AVERTISSEMENT

LIENS

Code la Propriété Intellectuelle – Articles L. 122-4 et L. 335-1 à L. 335-10
Loi n° 92-597 du 1er juillet 1992, publiée au *Journal Officiel* du 2 juillet 1992

http://www.cfcopies.com/V2/leg/leg-droi.php
http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Université Toulouse Capitole

École doctorale : **MITT**

Présentée et soutenue par

GODIN-DUBOIS Kevin

le 15/07/2020

Spéciation guidée par l'environnement: interactions sur des périodes évolutionnaires de communautés de plantes artificielles

---

Discipline : Informatique

Spécialité : Intelligence Artificielle

Unité de recherche : Institut de Recherche en Informatique de Toulouse (UMR 5005)

Directeur de thèse : Mr Yves DUTHEN, Professeur, Université Toulouse Capitole

JURY

Rapporteurs	Mr Pierre COLLET, Professeur, ICUBE Strasbourg Mr Francisco J. VICO, Directeur de Recherche, Université de Malage, Espagne
Suffragants	Mme Nathalie AUSSENAC, Directrice de Recherche, IRIT, Toulouse Mr Stefen BORNHOFEN, Maître de Conférence, EISTI, Paris Mr Frederic GARCIA, Directeur de Recherche, INRA Occitanie Toulouse Mr Christophe PRADAL, Chargé de Recherche, CIRAD, Montpellier
Directeur(s) de thèse	Mr Yves DUTHEN, Professeur, Université Toulouse Capitole Mr Sylvain CUSSAT-BLANC, Maître de Conférence, Université Toulouse Capitole

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par : *l'Université Toulouse 1 Capitole (UT1 Capitole)*

Présentée et soutenue le *15/07/2020* **par :**

KÉVIN GODIN-DUBOIS

Environment-Driven Speciation: Long Term Interactions in Artificial Plant Communities

---

JURY

PIERRE COLLET	Rapporteur
FRANCISCO J. VICO	Rapporteur
NATHALIE AUSSENAC-GILLES	Examinateur
STEFEN BORNHOFEN	Examinateur
FREDERIC GARCIA	Examinateur
CHRISTOPHE PRADAL	Examinateur
YVES DUTHEN	Directeur de thèse
SYLVAIN CUSSAT-BLANC	Co-directeur de thèse

École doctorale et spécialité :
MITT : Domaine STIC : Intelligence Artificielle

Unité de Recherche :
Institut de Recherche en Informatique de Toulouse (UMR 5505)

Directeur(s) de Thèse :
Yves DUTHEN et *Sylvain CUSSAT-BLANC*

Contents

List of Figures

List of Tables

Abstract

Artificial Life researchers have, for decades, created a plethora of creatures using numerous encoding schemes, motile capabilities and cognitive capacities. One recurring pattern, however, is that focus is solely put on the evolved individuals, with very limited environmental variations. In this work, we argue that more complex abiotic[1] constraints could drive an evolutionary process towards more robust and diverse regions of the genetic space.

We started with a complex morphogenetic model, based on K. Sims' directed graphs, which relied on the Bullet physics engine for accuracy and used 6 Degrees of Freedom constraints to connect pairs of organs. We evolved a panel of natural-looking plants which had to cope with varying resource levels thanks to a mobile light source and seasonal rain patterns. In addition to this experiment, we also obtained improved vertical growth by adding an artificial biotic[2] constraint in the form of static grass blades. However, the computational cost of this model precluded scaling to a population-level evolution and was reduced in the successive experiment, notably by removing the physical engine. This led to the exploration of co-evolution on single-species populations which, thanks to our Bail-Out Crossover (BOC) algorithm, were able to self-reproduce. The resulting populations provided valuable insight into the mechanisms of self-sustainability.

These were put to action in an even more straightforward morphogenetic model inspired by the work of Bornhofen. Due to its lightweightness, this allowed for both larger populations (up to thousands of individuals) and longer evolutionary periods (100 years, roughly 5K generations). Our first experiment on this model tested whether text-book cases of speciation could be reproduced in our framework. Such positive results were observed thanks to the species-monitoring capacities of APOGeT, a novel clustering procedure we designed for online extraction of species from a genealogic tree. This drove us to a final experiment in which the environment was controlled through Cartesian Genetic Programming thus allowing the automated evolution of both the population and abiotic constraints it is subjected to. Through a variation of the traditional $1 + \lambda$ algorithm, we obtained 10 populations (evolved group e) which had

[1]The abiotic component of an ecosystem are non-living parts of the environment that has an impact on living organisms

[2]The biotic component of an ecosystem are the living entities of which it is composed and which have an impact on the environment and other living entities

endured in harsh and unpredictable environments. These were confronted to a control group c, in which the constraints were kept mild and constant, on two types of colonization evaluation. Results showed that the evolved group was heterogeneous with half of e consistently outperforming members of c and the other half exhibiting worse performances than the baseline. We also found a very strong positive correlation between catastrophic drops in population level during evolution with the robustness of their final representatives.

From this work, two conclusions can be drawn. First, though the need to fight on both the abiotic and biotic fronts can lead to worse performances, more robust individuals can be found in reasonable time-frames. Second, the automated co-evolution of populations and their environments is essential in exploring counter-intuitive, yet fundamental, dynamics both in biological and artificial life.

Résumé

Depuis des décades, les chercheurs en Vie Artificielle on créé une pléthore de créatures en utilisant de multiples schémas d'encodage, capacités motrices et aptitudes cognitives. Un motif récurrent, cependant, est que la focalisation est centrée sur les individus à évoluer, ne laissant que peu de place aux variations environnementales. Dans ce travail, nous argumentons que des contraintes abiotiques[3] plus complexes pourraient diriger un processus évolutionnaire vers des régions de l'espace génétique plus robustes and diverses.

Nous avons conçu un modèle morphologique complexe, basé sur les graphes orientés de K. Sims, qui repose sur le moteur physique Bullet pour la précision et utilise des contraintes à 6 Degrés de Liberté pour connecter les paires d'organes. Nous avons ainsi évolué un panel de plantes à l'aspect naturel qui devaient survivre malgré des niveaux de ressources variables induits par une source de lumière mobile et des motifs de pluies saisonnières. En plus de cette expérience, nous avons aussi obtenu une meilleure croissance verticale en ajoutant une contrainte biotique[4] artificielle sous la forme de brins d'herbe statiques.

La complexité de ce modèle, cependant, ne permettait pas la mise a l'échelle d'une évolution de populations et a donc été réduit dans l'expérience suivante, notamment en supprimant le moteur physique. Cela nous a amené à l'exploration de la co-évolution de populations composées d'une unique espèce et ayant la capacité de se reproduire de manière autonome grâce à notre Bail-Out Crossover (Croisement avec Désistement). Bien que les populations résultantes n'ont pas démontré un grand intérêt pour cette aptitude, elles ont néanmoins fourni d'importantes informations sur les mécanismes d'auto-reproduction.

Ceux-ci ont été mis en action dans un second modèle inspiré des travaux de Bornhofen. Grâce à sa légèreté, cela nous a permis de traiter non seulement de plus grandes populations (de l'ordre de milliers d'individus) mais aussi de plus longues périodes évolutionnaires (100 années, approximativement 5000 générations). Notre première expérience avec ce modèle s'est concentrée sur la possibilité de reproduire des cas d'école de spéciation (allopatrique, parapatrique, péripatrique) sur cette plate-forme. Grâce à APOGet, une nouvelle procédure de regroupement pour l'extraction en parallèle d'espèces à partir d'un arbre généalogique, nous avons pu affirmer que le système était effectivement capable de spéciation spontanée.

[3] C'est à dire extérieures aux créatures

[4] Imposée par d'autres formes de "vie"

Cela nous a conduit à une dernière expérience dans laquelle l'environnement était contrôlé par de la Programmation Génétique Cartésienne (CGP), permettant ainsi une évolution automatique d'une population et des contraintes abiotiques auxquelles elle était confrontée. Par une variation du traditionnel algorithme $1 + \lambda$ nous avons obtenu 10 populations finales qui ont survécu à de brutales et imprévisibles variations environnementales. En les comparant à un groupe contrôle c pour lequel les contraintes ont été maintenues faibles et constantes, le groupe évolué e a montré des performances mitigées: dans les deux types de tests, une moitié de e surpassait c qui, à son tour, surpassait la moitié restante de e.

Nous avons aussi trouvé une très forte corrélation entre les chutes catastrophiques de population et la performance des évolutions correspondantes. Il en résulte que l'évolution de population dans des environnements hostiles et dynamiques n'est pas une panacée bien que ces expériences en démontrent le potentiel et souligne le besoin d'études ultérieures plus approfondies.

Chapter 1

Introduction

Artificial lifeforms are a human concern dating back to the antiquity but which has only soared in recent decades. The intent behind the development of such creatures differ from one researcher to the next, ranging from swarms of task-specific mini-robots to autonomous anthropomorphic entities. More often than not, preliminary iterations are performed *in-silico* to diminish the material and temporal costs of blatantly non-functional prototypes. But not all research in ALife is concerned with physical instantiation and some work are undergone solely in a virtual world (Soft ALife, Hard ALife, Wet ALife ...).

There again, differing avenues can be discerned with some simulations being designed to explore mechanisms that could scarcely be experimented upon in the biological world such as the cell-level behavior of cancer or the genetic dynamics of populations over millions of years. Other approaches attempt to generate meaningful constructs with a bottom-up paradigm in which the complexity at the concerned layer (e.g. behavior) is obtained through the non-additive interaction of its low-level components (e.g. artificial neurons).

In this work we will concern ourselves with the latter kind of experiments, exploring the generation of life-as-could-be artificial plants and studying the evolutionary trends of autonomous populations of such plants. This aims at providing not only a corpus of knowledge on the morphological and behavioral aspects of such individuals but also at producing a framework for the study of broader phenomenon on the geological scale.

1.1 Current issue

More specifically, we argue, as will be detailed in the next chapter, that whenever artificial creatures are generated they are often embedded in relatively plain environments. For instance, the seminal work of [Sims 1994a] only involved a flat environment, though the presence of a competing organism introduced biotic complexity. Digital ecosystems, thanks to their computational lightweightness, are foremost in the literature as the type of system in which environmental variations are experimented upon. Multiple articles (e.g. Canino-Koning et al.

2016; Luo et al. 2019; Nahum et al. 2017) have shown that dynamical conditions have a positive effect on the populations' robustness, though in these cases the variations were limited to different sets of experimenter-designed fitness landscapes.

Other works have performed similar experiments in more complex settings such as the niche-building creatures of [Chiba et al. 2017] or with 3D plants [Bornhofen et al. 2011; Eloy et al. 2017]. In these cases, however, the environmental dynamics were still under indirect supervision of the experimenter which defined the bounds and directions of such variations. By building upon these work, we aim, in this thesis, at producing a more flexible method for the use of non-homogeneous, non-static environments.

Indeed, in the same manner that the manual design of robots has been partially replaced by a bottom-up approach through automated evolution, we argue that abiotic constraints could be managed by an autonomous "agent" which could not only exhibit plausible strategies in manipulating the various levers at its disposal but could also discover novel and counter-intuitive types of dynamics.

1.2 Contribution

With this in mind, we explored a minimalist set of dynamics (sun, water) coupled with a complex morphological controller. From there artificial plants with drought-resistant capabilities were obtained [Dubois et al. 2017] some exhibiting natural-looking morphologies while other ventured into more original regions of the genetic space. While capable of producing elaborate body plans with a life-like demeanor (thanks in large parts to their continuous growth), we selectively trimmed portions of this model in order to move onto larger populations of simultaneously cohabiting individuals. Embedded with genetically-controlled parameters managing their reproduction patterns these creatures have been used to determine important characteristics of self-sustaining ecosystems [Godin-Dubois et al. 2019b].

A further simplification was performed when moving from populations evolved through a classical evolutionary algorithm to autonomous populations which had to find ways to survive without external supervision or help. In order to validate the framework we were building, these plants were subjected to hand-crafted equations simulating text-book case of conditions for the emergence of speciation. These were found to behave appropriately [Godin-Dubois et al. 2019c], with the individuals managing to survive in partially hostile conditions. More interestingly, the dynamics observed at the species[1] level were found to have remarkable levels of complexity with, depending on the run considered, alternating periods of cohabitation, competition, migration.

[1] "group of potentially interbreeding natural population reproductively isolated from other such groups" [Singh 2012]

These species were extracted, in parallel of the simulation, through a clustering tool named APOGeT [Godin-Dubois et al. 2019a] which transforms a stream of genomes into a user-controlled schematic view of the phylogenetic hierarchy. Through the use of representatives as the descriptors of a species' internal variances and key features, this tool was shown to accurately describe the temporal dynamics of a population at the species level.

Both of these milestones led to the actual design of the EDEnS framework, in which environments are controlled by evolvable structures, Cartesian Genetic Programming in this case, while the artificial creatures are left, unsupervised, to fend for themselves. The framework drives the population's evolution by attempting multiple variations of a single controller in order to find the one which most satisfies a given set of fitnesses. Thus, artificial selection is effected at the level of the ecosystem, instead of the individual, leading to the exploration of more diverse regions of the genetic space than would be obtained by a creature-centered approach. The resulting plants were shown to outclass, within a reasonable margin, competitors from a control group obtained in an excessively hospitable environment. This work also shed light on some of the desirable external dynamics one should have in order to promote robust behavior in a population.

1.3 Structure

In the following chapter 2 "A slice of Artificial Life" we will broach the subject of Artificial Life, focusing on a subset of its corpus: virtual creatures both in isolation and combined into large-scale ecosystems. Through this overview, we will describe some of the most salient work on the matter starting with the generation of body plan and, in the case of animals, the associated behavioral controller. We will then explore more complex systems in which individuals cohabit simultaneously in the same environment, some with collaborative goals other with predatory intents. This will be concluded by a synthesis on the position of environmental dynamics in the current literature and a formulation of how (and why) such dynamics should be thoroughly investigated.

This is a followed, in chapter 3 "Isolated evolutions", by a the use of a 3D physics engine (Bullet) to generate an environment with a dynamic light source (sun) and randomly-pattern rain falls. Both conditions act as a source of stress for the plants as they must devise ways to cope with fluctuating water availability, periods of obscurity (night) and low light level (winter). Furthermore, each individual is composed of a (potentially large) number of organs which accomplish a specific metabolic function (water uptake, photosythesis, storage) and are interconnected by breakable fixed 6DoF constraints. This imposes a further difficulty to the plants which must not only plan for a functional morphology but must also ensure that intermediate steps are structurally sound. Two types of morphological controllers are presented with the advantages of both being highlighted.

In the subsequent chapter 4 “Limited Co-evolution”, this model is transformed into the basis for a population-wise evaluation of the genomes. By removing the computationally costly constraints between organs, simulation times were improved so as to allow for the simultaneous management of up to 100 plants. These were fitted with a genome-regulated method of autonomous reproduction which was used to derive necessary conditions for the emergence of self-sustainability in unsupervised plant ecosystems.

Based on this reproduction scheme a tool for phylogenetic monitoring was devised and is detailed in chapter 5 “Phylogenetic monitoring”. Through the use of representatives, APOGeT is able to parse, from a stream of genomes, a data-rich structure describing the species interaction in terms of anagenesis[2] and cladogenesis[3] while allowing for the extraction of each species’ internal variations. Key parameters of this tool are described, before moving on to chapter 6 “Speciation Test-bed” which leverages the phylogenetic monitoring capabilities of APOGeT to assert that in text-book cases of environmental dynamics the appropriate speciation patterns occur.

Given the larger scale of this experiment, a lighter model was used which relied on 2D L-Systems to achieve small enough simulation times so as to permit thousands of individuals to coexist for a hundred years. Detailed observation of the phylogenetic trees thus obtained showed that the expected speciation patterns did occur, in addition to obtaining strikingly complex inter-species dynamics.

Chapter 7 “Timelines Exploration” is concerned by the deployment of the complete EDEnS framework, in which evolution is indirectly controlled via parallel evaluation of alternative environmental controllers, implemented as Cartesian Genetic Programming. This experiment aims at determining whether individuals produced in such varying environments do outperform competitors living in more lenient conditions. After detailing the specifics of the algorithm, we explore and compare both types of populations on a number of different dimensions: robustness, morphology, strategies. Analysis of the relationship between such robustness and the dynamics of the environmental controllers throughout the 1K simulated years reveals counter-intuitive yet promising results.

A summary of the contributions realised by this work and its desirable extensions are used to conclude this manuscript in chapter 8 “Conclusion”.

[2]Gradual evolution of a species that continues to exist as an interbreeding population

[3]Evolutionary splitting of a parent species into two distinct (reproductively isolated) species

Chapter 2

A slice of Artificial Life

Abstract Artificial life is a very active research field subdivided in myriads of specific topics. This work being concerned with the impact of environmental dynamics upon artificial plants morphologies and behavior, we explore a particular subset of this field: morphogenetic engineering and autonomous ecosystems. The former is concerned with the various models developed to encode for both plants and animals from Lindenmayer's Systems to Sims' directed graphs. Lower-level representations are also broached including Genetic Regulatory Networks (GRN) or Composite Pattern-Producing Networks (CPPN). The latter encompasses all simulations in which population is maintained, preferentially in an autonomous manner, through a form a natural selection in the sense of differential allelic survival through phenotypic expression. This includes digital ecosystems, in which individuals are computer programs, and virtual environments ranging from unicellular organisms to forest and "arenas". Synthesis of this review shows that environments are seldom given a major role.

Résumé La Vie Artificielle est un champ très actif subdivisé en une myriade de sous-genres. Ce travail étant focalisé sur l'impact des dynamiques environnementales sur la morphologie et le comportement de plantes artificielles, nous allons explorer des pôles spécifiques à ce domaine : l'ingénierie morphogénétique et les écosystèmes autonomes. Le premier se concentre sur les divers modèles développés pour encoder aussi bien des animaux que des plantes: des Systèmes de Lindenmayer jusqu'aux Graphes Orientés de Sims. Des représentations de plus bas niveaux telles que les Réseaux de Régulation Génétiques (GRN) ou les Réseaux de Production de Motifs Composés (CPPN) seront aussi abordées. Le second englobe toutes les simulations dans lesquelles la population survit face à une forme de sélection naturelle dans le sens du maintien allélique différentiel au travers de l'expression phénotypique. Cela inclut les écosystèmes digitaux, dans lesquels les individus sont des programmes informatiques, mais aussi les environnements virtuels: des organismes unicellulaires, en passant par les forêts ou les "arènes". La synthèse de cette revue met en lumière que l'environnement ne joue que rarement un rôle majeur.

2.1 Foreword

Though the notion of artificial life-forms is one we can trace back to the antiquity, the first formalisation of the Artificial Life (or A-Life) field of research was made in [Langton 1986]. More of a statement of purpose than a definition, Langton phrases its intent thus.

> *"By synthesizing 'life-like' behaviors in the study of artificial life, we want to try to distinguish between the relevant and irrelevant details of life's biochemical implementation in order to uncover the 'molecular logic' of life. The ultimate goal of the study of artificial life would be to create 'life' in some other medium, ideally a virtual medium where the essence of life has been abstracted from the details of its implementation in any particular hardware. We would like to build models that are so life-like that they cease to be models of life and become examples of life themselves."*

This leaves large room for interpretation as to what exactly falls within the domain of A-Life itself. While the term "virtual" could be taken at face value to solely mean production residing inside computers, one cannot deny that other human-constructs such a robots are also artificial. Further confusion arises when considering how to label such productions with respect to their "liveliness". While the protagonist in the 2001 movie "A.I. Artificial Life" looks and sounds alive, they are definitely not so in the sense we are used to. Indeed the definition of Life is by no means as rigid as we would like, as biological examples abound with no intrinsic reproduction system or cellular atomicity.

The matter of defining Life as a generic principle, independently of the specific instances we can find in our single earthly sample, is made that much harder when taking in consideration A-Life productions. As emphasized in *The MIT Encyclopedia of the Cognitive Sciences*:

> *"The claim that even virtual creatures in cyberspace could be genuinely alive is called strong A-Life, in analogy to strong AI. Most A-Lifers reject it (but see Langton et al. 1989 and Ray 1993). Or rather, most reject the view that such creatures can be alive in just the same sense that biological organisms are, but allow that they are, or could be, alive to a lesser degree."*

Given the nature of the work presented in this manuscript, both in this literature review and in the original research, it is enough for the reader to keep these questions in mind when asserting the contribution of a given piece of research.

Now leaving these philosophical considerations, we reassert once more the broadness of the field of Artificial Life. In our opinion, four major dimensions can be used to categorize its productions:

Medium onto which the life-form is implemented whether it is the physical world through robotics or molecular chemistry or a virtual one.

Morphological degrees of freedom, which could roughly be translated into the complexity of its possible shapes.

Behavioral capabilities from purely reactive individuals to fore-sighted, planning ones.

Scale at which the produced system works: from single individuals to complete self-sustaining ecosystems.

As our work, in this manuscript, is focused on plant growth throughout evolutionary periods, the subset of posterior research presented here will be those using a virtual medium with either non-trivial Morphological or Scale components as illustrated in figure 2.1. Additionally, only systems in which a form of selective, cumulative improvement is used are considered which accounts for Genetic Algorithms (GA), Evolutionary Algorithms (EA) but also the Natural Selection process found in self-sustaining ecosystems.

The remainder of this chapter will thus observe how solitary individuals have been produced through increasingly autonomous process (section 2.2) before diving into the complexities of virtually-instantiated ecosystems populated by digital or simulated individuals (section 2.3). This will be concluded by highlighting the specific contributions made in by this corpus with regards to morphogenesis, species dynamics and the effect of the environment on large-scale evolution.

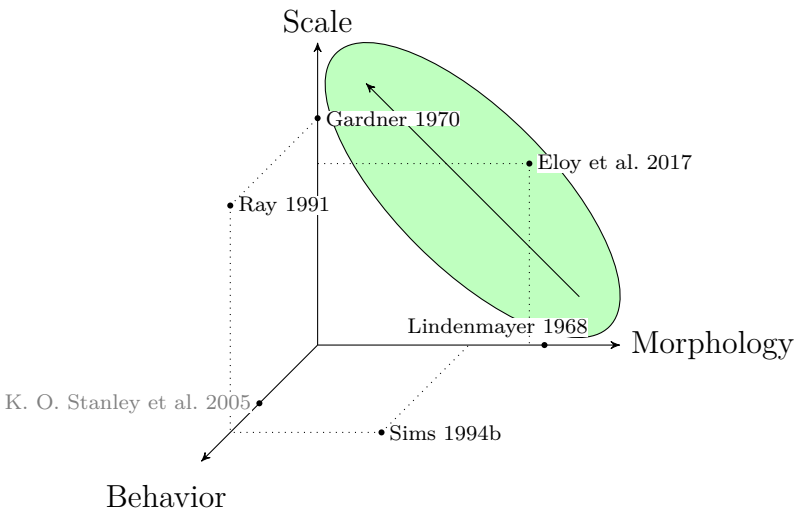
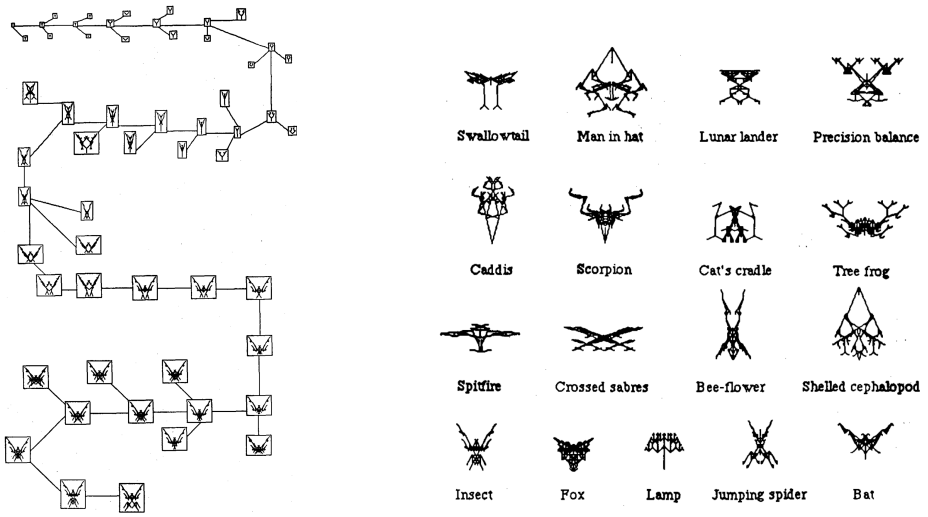


Figure 2.1: Schematic view of the field of (virtual) Artificial Life and our position in this work (in green).

2.2 Creatures



(a) Cumulative selection (b) Diversity panel

Figure 2.2: Dawkins' Biomorphs [Dawkins 1986]. Emergence of complexity even when an evolutionary process is guided by an "untrained" hand

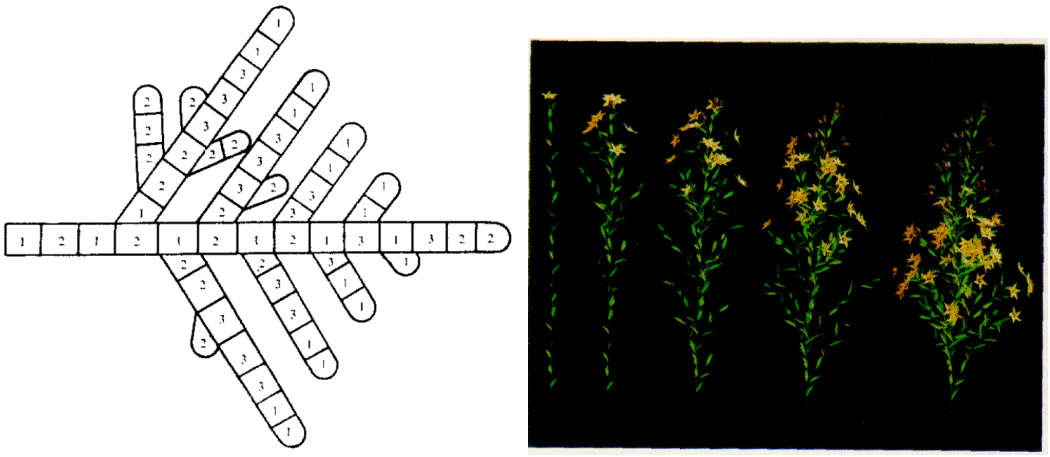
Designed as a pedagogic tool for explaining the power of cumulative selection in producing complex artifacts, Dawkins' Biomorphs [Dawkins 1986] were one the first instances of virtual artificial creatures.

Their "growth" process is akin to a recursive branching with a small number of integral-valued genes controlling various aspect of the final shape (recursion, width, height ...). The elegance in this simplicity, as shown in figure 2.2, was that the resulting genetic space was complex enough to generate a large panel of morphologies. The evolutionary process was named a "blind watchmaker" for, in this instance, a user performs the actual selection on the competing phenotypes. At each generation the selected shape is duplicated, with mutation, to produce a new population of biomorphs that the user can filter out by picking its favourite one. In this manner, only 29 such selection steps are enough for the single dot at the top of figure 2.2a to "evolve" into the insectoid shape at the bottom.

Such a show of force was inspirational in demonstrating the power of evolution, not only in its familiar, biological substrate but in the digital world as well. This section will thus be devoted to sampling the previous researches in artificial creatures production by clustering them based on the morphological controller used to map genotype from phenotype.

2.2.1 Lindenmayer Systems

First introduced in 1968 by Lindenmayer [Lindenmayer 1968], these rewriting rule systems have shown great capabilities in producing plausible plant mor-



(a) [Lindenmayer 1968] (b) [Prusinkiewicz et al. 1988]

Figure 2.3: Early L-Systems' generative capacities

phologies. Though in the initial publication, the focus was on low-level growth and differentiation of cells in a somewhat continuous morphogenesis process (see figure 2.3a) they have quickly been used to generate more complex shapes (fig. 2.3b).

In essence an L-System is a tuple $L =< T, N, X, R >$, where T and N are the sets of terminal and non-terminal symbols, respectively. $X \in N^d$ is the axiom (where d is commonly set to 1) and R the set of rewriting rules. These have the form $A \to S$, denoting that $A \in N$, the premise, is to be replaced by $S \in (T \cup N)^*$. Thus starting from the axiomatic string X and by performing multiple iterative steps of simultaneous replacement, the resulting string can greatly increase in size and complexity. The different symbols in T and N are interpreted as drawing instructions, thus translating a linear string of symbols into a 2 or 3 dimensional structure.

Most commonly this is done through a turtle-like robot which parses one symbol after another, leaving under most circumstance a "trace" of its displacements. Thus in order to produce non-linear patterns the set of terminal symbols generally contains rotational operators (one set in 2D and two in 3D) and branching marks, allowing a diverging subpath to be created. Numerous variants of varying complexity have been developed onto top of this basic functionality to encompass a broader array of possibilities: parametric L-Systems manipulate non-terminal symbols as function-like object to which parameters can be provided and computed upon, context-sensitive L-Systems have premises of more than one symbol and stochastic L-System allows for multiple use of the same premise with varying levels of occurrence. Due to this expressive power they have been extensively used to model plant morphologies e.g. for the population of virtual worlds as illustrated by the work of [Corchado et al. 2009] which selected plants for their capacity to grow towards and collect light.

Additionally, these L-Systems are sufficiently robust to replicate the mor-

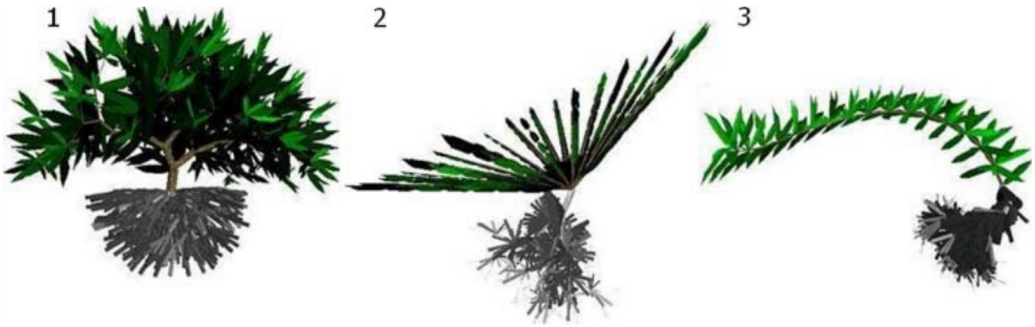
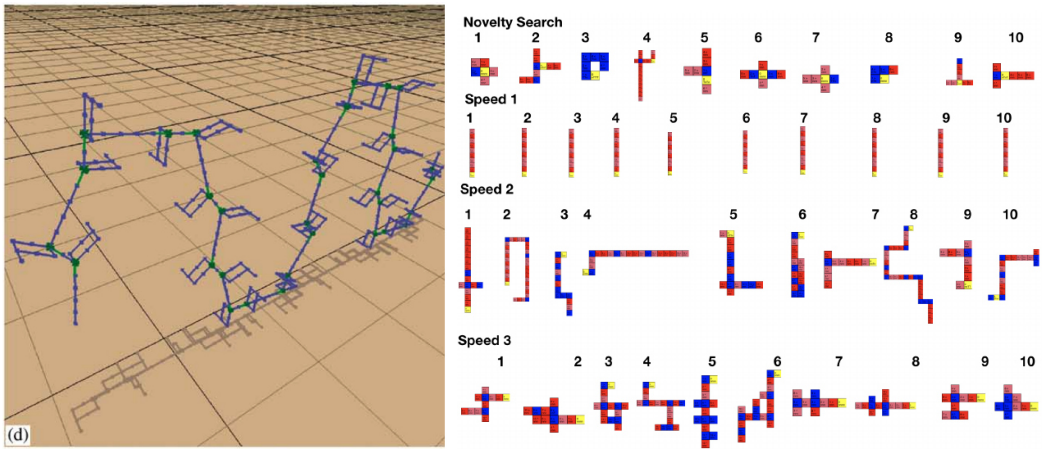


Figure 2.4: Virtual plants with natural strategies [Bornhofen et al. 2009]

phological aspects of natural-plants, as was investigated by [Bornhofen et al. 2009], when complemented by appropriate metabolic equations. As shown in figure 2.4, the obtained individuals bear striking resemblances, in terms of survival strategies, to their biological counterparts. The leftmost "bush" can be found in all sizes in nature (e.g. Beech tree) while in the center the radial disposition is a sensible solution for maximizing collective illumination, as strategy found in flowers and some types of fern. The elongated shape of the right-most individual is more akin the "foraging" plants such as the parasitic ivy. In this case the set of symbols is no longer interpreted solely as drawing instruction but carry information on a given organ's position, type, etc inside the plant's hierarchy. But, though the paradigm is commonly thought of as intricately linked to the creation of branching plant-like structures, they have also been used for other purposes such as animal generation.

Indeed in [Hornby et al. 2001] parametric L-System have been used to produce mobile creatures. As illustrated in figure 2.5a, the body plan was generated through repetitive replacement of the axiom. Control over the limbs' motion was hard-coded into the L-System's through a set of four rotation-capable joints. Through a fitness function rewarding displacement of the creature's center of mass and a genetic algorithm, motion-capable creatures were obtained that displayed a variety of motion strategies amongst which rolling, undulating and flipping.

Similar work was done more recently in [Miras et al. 2018] with the objective of evolving robots that could be deployed into the real world thanks to modular robotic parts. In this case morphology and controller are only loosely connected in the sense that both are derived from the same genotype contents but with a more complex, indirect, mapping. Indeed, when applying the L-System to the derive the axiom into the final phenotypical string two sets of commands are thus produced: those that affect the positioning of the building blocks, the morphological components, and those that are concerned with connecting neural elements to produce the brain of the creature. The evolution process was focused on the maximisation of speed and novelty (describe later on or see [Lehman et al. 2008]). The resulting animals exhibited once more a broad



(a) Hornby et al. 2001 (b) Miras et al. 2018

Figure 2.5: Applications of L-Systems to motile creatures generation

range of gait from rolling and crawling to walking.

But L-Systems have also been used in more specific tasks than whole-body morphogenesis: in [Prusinkiewicz et al. 2013], the authors investigate the use of such a grammar in controlling the diffusion of hormones to reproduce differentiated growth processes. Through context-sensitive, parameteric L-Systems controlling the rate of division they obtained a generative methodology resulting in the same shape as that biologically obtained in the ivy leaf (figure 2.6).

2.2.2 Graphtals

In 1994, K. Sims published a pair of seminal articles [Sims 1994a,b] in which he presented creatures of unprecedented complexity which devolved in three dimensional environments. The morphogenetic controller was based on directed graphs (also referred to as "Graphtals") in which nodes contained all the information required to generate a given organ type. This include, but is not limited to, the dimensions, joint-type and an internal neural network. Directed links coded for the creation of organs on one another in a hierarchy of structural components. They also contained data pertaining to the relative position and orientation, the scaling and possible reflection. The use of such repetition allowed for compact encoding of complex, yet regular, structures as shown by the hand-designed examples of figure 2.7a.

As in [Hornby et al. 2001], the linkage between body and brain was straightforward given that each morphological portion of the creature was also endowed with its own internal network. These were combined with one another, and an additional network that could serve as a control center, to produce the complete neural system of the creature (fig. 2.7b). As with the body, the repetitiveness of innervation allowed for complex control over multiple parts by virtue of the

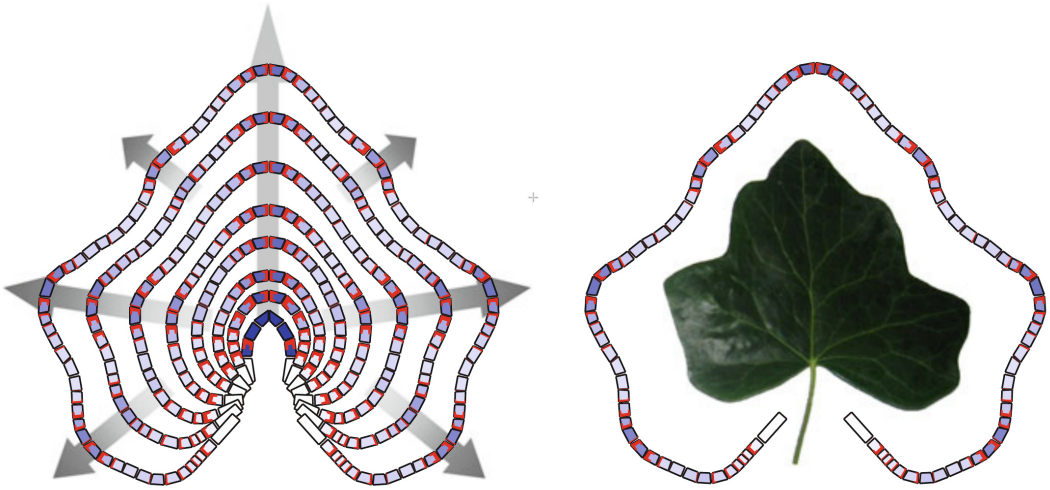


Figure 2.6: Differentiated growth of a cell structure can reproduce ivy leaf [Prusinkiewicz et al. 2013]

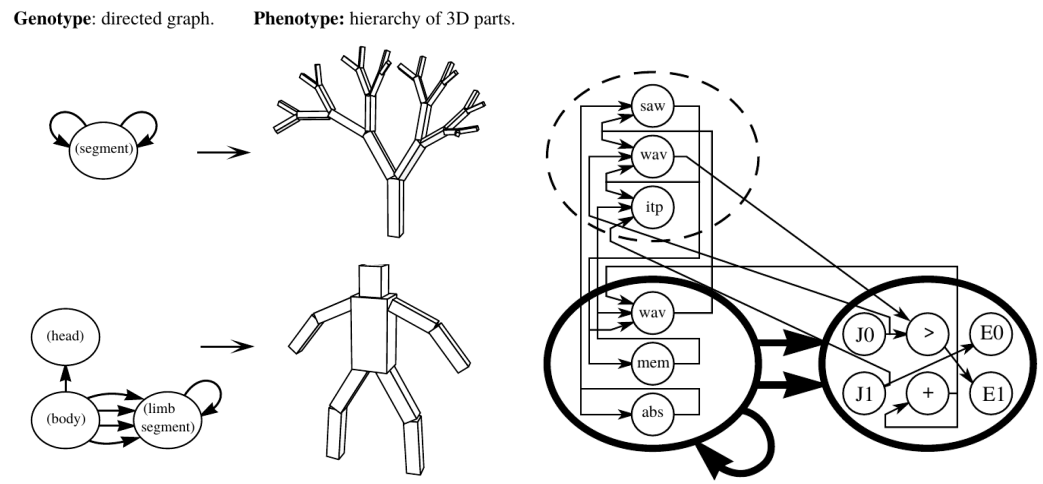
reuse of similar components.

These creatures were embedded in 3D environments and, in the first paper, were evolved solely for their motile capabilities: swimming, walking, jumping and following. They were obtained through a Genetic Algorithm in which the fitnesses measured the distance traveled during evaluation according the specific criterion under examination. The evolution of graphtals is relatively easy to code for, thanks to they being composed on numerous small bits, the contents of which being elementary. Both mutation and crossover of graphtals were thus implemented.

The resulting creatures showed a wide range of behavior most notably for the swimming portion of the evolution in which convergence with biological products was observed. Indeed in figure 2.8a, one can see snake-like individuals with wavy motions or others using fin-like appendages to propel themselves forward. More unnatural-looking objects were also obtained such as the three-stories creature (third of second row). Similarly, the evolution of motion on a solid ground provided a panel of individuals ranging from the gorilla (rightmost third row) to more surprising strategies involving repetitive jumping thanks to an inertia-producing appendage at the front.

The second paper dwelled upon the emergence of competitive capacities: two individuals separated by a small cube were rewarded for the control they exerted over it, i.e. how much they were in contact with it with respect to their opponent. This instantiation of the Red Queen Effect[1], used an all vs.

[1] The Red Queen Effect is named in reference to a response made by the Req Queen in "Alice in Wonderland": In this place one must run constantly to stay in the same place. In biological terms this refers to an evolutionary dynamic between pairs of competing populations for which any improvement from one of the competitor is matched, at a latter date, by the other. Examples of such a dynamic include predator/prey configurations where every



(a) Genotype to morphology mapping (b) Nodes contain inter-connectable ANN

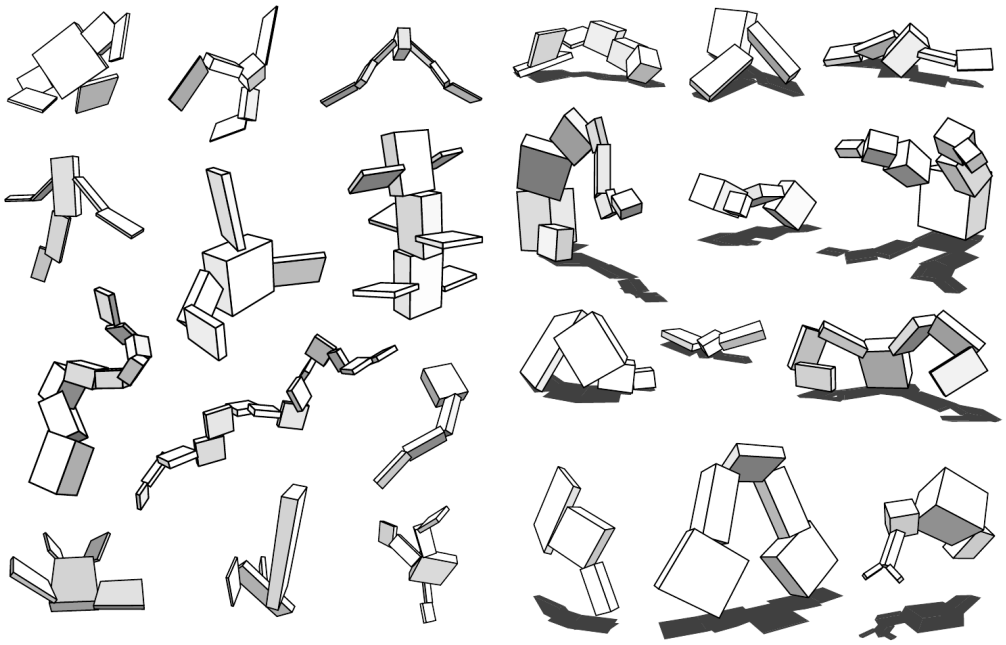
Figure 2.7: Overview of the different parts of a Graphtal: nodes contain morphological and neural data and links specify where each suborgan grow

best type of tournament to co-evolve two species against one-another. The tournament type was devised to limit the number of evaluation from $n(n-1)$ down to n and determine the fitness of an individual from a species by pitting it against the champion of the other species at the previous generation. In this manner, not only is the number of evaluation kept small, but each improvement found by one species can be observed until a counter-acting mutation can be leveraged.

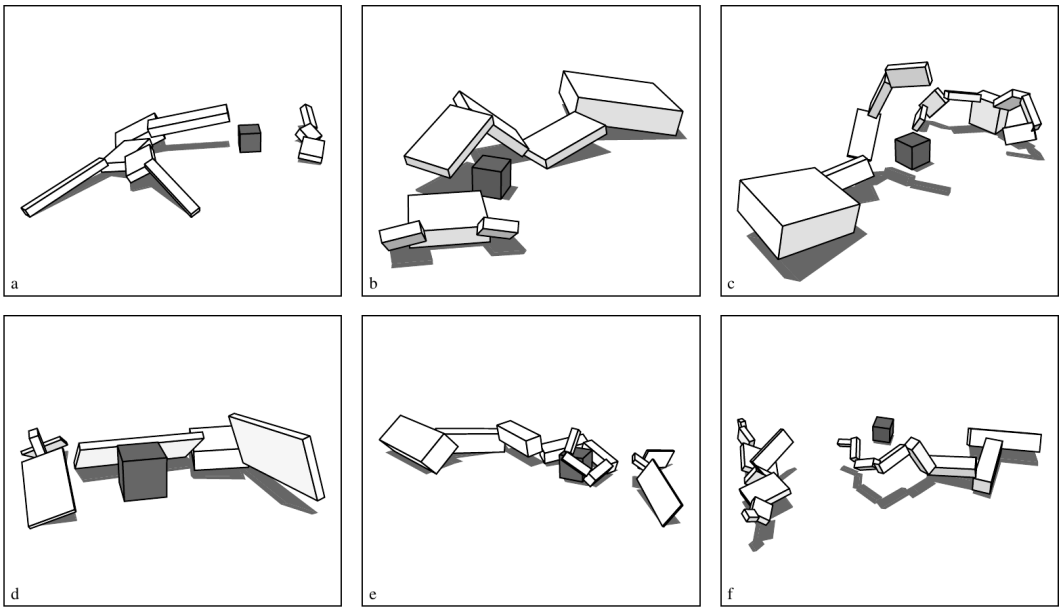
The resulting evolutionary dynamics ranged from complete domination of one species, to uneasy coexistence with each species gaining the upper hand alternatively. Single-species experiments were also performed but deemed of lesser interest by the author. Indeed when looking at the sample of duels from figure 2.8, one can see that using two species has a major advantage: the emergence of "complementary" strategies. Taking the frame labelled (c) as an example, the independent gene-pools allowed for the apparition of the long-reach strategy to be pitted against a variation of the gorilla. Given the formulation of the fitness function, touching the cube is only a viable strategy as long as the opponent fails to do so. As a matter of fact, as soon as both species understand how to properly collide the objective successful strategies are those that minimize the capacity for the opponent to reach said objective: pushing the cube (or the opponent) away, covering it...

Given the generic definition of Sims' graphtals (hierarchy of boxes with repetition) they have seen much reuse in subsequent literature. One such example are Ray's Virtual Pets [Ray 2001]. In this work, the morphological components are also described by a directed graph of boxes containing both

increase in the prey's capability to escape death is quickly counteracted by the predator, and oppositely



(a) Creatures evolved for swimming, walking and jumping [Sims 1994b]



(b) Creatures fighting for control over a cube [Sims 1994a]

Figure 2.8: Sample of diversity from Sims' creatures.

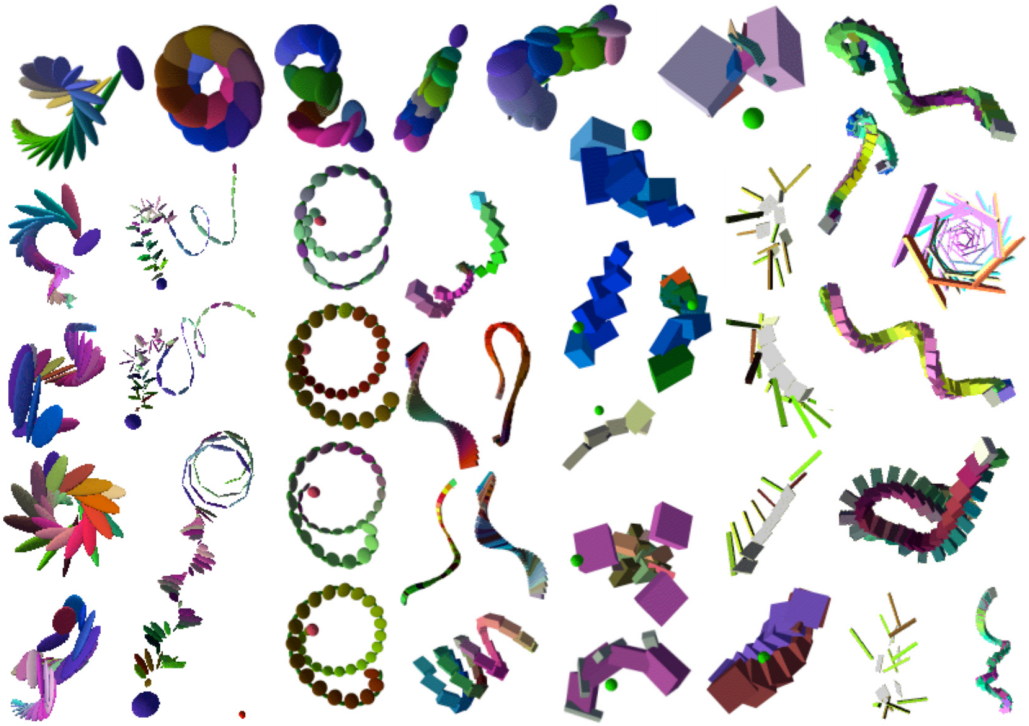
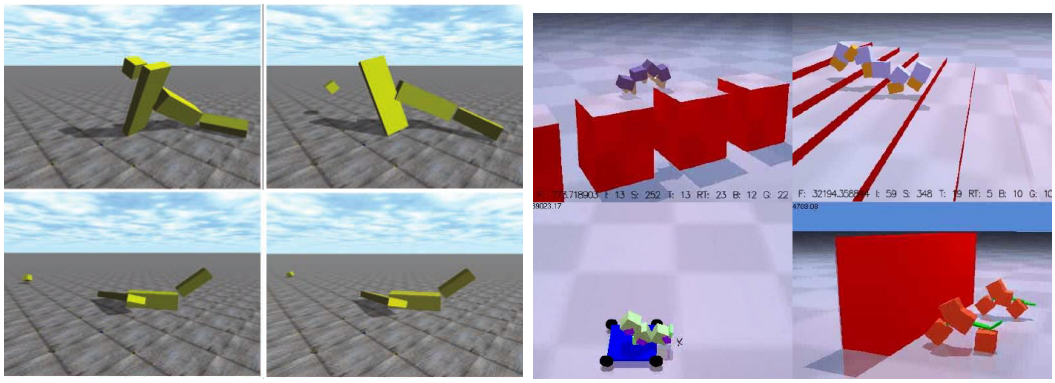


Figure 2.9: Virtual pets with a graphtal-based morphology [Ray 2001]

shape and neural data. The main difference lies in the objective of the experiment: while Sims' creatures were designed to perform specific tasks those of Ray were evolved solely for aesthetic purpose, to serve as pets. Thus, as in Dawkins' Blind Watchmaker, a user was the driving force of the selection process. The resulting creatures were capable of motion but much more as an afterthought, a means to show them behaving in a seemingly meaningful manner. More importantly, from the author's point of view, was the emotional response provoked in user watching such creatures engage in idle activity. Thus much more attention was spent on the external appearance of these pets: not only is each block color-coded but the neural system also affects this palette to produce patterns upon their surface, though only static snapshots are visible on figure 2.9.

Improvement upon the capabilities showed by the initial graphtal-based creatures was also investigated as in [Chaumont et al. 2007] where the authors studied the evolution of catapults. These differed from more traditional walkers by their requiring some portions of their morphology to remain constant throughout evolution. Indeed, the "projectile" is modeled as a regular body block with no neural component and is connected to an arm that serves as the root of graphtal. Both are unchanged by mutations and recombination. The projectile is disconnected from the arm either half a second after the start of the evaluation or upon receiving a signal from the central neural system. Throwing was experimented in both unguided and directional manners. Besides



(a) Projectile throwers [Chaumont et al. 2007]

(b) Climbers, skaters, pushers [Lassabe et al. 2007]

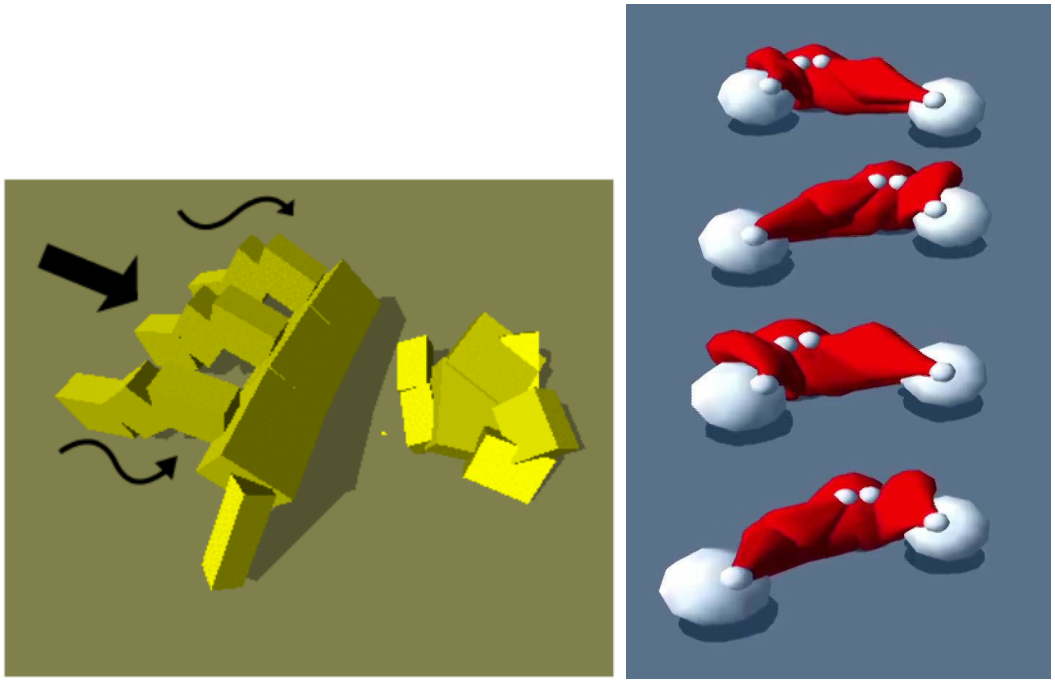
Figure 2.10: Graphtal-based creatures with more extensive capabilities

discovering the catapult-like behavior (fig. 2.10a), some individuals exhibited more "exotic" methods such as the drop kicker (releases the projectile before colliding strongly with it) and the acrobat (stands on the projectile and times the release with the optimal moment of its "summer-sault").

In a similar trend, the work of [Lassabe et al. 2007] explored a different set of motile capabilities: gap-crossing, stair-climbing, wall pushing and even skateboarding. The morphological basis remain unchanged, that is the final body plan is still extracted from a directed graph, but in this case control is effected through composition of patterns. Through a system of classifiers, patterns in $[-1; 1]$ are selected from a hard-coded database of 1K random items and can be adjuncted to one another when they exhibit sufficiently close boundary conditions. A series of individual experiments of increasing difficulty show the large potential of this methodology though it remains open whether such diversity of behavior results from the use of composite patterns or if they could have been obtained in the initial experimental setup.

Another author which first attempted to reproduce the results from Sims' experiments [Miconi 2008b], also expanded upon the model of co-evolution. Interested as they were by the cube competition, they built a low-level form of conflict which relied on computing forces magnitude and directions in case of block-block collision to determine the relative roles of aggressor and victim. Thus, in this work, the fitness is actually more the result of a boxing contest rewarding high-damage, high-defense creatures. All other components are similar to that of the seminal publication including the form of the two species' co-evolution (Last Elite Opponent). Resulting creatures showed flail-like strategies where long connected segments are used to sweep in the direction of the opponent. Instances of steam-roller (figure 2.11a) where also found where a single large block is pushed by numerous appendages to maximize speed of collision.

To conclude this section on graphtals, we turn our attention to an attempt



(a) Evolved predation [Miconi 2008b] (b) Soft muscles [Lessin et al. 2015]

Figure 2.11: Additional instance of extended graphtals

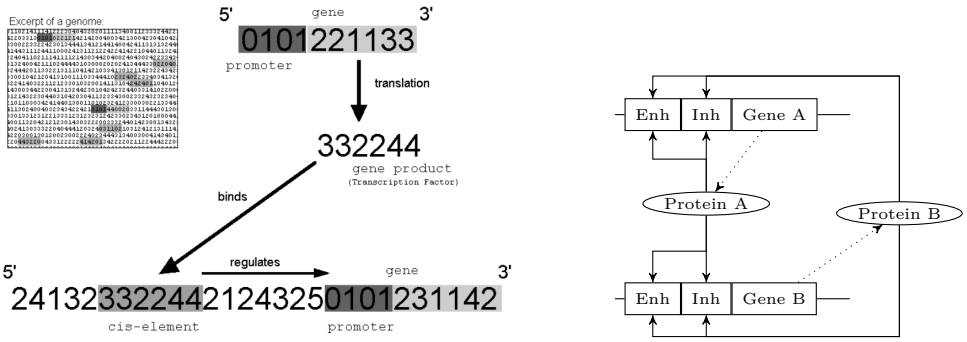
a producing less "blocky" demeanor. The rationale behind the work of [Lessin et al. 2015] is that biological animals tend to be frames of bones actuated by soft, fleshy parts. Quite unlike other instances of these creatures, the authors leveraged the capabilities of the PhysX engine to model soft bodies and, through neural-controlled stretching stiffness, achieved contractable muscle-like objects. This led to creatures with wildly different manners of locomotion as those obtained in previous settings as illustrated by the "crawler" of figure 2.11b.

2.2.3 Genetic Regulatory Networks

Another approach to the generation of virtual creatures is the use of lower-level building blocks akin to cells. Indeed in both previous methodology the atomic components of an individuals were capable of potentially complex actions. Through asymmetrical cell division and specialization, one could, in theory, reproduce the panel of diversity generated by biological morphogenesis.

