forming groups, individuals migrating to find sources of food, species emerging, etc.) are responsible for enormous changes in this distribution pattern over time. Various population structures and different–migration patterns (i.e. long term global movements of populations across the virtual world) may emerge since the world is large enough to allow them. More precisely, if an individual moves in one direction with its maximum speed, it can only encounter less than half of the world during its life time. In previous EcoSim studies, the use of behavioral models has resulted in a non-random distribution of individuals into populations/species that contain individuals with similar genomes (Golestani, Gras, and Cristescu 2012). It has also been proven that data produced by EcoSim indicate –multifractal properties similar to those observed in real ecosystems (Golestani and Gras 2011; Seuront, Schmitt, and Lagadeuc 1997), and Spiral waves of predator-prey interactions are an example of such an observation. Strong and robust Spiral waves created in complex and dynamic biological systems are considered a frequent phenomenon in complex ecosystems (Rohani et al. 1997). For example, self-organized Spiral patterns have been detected in snowshoe hares in real ecosystems of Northern Canada (Bascompte, Solé, and Marti 1997).

This Spiral pattern can be explained as it occurs via the system of interacting factors within EcoSim. Predation pressure combined with the pressure to search for food and potential mates plays a fundamental role in migration pattern~~s~~ formation in EcoSim. Prey near the wave-break have the capacity to escape from predators sideways. As a result, the prey subpopulation may find itself in a safe region far from predators. Consequently, a prey population can disperse greatly in a predator-free zone, thereby forming a circularly expanding subpopulation. Subsequently, these new subpopulations of prey and predators will go through the same spiral formation leading to the development of a second scale of subpopulations (Golestani and Gras 2012). Accordingly, this process will repeat over and over leading to the emergence of self-similarity (Biktashev et al. 2004) in the spatial distribution of individuals.

A1.4.3 Adaptation

Individuals carry a haploid genome of maximum length of 390 sites, where each site (gene) corresponds to an edge between two nodes of the FCM. However, to allow evolution, many edges have an initial value of zero, and only 114 edges for prey and 107 edges for predators are set at initialization. An additional site is applied to code for the amount of energy transmitted from the parent to its offspring at birth. Each gene follows the continuum-of-alleles model and can take values between -12 and +12. These alleles represent the strength of the positive or negative influence of one node on another, such as the strength of the association between a level of hunger and the tendency to feed. Each offspring acquires its genome from its parents and this genome is a combination of its parent's genome with some possible mutations. Genetic recombination through crossover was also included in this model, which allows epistasis (e.g., multiple stimuli can influence a given drive) rather than pleiotropy (each gene influences only one link between nodes). To model simple linkage, alleles are transmitted by blocks so that for each node the values of all its incident edges (in edges) are transmitted together from the same randomly chosen parent (i.e., no recombination among genes for edges to a given node). The probability of mutation is 0.005 per gene and per time step, and the effect of a given mutation is drawn from a normal distribution N (0, 0.1). In addition, a new gene (a new link between nodes) can arise or be lost at a per-generation per-gene probability of 0.001. Accordingly, new genes can emerge from the 265 initial edges of zero value.

A1.4.4 Fitness

In order to quantify the capability of each individual to survive and contribute to reproduction, fitness function was defined as the age of death of the individual plus the sum of the age of death of its direct offspring. The fitness of each species is quantified based on the average fitness of its individuals. Fitness was a post-processing computation, which means it was not considered during the simulation.

A1.4.5 Prediction

There is no learning mechanism in this model and individuals do not learn anything during their lifespan so they cannot predict the consequences of their decisions. The only information available for an individual to make a decision is coming from its perceptions at a particular time step and the values of the activation levels of the sensitive and motor concepts at that step. In fact, activation levels are never reset during an individual's life so that its current state depends on all previous states, meaning that the individual has a basic memory of its own past that will influence its future behavior.

A1.4.6 Sensing

Individuals are capable of sensing their local environment inside of its vision range. More precisely, each prey individual is able to sense its five closest predators, its five closest cells with food units and its five closest mates within its range of vision, as well as the number of primary resource units and the number of potential mates in its current cell. Additionally, each individual can also detect its current level of energy. It should be clarified that in this model the concept of perception differs from the concept of sensation; sensation is the real value coming from the environment, whereas perception is sensation modified by an individual's internal state.

A1.4.7 Interaction

Reproduction is considered the only action requiring a coordinated decision of two individuals. In order to have successful reproduction, the two mating partners need to be in the same cell, have enough energy, and choose the reproduction action. Moreover, 'enforced reproductive isolation' (where reproduction fails [without any extra cost] if the two mating partners are genetically too dissimilar) was modeled in one of our experiments. Predation is another type of interaction among individuals. A predator could perform a successful hunting action provided that it reaches the cell of its prey. At any hunting event, two units of meat are produced, one consumed by hunter (consequently, the predator's level of energy is built up by one unit of meat energy) and one added to the cell as a unit of meat. Competition for food is another type of interaction among individuals. As an example, when there is only one unit of food resource available in one cell but there are two individuals in that cell who want to eat that unit of food~~;~~, the individual that is younger will win the competition. In other words, "senescence" (where older individuals have decreased performance relative to younger individuals) is modeled in EcoSim. However, relaxing this constraint does not affect our results.

A1.4.8 Stochasticity

To create variability in our simulations, several processes were applied to generate stochasticity. For instance, at the beginning of the simulations, the number of grass units for each cell is determined through a uniform random distribution (a value between 1 and MaxGrass).

Additionally, the maximum age of an individual is determined randomly at birth from a uniform distribution centered at a value depending on the individual's type (Table A1). Stochasticity is also included in several kinds of actions of individuals, such as evasion and socialization. For instance, if there is no predator in prey's vision range, the direction of movement will be random. Moreover, the direction of the exploration action is always random. However, an investigation was defined by Golestani et al. (2010) to explore the level of randomness in EcoSim through testing the hypothesis that chaotic behavior (one signal of non-randomness) exists in time series generated by the simulation (Golestani and Gras 2010). Their results indicated that the overall behavior of the simulation creates patterns that are non-random, representing ~~a~~ complex biological systems (Kantz and Schreiber 2004).

A1.4.9 Collectives

The concept of species in this model is implemented through the genotypic cluster definition (Mallet 1995), ~~a~~ where a species is a set of individuals sharing a high level of genomic similarity. As a result, each species is then associated with the average genetic characteristics of its members (called the 'species genome' or 'species center'). Over time, a species will progressively contain individuals that are increasingly genetically dissimilar up to an arbitrary threshold, where the species splits. This speciation event is inferred from a 2-means clustering algorithm (Aspinall and Gras 2010), which determines clusters of individuals that are mutually most similar. After splitting, the two sister species remain similar enough that hybridization can occur until their genomic distance becomes at least half of the speciation threshold (in the model with enforced reproductive isolation). This information about species membership is only a label. It is not used for any purpose during the simulation (e.g. there is no species recognition) but only for post-processing analysis of the results.

A1.4.10 Observation

EcoSim produces a huge amount of data at each time step, including the number of individuals, the characteristics of each individual, and the status of each cell of the virtual world. Individual characteristics include spatial position, level of energy, choice of action, species identity, parents, FCM, etc.

A1.5 Initialization and Input Data

At initialization, the grass was randomly uniformly distributed (i.e., no divergent selection was imposed across space) and all individuals were genetically identical (with a user defined genome). Other parameter values used in this study are presented in Table A1.

Table A1. Values for user-specified parameters

User specified parameters	Used value
Initial Number of Prey	12000
Initial Number of Predators	4900
Initial Grass Quantity	5790000
Maximum Age Prey	46
Maximum Age Predator	42
Prey Maximum Speed	6
Predator Maximum Speed	11
Prey maximum Energy	650
Predator maximum Energy	1000
Distance for Prey Vision	20
Distance for Predator Vision	25
Reproduction Age for Prey	6
Reproduction Age for Predator	8

A1.6 Sub-models

A1.6.1 General

At any given time step, each individual performs one unique action based on its perception of the environment. EcoSim iterates continuously and each time step consists of the computation of the nodes’ activation levels within the FCM of each individual (the initial values of the edges of the FCM are given in Table A2 for prey and Table A4 for predators). This, in consequence, leads to the choice and application of an action for every individual. Each time step also includes the update of the world: emergence and extinction of species and growth and diffusion of grass, or decay of meat.