Such a paradigm was addressed by the conception of Genetic Regulatory Networks (GRNs) which mimic part of the complex intra-cellular machinery. Though multiple models have been developed in the last two decades, the first contributions could be traced by back to two models: [Banzhaf 2003; Reil 1999].

In both, GRNs are defined as strings of bits which, when appropriately parsed, can be translated into a network of interacting gene-protein pairs. In [Reil 1999], genes are defined by a single component, as illustrated by figure



(a) Pattern [Reil 1999] (b) Distance [Banzhaf 2003]

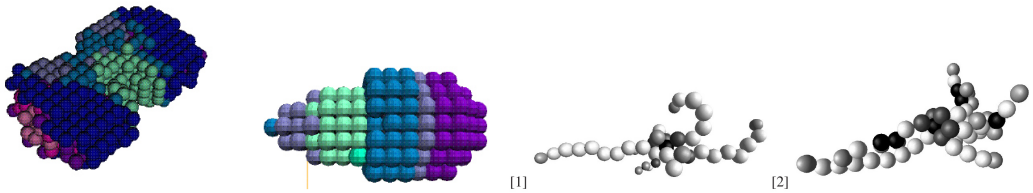
Figure 2.12: Two matching types for GRNs: pattern and distance

2.12a. The transcription factor, or gene product, is obtained by increasing each digit of the gene sequence. The resulting effect upon the other genes is obtained by collecting, from the rest of the genome, the list of matching sequences. These are then be defined as the regulatory elements of the gene directly after (but not necessarily contiguously so). Thus each gene is under the indirect control of up to n such elements.

The approach used in [Banzhaf 2003] differs in a number of major ways. First, the gene product is coded by 5 32-bits integers which provides a pattern of 160 bits instead of the 6 digits previously seen. The translation from this collection of bits into the expressed protein is performed by a majority rule. Second, to each gene exactly one inhibitor and one enhancer sites are defined, both also of 32 bits in length. This allows, when comparing proteins with each of these sites, for a more continuous definition of "matching". Through a XOR operation between both strings, the resulting score can go from 0 (both string are identical) to 32 (every bit in one string is flipped in the other). Thus, instead of a strict pattern matching, this model allows for more diffuse interactions between pairs of gene-proteins as illustrated by figure 2.12b where the resulting graph is strongly connected. In this instance, the effect of a single protein upon a gene is thus a function of its matching with the corresponding site (inhibitory/enhancing) resulting in complex networks of interactions.

Both of these models however show only limited practical utility in producing artificial creatures. Indeed one of the first attempt at such low-level morphogenesis was done in [Eggenberger 1997], even before the previous two. Though closer to the affinity-based model of [Banzhaf 2003], subtle differences can be found notably in the manner in which genomes are created. In this work, each word of n integers is interpreted as a gene, either a regulatory (which will control downstream units) or structural (which produces morphogenes). The matching between a transcription factor and a given gene is also continuous resulting in massively interconnected networks of interaction. One other crucial difference lie in the consequences of structural gene productions.

Unlike the previous theoretical models which were only concerned with the



(a) [Eggenberger 1997] (b) [Bongard et al. 2003]

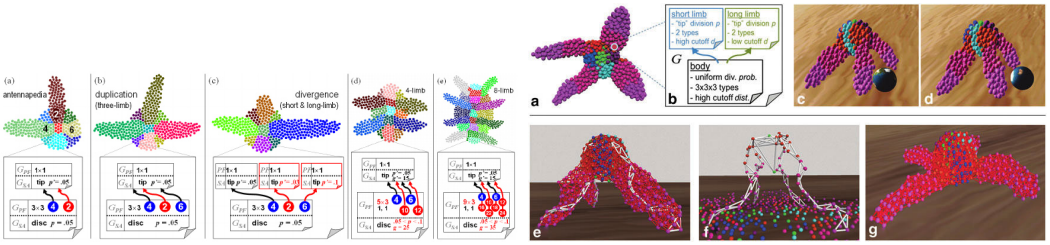
Figure 2.13: Using GRNs for fine-grained morphogenesis

internal dynamics of protein concentration, cells are expected to perform certain specific actions: adhesion, division, apoptosis, etc. This is coded in the first three integers of each gene allowing for 6^3 possible classes of action. Additionally, transcription factors undergo a process of diffusion to affect not only the current cell but also its neighbors, thus allowing, through cell differentiation, complex structures to emerge as illustrated by figure 2.13a. In these examples, evolution of the GRNs was guided towards production of bilaterally symmetric shapes thanks to globally placed morphogens.

In [Bongard et al. 2003], a similar approach is explored with yet another model of artificial genetic regulatory network. Here the genome consists of 100 floating-point values in $[0, 1]$ with all those below a specific threshold designated as promoter sites. A further seven values are used to produce complex behavioral rules regulating the specific pairwise interactions between genes and the products they generate.

Starting from a single cell, the GRNs is stepped through time and allowed to perform its protein-generating routines. Specific products induce a growth in these units which, upon doubling in size, split into two new units of default radius. In addition to the cellular events, a neural network is simultaneously grown using different outputs of the same regulatory network. After a process of artificial selection over 200 generations, the resulting creatures were able to move towards a target object. The evaluation is decomposed into two unequal portion: morphogenesis, during which the organism develops its cellular and neural structure and operation, where the ANN is activated and motion is tracked. Larger individuals were obtained when changing the task from simple motion into a block-pushing contest in which size does matter, given the relative weight of the target object (fig. 2.13b).

Large-scale structures, with thousands of cells collaborating in generating a global shape with limbs, are much harder to obtain using low-level GRNs. To address this issue, the author in [Doursat 2009] investigates the use of modular genomes in the form of directed graph, each node containing both Self-Assembly (SA) and Pattern-Forming (PF) instructions. The former is concerned with controlling cell proliferation through the probability of division, cell-to-cell adhesion and, indirectly, maximal range. The latter encodes the GRN proper which, by relying on PF-I gradients, can trigger cell differentiation. Though in this proof-of-concept model the GRN is kept relatively simple, having the



(a) 2D proof-of-concept [Doursat 2009] (b) Semi-rigid robot [Doursat et al. 2014]

Figure 2.14: Hierarchical GRNs for large scale 2 and 3D structures

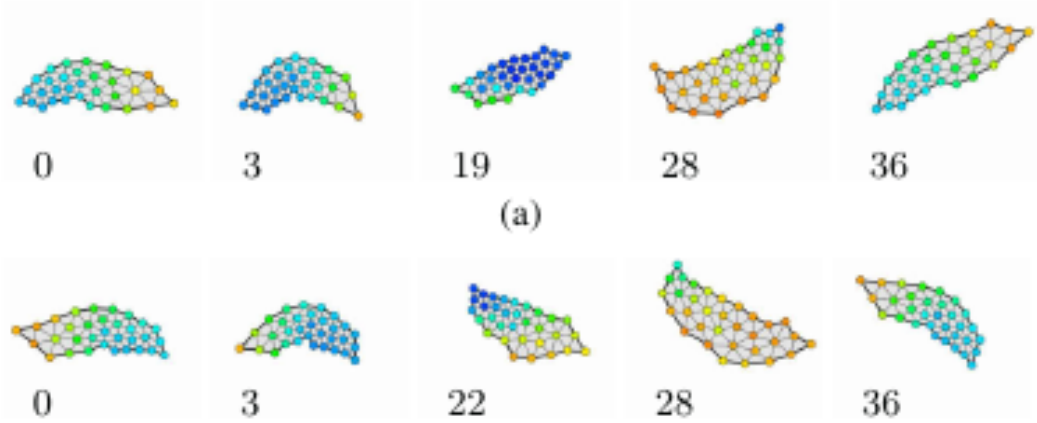


Figure 2.15: Contraction induced by internal GRN [Joachimczak et al. 2013]

hard-coded structure of a three-layer feed-forward network, results show that complex structures can be grown nonetheless. More surprisingly, the examples in figure 2.14a were obtained by manual exploration of the model's parameters. Due to their complexity most models require a form of artificial evolution in order to reach interesting regions of the genetic space but, in this approach, structuration and modularity alleviate part of this need.

An extension of this work was done in [Doursat et al. 2014], where the sane hierarchical GRNs were used to grow 3D robots with different types of cells: muscles, joints and bones. In a physically plausible environment, powered by the ODE engine, these creatures move by relying on the diffusion of cellular contractions to the underlying skeletal structures. Such a motion was, at the time of this work, performed by predetermined schedules but could, in theory, be handled, just as well, by any other form of neural-like controller. In this case only limited evolution was performed, notably with respect to the size of the creature and its limbs.

Following a benchmark evaluation of Banzhaf-like GRNs to a 3D version of the french-flag in [Joachimczak et al. 2011], Joachimczak et al. proceeded to their application to the control of both morphological and behavioral features. In [Joachimczak et al. 2013], the authors develop, from a single cell, two dimensional individuals which are animated by spring-like connections between

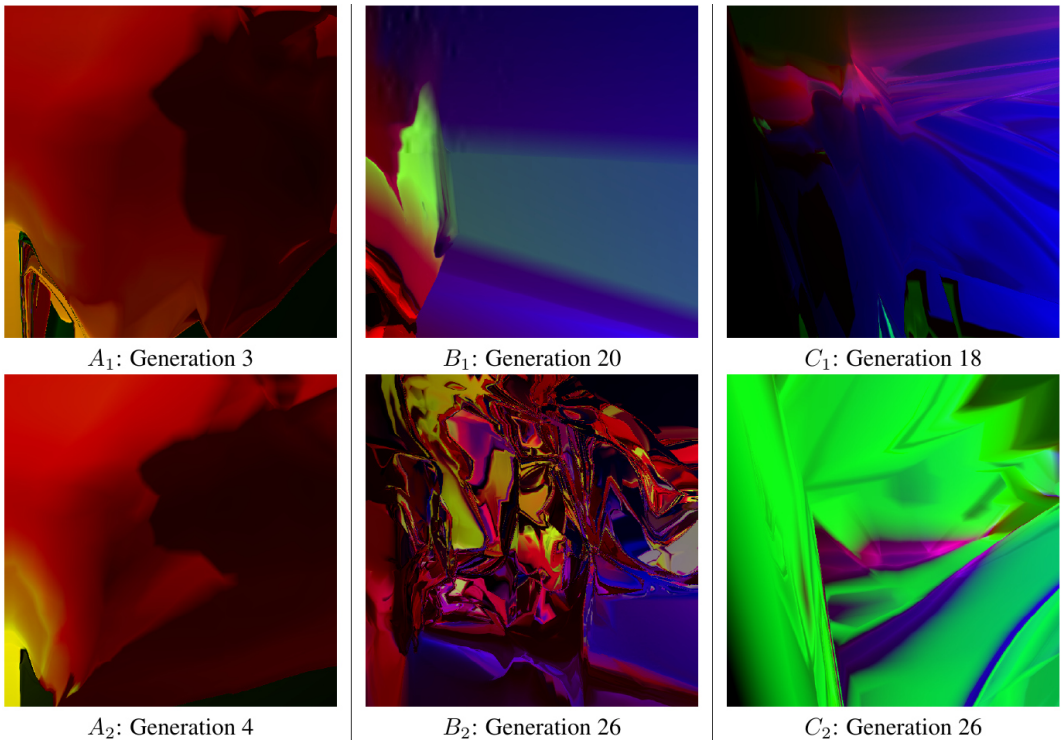
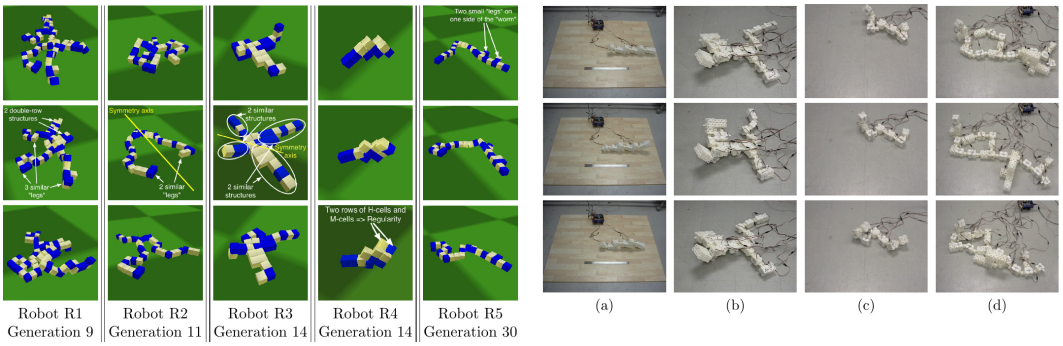


Figure 2.16: Protein interaction viewed as pictures [Cussat-Blanc et al. 2012]

cells. Two pairs of gene product control the expansion/contraction of each cells' springs thus allowing for the same GRN to determine the cellular shape but also the reaction to stimuli. Indeed, evolved individuals were selected for their capacity to swim through a fluid medium and eventually reached a food droplet that emits a continuous, chemical signal. Such an individual is shown in figure 2.15 with the pattern of cellular contraction shown as as a color gradient (red for expansion and blue for contraction) while performing a left (top) or right (bottom) turn. Though not all evolutionary runs managed to reach such robust behavior, this work clearly showed that such a form of chemiotaxis was entirely within the capabilities of Genetic Regulatory Networks.

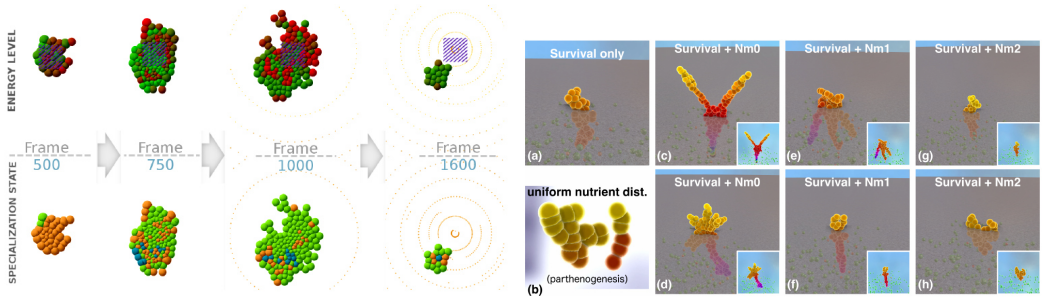
Indeed the fully connected network resulting from affinity-based gene-protein interaction allows for untold complexity in the dynamics of proteins concentration. One manner in which this complexity can be observed directly is through the production of RGB picture, as done in [Cussat-Blanc et al. 2012], where the relative position of a pixel is fed to a GRN which, in turn, outputs a color. A few examples of such images can be seen on figure 2.16 where the controller is quite similar to that previously described: gene have an id, enhancer and inhibitor integer which interact with one-another according to exponentially decay differential equations. The evolved picture were obtained by application of a Blind Watchmaker algorithm and show remarkable regularities and diversity. Additionally, given the use of relative coordinates such pictures are infinitely scalable.



(a) In virtual space (b) Physical instantiation

Figure 2.17: 3D printable robots controlled by a periodic oscillator [Cussat-Blanc et al. 2014]

A few years latter, the same authors investigated how such GRNs could be used to produce the body plans for 3D printable robots [Cussat-Blanc et al. 2014]. Relying on a few generic building blocks (structural and hinge-motors) with reasonable price-tags, they coupled virtual evolution, so as to obtain viable locomotive strategies, with manual assembly to instantiate their creatures into the physical world (fig. 2.17). In addition to the regulatory network previously described, these also relied on multiple morphogen (encoded as 3 or 4 dimensional Bezier curves) that are fed as inputs. These are dependent on both the age of the evolving creature and the identificator of the cell in which the GRN is stepped so as to provide different specialization trajectories in different portions of the body. After a developmental period during which cells divide and differentiate, the resulting creature is then evaluated in a physical environment (based on the Bullet engine [Coumans et al. 2013]). Motion is obtained through periodic oscillation of their hinge-motor blocks. One of the key feature of this approach is that, by relying on generic parts, the manual assembly step required to translated the evolved body plan into the physical world could, at some point, be completely automated thus allowing human-free generation of task-specific robots.



(a) Defense vs. toxins [Disset et al. 2014] (b) Resource collection [Disset et al. 2016]

Figure 2.18: Functionality from shape

This GRN model was further used in [Disset et al. 2014] to produce reactive creatures that develop responses to varying levels of stress. This stimulus takes the form of increasing levels of toxic proteins that degrade the integrity of affected nutritive cells. Another type of cells show stronger resistance to such toxins hence their being called defensive. The evolved GRN thus have to balance between nutrition collection and survival capabilities mostly obtained through a continuously growing structure that moves away from the initial center (where the concentration in toxins is higher). A second experiment with harsher resource levels was performed which required, and indeed obtained, the use of storage cells to survive increasingly longer periods of starvation (fig. 2.18a.

By extending the simulation platform MecaCell, the same authors also explored the generation of 3D plant-like structures [Disset et al. 2016]. These were embedded in a two-fold environment: light could be gather from the top layer while nutrients hot-spots were scattered throughout the lower portion. Thanks to the use of a novelty metric, the authors were able to evolve viable strategies some of which were, surprisingly, using a form of parthenogenesis to increase their effective range of collection.

2.2.4 Composite Pattern-Producing Networks

Further removed from bio-mimetism, Composite Pattern-Producing Networks have been introduced in [K. O. Stanley 2007] as context-free method from morphological development. In terms of artificial creations these have been mostly used in picture generation, as further explored in [Lehman et al. 2012], the complexity of which is in par with those generated through GRNs (fig. 2.19).

These rely the building of a single function by aggregation of multiple simpler primitive such as absolute value, gaussians or sinusoidals. In a similar manner to that of Artificial Neural Network, inputs are assigned to the local values of the corresponding variables, the pixels' absolute locations in the case of images. By a forward-feeding process these are transformed through potentially numerous functions to produce the output: the gray value. One of the salient point of such a method is its complete independence from local information, that is no lengthy developmental procedure is required. Whether or not this is a desirable feature, however, is beside the point of this manuscript.

An early application of this methodology to the generation of proper artificial constructs was undertaken in [Auerbach et al. 2010]. In this work, the developmental model used the CPPN to query, for a cloud of position surrounding an initial cell, the matter density at this position. Above a given threshold a "daughter" cell was created with the given density. The procedure was then performed recursively until a maximal proliferation value was reached. Unlike in the CPPNs' statement of purpose, however, local information was used besides the (x,y,z) coordinates of the potential cell site: angles from

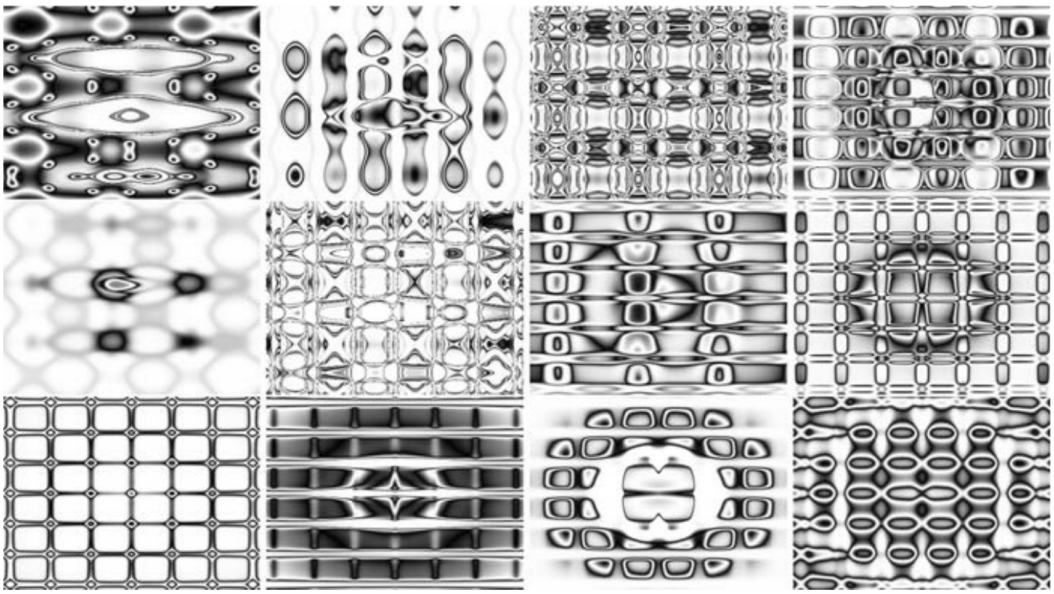
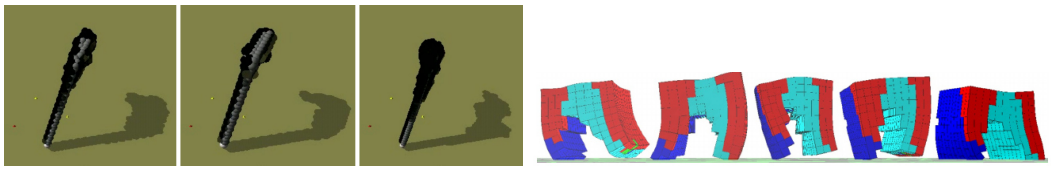


Figure 2.19: CPPN-generated picture show large degrees of complexity and regularity [K. O. Stanley 2007]



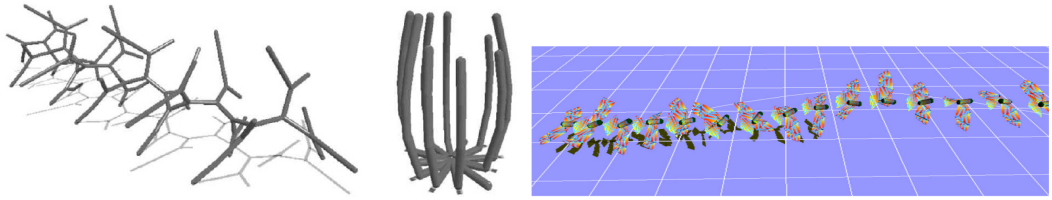
(a) Multi-scale structure (391, 200, 582) [Auerbach et al. 2010]

(b) Soft-bodied robot with horse-like gait [Cheney et al. 2013]

Figure 2.20: Application of CPPNs to the morphological challenge

parent and grand-parent cells and depth. These CPPN were evolved through CPPN-NEAT, the algorithm for iteratively increasing the complexity of a CPPN also introduced in [K. O. Stanley 2007], for their capacity to produce falling structures. As in the case of the GRN, such a model was shown to be scalable (fig. 2.20a) with instantations of the same genome with different cellular resolution being straightforward.

Another use of CPPNs, with a strong emphasis on physical instantiation, was performed in [Cheney et al. 2013]. Here, the inputs solely consisted, in addition to the 3D position of the voxel, of its distance to the center of the creature. Additionally, five outputs were present: one also coding for the presence or absence of matter at the given coordinate and the other four describing the type of voxel to create (periodic muscle, phased periodic muscle, passive and structural). Given the limited reliance on local information, given that no data on the other cells' state is required, this model is more morphological than developmental: a cube of size 10^3, is uniformly sampled by the CPPN with disconnected peripheral components being removed from the morphology



(a) The framestick framework [Komosinski 2003]

(b) Path-following flying creatures [Shim et al. 2004]

Figure 2.21: Examples of atypical phenotypes with repetition

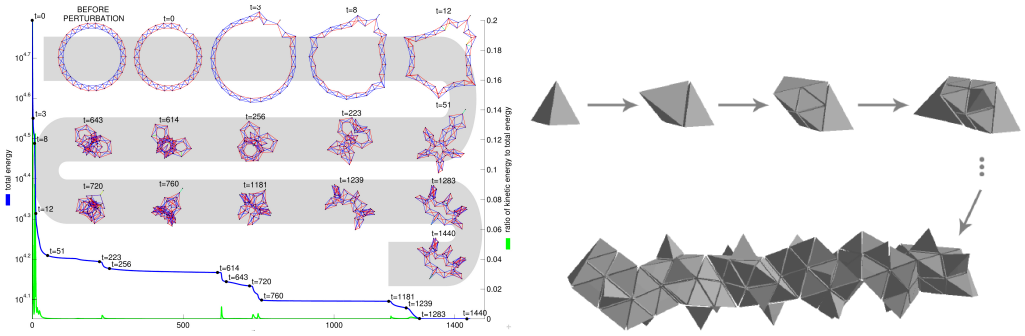
during a post-processing phase. Trough the same evolutionary algorithm (CPPN-NEAT), creatures capable of directed locomotion were found, with a large panel of different strategies, some of which are similar to those from [Sims 1994b].

2.2.5 Other morphological controllers

Besides these well-defined model, some noteworthy tangential approaches have also been made in terms of creature generation. One such instance is the work of [Chavoya et al. 2006] in which Cellular Automata were evolved to produce specific geometrical primitives in 2D (square, diamond, triangle, disk) and 3D (cube, sphere). Another un-classifiable example is that of the Framesticks framework [Komosinski 2003], which is not concerned with a single morphological developmental model but, instead, provides the low-level architecture for exploring numerous such models in, optionally, physical environments. As such there is no clearly defined genetic space, the creatures shown in figure 2.21a being examples of a developmental strategy.

Semi-fixed morphologies

Other models used fixed morphology to explore specific aspect of locomotion such as [Shim et al. 2004] in which a rigid base morphology composed of a body and two wing roots was evolved to produce a panel of (symmetrical) wing shapes. The evolved individuals showed, initially, limited capacity for active flight preferring instead to rely on a gliding behavior but some showed sufficiently robust flapping patterns to follow a given trajectory (see fig. 2.21b). Another such instance is the work of [Olson et al. 2016] in which the investigation of predator/prey dynamics was solely focus on eye placement and vision angle. The morphological components were thus limited to these two variable, nonetheless allowing for the emergence of a coevolutionnary cycle alternating between predators with coarse/focused retina and dispersed/swarming preys.



(a) Perturbation-based invagination [Fernández et al. 2012b]

(b) Tetrahedron subdivision [Rieffel 2013]

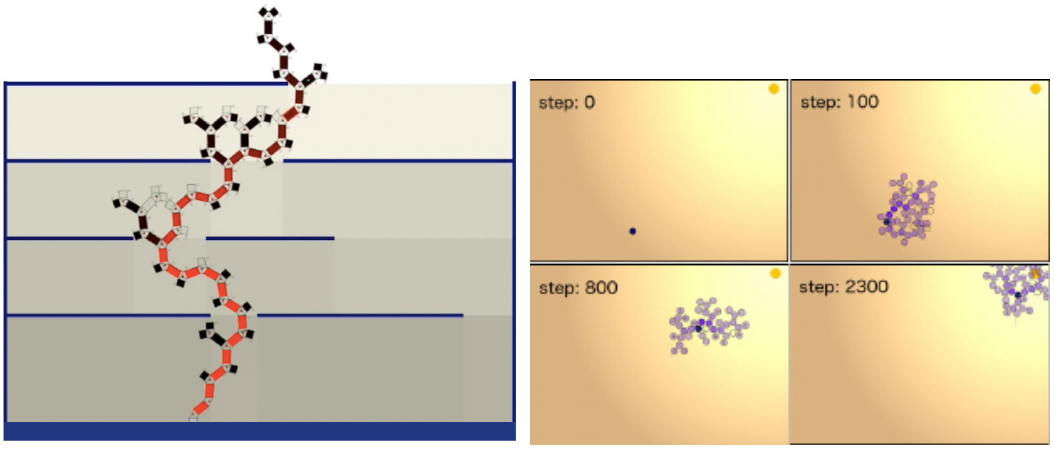
Figure 2.22: Exotic genotype-phenotype mappings

Novel controllers

Unlike the previously described literature some work investigated widely different models for morphological development as illustrated by [Fernández et al. 2012b] which used a variable-length series of perturbations to trigger a from of invagination in a circular sheet of 2D cells. Each of these have 6 elastic links controlling its quadrilateral shape and the perturbations are defined by the subset of cells and links they are applied to, in addition to the type of its interference (increased stiffness, resting length). When applied to a small disk of 26 cells, complex morphologies, similar to Dawkin's Biomorphs, are obtained. However, in this developmental model, the evaluation was designed to produce multiple intermediary steps, as shown in figure 2.22a, resulting in a morphological process with strong similarities with biological embryogenesis.

Similarly, [Rieffel 2013] is devoted to a face encoding grammar that are used to grow tetrahedral meshes. These associate a given face with an action, in similar manner to the working of a L-System. The likeness stops there, however, as in this case the action is not based on rewriting rules but on division procedures. Through relabelling (face X is renamed Y), growth (a new tetrahedron is placed on face X) and division (face X's tetrahedron is subdivided into four smaller parts) the authors evolve motile creatures. These are animated by periodic variation in their meshes' stiffness and perform better, in terms of morphology and behavior, when the evolutionary process slowly raises the number of allowed rewrites.

Another recent model, the Vascular Morphogenesis Controller, was designed to model the branching mechanisms in plants thanks to the use of "successin", in analogy to auxin. Presented in [Zahadat et al. 2017a] the model showed that, by rewarding efficiently performing portions of the structure, differential growth was obtained which allowed the plant to cope with varying harshness (fig. 2.23a) or non-uniform conditions (partial occultation, maze traversal). This model was also applied to the emergence of a vegetative form of motion in [Zahadat et al. 2017b] (fig. 2.23b) where module production is biased towards



(a) Vegetals (b) Animals

Figure 2.23: VMC for vegetals and animals modeling [Zahadat et al. 2017a,b]

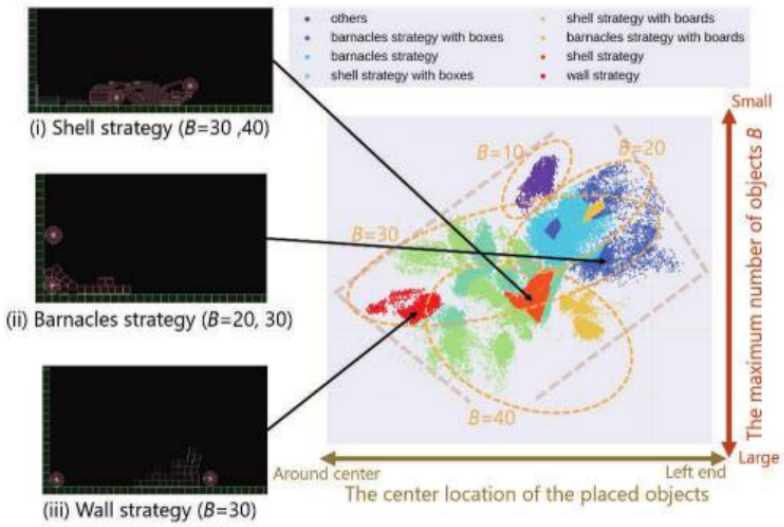


Figure 2.24: Extended phenotypical expression in prey/predator evaluation [Chiba et al. 2017]

the target point. By promoting death in the furthest modules and, conversely, growth in the "forward" direction, creatures with similar capabilities to those presented in [Disset et al. 2014] were obtained.

Long-reaching genomes

But, as highlighted in [Dawkins 1982], a creature's body is not the only manner in which genes can interact with their environment to maximize their change of propagation. Such a experiment was performed in [Chiba et al. 2017] with a study focused not on the creatures themselves, which were simple disks, but on the constructs the prey individuals were able to "build" in order to protect themselves from the predators. By using ANNs to control the decision of "shell-building", multiple strategies evolved which the authors link to biological examples (shell, barnacle or wall as shown in figure 2.24).

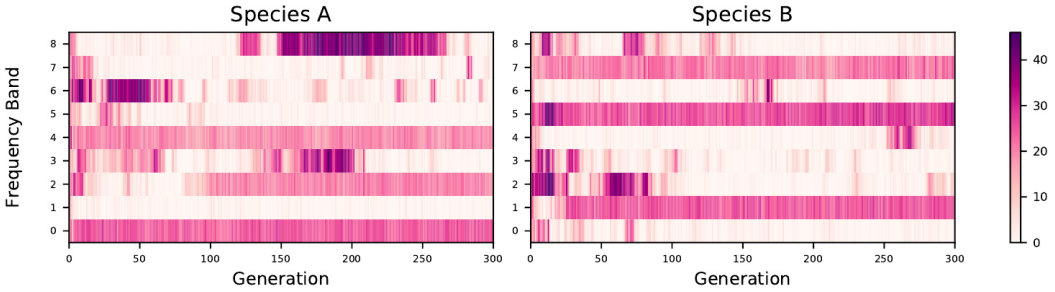
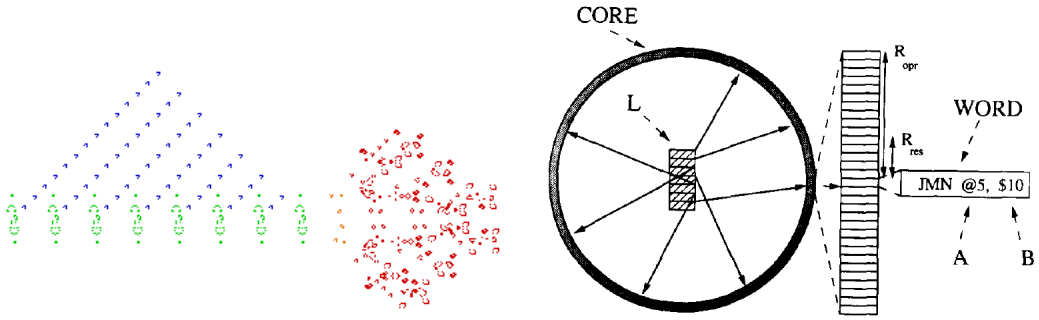


Figure 2.25: Soundscape ecosystem [Kadish et al. 2019]

Communication also is an instance of the long reach of the gene which is explored in [Kadish et al. 2019] through a soundscape ecology. In this article, behavior is also controlled by ANN with the direct encoding of the NEAT model, the main difference resting with the expressed phenotype which is a sound pattern on a 3 out of 9 available bands. The objective for individuals is then to correctly determine if the contents of the message and whether or not it is conspecific. Co-evolution of two hard-coded species cohabiting in the same soundscape showed that niching did emerge as illustrated in figure 2.25 where the random use of all 9 channels is quickly replaced by exclusive focus on 3 of them, per species.

2.3 Ecosystems

All of this literature has shed some light on the mechanisms by which individual morphology can develop from both direct and indirect encodings. These however were generally limited to a few creatures at a time with no purposeful interactions between them. In this section, we will devote ourselves to examining some preeminent work on larger-scale ecosystems. Our focus will be on self-sustaining systems, i.e. those for which the reproduction process is an integral



(a) The Game of Life [Gardner 1970] (b) Coreworld [Rasmussen et al. 1990]

Figure 2.26: Seminal instance of self-sustaining ecosystems

part and which are concerned with the evolutionary dynamics at the individual and upper scales. Thus, articles presenting advances in the generation of ecosystems for Virtual World purposes (e.g. [Damer et al. 1998; Deussen et al. 1998; Steinberg et al. 1999]) will not be detailed.

Surprisingly one of the first instance of such self-sustaining virtual ecosystems stems from a bare-bone implementation of cellular automata: the Game of Life. Initially described in [Gardner 1970], this system only relied on coarse local rules of the form $X \rightarrow Y$ with X describing the number of surrounding cells flagged as "alive" and Y being the next state for the current cell (either dead or alive). Nonetheless sufficient complexity was encoded in this system to see the emergence of multiple levels of self-organisation: from static structures to the well-known motile "gliders". In rarer cases very large-scale objects were observed some of which of being capable of manufactoring smaller types of others structures as in figure 2.26a where a red "puffer-type breeder" leaves a trail of green "glider guns" which, in turn, produce a constant stream of individual gliders.

One of the most straightforward type of artificial ecosystem to implement on a computer as those of a digital nature, that is to say those in which individuals are self-copying bits of machine code that, by virtue of their self-sustaining capabilities can generate an ecology of computer programs. One of the first attempt in such a direction was done in [Rasmussen et al. 1990], in the VENUS framework (fig. 2.26b), which used a small set of instruction in combination with multiple addressing modes. However, it suffered from a relatively brittle instruction set and, as such, was not thoroughly further explored. It, however, inspired a number of similar implementations which will be the subject of the next section.

2.3.1 Digital

Following in the footsteps of Coreworld, Tierra was introduced in [Ray 1991] which differed on a number of points notably the instruction set. In order to

provide robust primitives for these digital organisms there is only one addressing mode: patterning. With this paradigm, every jump instruction (`jmp`, `jmpb` and `call`) is followed by a sequence of no-operation (either `nop0` or `nop1`). This binary pattern is searched in both direction (only backward in case of `jmpb`) around the call location for a complementary match which does *not* need to be inside the individual's own program. For instance the `jmp 0010` instruction in the second block of the ancestor creature (figure 2.27), sends the instruction pointer (ip) back to the beginning of this block (pattern `1101`), thus generating an infinite loop of reproduction. When such pattern cannot be found in a reasonable range, the instruction is ignored and error condition is stored for latter reference by the "reaper".

In this manner the instruction set is not only kept small (32 in total compared the potential 10^{11} of Coreworld's redcode language when considering permutations with operands) but is also less brittle: given the small size of the pattern space, mutations are more likely to result in valid targets than when using direct addressing by integers. In addition, individuals in Tierra have a sort of semi-permeable membrane: while a program has complete rights on its associated memory region (read, write, execute) it cannot write outside of this area, though it can still read and potentially execute instructions found in a broader region. One exception to this rule is during reproduction: the "mother" program can request a secondary memory slot in which to write its "daugther" code (instruction `mal`). This area is cut loose (and turned into a proper individual) by use of the `divide` instruction.

The external control loop consists of a sequencer and a reaper. The former doles out small time slices from the global CPU to each of the creatures' virtual CPU thus approximating parallelism, with the possibility to favor either small or large genomes. The latter is responsible for removing individuals from the population upon reaching a threshold of globally used memory. Its works as a file (first in, first out) with the added twist that creatures can move towards either end depending on the pertinence of their behavior. Generating error codes from a failed `jmp` for instance, move the offending creature up by one position.

Finally, randomness is present in these system in multiple forms: background noise, fallible copy and flawed execution. The first of these flips individual bits selected from the entire memory range at a low rate (one each 10^{-4} instructions). Errors in the copying process are similar to traditional mutational operators in the sense that they are applied during reproduction (duplication in this case) of an individuals' genome. The rate is higher than for the other (10^{-3} per copied instruction). Additionally, most executions can suffer from minor variations in the computation they perform, e.g. incrementing by two or 0-shifting. Thus sources of randomness are numerous in Tierra, though the authors mentions that all but the fallible copy are unnecessary to generate genetic change and evolution after a sufficient state of complexity is reached.

Starting with a single instance of the hand-written ancestor displayed on the

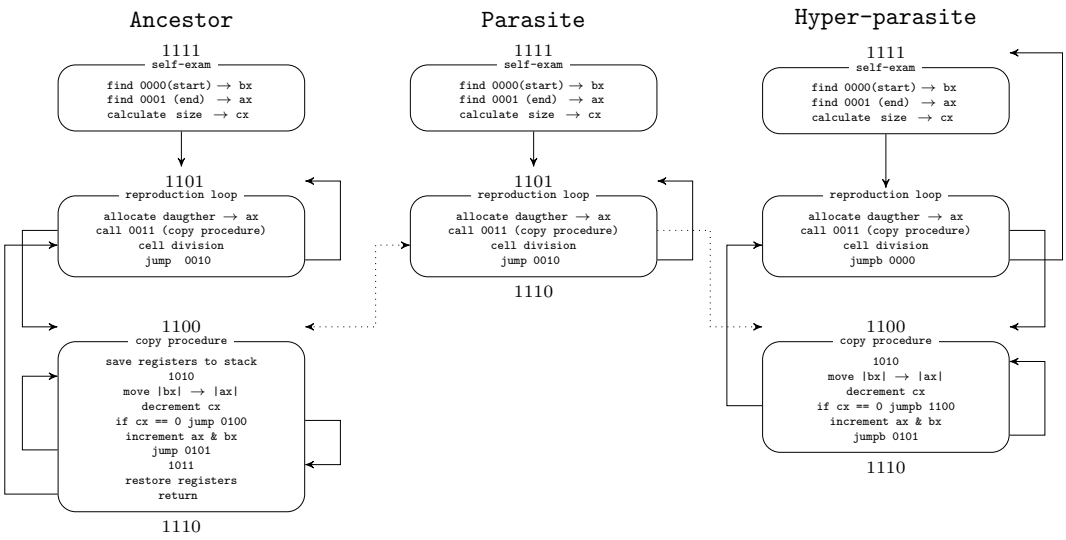


Figure 2.27: Incremental levels of parasitism in Tierra (reproduced from [Ray 1991]). Dotted lines indicate parasitic relationships between individuals (see appendix B for details).

left of figure 2.27, the soup is quickly filled by its descendants. As mutations accumulates, the perceived size of the individuals, which measured in the first block and stored in cx for latter use in the daugther cell allocation, start to change. This, in turn, induces the emerge of more complex strategies than the straightforward duplication of the ancestor: parasitism, hyper-parasitism, social hyper-parasitism, etc. The first two instance are illustrated in the same figure with their specifics being detailed in appendix B.

This model was expanded upon by the introduction of a `split` instruction in [Thearling et al. 1996] which allowed programs to discover, and make heavy use of, multi-threaded behaviors. Additionally, it was also used to study the positive effect of frequency-dependent predation in the maintaining of two prey species in [Shao et al. 2010]. In this study, predators needed to consume exactly $m = 6$ prey before being able to reproduce. Consumption being modeled as one organism obtaining a portion of another's CPU time. The frequency-dependent behavior was designed in such a way that the more a given predator eats a given prey type, the more likely it is to do so again, given the choice. The resulting dynamics showed Lotka-Volterra-like cycles, observed in biological life, in a numerical medium.

Another platform for digital ecosystem is Avida, first introduced in [Adami et al. 1994]. Inspired by the work done on Tierra, the individuals in this system are also computer programs whose genotype is variable length string of binary code (using a different, customizable, language). Unlike its predecessor, however, physical instantiation is done on a 2D toroid grid with a single cell containing either one or none organisms, instead of an instruction-based partition. The reproduction mechanisms is kept mostly identical in terms of individual steps

(size computation, memory allocation, duplication loop, division) with the difference that the newly born "daughter" is not placed in a system-determined location but in the oldest immediate neighbor of the "mother". Environmental noisiness were also simplified with only point-mutation occurring during the copying process and a deterministic instruction set.

Avida also makes use of a scheduler to distribute slices of CPU times to each of the individuals' virtual CPU though with an additional constraint resulting from one of the largest difference with Tierra. Indeed, in Avida, individuals will receive an environmental feedback depending on their performance. This is implemented in terms of increased or decreased relative CPU time conditioned by whether the program was deemed efficient or not. While this makes the task of the scheduler more complex, given that two individuals with wildly differing efficiency might even more strongly differing number of instruction processed per tick, it also allows for a clear-cut evaluation of the individuals in the system. In this manner, the only pressure not only comes from the biotic component of the ecosystem (i.e. the other programs) but also from the experimenter-designed abiotic component. Additionally, the reproduction mechanism precludes the need for a "reaper" as older programs will spontaneously die out when replaced by newer ones. The authors highlight this aspect by pointing out that, in case of this reaper queue every program was interacting with every other due to the re-ordering operations. Such a situation does not occur in Avida and information dispersal is thus only slow and local.

The punishment/reward principle is based on numerical and logical functions that an individual must perform. First it has to retrieve from the environment an appropriate number of input values for the given operation (addition, in the case of [Adami et al. 1994]). Then using the primitives available in the instruction set it must build a computational tool chain to produce the desired output and write it back into the environment. Starting with [Ofria et al. 1999], much more operations were made available (up to 80 logicals), in addition to introducing parallelism-orientated instructions. Furthermore, is was shown in [Lenski et al. 2003] how complex function such as `EQU` (equal, see eq. 2.1) could be obtained by incrementally discovering simpler ones.

$$A = B \leftrightarrow (A \wedge B) \vee (\neg A \wedge \neg B) \tag{2.1}$$

Given that the only logical operator available in the instruction set is `NAND`, the up-hill climb required to reach `EQU`, which comprises more than 20 nands in its naive implementation, is a steep one. Thus evolution in environments in which intermediate functions (e.g. NOT, NAND, OR) were rewarded were found much more likely to produce individuals capable of performing the more complex one than environments without rewards or that only rewarded `EQU`.

In [Fortuna et al. 2013], environmental complexity was further increased in the opposite direction: that of the biotic component. By introducing two types of individuals as ecological seeds they studied the dynamics of parasite-hosts co-evolution. In this work, the interaction with the abiotic is still performed

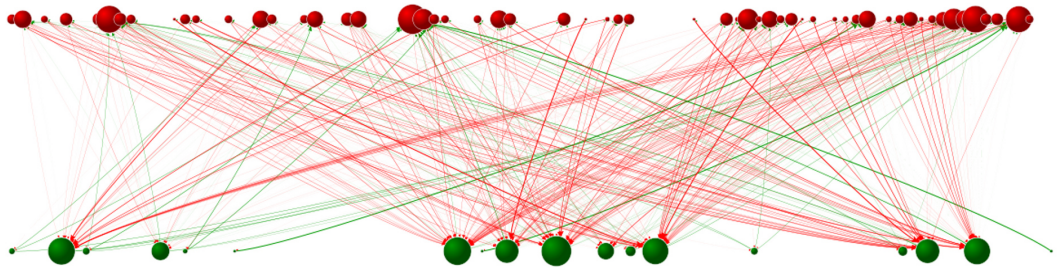


Figure 2.28: Co-evolving networks of host-parasite relationships [Fortuna et al. 2013]

through logical tasks with the same multiplicative advantages as previously described. However, parasites can steal CPU cycles from host with which they share common tasks thus enacting a contradictory push. From this implementation complex networks of co-evolving populations were observed as illustrated in figure 2.28.

But the non-living part of the environment can also be increased, especially given the link between genomic and environmental complexity mentioned in [Adami et al. 2000]. In the Avida platform, multiple such experiment have been performed with [Canino-Koning et al. 2016; Lalejini et al. 2017] focusing on changing the balance of reward and [Luo et al. 2019] on the introduction of cataclysmic periods. In all such cases robustness was found to be stronger than when compared with simpler or static controls. A further improvement was undertaken in [Nahum et al. 2017] where, in addition to dynamical environmental conditions, individuals affected local properties by consuming/producing resources. Though only under a limited number of punctually variations, the ecosystems with both dynamical and heterogeneous conditions were found to "liberate the population trapped on a suboptimal peak".

2.3.2 Unicellulars

By allowing more complex individual behavior, at the cost of increased CPU consumption, ecosystems with more elaborate types of creatures have also been studied with, in this section, the case of unicellular individuals. The Aevol framework [Knibbe et al. 2005] is such an instance in which the genomic component is a double stranded string of bits.

The mapping from an individual's genome into its phenotypical expression is a complex one implying three successive phases: transcription, translation and functional interaction. The first step consists in collating all potential promoter sequences, up to the corresponding terminator, from the genome and, for those similar enough with a long consensus sequence, their expression level l is determined. This depends on the actual distance previously computed and, as will be seen later, conditions the phenotypical importance of the associated

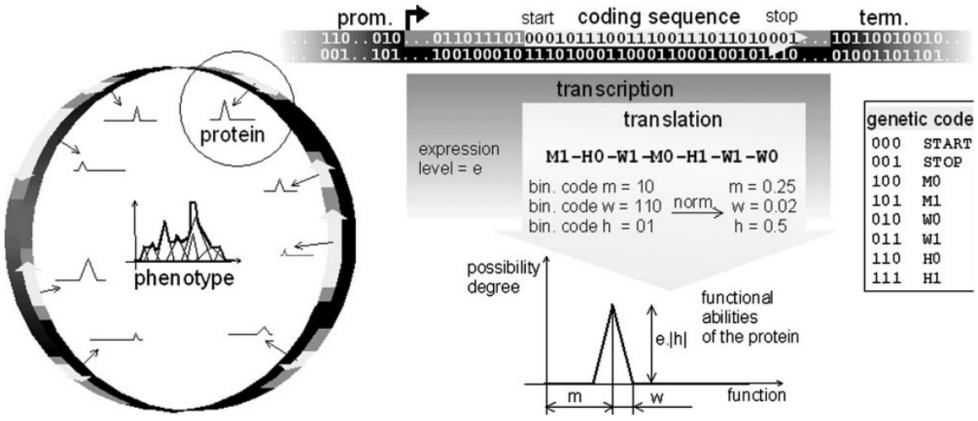


Figure 2.29: Genotype to phenotype mapping in Aevol [Knibbe et al. 2007]

protein.

The translation phase processes all transcribed region is search of start-stop signals, each triplet of bit inside these being interpreted, through a look-up table, as a given protein. As shown in figure 2.29, each such protein is characterised by three components: its mean (m, shift from origin), width w and height H. While the first two are directly derived from the previously read codons, the last is composed of both the genome-encoded value and the expression level described above: $H = l|h|$. Finally, all individual proteins are turned into a single phenotype determining the organisms' behavioral footprint by combining those for which h is positive and subtracting those with a negative h from the result.

The rest of the framework is composed of a traditional steady-state genetic algorithm (SSGA) in which individuals are selected for the matching between their phenotype (combined protein expression) with that of a given, user-determined, environmental pattern. Two types of mutation are modeled in Aevol: punctual events (switching, insertion, deletion) and rearrangements of large genomic segments (deletion, duplication, translocation, inversion). One of the atypical aspect of this work is that individuals have no physical position, indeed they have no physical demeanor besides expressing their protein production profile. Thus, in Aevol, evaluating a single generation is a straightforward business, thus allowing for the exploration of long evolutionary periods.

Indeed, in [Knibbe et al. 2005], the authors explored such an independent evolution for 30K generations which, from initial populations of random 5 Kbp-long genomes, showed the effect of mutation rates on genome size. One such example is provided where high mutation rates resulted in a small genome (980bp, 64% coding) and lower rates induced even larger genomes (21729bp, 11% coding). This result was further confirmed in [Knibbe et al. 2007] with an even more contrasted divergence.

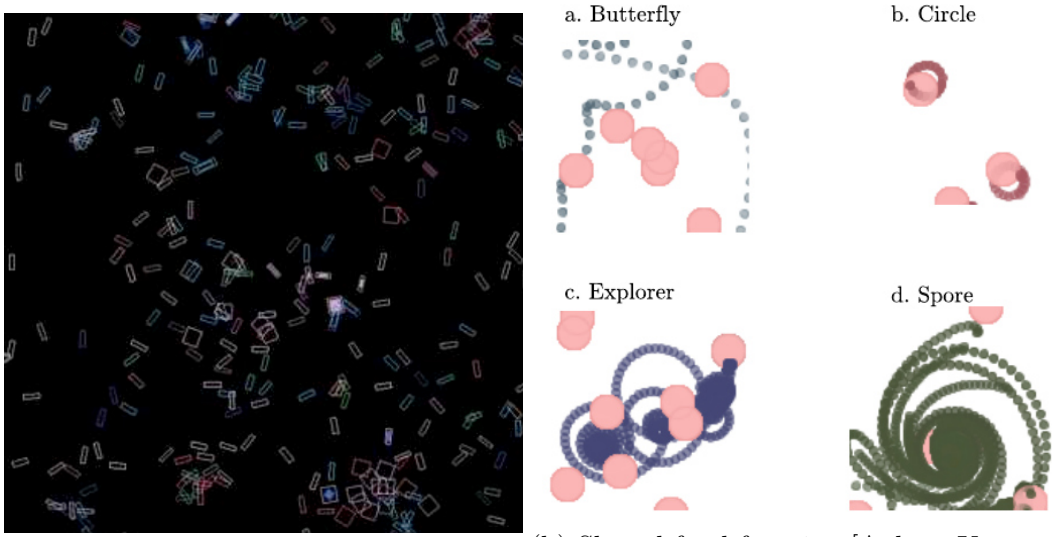
This platform was also used to study the mechanisms of cooperation though

with some modifications to the dynamics of the system [Misevic et al. 2012]. This cooperation is modeled as organisms being able to secrete a "public good" component into the environment. The expression pattern of individual's genome was thus partitioned in two: the left-hand portion still accounting for the metabolism (which determines the fitness w.r.t. the environment) and the right-hand side controlling the amount of public good is generated. Additionally, individuals are no longer behaving as is floating in a well mixed solution but instead have a physical position in a rectangular grid with periodic boundaries. Reproduction is also affected with offsprings being placed in the 3x3 Moore neighborhood of their parent. The public good has a positive effect on the fitness of individuals in contact with it, however it does come at a production cost. Indeed, given the structure of the framework, one individual's production of public good will only affect future generations. Never itself.

The authors were, thus, able to observe of the emergence of cooperation, through the generation of this costly product. It was found that such cooperation was stronger when the public good diffuses and degrades at high rates. This work was followed soon after by another [Frénoy et al. 2012] which showed that historical conditions have very limited impact on the latter dynamics of such a cooperation. Indeed populations submitted to numerous (high) cost of public good production were shown to converge back to similar behavior when the cost was reduced back to more tolerable levels. Thanks to bio-mimetic dynamics resulting from the genomic encoding scheme it was also possible to study, in [Frénoy et al. 2013], how populations of cooperative individuals could prevent the emergence of "cheaters", that is organisms which only profit from the public good without ever producing any. The authors linked entangled genomics with such a resilience, a dynamic in which any single mutation affecting the protein-producing portion of the genome also affects the metabolism, thus reducing the global effectiveness.

But Aevol is not the only system in which point-size individuals were studied. On such work is the disease/host dynamics investigated in [Dorin 2005] in which both types of individuals interact independently, with wildly different time frames. Indeed, hosts are color-coded agents of rectangular shape that freely move and mate in a continuous 2D toroidal environment. By contrast, diseases are only capable of surviving inside another agent and propagate from one to another based on the matching between their colors. One of the most salient conclusion from this work, is the effect diseases have on the affected genomic portions of the host. Indeed, as shown in figure 2.30a, individuals cannot converge on the same color panel for fear of an epidemic. This results in continuous exploration of diversity, kept in check through sexual selection, thus showing the "beneficial" effect of disease on hosts.

A more classical form of predation was investigated in [Erdei et al. 2013] with single-cell individuals being able to choose, through their internal GRN, which role to play in the predator/prey competition. Inert food sources are randomly placed in the environment and replenished at a diminishing rate to



(a) Disease/Host [Dorin 2005]

(b) Shared food foraging [Aubert-Kato et al. 2015]

Figure 2.30: Instances of "bacterial" ecosystems

slowly increase selective pressure. On the other side of the coin, predation was not a completely safe strategy due to an increased metabolic cost and, given that both preys and predators can sense each other, variable levels of success. It was indeed shown that both foraging and evasion (of predators) did emerge. Hunting behavior, however, was not observed which can be explained by the life/diner principle: in this work both types of individuals can feed on food sources which are nice enough not to move.

In another direction, that of discrimination, ambivalent environments were developed in [Ouannes et al. 2014] using a quorum sensing[2] mechanism further developed in [Djezzar et al. 2019]. These also contained sources of toxins which, unlike proper food, have a detrimental effects on the bacteria's metabolism (i.e. loss of energy). By asexual reproduction (upon reaching an energy threshold) these individuals colonize their environment and slowly evolve the capacity to actively pursue food sources while avoiding toxic sources thanks to their internal chemotaxis network. A similar approach was followed in [Aubert-Kato et al. 2015] with a special emphasis on the emergence of food management strategies. Indeed, in this model, food sources can be exhausted by excessive consumption. In such an event another will be created in some other, random, position which still leaves the greedy agent with further foraging as its only solution. Multiple dominating strategies were found, a sample of which is shown in figure 2.30b, which can be placed on the greedy/frugal spectrum. While behaviors such as the Butterfly or Circle have a very conservative management of the food sources, never staying long enough to dangerously reduce their

[2]Ability to detect and respond to cell population density through regulation of specific genes

reserves, other, more aggressive, demeanor were also found as the Spore which consumes all resources until near exhaustion before releasing a mass of offspring in all directions.