Table A2. Initial FCM values for prey (See Table A3). Each prey individual has an FCM representing its behaviour. At the beginning of simulations (the first time step), all prey individuals have an initial FCM. Through time, with operators like crossover and mutations, the FCMs of individuals evolve (Gras et al. 2015).

	FR	**HG**	**SP**	**CU**	**SD**	**ST**	**NU**	**ES**	**SF**	**SC**	**XP**	**WT**	**ET**	**RP**
PC	4	0	0	0.1	0	-1	1	0	0	0	0	0	0	0
PF	-4	0	0	0	0	0.5	-0.5	0	0	0	0	0	0	0
OC	0	0.5	0	-0.1	0.1	0.5	-0.5	0	0	0	0	0	0	0
OF	0	0	-0.4	0.2	-0.2	-0.7	0.7	0	0	0	0	0	0	0
FC	0	0	0.5	-0.1	0.1	0.5	-0.5	0	0	0	0	0	0	0
FF	0	0	-0.4	0.2	-0.2	-0.5	0.5	0	0	0	0	0	0	0
EL	0.4	4	-1.5	0	0	-2.2	2.2	0	0	0	0	0	0	0
EH	0	-1	1.5	0.2	-0.2	1.5	-1.5	0	0	0	0	0	0	0
OH	0	-0.2	0	-0.3	0.3	1.1	-1.1	0	0	0	0	0	2.6	0
OL	0	0.2	0	1	-1	-1.1	1.1	0	0	0	0	0	-4	0
PY	0	0	0	-0.4	0.4	0.5	-0.5	0	0	0	0	0	0	1.5

PN	0	0	0.5	0.3	-0.3	-0.8	0.8	0	0	0	0	0	0	-4
FR	0.5	0	0	0	0	0	0	1.5	-0.8	-1	0.3	-1	-1	-1
HG	0	0.3	0	0	0	0	0	-0.8	2.1	-0.7	0.7	-0.5	4	-1.8
SP	0	0	0.2	0	0	0	0	-0.2	0	1.5	0.5	-0.3	-0.4	3
CU	0	0	0	0.1	0	0	0	-0.1	0.5	0.3	1.5	-0.2	-0.3	-0.2
SD	0	0	0	0	0.1	0	0	0	-0.5	-0.3	-1.2	0.2	0.3	0.2
ST	0	0	0	0	0	0	0	-0.1	-0.8	-0.2	-2	1.5	0.8	0.7
NU	0	0	0	0	0	0	0	0.4	1	0.2	2	-1.2	-0.7	-0.7
ES	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SF	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SC	0	0	0	0	0	0	0	0	0	0	0	0	0	0
XP	0	0	0	0	0	0	0	0	0	0	0	0	0	0
WT	0	0	0	0	0	0	0	0	0	0	0	0.2	0	0
ET	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RP	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table A3. Prey/predator FCM abbreviation table. These abbreviations are used to present concepts of FCM in EcoSim, and have been used in other tables to show the values of these concepts (Gras et al. 2015).

NodeName	**Abbreviation**	**NodeName**	**Abbreviation**
Fear	FR	PredClose	PC
Hunger	HG	PredFar	PF
SearchPartner	SP	FoodClose	OC
CuriosityStrong	CU	FoodFar	OF
Sedentary	SD	FriendClose	FC
Satisfaction	ST	FriendFar	FF
Nuisance	NU	EnergyLow	EL
Escape	ES	EnergyHigh	EH
SearchFood	SF	FoodLocalHigh	OH
Socialize	SC	FoodLocalLow	OL
Exploration	XP	PartnerLocalYes	PY
Wait	WT	PartnerLocalNo	PN
Eat	ET	PreyClose	YC
Reproduce	RP	PreyFar	YF
ChaseAway	CA		
SearchPrey	SY		

Table A4. Initial FCM values for predators (See Table A3). Each predator individual has an FCM representing its behaviour. At the beginning of simulations (the first time step), all predator individuals have an initial FCM. Through time, with operators like crossover and mutation, the FCMs of individuals change (Gras et al. 2015).

	CA	**HG**	**SP**	**CU**	**SD**	**ST**	**NU**	**SY**	**SF**	**SC**	**XP**	**WT**	**ET**	**RP**
YC	0.7	0	0	-0.1	0	0.5	-0.5	0	0	0	0	0	0	0
YF	-0.5	0.7	0.1	0.4	-0.4	-0.5	0.5	0	0	0	0	0	0	0
OC	-0.5	0.7	0	-0.1	0.1	0.5	-0.5	0	0	0	0	0	0	0
OF	0.8	-0.2	0.1	0.2	-0.2	-0.6	0.6	0	0	0	0	0	0	0
FC	0	0	0.7	0	0	0.4	-0.4	0	0	0	0	0	0	0
FF	0	0	-0.5	0.3	-0.3	-0.4	0.4	0	0	0	0	0	0	0
EL	3.5	5	-1.2	0	0.2	-1.5	1.5	0	0	0	0	0	0	0
EH	-2	-3	1.4	0.3	-0.3	1	-1	0	0	0	0	0	0	0
OH	-1.5	0.3	-0.2	-0.3	0.3	1	-1	0	0	0	0	0	4	0
OL	1.7	0	0.2	1	-1	-1	1	0	0	0	0	0	-5	0
PY	-0.3	0	0	-0.4	0.4	0.8	-0.8	0	0	0	0	0	0	2

PN	0.3	0	0.5	0.3	-0.3	-0.8	0.8	0	0	0	0	0	0	-5
CA	0.2	0	0	0	0	0	0	1.5	-0.2	-0.4	0.3	-0.4	0	-0.4
HG	0	0.3	0	0	0	0	0	4	2.5	-1.2	0.3	-0.4	3.5	-0.8
SP	0	0	0.2	0	0	0	0	-0.8	-0.8	1.5	0.3	-0.5	-0.6	3
CU	0	0	0	0.1	0	0	0	0.3	0.3	0.3	1.5	-0.4	-0.3	-0.2
SD	0	0	0	0	0.1	0	0	-0.3	-0.3	-0.3	-1.5	0.4	0.3	0.2
ST	0	0	0	0	0	0	0	-0.8	-0.8	-0.2	-1.8	1	0.8	0.8
NU	0	0	0	0	0	0	0	1	0.8	0.2	2	-1	-0.6	-0.8
SY	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SF	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SC	0	0	0	0	0	0	0	0	0	0	0	0	0	0
XP	0	0	0	0	0	0	0	0	0	0	0	0	0	0
WT	0	0	0	0	0	0	0	0	0	0	0	0.2	0	0
ET	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RP	0	0	0	0	0	0	0	0	0	0	0	0	0	0

The complete FCM maps of prey individuals is given in Figure 2.2 in chapter two. It represents the FCM map for prey individuals in the presence of two different primary food resources (grass). At initialization time there is no meat in the world and the number of grass units is randomly determined for each cell. For each cell, there is a probability, probaGrass, that the initial number of grass units is strictly greater than 0. In this case, the initial number is generated uniformly between 1 and maxGrass. Each unit of food provides a fixed amount of energy to the agent that eats it. The prey can only eat grass, and the predators acquire their required energy either through hunting or scavenging. When a successful hunting action has occurred, a new meat unit is added to the corresponding cell, while another unit is consumed by the predator. A successful scavenging action performed by a predator leads to the elimination of one unit of meat from the corresponding cell. When a prey dies, the number of meat units in its cell is increased by 2. The number of grass units in a cell decreases by 1 when a prey eats, and the number of meat units decreases by 1 when a predator eats. The number of meat units in a cell also decreases at each time step by one unit due to decay, even if no meat has been eaten. For each cell of the world, if its number of grass units is greater than zero, half a unit is added per time step. Also, if a cell has zero grass units, but one of its eight adjacent cells contains grass, the same number of units is added with probability probaGrowGrass. With this mechanism, if the prey eats all the grass in one cell, the grass cannot regrow unless there still is grass in an adjacent cell. This models the problem of overexploitation of resources and the diffusion of resources through the world.

Each action has corresponding sub-model:

1. Evasion (for prey only). The evasion direction is the direction opposite to the direction of the barycenter of the 5 closest predators within the vision range of the prey, with respect to its current position. If no predator is within the vision range of the prey, the direction is chosen randomly, and the current activation level of fear is divided by 2. Then the new position is computed using the speed and direction of the prey.
2. Hunting (for Predator only). The predator selects the closest cell (including its current cell) that contains at least one prey and moves towards that cell at its current speed. If it reaches the cell, it kills one randomly chosen prey, eating one unit and having another unit of food added to the cell. If the speed of the predator is not enough to reach the prey, it still

moves at its speed toward this prey. Therefore, if the predator does not have enough energy to reach to the prey and/or its speed is not sufficient, the hunting action fails but a movement is performed in direction of the prey.

3. Search for food. The direction toward the closest food (grass or meat but not living prey) within the vision range is computed. If the speed of the agent is high enough to reach the food, the agent is placed on the cell containing this food. Otherwise, the agent moves at its speed toward this food.

4. Socialization. The direction toward the closest possible mate within the vision range is computed. If the speed of the agent is high enough to reach the mate, the agent is placed on the cell containing the mate but no reproduction action is performed, and the current activation level of sexual needs is divided by 3. Otherwise, the agent moves at its speed toward the mate. If no possible mate is within the vision range of the agent, the direction is chosen randomly.

5. Exploration. The direction is computed randomly. The agent moves at its speed in this direction. The activation level of curiosity is divided by 1.5.

6. Resting. Nothing happens.

7. Eating. If the current number of grass (or meat) units is greater than 1, then this number is decreased by 1 and the prey's (or predator's) energy level is increased by energyGrass (or energyMeat). Its activation level for hunger is then divided by 4. Otherwise, no action occurs.

8. Breeding. The process of generating a new offspring consists of the following steps. First, the conditions for successful mating are checked. Second, the value of birthEnergyPrey is transmitted to the offspring (with possible mutations) from one of the two parents, chosen randomly. Third, the edges' values are transmitted with possible mutations, and the initial energy of the offspring is computed. To model the crossover mechanism, the edges are transmitted by block from one parent to the offspring. For each node, its outgoing edges' values are transmitted together from that same randomly chosen parent. Fourth, the maximum age of the offspring is computed. Finally, the energy level of the two parents is updated.

A1.6.2 Speciation Sub-model

In order to model speciation for a given species *S*; first, the individual with the greatest Manhattan distance from the species' center (a vector containing the average of the gene values of its members) is detected. If this distance is greater than a predefined threshold for speciation, 2-means clustering is performed. Otherwise, species *S* remains unchanged. If clustering is to be performed, two new species are formed – one centered around a random individual in *S*, denoted I_r, and another centered around the individual in S that is farthest from I_r, denoted I_f. Subsequently, all remaining individuals in *S* are added to one of the two new species, whichever species the individual is more genetically similar to. After recalculating the centers for the two new species, this clustering process is repeated until convergence. After the 2-means clustering is completed, there are two new species, S_1 and S_2, whose members are subsets of the original members of *S*. The species closer to the original species *S* inherits the properties of *S*, such as the species ID and the ID of its parent species. Thus, one of the new species will continue to represent the original species, while the other one will represent a split-off of the original species.

A2. Adjusting EcoSim for Investigating Sympatric Speciation

As mentioned in chapter two, a second type of primary food resource (grass) was introduced so that prey individuals were faced with two different food options. As such, existing food chain structure (Figure A1, a) turned into (Figure A1, b).

Prey individual FCM maps in the single resource version of EcoSim had four sensitive and two motor concepts influencing by prey individual foraging behavior. Accordingly, after adding the new food resource, the prey FCM maps were modified from (Figure A2, a) to (Figure A2, b).

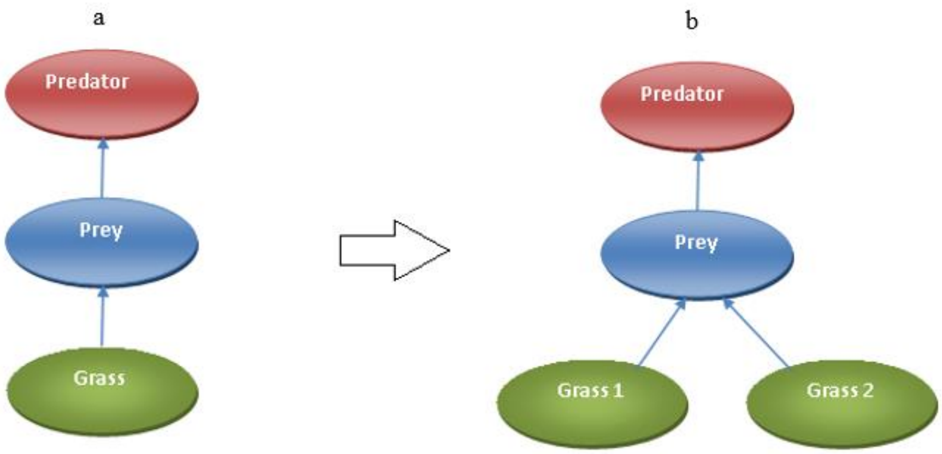


Figure A1. Regular food chain in EcoSim (a), and the new food chain in the modified dual resource EcoSim (b) (Bandehbahman 2014)

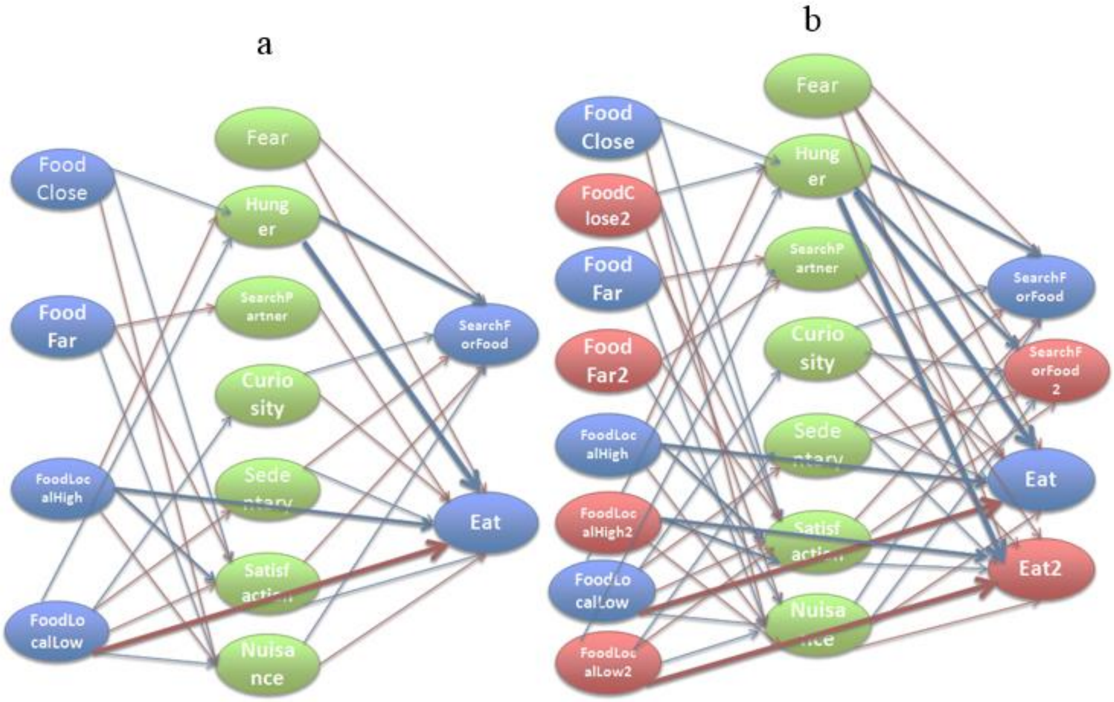
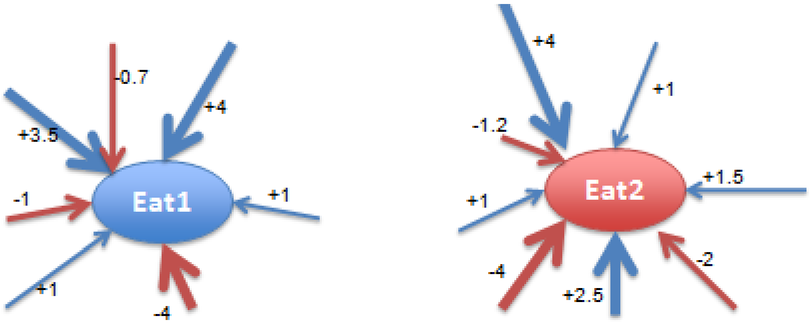