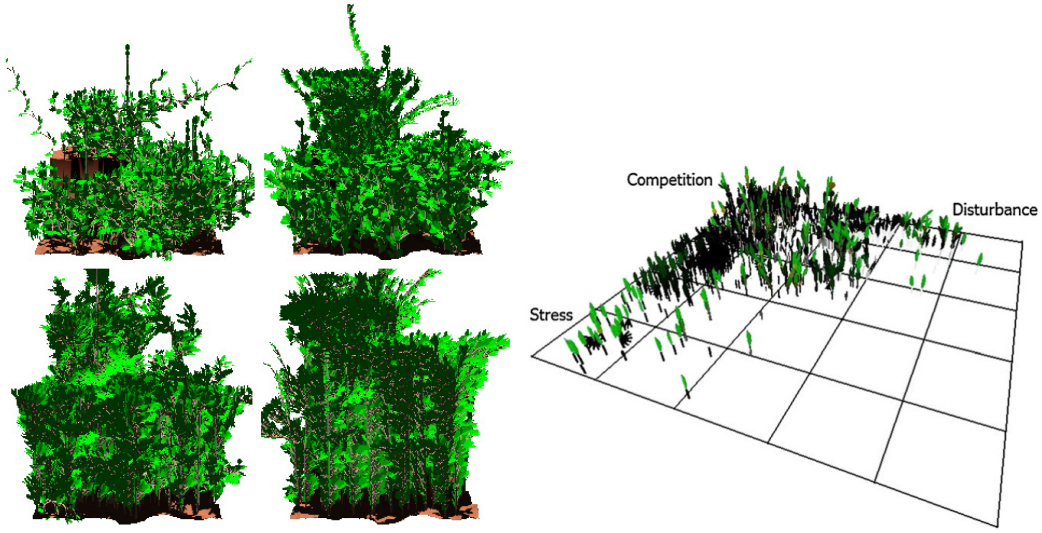
2.3.3 Vegetals

Naturally, more complex types of individuals have also been investigated in full-fledged ecosystems as in the case of the life-environment coupling of Daisyworld [Wood et al. 2008]. There flowers change their environment, not only by colonizing it but also by modifying the local temperature thus creating complex interaction feedbacks between biotic and abiotic components.

However, as was the case of digital and unicellular individuals, lack one crucial component of high-order life: morphology. One of the first instances of such a competitive evolution of individuals with morphological controllers as complex as those described in section 2.2 was introduced in [Ebner 2003]. As is often the case with plant morphogenesis, L-Systems were used to indirectly encode for each individuals' body plan. Reproduction is handled, externally, by the system with a fitness based on how much leaf surface is directly exposed to sunlight. Given that all plants are simultaneously cohabiting in the same 3D environment, this imposes a strong vertical competition, with smaller individuals quickly being removed from the population. A second portion of the fitness is concerned with minimizing the complexity of obtained morphologies by inducing a negative component which favors the use of branches over leaves. Such a configuration resulted in another instance of the Red Queen Effect with each generation requiring higher and higher heights in order to keep similar fitness levels. Indeed, plants were shown to grow up to the maximal theoretical height imposed (indirectly) by the fitness.

Similarly, the work of [Fernández et al. 2012a], also relied on external reproduction though the focus of this article was on the evolution of diversity in an infinite 2D ecosystem. In this case fitness F is also based on the amount of collected light l but the limiting factor is solely the number of branches f. The final expression $F = l/f^{\alpha}$ uses an external parameter α which conditions the harshness of the environment: the higher the value the harder it is for plants to produce large structures. As offsprings are placed near their parent, discoveries are local and multiple equally fit populations may co-exist. By experimenting under different values of harshness α two recurrent strategies were found: small, fast reproduction and large competitive morphologies which are similar dynamics to that of the ecological transition from newly colonized areas to old-growth forest.

A similar approach was undertaken in [Bornhofen et al. 2011], in which reproduction was left in the hands of the plants themselves. Thus the only fitness they were subjected was that of survival through propagation. The environment was also different with two sources of stress. Along one dimension minerals, one of the primary resources of these plants, were becoming increasingly infrequent



(a) Red Queen Effect [Ebner 2003] (b) CSR triangle [Bornhofen et al. 2011]

Figure 2.31: Convergence of virtual ecosystems on biological concepts

(down to 1/200 the initial concentration). On the other dimension "disturbance" events were increasingly more likely up to a probability of small plants patches being spontaneously destroyed. The interaction of these three components (self-reproduction, resource scarcity and extinction probability) led to the emergence of similar gradients of strategies as shown in figure 2.31b. In the upper, hospitable, corner plants are only in competition with each other and thus develop larger morphologies to outperform their neighbors. As resources levels drop, individuals become smaller and smaller, focusing their efforts into the fight with the environment while in the other corner they develop fast-reproducing strategies to quickly fill the gaps created by artificial destruction and prevent extinction. The last corner, too hostile for life, was left uncolonized. These dynamics were shown to be similar to the biological phenomenon of the CSR triangle[3].

An instance of vegetal ecosystem in which morphology is not controlled through Lindenmayer Systems can be found in [Eloy et al. 2017]. In this work trees, composed of a reserve, segments, foliages and seeds, are grown under the supervision of ANNs. Two such networks are used: one for the primary growth (new segments and seeds) and the other for secondary growth (increase in diameter). The plants extract resources from the environment by the photosynthesis process taking place in the foliages, which produces a photosynthate distributed to underlying segment and, if in sufficient quantities, stored in the reserve. This latter component is most useful when a given plant has to recover from major disturbance such as, in this work, strong wind

[3]Three common strategies are observed in plants, depending on their conditions: *Competition*, when resources are high and stress is low, *Stress-tolerant* when subjected to unpredictable variations and *Ruderals* when resources are scarce. [Grime 1977]

(a) Infinite 2D environment [Fernández et al. 2012a]

(b) Wind load and light sculpting [Eloy et al. 2017]

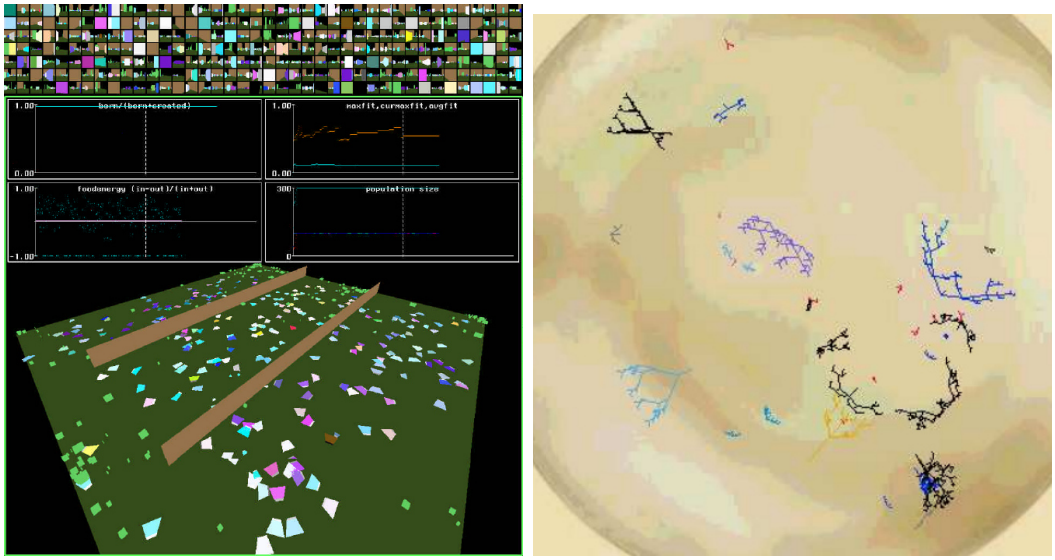
Figure 2.32: Self-organizing virtual forests

damages. In addition, starvation due to insufficient production of photosynthate can also lead to branch loss by a gradual weakening of its structure. Through a self-sustaining process in which plants reproduce asexually, the authors have modeled ecosystems of forest with notably life-like structure as the example tree of figure 2.32b clearly shows.

2.3.4 Animals

After studying point-sized individuals and immobile constructs with functional morphologies, the next logical step takes us to the exploration of animal-populated ecosystem. Amongst the first work addressing the co-evolution of motile individuals in a shared environment, is Polyworld [Yaeger 1994] in creatures are controlled through an Artificial Neural Network while morphology is limited to the size (which influences its storage capabilities). Seven predefined, high-level action can be chosen from, by the ANN, to determine, at a given timestep, the best course of action depending on its visual (ground-level colors) and internal inputs. In this work, the problematic of bootstrapping the system into a self-sustaining state is solved by use of a "fallback" steady state genetic algorithm which enforce a minimum population whenever required. This results, as shown in figure 2.33a, is fairly large population of cohabiting and competing artificial animals.

Similar work as also performed in [Gracias et al. 1997] with the same use of an SSGA as a way to kick-start autonomy. The ANNs used in this work were additionally empowered by Hebbian learning, though experiments showed that behavior was obtained through genetic evolution instead of online learning. Results additionally showed a positive feedback between environmental and



(a) Polyworld [Yaeger 1994] (b) Lifedrop [Metivier et al. 2002]

Figure 2.33: Animals with predefined high-level behaviors

behavioral complexity.

Combining both behavior and morphology was done in [Metivier et al. 2002]. Control is still performed by selection of high-level procedure including mating, fleeing or flocking. In this work, however, morphology is explicitly defined in a extension of Dawkins' Biomorphs (see figure 2.33b), presented earlier. Thus the evolutionary process works on both the body-brain pair when selecting individuals. In parallel, the authors designed a speciation model which, in response to varying stress levels, dilates or contracts species boundaries. Such a stress is generated by the pH of the environment which can range from 0 (hospitable) to 1 (terminally toxic). By raising the pH to .5 for small duration, a bottleneck effect was observed in which species count is drastically lowered. Thus the interaction of individual-centered evolution with external variation induced patterns similar to those of "punctuated equilibria" with rapid burst of diversification alternating with somewhat calmer periods.

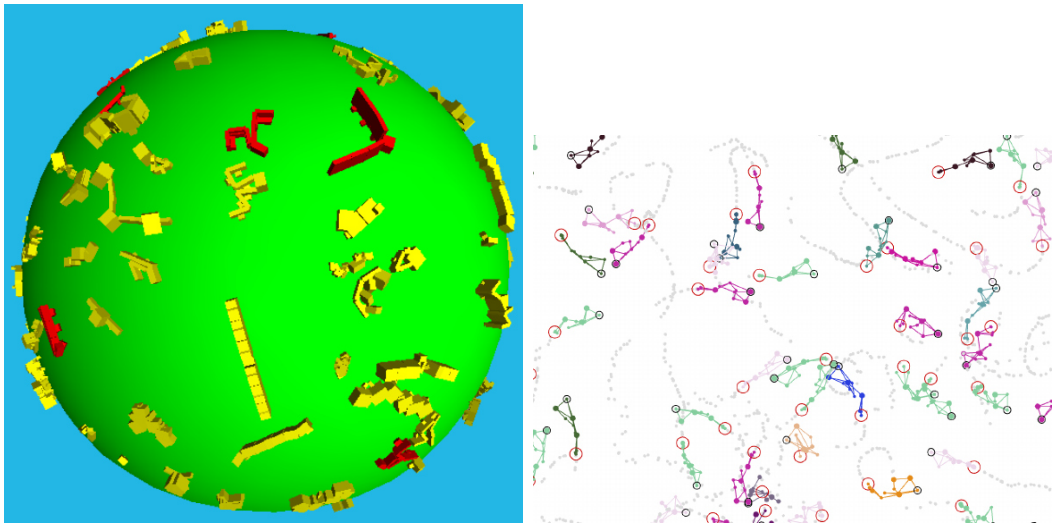
Morphology was further investigated in [Ventrella 2005] with the GenePool 2D ecosystem. In this work, individuals were composed of directly encoded parts connected to one another by hinge constraints. Motion was obtained by oscillation of these parts (according to a global, genetically controlled, clock) inside of a fluid medium. The central, essential, part of each of these creatures has a mouth at one end and a genital at the other: "the two goals in a swimbot's life". Control is performed by a state machine of high-level cognitive functions (concerned as advertised by food and mating) with the additional distinction of sexual preferences. When scanning its neighborhood for potential mates, a swimbot will evaluate each in turn based on its size, area or body pose. By this mechanism, sexual selection was allowed to emerge, though in a hand-written form, giving rise to a ecosystems populated by sitters (which do not move)

and breeders (which "transport" genes) when using the counter-intuitive "still" attraction criteria.

In [Pichler et al. 2008], the authors argue over the use of predefined high-level behavior: "such primitives are not necessary to evolve behavioral diversity even in a simple and homogeneous environment". To meet this objective, they used 2D creatures with a varying number of sensors and actuators. The former provided both internal and external information to the internal neural controller the complexity of which can be increased through mutation. The latter is responsible for motion by propulsing the agent forward. A minimalist structure is maintained to ensure that individuals have the (evolutionary) potential to survive: sensors for the current energy and reproductive depot and an actuator for the division process. In this model, individuals store, as given by the corresponding neural output, a certain amount of their energy in the reproductive depot which upon reaching a given threshold triggers an asexual reproductive behavior. These conditions were sufficient to reach a fairly high number of self-sustaining ecosystems (i.e. in which the SSGA was disabled). Using such low-level methodology still provided the experimenter with higher-level strategies such as the Drifters, Foragers, Avoiders or the more generalist All-rounder which discovered some of the fundamental primitives (obstacle avoidance, energy approach).

A three dimensional variation upon this work was undertaken in [Miconi 2008a] where the world was composed of sphere, the surface of which hosts competing creatures (figure 2.34a). The fundamentals of the developmental and interactions were laid down in [Miconi 2008b] which, as it has been described earlier, will not be further detailed. The salient point of this work lie in the potential for individuals to damage one another potentially to death. A reproduction mechanism was thus developed around this "interaction" in which fatally wounding another creature instantiates an asexual descendant of the "killer". The initial experiment resulting in a large dominance of the "roamer" strategy in which individuals would move as fast as possible to ensure being perceived as the aggressor. A second experiment was performed in which species barriers were artificially set up so as to prevent the emergence of a single dominating strategy. While the "roamer" strategy was still found in large number it also trigger a responsive counter: the "miniature". By diminishing the overall size of an individual it reduces the risks of being hit by a passing roamer (which rarely rely on sensory input and thus cannot change trajectory).

An alternative method for locomotion was experimented in [Turk 2010] which is computationally inexpensive, yet capable of complex motions: sticky feet. Creatures, in this work, are collection of point-masses connected by linear springs, undergoing periodic oscillation. In order to gain traction each point can have variable degree of friction, hence the name of "sticky" feet. Higher level control is obtained by closing a feedback loop between sensors on the segments and their properties (oscillation, stickiness). Each creature has one of its points designed as its heart and another as its mouth: the objective



(a) [Miconi 2008a] (b) [Turk 2010]

Figure 2.34: Individuals competing through shape and low-level actions

being to "eat" other creatures by placing one's mouth over another's heart. Such a capture is rewarded by triggering an asexual reproduction thus setting up a natural selection scheme favoring predation. Evolution in this system converged on reasonable regions of the genetic space with mouths being placed forward, on fast-oscillating segments, and hearts being protectively positioned at the rear. Additionally, hunting behavior did emerge with individuals actively seeking out other creatures' hearts.

Leaving aside morphological concerns, we note that late years have seen a resurgence of interest for boid-based swarm models which were introduced in [Reynolds 1987]. These creature were shown to exhibit plausible flocking behavior while only requiring three hard-coded rules: inter-individual avoidance, congregation and trajectory alignement. In [Harrington et al. 2017], such a model is used to test whether harsher environmental conditions, in terms of collision cost, results in an improved agent behavior. The metabolism of these boids is limited to an energy increase when colliding with a food source, a threshold for asexual reproduction and a loss of energy when colliding with another agent. Control is effected by GRNs, their outputs providing weigths for hard-coded rules (cohesion, separation, food attraction). Unbridled evolution over a fairly large sample of independent runs did produce more robust demeanor both in terms of agent avoidance and food collection.

In a similar approach, albeit with a different focus, boids are also used in [Witkowski et al. 2019] where ANN are evolved to react to communicative information diffused by other individuals. This information can be broadcast on two channels, which individuals scan, in all 6 cardinal directions. A feed-forward ANN receives these 12 inputs which are then processed and stored by 10 neurons layers, respectively the hidden and context. The output control the

Work	System	Ecosystem	Morphology	Continuous	**Heterogeneous**	**Dynamic**
Metivier et al. 2002	LifeDrop	X		X		**X**
Ebner 2003	-	X	X	X	**X**	
Lassabe et al. 2007	-			X	**X**	
Bornhofen et al. 2011	-	X	X	X	**X**	
Doursat et al. 2014	MapDevo3D		X	X	**X**	
Disset et al. 2014	SOMAS		X	X	**X**	
Ouannes et al. 2014	-	X		X		**X**
Aubert-Kato et al. 2015	-	X		X	**X**	
Canino-Koning et al. 2016	Avida	X				**X**
Chiba et al. 2017	-		X	X	**X**	
Eloy et al. 2017	MecaTree	X	X	X		**X**
Lalejini et al. 2017	Avida	X				**X**
Nahum et al. 2017	Avida	X			**X**	**X**
Zahadat et al. 2017a	VMC		X	X	**X**	
Luo et al. 2019	Avida	X				**X**

Table 2.1: Articles from this chapter with either heterogeneous or dynamical environments.

resulting rotation of the boid and eventually the emission of signal. From this architecture, swarming behavior were obtained by virtue of their facilitating resource access to a large population as well as cooperative behavior when one neural output determine the role played in an iterated Prisoner's Dilemna.

2.4 Summary and contribution of this work

From the articles presented throughout this chapter, we can gather one important piece of information: there is very little focus on the place of environmental constraints in the development of both isolated creatures and ecosystem. Summarized in table 2.1, is the subset of previously described work in which abiotic constraints have been used either to generate non-uniform or dynamical conditions.

The first noteworthy remark we can make is the relatively strong presence of ecosystems, even in the early stages. Out of the 16 articles that show marked environmental investment, only about one third is composed of solitary individuals. Indeed, in most cases, heterogeneous conditions are used to provide more complex, albeit localized, constraints on the capabilities of the evolved individuals. For instance, the stair climbing capacity obtained in [Lassabe et al.

2007] and [Doursat et al. 2014] directly derive from the use of tiered ground surfaces. Similarly, in [Zahadat et al. 2017a] the VMC (Vascular Morphogenesis Controller) is evolved in a maze-like environment to show its photosensitive capabilities.

Some other times the environment is less uniform than it first appears to be as in [Disset et al. 2014] where the creatures evolve the ability to flee towards the corner where toxin concentration is lower. Or some non-living component can be harnessed to increase one's survival capabilities as the defense-building individuals of [Chiba et al. 2017].

Nonetheless, varying environments are better leveraged by ecosystems for two reasons:

- Large population imply distribution over the local conditions, thus justifying their use.
- Large timescales allow for the use of dynamical properties of sufficient duration to be both effective and revertible.

The first point is well illustrated in [Ebner 2003] where different elevations are exploited by different portions of the population, thus accelerating the competition. Similarly, the use of two increasingly hostile dimensions in [Bornhofen et al. 2011] allowed for the emergence, in this same ecosystem, of radically different survival strategies.

One can note, however, the disproportionate presence of Digital ecosystem, especially with dynamical properties. Indeed given their lightweightness they are the most capable of study evolutionary trends on the scale of tens of thousands of generations. Thus, unsurprisingly, the only instance of both an heterogeneous and time-variant system comes from the Avida platform. Yet, this simplicity is also a drawback for amongst the examples presented here varying conditions boils down to either switching the palette of available operations or changing their associated rewards. In [Nahum et al. 2017], this is made more complex by introducing a shared resource consumption thus promoting self-regulatory dynamics at the population level.

With this in mind, we plan, in this work, to combine both morphological development with autonomous ecosystems evolution over evolutionary timescales. Figure 2.35 summarizes the main steps detailed throughout this manuscript. This is addressed, first, by investigating questions related to the self-sustaining properties of complex morphological models. In a second portion, a simpler and more efficient model is used to monitor the response to "blindly" changing environments. We develop an algorithm for coupling dynamical abiotic constraints with freely evolving populations which, when paired with our phylogenetic monitoring tool, allow for speciation tracking and study of long-term dynamics. The following chapter is thus devoted to the initial development of a developmental model with a broad range of capabilities.

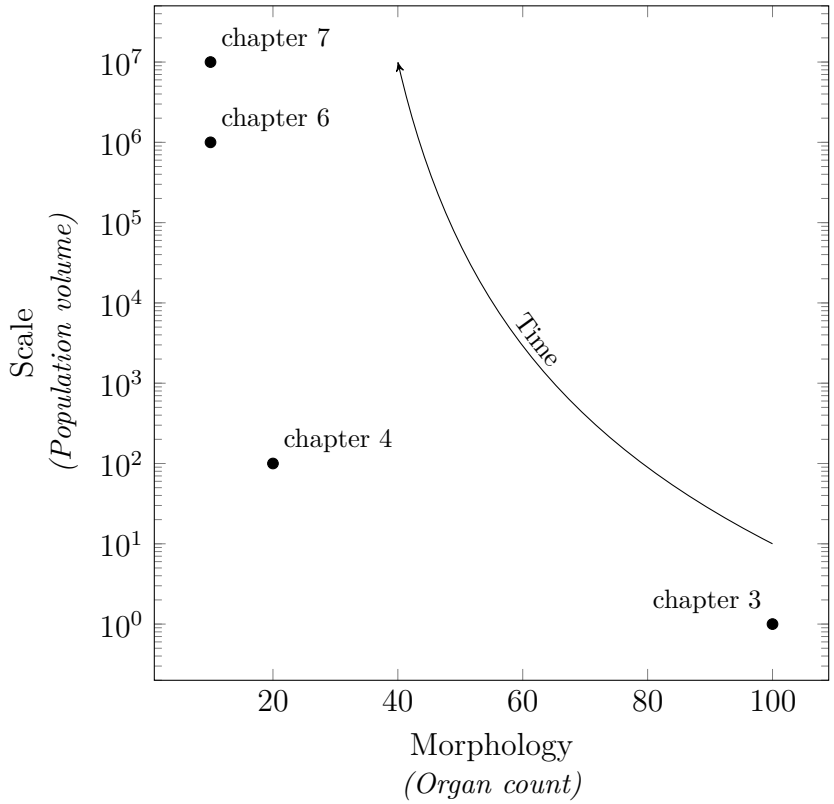


Figure 2.35: Distribution of this thesis' works. Population volume is the total number of individuals generated in a single simulation (counting autonomous reproductions but not parallel evaluations)

Chapter 3

Isolated evolutions

Abstract The first point we adressed in this work is that of artificial morphogenesis. To this end, we developed a generic model relying on genetic parameters as well as a physics engine. The former include a collection describing the plant's various organs alongside a morphological controller, responsible of the instantiation of the aforementioned organs. The latter provides a more complex frame of development as each intermediate state must be structurally sound. Resources extraction and growth are continuous and require robust strategies to cope the built-in variations in light and water availability. We observed that, amongst the two morphological controllers tested, the Graphtals showed more capable of producing diverse and plausible plants.

Résumé Nous nous sommes initiallement focalisés, lors de cette thèse, sur la morphogénèse artificielle. Dans cette optique, nous avons développé un modèle générique qui s'appuie aussi bien sur des paramètres génétiques que sur un moteur physique. Les premiers incluent un ensemble de données décrivant les divers organes des plantes et travaillent de concert avec un contrôleur morphologique qui instancie les organes susmentionnés. Le second fournit un environnement de développement plus complexe puisque chaque état intermédiaire se doit d'être structurellement stable. L'extraction de ressources et la croissance sont continues et requièrent donc des stratégies robustes pour faire face aux variations de lumière et d'eau. Nous avons ainsi observé que, parmi les deux contrôleurs morphologiques testés, les "Graphtals" se sont montrés mieux capables de produire des plantes crédibles et variées.

Our first concern in this work was the designing of a complex controller for morphogenetic engineering. To this end, this chapter is devoted to the specifics of a single individual's growth in a moderatly changing environment where water levels and light are heterogeneously present. A general model of an artificial plant's dynamics, partly inspired by the work of [Bornhofen et al. 2009], is introduced with two instantiations of morphogenetic paradigms.

The first was based on L-Systems, for their efficiency in producing believable plant-like structures has been well documented in the literature from [Prusinkiewicz et al. 1988] to [Fernández et al. 2012a]. They have also been used in the generation of morphologies for motile creatures as in [Hornby et al. 2001] which, given our longer term objective, validated them as a worthwhile starting point for our endeavor.

Considering initial results, detailed in section 3.3, of limited complexity and excessive linearity, we also devised another growth model based on the work of [Sims 1994b] on artificial animals. To the best of our knowledge, this was the first application of such a paradigm to plants, its capacity for modeling motile morphologies having been thoroughly researched (e.g. [Ito et al. 2013; Lassabe et al. 2007; Lessin et al. 2015; Miconi 2008b]) after this seminal publication. Details of this implementation of directed graph (hereafter named 'Graphtals') and the resulting morphologies are spelled out in section 3.4.

Beforehand, we give a general account of the common parts of this framework. We start, in section 3.1, by describing the environment in which those artificial plants are embedded through its major components: a dynamical light source loosely mimicking earth solar patterns, uneven water resources with time-dependent variations and potential (static) competition. We then detail the algorithm controlling the life-time of an individual (section 3.2) starting from its germination, growth, homeostasis and eventual death.

We conclude this section on isolated evolution by highlighting the most successful aspects of our approach, in this limited setting, while also drawing lessons on the improvements required to reach the more ambitious goal of long-term evolution.

3.1 Environmental model

3.1.1 Light

Of utmost importance for plant growth, because of its role in the photosynthesis of glucose by plants, the first component of our environmental model is the light source. Its polar position (R, θ, ϕ) is governed by the parameters enumerated in table 3.1 and updated as follows:

$$\theta(t+1) = \theta(t) + \frac{2\pi}{D} \tag{3.1}$$

$$\phi(t+1) = \phi(t) + a * \frac{2(\phi_s - \phi_w)}{Y} \tag{3.2}$$

	Variable	*Value*
R	Distance	100
D	Day length	100
Y	Year length	300
ϕ_s	Max altitude	$3\pi/8$
ϕ_w	Min altitude	$\pi/8$

Table 3.1: Dynamic light parameters

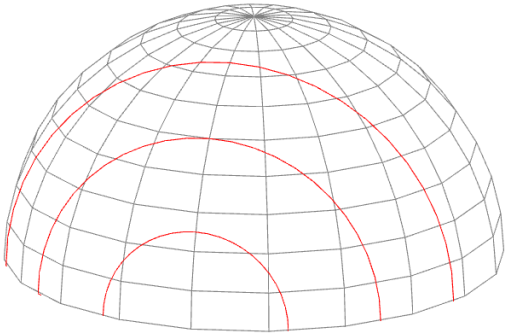


Figure 3.1: Typical sun trajectory

The daily rotation is performed, every simulation step, according to equation 3.1, describing a circular trajectory around the $\vec{z}$ axis at an altitude given by ϕ. This, in turn, is updated after every simulated day, as shown by equation 3.2, to alter the light's angle of incidence. The parameter a indicates whether the trajectory performed corresponds to the "winter" ($a = 1$) or "summer" ($a = -1$) season. The values used throughout this chapter are design to reproduce the seasonal pattern of a mid-hemisphere plot on earth.

We relied on these varying conditions to stimulate plant morphologies into producing more robust solutions than if using a stationary light source. Indeed, leaf placement is much less straightforward due to the daily motion of the sun but even more so because of its seasonality: leaves that performed well in summer might obstruct one another when sunlight comes from closer to the $\vec{x}\vec{z}$ plane. The specifics of how the plants interact with this evasive sun will be discussed in the corresponding section (3.2.1).

3.1.2 Water

The other variable of the abiotic component of our simulation is the water availability. To promote expansionist strategies and intra-root competition, we model it through a voxelisation of the ground portion as schematized in figure 3.2.

Initially, all voxels are empty until rainfall starts to accumulate on the surface. Those voxels are those that can collect rainwater, acting as the sources of the system. The below-ground portion is divided in as many parts, horizontally, and further partitioned in layers each being able to contain an amount of water proportional to its depth as formalised by equation 3.3. One should note that the bottom most-layer acts as a water table with a saturation of $1000L.m^{-3}$ indicating that every voxel is at most a cube of pure water.

$$S_j = \begin{cases} 2(j+1)/H & \text{if } j < H-1 \\ 1000 & \text{otherwise} \end{cases} \tag{3.3}$$

Surface voxels are not limited in the amount they can store but are, instead,

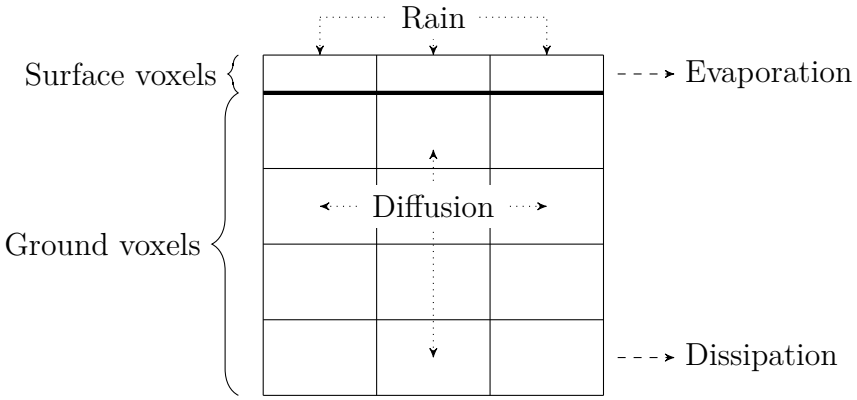


Figure 3.2: Water dynamics

Variable	Description	Value
R	Rain strength	(see text)
W, H, D	Ground dimensions	$20, 10, 20$
S_j	Saturation at depth j	(see eq 3.3)
V_i	Water at surface voxel i	$V_i \in \mathbb{R}^+$
$V_{i,j}$	Water stored in voxel i at depth j	$V_{i,j} \in [0, S_j]$
k_a, k_d, k_v	Diffusion rates	$.003, .03, .006$
d_s, d_b	Dissipation rates	$.025, .01$

Table 3.2: Water parameters

subjected to evaporation. Table 3.2 and equations 3.4-3.9 detail the parameters and interactions controlling the water's dynamics.

First, rainwater is diffused from the surface voxels to those directly beneath:

$$\frac{dV_{i,0}}{dt} = \min(k_a V_i, S_0 - V_{i,0}) \tag{3.4}$$

$$\frac{dV_i}{dt} = -\frac{dV_{i,0}}{dt} \tag{3.5}$$

A global diffusion is then enacted between every voxel of the ground portion according to equation 3.6, where N is the Von-Neumann neighborhood of a given voxel (top, bottom, backward, forward, left and right).

$$\frac{dV_{i,j}}{dt} = \max(S_j - V_{i,j}, 0) \sum_{(i',j') \in N} k_d(V_{i+i',j+j'} - V_{i,j})) \tag{3.6}$$

We complement this passive diffusion by increasing the amount exchanged in the downward direction, thus simulating a coarse gravity.

$$\frac{dV_{i,j}}{dt} = k_v(\min(V_{i,j-1}, S_j - V_{i,j}) - \min(V_{i,j}, S_{j+1} - V_{i,j+1})) \tag{3.7}$$

Finally, both the surface and bottom-most voxels are subjected to a proportional evaporation, the rate of which being constant for the latter. In case of the surface voxels, however, evaporation is dependant on the current rain pattern: given humid above-ground conditions the rate will be severely diminished when compare to drier settings.

$$\frac{dV_i}{dt} = \frac{d_s}{R} V_i \tag{3.8}$$

$$\frac{dV_{i,D-1}}{dt} = d_b V_{i,D-1} \tag{3.9}$$

The rain pattern is thus a crucial component of this system the specific implementation of which will be detailed later on, in the experimental section.

3.2 Common model

Despite using a couple of growth models for our artificial plants, a large common ground can be found in the algorithms controlling both metabolism and growth patterns. In this section, we detail such similarities while the specifics of each implementation will be elaborated upon in their corresponding sections.

3.2.1 Metabolism

From a metabolic point of view, plants are a collection of organs with a given shape and dimensions. Depending on their associated skills, they will serve various roles in the structure: water uptake, photosynthesis, structural resistance, etc. One of those most useful functions is that of resource extraction: indeed survival implies the ability to maintain internal reserves at comfortable levels throughout their whole simulated life.

Plants are initialized from a single seed, a spherical organ associated with the skill `RESERVE`, which provides the initial resources. The first two organs are also spawned at this time, one for each layer, so that they may serve as the starting point of the budding process (described in sections 3.3.2 and 3.4).

From this point, plants are updated according to the algorithm 1, which describes the common procedures for stepping a plant. With the exception of line 9, which depends on the growth model, details of the various equations will be provided in the following sections with table 3.3 referencing the numerous parameters used with, when applicable, their values.

"Store" light

The simulation step starts by querying for the available amount of light for each leaf l of the plant. In practice, this corresponds to performing a raycast between the center of mass of l and the current sun position. Though not capable of finer detection such as partial shadowing or opacity, this was deemed

Symbol	Semantic	Value	
		L-Systems	Graphtals
	Constants		
$\vec{E}$	Elements	{Water,Glucose}	
K_W	Water consumption	6	1
P_{spd}	Photosynthesis speed	.025	
P_{sat}	Photosynthesis saturation	.025	
$\vec{S}^W_{trv}$	Starvation speed (water)	$1e7$	$3e2$
$\vec{S}^G_{trv}$	Starvation speed (glucose)	$1e7$	$2e2$
	Genome-wide		
$\mathfrak{G}$	Growth speed	$[1e-3, 1]$	$[1e-3, 1e-1]$
	Organ-specific genetic fields		
$\vec{\mathfrak{A}}_o$	Allocation	$[0,1]^{\lvert E\rvert}$	
$\vec{\mathfrak{S}}_o$	Survival	$[0,1]^{\lvert E\rvert}$	
$\vec{\mathfrak{D}}_o$	Dimensions	$[.001, 1]^3$	
$\vec{\mathfrak{G}}_o$	Local growth speed	$[0,1]^3$	
$\mathfrak{G}^m_o$	Maximal local growth	$[1, 1000]$	$[1, 10000]$
ρ_o	Density	$[.001, 10]$	$[.01, 5]$
α_o	Adhesion strength	$[.001, 5]$	
	Organ-specific derived fields		
V_o	Volume	(shape dependent)	
M_o	Mass	$V_o\rho_o$	
S_o	Surface		
$\vec{p}_o$	position		
$\vec{v}^z_o$	Up vector	(of local system)	
$\vec{v}^S_o$	Sun's direction		
S^s_o	Photosensitive surface		
l_o	Perceived light		
s^e_o	stored amount	of element e	
m^e_o	maximal amount	V_o	
r^e_o	requested amount	$\vec{\mathfrak{A}}_o[e](m^e_o - s^e_o)$	
a^e_o	available amount	$s^e_o - \vec{\mathfrak{A}}_o[e]m^e_o$	
t^e_o	excess amount	$s^e_o - \vec{\mathfrak{S}}_o[e]m^e_o$	

Table 3.3: Common parameters of a plant's metabolism

```
   Data: e, environment
 1 for p, plant do
 2 |   p.storeLight(e);
 3 |   p.storeWater(e);
 4 |   p.transport();
 5 |   p.manageSeed();
 6 |   p.photosynthesize();
 7 |   p.consumeResources();
 8 |   p.processOrganDeath();
 9 |   p.produceOrgans(e) ;                    // Model specific
10 |   p.growOrgans();
11 |   p.updateConstraints(e);
12 end
13 e.step();
```

Algorithm 1: Generic algorithm of a simulation step

an acceptable compromise considering the objective of extending this model to multiple individuals.

Whenever the ray does succeed in connecting both the leaf and the sun, the actual amount of light gathered is given by:

$$l_l = max(0, \vec{v}_l^S \cdot \vec{v}_l^z) p_l^s \tag{3.10}$$

That is, we consider the photosensitive surface of area p_l^s of the leaf to be perpendicular to the up vector of the organ, in its local coordinate system. In this manner, maximizing light exposition need not only large dimensions along the $\vec{x}$ and $\vec{y}$ components but also a correct position with respect to the sun. Given that this target keeps on moving throughout the simulation, this is intended as an incentive to produce robust leaves distribution strategies.

Water uptake

The next action performed at the beginning of a plant's simulation step is the extraction of water from the ground voxels. Indeed, only those deeper than the surface can be "mined" with the depth-dependent storage capacities allowing for heavily consuming plants to find the necessary reserves by growing deeper root networks. Both extreme morphologies could thus emerge from this model: shallow, rain-synchronized water extraction or deep water-table mining.

The actual uptake equation is given by:

$$\frac{ds_r^W}{dt} = \min(water(p_r)S_o, m_r^W - s_r^W) \tag{3.11}$$

The whole surface of the root is taken into consideration and absorbs from the voxel situated around the organ's mean position p_r. Necessarily, this is

bounded by the actual storage capacity left inside the root itself. Additionally, this uptake impacts the water availability of the corresponding voxel in an inverse manner thus preventing access to infinite resources.

Given the fact that evaporation is already taken into account at the level of the environment and that, in most cases, water uptake is largely negligible with respect to the volume of water stored in each voxel this dynamic was shown to have little effect on the artificial plants' dynamics.

Transport

With newly stored water available in the root sections of the plants and light accessibility computed for each leaf, all that stands before glucose production is to actually transport this water up to the leaves. The set of equations below details the process of resources transportation through a plant-centered approach:

$$a^e = \sum_o a_o^e \tag{3.12a}$$

$$r^e = \sum_o r_o^e \tag{3.12b}$$

$$k^e = min(\frac{r^e}{a^e}, 1) \tag{3.12c}$$

$$\frac{ds_o^e}{dt} = \frac{a^e r_o^e}{r^e} - k^e a_o^e \tag{3.12}$$

We start by computing a^e and r^e the total amount of resource e available and requested, respectively. These aggregated values are of importance when considering the genomic field they rely upon: $\vec{\mathfrak{A}}_o$. The first effect of this field is to restrict the amount of resources an organ is ready to dispatch to others considering its own internal reserves: a maximal value of 1 would describe a sink such as a flower or a fruit while, at the other extreme, a value of 0 would produce a "pipe" through which resources flow. These values are independent for each of the two elements considered in this work. Additionally, $\vec{\mathfrak{A}}_o$ also has an effect on the requested amount of nutrients by weighting the urgency of receiving said resources. Through this field, organs can thus tailor their tendency towards consumption or production.

We then define the diffusion coefficient k which when $a^e < r^e$ is simply equal to 1. In the other case, k is reduced so that only the requested amount of nutrient flows throughout the plant which would help promote ruderal behavior by honoring the values of $\vec{\mathfrak{A}}_o$.

Each organ will thus receive a portion of the globally available reserve, the size of which will depend on the weight of its individual value of r_o^e with respect to the total requested amount. At the same time, those organs whose reserves exceed the allocation threshold will see them reduced by that same amount, thereby sharing their resources surplus with more demanding relatives.

Seed management

The seed is the first of all organs to be created and its objective is to provide the plant with the initial resources to generate a self-sufficient morphology. As such it is consumed at the beginning of the individual's lifetime with its size decreasing with respect to the fullness of its reserves as defined by:

$$scale_s = \frac{\sqrt[3]{\frac{3 \max_{e \in E} s_s^e}{4\pi}}}{\vec{\mathfrak{D}}_s[1]} \tag{3.13}$$

This ensures that the seed's storage capacity (volume) is just large enough to contain the largest of its current nutrient reserve. Upon exhausting any of its resources the seed is removed from the plant and the remaining organs are updated to reflect this change.

Photosynthesis

One of the core components of the plant's metabolism is the ability to transform water, extracted by the roots, into glucose inside those leaves that are directly exposed to "sunlight". This is performed in a centralized manner similar to that used for transportation, as detailed by the following set of equations.

$$t^W = \sum_o t_o^W \tag{3.14a}$$

$$l = \sum_l l_l \tag{3.14b}$$

$$\frac{ds^G}{dt} = \min(P_{sat} l, \sum_l \frac{P_{spd} l_l t^W}{K_W l}) \tag{3.14c}$$

$$\frac{ds_o^W}{dt} = \frac{-K_W \frac{ds^G}{dt} t_o^W}{t^W} \tag{3.14d}$$

$$\frac{ds_o^G}{dt} = \frac{\frac{ds^G}{dt}(m_o^G - s_o^G)}{\sum_o m_o^G - s_o^G} \tag{3.14e}$$

We start by determining how much water is available (t^W) by requesting from each organ just how much excess they contain. This rely on the other key genetic field controlling the metabolism: $\vec{\mathfrak{S}}$. By controlling the survival threshold of individual organs, the plant can decide which of its portions is the most critical to its continued welfare and where to extract resources from. Indeed, as specified in table 3.3, these parameters influence the amount of nutrients an organ considers as vital for its own survival, refusing to indulge in any activity (other than resource reception) that might otherwise deplete its stores.

In the same way, we compute the total photosensitive area l of the plant directly exposed to sunlight at the current time step which allows for the

computation of the total glucose production. The first part of the equation 3.14c is a limiting parameter that prevents extremely efficient individuals to produce an excessive amount of glucose at once while the second part performs the actual computation.

Organs stores are then updated: consumed water is removed, according to the organ's participation capabilities, and newly produced glucose is stored proportionally to the available storage capacities.

Resources consumption

Once every organ has been given a chance to participate in the extraction and distribution of resources, only consumption remains. This is modeled as a form of turnover of old tissues, though biomass itself is not an explicit variable. Instead, we consider each resource storage and straightforwardly reduce it, for each organ, with respect to its current volume and skill according to:

$$\frac{ds_o^e}{dt} = \frac{V_o}{S_{trv}^e} \text{ cost}(o) \tag{3.15}$$

Where cost(o) is defined in table 3.4. Through this consumption, organs with low incoming resources flow might see their storage levels reach zero or less, in which case they are disposed of.

Organ death

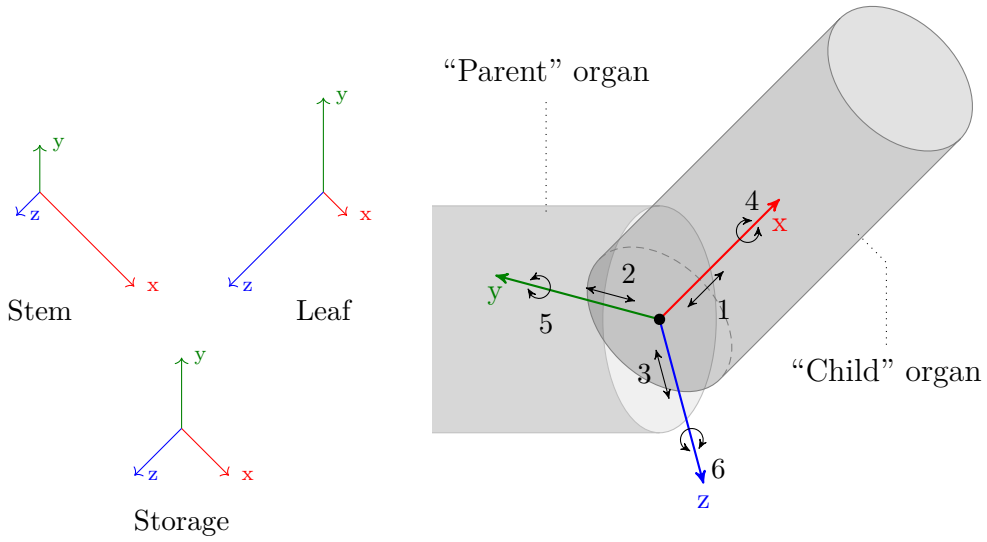
By collecting the set $D = \{o, \text{organ}/\exists e \in E/s_o^e \leq 0\}$ of organs with insufficient resource levels, we can determine which parts of the plant are dead at the current time step. These organs are then deleted from the individual along with all their "children", i.e. the organs that have been created along its surface. The remaining resources stored in these organs, those found starving as well as those below disconnection points, are also destroyed, the implementation of an ecosystem with a closed loop of nutrients being left as an extension of this work.

3.2.2 Growth

The other key aspect of the metabolic dynamics of our artificial plants are the individual growth of its various organs which, starting with humble dimensions, can reach much larger sizes at maturity. This rely mainly on the two genetic parameters controlling growth: $\vec{\mathfrak{G}}_o$ which defines the local growth factor of an organ and $\mathfrak{G}$ which imposes a global scaling on all local values.

Anisotropic growth

At the end of the stepping procedure of a plant, each of its organs has the opportunity to grow in size if two constraints are met. First, the organ must



(a) Examples of anisotropic growth (b) Fixed 6DoF constraints

not be in a state of survival, i.e. all of its nutrient stores must be above the surviving threshold defined by $\vec{\mathfrak{S}}_o$. In other words, only those organs for which $\forall e \in E, t_o^e \geq 0$ are considered.

To be qualified for growth, an organ must also be under the maximal factor allowed for its shape. That is, for a local scaling of $scale_o[1]$ in the $\vec{x}$ dimension, we maintain $scale_o[1] \leq \mathfrak{G}_o^m$.

In case both constraints hold, the growth variation itself is computed through the following straightforward equation:

$$\frac{dscale_o}{dt} = \mathfrak{G}\vec{\mathfrak{G}}_o \tag{3.16}$$

It is worthy of notice, though, that the local growth component is a 3D vector thus allowing for non-uniform rates depending on the dimension, effectively resulting in an anisotropic growth. As illustrated in figure 3.3a, this allows for the specialisation of growth pattern according to the functionality of the considered organ. For instance, leaves would gain from increasing their photosensitive surface while a uniform growth would be sufficient for storage organs.

Given that the initial dimensions and growth related components are managed independently from one another, the range of final shapes that can be generated is extremely high.

Physical constraints

The final procedure, performed during a single plant's step, is the management of the physical constraints between pairs of organs. We relied on the physics engine Bullet designed by [Coumans et al. 2013] to provide ready-made physical constraints as well as the raycast performed during the first step (light collection). Indeed, our initial goals were to use this engine to model fine-grained animats.

Even though, during the course of this work, we focused solely on plant evolution, it still allowed for physically inspired morphologies although with some limitations (gravity, for instance, is not enabled).

We used fixed 6DoF (Degrees of Freedom) constraints to join the various components of our plants into a network of interacting physical objects. Figure 3.3b illustrates the six dimensions locked into place by such constraints. Given $\vec{p}_A$ and $\vec{p}_B$ the connection point on the first and second object, respectively, the local coordinates system at the points is actively maintained perfectly aligned. Indeed constraints 1 to 3 represent the translation between the origins of those two systems which are kept null. Similarly, the next 3 constraints correspond to the rotation around a given axis which is also kept null.

Through Bullet, we used these constraints at the plant's initialisation to anchor it to the ground, through its seed, first shoot and root organs. These are also used to tie pairs of organs connected by a parent-child relationship. Given that they aim at preventing all forms of motion, these constraints are symmetrical in their action, but they are not, however, in terms of definition. The frame of reference, i.e. the coordinates system, that is maintained through the use of these 6DoF constraints is based on that of the second object: the organ connect to the ground or the "child".

Additionally, given that most morphological operators tend to produce closely positioned "sibling" organs (i.e. organs connected to the same "parent"), especially in the case of L-Systems, we do not perform collisions detection between them. All other organs, however, are subjected to this avoidance algorithm so as to prevent interpenetrating morphologies. Due to their being embedded in the ground, the inertia of root organs is deactivated.

This complex interaction of morphology and physical constraints leads to the apparition of potentially large forces between parts of the plants. Naturally, physics engines devised with characters modelling are, on some level, ill-suited to the management of such use cases. There are two methodologies to tackle such a dilemna: using a physic engine with a built-in focus on growth-related constraints (such as continuous repulsion or adhesion) or, as was selected in this work, delegating the responsability of stability to the genome. Thus, the morphogenetic process must also take this parameter into account to ensure that growth can occur as best as possible. A task made harder by the fact that the 6DoF constraints used to connect organs to one another do not have an infinite resolution capacity.

Indeed, to maintain both frames of reference aligned, Bullet performs a number of small-step resolutions to gradually return to the state defined as equilibrium. However, when confronted with a larger number of constraints, some of which potentially contradictory with others, finding a suitable stable point might be difficult if not impossible.

We detect such cases by computing, regularly, whether the difference between the theoretical and actual coordinate system deviate from one another in too large a fashion. Details of this implementation will not be discussed here.

Skill	Shape	Function	Cost
PHOTOSYNTHESIS	Box	Glucose producer (leaf)	1.5
REPRODUCTION	Sphere	(unused)	1.25
RESERVE	Sphere	Resource storage (bulb/seed)	.5
STRUCTURAL	Capsule	Load-bearer (stem)	.75
BURROWING	Capsule	Water extraction (root)	.625

Table 3.4: L-System skill set and associated shapes

The other case where constraints may be found wanting is when the impulsion required to realign both frames of reference is above a given threshold. This is under control of yet another genetic parameter: α_o. Through this field, organs can decide how much they tend to "stick" to others thereby forcing crucial parts of the anatomy to maintain coherence despite their conditions or, on the contrary, allowing less important portions to be shed as required. The actual threshold is given by equation 3.17 where we can see that larger objects have a larger adhesion strength which compensates for their heavier mass (and thus heavier physical response).

$$I_{AB} = min(\alpha_A M_A, \alpha_B M_B) \tag{3.17}$$

Once a constraint has been found as no longer viable, it is broken and the child organ to which it is connected to is deleted from the plant as previously described in the case of starvation. Thus a good morphogenetic plan must take into consideration not only the final shape of the plant but also the load distribution and growth dynamics in order to maximize its efficiency and limit potentially harmful break events.

3.3 L-Systems

The developmental model used in this first part of our experiments on morphogenesis relies on context-free, deterministic L-Systems. As mentioned in the opening paragraphs of this chapter, this formalism has already shown its proficiency in producing plausible plant morphologies. As in [Bornhofen et al. 2009], each symbol of the rule-set is associated with a number of variables most of which having already been mentioned in the previous sections.

The unaddressed question of what purpose each organ type serves is resolved by a dedicated field: the skill. After preliminary experiments, we decided to settle for a single skill per organ considering that morphologies tended to degenerate into single omnipotent organs. This skill, in turn, determines the shape the organ will assume as enumerated in table 3.4.

Additionally, a 3D vector in $[0, 1]^3$ is used to control the color of the corresponding organs with no impact of the behavior of the plant. This field only serves an aesthetic purpose and will, thus, not be further detailed.

Such a repository of symbol-to-organ is built for each portion of the plant's genome: one for the surface (shoot) and one for the underground (root). A number of portion-level parameters provide the additional information required to perform the morphogenetic process. Alongside the global growth scaling $\mathfrak{G}$, discussed previously, is the rotation angle $\in [\pi/12, \pi/3]$, the maximal derivations count $\in [0, 5]$, the growth period $\in \mathbb{R}^+$ and the rule-set.

3.3.1 Initialization and mutations

Randomly generated genomes contain, for each section, a minimalist rule-set barely capable of producing viable morphologies. For the shoot portion, this corresponds to the axiom `A[+a]` and the single rule `A -> [A+a]` with `A` a structural organ and `a` a leaf. The root rule-set is even smaller with only the axiom `A[-A][+A]` being defined where A is a root organ. All other genetic fields are randomly assigned in a subset of the ranges previously described.

The matter of mutating a genome boils down to point-wise mutation which can be classified into three categories depending on the type of the value mutated:

Integers the new value is an increment or decrement of the previous

Decimals a gaussian noise is added with a mean of zero and a standard deviation equal to $(M - m)/10$ with m and M being the lowest and highest allowed values, respectively.

Complex requires a more involved algorithm as in the case of the rule-set

Indeed, mutating the shapes repository is as simple as altering a single value inside a randomly selected shape, while introducing variation into the L-Systems requires a more elaborate procedure. The actions that can be performed at the set-level are addition, duplication and suppression of a complete rule. The former two are limited by a maximal number of allowed non-terminal (3 throughout these experiments). In both cases, an unused non-terminal is assigned to the corresponding derivation (null in case of an addition and the content of the duplicated rule otherwise) with a random minimal scale σ. If the selected non-terminal is not associated with a shape, then a random one is generated. The case of suppression is even more straightforward and self explanatory, with the exception that the corresponding shape is kept in the repository thus allowing further reuse and genetic drift.

Mutating a single rule is a more cumbersome affair with five different possible actions: suppression, duplication, replacement, permutation and insertion. Non-terminal symbols are always encapsulated between branching control characters and the maximal number of different terminal symbols is, as for the non-terminals, set to 3.

Given the preliminary state of these experiments, we did not devise a crossover procedure neither for the L-System nor the shapes repository.

3.3.2 Morphogenesis

Obtaining a morphology from a L-System and an axiom is not a new concept and will thus only be described in passing. The most important feature of this development is that the process of organ creation is a local one, unlike its traditional generative approach. The objective in this work being not only to produce interesting, and functional, shapes but to actually make them grow from a single seed. Thus in a similar fashion to the one used by [Bornhofen et al. 2009], we consider the organs associated with non-terminal symbols as the apexes of our plant, i.e. those points from which more organs can be generated.

In order to put an upper bound on the complexity (and computational cost) of the generated individuals, rules can only be applied a fixed number of time, the value of which being controlled by a genetic parameter (max derivation count, previously mentioned). An organ o associated with the non-terminal symbol S will only be derived into the collection of organs specified by the corresponding rule if two conditions are met. Not only must o be in a healthy situation (i.e. not starving, as described in the section on growth) but it must also have increased in size up to a rule-dependent factor $minScale \in [1, 3]$. The first constraint serves as a regulating process, allowing the plant to stop further growth in case of excessively unfavourable external conditions. The second emulate the growth pattern of natural plants where complexification of the morphology is dependent on both time and internal resource levels.

The derivation process itself simply consists of parsing the rule's successor, generating organs for each (non-)terminal encountered and processing rotations and branching instructions. An overview of the interactions between the different components of this system is shown in figure 3.4

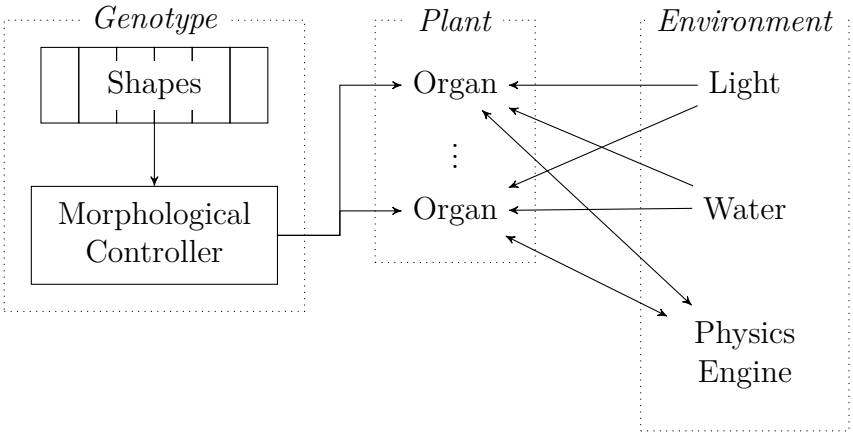


Figure 3.4: Interactions between the different components of the system

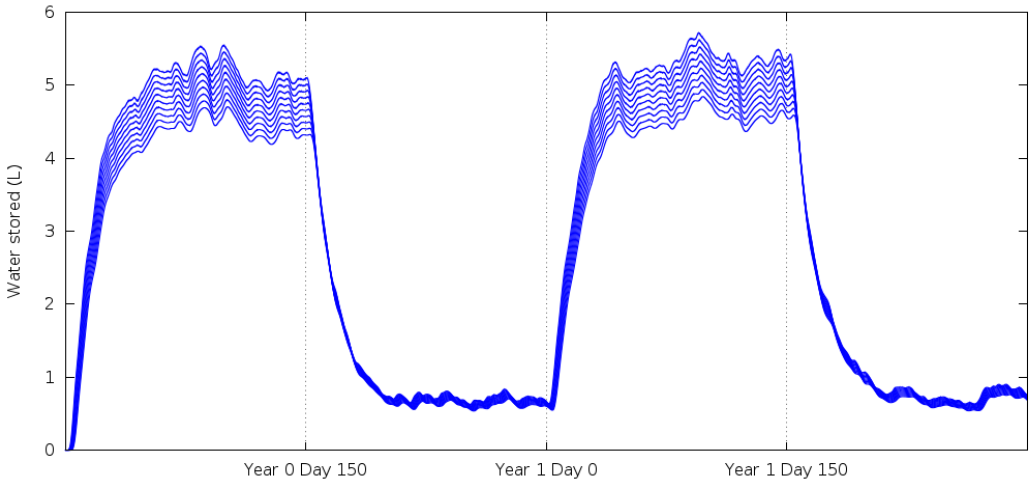


Figure 3.5: Water levels at each depth throughout the 2 years

3.3.3 Experimental protocol

This first experiment was designed to test capabilities of this model to produce functional morphologies in the face of mildly hostile external conditions. To this end, the environment was instantiated with the light pattern described previously and a semi-random seasonal rain pattern as defined by the equations 3.18 and 3.19.

$$P(S|Y) = \begin{cases} .1 & \text{if } Y < .5 \\ .01 & \text{otherwise} \end{cases} \tag{3.18}$$

$$R(Y) = \begin{cases} \mathcal{U}(.375, 1.5) & \text{if } \mathcal{B}(P(S|Y)) \\ 0 & \text{otherwise} \end{cases} \tag{3.19}$$

In these S denotes the event "raining" which is more likely to occur in the first half of each year. Given a coin toss, i.e. a test on a Bernoulli distribution with a probability $P(S|Y)$ of success, a positive result implies rain occurring at the current time step, the amount being obtained by an uniform distribution.

In this experimental setting, we simulated 2 years of survival, which is equivalent to $N = 60000$ simulation steps in a square 20 meters wide environment, 10 meters deep. The resulting dynamics of the water reserves, shown in figure 3.5, exhibit a seasonal cycle as induced by the rain equations. Each line represents the cumulative water level, i.e. that of the voxel itself and all underneath, and we can see that, starting from a dry environment, the simulations enter a "wet season" which lasts half the year. Evaporation, dissipation and reduced rain amount induce a drastic change in available water levels over the course of the second half of each year resulting in a "dry season". We call this pattern semi-random because, from the standpoint of the plant, it is not possible to determine rain amount at a given time step. From an experimental point of

view, however, each plant will be subjected to the exact same pattern to allow for comparison.

In terms of evolutionary process, we subjected a population of 200 random individuals for a duration of 50 generations to an evolutionary algorithm, i.e. individuals were only mutated, never crossed. We relied on the GAGA[1] library, developed in our team, to perform the actual evolutions with a pareto tournament selection of size 3. We used two fitnesses to guide the process: the first promotes glucose production (eq. 3.20) and the second is a novelty metric relying on a number of behavioral descriptions (equations 3.21a-3.21c) to characterise original solutions.

$$F_G = \frac{2}{N(N-1)} \sum_{i=1}^{N} iG_i \tag{3.20}$$

$$norm(l, x, h) = max(l, min(x, h))$$

$$N_S = \frac{S}{N} \tag{3.21a}$$

$$N_G = norm(0, F_G, 1) \tag{3.21b}$$

$$N_{S_j}^L = norm(-1, AABB(j, L), 1) \tag{3.21c}$$

With G_i the amount of glucose produced at time step i, S the number of time steps the plant survived for and $AABB(j, L)$ the size of the bounding box occupied by the layer L of the plant (root or shoot) in the j^{th} dimension.

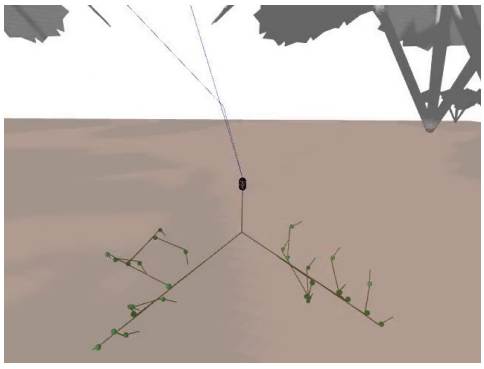
Initial results mostly produced individuals that either degenerated into quiescent minimalist organism or very small morphologies. The former only relied on the initial resources of the seed for survival while the latter developed a limited number of organs, though enough to ensure their survival. To alleviate this problem and stimulate the emergence of more complex morphologies we introduced artificial plants into the environment. These are hard-coded, static members of the environment and only influence the evaluated plant on one point: light availability.

3.3.4 Obtained morphologies

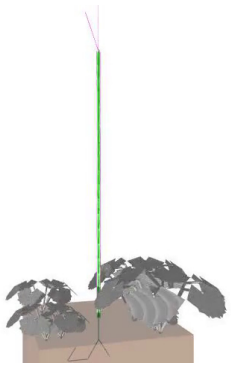
The presence of such artificially induced vertical competition did result in a better average complexity with individuals striving to reach higher altitudes. A small sample of obtained morphologies is displayed in figure 3.6 with the top row showing snapshots of the growth process. At the bottom these instances characterise the typical distribution of shapes observed during this experiment: either very thin with recursive rules or excessively linear, as a by-product of the fitness.

[1] https://github.com/jdisset/gaga

(a) Continuous Growth with rule derivation



(b) Storage organs



(c) Massively vertical

Figure 3.6: Sample of morphologies. Top row: the growth process with the successive application of L-System rules. Bottom row: typical examples of the morphologies obtained with strongly linear body plans and extremely thin organs.

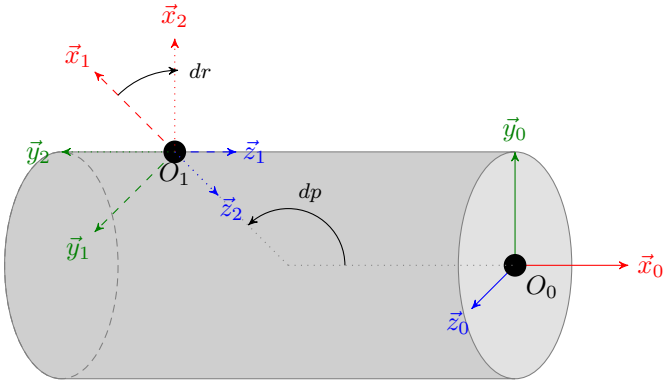


Figure 3.7: Effect of the two rotational parameters *dp* and *dr*

3.4 Graphtals

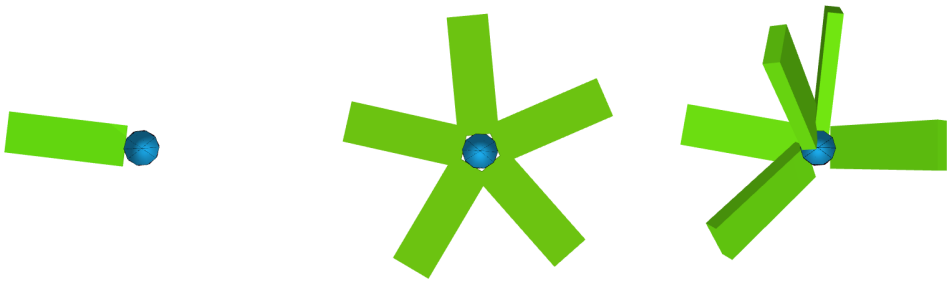
Given this over-representation of linear shapes, we devised another developmental model, this time derived from the directed graphs designed in [Sims 1994b]. In this case, the shape repository is an integral part of the morphological controller, each such shape being considered as a node in a directed graph. Links from a node to another indicates a parent-child relationship, i.e. the latter will be produced by the former given favourable conditions.

The content of each shape data is identical to those previously used but for an additional parameter, *medium* which indicates whether the organs should grow above or below ground (an incorrectly placed organ is instantly removed). Indeed with these Graphtals, individual controllers for each layer of the plant is no longer required. Instead a single one is used to produce all of the organs of the plant according to a set of rules similar to that of their initial introduction.

In addition to defining a parent-child relationship, links between node pairs also carry both control and positional information. The former includes the same *minScale* parameter which inhibits sub-organ production until a sufficient size increase has been reached. The *recursivity* field has been made local thus allowing each recursive connection to specify how often it should be triggered. The last control variable is the *relativeScale* $\in]0, 1]$ which impose a limit on how much size the child organ can gain with respect to its parent, i.e. with V_p and V_c the volume of the parent and child organs, respectively, we maintain $V_c \leq relativeScale\ V_p$.

The other fields concerns themselves with the placement of the sub-organ(s) by specifying the position *dp*, relative rotation *dr* and potential *effect*. The first two are quite straightforward and are implemented as quaternion[2]-based rotations as illustrated in figure 3.7.

[2] Unit quaternions, also known as versors, provide a convenient mathematical notation for representing orientations and rotations of objects in three dimensions. Compared to Euler angles they are simpler to compose and avoid the problem of gimbal lock. Compared to rotation matrices they are more compact, more numerically stable, and more efficient



(a) none (b) radial (c) random

Figure 3.8: Pattern-producing effects

In this model, the local coordinate system of organs places $\vec{x}$ along the major dimension. For these rotational operators we first consider this coordinate system, shifted along the length of the organ so that it rests at its extremity, called $C_0 = (O_0, \vec{x}_0, \vec{y}_0, \vec{z}_0)$. The parameter dp is a quaternion with no rotation around the $\vec{x}$ axis, its goal being to indicate the direction in which to place the child organ with respect to the parent's center of mass. This produces the secondary coordinate system $C_1 = (O_1, \vec{x}_1, \vec{y}_1, \vec{z}_1)$ which is further modified by the relative rotation dr. Indeed considering that the initial orientation of the child organ (in $C1$) is defined by its location on the parent's surface, this second parameters allows the genotype to independently choose the direction this sub-organ should be facing. This results in a final coordinate system $C_2 = (O_2, \vec{x}_2, \vec{y}_2, \vec{z}_2)$ which serves as the basis for the newly spawned organ. In this case rotations around the $\vec{x}$ vector are allowed so that all organs would not lie on the same plane which, though occurring in nature, is more an exception (e.g. fern) than a rule (e.g. most trees).

Finally, in an approach similar to that used in [Sims 1994b], we endowed our link data with pattern producing capabilities, the so-called *effect*. Given that we work with plants, instead of animals, we selected a different set of instructions compared to that of Sims.

Figure 3.8 enumerates the possible patterned produced by this field for a link $n_1 \rightarrow n_2$ with n_1 the blue, parent organ and n_2 the green, child organ. The first effect is a no-operation, i.e. a single organ is produced with its orientation directly determined by dp and dr. The second is controlled by a vector $\vec{v}$ and a number of iterations n. In addition to the first instance of n_2 another $n-1$ are also generated, each placed on same plane defined by the origin of the local coordinate system and $\vec{v}$. These copies are uniformly placed with a distance between each other of $2\pi/n$ which allows for a very compact encoding of regular morphologies such as those exhibited by flowers.

Lastly, the random effect is also parameterized by a repetition count n but its other is the seed s for a random number generator (RNG). Indeed each additional organ is a clone of the base one though with a random position on the parent's surface so as to easily code for repetitive structure with no

straightforward organisation (e.g. tree branches). The use of a stored RNG seed allows for determinism in the chaotic branching produced: though each application is likely to generate different positions, they will stay identical for each instantiation of the same genotype. In this way the random sequence produced is indirectly a fixed part of the genome, given the same implementation.

All other aspects of the plant life-cycle (growth, metabolism ...) are identical to that of the L-System variant with the exception of a few variables already mentioned in table 3.3.

3.4.1 Morphologies

The same holds for the evolutionary process used to generate test morphologies, that is the rain-pattern, fitnesses and parameters are identical. In a similar fashion, we performed evolutions in both an empty and grass-filled environments. In this case the competition is provided by more simple constructs, mimicking slightly bent grass blades. There were a hundred of these, placed and rotated randomly but consistently across runs.

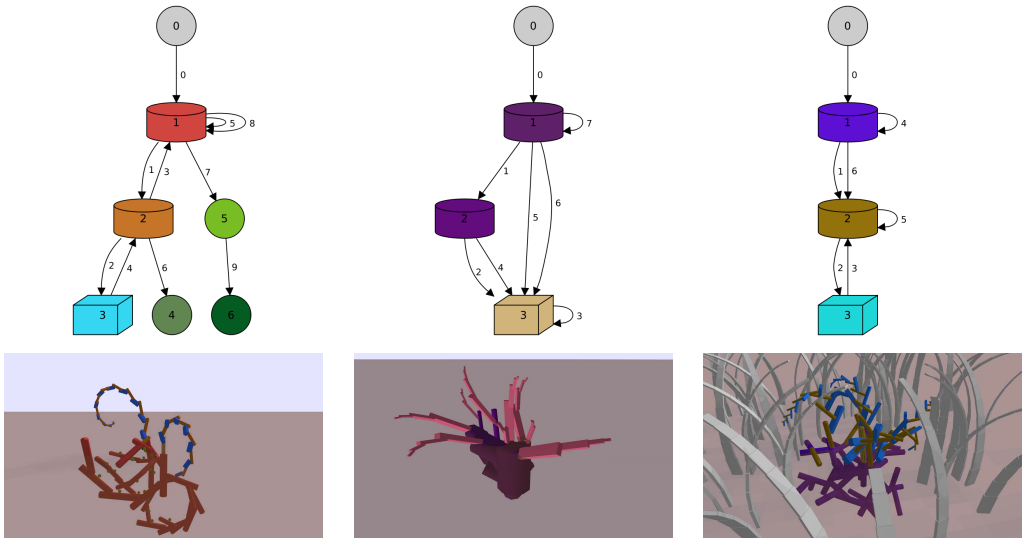
The first difference that we noticed, when compared to L-Systems, is that somewhat complex morphologies emerged even without the presence of artificially induced competition. As displayed in the first two pictures of figure 3.9a, we can see a regular structure (on the left) which also uses reserve organs to increase its storage capabilities. The second plant produced a more compact morphology with its root structure appearing as a single amalgamate while its broad dispersal of leaves allowed for a very efficient light gathering surface.

Figure 3.9b provides a panel of the diversity generated by this encoding ranging from the simple shapes exhibited by the upper left plant up to the amorphous congregation of the upper right. Some of the obtained morphologies, although perfectly valid given the experimental settings, are quite unfamiliar as the snake-like root trail of the seventh plant or, in the case of the last one, root-producing leaves.

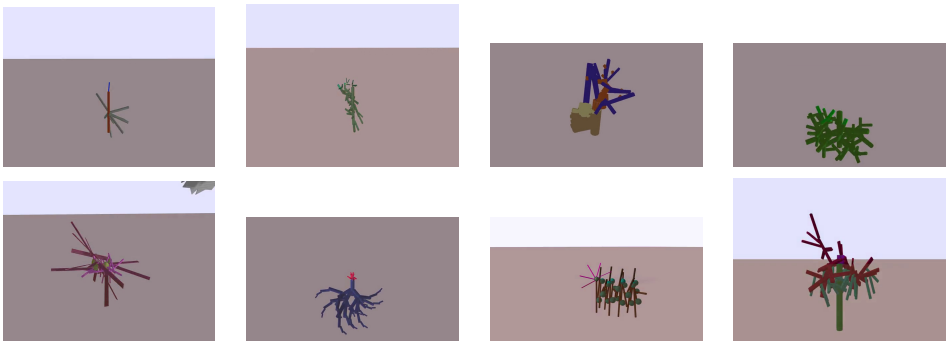
In the case of the artificially induced vertical competition similar trends can be seen with the individual at the rightmost side of figure 3.9a exhibiting complex structure though with a strong emphasis on upward growth when compared to the "plain" case. Similar observations can be made for the additional samples presented in figure 3.9c, where all morphologies show a definite trend to out-grow its static competitors.

Furthermore, the graphs displayed at the topmost portion of figure 3.9a show the genotypic contents of the individuals below in terms of nodes and connections. There, we can see that all have picked up on the utility of recursive links in the developmental process. More importantly, however, a recurring pattern, discovered frequently and independently across evolutionary run, is exhibited by the left and right-most genotypes.

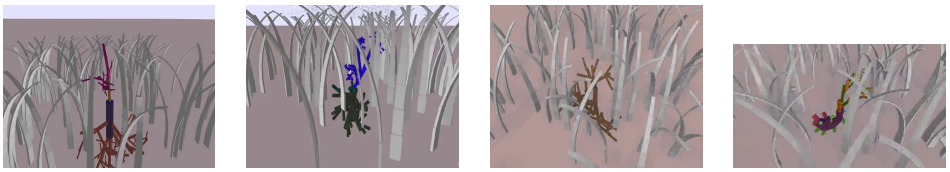
Indeed, while there is a regulatory parameter (*recursivity*, as detailed earlier) designed to prevent excessive organ duration the developmental, a solution for



(a) Genotype and phenotype



(b) Empty ground case



(c) Competition case

Figure 3.9: Samples of morphologies obtained via Graphtal encoding. Videos of these individual's development can be seen at `https://vimeo.com/showcase/5075634`. The specific on the top row's graphs are detailed in the following chapter.

near-unlimited growth lies in the use of recurrent connections (e.g. $2 \leftrightarrow 3$ for the rightmost plant). In this manner the limitation is never triggered and the individual can produce much more organs that allowed for, though with the compromise of these strangely alternating shapes: in this case leaf-stem-leaf.

3.5 Conclusion

While this series of experiment resulted in a publication at the Morphogenetic Engineering Workshop at ECAL 2017 [Dubois et al. 2017], their heavy focus on shape generation ultimately was a source of concern. The use of a physics engine, to model fine-grained constraints between organ pairs, did produce more robust morphologies but not without a heavy CPU cost. Additionally, the number of constraints to solve for a single time step resulted in some organisms exhibiting very strong instabilities[3], the roots of which lying in the engine's struggle to satisfy such a large set of equations.

In terms of differences between L-Systems and Graphtals, one can note the morphological differences visible in both panels (figures 3.6 and 3.9). Indeed while in the former case the individuals are elongated with stark rotations, in the latter case the shapes are fluid and natural-looking. Two reasons come to mind to explain this divergence: first the angle of rotation is a organ-local genetic parameter in the Graphtals whereas it is a global constant for the L-System. Second the application of one "step" of derivation is of much lower granularity in the former case with the depth increasing by at most one organ whereas in the latter case multiple organs can be added in succession which might prove a problem in such a constraint environment. With regards to these results, we settled on further use of the Graphtals for the future experiments.

Furthermore, the lack of structural plausibility exhibited by a number of the individuals presented here (including but not limited to stems growing over leaves) raises the question of whether or not this is a desirable result. Indeed concerned as we are by evolutionary trends, the underlying population need not be limited to life-as-we-know-it especially on such minor points.

A more limiting factor is the computation time involved in evaluating a single individual: from a median duration of about a hundred milliseconds up to 5 seconds for the most complex or morphologies. In its current state, this model would not stand the scaling both in the size of the population and the simulated duration, both of which needing at least three more orders of magnitude.

This point will be addressed in parts in the next chapter, where we perform our first experiments on populations of individuals with a common survival objective.

[3] See `https://vimeo.com/223741643` for an individual whose branches seem to move as if by the action of the wind, while only physical instabilities are responsible.

Chapter 4

Limited Co-evolution

Abstract In the previous chapter we have shown how Graphtals can be used to encode for artificial plants morphologies when evaluated in isolation. In these pages we demonstrate how to apply such a model to the evolution of populations cohabiting in a shared environment. These plants are endowed with mating capabilities, combining genetic parameters with environmental factors to guide the reproduction process. In an exploratory experiment we show how diverse types of fitness functions guide evolution towards different types of strategies, with the multi-objective criterion resulting in more robust solutions. We also identify key characteristics for promoting the emergence of self-sustaining systems: a) seed dissemination upon plant death provides a gradient towards autonomous abscission[a] and b) reduced life expectancy helps escape the local minimum of individual survival versus community wellfare.

[a]Spontaneous shedding of a body part such as leaves or fruits

Résumé Dans le chapitre précédent, nous avons démontré comment les Graphtals peuvent être utilisés pour encoder des morphologies de plantes artificielles lorsqu'elles sont évaluées de manière isolée. Dans les pages suivantes, nous montrerons comment appliquer un tel modèle pour l'évolution de populations qui cohabitent dans un environnement partagé. Ces plantes ont la capacité de se reproduire, en combinant des paramètres génétiques avec des facteurs environnementaux pour guider le processus de fécondation. Dans une expérience exploratoire, nous rapporterons comment les différents types de fitness guident l'évolution vers diverses stratégies et tout particulièrement les critères multiobjectifs qui conduisent à des solutions plus stables. Nous identifierons aussi les caractéristiques clefs qui promeuvent l'émergence de systèmes auto-suffisants : a) la dissémination des graines à la mort de la plante permettant un gradient vers l'abscission autonome, b) une espérance de vie réduite facilitant la sortie du minimum local de survie individuelle au profit d'une meilleure persistance de la communauté.

Given the morphologies, of reasonable complexities, obtained in the previous experimental setting, we deemed our model ready for the next step towards obtaining autonomous ecosystems: reproductive behavior. In this chapter we detail first the modifications effected upon the genomic, metabolic and developmental parts of the individuals life cycle (section 4.1). Follows the changes in environmental dynamics and the population setup (section 4.1.5). We then expand upon the experimental protocol devised to promote self-reproduction in a small population of artificial plants and resulting morphologies and dynamics (section 4.2). We conclude by drawing lessons from the observed trends and the necessary modifications they imply.

4.1 Self-reproducing vegetals

Moving on from single individual evolution to this first attempt at a self-made ecosystem required a number of changes throughout the framework described in the previous chapter most of which related to the plants' demeanor. Though the bulk of the genetic substrate remains mostly unchanged, a few fields have been added and some of the most computationally costly ones have been trimmed off. One especially noteworthy difference with the previous setting is the use of a single topology as the starting point of all ulterior evolution.

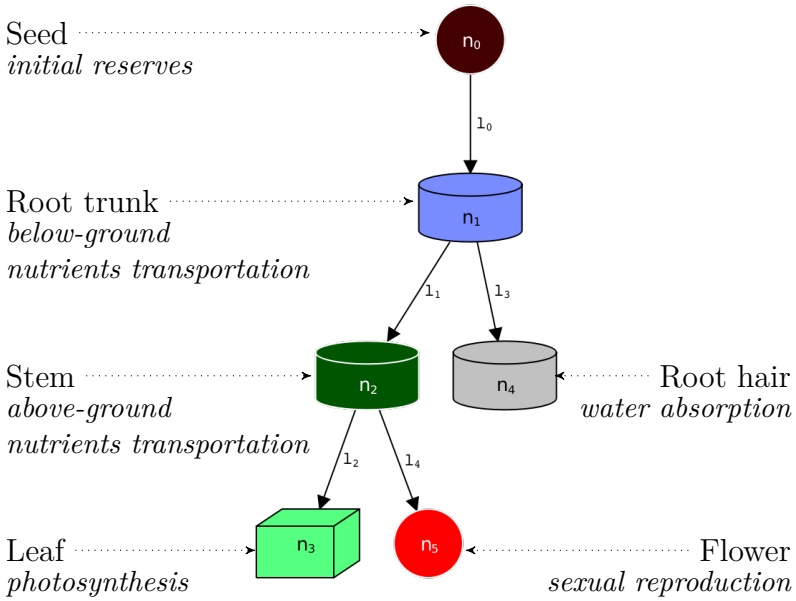


Figure 4.1: Base graphtal in all following experiments

As depicted by figure 4.1, we rely on a minimalist structure that can produce all organs type including the all-important flower, crucial component of the reproduction process. Individual fields inside each of the nodes and links are still subjected to random initialisation (including the seed though with some

limitations). The behavior of individuals plants in also altered in a number of small ways as the broad view offered by algorithm 2, presented latter on, shows.

The procedure for nutrients transportation has been improved, as described in 4.1.1 which increased its complexity both in terms of organ-organ interaction and computational cost. Oppositely the updating of constraints at the end of the plant step has been completely removed so as to drastically reduce the load imposed on the physics engine[1]. Indeed the 6DoF constraints between organs that we used in the previous chapter, being one of the major computational bottleneck, have been relaxed in favor of a lighter approach for organ production as will be detailed in section 4.1.3.

```
   Data: e, environment
 1 for p, plant do
 2 |   p.storeLight(e);
 3 |   p.storeWater(e);
 4 |   p.transport();
 5 |   p.manageSeed();
 6 |   p.photosynthesize();
 7 |   p.consumeResources();
 8 |   p.processOrganDeath();
 9 |   if p is male then p.tryReproduce();
10 |   p.produceOrgans(e);
11 |   p.growOrgans();
   |   p.updateConstraints(e);
12 end
13 e.step();
```

Algorithm 2: Algorithm of a simulation step. Modified lines are indicated by their color (green: added, orange: modified, red: removed).

Similarly the growth process of individual organs (section 4.1.2) has also seen some trimming which, though reducing the range of growth dynamics, further improved the efficiency of our simulation. Last and not least, the plants were endowed with self-reproductive capabilities through a number of genetic fields. As we argue in section 4.1.4, these, alongside the control algorithm, are a novelty in this field of research.

4.1.1 Metabolism

Most notable amongst the numerous variations between these and the previous experimental settings is the switching from a global resource allocation to a

[1] As a result 2 years simulation of a single individual went down to a few tens of milliseconds, an order of magnitude lower than in the previous setting

Skill	PHOTOSYNTHESIS	STRUCTURAL	ROOT TRUNK	ROOT HAIR	RESERVE	SEXUAL
Value	.5	1	1	.5	.75	.5

Table 4.1: Initial diffusion rates with respect to an organ's skill

local diffusion algorithm. Indeed, instead of a share pool of nutrients, each organ was now put "in charge" of the resources it contains thus allowing the apparition of delay between producer and consumer.

Given a source organ A we define its neighborhood N_A as its parent organ (if any) and all sub-organs (those it has generated through graphtal derivation). Thus, given the variables defined in table 3.3, we can determine which organs in N_A are in need of nutrients by examining their "fullness" defined, with respect to s^e and m^e the current and maximal storage, as:

$$f_o^e = \frac{s_o^e}{m_o^e} \tag{4.1}$$

which is the ratio of their storage capacity actually containing a given resource e. From there we can further refine the definition of N_A by only considering these neighbors with sufficiently low fullness of e when compared to the organ A of reference. This results in the filtered set $\bar{N}_A^e$ whose formal definition is:

$$\bar{N}_A^e = \{o \in N_A / f_o^e < f_A^e\}$$

With this set, we can determine the distribution of resources based on the individual needs of these neighboring organs and the available amount or nutrients A can share. Thus the quantity de_{AB} of element e transferred from A to B is given by

$$de_{AB} = k_{AB}^e * \frac{a_A^e * (1 - f_B^e)}{\sum\limits_{o \in \bar{N}_A^e} 1 - f_o^e} \tag{4.2}$$

The variable $k_{AB}^e \in [0, 1]$ is a refraction coefficient which accounts for one of the extensions we added to the model. In essence, it codes for a form of backpropagation in which organ are marked for their contribution to the plant welfare, according to their capabilities and place in the topology in a similar manner to that described in [Zahadat et al. 2017a]. In practice this should allow plants to adapt during their lifetime to varying environmental conditions which, given the current objective of larger populations, would prove useful for coping with both the abiotic and biotic components of the ecosystem.

We maintain a list of all connections between organs (i.e. all parent-child relationships) with their corresponding value of k_{AB}. Initially those are set according to the contents of table 4.1 by simple product, that is to say for a pair of organs A and B of initial diffusion rates k_A and k_B, respectively, then $k_{AB} = k_A k_B$.

The backpropagation itself only takes place after all local diffusions have been performed for the current step and starts by considering all producer organs for a given element e (i.e. leaves for glucose and root hairs for water). These are marked for their *direct* contribution to the plants welfare in virtue of the accumulation of the appropriate resource in the current time step. That is with $s_o^e(t)$ and $s_o^e(t-1)$ the stored amount of element e by organ o at the current and previous time step, respectively, we can define the efficiency in terms of resource collection through equation 4.3.

$$m^e(o) = \frac{s_o^e(t) - s_o^e(t-1)}{V_o} \tag{4.3}$$

This mark is thus normalized to discard the effect of size, only concentrating onto the effective variation in storage. In this manner, a producer organ which gathers only limited amounts of resources (an obscured leaf for instance) will receive a low score, while if its stores are being depleted its mark will further worsen, going into the negatives.

The mark for non-producing organs of element e is given by aggregating the scores of their sub-organs to provide a comprehensive value of its relevance in the diffusion network. This cumulative metric $cm^e(o)$, for a given organ o of with N_o the set of its sub-organs, is defined as:

$$cm^e(o) = \frac{P_o^e m^e(o) + \sum\limits_{o' \in N_o} cm^e(o')}{P_o^e + |N_o|} \tag{4.4}$$

$$\text{with } P_o^e = \begin{cases} 1 & \text{if } o \text{ is a producer of element } e \\ 0 & \text{otherwise} \end{cases} \tag{4.5}$$

Additionally, a non-producing organ for which $|N_o| = 0$ will be allocated a score of 0. In all other cases, $cm^e(o)$ determines how much the network rooted at o contributes to the replenishment of the distributed resource storage by affecting larger scores to those parts of the plant that perform better. Indeed, the diffusion coefficient between two organs A and B will then be updated as follows:

$$\frac{dk_{AB}^e}{dt} = 1 - cm^e(B)\delta_W \tag{4.6}$$

where δ_W is an inertia factor. This parameter is located inside the global portion of the plant genome and subjected to mutation. Initially it is set to 0 so that the whole backpropagation process is bypassed until such mutations occurs as to provide the first attempts at calibration.

Globally this method was designed to allow plants to accommodate environmental conditions that could not be directly under the control of the morphological components of the genomes either because of the random rotation imposed upon the plant at initialization or an excessively competitive neighborhood.

4.1.2 Organ growth

Another area of the plants' life cycle that was modified in-between these experiments, this time with a focus on reducing computational cost, is the growth process of individual organs. As previously, an organ can only grow in size if it does not exceed the genetically defined maximum. In this version, however, this upper bound is based on relative scales instead of volume, i.e. we directly compare the amount of growth with the limit, instead of a derived figure. The anisotropic growth described previously which allowed organs to alter their shape ratios in a wide manner has also be replaced by a more straightforward uniform scaling.

Indeed, instead of a distributed algorithm, the updated version is controlled by plant-wide parameters namely $\mathfrak{G}$, previously used to further regulate local growth values, and $\mathfrak{G}^m$ which defines the maximal scaling value attainable as described by:

$$1 + (\mathfrak{G}^m - 1) \tanh\left(\frac{3t}{\mathfrak{G}}\right) \tag{4.7}$$

in which t is a variable local to each organ which counts the number of times such a growth process has been applied to it. Additionally those computations are only performed at midnight every day, that is, with the current parameters, every 100 simulation step.

All in all, despite these modifications, individual organs regain a measure of control over their individual growth patterns though much of the computational load is diverted at the plant-level thus providing much need reductions of simulation times.

4.1.3 Budding process

The last alteration effected on the model concerns the actual production of new organs. In the previous settings these where handled by so-called "buds" which had only an abstract existence, i.e. these where an algorithmic constructs used to store the location and orientation of soon-to-be-instantiated organs.

Here they become an integral part of the plant morphology in the sense that as soon as an organ is created the list of its potential budding points is extracted for storage in a plant-wide container. There, they are matured for a complete day during which each will accumulate information about its direct surroundings with the ultimate goal of fine-tuning the genetic recipe to better suit the local conditions.

In our case, these comprise gravity, neighborhood and light although all but the last are determined by examining the surroundings of a bud at the moment it is transformed into an actual organ. The special case of light comes from the fact that not only is the light source dynamic throughout the day but it is also highly dependant on all neighboring organs (whether or not they belong to the same plant). As an effort to provide accurate description of the dominant

incoming directions, rays are cast every simulation step for a day with each positive hit (i.e. the bud is directly illuminated by the sun) being integrated into an average vector. This provides plants who rely on phototropism an indication of which way to grow so as to maximize its exposed leaf surface.

Similarly both other forms of tropisms result in a stimuli vector, the difference being that their computation can be done as late as required. This results in a collection of three vector $\vec{s_l}$, $\vec{s_g}$ and $\vec{s_n}$ indicating the stimuli direction for the phototropism, gravitropism and "densitropism", respectively. By definition, $\vec{s_l}$ will point towards the averaged sun position, as seen by the bud itself, while $\vec{s_g}$ will always point downward and $\vec{s_n}$ in the direction with the lowest density of organs.

Each link in the genome is endowed with as many coefficients for regulating the corresponding form of tropism denoted w_t for t a particular stimuli. Initially all values are set to 0, as in the case of the metabolic backpropagation, and mutation and selection must determine the pertinence of actually using this mechanism. Indeed we may expect genetic drift to have a strong impact on the non-coding (or rather neutral) alleles such as the use of phototropism in an (underground) root hair for which $\vec{s_l} = \vec{0}$, by definition.

The net result is that given $\vec{v}$ the original direction the organ was to be facing and $\vec{q_l}, \vec{q_g}, \vec{q_n}$ the quaternions transforming $\vec{v}$ into the corresponding stimuli vector, we can define the variation in orientation as:

$$\vec{q} = \prod_{t \in \{l,g,n\}} w_t \vec{r_t} \tag{4.8}$$

The relation between $\vec{q_l}$ and $\vec{r_l}$ being one of normalization, indeed the angle represented by $\vec{r_l}$ is guaranteed to be at most $\pi/8$ to avoid large reorientation due to strongly divergent stimuli. This additional rotation is applied after those described in the previous chapter thus allowing plant to enact fine-tuning of their organs position in response to their local conditions.

Another drastic measure taken to reduce the computational cost of this model when deployed to larger population was to completely remove the 6DoF constraints (managed through the Bullet physics engine) between organ pairs. Thus instead of connecting a child organ to its parent and leaving to Bullet the task of maintaining, if at all possible, their relative system coordinates in appropriate positions, we perform much of this work by hand. While this implies additional computation to keep each organ at its appropriate place in the plant structure, this also removes the most expensive bottleneck of our previous simulations.

Plant thus become, from the physics engine point of view, static objects composed of a collection of sub-shapes instead of the uncomfortable hierarchy of (sometimes contradictory) constraints that required extensive computations to keep in check. Thus, given that organs can no longer be torn off as a result of a failing connection, the *solidity* parameter previously described has been removed.

Thus, to create an organ, we could no longer rely on the physics engine to handle organ-organ intersections and instead perform a collision test to determine if enough space can be found at the target location. As in the previous chapter this test takes into account neither the parent nor the siblings of the organ for which the test is performed.

4.1.4 Autonomous reproduction

One of the most powerful tools available to Life is its ability to adapt through the process of natural selection. Over the course of history numerous propagation scheme have been developed. In this work, we chose to focus on sexual reproduction because of its greater degree of interactions and inter-species diversity.

To this end, we included plant-wide genomic components devoted to reproduction: the gender, compatibility metrics (optimal genetic distance μ, inbreed tolerance σ_i and out-breed tolerance σ_o) and sexual organs previously mentioned (organ skill `REPRODUCTION`). The global outline of the process is given by algorithm 3.

Data: P, set of plants
$M \leftarrow \{p \in P / p \text{ is male}\}$;
for $m \in M$ **do**
  $G_m \leftarrow genotype(m)$;
  for s_m, *stamen* $\in m$ **do**
    $f \leftarrow$ random female, with pistils, in range of m;
    $G_f \leftarrow genotype(f)$;
    $A \leftarrow align(G_m, G_f)$;
    $d \leftarrow distance(G_m, G_f, A)$;
    $C_{mf} \leftarrow compatibility(d, G_f)$;
    if *random toss with probability* C_{mf} **then**
      delete s_m;
      $p_f \leftarrow$ random pistil from f;
      $p_f \leftarrow$ Fruit(Mutate(Cross(A, G_m, G_f)));
    end
  end
end

Algorithm 3: Mating process

Every simulation step, each male plant will search potential mates in its vicinity. In this case, it is equivalent to finding those females in collision with a large sphere of radius $R = 50L$, with L the diagonal length of the plant's AABB[2]. Any such individual will then be queried for mating by the

[2]Axis Aligned Bounding Box

male, submitting its genome to a compatibility computation performed by the solicited female.

The first step in this procedure consists in aligning both genomes to find their common portions. For elementary fields, such as the plant-wide growth speed $\mathfrak{G}$ for instance, this is trivially done. Aligning the graphtals components however is a much more involved undertaking, performed as follows.

Consider two genomes $G_1 = (N_1, L_1)$ and $G_2 = (N_2, L_2)$ with N_i and L_i the nodes and links collections, respectively. We further define $l_1 \in L_1$ and $l_2 \in L_2$ and for l, a link, l_i^{id} refers to its identifier and l_i^o, l_i^i to its source and target node, respectively. From there the alignment consists in forming three subsets by searching for match between the links in both genomes.

$$\begin{aligned} M_a &= \{\{l_1, l_2\} / l_1^{id} = l_2^{id} \wedge \ l_1^i = l_2^i \wedge l_1^o = l_2^o \ \} \\ M_{ia} &= \{\{l_1, l_2\} / l_1^{id} = l_2^{id} \wedge (l_1^i \neq l_2^i \vee l_1^o \neq l_2^o)\} \\ M_i &= \{l_1 / \nexists l_2 \in L_2, l_1^{id} = l_2^{id}\} \cup \ \{l_2 / \nexists l_1 \in L_1, l_2^{id} = l_1^{id}\} \end{aligned}$$

The first, M_a, is the (match) set of links that were found identical in both individuals, that is with the same id, source and target nodes. The second set, M_{ia}, comprises those link pairs that did not match exactly because of differences in the components they connect. Given the complexity of the derivation procedure such minimal change could, nonetheless, have large impact on the produced morphologies and such cases are thus labelled separately. Finally M_i collects all remaining links for which no equivalent could be found in the other genomes. Whether as the result of introduction of new budding types or deletion of old ones, these instances are the most striking cases of diverging morphologies. As a result of this graphtal-aligning procedure, we obtain an overview of the topological similarities (and differences) between the genotypes of both individuals.

This, in turn allows for the computation of the genetic distance metric d. As previously, elementary fields are just as trivially processed: assuming that v_1 and v_2 correspond to the values of such an elementary field in both genomes which is bounded by $[v_{min}, v_{max}]$, the distance between the two genomes with respect to this field is given by:

$$\mathfrak{D}(v_1, v_2) = \frac{|v_1 - v_2|}{v_{max} - v_{min}} \tag{4.9}$$

More complex fields such as containers or structures simply average the differences between their individual components. For a field f comprised of the sub-fields $f^1, ..., f^n$ the distance between two such instances f_1, f_2 is thus:

$$\mathfrak{D}(f_1, f_2) = \frac{1}{n} \sum_{i=1}^{n} \mathfrak{D}(f_1^i, f_2^i) \tag{4.10}$$

Computing the distance for the graphtal topology, however, is a more complex task, though only marginally thanks to the previous alignment procedure.

Indeed for all pairs $\{l_1, l_2\} \in M_a$ we can define the distance between both links as:

$$\mathfrak{D}(l_1, l_2) = \frac{1}{3}(\mathfrak{D}(l_1^i, l_2^i) + \mathfrak{D}(l_1^o, l_2^o) + \mathfrak{D}(l_1^d, l_2^d)) \tag{4.11}$$

where the internal calls to $\mathfrak{D}(l_1^*, l_2^*)$ fall into the case described by equation 4.10 and l^d correspond to the data associated with each link (rotations, effects ...). Similarly, pairs in M_{ia} are submitted to the same procedure although, in this case, distance are likely to be much more pronounced due to the fact that the compared nodes may be drastically different. Finally, all unaccounted for links are counted as maximally distant resulting in the following equation for distance between two graphtals:

$$\mathfrak{D}(G_1, G_2) = |M_i| + \sum_{\{l_1, l_2\} \in M_a \cup M_{ia}} \mathfrak{D}(l_1, l_2) \tag{4.12}$$

Once the distance between both genomes is defined, the female individual of the potential mating pair transforms this somewhat objective metric into a much more subjective one: the compatibility $\mathfrak{C}$. This relies on the genetic components mentioned in passing at the beginning of this section namely the optimal genetic distance μ, inbreed tolerance σ_i and outbreed tolerance σ_o. These are combined to produce a compatibility function (equation 4.15) that performs a subjective assessment, from the female point of view, of how likely producing an offspring with the given partner would be a good investment of its limited resources.

$$\sigma = \begin{cases} \sigma_i & d < \mu \\ \sigma_o & d \geq \mu \end{cases} \tag{4.13}$$

$$d = \mathfrak{D}(G_1, G_2) \tag{4.14}$$

$$\mathfrak{C}(G_1, G_2) = e^{-\frac{(d-\mu)^2}{2*\sigma^2}} \tag{4.15}$$

This function has the general shape depicted in figure 4.2 where, *for an optimal distance $d = \mu$, the resulting compatibility is maximal* (i.e. 100%). In all other cases not only will the compatibility be strictly lower but furthermore, due to the presence of two standard derivations, the slope of the exponential decay is likely to be different from one side of μ to the other. *This enables speciation patterns to both emerge, with population clusters moving away from one another in the genetic space, and evolve with, for instance, species reducing the outbreeding tolerance to consolidate those alleles found most useful in their environment.*

Which brings us to the last steps of the reproduction algorithm. Given c, the compatibility the female marked the male with, we perform a biased coin toss of success probability c which, in case of a failure, results in a no-operation. Whenever this test passes, however, the stamen that initiated the mating attempt is destroyed with all its resources so as to limit the number of reproduction an individual can engage in. On the female side, a random pistil

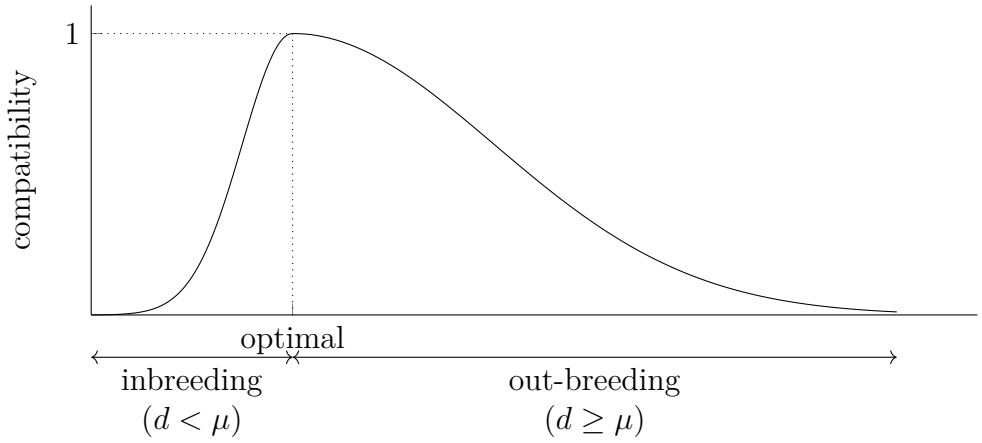


Figure 4.2: Genetic compatibility function

is selected an replaced by a fruit which contains the child genome obtained by selecting either alleles for elementary fields and performing an additional mutation.

The procedure generating an offspring graphtal from two parents is relatively straightforward, once more thanks to the alignment previously performed. Indeed the contents of those links that have been matched are randomly selected from one parent or the other. Partially matching links (i.e. those for $l_1^i \neq l_2^i \vee l_1^o \neq l_2^o$) undergo the same treatment but for their target and source nodes which, being potentially different, are also randomly selected. Links only found in one parent are added with a probability of .5 to the offspring. Nodes found in both parents are also crossed down to their elementary components and all nodes required, as a result of selecting a mismatching or partially matching link, are also added. Finally, nodes which cannot be produced indirectly by the seed are removed from the graph as an effort to avoid bloating (which is still possible depending of the production values stored in the links, for instance).

This crossover operator differs from those commonly found in the literature (e.g. [Sims 1994b;Bonfim et al. 2005; Disset et al. 2016]) on three points: 1) it can fail early on, 2) is biased by the *female* genome and 3) has low resistance to large structural differences. The rationale behind point 3 is that, instead of devising a robust operator that can produce a somewhat viable offspring from two completely unrelated individuals, a minimalist alignment procedure is better suited to sexual reproduction of same species creatures in which the population is mostly homogeneous. Indeed, point 1 guarantees that the more both genomes are different the less likely it is that crossing will be attempted at all. The decision of aborting or proceeding with the reproduction is left to the female individual as, in this sexual scheme, it will have to provide the fruit in resources. Given its capacity to abort a reproduction based on genetic parameters of proximity, we designed this operator as the Bail-Out Crossover (shortened into BOC).

As is the case with the tolerance parameters σ_i and σ_o, the optimal distance

is also subjected to evolution especially given that our genomes are only loosely bounded in terms of nodes and links count. Indeed large genomes should have either a proportionally higher value of μ or a more homogeneous population to cope with this increase in maximal genetic distance. All in all this scheme allows small plants population to form spontaneously and, by themselves, specify the search space in which to mutate their alleles. Indeed this operator was designed with bio-mimetism in mind by trying to reproduce the manner in which spores that do not rely on "magic-bullet" insects effectively cope with the problem of inter-species mating.

Once generated, fruits behave as any other organ, accumulating resources and increasing in size. Every simulation step, however, they are given a chance to separate from their plant and start producing a new individual. This event only occurs under two conditions: the fruit needs to have reached its maximal size and it must be fully stocked in both nutrients. When such constraints are met the fruit f is detached and displaced on the $\vec{x}\vec{z}$ plane by a vector $\vec{dp} = (\theta, R)$. The angle θ is uniformly picked in the $[0, 2\pi]$ range while R follows a truncated normal distribution of mean 1cm and standard derivation $S/8$ with S the environment's diagonal size (). The resulting distance is further clamped to remain inside the $[1cm, R_{max}]$ range where $R_{max} = S/(2 * \rho_f)$.

Given the number of potential neighboring plants, insertion of the newly produced plant is not guaranteed to succeed. Up to 10 attempts are made to find a collision-free position in the vicinity of the selected position. If no such position can be found the insertion is aborted and the genetic material is lost. Additionally the initial resources levels available to the plant will be, contrary to those initially populating the environment, directly defined by that of the fruit. Indeed it will become the seed of the offspring, thus encouraging plants to store a generous amount of nutrients in the fruits to ensure germination.

4.1.5 Environment & Ecosystem

In case of the abiotic component of the ecosystem much is similar to the previous experimental setting especially with respect to the light. The water cycle, however, has been simplified to further improve simulation times: water levels are now stable throughout the simulation with both diffusion and consumption being removed from the model. The ground size was also reduced to half the size it previously had (down to 10 meters wide and 5 meters deep). Furthermore, given that the ground surface is expected to be covered in plants, no components were added to promote vertical competition, this being left to the discretion of individual evolutionary trends.

The complete ecosystem is composed of both a description of the environment (currently not evolved) and a set of plants 'templates'. In these experiments a single template is considered. The procedure to translate these templates into a densely populated ecosystem is straightforward. First the environment is divided into as many cells as the requested number of plants

(100 in this experiment) and the largest seed size is tested against half the cell size. If this fails, the whole ecosystem is deemed non-viable and the simulation is aborted, thus preventing plants from having too large initial reserves. Otherwise, each cell is subdivided once more in four and a plant is placed in a single subcell with a random genome from the set of templates and a random vertical rotation. This leaves enough room for autonomous reproduction to place offspring even when the initial population has not entirely died out. For determinism purposes, every random number used during the simulation (plants position, rotations, iterations, etc.) is generated from a fixed seed provided by the genotype (not evolved but randomly set).

4.2 Colonization dynamics

4.2.1 Evolution protocol

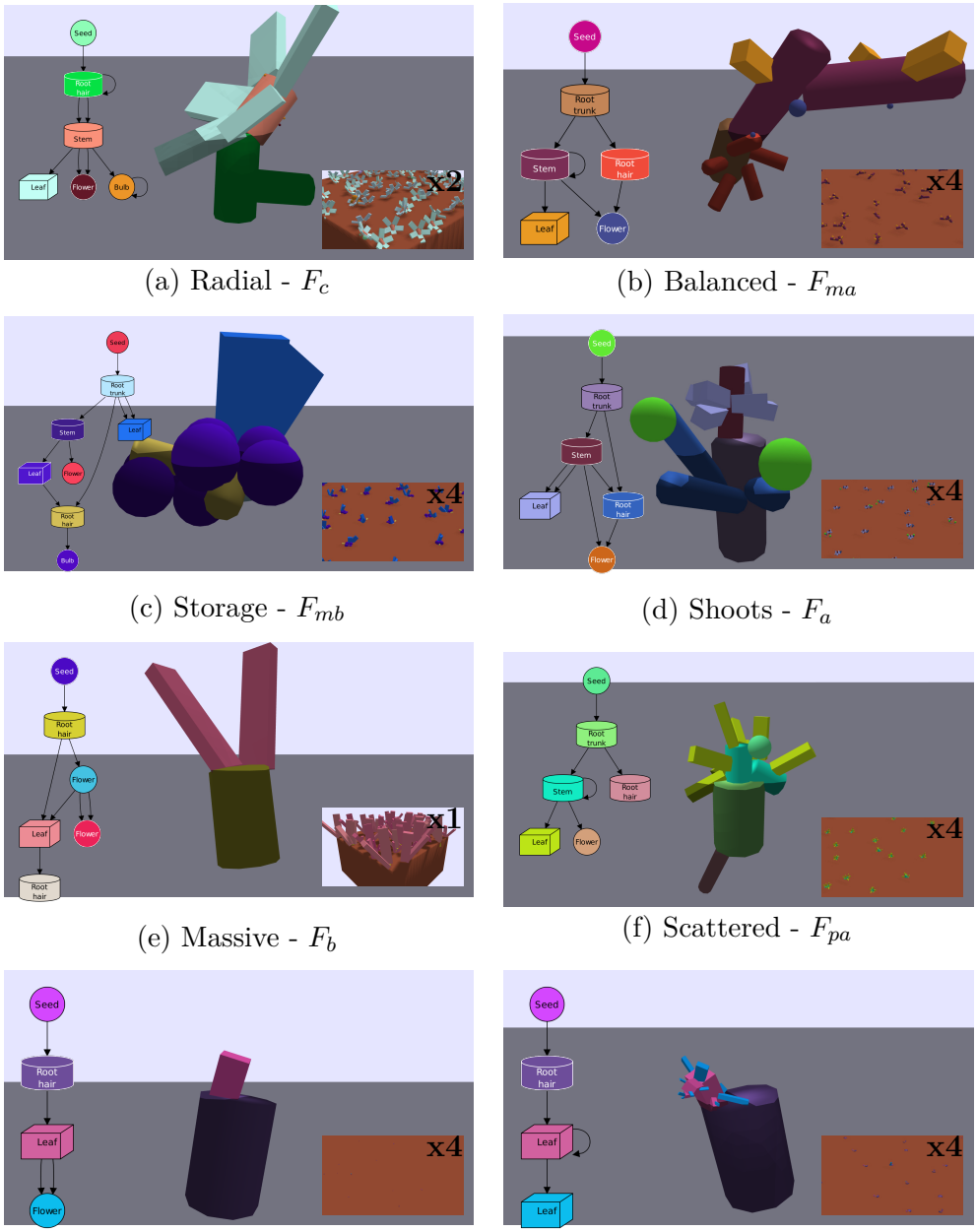
This work comes within the scope of studying long-term evolutionary trends especially in elaborate 3D ecosystems. However, evolving, from scratch, such systems with a non-trivial degree of complexity would require a prohibitive amount of computational resources. Stemming from this intent, the following experiment was designed to generate usable individuals to seed an environment with. Viable plants would thus have to develop strategies to both survive and reproduce so that their genetic material does not die off.

The evolution protocol relied on evolution programming where plants' genomes underwent single point, equiprobable, mutation on all of the fields involved. Evaluating a genotype implies populating an empty environment as described previously and then stepping back for a maximum of $N = 60000$ simulation steps (2 simulated years) to see whether autonomous dynamics would emerge.

In order to limit the search space to the genetic fields of the plants, the environment was kept constant in all runs. As we aimed for both efficiency and diversity, we devised a range of fitness functions F_* which evaluate the plant population P as described below.

$$
\begin{aligned}
\nu &= \frac{1}{N|P|} \\
F_b &= \nu \sum_{t \in N} \sum_{p \in P} \text{biomass}(p,t) \\
F_p &= \nu \sum_{t \in N} \sum_{p \in P} \text{production}(t,p) \\
F_c &= \frac{\nu}{W^2} \sum_{t \in N} \text{surface}(t) \\
F_a &= \nu \sum_{p \in P} \text{lifespan}(p) 2^{g-\alpha_p}
\end{aligned}
$$

where surface(t) corresponds to the total surface covered by plants at time t and α_p is designed to provide a smoother gradient towards reproduction. Plainly



(a) Radial - F_c

(b) Balanced - F_{ma}

(c) Storage - F_{mb}

(d) Shoots - F_a

(e) Massive - F_b

(f) Scattered - F_{pa}

(g) Reproduction - F_{ma}

(h) Survival - F_{mb}

Figure 4.3: Examples of the morphologies developed. (a) to (g) are at the 20th day and (h) is at the 30th.
From left to right: Genome, Single individual, Ecosystem (zoomed on central individual with factor in upper-right corner).
The fitness that produced this individual is indicated in the caption with F_{ma} indicating the age criterion of the multi-objective fitness F_m.
Videos of these individuals' full ecosystem can be seen on `https://vimeo.com/album/5075632`.

		Groups				
		G_b	G_p	G_c	G_a	G_m
Fitness	F_b	X				X
	F_p		X			X
	F_c			X		X
	F_a				X	X
	N_{ov}	X	X	X	X	X

Table 4.2: Evolution runs with different fitness allocation

put, these aim at producing plants which are: F_b) large, F_p) many-leaved, F_c) wide, F_a) fast reproducers. Given that every fitness is likely to be exploited into non-desired behaviors, a fifth one F_m is introduced that evaluates genomes on all four criteria at the same time. Furthermore, in order to prevent local optimum a *novelty metric* is used as proposed in [Lehman et al. 2008]. An individual's 'footprint', i.e. its synthetic behavioral description, is (F_a, F_b, F_p, R, G, S), with R the number of successful autonomous reproductions, G the number of autonomous generations and S the seed size.

An autonomous *reproduction*, in this context, is defined as two individuals (m, p) embedded in a simulation deciding on generating an offspring through the process described in algorithm 3. The autonomous *generation* g_c of such an offspring is $g_c = max(g_m, g_p)+1$, where g_m and g_p is the autonomous *generation* of the mother m and father p, respectively. Thus, in F_a, survival is greatly rewarded for individuals from latter generations. At the same time the α_p parameters penalizes ecosystem unfit for autonomous reproductions. Initially set to 0 it is incremented if any of the followings are true: no females, no males, no fruits to which we further add 1 − the maximal fruit maturity. In this manner ecosystem can gradually improve their score towards F_a by first finding the necessary prerequisite to reproduction before attempting any optimization upon it.

In each scenario, plants are evaluated on two to five criteria using a tournament selection where 3 participants are randomly selected from the population and compete on a random objective as described in [Disset et al. 2016] (see table 4.2 for a comprehensive list). Ten runs per fitness were dispatched on a cluster of Bi-Intel(r) IVYBRIDGE 2,8 Ghz 10-cores and were re-launched as soon as they completed an evolution (250 generations) with a maximal, total, duration of five hours.

4.2.2 Morphologies

While three out the five fitnesses performed an average of two evolutions in the given time frame (2.4 for F_a, 2.5 for F_c and 1.8 for F_m), the remaining two behaved very differently: while F_b produced 8.4 'champions' per run, F_p did

not manage to bring a single one to the 250 generations threshold (the best having been stopped at the 199th and the worst at the 93rd). This can be explained by observing the evaluation times of those final individuals which range from 12 milliseconds up to 10+ minutes.

In order to gain a better understanding of the situation, we manually examined the phenotype of the 40 best champions (out of a total of 215) that is 5 for every single objective fitness and another 5 for each criterion in F_m. Summarized in figure 4.3 are the morphologies of those creatures we found the most interesting.

As one can see these evolutions produced very different strategies to cope with the environment and their respective fitness. Variation in the sun's position and the plants' relative orientation led to either having large leaves so that production is maximized during short favorable moments (4.3a,4.3c,4.3e), or numerous, evenly spread, leaves so that sunlight can be efficiently gathered throughout the day by different parts of the plant (4.3d, 4.3f, 4.3h).

Root morphology was not thoroughly investigated, due to the uniform water distribution exerting only very limited evolutionary pressure, and most individual manage with a simple root trunk connected to a handful of capillary tubes (4.3b, 4.3d, 4.3f). Some even went as far as to completely forsake the former (4.3a, 4.3e, 4.3g).

As the autonomous reproduction process starts from flowers, their growth is of utmost importance for a species' permanence. All plants except two from figure 4.3 actually generate at least one such organ, though only 4.3g, 4.3e, 4.3a manage to bring them to maturity.

4.2.3 Strategies

From these morphologies and their associated dynamics graphs, we can extract three main strategies: quiescence, expansion and reproduction as illustrated in figure 4.4. The first one (in red) is quite straightforward in its survival method. One can see on the graphs that after a short burst of activity, early on in the simulation, this type of individuals goes into a quiescent state, keeping its metabolic value in a comfortable range so that most plants make it to the end. The expansionist (in blue) however, adopts a radically different approach: instead it tries to reach as fast as it can a mature state which can, depending on the plant, take up to a full year. This allows the ecosystem to compensate for the extremely high mortality rate: in the example depicted, 96% of the population dies in the first hundred days. Finally, the reproduction strategy (in green) relies on having the smallest possible morphology, i.e. a small seed and a single root hair directly connected to the leaf. Resources are mostly directed towards producing mature fruits as quickly as can be, thus maintaining a population in a safe range ($[60, 80]$ in this case).