Figure A2. A part of the prey individuals' FCM associated with grass consumption by prey in the single resource version of EcoSim (a) and in the dual resource version of EcoSim after introducing a new food resource and adding the new concepts (in red) (b). Note that the width of each edge shows the influence value of that edge and the color of an edge shows inhibitory (red) or excitatory (blue) effects (Bandehbahman 2014).

A3. Species Categorizing Algorithm

A3.1 FCM-clustering Approach

The FCM behavioral model was one of the approaches employed in this study to categorize existing species. In this approach, the weighted sum of all edges with some influences on Eat1 and Eat2 Motor concepts were independently quantified. Accordingly, these values were compared with a fixed threshold to measure the extent of preferential behavior expressed by prey individuals toward different food resources. This process has been thoroughly clarified in chapter two (section 2.2.4.1). An example of this process is summarized in Figure A3.



Weighted sum of Eat1= (-4)+(+1)+(+4)+(-0.7)+(+3.5)+(-1)+(+1)= 3.8

Weighted sum of Eat2=(-4)+(+2.5)+(-2)+(+1.5)+(+1)+(+4)+(-1.2)+(+1)= 2.8

Difference= 3.8 – 2.8=1 > Threshold=0.5 ----> The species Belongs to Group1

Figure A3. The evaluation of the weighted sum of all incoming edges to Eat1 and Eat2 actions to determine species' group (Bandehbahman 2014)

Distribution of food resource preference for Food 1 and Food 2 for all populations of prey individuals from one run of the simulation is indicated in Figure A4. The horizontal axis represents time steps, while the vertical axis represents the percentage of prey belonging to each group. As it can be observed, from time step 17400, the prey population starts to be divided into three separate groups with a significant proportion of the population belonging to both Group 1 and Group 2. This observation is used to determine the time steps for which the analysis of the four criteria should be performed.

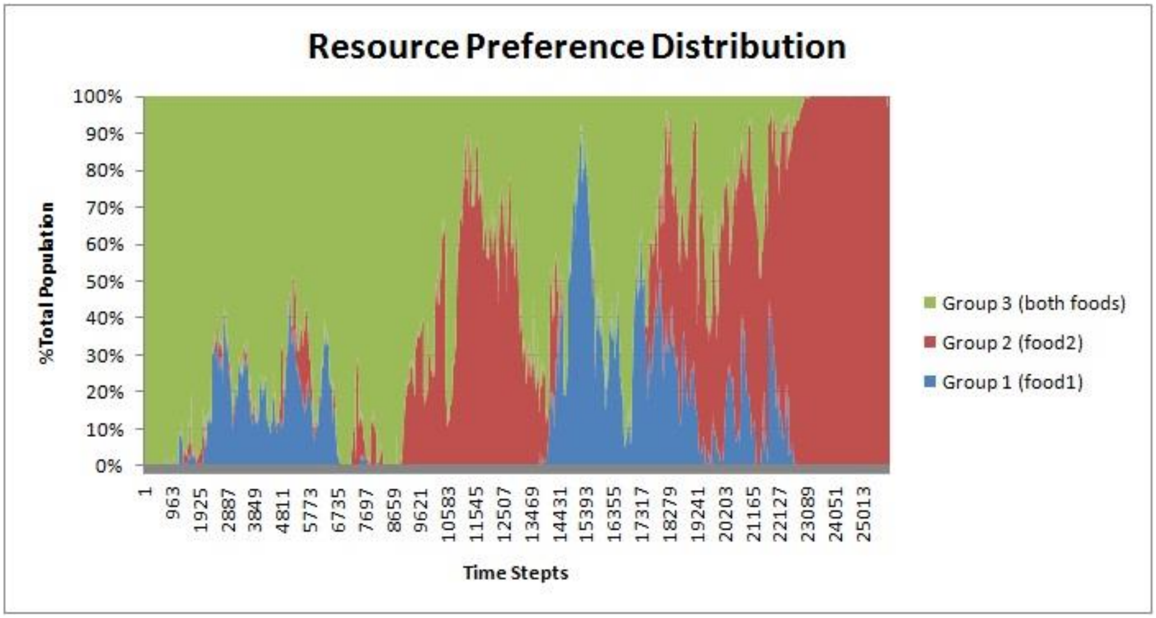


Figure A4. Food resource preference distribution for Food1 (blue), Food2 (red), and Both foods (green). Each individual preference from the total prey population is calculated for the duration of the simulation based on their FCM model (Bandehbahman 2014).

A3.2 Action-Perception Clustering Approach

Following this approach, species' actual behaviors were used for the purpose of categorizing existing species in the study. More precisely, each species were carefully considered in terms of the extent of the performed Eat1 and Eat2 actions as well as its average perception of Food1 and Food2. Five simple logical rules were employed to assess these two parameters (Table A5).

Table A5. Five rules of Action-Perception Clustering (Bandehbahman 2014)

Rules	**Rate of eating actions (Eat 1 & 2) for the individuals of each species ***	**Abundance of different food types (1 &2) ****	**Description**	**Species group**
1	If the rate of Eat 1 is significantly greater than the rate of Eat 2	While the abundance of Food 2 is higher than Food 1	Despite the high availability of Food 2, individuals of this species show a greater preference for Food 1 consumption rather than Food 2	Then this species is categorized as group 1
2	If the rate of Eat 2 is significantly greater than the rate of Eat 1	While the abundance of Food 1 is higher than Food 2	Despite the high availability of Food 1, individuals of this species show a greater preference for Food 2 consumption rather than Food 1	Then this species is categorized as group 2
3	If the rate of Eat 1 and Eat 2 are almost equal	While the abundance of Food 2 is much greater than Food 1	Although the abundance of Food 1 is significantly lower than Food 2, individuals still consume this food resource (Food 1) at the same rate of the consumption of more available food resource (Food 2). This means that this species expresses increased preference for Food 1	Then this species is categorized as group 1
4	If the rate of Eat 1 and Eat 2 are almost equal	While the abundance of Food 1 is much greater than Food 2	Although the abundance of Food 2 is significantly lower than Food 1, individuals still consume this food resource (Food 2) at the same rate as the consumption of more available food resource (Food 1). This means that this species expresses increased preference for Food 2	Then this species is categorized as group 2
5	The species that were not assigned to any group based on the four previous rules were assumed to not be specialized on any specific resource (not showing any preferential behavior)			Then this species is categorized as group 3

* In order to be able to claim that the rate of one eating action is higher than the other, a threshold was applied for the minimum required differences between the rate of Eat 1 and Eat 2. This threshold has been defined so that the rate of one eating action should be twice as high as the other one to be counted as significantly greater.