It is interesting to compare these behaviors with those obtained in [Bornhofen et al. 2011] where varying environmental factors led to the emerge of

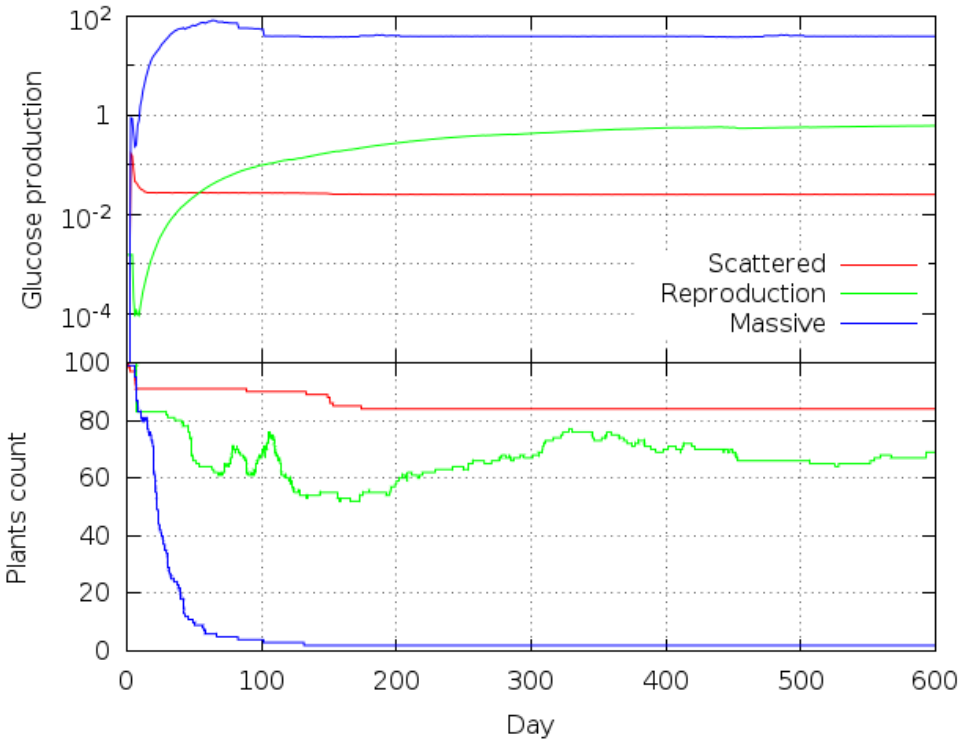


Figure 4.4: Typical examples of the three strategies' dynamics. Individuals are taken from figure 4.3 with the red, green and blue curves corresponding to 4.3f, 4.3g and 4.3e

the CSR triangle [Grime 1977]. Indeed, a plant population under favorable conditions should evolve towards individual competition while with decreasing resources availability a slower, more conservative, metabolism is expected. If exposed to recurrent, localized uncontrolled deaths, the ruderals would thrive with their fast life cycle and colonization approach.

The fact that all three strategies emerged within identical environments shows that, on the one hand, the genetic search space is large enough to contain very distinct viable genomes even before being subjected to an evolutionary process. On the other hand, it also warns about a possibly *too* large search space with functional genotypes separated by wide gaps of unfit combinations.

4.2.4 Influence of evaluation criteria

We now turn our attention to the evolution procedure itself and more specifically the contribution of our fitness functions set. The diversity of criteria used induced a similar amount of variability in the obtained genomes as one can see the range of morphologies and behavior obtained. However, all fitnesses did not perform equally both in terms of complexity (see figure 4.5) and relevance. Indeed, while F_b produced plants that could grow at a sustained pace, they proved quite simplistic, morphologically speaking, with almost 60% of the

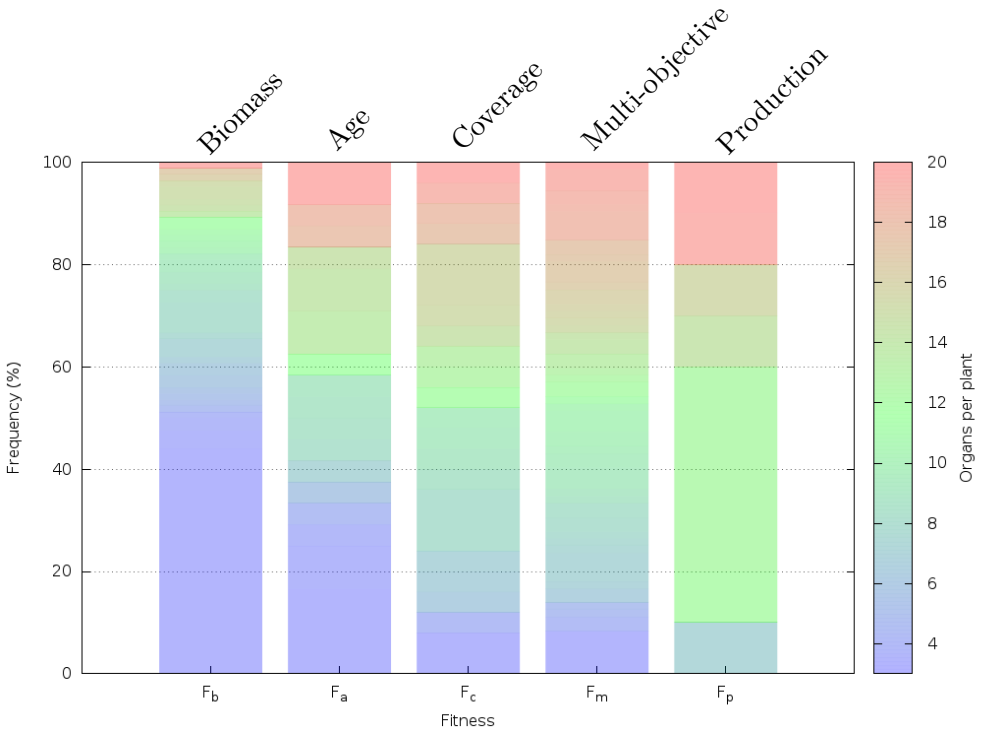


Figure 4.5: Repartition of average organ count per fitness across all runs.

champions being comprised of less than five organs. Given that every random graphtal starts with this specific amount, it shows that evolution discovered that the bigger one wants to grow, the smaller the genotype.

A similar trend can be observed in F_a, though with a slight offset caused by the necessity of having sexual organs. The global strategy for this fitness is as described in details for individual 4.3g: small genome, small plant, fast mating. This tendency is reversed in F_p, with no instance in the 'Minimalist' section of the phenotypic space. Indeed, as glucose production requires both efficiently positioned leaves and sufficient water uptake, evolution favored genotypes with repetitive structure. On the downside, this also led to extremely long evaluation times which prevent all runs to reach 250 generations with no overwhelming advantages over the competing fitnesses.

The coverage-oriented evolutions performed by F_c led to a more balanced distribution of organ count between minimalism and over-complexification. While this is the less biologically inspired criterion, it proved more robust to being exploited by the evolutionary algorithm and, paradoxically, brought more life-like individuals about, such as 4.3a. Finally, the multi-objective fitness F_m generated more all-rounder creatures, that did not suffer from over-optimization. Indeed when looking at 4.3c, 4.3b or 4.3f, one can observe plausible morphologies made functional by the contradictory pull of all individual fitnesses. Furthermore, it settled in a complexity landscape similar to that of F_c, though with less exploration of the uppermost region.

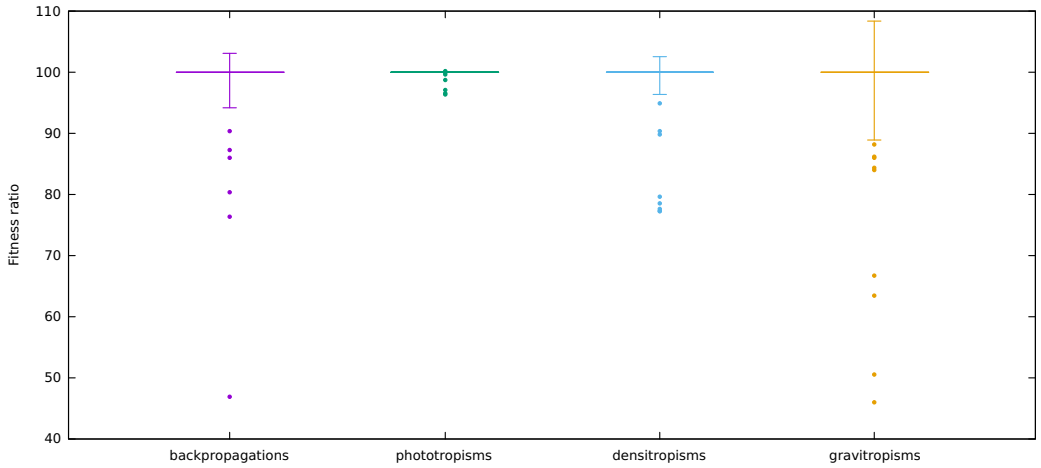


Figure 4.6: Impact of the adaptive mechanisms on the fitness. More details can be found in appendix D

4.2.5 Adaptivity

Another point of interest was the use of the backpropagation and various tropisms to sense the environment both for nutrients and competition. The population was dominated by genotypes with non-zero values which, given the initial null value, demonstrated an exploration effort on the part of the evolutionary algorithm. This, however, did not imply that they had any impact on the phenotype (e.g. a photo-tropism on an underground link).

To measure how well the evolution managed to find meaningful values for these adaptive behaviors, we reevaluated each genome after disabling either the backpropagation (plant level) or all tropisms for a given stimulus (link level) and measured the resulting variation in fitness. That is for F_1, F_2 the fitnesses obtained with the base and altered genomes, respectively, we computed F_2/F_1 to determine how much of a loss such deactivation induced. As can be seen on the boxplots of figure 4.6, the average InterQuartile Range of the fitness variation observed for the organisms that exhibited non-zero values is tremendously low (0.07%). In fact almost all reevaluations were squeezed into a narrow band around 100% (i.e. no fitness variation).

For the handful of individuals that performed differently, we can see that the photo-tropism is the least meaningful one with the most impacted simulation still producing 96% of its original performance. This could, in part, be explained by the relatively uniform access to sunlight in the environment where vertical competition did not have time to develop.

On the other side of the spectrum, both gravitropism and backpropagation proved proportionally more useful with a maximal fitness drop exceeding 50%. The former was used in its negative form to provide an upward growth of above-surface organs thus increasing light absorption while the latter was used to redirect resources towards expansion by reducing the communication

canals with less growth-oriented parts of the plant. Furthermore, in the case of individual 4.3e, loss of this component induces a rapid decline by the 150th day which leads to an untimely extinction event a year later, showing that for some elements of the genetic space this parameter can be a vital component of the metabolism.

4.2.6 Reproduction

Given low initial performances, in terms of self-reproduction, we extended our experimental protocol to test for two hypotheses on the emergence on self-sustainability:

Hypothesis **F** Fruit dissemination should be supported by the system until autonomous abscission[3]can stabilise.

Hypothesis **D** Plants should be stressed by the unavoidability of their deaths.

The former is implemented through a collection algorithm that retrieves any fruit disconnected from its plant (either voluntarily or through a parent's organ death) and proceeds to its dissemination through the usual algorithm. The latter is emulated by extending the simulation duration s_d when in presence of self-reproductive behavior. That is, for a number g of autonomous generations, the allotted number of years is $s_d = min(10, 2max(g, 1))$. This allows genomes exhibiting self-reproduction capabilities to reach much higher fitness values thus increasing their chance of producing offspring in the next generation of the evolutionary program.

We tested all four combinations to see how this would influence the capacity to develop self-reproduction (fig 4.7) using the following notation:

(f,d) Baseline (as in the previous setting)

(F,d) Fruits dissemination is facilitated

(f,D) Multiple generations are rewarded

(F,D) Both facilities are active

The initial conditions (f,d) proved quite detrimental for the emerge of self-sustainability: indeed, aside from a single outlier reaching 7 autonomous generations, 93% of the individuals obtained do not reproduce at all. Only introducing adaptive simulation duration (f,D) reduces this number to 84% and creates a secondary behavioral cluster around the 4th generation. This dynamic is inverted when only collecting immature fruits with only 54% of non-reproducing runs and 38% reaching the first generation. Finally when both features (F,D) are enabled the threshold of less than 50% of infertile individuals is crossed, albeit slightly, and peak performance is at absolute maximal across

[3]The process by which a plant spontaneously shed parts of its structure as with dead leaves or ripe fruits.

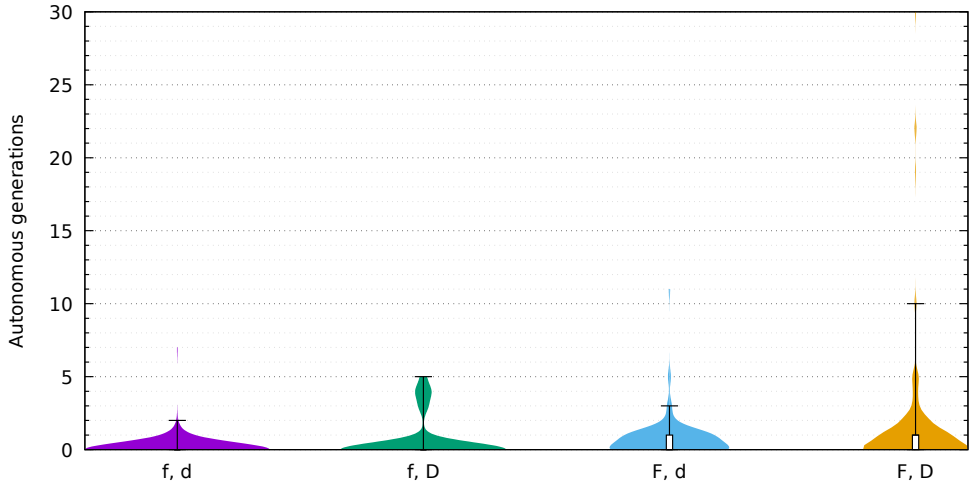


Figure 4.7: Violin plots of the number of autonomous generation per run type (see the text for details). The colored area displays the kernel density overlayed with a boxplot with whiskers spanning the 95% confidence interval.

all alternatives (19th, 22nd and 30th autonomous generations obtained in the 10 years allotted time-frame).

Based on the capacity to produce self-reproducing plants and the results of T-Test evaluations we can surmise that run (f,d) is outperformed by every other alternative (p-value ≤ 0.001). Additionally, while no significant differences were detected between runs (f,D) and (F,D) the last one (F,D) shows better results (p-value < 0.01) than both of them.

In order to further understand the dynamics behind these differences in self-reproductive behavior, we investigated what 'checkpoint' individuals tended to stop at (fig 4.8). The different categories are

- None: No sexual organs were produced by the plants
- Flowers: Some were produced but never fecundated
- Fruits: Seeds were produced but never planted
- Repro.: Self-reproduction occurred.

At first glance, we can note different locations for the point(s)-of-failure depending on which Hypothesis was enabled. While run (f,d) seems to struggle at every checkpoint, loosing almost a third of its population each time, run (f,D) understood the importance of producing fruits but rarely found how to disseminate them into the world. Both (F,d) and (F,D) show no particular problem on this point due to the algorithm taking the lead when necessary and, instead, are clustered between individuals that do not attempt to reproduce and those that succeed.

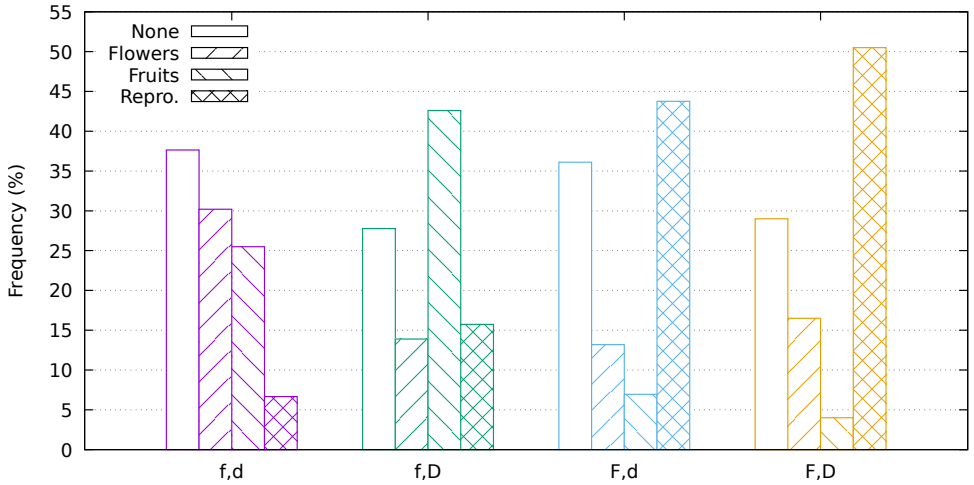


Figure 4.8: Repartition of checkpoints for autonomous reproduction. Colors are the same as 4.7, for details of the checkpoints see text.

In the end, both hypotheses are verified by these additional runs as, on the one hand, emphasizing the need to circumvent individual death promotes self-replication while, on the other hand, providing a fallback mechanism, until self-controlled abscission can stabilise, leads to fruit generation.

4.3 Conclusion

In these experiments we improved upon the mode described in the previous section, reducing the computational cost of the individuals to shift the point of focus towards larger populations. Though this also resulted in less complex interactions between the plants and the environment, this allowed for the emergence of inter-plant dynamics that would have required prohibitive simulation times to achieve in the original version. A more concise and self-contained version of this work can be found in the corresponding article [Godin-Dubois et al. 2019b], with the videos of the exhibited individuals being available at `https://vimeo.com/album/5075632`.

By devising a polyvalent reproduction scheme, especially with respect to the compatibility function, we took the first step towards autonomously speciating populations of artificial plants. The lack of self-sustainability observed, notably in the baseline conditions, was thoroughly investigated and underlined the need to provide a strong incentive to escape the local minimum of survival-oriented solutions. Indeed both hypotheses regarding improvement of this low performance were validated and confirmed that, in the current setting, self-reproduction was a much harder task that initially planned.

Furthermore, we observed that the genetic space was large enough to provide an equally large number of viable demeanor, albeit most functioning as local optimum, preventing the obtained individuals to evolve much further. Given the

objective of this body of work, both this problem and that of self-sustainability needed to be addressed in order to make definite progress towards the validation of our environment-driven evolution hypothesis, especially when considering the somewhat static state of our current abiotic component.

In the next two chapters, we tackle these issues first by considering how to monitor evolutionary dynamics over sufficiently long time ranges so that individuals can no longer be the unit of interest. We then use this methodology to explore a radically different model of ecosystem able, within reasonable times, to produce such dynamical environments and draw conclusions on the viability of this approach.

Chapter 5

Phylogenetic monitoring

Abstract In artificial life, expressing the relationship between different clusters of individuals is constrained, unlike its biological counter-part, by an excess of information not least of all the complete genealogical tree. In the APOGeT tool we allieviate such concerns by processing such parent-child relationships as a stream of data from which a phylogenetic tree is built in real-time alongside the monitored evolution. It relies on the concept of representative-set, a small collection of genomes that best described the diversity and boundaries of a species and is capable of describing both anagenesis and cladogenesis events. Visual tools have also been developed to provide easy-to-read rendering of the inherently complex resulting trees. A discussion of the three key parameters of the algorithm concludes the section and provides hindsight on how to tweak the tool to better suit a given combination of genome/experiment/expected dynamics.

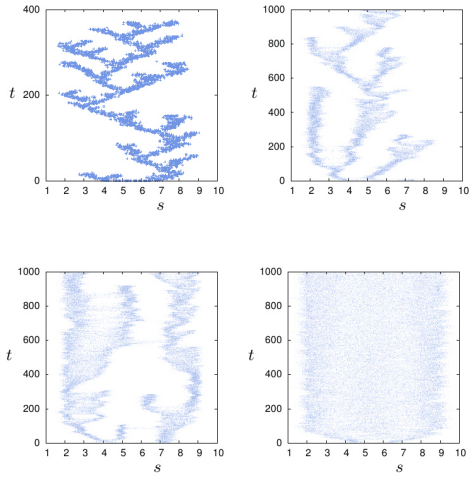
Résumé En Vie Artificielle, la difficulté à mettre en évidence les relations entre différents groupes d'individus est contrainte, à l'inverse de la Biologie, par un excès d'informations, notamment par la disponibilité de l'intégralité de l'arbre généalogique. Grâce à l'outil APOGeT, nous atténuons le problème en traitant ces relations parent-enfant comme un flux de données à partir duquel un arbre phylogénétique est construit en temps réel, en parallèle de l'évolution monitorée. Il s'appuit sur le concept d'un ensemble de représentants, une collection de génomes succincte qui décrit au mieux la diversité et les limites d'une espèce et qui est capable de décrire les instances d'anagenèse et de cladogenèse. Des outils visuels ont aussi été développés pour fournir un rendu facile à lire de la complexité des arbres phylogénétiques résultants. Une discussion relative aux trois paramètres clefs de cet algorithme vient conclure cette section et apporte une plus large compréhension pour ajuster l'outil au mieux selon les combinaisons de génomes/expérience/dynamiques attendues.

Speciation is one of the cornerstones of Darwnian evolution, through which homologous population differentiate into niche-fitting, diverging phenotypes. Though of great interest in researches on biological evolutionary dynamics, many of the underlying phenomenon remain either unexplained or lack consensus from the community at large even on something as primordial as the definition of a species. Indeed [Singh 2012] enumerates more than twenty definitions across time and fields. Similarly the mechanisms inducing speciation are subjected to similar controversy with the previous author cataloging seven modes with some overlapping, while [Butlin et al. 2008] rejects the very concept of quantifying such a continuous class of phenomenon:

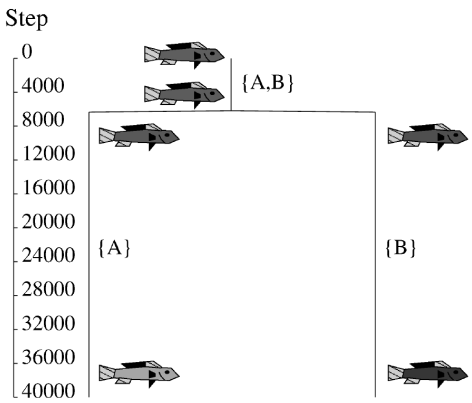
> *"This classification is unsatisfactory because it divides a continuum into discrete categories, concentrating attention on the extremes, and it subordinates other dimensions on which speciation processes vary, such as the forces driving differentiation and the genetic basis of reproductive isolation."*

Artificial Life experiments can come in handy in such situations either by allowing exploration of extremely long phenomenon in reasonable time or by expanding the search space to novel situations to infer general rules. There has indeed been a number or work on the concept of speciation, one of the first being more akin to an exploitation of the niching process by which an individual, by virtue of its adaptation to specific external conditions, can thrive where cousin species cannot. In this context, species are artificially created and maintained as repositories of diversity for generic optimization algorithms as in [Oei et al. 1991] where niching is combined with tournament selection to produce multiple sub-populations of solutions. Similarly they have been used in the various "Augmenting Topologies" algorithms of both NEAT [K. Stanley et al. 2002] and GRNEAT [Cussat-Blanc et al. 2015]. In both instances fitness sharing is used in combination with an adaptive species count to maintain diversity, thus preventing premature convergence. A single representative is maintained from the previous generation to synthesise the genomic characteristics of the species. The difference comes from the compatibility function used to determine whether to put two genomes in the same species: historical markers, that is mutation identificators, in NEAT, and the actual protein-protein distance in GRNEAT.

However speciation as a property of a living population is more closely related to the work we set out to do in this manuscript. Indeed there has been research on observing multiple species and their interaction in a number of settings. The seminal work of [Sims 1994a] used such a paradigm where hard-coded pairs of species competed with one another in an implementation of the Red Queen Effect. Co-evolution of population were also used in [Graham et al. 2007] with experiment-designed migration events from the original niche to an foreign one. Confrontation was devised in [Miconi 2008a] in his "Evosphere" where the two species had to thrive by defeating their competitors either by direct confrontation or successful evasion.



(a) Phenogenealogic trees [Hamann 2013]



(b) Post-processed phylogenetic tree [Clement 2006]

Figure 5.1: Examples of artificial phylogeny in the literature. See text for references

More rarely researches involved the unbridled emergence of multiple species with only the speciation mechanism under (indirect) supervision of the experimenter as the work of [Metivier et al. 2002] in LifeDrop which relied on stress to trigger reproductive isolation.

Further complexity arises when trying not only to obtain speciation but to monitor its dynamics throughout evolution. In small case systems, it is possible to visualize the whole population across time with respect to appropriate parameters so that patterns are clearly visible. For instance, in [Payne et al. 2007] a single allele is tracked through space and time, however this method is clearly not trivially scalable.

A more promising approach was performed on a model inspired by the natural system of finches on the Galapagos Islands [Hamann 2013, 2015; Woehrer et al. 2012]. In this work, there is no explicit speciation model. Instead the only variable of interest (beak size) is undirectly tied to the reproductive success of the individuals. Cladogenesis[1] is, thus, expected but not enforced. Its appearance is monitored through the use of the so-called phenogenealogic trees illustrated in figure 5.1a where the evolution of beak size is plotted against time, showing a continuous range of speciation patterns: from the clearly-marked at the top left to the uniform, single-species, at the bottom right.

Another perspective was put forth by [Clement 2006] in A-Life experiments modelling the speciation capabilities of African Cichlid fish. As depicted on figure 5.1b, the author produces phylogenetic trees, in the typical sense of the term, of the evolved population. The major drawback of this approach, however, was the use of the complete genealogic tree to perform a post-simulation

[1]Process by which a species produces a secondary strand

clustering algorithm which implies heavy computational and memory usage.

In this work, as we needed to monitor long-term evolution of unbridled populations, we set out to designing a polyvalent tool that would perform its task online, i.e. during the simulation, with limited overhead cost. First and foremost we rely on a fuzzy definition of species as worded in [Singh 2012]:

> *"[...] group of potentially interbreeding natural population reproductively isolated from other such groups."*

The specifics of how we model a single species, assign genomes to it and decide when to trigger cladogenesis events are detailed in following sections.

5.1 Core

5.1.1 R-Set

Using centroids or other forms of aggregate data is of particular relevance when performing unsupervised clustering of large databases (see [Zhang et al. 1996] for instance). However, this is only relevant in those cases where the manipulated data follows an appropriate distribution. Centroids, for instance, require that the mean of the observations has some form of semantic in the same space from which these observations where drawn.

When manipulating discrete values such as integers or even enumeration, such an aggregation becomes essentially meaningless. Furthermore, the assumption that the underlying distributions are of a simple form (e.g. normal) does not necessarily holds, especially in the context of this work, where speciation[2] implies multiple clusters of well-explored genetic regions.

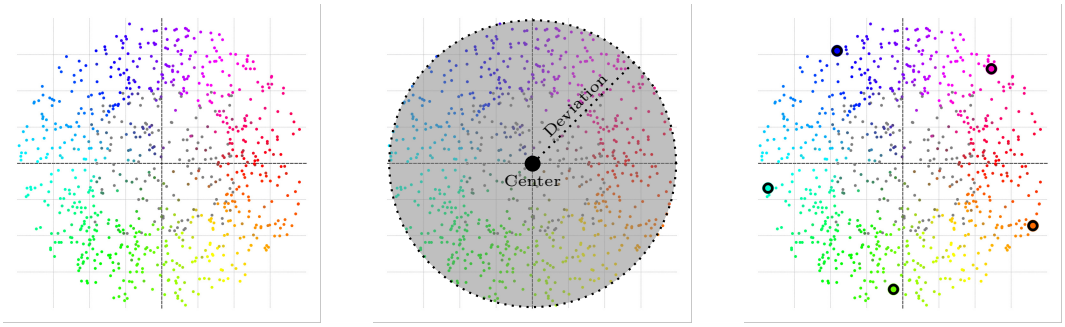
We also argue that the speciation model used in [Cussat-Blanc et al. 2015] would fail to encompass certain species boundaries composed of heterozygous populations. Indeed the species shown in figure 5.2 cannot be accurately described by any single data point: its main characteristic is its internal diversity. In fact, individual variation is a prerequisite of natural selection:

> *"[...] I look at individual differences, though of small interest to the systematist, as of high importance for us, as being the first step towards such slight varieties as are barely thought worth recording [...]" [Darwin 1859, p. 63]*

Concept

Thus individual differences should be considered as of relevance to the evolutionary process as they might be indications of potential speciation directions or plain fine-tuning to local conditions. To address this challenge of maintaining

[2]Evolutionary process by which populations evolve to become distinct species



(a) Heterozygous species (b) Centroid (c) Representative set

Figure 5.2: Heterozygous species can be compactly and accurately described by a representative set of sufficient size

information on the allelic diversity of a species we rely, in this tool, on a collection of representatives: the R-Set. The main dynamics of this concept are detailed in figure 5.3

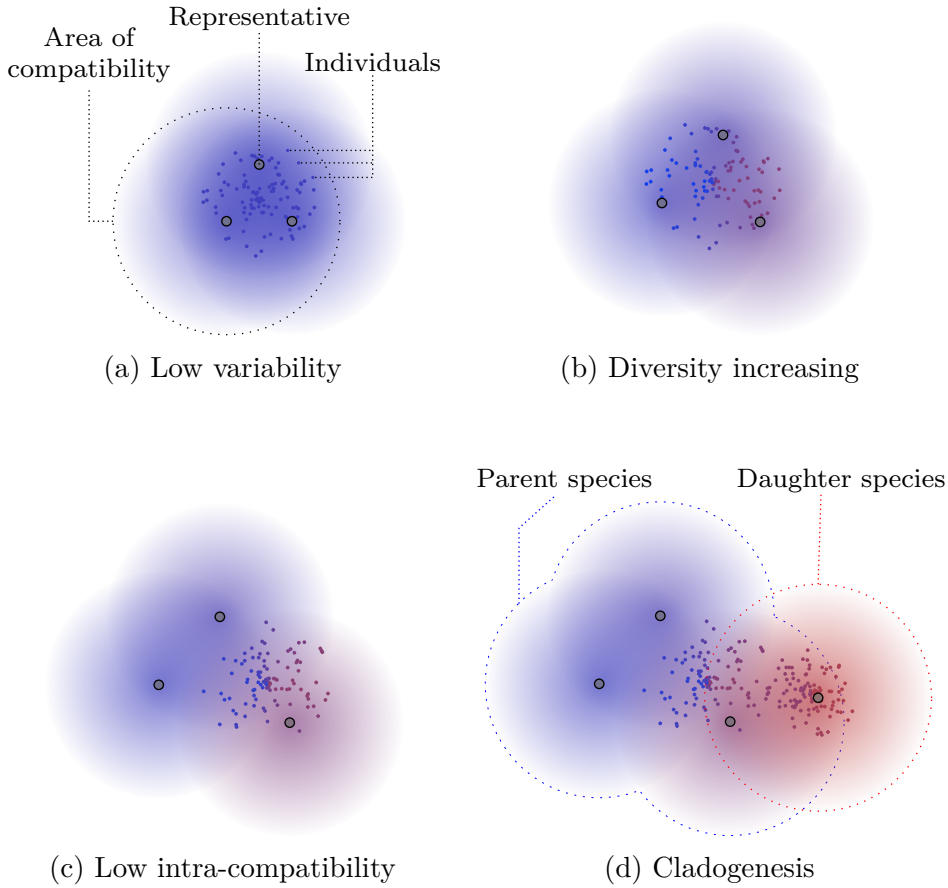
Our model assumes an initial, somewhat uniform, primordial collection of individuals which forms the Last Universal Common Ancestor (LUCA[3]) population. Indeed in all experiments described in this work, the LUCA is a single progenitor genome which is cloned (without mutation) to produce the first plants.

In figure 5.3a, a cluster of such individuals is represented by blue dots. These abstractions, by virtue of being close together, do not have much inter-individual diversity, thus forming a tightly-bound species. In addition, the larger, gray, dots represent the member of the R-Set, just as uniformly placed as the population they describe. Each representative has a so called *Area of Influence* (AoI), i.e. the region in genetic space inside which it can find suitable mates (according to a compatibility function $\mathfrak{C}$, described later on). The intersection of each of these AoI then defines the subset of genetic space in which this species resides, for a given threshold of overlapping.

But this method is meant for online observation of speciation patterns and thus the dynamics of the underlying population impact the species boundaries, through the R-Set. Indeed, in figure 5.3b, we can see that, as members of the considered species become more different from one another, their representatives change as well to reflect this variation in 'typical' features. As more and more diversity emerges, so does the R-Set grow, but only up to a point. One can see in figure 5.3c that the R-Set does not only describe the currently alive individuals of a species, but also their evolution across time. As the starting point was solely composed of blue centered dots, so does the R-Set maintains a trace of this information.

Hence, after a sufficient degree of evolution of a species contents, its bound-

[3] The Last Universal Common Ancestor is the most recent population of organisms from which all currently observable organisms have a common descent



(a) Low variability

(b) Diversity increasing

(c) Low intra-compatibility

(d) Cladogenesis

Figure 5.3: Implicit species boundaries by means of an R-set

aries will become overly stretched. At this point, diverging individuals will no longer represent new features of the current species but, instead, the seedling of a new subspecies as illustrated in figure 5.3d. The population of red dots, being too far outside the species boundaries of the parent species, are being placed into a newly created daughter species, thus describing a cladogenesis event.

Implementation

More formally, given a genetic space G and a distance function $D : G^2 \to \mathbb{R}^+$ between two genomes, we defined the update procedure of an R-Set RS of target size K as detailed in algorithm 4

Where the individual contribution of the evaluated genome g with respect to an existing representative e_i is given by:

$$\begin{aligned} C(RS, r, g) = &- \min_{r_i \neq r} D(r, r_i) \\ &+ \min_{r_i \neq r} D(g, r_i) \end{aligned} \tag{5.1}$$

We first determine how much more distant g is from the rest of the R-Set's contents. Whenever we find g to be more deviant from the norm than is currently represented, we compute the member of the R-Set that induces the least amount of variation with respect to g and replace it.

This straightforward method allows for a linear (in K) amount of computation of the genetic distance, when caching the results from previous insertion. However, the number of actual comparisons is on the order of $O(K^2)$.

5.1.2 Species affectation

Thanks to our reliance on a representative set, affectation of an individual genome to its appropriate species becomes a relatively straightforward procedure, as outlined in algorithm 5. This relies on a compatibility function (hereafter denoted $\mathfrak{C}$) in the same form as the one implemented by our Bail-Out Crossover (BOC, as defined in section 4.1.4). This function is in the form $\mathfrak{C}$: $G^2 \rightarrow [0, 1]$, that is we can query the compatibility between any pair of genomes in G. As $\mathfrak{C}$ is not expected to be symmetrical (and indeed the implementation used with the BOC operator is not) we rely on the cross compatibility as defined in equation 5.2.

$$C_x(g_0, g_1) = min(C(g_0, g_1), C(g_1, g_0)) \tag{5.2}$$

Indeed we interpret, in the definition of biological species as quoted in the forewords of this chapter, a *"potentially interbreeding natural population"* as *individuals with a cross-compatibility above a given threshold*, T. From this we derive the matching score of a given genome G with the representative set RS of a target species as:

$$matching(g, RS) = -T + \frac{1}{|RS|} \sum_{r_i \in RS} C_x(g, r_i) \tag{5.3}$$

The insertion procedure then boils down to determining which amongst the candidate species is the best match for the genome we want to insert. We

Data: g, genome; RS, R-Set
if $|RS| < K$ **then**
  $RS \leftarrow RS \cup \{g\}$;
else
  $c \leftarrow \max_{r_i \in RS} C(RS, r_i, g)$;
  if $0 < c$ **then**
    $r \leftarrow \arg\max_{r_i \in RS} C(RS, r_i, g)$;
    $RS \leftarrow (RS \setminus \{r\}) \cup \{g\}$; `// Replace` r `by` g
  end
end

Algorithm 4: R-Set update procedure

```
1  affectSpecies(g: genome, T: phylogenetic tree)
       /* Determine if parent species is a match                    */
2      p ← parentSpeciesId(g);
3      S_p ← node(T, p);
4      S ← S_p ;                                  // Target species
5      M ← matching(g, S_p.rset);
6      if M < 0 then
           /* Otherwise search in subspecies                        */
7          matches ← (S_i ∈ S_p.subspecies / 0 < matching(g, S_i.rset));
8          if |matches| > 0 then
9              S ← matches_1;
10         else
               /* Otherwise defaults to a new species               */
11             S ← newSpecies() ;
12             S_p.subspecies ← S_p.subspecies ∪ {S};
13         end
14     end
15     updateContents(S,g);
```

Algorithm 5: Species affectation. See text for details

start by retrieving the node corresponding to the parent of g for the hypothesis underlying our clustering algorithm is that individuals will either be in the same species S as their parents or in a *direct* subspecies S_i. While the former makes sense from a biological point of view, the latter is more counter-intuitive. Indeed, given that we aim at transforming an altogether continuous genealogic tree into a more discrete phylogenetic one, we end up erecting species boundaries between direct descendants. However, though by definition arbitrary, this is a problem intrinsic to any algorithm that attempts to categorize collections of elements for which no natural partition exist.

We first consider whether S is a correct matching: if so the procedure is complete and, for an R-Set size of K, this required exactly K comparisons. If the score for S is lower than 0, however, we need to iterate amongst the subspecies. From there two situations can occur: if there exists at least one species with an above-zero matching score, we immediately stop the search and consider it the target species. If none can be found then a new, atomic, subspecies is created with g as its (currently) sole member and representative. With time those skeleton-species will be populated by more individuals as evolution proceeds in similar directions. There is however the possibility of producing 'stillborn' species as mentioned in more details in section 5.2.2.

The procedure ends by an update of the internal fields of the target species including, but not limited to, the R-Set.

5.1.3 Hybridism

Despite its straightforwardness, this algorithm suffers from one major drawback: the implicit postulate that individuals will always be bred from same-species parents. In the previous section we already mentioned one case where this would not hold: genetically similar individuals across an arbitrary species barrier.

Another such case, which proved rather common in our experimental settings, is that of an actual inter-species crossing. That is individuals from more distantly related species that successfully produce offsprings, however infrequent. Such occurrences stem from our reproduction algorithm which does not imposes harsh boundaries an indeed reaching a compatibility of zero is only possible trough numerical approximations. Nonetheless, this pattern was deemed worthy of investigation and we devised a limited number of alterations to the basic algorithm to cope with such cases.

Though the general structure of the algorithm remains unchanged the specifics of the first two portion needed much adaptation. Algorithm 6 shows the updated version with bracketed line number indicating some modified or new content. Indeed the code is written so that individuals from same-species parents will follow the same path as in the previous version. However, whenever $p_m \neq p_f$ we first need to determine if one of the parents' species is a good enough match. In that case, the best match is defined as the target species and the update procedure can resume as previously, with the exception of the additional c parameter which will be discussed further on.

If neither parent species are valid candidates, we search as previously in the *direct* subspecies and stop at the first one with an above zero score. It is worthy of note that the iteration procedure alternatively examine subspecies belonging to both parents so as no to give undue advantage to the species of the maternal parent. Similarly the creation of a new species when no sufficient matching could be found follows the same pattern as previously though now we assign the subspecies to the parent with the highest score with the assumption that subspecies are more similar to their ancestral species than to "cousins". This holds in the general case where no evolutionary convergence exists.

But being able to associate a hybrid individual to an appropriate species is only part of the problem. Indeed, this hybridation process is in itself an interesting subject of inquiry. To this end, we keep track when processing each genome of the provenance of their alleles or, in other words, their parents' species. This is the role of the variable c, affected at line 4 in the algorithm: we record one additional contribution for each of the ancestors of the individual.

With such information, we can build for each species the collection of "allelic contributions" which indicates the frequency with which genetic material is being used by such other species. More formally, this is an ordered set of (S, c) pairs where the former value is a species identificator and the latter the contribution count. The contribution pool for a given species S_i is thus given

```
1  affectSpecies(g: genome, T: phylogenetic tree)
       /* Determine if parent species are a match                    */
[2]    p_m ← parentSpeciesId(g.mother);
[3]    p_f ← parentSpeciesId(g.father);
[4]    c ← ((p_m, 1), (p_f, 1));
[5]    S^p ← {node(T, p_m)};
[6]    if p_m ≠ p_f then
[7]        S^p ← S^p ∪ {node(T, p_f)}
[8]    end
[9]    for S_i ∈ S^p do M_i ← matching(g, S_i.rset);
[10]   b ← arg max matching(g,S_i.rset);
[11]   S ← S_i;
[12]   if M_b < 0 then
           /* Otherwise search in the direct subspecies              */
[13]       Sub ← ⋃ S_i.subspecies;
[14]       matches ← (S_j ∈ Sub / 0 < matching(g, S_j.rset));
15         if |matches| > 0 then
16             S ← matches_1;
17         else
               /* Otherwise defaults to a new species               */
18             S ← newSpecies() ;
[19]           S_b.subspecies← S_b.subspecies ∪{S};
20         end
[21]   end
[22]   updateContents(S,g,c);
```

Algorithm 6: Species affectation with hybridism management. See text for details

by:

$$C_i = \{(S,c)_1, (S,c)_2, \ldots, (S,c)_n\} \quad c_1 \geq c_2 \geq \cdots \geq c_n \tag{5.4}$$

In this manner, we have at all time enough information to infer the interaction graph of all species. However, in many regards, the increase in information provided by graphical relationships has similar impact on the complexity of the algorithms involved for data extraction and processing. By prioritizing clarity, we decided to keep this tool as a phylogenetic tree extractor and, instead devised from the contribution pool the means to define the parent species of a node S_i in a robust manner even in the face of variable gene-flow.

To this end, we consider the first member of $(S_j, c_j) \in C_i$ for which $S_i \neq S_j$. Given that C_i is ordered by the number of contributions, this equals to selecting the species, different from the one under scrutiny, that provided the most genetic material. We define the current parent of S_i as this species: indeed

during the initial steps of a cladogenesis most deviant individuals result either from pure breeding for the ancestral species of by hybridation between both parent and children species.

Additionally, if the set C_i evolves and the major contributor for S_i changes, due to underlying dynamics in the gene-flow, we re-parent it under the newly computed parent species. It follows that, in our phylogenetic model, a parent-child relationship denotes not necessarily that one was obtained from the other but rather that incoming allelic material was predominantly obtained from this source.

This concept of *major contributor* allows the use of a tree representation of the phylogenetic data while, at the same time, maintaining partial information on the hybridation process.

5.2 Extensions

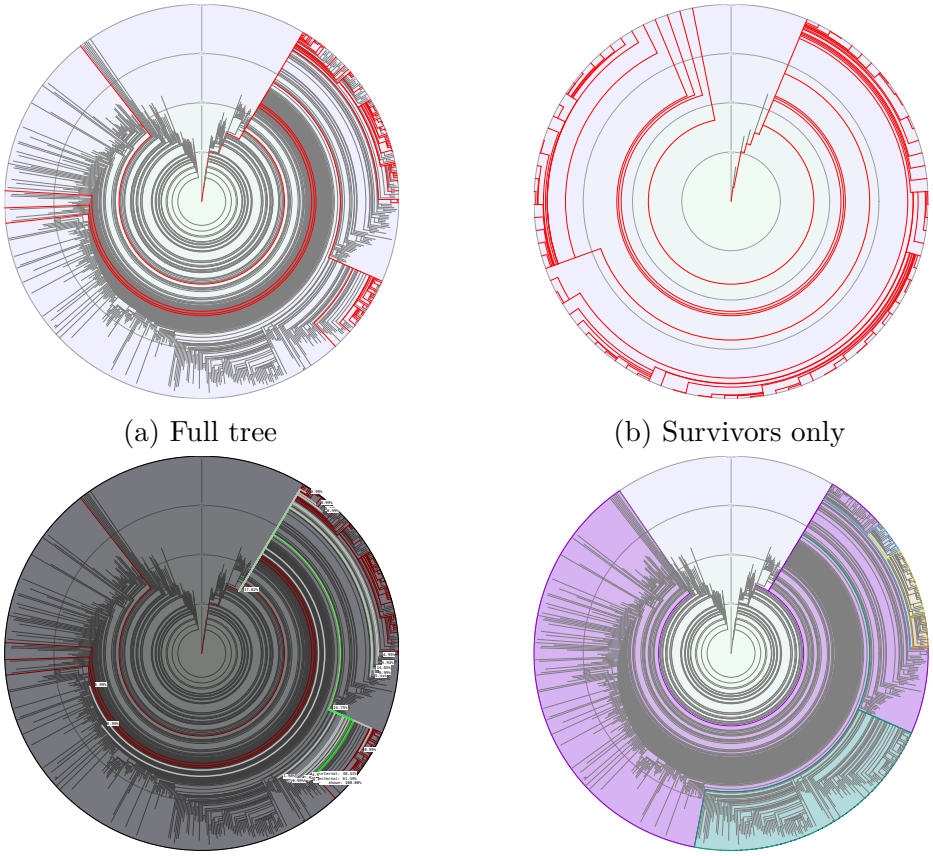
5.2.1 Visualizations

In addition to the core components of APOGeT, a collection of visualization tools have also been designed. The pictures in figure 5.4 display some of the graphical capabilities of our library.

The most basic of these tools is the raw display of all species encountered during the simulation as shown in figure 5.4a. Each node condenses a number of elements that both specify its place in the phylogenetic tree and provides information about the species. The most obvious of these are the apparition date (in ticks) which is used to position the node itself, starting from the initial node at date 0. The last observation date stores either the last timestep at which the species was observed or the current date for those that still exist in the environment. From this data, we plot the (radial) timeline of each species while, on the other side, and arc connect the node with its parent, at the recorded time of cladogenesis.

In addition, we single-out living species, that is those whose timeline reaches the outer boundaries. This allows us to defined species on the 'survivor path' as the set of species that are either alive or an ancestor of one such species. The paths corresponding to these are, by default, displayed in red to show, at first glance, the different speciation path with respect to the global pattern. Individual nodes can also be restrictively displayed based on this criteria to provide a simplified, albeit easier to process tree, as illustrated on figure 5.4b. Given that only those species who participated in the final state of the population are shown, the cladogenesis event are made easier to observe with respect to one another.

As mentioned in the section on hybridism, we keep track of the gene-flow between species which we can extract, not only in a data-processing manner but for graphical display as well, as shown on figure 5.4c. For the species under scrutiny, in the lower left quadrant, the various incoming allelic material is



(a) Full tree (b) Survivors only

(c) Hybridism (d) Taxonomic levels

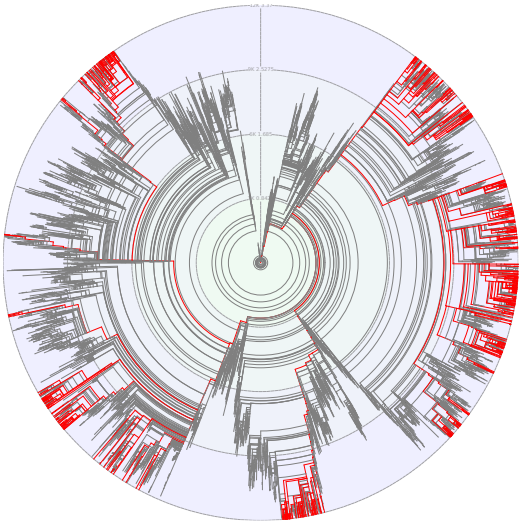
Figure 5.4: Examples of visualisations

represented by path of increasingly saturated green with the actual amount printed near the corresponding species. In this manner, we can, with a glancing look, determine which portions of the simulation-wide gene-pool contributed to that of the investigated species and in which proportions. We also detail how much gene mixing was performed internally and externally.

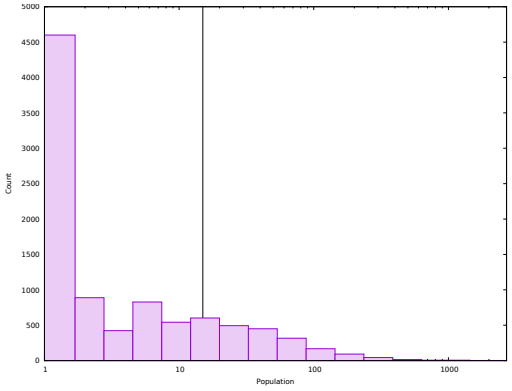
The last visualization tool is one of taxonomic levels visualization (fig. 5.4d). Through this functionality one can track a collection of species throughout the time and easily visualize the relationship between various parts of the phylogenetic tree. The increasingly larger taxa are automatically built from the ancestral relationships between each tracked species, allowing for the monitoring of the repartition of a given subset of the global speciation pattern.

5.2.2 Stillborn trimming

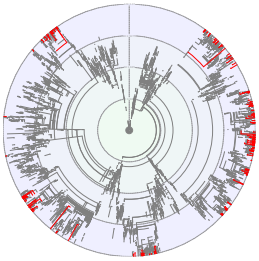
Throughout the various instantiations of our phylogenetic trees in multiple experimental settings, we observed a recurring pattern. Not only does it hinders the clarity of the produced data but it also exacts a heavy toll on the computational cost of the species affectation protocol.



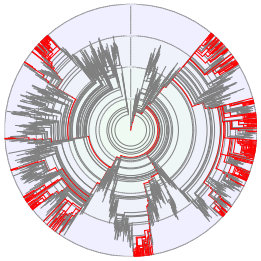
(a) Phylogenetic tree with multiple evolutionary trajectories



(b) Species population count repartition



(c) Under-populated



(d) Information-carying

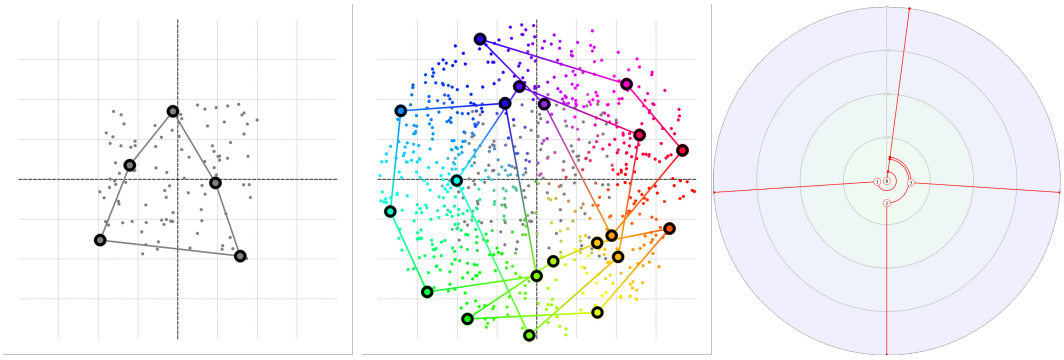
Figure 5.5: Bloating pattern: the majority of species are of no relevance to the evolutionary process

We assimilate this as a bloating process as illustrated by figure 5.5. From the phylogenetic tree obtained at the end of a simulation, we extracted the repartition of species with respect to the number of individuals they have been assigned to in figure 5.5b. As one can see from this histogram, most of the species (77.3% in this exemple) contained less than 15 individuals, which was the size of R-Set. This only leaves 22.7% of relatively well defined species.

More strikingly, when looking at the lower row, we can see that the observed phylogenetic tree can be decomposed in two portions (fig. 5.5c & 5.5d), the under-populated species and those containing information on the speciation process. Indeed the pattern shown by the complete tree is almost identical to that of the smaller subset, indicating that the "bloating" part plays no functional role in the simulation.

This might only be a side-effect of the mutation operators that, by producing strongly diverging offsprings, force the clustering algorithm to encapsulate these mutants into "atomical" species. Nonetheless, we devised a counter-measure aptly named *stillborn-trimming* to limit the proliferation of such occurrences.

The procedure examines, every T_n steps, all of the current leaves of the phylogenetic tree for which the R-Set contains less than K_t elements. For such a species S_i with first and last recorded occurrences being given by t_f and T_l, respectively, we compute the lifespan ($s_l = t_l - t_f$) and "deathspan"



(a) Initial configuration (b) Final configuration

Figure 5.6: Correct parameters successfully extract four well defined species

($s_d = t - t_l$). From there we determine whether S_i is a "stillborn" species by testing whether

$$max(M, s_l * D) < s_d$$

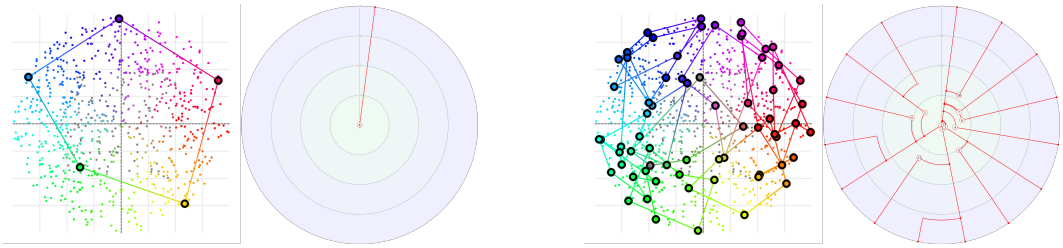
with $D \geq 1$ the trimming delay, indicating how soon after extinction a species should be culled, and M the minimal number of steps before a species is considered by the procedure. Nodes falling under this condition are removed from the tree, all of their contents being destroyed to focus the computational and memory footprint towards the so-called information-carrying species.

5.3 Key parameters

Though only mentioned in passing, up to now, this algorithm relies mainly on three parameters to perform its clustering task in optimal conditions: the belonging threshold T, representative set size K and compatibility function $\mathfrak{C}$. In this section we detail how they impact the process of segmentation into distinct species and how best to set them up.

Beforehand, we bring attention to figure 5.6 which will serve as the baseline for illustrating a functional parameter set. The initial configuration, on the left-hand side, shows a handful of individuals whose genotype is the tiniest of all: a 2D position in $[-1, 1]^2$ and an RGB color in $[0, 1]^3$. Given their proximity and uniformity of color they form a single species, with the larger dots corresponding to representatives and the lines between them highlighting their affiliation to the same set.

On the right hand side, is displayed the final state of this calibration protocol: a few thousands individuals have been generated following a sympatric speciation pattern and we can see, on the far right, that four additional species have been detected. In the population these species correctly segregate individuals of different colors and positions with the R-Set boundaries clearly discernible despite overlapping of neighbouring species.



(a) $T = .125$ (b) $T = .5$

Figure 5.7: Too small or large values result in too loose or too harsh species boundaries

The parameters used for this baseline speciation are:

$$T = .25$$
$$K = 5$$
$$\mathfrak{C} : BOC^4$$

5.3.1 Belonging Threshold (T)

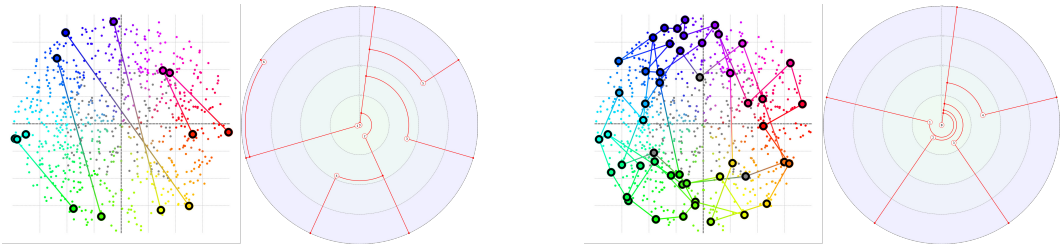
The first parameter of interest is the belonging threshold T which governs the tolerance to variation inside a species. It is also the easiest one to set given its relatively loose tying with problem-specific implementations.

It is interpreted as the minimal overlapping density allowed to consider an individual a member of a given species: as each representative radiates an "area of compatibility", whose shape depends on the compatibility function $\mathfrak{C}$ used, only those regions where the average coverage is above the specified threshold will be seen as inside the species boundaries.

As a matter of illustration figure 5.7 shows the same simulation as presented in the baseline setting with this single parameter altered. One can see that too low values will lead to lenient inclusion of remote individuals those reducing the number of computed species while, conversely, restrictively high values will enforce a very strict segregation scheme. While the looseness of species boundaries is dependent on how much internal variation is expected by the experimenter, one should be mindful that it might also have adverse effect as visible in the strict $T = .5$ case. The size of a phylogenetic node mirrors the fullness of its representative set thus the small peripheral species on the tree denote potential stillborns with very limited information-carrying capacity as previously described in section 5.2.2.

Empirical evaluations found that a value of $T = .25$ was a valid compromise between leniency and harshness, producing reasonably dense phylogenetic trees in actual experimental setups.