****** Likewise, another threshold was used for the differences between available resources, to find out whether their abundances are approximately equal, or if one of them is more available than the other.

Figure A5 presents the output of these species categorizing algorithms for one simulation, as an example of the resource preference distribution of all prey individuals based on their completed eating behaviors and their perception of available resources in their environment. The horizontal axis represents the time steps, while the vertical axis represents the percentage of prey belonging to each group. According to this figure, starting from around time step 21000, a significant proportion of the prey populations belong to both groups one and two. This provides an approximate time step to consider-for indicators of sympatric speciation (exploring the four required criteria on those species).

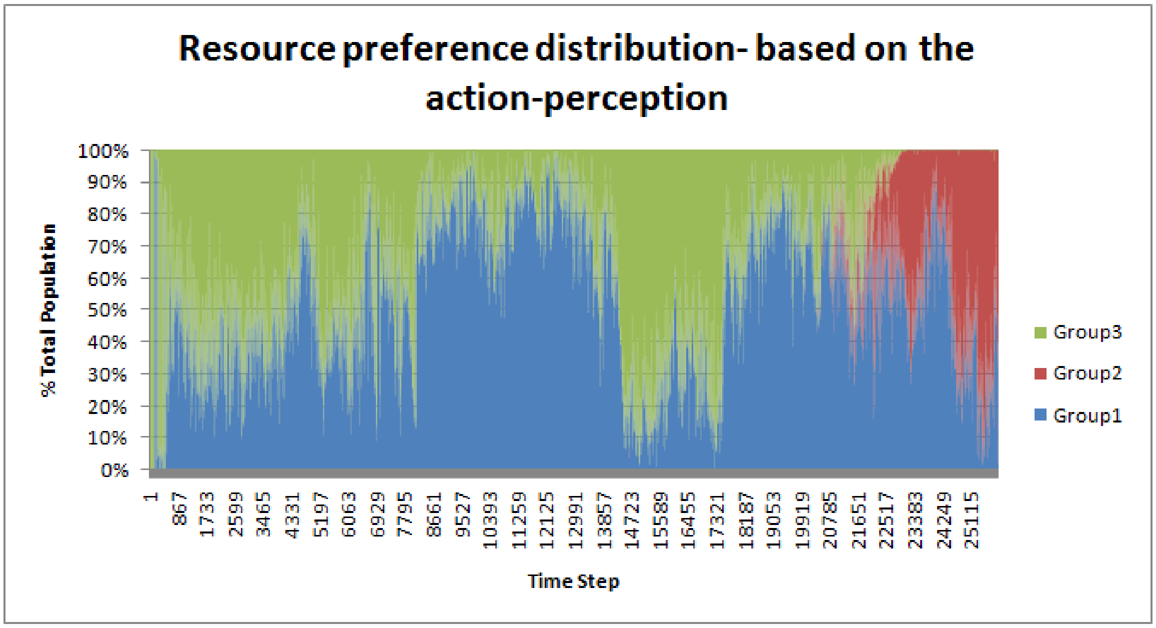


Figure A5. Resource preference distribution based on the action-perception for Food 1 (blue), Food 2 (red), and Both resources (green). Each individual's preference from the total prey population is calculated for the duration of the simulation based on their real eating behavior and their perception about the local food available (Bandehbahman 2014).

A4. Verifying Required Criteria

Overlapping geographic ranges: Figure A6 represents the geographical distance between the populations of two sister species right after the occurrence of their speciation event. The horizontal axis represents the time steps since speciation, while the vertical axis represents the distances. The distance between the two populations of sister species at the time of speciation is very small. There are at least 200 individuals from one species-living in the same cell rather than with any individuals of the other species, and the total average distance between all the individuals of the two populations is about 10 cells. This means that these sister species have been living in a common geographical area. Hence, the third required criterion is met for this set of sister species.

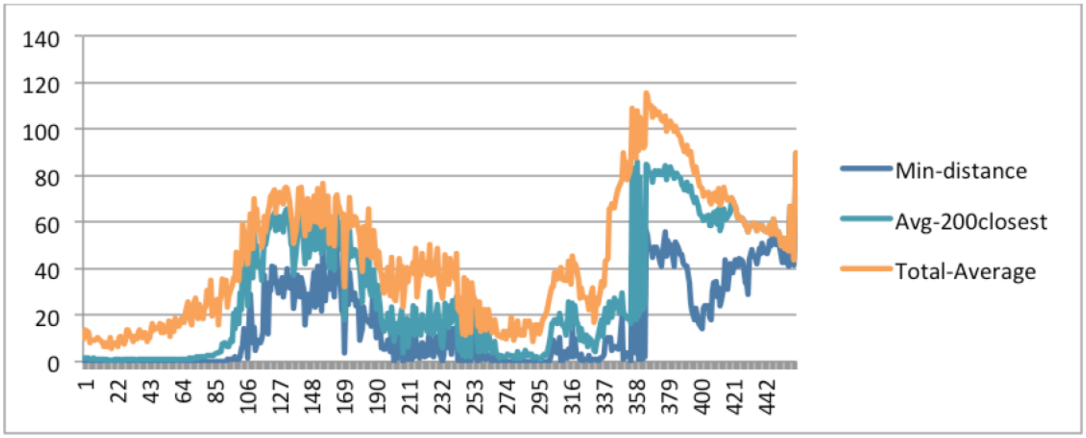


Figure A6. The minimum distance, the average distance of the 200 closest individuals, and the average distance between all the individuals corresponding to two sister species at the speciation event and through subsequent time steps (Bandehbahman 2014).

A5. Attribute Selection

In order to generate the initial dataset, all the attributes of each species were selected. These attributes covered a wide range of information about each species, from some general information (such as: population size of each species, their interbreeding ratio, and the amount of energy transferred to the offspring) to some behavioral specifications (such as: the rate of choosing different actions, and their perception of their environment). A complete list of the initial attributes applied to create the datasets is summarized in Table A6. A brief description of each attribute is also provided in this table. Starting with 81 attributes (Table A6), we applied different attribute selection methods to identify the best set. Table A7 represents the complete list of attributes after applying attribute selection methods. The attributes highlighted in red in this table were eliminated as they had a low score in the GreedyStepwise+Csf method and consequently, were not selected by the BestFirst+Csf method. As a result, 29 attributes remained.

Table A6. List of initial attributes used to create the datasets, and a short description about each attribute (Bandehbahman 2014)

ID	Attribute	Description
1	nbSpecies	Total number of currently alive
2	nbIndividual	The total prey population size
3	individualRatio	Species population size , divided by total population size
4	birthRatio	Total number of new born individuals, divided by species population size
5	interBreedingRatio	Number of interbreeding events (new born individuals with parents from different species), divided by the species population size
6	deadRatio	Number of dead individuals, divided by the total number of individuals in that species
7	deadAgeRatio	Number of dead individuals due to old age, divided by total number of deaths in the species
8	deadEnergyRatio	Number of dead individuals, due to lack of energy, divided by total number of deaths in the species
9	deadKilledRatio	Number of killed individuals, divided by total number of deaths in the species
10	deadAge	Average death age in a species
11	deadEnergy	The average energy of dead individuals in a species
12	Entropy	Diversity of alleles for all loci based on an entropy calculation
13	diversitySpatial	Dispersal level of individuals based on the average distance towards the species center
14	diversitySpatialRatio	The square roots of sum of the square of actual distances of each individual from the species center, divided by the total number of individuals
15	distEvol	Average genetic distance between the reference genome (origin) and the current genomes
16	stateOFbirth	The amount of energy transferred to the child from parent at the birth time
17	Age	The average age of individuals in the species
18	Energy	The average energy of individuals in the species
19	Speed	The average speed of individuals in the species
20	Compactness	The average number of individuals per cell
21	nbArc	Average number of arcs (genes) in the FCM of individuals
22	act_EscapeRatio	Percentage of population that chose Escape action
23	act_SearchFoodRatio	Percentage of population that chose search for food 1action and succeed
24	act_SearchFoodFailedRatio	Percentage of population that chose search for food 1action and failed
25	act_SearchFood2Ratio	Percentage of population that chose search for food 2 action and succeed

26	act_SearchFoodFailed2Ratio	Percentage of population that chose search for food 2 action and failed
27	act_SocializeRatio	Percentage of population that chose socialization action and succeed
28	act_SocializeFailedRatio	Percentage of population that chose socialization action and failed
29	act_ExplorationRatio	Percentage of population that chose exploration action
30	act_WaitRatio	Percentage of population that chose wait action
31	act_EatRatio	Percentage of population that chose eat 1 action and succeed
32	act_EatFailedRatio	Percentage of population that chose eat 1 action and failed
33	act_Eat2Ratio	Percentage of population that chose eat 2 action and succeed
34	act_EatFailed2Ratio	Percentage of population that chose eat 2 action and failed
35	act_ReproduceRatio	Percentage of population that chose reproduction action and succeed
36	act_ReproduceFailedRatio	Percentage of population that chose reproduction action and failed
37	reprodFailed_age	The average age of individuals which failed to complete the reproduction action
38	reprodFailed_energy	The average energy of individuals which failed to complete the reproduction action
39	parent1_reproductionAge	The average age of parents 1 for the reproduction action
40	parent1_reproductionEnergy	The average energy of parents 1 for the reproduction action
41	parent2_reproductionAge	The average age of parents 2 for the reproduction action
42	parent2_reproductionEnergy	The average energy of parents 2 for the reproduction action
43	DistMating	The average genetic distance between mates
44	reasonReproduceFailed_Energy	The amount of unsuccessful reproduction actions due to lack of energy, divided by the total number of unsuccessful reproduction actions
45	reasonReproduceFailed_NoPartner	The amount of unsuccessful reproduction actions due to no available partner, divided by the total number of unsuccessful reproduction actions
46	reasonReproduceFailed_PartnerEnerg	The amount of unsuccessful reproduction actions where the reason is that partner does not have enough energy, divided by the total number of unsuccessful reproduction actions
47	reasonReproduceFailed_PartnerActed	The amount of unsuccessful reproduction actions where the reason is that partner has already acted, divided by the total number of unsuccessful reproduction actions
48	reasonReproduceFailed_PartnerAction	The amount of unsuccessful reproduction actions where the reason is that partner has chosen a different action, divided by the total number of unsuccessful reproduction actions

49	reasonReproduceFailed_PartnerDist	The amount of unsuccessful reproduction actions where the reason is that partner distant is greater than distance mating threshold, divided by the total number of unsuccessful reproduction actions
50	concept_predClose	Average activation level of predator-close concept
51	concept_predFar	Average activation level of predator-far concept
52	concept_foodClose	Average activation level of Food1-close concept
53	concept_foodFar	Average activation level of Food1-far concept
54	concept_foodClose2	Average activation level of Food2-close concept
55	concept_foodFar2	Average activation level of Food2-far concept
56	concept_friendClose	Average activation level of friend-close concept
57	concept_friendFar	Average activation level of friend-far concept
58	concept_energyLow	Average activation level of energy-low concept
59	concept_energyHigh	Average activation level of energy-high concept
60	concept_foodLocalHigh	Average activation level of local food1-highconcept
61	concept_foodLocalLow	Average activation level of local food1-low concept
62	concept_foodLocalHigh2	Average activation level of local food2- high concept
63	concept_foodLocalLow2	Average activation level of local food2- low concept
64	concept_partnerLocalYes	Average activation level of partnerlocal-yes concept
65	concept_partnerLocalNo	Average activation level of partnerlocal-no concept
66	concept_fear	Average activation level of fear concept
67	concept_hunger	Average activation level of hunger concept
68	concept_searchPartner	Average activation level search for partner concept
69	concept_curiosity	Average activation level of curiosity concept
70	concept_sedentary	Average activation level of sedentary concept
71	concept_satisfaction	Average activation level of satisfaction concept
72	concept_nuisance	Average activation level of nuisance concept
73	concept_escape	Average activation level of escape concept
74	concept_searchFood	Average activation level of search for food1 concept
75	concept_searchFood2	Average activation level of search for food2 concept
76	concept_socialize	Average activation level of socialize concept
77	concept_exploration	Average activation level of exploration concept

78	concept_wait	Average activation level of wait concept
79	concept_eat	Average activation level of eat1 concept
80	concept_eat2	Average activation level of eat2 concept
81	concept_reproduce	Average activation level of reproduction concept

Table A7. List of attributes and the result after applying attribute selection methods. The attributes highlighted in red were removed at the first step (Bandehbahman 2014)

ID	Attribute	Ranker +InfoGain	bestFirst +Cfs	Greedy Stepwise +Cfs	Genetic Search +Cfs
15	distEvol	1	●	1	100%
21	nbArc	2	●	4	100%
16	stateOFbirth	3		10	80%
76	concept_socialize	4	●	3	100%
31	act_EatRatio	5	●	2	90%
38	reprodFailed_energy	6	●	5	40%
74	concept_searchFood	7	●	7	70%
80	concept_eat2	8		9	60%
69	concept_curiosity	9		12	90%
60	concept_foodLocalHigh	10	●	6	70%
61	concept_foodLocalLow	11		16	70%
33	act_Eat2Ratio	12	●	8	100%
63	concept_foodLocalLow2	13			90%
62	concept_foodLocalHigh2	14		18	10%
1	nbSpecies	15		20	0%
72	concept_nuisance	16			40%
68	concept_searchPartner	17		17	0%
40	parent1_reproductionEnergy	18			40%
71	concept_satisfaction	19			10%
59	concept_energyHigh	20			20%
58	concept_energyLow	21			0%
29	act_ExplorationRatio	22			70%
67	concept_hunger	23			80%
32	act_EatFailedRatio	24		11	50%
81	concept_reproduce	25			10%
18	Energy	26			40%
78	concept_wait	27			20%
70	concept_sedentary	28			10%
12	Entropy	29		13	50%
42	parent2_reproductionEnergy	30			40%
75	concept_searchFood2	31			0%
36	act_ReproduceFailedRatio	32			30%

19	Speed	33			20%
11	deadEnergy	34			60%
10	deadAge	35			80%
25	act_SearchFood2Ratio	36			50%
46	reasonReproduceFailed_PartnerEnerg	37			30%
22	act_EscapeRatio	38			0%
47	reasonReproduceFailed_PartnerActed	39		19	10%
2	nbIndividual	40			20%
79	concept_eat	41			10%
5	interBreedingRatio	42			30%
6	deadRatio	43			10%
50	concept_predClose	44			20%
51	concept_predFar	45			40%
77	concept_exploration	46			20%
17	Age	47			40%
13	diversitySpatial	48			20%
64	concept_partnerLocalYes	49			30%
65	concept_partnerLocalNo	50			0%
14	diversitySpatialRatio	51			80%
35	act_ReproduceRatio	52			20%
4	birthRatio	53			30%
66	concept_fear	54			40%
3	individualRatio	55			30%
20	Compactness	56			10%
27	act_SocializeRatio	57			10%
73	concept_escape	58			0%
34	act_EatFailed2Ratio	59			40%
41	parent2_reproductionAge	60			0%
23	act_SearchFoodRatio	61			0%
48	reasonReproduceFailed_PartnerAction	62			30%
28	act_SocializeFailedRatio	63			30%
30	act_WaitRatio	64			10%
7	deadAgeRatio	65		15	20%
49	reasonReproduceFailed_PartnerDist	66			10%
43	DistMating	67			20%
39	parent1_reproductionAge	68			20%
37	reprodFailed_age	69			10%
44	reasonReproduceFailed_Energy	70		14	40%

56	concept_friendClose	71			70%
57	concept_friendFar	72			0%
9	deadKilledRatio	73			50%
8	deadEnergyRatio	74			10%
45	reasonReproduceFailed_NoPartner	75			10%
54	concept_foodClose2	76			0%
55	concept_foodFar2	77			0%
53	concept_foodFar	78			0%
24	act_SearchFoodFailedRatio	79			0%
26	act_SearchFoodFailed2Ratio	80			0%
52	concept_foodClose	81			0%

A6. Classification, Specific Rules Associated to Each Run

Classification using J48 returned a decision tree for each data set, with each leaf being a rule assigned to a specific class. Figure A7 – A11 respectively represent the decision trees related to datasets from run #1, #3, #4, #5, and all the datasets combined together.