[4]Bail-Out Crossover. See section 4.1.4 for its definition



(a) $K = 2$ (b) $K = 10$

Figure 5.8: Incorrect R-Set sizes result in poorly defined species

5.3.2 R-Set size (K)

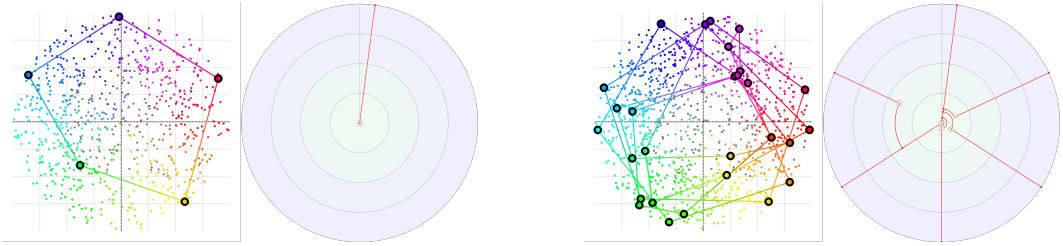
Given that species are solely described by their representative set, its size K is of utmost importance in producing intelligible yet accurate modeling of the underlying speciation process. Unlike the belonging threshold, K is strongly domain-dependent or rather genome-dependent. Indeed in order to accurately describe a heterogeneous population of individuals as sufficiently similar to be called kin, the internal variance of the "gene-pool" imposes an accordingly large collection of typical points.

Considering a most trivial 1D genome of real values, one can see that a minimum of two representatives are necessary to correctly encompass the different possibilities of the various subpopulations. However, the argument also works in reverse: if we were to assign too large a value, say 5, then this redundancy would not only be irrelevant, it might very well be detrimental.

Indeed the R-Set is the sole computational component of this phylogenetic tool: increasing its size necessarily increases the amount of comparisons needed at every genome assignation and the memory footprint of the tree. Furthermore this might throw off the search by imposing undue generality on the species model. As a natural extension of this model, the R-Set not only encompasses the current allelic distribution across the population but also maintains, as a side-effect, a trace of previously used combination. In this sense large values of K will tend to limit the capability to model anagenesis[5], i.e. the ability to detect when a species becomes self-alienated.

Such effects can be seen plainly on the examples of figure 5.8: two representatives is a large undershoot for modelling continuous 5D genomes resulting in very poorly defined species with much uncalled for overlapping. On the right side, the value of $K = 10$, though double the baseline, performs reasonably well with only slightly shifted cladognesis events and an offshoot species, for double the space and computation. This goes to show how easy it is to underestimate the value while the upper bound is much more elusive even in such a test-case scenario. Our experiments, as mentioned in the corresponding annexes, used a value of $K = 15$ representatives which, given the relatively high dimensionality of the involved genomes, was a low estimation.

[5]Whereby a species diverges from its ancestral root, without branching.



(a) $C = 1$ (b) $C = \mathbb{U}(0, 1)$

Figure 5.9: The compatibility function defines the underlying (individual-level) speciation process

5.3.3 Compatibility function ($\mathfrak{C}$)

The last and most problematic of the parameter is the compatibility function $\mathfrak{C}$ itself. As it is tasked with describing the individual-to-individual relationship of interbreeding capabilities and thus the speciation pattern it designs is the keystone of this phylogenetic extraction tool.

Our experiments relied on the function designed for the Bail-Out Crossover, presented in section 4.1.4, which showed both robustness and autonomy but any function in the form $G^2 \to [0, 1]$ could potentially be used. This extends to the alignment methods used in both NEAT and GRNEAT which could then include automated phylogeny monitoring over their optimization process. Similarly, alignment of sequences from Computational Biology could be used to provide low-level perspectives on the species dynamics from the DNA strands point of view.

A word of caution seems nonetheless necessary as $\mathfrak{C}$ will directly determine how well APOGeT is able to extract species, though rarely obviously so. Two very extreme cases of poorly performing compatibility functions are depicted in figure 5.9. On the left, the always-on version returns a constant maximal compatibility between any pair of genomes, resulting in no speciation process being detect: a correct answer on the part of the tool, for indeed no speciation can occur in such context. On the right side, compatibilities are randomly defined, modeling a chaotic system where no pattern ought to be discerned. As expected the species obtained reflect the encoded dynamics: individuals are anarchically bundled into poorly defined and massively overlapping boundaries.

Though caricatural these examples show that the design of the compatibility function is, under usual circumstances, not a trivial task and is highly dependent on the genomic representation and the experimenter's familiarity it has with it.

5.4 Conclusion

In this section we described a tool for monitoring large population of individuals over equally large evolutionary time frames. One of its strongest features is that it does not rely on any particular genomic structure. This makes it equally

capable of processing L-System-encoded plants or Sims-like combinations of Graphtals and neural controllers. Indeed the key algorithmic components only rely on the notion of genetic distance which has already been investigated for a number of substrates including Artificial Neural Networks K. Stanley et al. 2002, Genetic Regulatory Networks Cussat-Blanc et al. 2015 or the Biomorph-inspired creatures of Metivier et al. 2002. Actual computation of par-wise comptability is performed via a triplet of genetic parameters which can be easily introduced in most genetic encodings. In addition, the parameters one must understand to use this tool are only loosely tied to the specifics of the genetic encoding used: the distance/compatibility functions need only provide consistent metrics, the envelope size depends only on the size of the genetic space and the belonging threshold on the radius of species boundaries.

The second desirable property of APOGeT is that it is meant to run alongside a given simulation or genetic algorithm, thereby alleviating the need to store a complete genealogic tree. Indeed, in the following experiments, it was used to produce speciation data on the evolution of up to tens of millions of individuals. Post-processing such a large amount would have required a prohibitive amount of computational power which, thanks to temporal information, was not required with this methodology. Moreover, the graphical components developed alongside the main algorithms provide a straightforward manner to visualize different aspects of the processes both during and after the fact.

We argue that such a tool could be used by a broad audience of researchers first and foremost by those of the field of Artificial Life in the context of which it was developed but that its scope is potentially larger. Indeed, while in the cases presented in this section, speciation was viewed as the result of an autonomous reproduction process, the algorithm is theoretically just as suited to handle data originating from a genetic or evolutionary algorithm.

Chapter 6

Speciation Test-bed

Abstract As was shown in the previous experiments, reaching simulated durations of sufficient length to observe evolutionary trends requires a much more straightforward approach. It follows that we reverted to using L-System in a 2D environment which allowed populations of a few thousands of individuals reproducing autonomously and sexually for thousands of generations. This specific setting explored, on text-book cases of speciation, how to drive evolution through environmental variations. Different replicates were subjected to total geographical isolation, niche formation and partial isolation in order to assert that these resulted in allopatric, parapatric and peripatric speciation, respectively. Thanks to the monitoring capabilities of APOGeT, we showed that not only did the expected patterns emerge but also that low-level observation of individuals species-species interaction showed intricate levels of complexity.

Résumé Comme démontré dans les expériences précédentes, atteindre des durées de simulations suffisantes pour observer les tendances évolutionnaires requiert une approche beaucoup plus directe. Il s'ensuit donc que nous somme revenus vers des L-Systèmes dans un environnement 2D permettant à des populations sexuées de milliers d'individus de se reproduire autonomement durant quelques milliers de générations. Cette configuration, s'appuyant sur des exemples typiques de spéciations, a permis d'explorer comment diriger l'évolution au travers de variations environnementales. Différent duplicats ont été soumis à une isolation géographique totale, à l'édification d'une niche ou à l'isolation partielle dans le but de prouver qu'elles résulteraient en une spéciation allopatrique, parapatrique et péripatrique, respectivement. Grâce à la capacité de monitorage d'APOGeT, nous avons non seulement démontré que les motifs attendus ont en effet émergé mais aussi que l'observation des interactions interespèces a révélé des niveaux de complexité entremêlés.

Now equipped with a tool for monitoring population-scale dynamics over long periods of evolution, we turned our attention onto a more autonomous process, akin to natural selection. To this end, we drastically reduced the complexity of our individuals' morphogenetic capabilities (below) and that of their environment (section 6.2) by downgrading to 2D.

This shift of computational burden from the individual to the population allowed for simulating much longer period of time with larger population count. The experiment described at the end of this chapter (section 6.3) shows that even when using text-book environmental patterns, much complexity is produced both at the plants and species levels (section 6.4).

6.1 Plants Model

In order to simulate larger plants populations over longer periods the individual's modeling has been condensed into a much simpler set of components: a pair of Lindenmayer Systems controlling the underground and surface portions, a set of portion-wide compartment for resource storage and a limited number of metabolic coefficients to control the plant-wide behavior. This approach is similar to that used by [Bornhofen et al. 2011] to efficiently model biological classes of plants and their interaction in a limited environment.

The complete genomic contents of the base individual used in the latter experiments is displayed in figure 6.1. The fields groups are:

- I) L-Systems, the morphological basis of the individuals, described in section 6.1.1.
- II) Metabolism, which controls the interaction with the environment, detailed in section 6.1.2.
- III) Miscallenious, parameters used at various stage of the plant's life, discussed below.
- IV) Self-Reproduction, whose specific of this experiment are outlined in section 6.1.3.

6.1.1 L-System

The choice of L-Systems over the previously used 'Graphtals' was motivated by the fact that they are computationally cheaper especially because of their focus on large organ complexes. Moreover, we were still intent on obtaining, after evolution, insights into the strategies involved which prevents the use of more black-box models such as Genetic Regulatory Networks [Disset et al. 2016] or Vascular Morphogenetic Controllers [Zahadat et al. 2017a].

The L-Systems used in this experiment are deterministic, context-free and share the same set of control instruction. Left and right rotations are triggered by the $+$ and $-$ symbols, respectively. All symbols enclosed in brackets denote a branch connected to the organ directly preceding the opening bracket. The

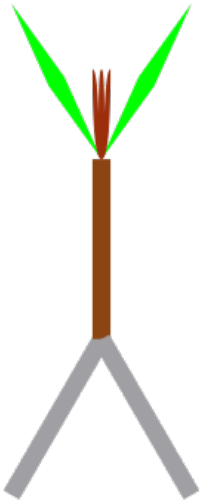
(I)
```
"root": {
  "recursivity": 2,
  "rules": { "S": "S -> [+h][-h]" }
},

"shoot": {
  "recursivity": 2,
  "rules": { "S": "S -> s[+l][-l]f" }
},
```

(II)
```
"metabolism": {
  "conversionRates": [ 0.9, 0.9 ],
  "deltaWidth": 0.2,
  "growthSpeed": 1.0,
  "resistors": [ 1.0, 1.0 ]
},
```

(III)
```
"dethklok": 10,
"fruitOvershoot": 1.1,
"temperatureOptimal": 10.0,
"temperatureRange": 10.0,
"seedsPerFruit": 3,
```

(IV)
```
"cdata": {
  "gen": 0,
  "id": 0, "p0": -1, "p1": -1,
  "sex": 0,
  "mu": 2.0, "si": 2.0, "so": 2.0
}
```

(a) Genotype



(b) Phenotype

Figure 6.1: Base plant for the speciation test-bed experiment

set of non-terminal symbol is $T = \{A, \ldots, F, S\}$. The non terminal S is treated specially as it models the undeveloped (seed) state of the corresponding plant portion.

The shoot manages the above-ground portion of the plants' structure and uses the terminals s (stem), l (leaf), f (flower) while the below-ground compartment instead relies on t (root trunk) and h (root hair). Both stem and root trunk organs have a structural and storage role in the plant. Leaves collect light from the environment, root hair extract water from the underground layer and flowers engage in sexual reproduction, their behavior depending on the sex of the plant. Additionally, a special terminal g models fruits, in the upper compartment, and results from a successful reproduction event.

The possible mutations for the rule set and their theoretical rate of occurrence is given in table 6.1. When applying these punctual mutations, a number of restrictions apply to maintain the system in working conditions. First, the rule-set can never be empty. In the event that there exists only one rule, then it must be the one derivating from the non-terminal S. This ensures that plants will always attempt to germinate (though no restriction are imposed on the contents of this rule). Additionally, The number of non-terminal symbols is limited to six ($A...F$, not counting S) and rule addition is inhibited when no more symbols are available. Moreover, rules are limited to a maximum of $M = 4$ non-control characters ($\notin \{+, -, [,]\}$) so that derivations must occur when aiming for complex morphologies. This also provides a more plausible developmental pattern by prohibiting the production of very large plant sections in a single step. Finally, the number of replacements a plant can perform for a given compartment is limited to a small value $D \leq 5$ (the field named

Rate	Type	$Before$	$After$
7.7%	Duplication	$S \to slf$	$S \to sllf$
31.1%	Replacement	$S \to slf$	$S \to sll$
3.8%	Suppression	$S \to slf$	$S \to s\emptyset f$
3.8%	Branching	$S \to slf$	$S \to s[l]f$
23.3%	Swapping	$S \to slf$	$S \to sfl$
20%	Addition	$S \to slf$	$S \to sAf$ $A \to l$
10%	Suppression	$S \to sAf$ $A \to l$	$S \to s\emptyset f$

Table 6.1: L-Systems mutations

recursivity in figure 6.1a), itself subject to mutations, which bounds the number of symbols in the derived phenotype to M^D.

Crossover is implemented in a most straightforward fashion, with a 50% chance of a child inheriting any rule from its parents. In case of a match (i.e. the premise exists in both) then one of the derivation is selected. In case of a mismatch, when a given premise is only found in one of the parents, a coin toss decides upon the insertion into the children rule-set. The rationale behind this simple crossing procedure is the same as defended in section 4.1.4: a heavy operator capable of handling dramatically different topologies is not required, given that it is the plants' responsability to impose genetic barriers with distant genotypes.

In order to use the L-Systems as the controllers for a developmental model, the transition from a seed into a plant is designed as a local process. Indeed each apex (non-terminal) acts as a derivation point that produces the right hand side of its corresponding rule when local conditions are appropriate.

These are twofold: part of the dry biomass produced by the plant's metabolism (see next section) is stored in these special organs until the reserves are large enough to 'pay' for the derived organs. The cost for a given rule $R \to s_1 \dots s_n$ is given by equations 6.1-6.2 where $width_0(s)$ and $height_0(s)$ are the initial dimensions for a given symbol (see E for the complete list of values).

$$mass(s) = \begin{cases} width_0(s) height_0(s) & \text{if } s \text{ is a terminal symbol} \\ 0 & \text{otherwise} \end{cases} \tag{6.1}$$

$$cost(R \to s_1 \dots s_n) = \sum_{1 \leq i \leq n} mass(s_i) \tag{6.2}$$

Second, to promote inter-individual competition, collisions are prohibited and a rule can only be triggered when it does not conflict with any other already existing organ (see section 6.2.2 for more details). However, one must note

Constants	
k	assimilation rate
J	saturation rate
f	resource cost
l	life cost
m_{Tr}, s_{Tr}	temperature range regulation
Genetic fields[1]	
g_s	Growth speed
δ_w	Structural organ width variation
μ_T, σ_T	Plant's temperature parameters
R_E	Resistor for transportation of element E
Environmental conditions	
P	plant's position
T	temperature at P
X^L	Biomass for layer L
R_E^L	Reserve in layer L of element E
C_E^L	Concentration in layer L of element E
T^-	1 if $T < \mu_T$, 0 otherwise
T^+	1 if $T > \mu_T$, 0 otherwise
w_h	Water around root hair h
s_h	Surface of root hair h
l_l	Length of leaf l exposed to the sun

Table 6.2: Metabolic variables

that these apexes have no impact on the plant's phenotype (i.e. they do not participate in collision detection and have no production cost).

6.1.2 Metabolism

Similar to [Bornhofen et al. 2011], plants in this model have three 'reservoirs' per compartment: one for water, which is extracted by root hairs h below the surface, one for glucose, produced by photosynthesis from leaves l, and one for dry biomass generated by converting these nutrients.

In addition, the effects of external temperature are taken into account at multiple stages of the metabolic dynamics whose control parameters are detailed in table 6.2. Given the bell curve function of mean m and standard deviation s

$$gauss(x, m, s) = exp^{-\frac{(x-m)^2}{2s^2}} \tag{6.3}$$

a plant's heat efficiency at temperature T is defined as

$$h_{eff}(T) = gauss(T, \mu_T, \sigma_T)\ gauss(\sigma_T, m_{Tr}, s_{Tr}) \tag{6.4}$$

The left-hand side part of the equation impedes the metabolism as T goes further from the plant's optimal temperature μ_T while the right-hand side part regulates the tolerance range σ_T so that it cannot grow unchecked. Indeed,

[1] With respect to the genotype from figure 6.1a, g_s corresponds to the *growthSpeed*, R_E to the *resistors* array, δ_w to *deltaWidth* and μ_T, σ_T to *temperatureOptimal* and *temperatureRange*, respectively.

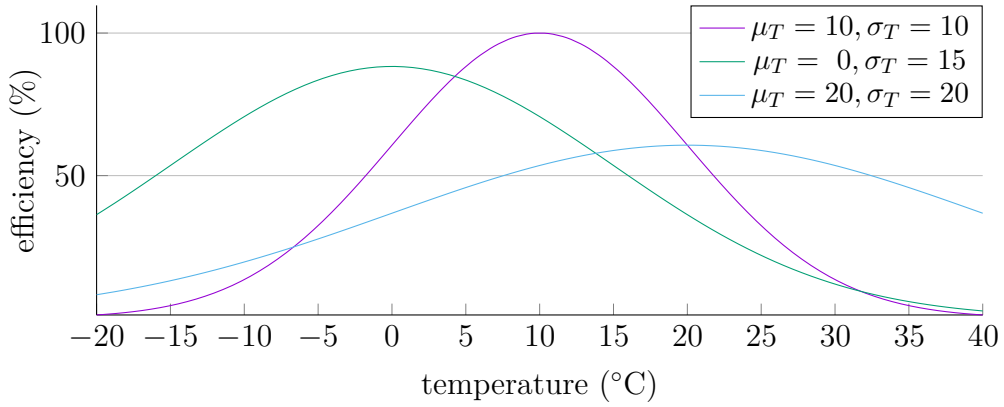


Figure 6.2: Samples of heat efficiencies for a handfull of parameters using, as in the following experiments $m_{Tr} = s_{Tr} = 10°\text{C}$. One can see that, by favoring a larger range of acceptable temperatures, peak performance is conversely reduced.

the individuals must strike a balance between resilience to greatly varying temperatures (at the cost of average efficiency) and optimization for specific environmental conditions (at the risk of extinction should these change too much) as illustrated on figure 6.2. This impacts water uptake as, the lower the temperature is below μ_T, the less a plant can absorb water through its root hairs:

$$U_W(T) = \frac{T^-(h_{eff}(T) - 1) + 1}{1 + C_w^{rt} J} \sum_{h,\text{root hair}} k w_h s_h \tag{6.5}$$

Nonetheless, the root compartment shares a portion of its water reserve to the shoot, according to the relative concentrations and transport resistors:

$$T_W = \frac{C_W^{rt} - C_W^{sh}}{\frac{R_W}{X^{rt}} + \frac{R_W}{X^{sh}}} \tag{6.6}$$

Leaves in the upper layer with direct access to sunlight then produce glucose and similarly to equation (6.6) transports part of it to the lower layer:

$$U_G = \frac{1}{1 + C_G^{sh} J} \sum_{l,leaf} k l_l \tag{6.7}$$

$$T_G = \frac{C_G^{sh} - C_G^{rt}}{\frac{R_G}{X^{rt}} + \frac{R_G}{X^{sh}}} \tag{6.8}$$

When placed under too hot environmental conditions, plants will additionally experience water loss through transpiration:

$$R_W^{sh}(T) = (1 - T^+ h_{eff}(T)) R_W^{sh} \tag{6.9}$$

Extreme temperatures can lead to a complete drain of their shoot water reserves in a day. Plant tissue turnover is modeled by continuously transforming

part of the biomass in to wastes:

$$W^L(T) = l(2 - h_{eff}(T))X^L \tag{6.10}$$

External conditions influence this as well by inflicting upon plants under uncomfortable temperatures up to 200% the rate of cellular decaying experienced by siblings under a more favorable climate. Finally, both glucose and water reserves are consumed to generate new biomass which is allocated to the various sinks (flowers, fruits, stems and root trunks) in the plant:

$$\dot{X}^L(T) = g_s X^L C_W^L C_G^L - W^L(T) \tag{6.11}$$

One should note, however, that, whenever wastes production exceeds dry biomass renewal, $\dot{X}^L(T)$ will be negative. That is, sinks will *lose* biomass causing them to shrink. This leads to their death as soon as their individual biomass is completely depleted, removing them and their subtrees from the plant. Starvation is, thus, one of the possible cause of death for an individual: when all of its sinks are destroyed the plant itself is considered dead. Senescence is the other one, as determined by an evolved genetic field (*dethklok* in figure 6.1a), thus preventing immortal phenotypes from monopolizing the environment.

The structural sinks (stem and root trunk) consume the produced dry biomass to increase in size depending on their position in the plants hierarchy. Given an organ o at depth d in the structure (i.e. the subtree rooted at o has an height of d when considering structural organs only) its expected biomass is:

$$width(d) = w_0(1 + \delta_w)^{d-1} \tag{6.12}$$

with $w_0 = 0.01$ the initial width for an organ (see annexe E for the complete list of values). These additional biomass is used to increase the sink's width so as to produce more realistic morphologies with bulky trunk segments at the base of the plant and thinner branches near the canopy.

The flower and fruit sink types use the stored resources towards self-reproduction and will, thus, be detailed in the following section.

6.1.3 Self-reproduction

Given that, in this experiment, plants receive no help from an external evolutionary algorithm, they must be capable of efficient self-reproduction. The initial genome presented in figure 6.1b has such capabilities, though with a large room for improvement.

The first step in producing offspring is by the development of a flower organ. In both sexes, these share the same initial step of dry biomass collection (3 times their initial biomass) in order to reach maturity. Once this state is reached, flowers will act differently depending on the gender of their plant. Stamens 'actively' seek out potential parterns in their vicinity through a collision procedure described in section 6.2.2.

Upon pairing of both a mature male and female plant, the crossing procedure is similar as the one presented earlier in the case of our 3D experiments (section 4.1.4). Whenever the bail-out crossover returns positive a litter of offspring is produced, the size of which being controlled by the *seedsPerFruit* genetic parameter. The pistil is then replaced by a fruit (the special terminal symbol g) which must accumulates more biomass before being torn off from its plant.

For every offspring genome, their corresponding shoot and root L-Systems contain an initial rule derived from the seed non-terminal S. In order to guarantee germination, the parent must ensure that each offspring has enough starting resource for this rule to be triggered. This imposes that the total cost for a fruit containing n genomes each with an initial rule R_i^{sh} and R_i^{rt} for the shoot and root, respectively, is given by equation 6.13.

$$X_t = f_O \sum_{1 \leq i \leq n} cost(R_i^{sh}) + cost(R_i^{rt}) \tag{6.13}$$

The term f_O corresponds to the genetic field *fruitOvershoot* and serves as a safeguard against overcrowding. Indeed, in the event that a seed cannot germinate immediately after being torn off its parent plant, its internal reserves will slowly deplete themselves as a result of waste production. Thus it is possible for a seed to die of starvation even before sprouting its first organs. The *fruitOvershoot* thus codes for carefulness in the face of a potentially highly competitive biotic environment.

Actually disseminating these seeds into the environment can happen in one of two ways. The most obvious one is that upon reaching maturity (i.e. accumulating all the required biomass), the fruit is torn off its plant and each seed is placed at a distance dx taken from a normal distribution with a mean of $\mu = 1 + 5h$, where h is the altitude of the fruits with respect to the ground. The standard deviation σ is a third of the mean and no values can be off by more 4σ.

This restriction comes from the sampling procedure involved in determining the validity of each potential location. As seeds are modeled after air-borne objects, reaching higher altitudes is less likely than dropping down. It stems from this assertion that for a voxel v at distance d from the fruit and relative height h the probability to plant a seed at this location is given by equation 6.14

$$P(v) = gauss(d, \mu, \sigma) gauss(h, 0, \sigma_h) \tag{6.14}$$

with σ_h a control parameter defining exactly how hard it is, for a seed, to climb upward.

The other mean of dissemination is upon plant death: all pending fruits are broken down and their reserves fairly distributed between individual seeds. The insertion position is also determined via equation 6.14.

6.2 The environment

6.2.1 Variables & dynamics

So that individuals can be subjected to a large range of dynamical abiotic conditions, the environment can produce changes along three dimensions:

Topographical y, with seeds being much harder to disseminate onto higher ground (see eq 6.14)

Hygrometric w, water availability which, being the primary resource of the plant's metabolism, is of utmost importance

Temperature t, with equations (6.4-6.10) showing the impact of uncomfortable temperatures on an individual's welfare.

The system is designed so that one can easily plug any kind of controller between the input D, Y, x, y, w, t and output $\dot{y}, \dot{w}, \dot{t}$ variables where D is the relative time in the current year ($\in [0:1]$), Y the relative time in the planned simulation duration (same range) and x the position in the environment. All other values have range $[-1:1]$. In this proof-of-concept experiment, we resorted to a simple expression parser to easily define straightforward validation scenarios.

Additionally, a pair of constraints $C0$, $C1$ (controlled by the genomic coefficients c_0, c_1) is used to post-process the outputs of the environmental controller so as to provide more plausible correlations between physical dynamics:

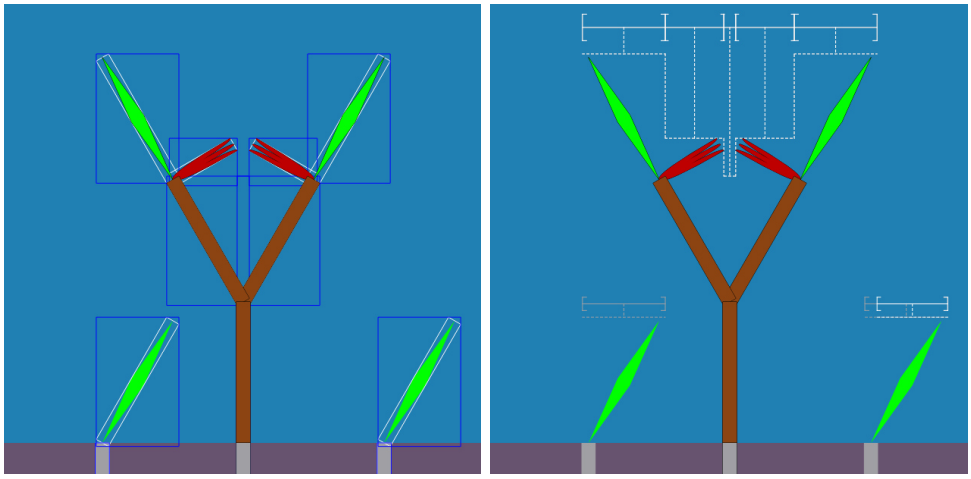
$$C0 : \hat{t} = -c_0 \max(0, y) + (1 - c_0)\dot{t} \tag{6.15}$$

$$C1 : \hat{w} = -c_1 \min(0, t) + (1 - c_1)\dot{w} \tag{6.16}$$

that is temperature decreases linearly with an increase in altitude and water evaporates more (and thus also decreases) as temperature raises.

6.2.2 Physics engine

In order to make the most of our very specific experimental conditions (2D, guaranteed intersection with the origin, simple geometrical primitives, no self-triggered motion), we designed a bare-bones physics engine from the ground up. Its collision detection algorithm are divided into three categories: determining which parts of a leaf are *directly* exposed to the sun, finding the pistils in range of a given stamen and detecting all manners of plant collision (inter-plant, intra-plant ...)



(a) Bounding boxes (b) Canopy

Figure 6.3: Light exposure detection algorithm

Light

Light availability is of paramount importance for the photosynthesis of glucose and thus plant welfare. As a bio-mimetic prerequisite, this resource is not accessible without appropriate morphology. Though one obviously need leaves to produce glucose they are often, in the biological world, backed by a collection of supportive structure designed to maximize efficiency.

With this in mind, we implemented a very straightforward algorithm that accurately divides the incoming 'sunlight' onto each exposed surface (fig 6.3). First, on the left-hand side the blue and white boxes surrounding each organ are their AABB and OBB[2], respectively. Indeed, for efficiency purposes, all organs are modeled as an oriented rectangle. The shapes chosen for rendering are arbitrary but still contained inside this OBB.

On the right-hand side is the same situation showing the graphical debugging of our 'canopy' algorithm. The procedure is twofold: for each plant, we start with a canopy evaluation in isolation before merging the upper layers of each AABB collision. The underlying algorithm is a 2D linesweep where we keep track of which organ occupies the topmost position, which allows for partial exposition. For instance the stems in figure 6.3b are almost completely covered but a limited portion is detected as exposed to the sunlight. The result is, for each plant, a collection of item each referring to a non-null portion of those organs that form the plant's canopy.

These items are displayed as the range, in gray, at the top of the figure. The second part of the algorithm then simply consists of repeating the same procedure for each pair of AABB colliding plants. In this pass, we only need to compare the canopies of pairs of plants, thus strongly limiting the computational cost. This produces another collection of items, which takes the

[2]Oriented Bounding Box

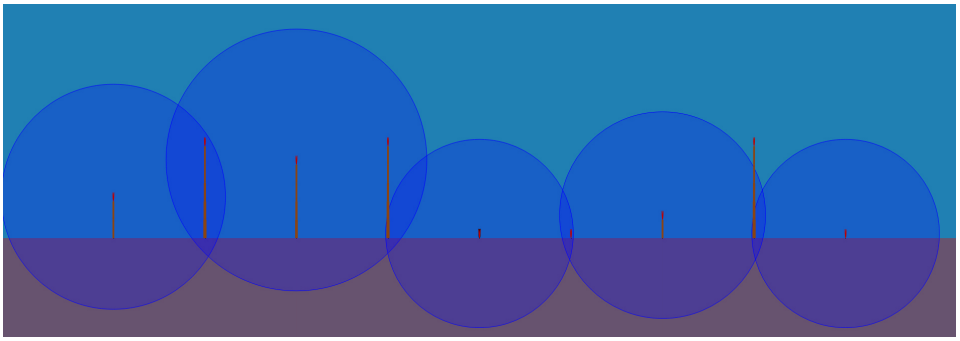


Figure 6.4: Reproduction area as a disk whose radius depends on the stamen's height. Picture shows 10% of actual value for clarity

local environment into account, resulting in the white colored items.

One can see that through this algorithm, vertical competition over light can emerge through processes of partial or even complete occlusion. Indeed the left-most plant is unlikely to survive under its current conditions while the right-most will only be slightly inconvenienced.

Reproduction

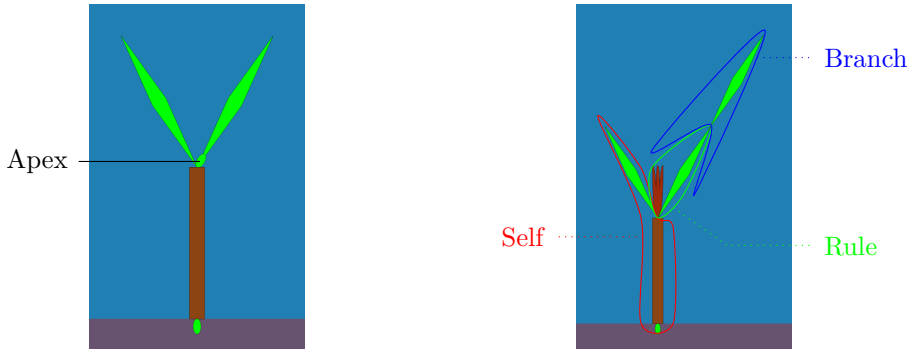
As previously mentioned, the reproduction of individuals requires stamens to be able to find potential pistils in their vicinity. In our physics engine, each female flower f is represented by a disk of radius $R(y) = 5(y+1)$ meters, where y is the relative altitude of f. As illustrated in figure 6.4, this corresponds to a rough approximation of air-borne sexual material dissemination. Thus the higher a plant grows, the larger its effective reproductive area, though the choice of smaller morphologies is still viable but requires close proximity.

When searching for mates, we instantiate another disk with radius $R(y)$ for the male flower and, through an horizontal sweep, collect every female bounding disk found for application of the bail-out crossover. This allows for a very efficient method for finding pistils in the reproductive area, while providing a continuum of strategies between large plants with a broad seed dispersion pattern and smaller ones that rely on close proximity to kin.

Collisions

The last part of our physics engine is concerned with collisions detection between organs in order to maintain some measure of physical plausibility. Indeed preliminary experiment show the tremendous capacity of these 'naturally' evolving plants to produce very efficient morphologies by stacking organs on top of one another.

The procedure for deriving a collection of organs from an apex is, as previously mentioned, a local process. Every day (10 simulation steps), every non-terminal organ A is queried, in a random fashion as to promote robustness,



$s[+l][-Al]$ $s[+l][-[+f]ll]$

(a) Initial plant (b) After derivation

Figure 6.5: Plant section when applying $A \rightarrow [+f]l$. See text for details

for a potential derivation. In those cases where there is a rule for the corresponding premise, we instantiate these new organs and generate a clone of the subtree rooted at A as shown in figure 6.5.

The *Self* section corresponds to those part of the plant, of the same layer, left untouched by the derivation attempt. The *Rule* is the collection of newly created organs, aiming at replacing A and the *Branch* is the portion of the plant that includes the *Rule* and all of the subtree under A. This labeling highlights the specific sets of organs that we must test in order to prevent the more flagrant violation of real-world physics law.

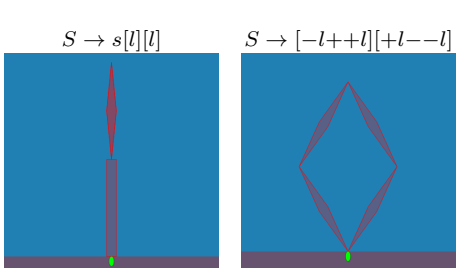
Based on this sectioning of the morphology of a growing plant we derived four classes of collisions as illustrated in figure 6.6.

First and foremost, in all following collision tests we rely on a linesweep algorithm for performing the broadphase collision between organs. Candidates are then evaluated through the Separating Axis Theorem (SAT) which is highly simplified thanks to our exclusive reliance of rectangular shapes.

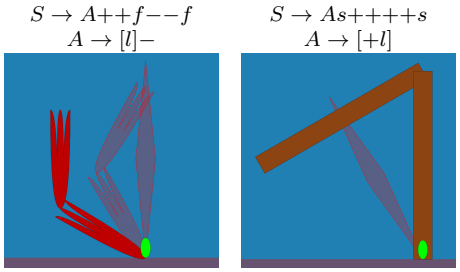
Our first class addresses those rules that produce colliding organs (figure 6.6a). On the one hand, this case is the largest concern for it can quickly produce phenotypes with a very large number of 'layers' while occupying a minimal space: a very efficient, albeit undesirable, strategy in our simulations where space is one of the most limiting resources.

On the other hand, it is also the simplest to solve given that we only have to test for collision between a very small number of organs. As mentioned in section 6.1.1, the number of non-control characters in each rule is bounded by $M = 4$. Thus given that we rely, once again, on a linesweep algorithm, the complexity of determining class 1 collision is $O(M)$.

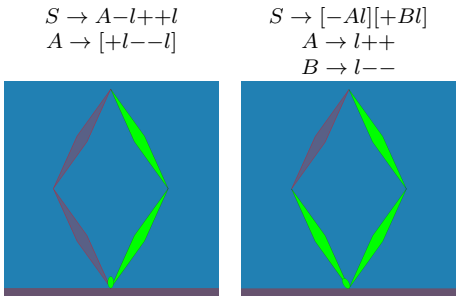
Additionally, these rules are definitely never going to be derivable: whatever the context maybe a rule resulting in a class 1 collision will always fail. Thus when detecting this kind of behavior we remove the offending apex from the plant, effectively transforming the rule into $A \rightarrow \emptyset$.



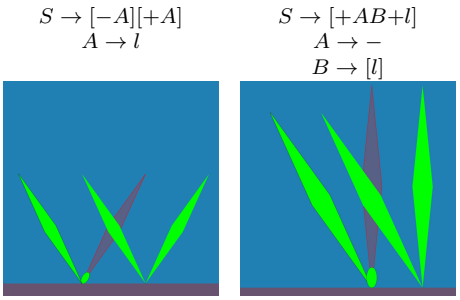
(a) *Rule* / *Rule* (**C1**)



(b) *Rule* / *Branch* (**C2**)



(c) *Rule* $\cup$ *Branch* / *Self* (**C3**)



(d) *Rule* $\cup$ *Branch* / *Plant* (**C4**)

Figure 6.6: Collision classes. Organs in red denote rejected derivation attempts

A slightly more involved case is the one described in figure 6.6b: the newly created organs intersect with the subtree under A. Though still capable of creating overlapping morphologies, the problem of class 2 collision lies in the fact that we no longer have a reasonably low upper bound for the number of objects we need to examine. Indeed we look for any collision between organs in *Rule* (with $|Rule| \leq 4$) and those in *Branch* (with $|Branch| \leq M^D$) which results in a complexity of $O(M + M^D)$.

Unlike the previous class of collision, these are dependent on the context surrounding the derivation attempt: differences in the *Branch* portion (e.g. through the expansion of another apex) can lead back to stable (non-colliding) topologies. Thus apexes falling into this category are left untouched, until the next attempt.

The final case of internal collision occurs when the modified part of the plant (the union between *Rule* and *Branch*) intersects the rest of the plant (the untouched part, *Self*) as shown in figure 6.6c. Though the sets of organs to examine differ, the same remarks can be made: undesirable morphologies, number of comparison bounded by $O(M + M^D)$ and context-sensitiveness.

Inter-plant collisions are illustrated in figure 6.6d and the same manner of broadphase iteration, though the sets are now from different individuals. For the plant attempting a derivation this is the same set as for class 3 collisions. The set of organs to test against, however, is now bounded by the number of plants in the environment, though the morphologies required for such a case to

occur are very unlikely to be attained.

Still, this class of collisions is the most computationally expensive and is optimized at the level of the initial broadphase (between the plant and its neighbors) by exploiting the specifics previously mentioned. Indeed we know that all plants must cross the ground plane and that motion is impossible: this allows for a very efficient caching of the pair of plants whose AABB are intersecting.

There, again, the derivation is context sensitive: further derivation attempts might provide with a positive result once the competition disappears. In fact, this is one of the limiting factors for seed germination. When population density is high, seeds, which consist only of the non-terminal S and by consequence trigger no collision, tend to fall into already occupied patches of ground. This local derivation with retrials imposes a period of quiescence which then leads into a thriving growth once the competing organisms die off.

6.3 Test protocol

Before diving into the exploration of environment-driven evolution, we tested our framework on text-book cases of environmental dynamics known to produce speciation. More specifically we used hand-written dynamical equation tailored so that, starting from the uniform environment at the top of figure 6.7, each group of simulations ended in one of the configuration displayed on lower line.

All simulations started with the same plant genotype (fig. 6.1a) and, besides the dynamics equation, the same environment. Its size was 100 meters wide and 50m high, which, given an initial plant dimension on scale of the decimeter, provides plentiful space for colonization[3]. Simulated durations are:

- 10 steps per day
- 100 days per year
- 100 years per simulation

Upon initialization 100 clones, derived from the primordial genome, are uniformly placed around the center of the environment with a uniform spacing of half a meter. These individuals are then left to their own devices, in a form of 'natural' selection in the sense that there are no outside interference: any evolutionary pressure is the result of the interaction between the biotic and abiotic components of the ecosystem.

In this experiment, we are interested in whether or not *strong* speciation occurred, that is we are more focused on the apparition of reproductively isolated species than of varieties. To this end, we defined two metrics.

[3]Upper bounds is 10K mature plants (i.e. with at least one terminal organ) not counting for un-germinated seeds

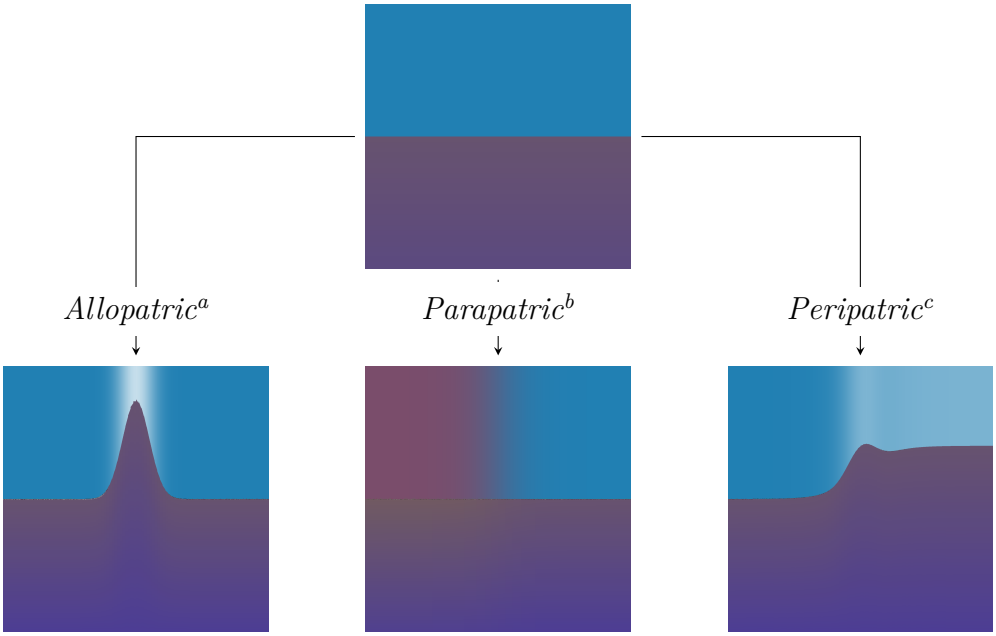


Figure 6.7: Top: initial conditions: altitude of 0, temperature of 10°C and hygrometry of 0.5. Bottom row shows the final state. From left to right: complete geological isolation due to mountain formation (37.5m high and 20m wide), niche formation via desertification of left-hand side (up to 40°C on the left edge) and plateau formation resulting in partial geological separation.

[a] Mode of speciation that occurs when biological populations of the same species become isolated from each other to an extent that prevents of interferes with gene flow. Genetic drift and different evolutionary pressures on each sub-population lead to cladogenesis preventing reversion to a single species should those sub-populations meet again

[b] Mode of speciation that occurs when two sub-populations of the same species evolve reproductive isolation from one another while continuing to exchange genes

[c] Mode of speciation where an a new species is formed from an isolated peripheral population. The primary difference with allopatric speciation is that one the population is smaller than the other

The absolute compatibility between species A and B at a given timestep with $P^A = \{P_1^A \dots P_n^A\}$, the female plants of species A, and $P^B = \{P_1^B \dots P_m^B\}$, the male plants of species B, is:

$$c_a(A,B) = \frac{1}{|P^A||P^B|} \sum_{f \in P^A} \sum_{m \in P^B} c_{f,m} \tag{6.17}$$

where $c_{f,m}$ is the compatibility as computed according to both individuals' genetic distance and the parameters $(\mu, \sigma_i, \sigma_o)$ of f's genome. We thus obtain the average compatibility between possible mating pairs of each considered species. We then derived from $c_a(A,B)$ the relative compatibility as follow:

$$c_r(A,B) = \frac{c_a(A,B)}{c_a(A,A)} \tag{6.18}$$

which provides a normalized metric whose comparison between different reproductive trends or even simulations is more straightforward.

6.4 Results

Results across all three experiments are summarized in figure 6.8 with an uneven number of replicates: 13, 12 and 11 for the allopatric, parapatric and peripatric, respectively. This corresponds to the subset, from 20 runs per protocol, that neither immediately go extinct nor failed to reach the 100th years, in the allotted 10 hours timeframe. Note that, given the definition of $c_r(A,B)$, the minimal worse and maximal best relative compatibility is 100%. Indeed, the worst case scenario would be having all values clustered at, or very close to, 100% which would show a striking lack of speciation. Given that this

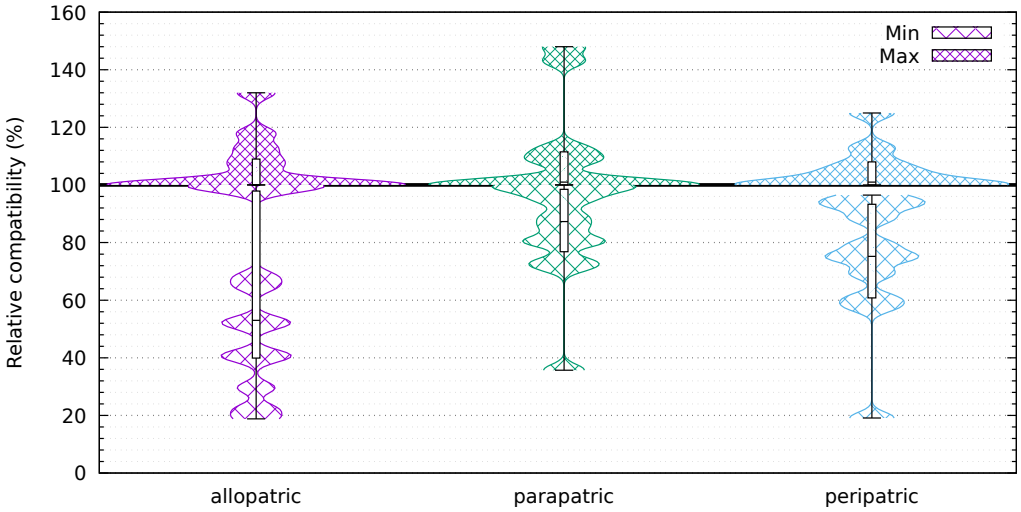
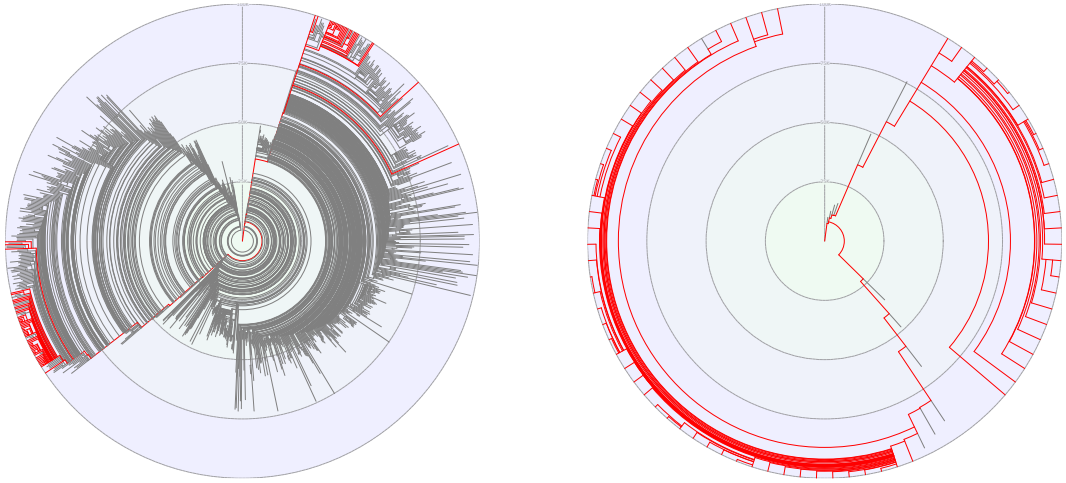


Figure 6.8: Speciation results for the three experiments



(a) Complete (b) Simplified

Figure 6.9: Phylogenic tree for the lowest c_r at the 100th year

is not the case, we can safely conclude that *some* did occur, which will be explored in the following sections.

6.4.1 Allopatric speciation

$$\begin{aligned} c_0 &= c_1 = 1 \\ \dot{t} &= .75 sin(.5Y\pi)\ gauss(x, .5, .05) \\ \dot{y} &= \dot{w} = 0 \end{aligned} \tag{6.19}$$

Our first test case is focused on the most simple mode of speciation: complete geographical isolation. To this end, our environment, otherwise uniform, slowly grows a mountain in its center according to the parameters described above. This gradual process produces, at the end of the simulation, a topographical barrier 37.5 meters high and 20 meters large. As seeds have difficulty reaching higher places this effectively prevents cross-reproduction between individuals from either side.

As seen in figure 6.8, speciation did occur in this experiment, however aggregated data can only show a coarse picture. To this end, focused our attention on the most successful run, in terms of *strong* speciation. Its minimal relative compatibility is of $c_r = 18.8\%$ that is to say there is pair of species that is only capable of fecundating one another one out of five attempts. On the opposite side the maximal value is $c_r = 100\%$ indicating that for every species in the final population, none is more compatible with itself than... itself.

The phylogenetic tree obtained from this run shows a very strong divergence (fig. 6.9a). One can clearly distinguish the two species clusters stemming from the geographical separation with the lower part of the rightmost one failing to provide viable species past the 75th year. Unfortunately, however complete this graph may be, it is too densely packed with extinct species to provide much

(a) Most complex (b) Divergent

Figure 6.10: Morphologies show limited complexity

insight on when speciation did occur. We thus favor the simplified version in figure 6.9b which only shows the species on the survivor paths (those direct ancestors of currently alive individuals). Then we can easily see that very early in the simulation, around the 10th year, two species branched off from the main branch and, due to the harsh topographical barrier, went on to further speciate in their own isolated plot of earth.

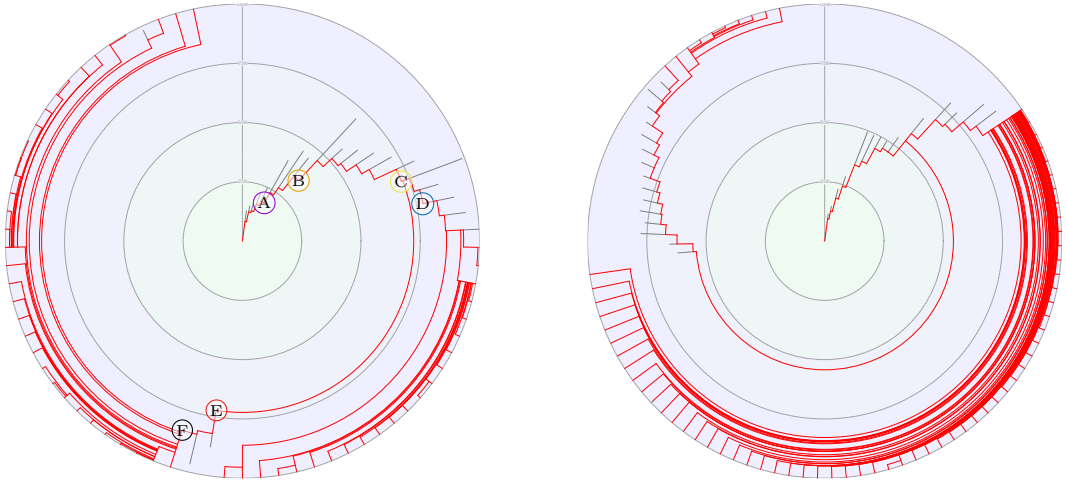
In order to better understand the type of genomic difference between individuals from different species we examined the morphologies produced during these simulations. However, as can be seen in figure 6.10a, even the most complex one is a far cry from what we could expect from an L-System. Indeed, always the minimalist one, natural selection only produced that which is essential and plainly ignored the structural organs (stem s and root trunk t), instead focusing its efforts on extracting nutrients from the environment (root hairs h, in gray, and leaf l, in green) in order to grow the maximal amount of flowers (f in red) so as to maximize its reproductive potential.

Still, some degree of morphological divergence were observed from individuals in the same simulation with sample plants from figure 6.10b being representatives of the most populated species on the left and right side of the mountain for a run with a good speciation score (minimal $c_r = 29.6\%$). Obviously, given the depth of structural complexity, these differences are not as striking as one could wish for. Thus, the non-uniform locusts are to be found in other parts in the genome (metabolic values, compatibility functions, ...) where direct observation is much less straightforward and is left to actual timelines exploration (see chapter 7).

6.4.2 Parapatric speciation

$$\begin{aligned} c_1 &= 1 \\ \dot{t} &= .4sin(.5Y\pi).5(tanh(8(.5-x))+1) \\ c_0 &= \dot{y} = \dot{w} = 0 \end{aligned} \tag{6.20}$$

A slightly more complex scenario involves the gradual apparition of a niche with no geological separation from the rest of the environment. This implies that a contact zone exists between the two parts and thus that gene flow is



(a) 'Worst' (b) Best

Figure 6.11: Phylogenic trees for the parapatric runs with the most extreme speciation at y=100.

not restricted by the abiotic component: speciation is left in the hands of the individuals themselves.

As illustrated in figure 6.7, the left-hand side of the plot undergoes a gradual warming effect which, given the activation of constraint $C1$, also reduces the amount of available water.

Once more we refer to the relative compatibilities shown in figure 6.8 to assert that this experiment also produces divergences and clustering, though more limited in range. Indeed, in this case, the apparition of a reproductive barrier is mostly dependant on the emergence of the reproduction schemes able to enforce isolation. We thus expected to see a period of cross-breeding before the speciation takes sufficient hold. One should also note that some simulations failed to colonize the harsher portion of the environment, thus degenerating into an evolution in uniform abiotic conditions.

The survivor-only version of our phylogenetic tree is displayed in figure 6.11b for the best scoring simulation ($c_r \in [35.7\%, 108\%]$) and it shows that the branching event that produced the two main strands occurred much later than in the previous experiment (slightly after the 50th year). Furthermore, the species density of these two branches is quite dissimilar with only the upper left portion accounting for those found off the desertic side. We could thus conclude that, to a weaker extent, the parapatric experiment successfully induced speciation.

However, the case of the worst scoring simulation ($c_r \in [89.1\%, 143\%]$) is much more interesting when looked at in more details. Indeed these c_r values show that not only reproductive isolation did not emerge in any significant proportion (even the term varieties might be too strong a word) but, on the contrary, there are cases of intense outbreeding: the 143% maximal relative

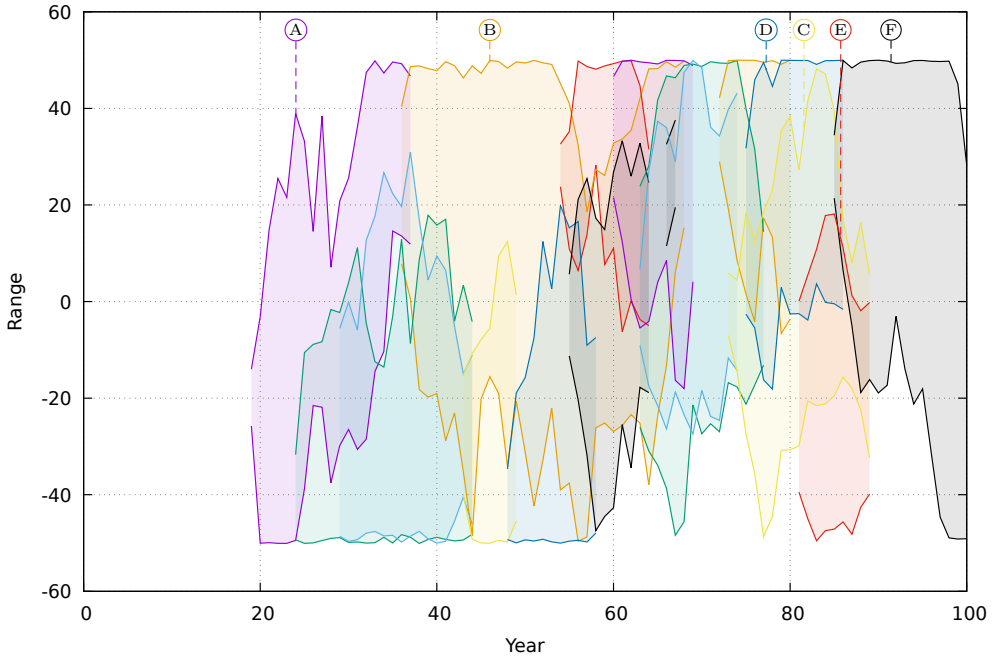


Figure 6.12: Colonization ranges for the 16 most populated species in a 'negative' run. The vertical span of a particular species indicates the range in which it has living individuals at the end of the corresponding year.

compatibility indicates that for at least one species it is 1.5 times more likely to reproduce with member of a foreign species than with more closely related mates.

We surmised that these results should come from a desertic species trying to gain ground into the temperature region by assimilating existing species and thus decided to look at the dynamics of colonization. The spatial distribution of the 16 more populated species generated by this 'worse' simulation are summarized in figure 6.12. The height of a region depicts the range over which a given species has individuals alive at the end of the corresponding year which is why ranges can and *do* overlap.

Even broad analysis shows that, contrary to our hypothesis, the simulation has not degenerated into a champion-dominated situation. In fact, various dynamics emerge as time goes by and temperature diverges in the desert (lower part of the graph) and temperate regions (upper part). During the first 18th year population count is too low to appear in the graph until species A emerges from a small region of the desert ($x \in [-26, -14]$). From there, it quickly grows in range during the next years, colonizing the whole region and sending onward 'scouts' in the more temperate zone. This leads to migration, over the next decade, into the environment's temperate portion where it is quickly overtaken by species B, an indirect descendant (see fig 6.11a).