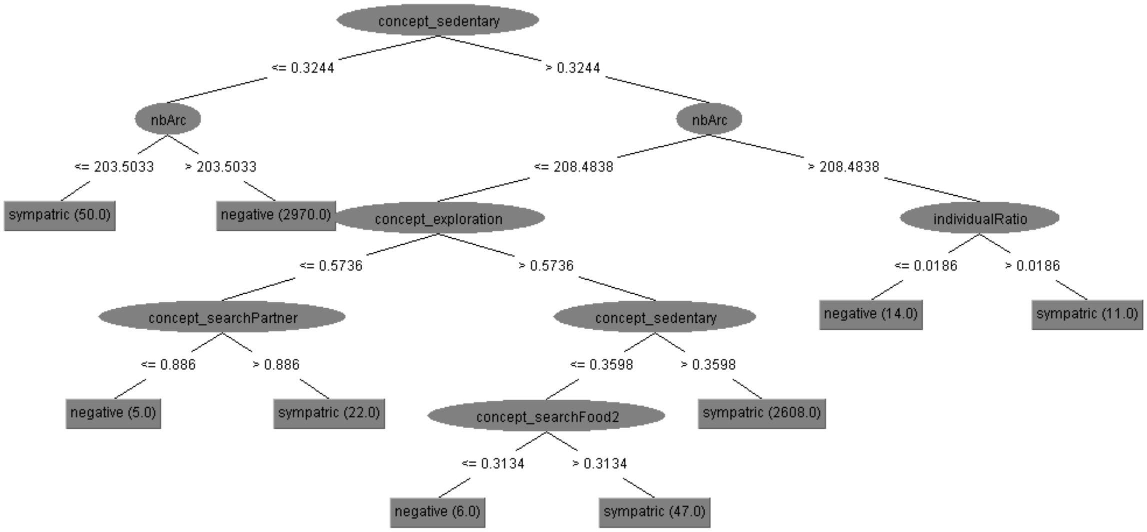


Figure A7. Decision tree corresponding to Run #1 with 9 rules (Bandehbahman 2014)

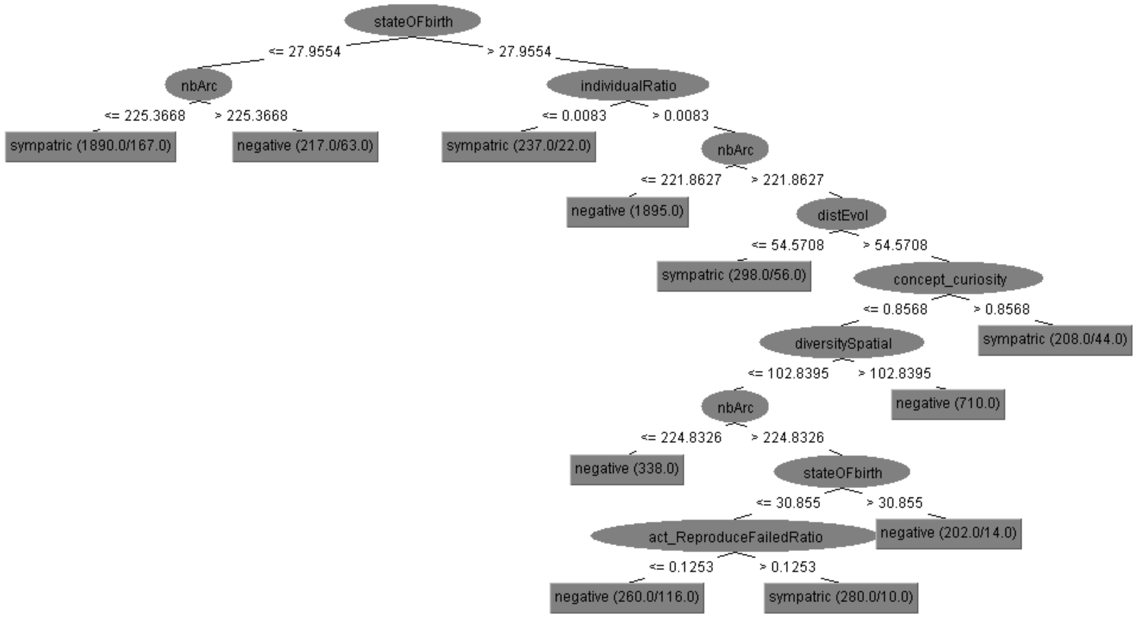


Figure A8. Decision tree corresponding to Run #3 with 11 rules (Bandehbahman 2014)

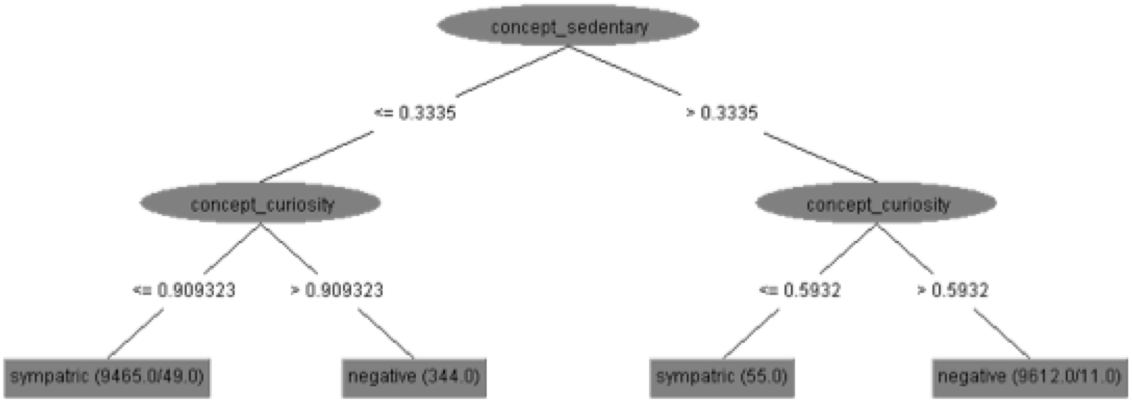


Figure A9. Decision tree corresponding to Run #4 with 4 rules (Bandehbahman 2014)

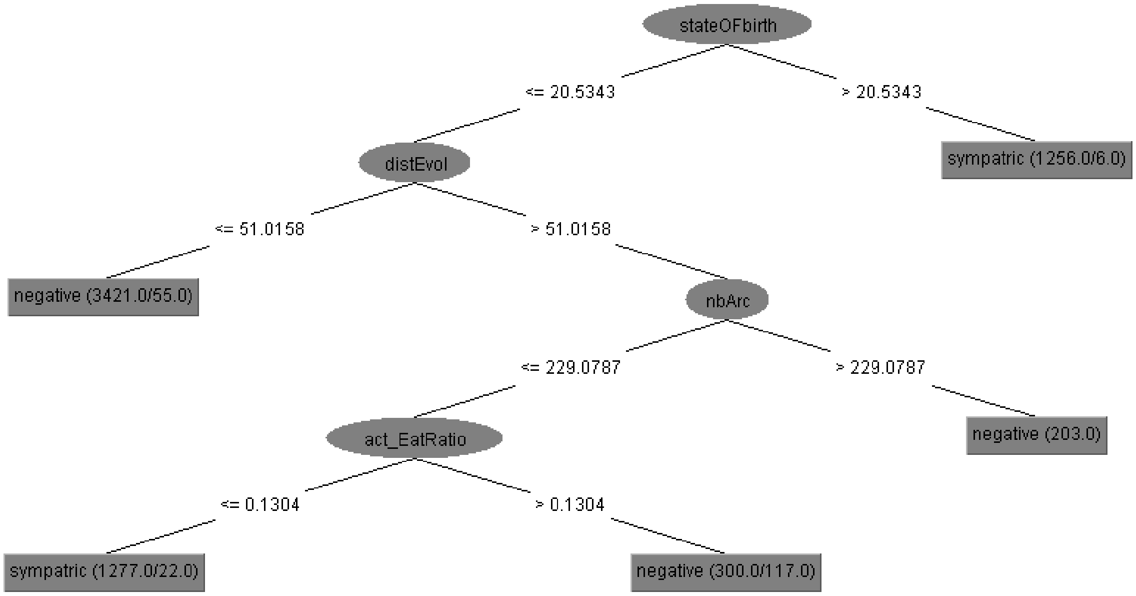


Figure A10. Decision tree corresponding to Run #5 with 5 rules (Bandehbahman 2014)

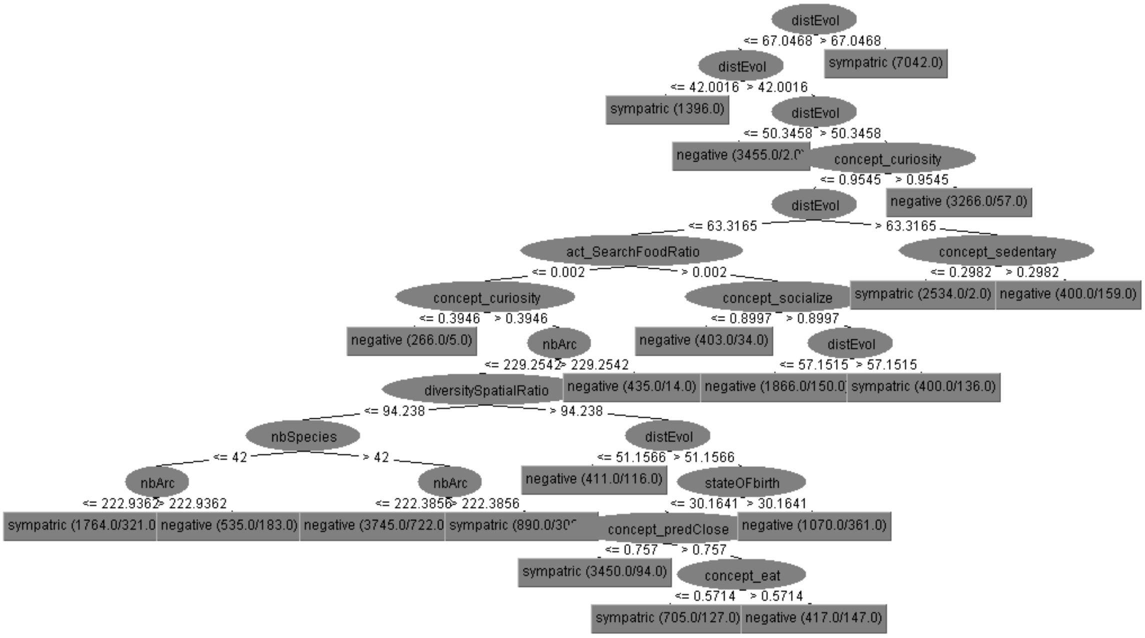


Figure A11. The decision tree returned by J48 classifier on all the datasets combined together, with 11 attributes and 20 rules (Bandehbahman 2014)

A7. References

Aspinall, Adam and Robin Gras. 2010. “K-Means Clustering as a Speciation Mechanism within an Individual-Based Evolving Predator-Prey Ecosystem Simulation.” Pp. 318–29 in *Active Media Technology*. Springer.

Bandehbahman, Sara. 2014. “Investigating The Effects of Food Chain on Sympatric Speciation Using ECOSIM.”

Bascompte, Jordi, Ricard V Solé, and Norbert Marti. 1997. “Population Cycles and Spatial Patterns in Snowshoe Hares: An Individual-Oriented Simulation.” *Journal of theoretical Biology* 187(2):213–22.

Biktashev, V. N., J. Brindley, A. V Holden, and M. A. Tsyganov. 2004. “Pursuit-Evasion Predator-Prey Waves in Two Spatial Dimensions.” *Chaos: An Interdisciplinary Journal of Nonlinear Science* 14(4):988–94.

Golestani, Abbas and Robin Gras. 2010. “Regularity Analysis of an Individual-Based Ecosystem Simulation.” *Chaos: An Interdisciplinary Journal of Nonlinear Science* 20(4):43120.

Golestani, Abbas and Robin Gras. 2011. “Multifractal Phenomena in EcoSim, a Large Scale Individual-Based Ecosystem Simulation.” Pp. 991–99 in *Int. Conf. Artificial Intelligence, Las Vegas*.

Golestani, Abbas and Robin Gras. 2012. “Identifying Origin of Self-Similarity in EcoSim, an Individual-Based Ecosystem Simulation, Using Wavelet-Based Multifractal Analysis.” Pp. 1275–82 in *Proceedings of the world congress on engineering and computer science,* vol. 2.

Golestani, Abbas, Robin Gras, and Melania Cristescu. 2012. “Speciation with Gene Flow in a Heterogeneous Virtual World: Can Physical Obstacles Accelerate Speciation?” *Proceedings of the Royal Society B: Biological Sciences* 279(1740):3055–64.

Gras, Robin, Didier Devaurs, Adrianna Wozniak, and Adam Aspinall. 2009. “An Individual-Based Evolving Predator-Prey Ecosystem Simulation Using a Fuzzy Cognitive Map as the Behavior Model.” *Artificial life* 15(4):423–63.

Gras, Robin, Abbas Golestani, Andrew P. Hendry, and Melania E. Cristescu. 2015. “Speciation without Pre-Defined Fitness Functions.” *Plos One* 10(9):e0137838. Retrieved (http://dx.plos.org/10.1371/journal.pone.0137838).

Grimm, Volker et al. 2006. “A Standard Protocol for Describing Individual-Based and Agent-Based Models.” *Ecological modelling* 198(1):115–26.

Kantz, Holger and Thomas Schreiber. 2004. *Nonlinear Time Series Analysis*. Cambridge university press.

Kosko, Bart. 1986. “Fuzzy Cognitive Maps.” *International journal of man-machine studies* 24(1):65–75.

Mallet, James. 1995. “A Species Definition for the Modern Synthesis.” *Trends in Ecology & Evolution* 10(7):294–99.

Mashayekhi, Morteza, Abbas Golestani, Yasaman Majdabadi Farahani Farahani, and Robin Gras. 2014. “An Enhanced Artificial Ecosystem: Investigating Emergence of Ecological Niches.” Pp. 693–700 in *ALIFE 14: The Fourteenth Conference on the Synthesis and Simulation of Living Systems*, vol. 14.

Rohani, Pejman, Timothy J. Lewis, Daniel Grünbaum, and Graeme D. Ruxton. 1997. “Spatial Self-Organisation in Ecology: Pretty Patterns or Robust Reality?” *Trends in Ecology & Evolution* 12(2):70–74.

Seuront, L., F. Schmitt, and Y. Lagadeuc. 1997. “Multifractal Analysis of Phytoplankton Biomass and Temperature in the Ocean.” *Oceanographic Literature Review* 6(44):652.

VITA AUCTORIS

NAME: Maryam Karim Pour

PLACE OF BIRTH: Isfahan, Iran

YEAR OF BIRTH: 1982

EDUCATION: University of Yazd, BSc Natural Resources Engineering - Environment, Yazd, Iran, 2004
Tarbiat Modares University, MSc Environmental Engineering, Tehran, Iran, 2007
University of Windsor, MSc Environmental Science, GLIER (Great Lakes Institute for Environmental Research), Windsor, Ontario, Canada, 2016