Then starts a period of relative prosperity, where B has no real competition in its core range, so much that it regularly sends more 'scouts' back into the

desert, though without much success. This era ends past the middle of the simulation (50th year) where it must, once again, share space with multiple, newly born challengers. This chaotic period lasts until about the 80th year with only three dominating species left: D in the temperate region, E in the desert and C their ancestral species. In time, D spawns a final species, F, which in about a year colonizes and dominates the whole right-side part of the environment. It takes little more than a decade for its influence to grow over the rest of the simulation into the desertic portion. Thus from the 98th year onward F is firmly anchored as a polyvalent species capable of thriving in a range of heat/water combinations, though one can see the start of a downward trend in its original biome.

These dynamics are not without similarities with those produced by natural selection in the real world which goes to show that, despite the simplicity of both the environment and the morphological adaptations displayed by its inhabitants much complexity still emerged. They also throw a measure of doubt on the metric used to broadly classify the results: despite being anchored in the pragmatic definition that a species is a "group of inter-breeding individuals reproductively isolated", we can see that it produced at least one (and probably many more) false negative.

6.4.3 Peripatric speciation

$$\begin{aligned} c_0 &= 1 \\ \dot{y} &= .4sin(.5Y\pi)(.5(tanh(8(x-.5))+1) \\ &\quad + .5gauss(x, .5, .05)) \\ c_1 &= \dot{t} = \dot{w} = 0 \end{aligned} \tag{6.21}$$

For the sake of completeness, we briefly go into the details the last experiment performed: partial geological separation with niche subdivision which used the environmental parameters described above.

The right side of the environment rises slowly from sea level up to a 20m high plateau which, due to the activation of constraint c_0, is notably cooler than the adjacent lowlands. A small elevation in the center further separates both halves of the plot. This provides a more complex scenario which combines both of the previous approaches: on the one hand, the temperature differences stimulate generation of new shapes and exploration of genetic parameters while, on the other hand, the topographical separation limits gene flow, making it easier to keep true to the current evolutionary trend. In this case, however, the barrier is asymmetrical: as in the allopatric experiment, individuals at sea level have very limited chances to send seeds at such a remote altitude but plants on the plateau only have to cross the center elevation to disseminate their genetic material onto the lower half.

Given the intermediate nature of the setup, the fact that observed results, in terms of minimal/maximal relative compatibilities, are also intermediate

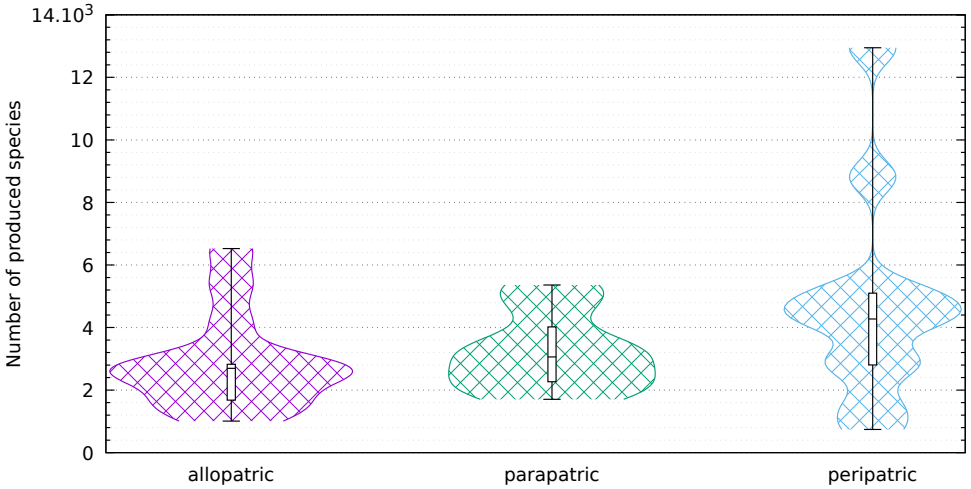


Figure 6.13: Number of observed species per experiment (in thousands)

does not come as a surprise. The topographical asymmetry induces a slightly more dispersed distribution of relative compatibilities than in the parapatric case, as seen in figure 6.8. Conversely, these trends are inverted when compared with the purely continuous simulations.

There is, however, a point on which we can differentiate this experimental setting from the others as depicted in figure 6.13 and annex F: the number of species.

Indeed the first produces an average of 2905 per run (1.1×10^6 plants, 849 generations), which is only marginally lower than the second one ($3228/1.09 \times 10^6/887$) and stays comparable with the third one ($4813/1.2 \times 10^6/898$). Even though these mean figures do not exhibit statistically significant differences, the distribution of values differ in a much more pronounced manner. While most runs for the allopatric speciation are clustered around the median and inter quantiles, runs in the peripatric experiment are more diffused, some reaching up almost to the next order of magnitude.

There is a similar trend with the number of generations but not the number of plants hinting that the lack of a strong geological separation promotes apparition of new species with roughly the same number of individuals by providing more noisy conditions.

6.5 Conclusion

In this experiment we set out to validate our framework for a natural selection algorithm controlled by the abiotic component of the ecosystem. To this end, we devised simple environmental settings that would mimic the natural conditions for known real-life cases of speciation.

Amidst the mass of data generated by our simulations, APOGeT managed to extract species trees which, when rendered into either their full or simplified

forms, were instrumental in determining whether speciation emerged from the underlying plant-controlled reproductions. In addition to these graphical analysis, the automatic tagging of individuals into species allowed for the use of population-level metrics, though, as mentioned, those used in this experiment did not prove as resilient as needed.

Although this process of natural selection did not feel the need to complexify the morphologies to any great extent, the dynamics exhibited on the species level were much more diverse and intricate, reminiscent of real-world ecosystem dynamics. An article resulting from this work was published in the proceedings of the 2019 Conference on Artificial Life [Godin-Dubois et al. 2019c].

This paves the way for a very broad number of future works divided into two categories: investigation and complexification. Indeed despite the minimalist approach used to generate the test environments, the complete range of dynamics, competitions and inter-dependancies could not be fully investigated here. Whether or not the situation described in the results of the parapatric experiment is a typical, favorable or below average case is left as an open question. The interested reader can refer to annex G, for the dynamics corresponding to those runs whose phylogenetic trees where presented here.

Furthermore, individual genetic fields were only briefly examined, mostly regarding morphologies. As we move towards longer simulated durations, to allow for appropriate exploration of the genetic space, we will need to develop methods for analyzing allelic trends across whole centuries. Indeed, much of the evolutionary adaptations will be located in numerical fields which will preclude manual observation.

Additionally, using hand-crafted equations for generating environmental dynamics is not the most generic way to tackle the problem of environment-driven speciation. To this end, the presented model will need to be extended by using an evolvable substrate (CGP, GRN, ANN) as the basis for the environmental controller. This will allow for the automated generation of ecosystems displaying wider ranges of demeanors whether related to well-known examples of real-life equivalents or diverging into unfamiliar directions.

Chapter 7

Timelines Exploration

Abstract Individuals complexity is irrevocably linked with that of the environment in which they are embedded and, as such, the former is an equally important component of any evolutionary process. We thus introduce the Environment-Driven Evolutionary Selection (EDEnS) algorithm in which environmental dynamics are used to drive a evolutionary process towards specific, user-targeted, regions of the genetic space. The novelty of this methodology is that selection is performed at the population level while action is performed by choosing a specific environmental controller that will specify, indirectly, future trends that will be favored in each individuals' survival. The adjoining experiment shows that top-ranking runs are able to outperform the control group on two different tasks of colonization. We also observe the positive impact of catastrophic population trimming through which evolution moves at a faster pace: either by removing all competing alleles for specific genes or by "leaping" from one region of genetic space to another.

Résumé La complexité des individus est irrévocablement liée avec celle de l'environnement dans lequel ils sont plongés et, en tant que tel, ce dernier est un composant essentiel des processus évolutionnaires. C'est pourquoi nous présentons l'algorithme de Sélection Evolutionnaire Guidée par l'Environnement (EDEnS), dans lequel les dynamiques environnementales sont utilisées pour conduire un processus évolutionnaire vers certains pôles spécifiques, ciblés par l'utilisateur, de l'espace génétique. La nouveauté de cette méthodologie est que la sélection est réalisée au niveau de la population tandis que l'action est effectuée par le choix d'un contrôleur environnemental donné qui va spécifier, indirectement, les futures tendances qui seront favorisées dans la survie de chaque individu. L'expérience adjointe montre que les séries les mieux notées sont capables de surpasser le groupe contrôle dans deux tâches de colonisation différentes. Nous observons aussi l'impact positif d'un élagage de population catastrophique au travers duquel l'évolution avance plus rapidement : soit en éliminant les allèles concurrents pour des gênes spécifiques soit en "sautant" d'un pôle génétique à un autre.

With the framework for simulating and monitoring the evolution of plant populations over significantly long time spans now set up, we move onto the meat of this work: Environment-Driven Evolution. In this chapter, we define an experimental protocol in which we guide a complete ecosystem (in terms of biotic and abiotic components) for a thousand years, with the controlling mechanisms being applied at the environment level. After mentioning a few minor modifications to the plants' morphological model, we describe the use of Cartesian Genetic Programming (CGP) to encapsulate environmental dynamics. We then detail the manner in which such an evolution is performed, with a particular attention to the selection method, before drawing conclusions on the relevance of such a method to the field of Artificial Life.

7.1 Model's extensions

Our first concern was to address the lack of vertical competition observed in the previous experiment. To facilitate the emergence of such growth, we drastically reduced the initial size of the structural organs (from a length of 10cm down to 1cm) and added a genome-wide parameter controlling their scale. In this manner, plant height could be slowly increased, throughout successive generations, by similarly slow increments of this parameter. Initially set to 1 in the primordial population so as to allow the discovery of functional body-plan, it has an upper bound of 10 which would result in stems and root trunks of equal size as those of the previous experiment.

However, this modification is minor with respect to that undergone by the environmental controller itself as will be shown in the following section.

7.1.1 Environmental controller

Indeed, when demonstrating the capabilities of our previous model to cope with large populations over reasonably long periods of time, we focused our efforts on the plant and phylogenetic aspects of the simulations. The object controlling the environmental parameters was thus kept simple: a hand-crafted function displaying "interesting" dynamics.

As we moved onto this experiment on Environment-Driven Evolution, we replaced such an arbitrary construct by an encoding with similar power of expression with the added capability of being, itself, evolvable: Cartesian Genetic Programming [Miller et al. 2000] hence referred to as CGP. Given a range of inputs $I_0, \ldots, I_{n_i}$ and outputs $O_0, \ldots, O_{n_o}$, this model produces a directed graph of computational nodes which, when translated into its phenotypical form, can be used as a function set of size n_o (see appendix H for the full list of elementary functions) . In this work, we used 100 internal binary nodes and the inputs/outputs are identical with the previous experiment (see section 6.2), although with none of the "plausibility" constraints C_0, C_1.

(a) Initial genotype (b) Initial (neutral) phenotype

Figure 7.1: Initial state of the CGP (left) and resulting phenotype (right)

Moreover, only one row is allowed with no recursive connections, as is common practice with such a model.

Input variables D and Y which define the temporal position in a given year and epoch, respectively, required special treatment in order to prevent brutal transitions. Thus, instead of a linear progression from -1 to 1 they are modeled as sinusoidals thereby ensuring a continuity between successive years/epochs.

Generation of random genotypes was designed to produce only neutral phenotype as illustrated by figure 7.1. This was performed by generating random connection between nodes according to the usual rank restrictions stating that every node n at rank i can only receive input from lower ranked nodes. Afterwards, the first computational node is artificially set to the zero function and connected to every output node, thus resulting in a constant null output over every such variable.

With this method every initial CGP has identical phenotype, i.e. starting from the average environment depicted in figure 7.1b: every query from the environmental controller will return the same neutral values as were fed in. However, the genotypes contain a large amount of variability thanks to the unexpressed portion (99 nodes with the current configuration values, all of which having random functions and input connections). In this manner, though producing identical initial phenotype, strong divergence might occur after a single mutation.

This operation is performed according to the *Accumulation* operator advocated in [Goldman et al. 2013]. Given a CGP composed of n_h internal nodes and n_o outputs, the number of atomic fields that can be modified is $N = 3n_h + n_o$. Indeed, for an arity of 2, each node is composed of 3 addresses: the first defining its function and the last two referencing previous nodes or inputs. In this manner, an atomic mutation consists simply in selecting of these N fields and altering its value according to its local bounds. However, this tends to produce neutral mutations especially given the initial states of our CGP: with 99 inactive nodes only 6 fields out of the 303 (1.9%) would have an impact on the phenotype. The idea behind the *Accumulation* operator is to perform atomic mutations until one of the "active" nodes is altered. This asymmetrical rates of modifications between "junk" and expressed DNA allows for faster convergence rates on task-focused experiments and was thus deemed relevant

in this experiment for two reasons: a) every phenotype requires full evaluation thus multiple evaluations of identical controllers is a wasteful consumption of resources; b) the controllers will only be mutated a very small number of times (see section 7.2) when compared to traditional CGP experiments.

Two additional constraints apply on the updating of individual patch values. The first is an inertia coefficient kept at the constant value of $\alpha = .95$ for all runs. This is devised to reduce the speed at which environmental values are modified so that, from the plants' point of view, external conditions do not change abruptly (e.g. from -20°C to 40°C in a $1/10^{th}$ of a day). Given a current value v and a requested new of $\hat{v}$, the actual value $\dot{v}$ is given by

$$\dot{v} = \alpha v + (1 - \alpha)\hat{v} \tag{7.1}$$

This method is of particular interest to enforce usefulness from functions such as *rand* or *step* which would, otherwise destroy the plant population through catastrophic, instantaneous variations. Additionally, given that each update of the ground surface requires moving every plants and updating their cached data in the physics engine, the topography is only changed in one every ten simulation steps (e.g. each day) whereas the other two variables (temperature and hygrometry) are continuously updated.

7.2 Experimental protocol

Given the objective of exploring long evolutionary periods, traditional methods of optimization or open-ended evolution were deemed ill-suited to our current purpose. In order to alleviate this problem, we devised an evolutionary algorithm for timelines exploration based on the $1 + \lambda$ process used, for instance, in CGP. After a thorough description of this method, we elaborate on the hypothesis we are testing against and the various components of this evaluation.

7.2.1 Evolutionary algorithm

One of the key aspect of this experiment is the need for a long continuous process of sufficient duration for the Natural Selection to produce meaningful results but, at the same time, computationally tractable. With the very reasonable objective of monitoring 1000-years long evolutions (ten times the duration used in the previous chapter), we could not hope to use a generational algorithm in which each pair of population-controller would be evaluated for the whole duration. Indeed, simulation times were already somewhat of a bottleneck in the speciation test-bed, thus a tenfold increase in duration would have disastrous results for even a small number of generations.

On the other hand, mutating controllers on a single timeline was shown to have different but similarly catastrophic results where one hostile set on environment dynamics would wipe the plant population out, thus effectively terminating the evolutionary process.

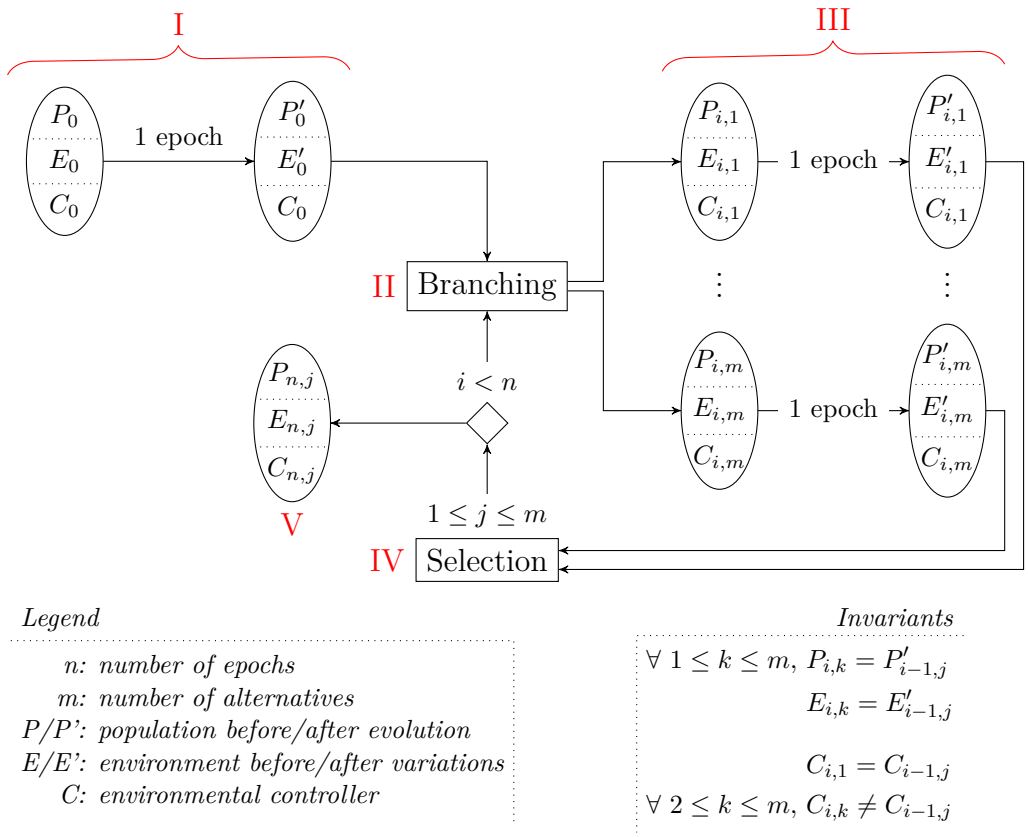


Figure 7.2: Environment-Driven Evolutionary Selection, see text for details.

As a matter of compromise, we settled on a hybrid approach: multiple, small evaluations that when merged together produce a continuous process of natural selection with potentially harsh environments but never cataclysmically so. One of the key aspect of this procedure is that plants are not directly subjected to artificial forms of selection (e.g. through fitness functions). Instead, whole ecosystems are considered when evaluating their relevance to the task at hand. External action is then performed at the level of the *environmental controller*, thus guiding the evolution through selective variations of the abiotic constraints. Summarized in figure 7.2 are the main steps of this Environment-Driven Evolutionary Selection (EDEnS):

Step I: initialisation From a base population produced in the same manner as in the previous experiment (single DNA duplicated 100 times and mutated) and the neutral environment presented in figure 7.1b, we perform a first epoch of uncontrolled evolution. In this case epochs are 4 years long with the same duration settings (10 ticks/day, 100 days/year). At the end of this short period (in evolutionary time), the plant population has undergone some change: with a maximal life expectancy of 10 days about 40 “natural” generations should have occurred.

Step II: mutation The provided ecosystem (either the result of the initialisation phase or the one selected in step IV) is duplicated, as-is, $m = 10$ times producing alternatives $a_1, \ldots, a_m$. This duplication performs a *deep-copy of the whole simulation* including, but not limited to, the plants (internal variables, organs ...) and the environmental variables. The controller is also duplicated but is additionally subjected to mutation for all alternatives except the first. In this manner even if all the new controllers generate unviable conditions, there is still an evaluation in which the population is expected to survive, given that it has done so in the previous epoch.

Step III: evolution As in the case of the initialisation, each alternative ecosystem is left free to evolve for 4 years. During this period, the controller is expected to behave in a non-neutral way, i.e. at least one of its outputs should exhibit a range of values different from zero either temporally or spatially. These variations in external conditions will change the focus of the evolutionary pressure imposed on the population. Indeed, in the initialisation phase most of the competition is against the biotic component of the ecosystem (i.e. inter-plants) with the leniency of the environment imposing only very mild abiotic constraints.

With the temperature reducing output capabilities, the hygrometry reducing water uptake and the topography incapacitating sexual reproduction, the nature of the game changes and so does the implicit fitness. Thus the population, as a whole, is expected to choose a different evolutionary direction to address these immediate difficulties. However, one should note that under all but the most dire climates inter-plant competition remains an ongoing challenge.

Step IV: selection Once all ten alternatives have completed their epoch, yet another combination of population-environment pairs is produced. Each is compared against all others according to a set of fitness (the specifics for this particular experiment are detailed in section 7.2.2) and a pareto front is computed. Every step i, one alternative, noted $a_{i,j}$, will be randomly selected from amongst the front to act as the "reality". From this, the new set of alternatives for the next epoch will be obtained, by looping back onto step 2, with $a_{i,j}$ as the reference for the population, environmental variables and controller.

Step V: result The loop from steps II to IV is performed for 249 epochs at the end of which a final triplet population/environment/controller is obtained. Given the dynamics of this algorithm both latter components are only of limited interest as they might only be the result of mutations occuring in a recent branching. Moreover, as the focus of this work is on the biotic component, the final population ($P_{n,j}$) will be the main focus of the following sections.

	Features
F_D	Genetic diversity
F_S	Inter-species standard deviation of genetic distance
	Controls
C_P	Population size
C_T	Simulation time

Table 7.1: Fitnesses used to guide the EDEnS algorithm

7.2.2 Fitnesses

The fitnesses we used to guide the evolutionary process are summarized in table 7.1 and can be classified into two groups. The feature-focused promote "interesting" characteristics in the underlying population while the control is tasked solely to maintain the evolution into comfortable value ranges.

The latter group is quite straightforward with C_T being equal to the opposite of the simulation duration (wall time) and rewards ecosystem that require shorter evaluation times. Indeed one of the drawback of our plant framework is the degeneration into extremely dense packs of tiny organism whose density across the whole spatial range tends to the theoretical maximum. The second fitness, C_P, has similar objectives:

$$C_P = \begin{cases} 2gauss(p, L, 120) - 1 & \text{if } p < L \\ gauss(p, U, 800) & \text{if } p > U \\ 1 & \text{otherwise} \end{cases} \tag{7.2}$$

Given an upper bound $U = 2500$, the function will exponentially decay down to 0 as the population grows larger than this threshold value. The *gauss* function being the same a defined in the previous chapter (the standard bell-curve, unnormalized). In this manner, overcrowding is selected against by the second portion of C_P, while the first, instead, is concerned with underpopulated ecosystems. With a lower bound $L = 500$, too small populations will be penalized more strongly: while overcrowding is a problem on a number of level, extinction is a much more dire threat in this algorithm.

Assuming that the control group of fitnesses works as expected, viable ecosystems will be produced though without any specific aim. This is addressed by the second group of fitnesses (F_D, F_S) which work by examining the plant population for desirable features. The latter is designed to promote speciation, first, by extracting every pair of potential mates belonging to different species:

$$M(P) = \{(f, m)/f \in P_f, m \in P_m, \text{sid}(f) \neq \text{sid}(m)\} \tag{7.3}$$

where P_f and P_m correspond to the partition of a population containing all female and male individuals, respectively while $\text{sid}(p)$ provided the species

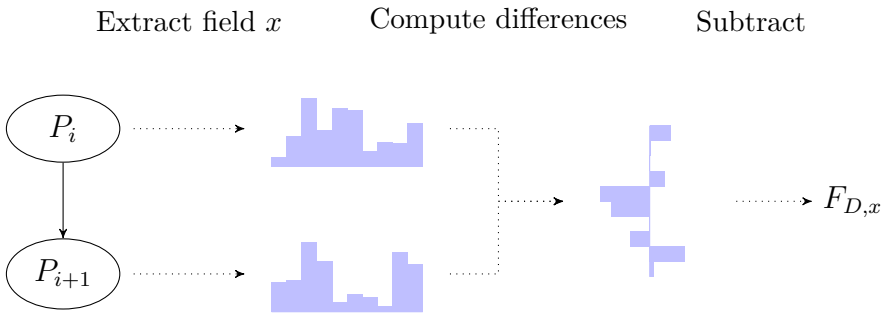


Figure 7.3: Genetic diversity through allelic distribution differences

identificator for an individual p. Thereby the score is obtained by computing the variance of genetic distance between these pairs:

$$\mu_{GD}(P) = \frac{1}{|M|} \sum_{(f,m)M(P)} \text{distance}(f, m) \tag{7.4}$$

$$F_D(P) = \frac{1}{|M|} \sum_{(f,m)\in P_f} (\mu_{GD} - \text{distance}(f, m))^2 \tag{7.5}$$

The variance is used to promote the presence of species that result either from a recent cladogenesis event or from long-standing speciation. Finally the last fitness F_D stimulates genetic diversity on an allelic level and is detailed below.

Genepool

To monitor allelic dynamics throughout the evolution, we relied on the concept of "genepool". Such a construct is built by parsing all genomes from a population and, for each field, converting them into histogram data. Indeed for discrete fields such as the number of seeds per fruits of the rules of each L-System, this is a straightforward process as the values already have a small, finite number of alternatives. Continuous fields such the growth speed, however, need first to be quantized into a more tractable number of "bins".

The rationale behind this methodology is to draw a parallel between different values of a given genetic field and allelic competition in natural DNA. Indeed, each alternative for e.g. the L-System recursivity can be seen as struggling against all other to be fixated into latter generations. Thus, by collecting the distributions of these alternatives amongst the population, we can gain a picture of the genetic space it is exploring. Such a genepool is collected both at the beginning (i.e. after step II) and end (before step IV) of each epoch and the genetic diversity of individual fields is computed through the differences between the corresponding histograms as illustrated by figure 7.3.

The genome-wide fitnesses F_D is then simply obtained by summing across all available genetic fields. Though slightly heavy in terms of computational cost this fitness was designed mainly to avoid local minimum where ecosystems

would settle into favorable sub-regions of the larger genetic space degenerating into a simple optimization task of the plant-level genomes.

7.2.3 Hypothesis

With the framework now comprehensively detailed, we can frame the hypothesis investigated during the following experiment:

Hypothesis H1 (H1): *EDEnS promotes robustness in the population*

To determine whether H1 holds we opposed two groups of ecosystems:

control c, 1000-years unsupervised evolution in perfectly neutral external conditions. That is for 10^6 (1000 years * 100 days * 10 steps) simulated steps plants competed against one another in the static environment depicted earlier (figure 7.1b). The only constraint imposed on these runs will thus come from the biotic component itself: straightforward plant-plant competition.

evolved e, subjected to EDEnS for 4 years epoch, with one devoted to the initialisation and the remaining 249 epochs to the exploration of alternative timelines and genepool optimization on conflicting objectives. Indeed, as the environments were dynamic, and unpredictably so, we expected the resulting populations to be able to cope more efficiently with a broader range of external conditions. Each evolved run had a branching factor of 10 with the first alternative, as defined earlier, using the same controller as in the previous epoch.

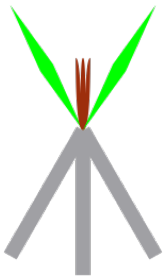
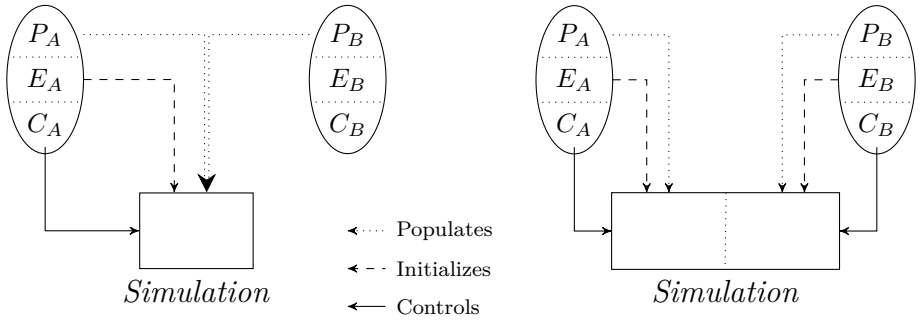


Figure 7.4: Base phenotype

Ten independent runs were performed for the evolved group e, while the control group c was only allotted 5 repetitions, due to its expected relative uniformity. This decision is further consolidated by the proportionally longer duration of simulation because of the lack of a limiting factor (fitness C_T in e).

All runs start with the same initial plant genome whose morphological component is shown in figure 7.4, the differences with that used in the previous experiment being both minor and noteworthy. In order to improve behavior (and thus limit excessive extinctions) in the face of potentially harsher external conditions in terms of temperature and water availability, we added a third root hair and removed the structural stem[1]. This results in a simpler and more production-focused morphology better suited to its future challenges.

[1] As observed in the previous experiment the increase in storage capabilities and potential gain in vertical competition was not deemed efficient enough by evolution. Given that individuals converged towards grass-like behavior, the initial genotype reflects this bias.



(a) Merged (b) Contiguous

Figure 7.5: Types of naturalisation evaluation, given the final populations, environments and controllers of runs A and B. a) Both populations (after filtering) coexist in the conditions given by A. b) Each population starts in its personal conditions, with no additional separation.

Similarly all base environments are identical (neutral) and the controllers also are duplicates. The only difference between the simulations lies in the seed used for their random number generators: in case of the *evolved group* there is only one such source of randomness: that which controls the selection of a given alternative on the pareto front. However, for the *control group* that seed is used to set the internal random number generator controlling all the chaotic aspects of the simulation: plants position, iteration order, reproduction ...

7.3 Evaluation

For all performed evolutions, we define c_i and e_i as the final ecosystem produced by the i^{th} run of the control and evolved group, respectively. When the distinction becomes necessary the exponents $^p, ^e, ^c$ will denote the population, environment (variables) and controller, respectively. That is c_2^p is the population of the second control run and e_5^c the controller of the fifth evolved run.

In order to transform the subjective value of "more robust" into quantitatively comparable metrics we subjected both groups to a pair-wise competition reminiscent of the process of naturalisation. If a population A when confronted to another population B ends up overwhelmingly dominating the test area, we can safely conclude that $A > B$.

With this in mind, we performed two types of such "naturalisation", summarized in figure 7.5, namely: Merged and Contiguous.

7.3.1 Merged naturalisation

Given two runs A and B, this form of naturalisation works by artificially injecting the population B^p into the whole A ecosystem. In practice, this imply that both A^e and A^c are kept as-is and only the populations are impacted.

Indeed placing plants from B^p into A would more often than not result in invalid interpenetrations between the two populations.

To bypass this problem, we first collect all seeds from A^p and B^p before planting them at their appropriate positions. In this manner, no interpenetration occurs and future collisions are resolved in the usual manner (i.e. growth is stunted until such a time as enough space is available).

Evaluation of success is then a straightforward procedure, or at least should be. While the biological metaphor implies that after some time we would just count how many of the remaining plants are descendants of either A^p or B^p, this does not necessarily hold in our model. Indeed, given the relative small size of our genotype space and the bail-out crossover for which reproduction rate between two individuals is always strictly positive (albeit potentially extremely low), hybridism is a very probable outcome. To drive the point further, for this test to work for one population against itself, we cannot simply count individuals: while the expected result from such a reflexive test would be close to null, the final population should be solely composed of hybrids of A^p and A^p. We thus resorted to a comparison of the respective proportion of inheritance from each initial populations.

To every plant in these evaluations, we assign a coefficient τ denoting how much genetic material they obtained from both populations. Thus at initialization every plant from the left-hand side and right-hand side populations are assigned $\tau = 1$ and $\tau = 0$, respectively. After each reproduction, the ratio τ_c for the newly born plant is computed from the parents ratios τ_f, τ_m by taking the average: $\tau_c = .5(\tau_f + \tau_m)$. The raw score of A versus B can thus be extracted from the population P by computing:

$$S = \frac{1}{|P|} \sum_{p \in P} \tau_p \tag{7.6}$$

in which we arbitrarily take *the left-hand side (host) ecosystem as reference*, the complementary score for the right-hand side (foreign) being simply $1 - S$. S then represents how much of the genepool of P originally belongs to A either by it having propagated new individuals in the environment or by contributing, through hybridism, to the genetic contents of the current population.

Once initialisation is complete we then let the evaluation run until either convergence or a deadline of 104 years is reached. The former is monitored by computing whether the absolute difference of scores between two consecutive years is less than an arbitrary threshold $T = .005$. If such a state is maintained continuously for three years, the simulation is deemed complete thus allowing for faster evaluation times in the face of overwhelming colonization.

7.3.2 Contiguous naturalisation

The previous evaluation highlights the capacity of a population in overtaking another given a home ground advantage to one of the competitor. In this sense,

the test is strongly biased towards the host population. Another method of defining robustness is to leave each ecosystem as-is thus preserving niches from both and instead gluing them together. This way, colonization is a slower, more active process requiring invaders to win against both the opposing population and environment. The metaphor behind this evaluation method is inspired by tectonic motion where different land masses come into collision with one another (e.g. India against the rest of Asia).

We initially used both ecosystems as defined in their final state but this was shown to induce undue advantages thanks to topological peculiarities in the environments. Indeed, quite independently of the underlying population, if the ground level on one side of the evaluation is much higher than on the other side then only elevated population will be able to send seeds into the other portion. When performing this form of naturalisation, we thus flatten out both environments to a zero elevation and deactivate the corresponding output from the controller thus effectively reverting to a neutral state for the topography dimension.

Score is computed in the same manner as the other method but this time highlights the invasive capabilities of the populations.

7.3.3 Scores normalization

In order to provide intelligible results, we performed a scaling and normalization of the scores obtained with both types of naturalisation to account for differences in population sizes. Indeed, if we denote S_0 as the relative proportions of population at the beginning of an evaluation, it is more often that not largely different from equilibrium (i.e. $S_0 = .5$). Thus the final score S_f is only relevant with respect to this initial value S_0. We thus define the normalized score $\hat{S}$ as:

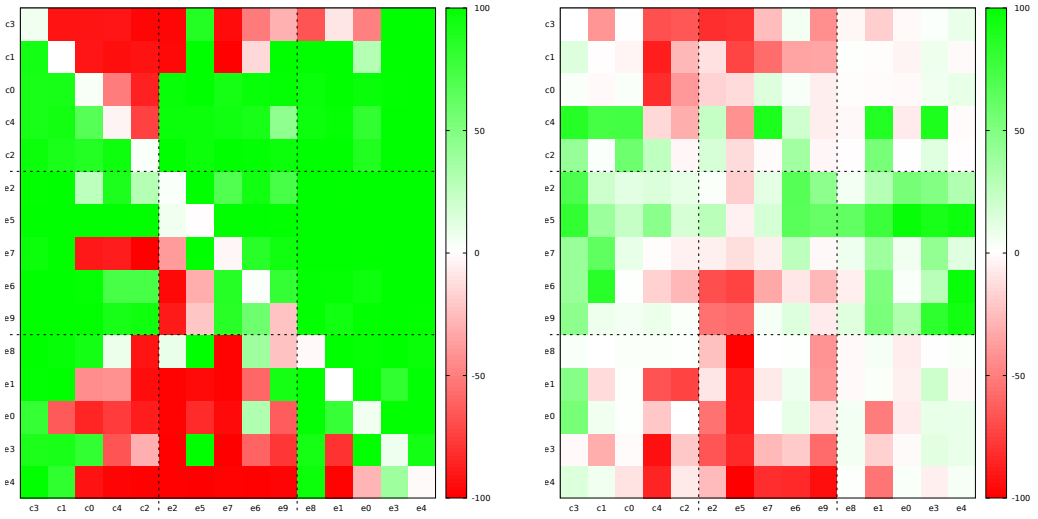
$$D = S_f - S_0 \tag{7.7}$$

$$\hat{S} = \begin{cases} \frac{D}{1-S_0} & \text{if } D \geq 0 \\ fracDS_0 & \text{if } D < 0 \end{cases} \tag{7.8}$$

This allows for clear interpretation of the results, however varied the initial distribution of populations: -100% indicates complete loss from the left-hand side while +100% represents an overwhelming victory. A neutral score of 0% stands for no variation between start and finish. Values between these extremes account for varying degree of success or defeat. Once more, the score for the right hand side is directly accessible, being only equal to $-\hat{S}$.

7.4 Results

In light of the evaluation methodology, the worst case scenario would be one were all competitions between the final ecosystems $c_0, \ldots, c_4, e_0, \ldots, e_9$ would result in close to zero scores. However, by looking at the pair-wise matrices in



(a) Merged (b) Contiguous

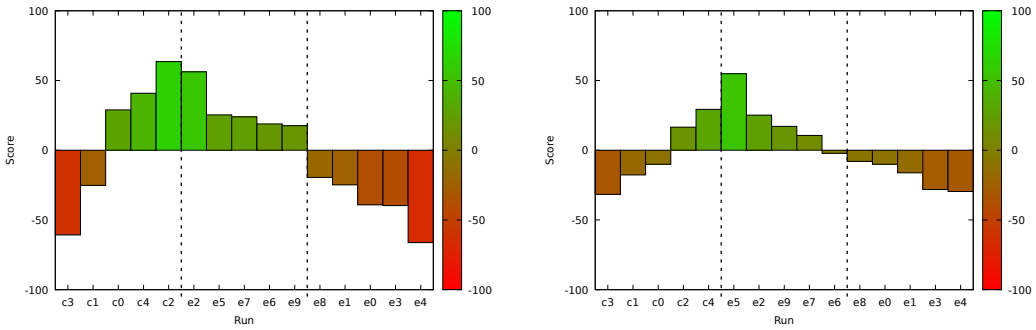
Figure 7.6: Scores for pair-wise evaluations in both scenarios. For every case, the score S is that of the left-hand side ecosystem (in rows) against that of the right-hand side ecosystem/population (in columns). Three extreme cases can be seen. White: stalemate with both populations maintaining themselves (e.g. diagonal of matrix a or c_0xc_3 in matrix b). Green: victory of left-hand side population (e.g. c_4xc_3 in both matrices or e_6xe_1 in matrix b). Red: victory of right-hand side population (c_1xc_4, e_8xe_5 ...). More details in appendix I

figure 7.6 we can see that this is not the case: the average score (in absolute value) for the Merged case is 81% (86.48 without the diagonal). Contiguous naturalisation was a much harder task and thus resulted in an average of only 30.27% even without accounting for reflexive confrontations. While on this subject, we can note that the diagonal is mostly neutral, especially when accounting from a reasonable amount of noise from the simulation.

We thus start by addressing whether or not H1 hold true in this experimental setting, before further exploring the contents of these ecosystems and tie dynamics of both plants and environments with the scores they obtained.

7.4.1 Hypothesis validation

Though reordered, so that the control and evolved groups are sorted in increasing and decreasing, order of their score, respectively, the raw amount of data prevents direct determination of which of these group performs better. We can nonetheless see a few definite differences: e.g. the control group follows a strict order relation is the sense that $c_3 < c_1 < c_0 < c_4 < c_2$. This is not as obvious in the case of the evolved group where though e_4 tends to loose most of its confrontations (red last row and green last column), there are cases where it survives (e.g. versus e_3) or even thrives (versus e_6).



(a) Merged (b) Contiguous

Figure 7.7: Aggregate scores showing the three subgroups

However, the question of whether or not EDEnS produced better populations than the control group is less straightforward to address. A measure of insight can be obtained by looking at the aggregate scores obtained by each run. That is for E, the set of final ecosystems, and $S_{a,b}$, the score obtained when ecosystem a competes against ecosystem b, we compute:

$$S_a = \sum_{b \in E} (S_{a,b} - S_{b,a}) \tag{7.9}$$

The summary of these colonization successes are displayed in figure 7.7 where runs are ordered in the same manner as the matrices of figure 7.6. There we can see that the relative success of the control group c versus the evolved one e depends on the task at hand. In the Merged naturalisation case, individuals are forced to share an environment to which only one of the competing population is, supposedly, adapted. Thus they are tested against not only on their capacity to survive in hostile conditions but also on whether or not they can out-reproduce their opponent. The alternative case, the Contiguous naturalisation, works at a slower pace: while a number of evaluation do not manage to gain a footing on the opposing ecosystem, those that do tend to have longer convergence rates: 43 years on average (out of a maximum of 104) as opposed to the 12.5 observed in the first alternative.

To formally test whether EDEnS did produce better population, we compared, for each run, the score obtained against members of group c versus those of group e. The group with the lowest value would then be identified as the hardest to beat, or conversely the strongest. As shown in table 7.2, a Wilcoxon test with this set of relative score leads to a strong rejection of H1 in case of the Merged naturalisation (first two rows). Results for the Contiguous type of evaluation are better but still largely above the threshold of .025 required by this one-tailed test. It follows that we cannot confidently say that EDEnS produced overall better populations.

If, however, we split the 10 individuals from e into two subgroups e^+ and e^- containing the five best and worse run, respectively, we can see an inverted

Test	Naturalisation	
	Merged	Contiguous
$c > e$	**.0134**	.8329
$e > c$	.9866	.1671
$e^+ > e^-$	**.0003**	**.0003**
$e^+ > e$	**.0003**	**.0003**
$e > e^-$	**.0003**	**.0003**
$e^+ > c$	**.0099**	**.0011**
$c > e^-$	**.0003**	**.0032**

Table 7.2: Statistical significance (p-values of wilcoxon test) for group ranking determination. The evolved group e can be split into two subgroups: e^+, containing the top-ranking 5 runs, and e^-, the rest. Bold values are significant at a .025 threshold and show that, even though both initial groups are of similar robustness, the subgroup e^+ consistently out-performs all alternatives.

trend. In both cases e^+ is found significantly better than the control group c which is, also consistently, shown to out-class e^-. From there we are faced with the problem of identifying what are the possible causes for such a broad range capabilities. Indeed, for EDEnS to perform satisfactorily, the key components of a good evolution needs to be identified: this will be the focus of the following sections starting with a brief overview of the morphological aspects of the produced populations.

7.4.2 Morphologies

In the previous experiment, we had commented on the strikingly simple amount of morphological complexity displayed by the evolved populations. Though, given the very fast initial life cycle, the 100 years they were allotted was large enough to allow for the exploration of a large panel of body plans. From the summary in figure 7.8, we can infer that the same dynamics were in effect in this setting.

Once more structural organs were deemed of no relevance by the evolution with no observed occurrence in any run at all epoch boundaries (i.e. every 4 years). This implies that the newly added component for inducing steady vertical growth was not used due to the lack of organs on which to act. Furthermore, the complexity, in terms of branching and vertical exploration, is on a similar level as previously observed. While e_1 and c_4 do produce high-reaching organs they seem to do so out of concern for their reproductive and seed dissemination ranges which are proportional to the flower's altitude. As a matter of fact most representatives tend not to cross the threshold of three vertical organs some, such as e_6, coping with only one level.

However, one should note that those representative were extracted from the

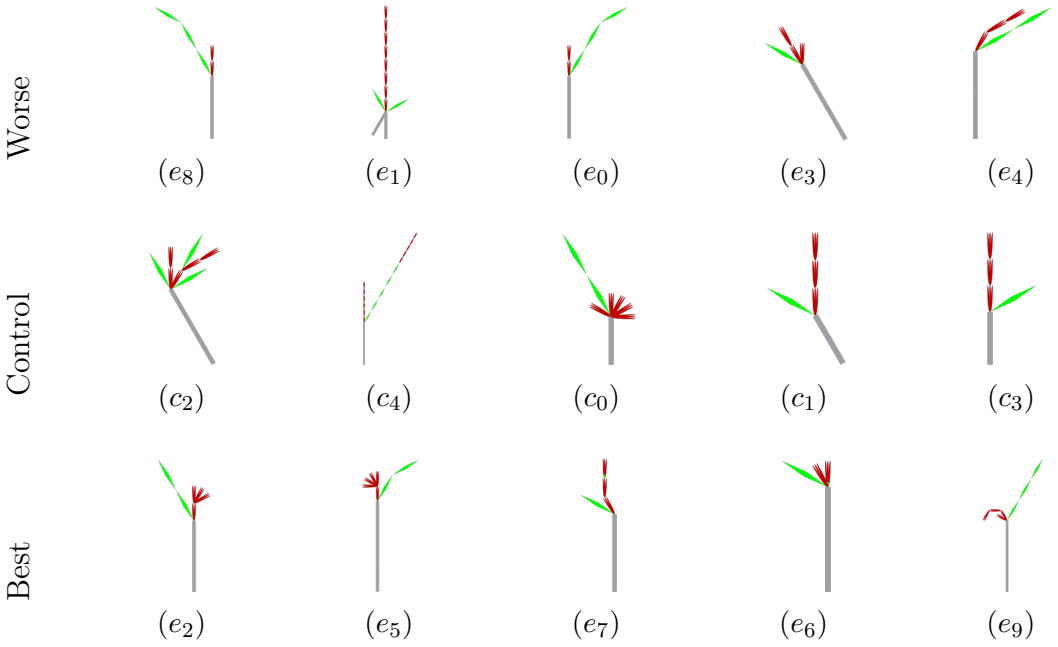


Figure 7.8: Morphologies of adult, un-fecundated plants from each run

final populations based on an arbitrary combination of frequency of occurrence and number of organs. Thus, though each gives a correct approximation of its source population, they do not represent the whole spectrum of possibilities (which can be seen with more details in appendix J). In fact, each population tends to be segmented into different maturity levels with, at the lowest, ungerminated seeds waiting for an opportunity to sprout. Then, depending on the level of interdependence between the L-System rules some might stay a long time in premature state e.g. with only the shoot or root finding enough space to grow.

Interestingly we can see that, out of these 15 representatives, multiple instances of independent convergences can be discerned. Even more surprisingly there is one such occurrence for each of the subgroups: e_8/e_0 for e^- with a perfect axial symmetry, c_1/c_3 with a slight difference in root orientation and e_2/e_5 in e^+ with more variation in terms of radial flowers count and root depth.

We can also note, as illustrated by figure 7.9, that body plans converge quickly. In fact, in the case of e_5, besides some small variations around the base morphology of one "stem" flower with 4 radial sprouts, no major change occurs after the 100th year. Furthermore, when looking more closely at this period, one can see that such an organisation was found (or rather became dominant) in a single 4-years epoch: between the 92nd and 96th years.

Thus in a such a short time a component as vital as the morphology has been fixed in the population and endured with almost no variation for the remaining 900 years. While this may indicate either a too small genetic space or the existence of excessively numerous local minimum (despite the expected

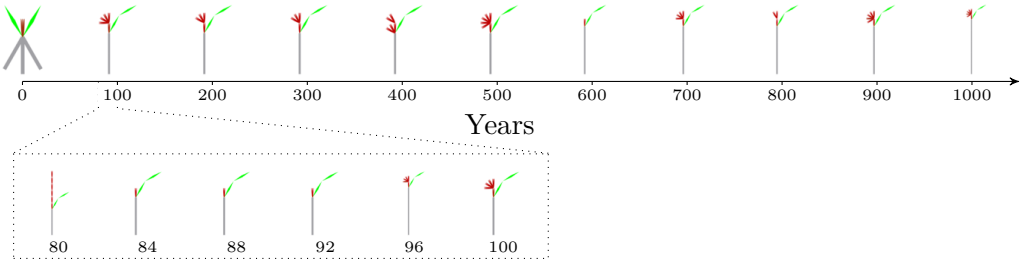


Figure 7.9: Morphological evolution of run e_5 and zoom on the 80-100 periods where convergence on the final morphology was achieved, specifically between years 92 and 96.

counter-action of fitness F_D), it may also be the result of the interaction with the environmental controller's evolution, that is an increase in external stress.

Nonetheless, there is not, at first glance, any striking differences between the morphologies of the different subgroups except in terms of flower disposition. Indeed, this type of organ being crucial for reproduction their placement should be one of the key components of success. While some relied on stacking flowers on top of each other, sometimes up to massive height as e_1, this strategy is a dangerous one: upon being fecundated all organs connected to the flower will be destroyed and their nutrients lost. Thus, given a lucky streak, c_1 might succeed in producing and disseminating 3 fruits but it might also only get the one chance if its first flower to be mated with is the lowest one.

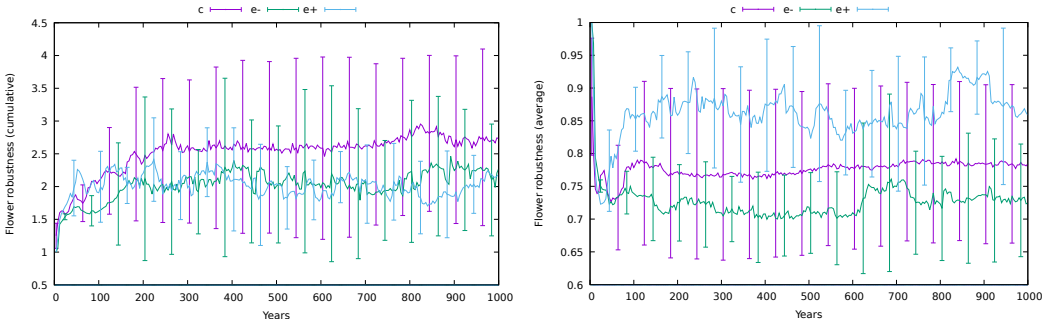
To test whether flower placement is indeed a valid measure of success we computed for each plant and each of its flowers a robustness metric, i.e. of flower independence. Given F_p, the set of flowers in a plant p and $F(o)$ the number of flowers in the subtree rooted at o (excluding itself) we define the robustness R_p of p as:

$$R_p = \sum_{f \in F_p} \frac{|F_p| - F(f)}{|F_p|} \tag{7.10}$$

$$R_p^n = \frac{R_p}{|F_p|} \tag{7.11}$$

with R_p^n being the normalized version of R_p, that is defining the robustness per flower of p.

As can be gathered on figure 7.10, there is a large variance in robustness across each group. Still, we can see that when grading plants based on the cumulative flower robustness, the control group fares better than both e^+ and e^- by a slight margin. We observe an inverse trend when comparing the grades of the normalized version where e^+ comes out first with scores between .9 and .85. Similarly c perform consistently better than e^-, the lack of environmental variations being clearly visible through the monotonousness of its dynamic. From these observations, we can conclude that, thanks to more lenient external conditions, the control group settled on a strategy of mass flower production



(a) Cumulative (R_p) (b) Average (R_p^n)

Figure 7.10: Flower robustness dynamics per subgroup

while subgroup e^+ optimized the position of each flower so as to maximize its success in the face of regularly hostile abiotic constraints.

7.4.3 Analysis of individual strategies

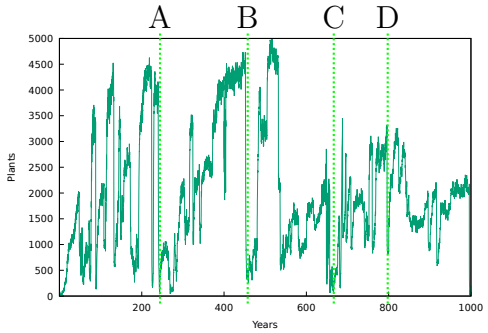
In order to gain a better understanding of the kind of events that produced and characterize the individuals from e, we detail, in this section, the specifics of one strongly performing ecosystem, e_5, and of the least successful e_4.

The case of e_5

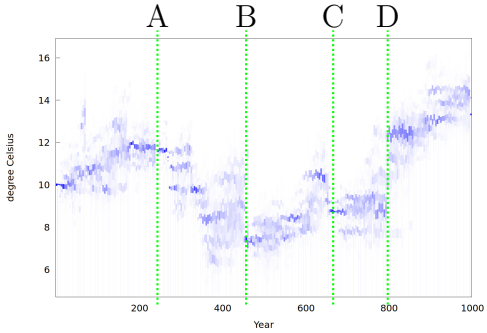
The fifth run from the evolved group showed great adaptive capabilities in both naturalisation evaluations: second and first (in e) in the Merged and Contiguous cases, respectively. It follows that we explored the dynamics of this ecosystem both in its final state and across its evolution in order to gain insight into the kind of events that may have caused such performances.

In figure 7.11, we display some of the most relevant aspects of the evolutionary process. In the top left corner is the dynamics in the plant population's size where we can see large variation corresponding to different controller providing changing sets of constraints to which the plants adapt, notably through selective trimming. Indeed, for the four variations labelled A,B,C & D on the graph, we can observe that right after the catastrophic loss of population, reproduction restarts very quickly, at least until another change in controller comes along.

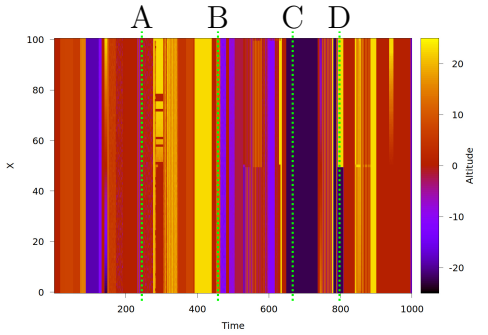
The top right corner picture shows the distribution of value for one crucial component of the plant metabolism: the optimal temperature μ_T. Given the range of environments the plants might face and the strong metabolic impact of excessive temperatures (see equations 6.5,6.9,6.10 for a reminder), fine-tuning of this field is one of the plants' primary challenge. The graph is organized as follows: for each epoch the distribution of value is extracted from the final population and quantized into a forward-facing frequency histogram. By merging the successive histograms thus generated we can observe the raw dynamics of a field even in the case of divergence (e.g. due to speciation).



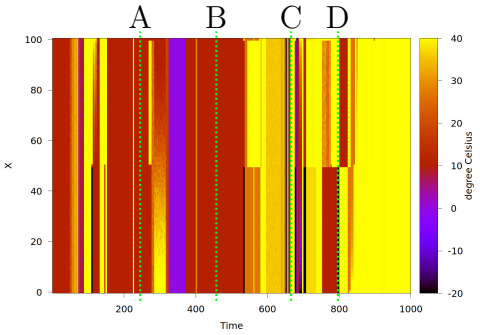
(a) Plants



(b) Optimal temperature

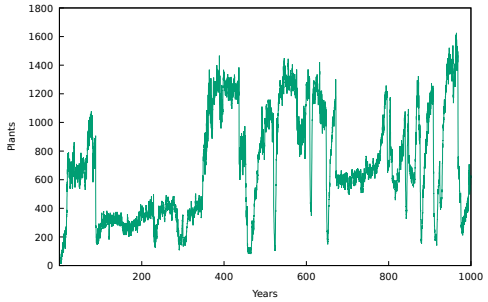


(c) Environmental topography

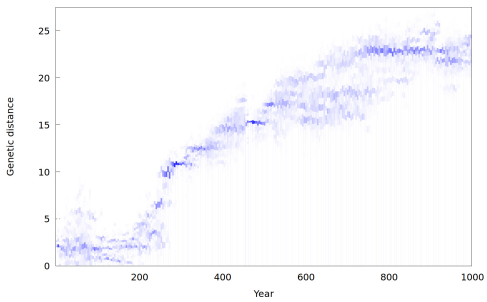


(d) Environmental temperature

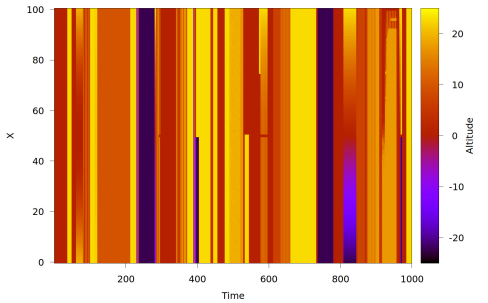
Figure 7.11: Dynamics of evolved run 5 (e_5). Major transitions indicated by markers A,B,C,D at years 240, 452, 664, 796, respectively.



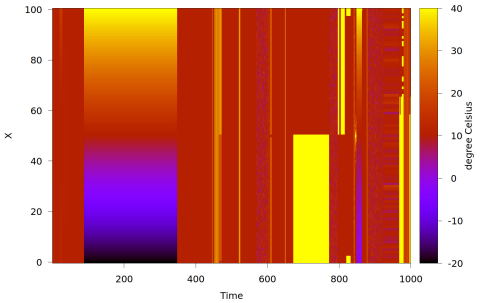
(a) Plants



(b) Optimal genetic distance

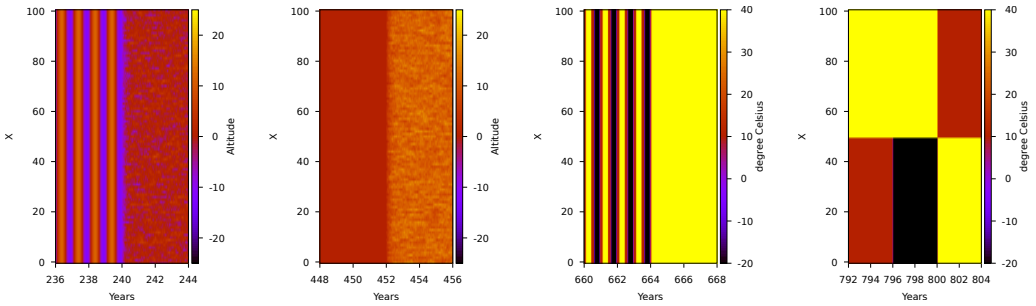


(c) Environmental topography



(d) Environmental temperature

Figure 7.12: Dynamics of evolved run 4 (e_4).



(a) Topography Epochs 59-60 (b) Topography Epochs 112-113 (c) Temperature Epochs 165-166 (d) Temperature Epochs 198-200

Figure 7.13: Details of the environmental dynamics around the markers defined in figure 7.11

From the initial value of 10°C, some exploratory variation appears early on until the 60th epoch (years 240-244, marker A) where the population enters a chokepoint. This results in a dramatic loss of allelic variability for this field which is soon compensated for but, this time, back down to about the initial value. Similar events occur at epochs 113 and 166 (years 452-456 and 664-668; markers B and C) with near-extinction drops in the population result in a tightening of the genepool around specific values whether through sheer luck or because of a better fit to the new environmental conditions. Marker D is set around another kind of transition pattern: from an average optimal temperature of about 9°C, there is a sudden jump, with none of the preliminary narrowing previously observed, leading to a population-wide change up to about 13°C. Given the relatively slow mutation rates (normal distribution of mean 0 and standard deviation .6) such a rapid variation could not have come from slow accumulation and must be put down the positive effects of the catastrophic trimming they faced.

Indeed, when looking at figures 7.11c and 7.11d, the whole diversity of variability introduced by the environmental controller is plain to see. All four previously mentioned markers can be linked to one of those two external variables, the water availability being somewhat withdrawn in this case. Figure 7.13 displays the environmental variables around each of these markers.

Surprisingly, markers A and B coincide with changes in the topography of the environment which, given the impact they had on the unrelated genetic field of temperature management, is nothing short of counter-intuitive. As a matter of fact, the transition from the 59th epoch to the 60th is marked by a stark change in the CGP output for the altitude variable as shown below:

$$\dot{y}_{59}(D) = tanh(max(D, D))$$
$$\dot{y}_{60} = rand$$

The smooth, periodic function which, given its spatial uniformity, is equivalent to a flat ground from the plants point of view is replaced by a widely chaotic use

of the random function (though the inertia factor tunes this down to tolerable levels). Due to the numerous interactions between topography and plant behavior (reproduction ranges, seeds dissemination, root/shoot competition), the exact reasons for the disastrous impact it had on the population is unclear. However, one thing is certain: from a comfortable occupation of the whole plot by slightly less than four thousands plants, the end of the 244th year sees a struggling population of only 138 individuals occupying only 15% of the available space.

Such a reduction in the genepool might be called biased for it favored certain alleles (i.e. allowed their fixation) even though their only quality might have been their location. A similar effect occurs around marker B with, once more, a smooth topography being harshly replaced by a randomly varying one.

$$\dot{y}_{112} = floor(\sqrt{rand})$$
$$\dot{y}_{113}(Y,t) = floor((rand < max(t,|Y|)))$$

Indeed, despite its use of the random function, $\dot{y}_{112}$ is effectively a constant function, outputting 0 across its whole temporal and spatial range. One mutation transformed it into something much more complex, although unlike $\dot{y}_{60}$ its randomness is turned towards positive values. The resulting bottleneck effect is thus slightly less traumatic for the plants which are "only" trimmed down to 363 individuals with multiple, small population centers.

Both other transitions were due to changes in the temperature output of the environmental controller, the first of which showing similarities to A:

$$\dot{t}_{165}(D,w) = step(min(round(D),w))$$
$$\dot{t}_{166} = 1$$

From a repetitive pattern using the build-in sinusoidal variations of D, which proved difficult but not insurmountable to the population, the output abruptly changes to a constant 40°C. In this case, there was only limited reaction from the population: already expecting small periods of extreme heat a portion was, by that time, suited to this kind of conditions, thus after a 2 year period of adaptation complete recuperation was well underway. This can be linked to the tightening of variability in the optimal temperature genetic field previously discussed.

We found the final major transition to be even more interesting not only because of the "jumping" behavior of the observed genetic field but also because it can be spread out across three epochs:

$$\dot{t}_{198}(x) = (Y^2 == x)$$
$$\dot{t}_{199}(w) = floor(\frac{1}{w})$$
$$\dot{t}_{200}(x) = floor(0 > \frac{1}{floor(x)})$$

For the last 20 years (up to the end of the 198th epoch), the temperature was constant with the left half at a comfortable 10°C and the right half at the maximal 40°C. Abruptly, the former dramatically plummets down to the minimum (-20°C) with no change in the latter. Unsurprisingly, this resulted in a massive extinction in the now glacial portion of the environment. When reaching the 200th epoch, the sudden exchange in temperature distribution further promotes adaptability and leads, after a recuperation period of 2 years, to more than half of the left-hand side being recolonized at the end of the 804th year.

From the observation of these diverse dynamics, it follows that for e_5 there were some positive catastrophes which by providing conflict pressures and tightening the genepool enhanced the capabilities of the population. In order to see whether or not this a recurrent event, we now turn to the least successful run of the evolved group e_4.

The case of e_4

The first difference with e_5 is the size of the population: with an average of about two thousands and peaks of more than twice that figure, we can consider this a decent density. On the opposite, the average for e_4 is a mere 700 with hardly more than 1300 individuals even at the best of times. Given the well-known effect of population size on genetic diversity and hence competitive success, this may go some way to explaining the poorer performances of this ecosystem. In addition, the bottleneck effect, previously identified, seems much less present in this case though this is also a consequence of the previous point: there is a much thinner margin for error with smaller groups.

This can be further extended to the variations of temperature as imposed by the environmental controller: though there still are some random epochs or even harsh transitions, they seem much sparser with more continuous periods between them. For instance, at the beginning of the evolution up to the 400th year, e_4^c has only visibly changed two times while e_5^c is closer to twenty. The same seems to hold for the hygrometry (not shown here) but not for the topography despite the pivotal role shown in the first two major transitions of e_5.

On the subject of genetic evolution, we observed no similar trend to that previously described for the optimal temperature field: the data only shows a random genetic drift. One other field that did undergo a strongly directed selection is that of the genetic distance. Indeed, from the default value of 2, we can observe a drastic increase starting shortly after year 200. With very little exploration of alternative alleles, except during the 600-800 period, this trend endured until it reached a maximum.

In fact the value of μ is high enough to transform each individual into its own species: the 604 plants at year 1000 correspond to 559 species (that is each contains 1.08 plants). The most reasonable explanation is an excessive compliance with fitness F_S which can easily be maximized by artificially in-

creasing the number of species. In light of this, the previous remarks are that much more interesting: it would seem that, when unconcerned by robustness, small populations with little variations both in terms of number of individuals and of temperature are a natural attractor. Of course this might also be the other way around, that is, due to its already challenging decision of a large optimal genetic distance, the population was unable to cope with more than the mildest of evolutions.

7.4.4 Population-level analysis

From the previous examples, we can see a number of dynamics to try and tie with success in the robustness evaluation. The objective of this section will be to look first at the global strategies in terms of relative efficiency before exploring which of the dynamics are indeed related to success in the naturalisation evaluations.

Strategies

As already noted during their introductions, both types of naturalisation require different strengths in order to win against the panel of competitors. Indeed, given the close-quarter conditions of the Merged case, reproduction efficiency is a feature on the same level of importance as tolerance to a large range of temperature. Similarly, in the Contiguous case, evaluations could very easily degenerate into stand-offs (and some do) where neither side succeeds in invading the other. Thus pure reproductive capabilities are not sufficient in this case, one must be able to thrive in both types of conditions at least slightly better than the opponent.

However, these are not the only strategies we can extract from these evaluations as summarized in table 7.3 where we only look at the rank obtained in a given task. In addition to *Rep* and *Col*, we can note that given the strongly anti-symmetric aspect of the Merged naturalisation we should expect similarly anti-symmetrical tendencies from the ecosystems. Indeed, high scoring runs can do so by resisting invasion (either through hostile environmental conditions or strong plant dynamics), a strategy labeled *Def* for it is obtained by looking at the row-wise scores (which are performed at home). The alternative is to be able to invade other environments (*Atk*) which corresponds to the column-wise scores (in which the population is injected into a foreign ecosystem).

In the case of the Contiguous naturalisation there is an expected symmetry for the only difference between evaluating A/B and B/A is which one is on the left side. Especially with no topography, there should not be a strong bias towards one relative position being more favorable than the other: *Sym* allows verification of this assumption. Finally, hybridism is a recurring occurrence given the use of our Bail-Out Crossover. As we observed that, to a certain extent, some evaluations relied heavily on such combination of the competing

Label	Semantic
Rep	Reproduction strength
Col	Invasion strength
Def	Defender, resists invasion
Atk	Invader, beats foreign constraints
Sym	Symmetry, performant on both right/left
Hyb_M	Hybridization tendency
Hyb_C	

Table 7.3: Strategies examined from the aftermath of both types of evaluation

populations to reach an equilibrium, we devoted Hyb_M and Hyb_C to grade the hybridization rates in the Merged and Contiguous cases respectively.

From there we obtain the grade table 7.4 which presents, for each run, its performance at the monitored task with respect to the rest of the sample.

General remarks As was already visible in figure 7.7, runs from the evolved group fare globally better than that of the c in the Contiguous naturalisation. This holds true even for the less efficient members of e^- although by a slimmer margin. Most notably, we can see that c_0 drops from the 4th position in the Merged evaluation down to the 10th while c_2, which outclassed every other run in the first case, ends at the end of the first third, seemingly swapping places with e_5. Once more we put this down to the differences in evolutionary stress imposed in both types of evolution. On the one hand, as the control group fought only against biotic constraints, optimisation was focused on out-performing competing plants (of the same species). The evolved group, on the other had to balance inter-plant competition with abiotic "hostility" thus promoting not only functionality but also the capacity to re-invade an environment with potentially wildly different conditions.

Defenders and Invaders We can also see, from the differences in rank for *Def* and *Atk*, that some individuals have very strong bias towards the type of strategy they can perform. Indeed while e_5 fares honorably in this evaluation type, it mostly does so by preventing invasion (i.e. wins at home) rather than by its capability to survive on a foreign soil. Oppositely, e_7 shows inverse tendencies, having more success in the invader role than the defender. This makes sense when looking at the environmental conditions of both ecosystems: while most of e_5 success is due to being able to survive in its own personal hell (40°C with 2/5 the base hygrometry), e_7 has more than hospitable parameters (10°C with twice the base hygrometry) although with a long history of more demanding environments.

Group	Run	*Rep*	*Def*	*Atk*	*Col*	*Sym*	Hyb_M	Hyb_C
e^+	e_2	2	3	2	3	6	7	8
	e_5	5	2	8	1	4	15	4
	e_7	6	8	4	6	13	9	10
	e_6	7	7	6	7	15	2	2
	e_9	8	6	7	4	7	5	1
c	c_2	1	1	1	5	9	12	7
	c_4	3	4	3	2	3	1	3
	c_0	4	5	5	10	8	3	11
	c_1	11	10	10	12	12	11	9
	c_3	14	14	14	15	10	4	5
e^-	e_8	9	9	12	8	14	8	15
	e_1	10	12	9	11	1	10	6
	e_0	12	13	11	9	2	6	13
	e_3	13	11	13	13	5	13	12
	e_4	15	15	15	14	11	14	14

Table 7.4: Ranking of each run in each type of strategy. Colors range from green (first rank) to green (last rank). Unabridged data set is available in annex K

Symmetry Surprisingly, we can note that some individuals do not perform similarly when evaluated on the left or right side of the Contiguous naturalisation. Though this can be explained in some cases by the morphological bias of the population (e.g. the "left-handedness" of c_0) for others, such as e_8, the improved performance when on the right side is at odds with its growth direction. However, further examination of appendix K.4 shows that scores variation range from .25% to 5.45%. Such a small magnitude is consistent with the expected internal variability of our system, especially in this type of evaluation where a pair of ecosystems interact with one another. As it stands, we can conclude that, within a reasonable margin, these populations have symmetrical performances.

Hybridism On a similar note the variation in reliance to hybridism is strikingly strong in some individuals: e.g. e_5 comes from being the worst hybridizer in the Merged case to the 4th place in the Contiguous case. While this could be explained by the fact that, in the latter, invasion is a much longer process thus leaving ample time for cross-population mating to occur and stabilize, others runs (c_0, e_0, e_8) exhibit a very strong bias in the opposite direction. As a matter of fact, hybridism can be a very efficient way to quickly adapt to a foreign environment by incorporating relevant alleles from the local population. However, it is not a robust strategy for the probability of obtaining said alleles

is low enough that a large number of hybrids must be attempted before a successful one is found.

In the end, each run, even those of the control group, seem to have found different means to reach a competitive state. Though some do fare better overall, the specific of each individual strategy hint at the number and complexity of the involved dynamics.

Successful plant variables

Thus, in order to determine what made a successful population, we examined whether some plant-related metrics could be correlated with success in either evaluation. Amongst the numerous variables examined, we only present here those that shed some light on the matter as summarized by table 7.5. For each studied pair of variable we investigated the Spearman correlation when computed on all 15 runs (c & e) and solely on the evolved group. Additionally, if a positive correlation is obtained for one type of evaluation, the correlation for the other type is also displayed, regardless of significance so as to provide comparison material.

In the spirit of concision the variables have been shortened according to the following glossary:

- Merg. & Cont., scores obtained in the Merged and Contiguous naturalisation evaluation methods, respectively
- BTP, the Birth rate per Tick per Plant (over the last year)
- R^n, the flower robustness metric as defined in equation 7.11
- Max, Maximal number of individuals (starting from the 4.5th year)
- SPF, Seeds Per Fruits (genetic parameter)
- σ_T, plant's acceptable temperature range
- $\bar{P_l}$, average plant count throughout the evolution
- $P_r, P_r^+, \ldots, \bar{P}_r$, perturbation metrics (equations 7.12-7.15, below)

Given $P_l(t)$ the population size at time t for the considered run, the following equations describe a "perturbation" metric which, shortly put, measure how much the plants have been disturbed by controller variations.

$$d_e = P_l(4(e-\nu)) - P_l(4(e+\nu)) \qquad P_r = \sum_{e=1}^{249} \frac{|d_e|}{249} \tag{7.12}$$

$$d_e^+ = \begin{cases} d_e & \text{if } d_e > 0 \\ 0 & \text{otherwise} \end{cases} \qquad P_r^+ = \sum_{e=1}^{249} \frac{d_e^+}{249} \tag{7.13}$$

$$d_e^- = \begin{cases} d_e & \text{if } d_e < 0 \\ 0 & \text{otherwise} \end{cases} \qquad P_r^- = -\sum_{e=1}^{249} \frac{d_e^-}{249} \tag{7.14}$$

Variables		c & e	e only
Merg.	Cont.	.821	.939
Merg.	BTP	.561	.333
Cont.	BTP	.136	.103
Merg.	R^n	.493	.394
Cont.	R^n	.521	.527
Merg.	SPF	-.314	-.782
Cont.	SPF	-.496	-.867
Merg.	σ_T	.043	.745
Cont.	σ_T	.386	.709
Merg.	Max	.579	.685
Cont.	Max	.429	.685
Merg.	$\bar{P}_l$	.589	.648
Cont.	$\bar{P}_l$	.421	.697

Variables		c & e	e only
Merg.	P_r^+	.600	.612
Cont.	P_r^+	.607	.600
Merg.	P_r^-	.589	.661
Cont.	P_r^-	.679	.733
Merg.	P_r	.625	.697
Cont.	P_r	.682	.733
Merg.	$\bar{P}_r^+$	.039	.273
Cont.	$\bar{P}_r^+$	.393	.273
Merg.	$\bar{P}_r^-$	.193	.927
Cont.	$\bar{P}_r^-$	.571	.927
Merg.	$\bar{P}_r$	.125	.758
Cont.	$\bar{P}_r$	.511	.794

Table 7.5: Spearman correlations between scores and population-dependent variables. Grayed-out values are not significant under a p-value $< .05$ threshold

$$\bar{P}_r = \frac{P_r}{P_l} \qquad \bar{P}_r^+ = \frac{P_r^+}{P_l} \qquad \bar{P}_r^- = \frac{P_r^-}{P_l} \tag{7.15}$$

With $\nu = .25$. In its most simple form, P_r is simply the average variation in number of individuals around each epochs. In case of runs from the evolved group, this measures the impact of potential changes in the environmental controller on the underlying plants dynamics. However, even though they are not subjected to these modifications, P_r can also be computed for the control group c thus providing a baseline of the background noise. The two variations P_r^+ and P_r^- focus respectively on the positive and negative changes in population size. The latter indicating the average drop induced by hostile sets of environmental dynamics while the former accounts for more hospitable conditions. Finally to mitigate the effect of the population size on the amplitude of these metrics, $\bar{P}_r, \bar{P}_r^+, \bar{P}_r^-$ are normalized by the average number of individuals.

Circling back to table 7.5, we can, first of all, see a strong correlation between the score obtained in one type of evaluation versus that obtained in the other. Unsurprisingly, the effect is more pronounced when only considering the evolved group which performed more consistently than c. Both BTP and R^n are only correlated in one out of four contexts, and even then, mildly so. Though these are not counter-intuitive (fast birth rate is good in Merged evaluation and flower robustness is a plus in the Contiguous case), they are mostly model-dependant and potential false positives.

The next two are not only more interesting but also seemingly more reliable: significant correlations for SPF and σ_T have a strong effect and are consistent in both types of evaluation. Furthermore, the fact that they show no correlation

when including the control group indicates that this trend is only present in populations resulting from EDEnS. Indeed, given the competition for resources, having a low number of seeds per fruit increase the chances of successfully generating enough biomass for each offspring to be viable. Similarly, through a steady increase in the acceptable temperature range σ_T, individual plants were most likely to survive variations which was the expected response. Combined with the observed biais of selecting warmer environments this effectively allowed populations to survive variations in the [10°C,40°C] almost unhindered.

Both of the last variables on the left side of the table are concerned with the relative sizes of the population throughout the evolutionary process and, although only moderately so, show a strong consensus. The score in Contiguous naturalisation, however, does not follow the same trend when tested against all 15 runs. This stems from the fact that, while a large population is an advantage for it allows access to a larger pool of potentially useful mutations, pure optimization is also a viable option when there are no catastrophic variations in external conditions. Indeed the most successful of runs from the control group chose different strategies with respect to this parameter: 1300 individuals in average (up to 2000) for c_4 while c_2 reaches up to 8000 with a mean figure around 2500. Such a wide variability is not observed in the evolved group: 4 runs from e^- never obtain populations of more than 2000 individuals, even in peak conditions as opposed to e^+ for which the average figure never falls below this threshold. This size effect is, in itself, not a novel concept mentioned as it was in [Darwin 1859, p. 120]. It is nevertheless a point in favor of our framework that it can exhibit such well-known natural dynamics.

To conclude on this section devoted to the population-related variables that can be linked to success in both evaluation types, we now turn our attention to the right column of table 7.5. Given that when examining stereotypical individuals we noted the seemingly strong impact of environmental perturbations, we expected to find some relationship between the score and the amount of such disturbances. At first sight, we may conclude that this assumption holds given the large number of statistically significant correlation observed. A different picture is painted, however, when considering that when including runs from the control group no correlation was expected. As they are not subjected to different environmental controllers their variation in population count are solely due to the randomness of their internal dynamics. Thus, when looking at the distribution of values (visible in appendices L.1a and L.4a), the fact that runs from e^- showed similar amount of perturbations as those from c came as a surprise. Indeed, this would imply that the worse performing members of e did not undergo an environment-driven evolution, rather more of a drift in varying environments.

However, considering the low population count of e^- we tested whether such a discrepancy was not caused by size effects, alleviated by $\bar{P}_r$ and its variations. With these less-biased metrics, results are much more plain to see: a very strong (and surprisingly uniform for both types of evaluation) correlation is detected

for the negative variations of population. Indeed with these normalized version of P_r, all three subgroups are clearly separated: the medians for $\bar{P}_r$ are .032, .075 and .108 for c, e^- and e^+, respectively. Thus while individual runs from e^- did not show as much perturbation as those the more successful subgroup e^+, they still are significantly above the background noise exhibited by c.

This explains the almost complete lack of correlation of the normalized version when tested on the 15 individuals. It follows that P_r^+, which was only significant when including the control group no longer shows any significance, thus rejecting the possibility of a link between transition from hostile to hospitable environments with robustness. Oppositely, $\bar{P}_r^-$ exhibits the *strongest correlation* in all tested variable pairs (with the exception of both score types), suggesting that this is a key component of success. Unsurprisingly, $\bar{P}_r$ is slightly less strongly significant due to its aggregation of two variables with diverging effects on the viability during naturalisation.

7.4.5 Environments and controllers

These variables, however, are resulting not only from population-level dynamics, on which we have no direct control in this framework, but also from environment-induced events which are the designed lever of action. In order to explain how the successes observed in the previous section can be linked to the dynamics and selection methods of EDEnS, we performed a similar analysis on environment- and controller-managed variables.

Once more we only present those that did show positive, consistent correlations with not only the scores in Merged and Contiguous naturalisation. With $v \in \{t, h, w\}$ an environmental variable type (temperature, heat, water, respectively), $s = .25$ a sampling period, and $V_v(i,t)$ the value of variable v in the patch at index i at time t (in year), the average amount of spatial variation is computed by:

$$\frac{dV_v}{di}(i,t) = V_v(i-1,t) - V_v(i,t) \tag{7.16}$$

$$S_v^\delta = \frac{s}{1000I} \sum_{j=1}^{1000/s} \sum_{i=1}^{I} \left| \frac{dV_v}{di}(i, sj) \right| \tag{7.17}$$

Similarly the number of times a environment with variable conditions is observed is:

$$S_v^\# = \sum_{j=1}^{1000/s} \begin{cases} 1 & \text{if } \exists i / \frac{dV_v}{di}(i, sj) \neq 0 \\ 0 & \text{otherwise} \end{cases} \tag{7.18}$$

Both metrics measure similar quantities but with a different focus: the former computes the intensity of the average difference between neighboring patchs while the latter is only concerned with annotating whether any such difference did occur. From a broader perspective we also monitor the global variation

Variable	ρ	
	Merg.	Cont.
S_h^δ	-.673	-.794
T_h^δ	-.661	-.648
T_h^σ	.818	.818
$S_t^\#$	.806	.806
S_t^δ	.770	.806
T_t^δ	.709	.770

(a) Consistently significant

Variable	ρ	
	Merg.	Cont.
S_w^δ	.333	.200
T_w^δ	.358	.297
T_w^σ	.067	-.127
$T_t^{\sigma,249}$	-.255	-.255
$T_h^{\sigma,249}$	.228	.419
$T_w^{\sigma,249}$	.012	.250

(b) Non statistically significant

Table 7.6: Spearman correlations between scores and environmental variables

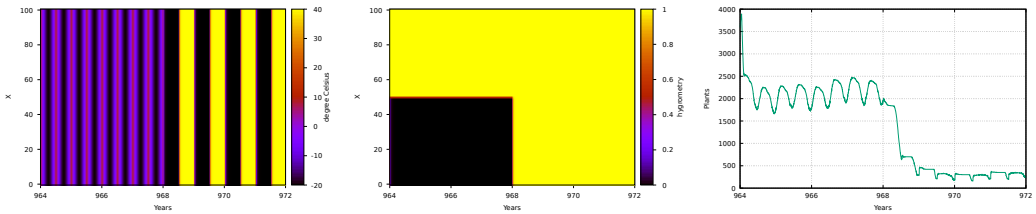
exhibited throughout the evolutionary process via the following equations:

$$T_v^\delta = \frac{1}{NI}\sum_{j=1}^{N}\sum_{i=0}^{I}\left|\text{scale}_v\left(\frac{dV_v}{dt}\left(i,\frac{j}{1000}\right)\right)\right| \tag{7.19}$$

$$T_v^\sigma = \frac{1}{NI}\sum_{j=1}^{N}\sum_{i=0}^{I}\left|\text{scale}_v\left(V_v\left(i,\frac{j}{1000}\right)\right)\right| \tag{7.20}$$

with $N = 1e6$ the number of simulation steps (10 steps per days, 100 days per years for a thousand years) and $\text{scale}_v(\text{x})$ a function transforming variable v back into the $[-1, 1]$ range. Through T_v^δ we can measure the gross amount of variation in a given dimension regardless of its type (temporal, spatial, between controllers). Given our assumption that dynamical features from the abiotic component of an ecosystem is one of the key aspect of its eventual success, we expect these to be relatively high. On a related approach, T_v^σ computes the average amount of divergence exhibited by the environments, i.e. how much they diverge from the initial, neutral state.

We summarized in table 7.6 the foremost correlations obtained between the previously defined metrics and the scores obtained in both types of evaluation. Starting with the left portion (table 7.6a), we can see, that both the temperature and topography were found to impact the resulting populations, though in a negative way in case of the former. Indeed, not only is there an inverse relationship between score and evolutions with large amounts of spatial variation (S_h^δ), but also in terms of gross variation (T_h^δ). However, this is somewhat contradicted by the strong link observed with divergence from the neutral state (T_h^σ) indicating that only excessive variations are deleterious while moderate levels of stress are more likely to be concomitant with robustness. In case of the topography, all correlations denote a positive influence, notably for S_t^δ and T_t^δ which seem to have a similar strength of impact. The presence of a strong link with $S_t^\#$ favors the assumption that the magnitude of the topological variations was not the deciding factor for this metric is only focused on the number of occurrences.



Temperature Hygrometry Population

Figure 7.14: Example of extreme variations (e_7 around epoch 242)

On the other side, table 7.6b shows that surprisingly the hygrometry did not play any consistent role in the evolution of robustness: none of the 10 tested variables, 3 of which are shown here, show any statistically significant results. Though punctually it induced some catastrophic variations in population (such as the last epoch of e_5), this indicates that, in our current implementation, water-related dynamics did not contribute to the success of the evolved runs.

Additionally, we can note that $T_v^{\sigma,249}$ was consistently not found significant for any combination of environmental variable and score. This variation of T_v^{σ} performs a similar computation of divergence from the neutral state but solely based on the data of the last epoch, that is the environment in which the evaluations are performed. Such a lack of correlation thus shows that despite the possible bias introduced by the combined evaluation of plant/environment the harshness of the latter is not an accurate way to predict the success of the former.

In light of these linkage between environmental dynamics and robustness we attempted to classify the various changes in controller-induced dynamics but quickly realized that such an ontology would be massive. Indeed, though the neutral state was the most frequently observed feature (11.5%, 27% and 18.7% for the topography, temperature and hygrometry, respectively) the rates of occurrence were much lower than expected before-hand. No link was found between the quantity of neutral states encountered during the evolutionary process and the scores of the resulting population, thus showing that the poorly fairing runs were not doing so for lack of environment-induced stress.

In fact quite the opposite occurred: out of the 5% strongest perturbations endured across all runs, 95 of them were negative ones (i.e. catastrophic trimming). Given the inertia of the CGP, it is unsurprising that blindly devising difficult environments is more likely that more hospitable ones. A large number of such transitions take very simple forms such as mild temperatures suddenly raising all the way up to 40°C, but some were much more intricate as in the example of figure 7.14. From a pattern of cold to glacial successions (with a period of half a year) the change of controller at the start of epoch 242 produces even harsher oscillations up and down the available temperature range with very short phase change durations. Such a drastic alteration of abiotic conditions is, luckily, compensated by a larger availability of water which allow

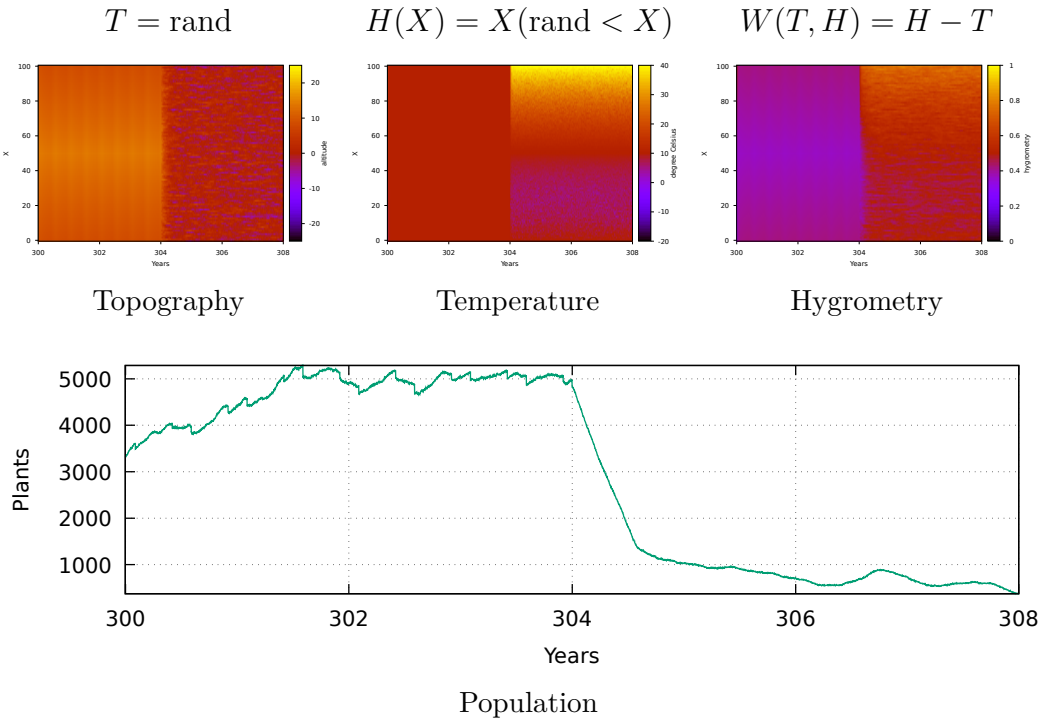


Figure 7.15: Creative compositions of the `rand` function (e_2 around epoch 76)

some plants to survive long enough for a stable pattern of population size to be restored.

Another area in which the evolutionary process proved surprisingly resourceful is in using the random function. Indeed due to its arbitrary aspect, one could expect that the environments exhibiting such wild variations would be shunned during the selection process, but, as illustrated by figure 7.15, this was not always the case. Indeed these plants showed remarkable resilience to such hostile dynamics: while the temperature exhibits a large-scale pattern, individual conditions are still quite erratic and, in case of the topography, the whole range is covered. Nonetheless, after a somewhat short period of intensive self-trimming, the population starts to level itself out though the final trends implies that a change of controller might be the only way to prevent extinction.

7.5 Conclusion and avenues of inquiry

In this section, we introduced a novel evolution method in which a population of individuals is subjected to external sources of stress so as to promote robustness in a manner hand-crafted environments would struggle to achieve. Through a variation of the $1 + \lambda$ algorithm, we generated a sample of 10 runs which when confronted to the control proved ambivalent in their capacity to cope with competition. Indeed, the lack of a clear-cut objective resulted in an unbridled evolution which prevented early convergence but at the same time precluded

the possibility of easy comparison between the final populations as well as giving a non-negligible advantage to the control group.

With only biotic constraints to compete against for a thousand years, individuals from this group have developed strategies solely focused on a plant versus plant struggle. Oppositely, the evolved group had to compromise between internal and external competition thus producing more all-round individuals who showed better proficiency in colonizing hostile territories. Despite these setbacks, we observed that the final set of ecosystem could be divided into two subgroups with radically different robustness.

While this might be a result of evolution falling into partial local minimum (i.e. in case of the morphology which was quickly settled), some other aspects of the plants' genomes underwent constant variation with respect to their changing environments. Foremost amongst them are the fields controlling the temperature management which was under considerable pressure from the corresponding abiotic constraints.

While individual examination of evolved run 5 (e_5) showed a great deal of desirable dynamics both in the population and its environment, such an extensive task could not be performed for all ecosystems. We can reasonable assume that most of the other runs would present similar amount of interesting demeanor, though a thorough investigation would require much involvement.

On a broader perspective, we found that negative perturbations were one of the key component of robustness: evolved runs in which catastrophic trimming were a relatively common occurrence were found to perform significantly better. While we cannot determine which way the causality links goes (assuming no third variable is involved), we do note that EDEnS was designed with just such a situation in mind. Indeed the counter-intuitive fact that massively stressing a population would improve its efficiency is not unreasonable in itself: defining exactly just how much and when to apply such stress is a much more intricate task to perform manually.

Additionally, we extracted a number of environment-level features whose role in guiding the evolutionary process were found relevant: the positive impact of topography, the negative one of the temperature and the global-level neutrality of the hygrometry. Though in case of the latter, we could nonetheless observe relevant local transitions in which the population has been deeply impacted. This highlights the need to redesign the water dynamics which, for the sake of simplicity, were kept very elementary (notably with regards to the infinite reserves). Even so, individual controllers with highly regular or extremely chaotic patterns were produced and used by this evolutionary algorithm.

In the end, EDEnS did prove its mettle by producing competitive populations which fared honorably in spite of unfavorable odds while, on the side, allowing for the investigation of general knowledge regarding the evolution of successful life-forms. A short version of this experiment has been submitted in Godin-Dubois et al. 2020.

Chapter 8

Conclusion

Throughout this manuscript we described the progress of our research from a morphology-focused model into an evolutionary framework. This section will present the more salient aspects of this work before broaching possible avenues of research either left open or requiring more attention.

8.1 Summary

In terms of scientific contribution, three points are most notable in the research we performed, namely the application of Sims' directed graphs to plants, the design of an autonomous reproduction scheme which allowed for automated speciation monitoring and the prototyping of an evolutionary algorithm in which the environment is a crucial, driving, component of the system. These will be detailed in turn in the following pages, starting with our least recent contribution.

8.1.1 Plants and Graphtals

Graphtals, historically used to generate motile creatures of staggering complexity in the seminal work of [Sims 1994b], have been strongly linked with animal morphology. The application of such a controller to plants was a novel approach which resulted in body plans that strongly differed from those obtained by more usual models. Though some showed structural regularities reminiscent of biological instances, more atypical morphologies were also obtained.

Embedded as they were in a dynamical environment where the light source varied in position and altitude across time and where rain patterns were providing fluctuating and unpredictable water availability, these artificial plants developed efficient foraging strategies mostly through the use of repetition. This set of constraints generally converged towards small, rampant morphologies and the induction of an additional source of stress, in the form of static grass blades, was required to obtain an artificial form of vertical competition which promoted the emergence of upward-reaching complexity.

8.1.2 APOGeT and the Bail-Out Crossover

The second contribution is two-fold as it entails not only the autonomous reproduction scheme but also the phylogenetic tool build around it. Through the use of a non-deterministic and non-systematical algorithm for determining whether two, potentially very different, individuals will produce an offspring as the result of an intercourse, we allowed for speciation barriers to spontaneously emerge in a fuzzy manner. Indeed genetic compatibility is not viewed as a binary operation but as a continuum, though with exponentially decaying rates.

By encoding such a function directly into evolvable genetic fields, these barriers can react to external changes either in the environment or the rest of the population. They can also adapt to changes in the internal dynamics of a given species especially in cases where the genome is of a varying size, thus requiring increasing values for the optimal genetic distance as complexity progresses. Furthermore, the independent management of inbreed and outbreed coefficients permits diverging strategies from diversity-promoting to closed species boundaries.

The compatibility function at the core of this algorithm was further leveraged to create APOGeT, a monitoring tool capable of generating phylogenetic trees of an evolving population. At the center of this tool is the concept of defining a species by a (small) set of its representative individuals hereby allowing the use of genetic representations with non ordinal values which preclude aggregation methods such as the mean. This, in turn, is used to generate a species boundaries in the genetic space, which is not necessarily composed of a single homogeneous subset.

By combining the representative set with the compatibility function of the BOC such a boundary is obtained by imposing a sufficient degree of similarity with the current "typical" features of the species. Managed by a user-specified threshold, this allows for varying degrees of strictness which influences the complexity of the resulting phylogenetic tree. Additionally, APOGeT handles both asexual and sexual reproduction modes, the latter requiring further treatment due to the potential presence of hybridism. Indeed, when an individual's parents do not belong in the same species the number of candidates nodes in which to insert its genome increases, resulting in a similar increase in complexity. By keeping track of such events, we are able to determine the exact proportion of genetic inheritance each species obtained from its contemporaries.

Graphical tools were also developed to enable straightforward visualization of the species' hierarchical structure and gross preliminary analysis. In the two previous chapters, these were used to illustrate the speciation capacities of our system.

8.1.3 EDEnS

As defined in our problematic, environments in virtual artificial creatures experiments were generally lacking in heterogeneity or dynamical properties. The Environment-Driven Evolutionary Selection is a novel attempt at producing a framework for exploring traditional ALife experiments with a stronger focus on the role of abiotic constraints in the evolution of complexity. This framework is derived from the $1 + \lambda$ algorithm, the difference lying in a parallel evolution of identical populations under different conditions.

This work started out by using hand-written equations for the values of three environmental variables: topography, temperature and hygrometry. By reproducing text-book cases of speciation pattern we were able to show that, under supervised conditions, the population were indeed diverging into their own, reproductively isolated, niches. In addition, the variation of environmental constraints were found to have a profound effect on the dynamics of the species to species competition.

The actual deployment of the complete algorithmic loop entailed a similar setup with the arbitrary equations being replaced by an evolvable CGP. Assisted by control-focused (population size, simulation time) and feature-focused fitnesses (species distance, genetic diversity), this implementation of the framework explored 1K years of autonomous evolution. Amongst the 10 repetitions, some runs found harsh conditions, alternated with more hospitable ones, which resulted in an overall stronger strategy towards resources and temperature management. The less robust populations were obtained in runs where the constraints were either too lenient, thus generating too small a push towards polyvalence, or too hostile, with too much effort spent in surviving transient conditions.

Robustness was determined by pairwise population competition of two types: one in which both are subjected to a single environmental controller and another where both populations retained their local conditions and had to invade the "foreign" portion of the environment. The evolved group performed differently depending on the type of evaluation with the first one favoring fast reproducers, more commonly found in the control group, while the second, having similar conditions to the evolutionary protocol itself, was better suited to dynamically stressed populations.

With this instantiation of the EDEnS framework we found that catastrophic transitions, in which large chunks of the population are destroyed in response to drastic environmental changes, were found to have a positive impact on the robustness of the resulting populations. This effect was strongly observed in the case of variations of temperature and more moderately so for topological dynamics. As this was the first use of this framework, much work remain notably with respect to the fitness functions used to guide the evolutionary process one of which (species distance) having negatively impacted the exploration because of an exploit of the compatibility function. Additionally, as will be elaborated further on, speciation was not observed in this experiment where

only one dominating species could be found at any time with a rather small number of subspecies being simultaneously explored.

8.2 Future work

With this work, we have explored two key aspects of virtual plants' evolution: the role of environmental constraints and the resulting speciation patterns. A number of specific points could be further improved in the current implementation such as moving towards a closed system, as opposed to the infinite reserves of this version. More external perceptions and effectors could also be devised to investigate reaction to seasonal patterns (though plants in the last experiment already found alternatives to cope with such dynamics) or a broader range of nutrient sources (most notably minerals).

More salient aspects, however, should require our attention in the near future either to reach the expected level of functionality or to pursue avenues of research ensuing from this prototype.

8.2.1 Bloating

The first point we address falls into the former category: while the unrestrained interaction between the compatibility function's genetic parameter and APOGeT allows for the autonomous emergence, and monitoring, of species barrier a negative feedback can appear.

One of the strongest instance of such a self-promoting degradation was observed in the previous experiment for one run in particular. The fitness which stimulated inter-species genetic distance was exploited by generating single-individual species through an exponential increase in the optimal distance between genomes. This resulted not only in an unusable phylogenetic tree but also required a large amount of computational power due to the number of species to test at each insertion. Such a behavior is, unfortunately, to be expected whenever the *monitoring* component of APOGeT is involved, even in such an indirect form, into the evolutionary process itself. Whether or not this precludes use of this tool in a directing role warrants further investigation as, if it could be made to work under such conditions, it would provide similar functionality as the niche maintaining algorithms, albeit with more flexibility.

Additionally, as in the case of hybridism, evolution rarely is a neatly segmented process. Instances can be found in which vertical inheritance is no longer the sole mechanism for genetic material propagation such as Horizontal Gene Transfer. As of the current version, this cross-species method of sharing genetic discoveries is not managed by the system but could be, theoretically, handled in the same manner as hybridism: a species-specific storage of incoming material could then be used to derive the weighted graph of contribution. With such an addition, species monitoring could be performed even in systems where the genome can no longer be viewed as a hermetic, atomic entity.

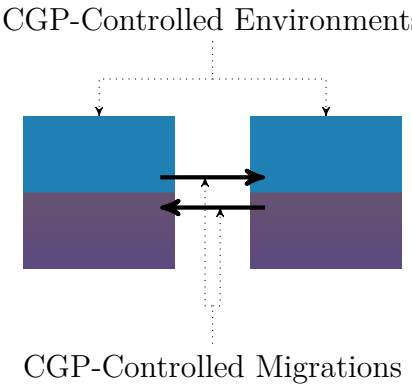


Figure 8.1: Enforcing allopatric speciation via CGP-regulated migrations

8.2.2 Lack of speciation

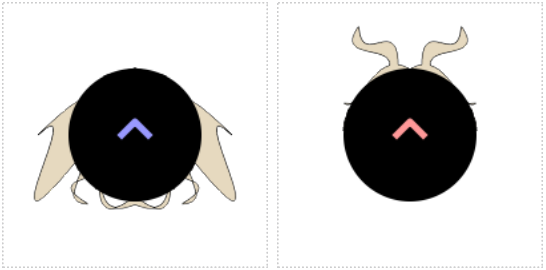
Though EDEnS was shown to perform as advertised, by generating more robust populations than those evolved in static, hospitable conditions, one salient point was found lacking. As mentioned above the number of concomitant species was overall low and generally resulted from recent cladogenesis events. This can be relatively easily explained away by the combination of three factors: efficiency-dependent stress variations, low environment size and lack of foresight.

The first is linked with the whole feedback loop of the framework, in this case enacted by the diversity fitness, which alternates between periods of intense stress and relaxation. As gathered in the previous chapter, whenever a population is in too brittle a state so as not to survive harsher conditions, the only viable alternatives available for selection as those in which the constraints are at most as difficult. These periods of calm provide the opportunity for a more diverse exploration of the genetic space around the currently used regions but, as soon as they prove resilient enough, the selection algorithm will tend to go with radically different conditions thus effectively trimming off all but the most successful species.

Coupled with the fact that, due to constraints on the computational cost, we kept the environment size sufficiently small so as to prevent simulation time explosion when faced with minimalist morphologies[1], it is difficult for the CGP to elaborate a physical separation that would promote allopatric speciation. This is further worsen by the fact that the environmental controllers selection does not operate on the same time-scale as the speciation process, thus preventing the planning (or discovery) of an appropriate solution for the emergence of cladogenesis. Indeed, even where such a solution to found it would most likely be replaced by another set of constraints in too short a time for it to have an effect.

A somewhat ad-hoc solution to bypass this problem would be to use different

[1]In the current setting 2500 plants is an average figure while (undesirably) optimized morphologies can reach above 10K



(a) Defensive (b) Aggressive

Figure 8.2: Lightweight morphological model for behavioral divergence

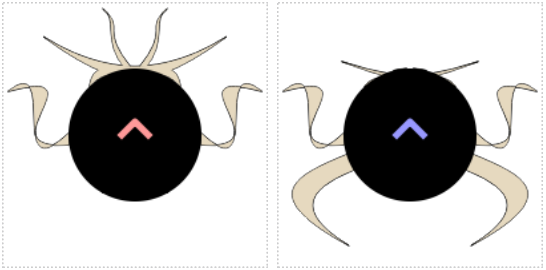
"islands" in which the evolutionary trends could freely diverge thereby imposing an impassible geological separation. This, however, implies some mechanism by which such independent evolutions could "rediscover" one another. Naturally, hand-crafted equations defining the migrations rates of a given population into another region could be devised in the same manner as equations for the control of environmental conditions have been used earlier on in this work. But this would limit the explored space of possibilities to those solutions we deem acceptable, in contradiction with the autonomous approach advocated in this framework. The resulting compromise would thus be as illustrated in figure 8.1 with multiple simulations in which a population is subjected to CGP-controlled constraints while the migration of individuals from one island to another would be under the supervision of a second CGP.

While such a methodology would further expand the complexity (in the negative sense of the term) of the framework by allowing for more undesirable combinations of conditions, it would also alleviate the need for extremely large experiments in which speciation can spontaneously emerge. Furthermore, it would also be of interest to study the impact of the connectivity of the islands on the development of the underlying communities: would fully connected networks be more favorable to uniformity? to competition? Would it be better to use von Neumann or Moore neighboorhoods?

8.2.3 Animal population

Another natural direction in which to orient future research is the use of motile heterotrophic creatures instead of the current autotrophic structures. As seen in chapter 2, there is an extensive amount of literature on such a subject, both in case of isolated evolution and of ecosystems. The challenge of this approach, however, would lie in applying the EDEnS framework to a different set of constraints.

Indeed the current environmental constraints would not make any sense for animals and would have to be replaced by more fitting ones such as food sources, ponds and obstacles distributions. Similarly, in order to obtain diverging demeanor, the creatures would need behavioral and morphological controllers



(a) Female (b) Male

Figure 8.3: Gene-controlled dimorphism for sexual selection monitoring

of sufficient complexity as to allow for a large viable genetic space but at the same time would require sufficient simplicity to cope with the population size and simulation duration necessary for the study of evolutionary trends.

Preliminary work has been engaged in this direction (see figure 8.2), in which a small number of cubic Bézier curves are used to generate "artifacts" on the surface of an otherwise smooth circle. In the figure, two such creatures are shown with light artifacts surrounding a darker center of mass. By exploiting a similar method to that of [Miconi 2008b], in which creatures fought against one another through the application of non-uniform collision damage, different types of strategies could emerge on the defensive/aggressive spectrum. For example, because of the lateral position of its artifacts which protect the softer central region, the creature in figure 8.2a could be describe as a herbivore. Oppositely, the creature in figure 8.2b has a much more offensive strategy in which its forward-facing "horns" could be used to inflict damage on an insufficiently protected prey.

With such a straightforward model, and the action of isolated speciation mentioned earlier, populations could diverge into different niches which would allow the study of the emergence of predation. Another aspect worthy of attention, which could easily be obtained from the same model, is that of the evolution of sexual dimorphism (figure 8.3).

Through gender-dependent differences of energy investment into their offspring, functional dimorphism could emerge to favor those individuals which, when of the correct gender, exhibit appropriate structures for the defense of these offspring. As illustrated, the rear portion of 8.3b could serve as some form of "shield" to prevent access to would-be predators. If associated with the corresponding behavioral patterns, members of the other sex would be free to specialize differently by saving on the cost of generating such costly structures.

But dimorphism could also be the result of a runaway sexual selection in which members of a given sex are favored for characteristics which not only do not participate in the survival of the individual but can even be detrimental. While such a phenomenon has been extensively observed and documented in nature, it only has limited counterparts in the virtual creatures domain and often comes with strong constraints e.g. [Ventrella 2005; Woehrer et al. 2012].

Although still in its infancy (see Godin-Dubois 2020 for a broad outline) this project could be one of the first to merge morphology and behavior into simulations of evolutionary scales from which we could study not only the individual dynamics and strategies but also that of their species.

Appendices

Appendix A

3rd party libraries

ID	Name	Utility
QT	Qt `www.qt.io`	Cross-platform GUI toolkit
BP	Bullet Physics SDK `github.com/bulletphysics/bullet3`	3D physics engine
GA	GAGA `github.com/jdisset/gaga`	Multi-objective genetic algorithms
CO	cxxopts `github.com/jarro2783/cxxopts`	Command line option parser
JS	json `github.com/nlohmann/json`	JSON for Modern C++
EX	exptrk `github.com/ArashPartow/exprtk`	Mathematical parsing and evaluation
CG	CGP-Library `www.cgplibrary.co.uk`	Cartesian Genetic Programming library

Table A.1: List of `C++` 3rd party libraries used throughout the thesis

Appendix B

Parasitism in Tierra

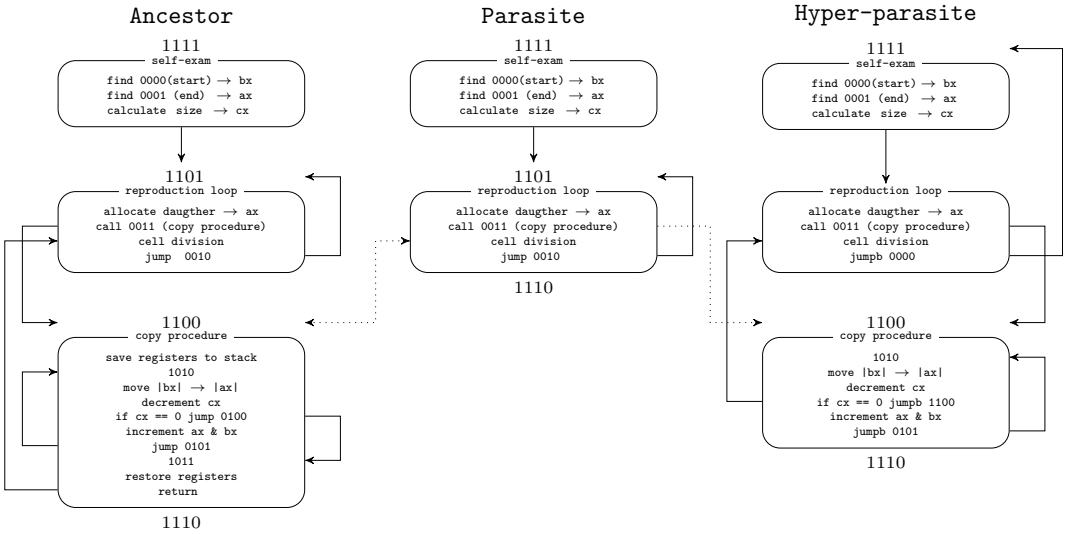


Figure B.1: Incremental levels of parasitism in Tierra (reproduced from [Ray 1991]). Dotted lines indicate parasitic relationships between individuals (see text for details).

The ancestor program can be cut into three components: an initial self-examination to determine its size, a reproduction loop which repetitively generates a daughter cell and a copy procedure which performs the low-level duplication of each of the genomes' instructions. This methodology is sufficient to indefinitely self-reproduce, however, its is not resilient to cheaters. The parasite has the same structure as the ancestor with one crucial difference: the end-of-genome marker is right after the reproduction loop. It follows that such a parasite does not have the required code to perform the copying process but, given the semi-permeability of individuals' membrane, this is not a problem: it just uses the code of a neighboring individual in which such a code fragment exists and is tagged by the same pattern.

Given its smaller size, it is much more efficient at reproducing than the

ancestor type but cannot dominate the population because of its reliance on other species' code. Additionally, it, itself, is not immune to more complex forms of parasitism as shown by the hyper-parasite. When a regular parasite P attempts to hijack the code of a hyper-parasite H all starts as in the previous case: each of its instructions is duplicated until reaching the end of the copy procedure. In this case, there is no `return` instruction which would send the ip back to the parasite's own memory region. Instead, the instruction pointer jumps directly to H's segment of cell division and further jumps back all the way to the start of H's program. There the self-examination process is performed, causing all subsequent copying to act upon the code of H and not P. Effectively, while P makes use of another code section (e.g. A), H actually steals CPU cycle from P by trapping it in a infinite loop where it works for the benefit of another.

Appendix C

Species Test-bed configuration files

Listing C.1: Configuration file the isolated evolutions' Config

```
              depHash_bullet: "a230c941f9"
             depHash_cxxopts: "70b9230639"
                depHash_gaga: "894e2393ee"
                depHash_json: "8424d10e45"
ecosystemDefaultSaveFolder: "evos/Ecosystem"
         gccCompilationDate: "vendredi 25 mai 2018, 08:52:58 (UTC+0200)"
       gitCurrentCommitDate: "Wed, 23 May 2018 10:55:16 +0200 (2 days ago)"
       gitCurrentCommitHash: "alife-2018-stage1-15-g6ab84a8-dirty"

       allowEmptySimulation: false
             immortalOrgans: false
                   logStats: true

           geneticCostLinks: 1.26491
        geneticCostMinLinks: 5
        geneticCostMinNodes: 6
           geneticCostNodes: 1.58114
                 globalDice: 1527231623725
       initialWaterReserves: 0.5
                  maxOrgans: 128
    minNoveltyForEcosystem: 0.01
nutrientsDiffusionBaseRate: 1
  photosynthesisSaturation: 0.025
       photosynthesisSpeed: 0.025
   photosynthesisWaterCost: 1
     reproductionAreaRatio: 100
              restrictWater: false
             simulatedYears: 2
       tropismMaximalAngle: 0.392699
       waterAbsorptionRate: 1

                mediumsMap: map(Medium, set<Skill>)
   Ground: Root trunk Root hair Reserve
      Air: Photosynthesis Structural Reserve Sexual

             producersMap: map(Element, Skill)
     Water: Root hair
   Glucose: Photosynthesis

skillConsumptionModifiers: map(Skill, float)
  Photosynthesis: 1.5
      Structural: 0.5
      Root trunk: 0.75
       Root hair: 1.5
         Reserve: 0.625
          Sexual: 1.25

       starvationDuration: map(Element, float)
     Water: 6
   Glucose: 4

          transportSpeed: map(Skill, array<float, 2>)
   Photosynthesis: [ 0.5 0.5 ]
      Structural: [ 1 1 ]
      Root trunk: [ 1 1 ]
       Root hair: [ 0.5 0.5 ]
         Reserve: [ 0.75 0.75 ]
          Sexual: [ 0.5 0.5 ]
```

Listing C.2: Configuration file the isolated evolutions' MutationBoundsConfig

```
     maxGrowthOnFirstDay: 10
            maxSeedScale: 100
mutationQuaternionRange: 1.0472
  randomQuaternionRange: 1.5708

             environment: map(MEnvironment, FieldBounds)
   Rng seed: [ 0, 4294967295 ]
       Size: [ 10, 10 ]
      Depth: [ 5, 5 ]

           graphtalData: map(MGraphtalData, FieldBounds)
    Growth speed: [ 7, 600 ]
     Gmax growth: [ 2, 2000 ]
    Width change: [ 0, (0) 1 ]
     Gdist optim: [ 0 (0.5), (10) 100 ]
   Gdist stddevs: [ [ 0.001 0.001 ] ([ 0.5 0.5 ]), ([ 0.5 0.5 ]) [ 100 100 ] ]
   Min fmaturity: [ 0 (0.25), (0.75) 1 ]

               linkData: map(MLinkData, FieldBounds)
        Min scale: [ 1 (2), 10 ]
      Recursivity: [ 1, 5 ]
   Relative scale: [ 1.17549e-38, 1 ]
         Tropisms: [ [ -1 -1 -1 ] ([ 0 0 0 ]), ([ 0 0 0 ]) [ 1 1 1 ] ]

               nodeData: map(MNodeData, FieldBounds)
   Allocation: [ [ 0 0 ], [ 1 1 ] ]
     Survival: [ [ 0 0 ], [ 1 1 ] ]
   Dimensions: [ {0.001,0.001,0.001} ({0.001,0.001,0.001}), {0.01,0.01,0.01} ]
      Density: [ 0.001, (1) 10 ]
        Color: [ {0,0,0} ({0,0,0}), {1,1,1} ]
   Nmax scale: [ 1, (10) 10000 ]

           radialEffect: map(MRadialEffect, FieldBounds)
   Re count: [ 1, 5 ]

           randomEffect: map(MRandomEffect, FieldBounds)
   Se count: [ 1, 5 ]
```

Listing C.3: Configuration file the isolated evolutions' MutationRatesConfig

```
       changeEffectProba: 1.0e-01
      randomAddLinkProba: 7.5e-01
  randomLinkAddNodeProba: 5.0e-01
terminalRecursivityProba: 2.5e-01

               ecosystem: map(MEcosystem, float)
     Templates: 1
   Environment: 0

             environment: map(MEnvironment, float)
   Rng seed: 0
        Sun: 0
       Size: 0
      Depth: 0
     Drawer: 0

              graphtal: map(MGraphtal, float)
       Add link: 0.05
       Del link: 0.05
   Mutate gdata: 0.1
    Mutate node: 1
    Mutate link: 1

           graphtalData: map(MGraphtalData, float)
    Growth speed: 1
     Gmax growth: 1
    Width change: 1
             Sex: 0.1
     Gdist optim: 0.5
   Gdist stddevs: 0.5
   Min fmaturity: 0.1

               linkData: map(MLinkData, float)
        Min scale: 1
      Recursivity: 1
           Effect: 1
     Relative pos: 1
     Relative ort: 1
   Relative scale: 1
         Tropisms: 1

               nodeData: map(MNodeData, float)
        Type: 0
       Skill: 0.1
   Allocation: 1
     Survival: 1
```

```
Dimensions: 1
   Density: 0.1
     Color: 0.1
Nmax scale: 1

         radialEffect: map(MRadialEffect, float)
Re normal: 1
 Re count: 1

         randomEffect: map(MRandomEffect, float)
Se count: 1
 Se dice: 1
```

Appendix D

Details of adaptive mechanisms effect on fitness

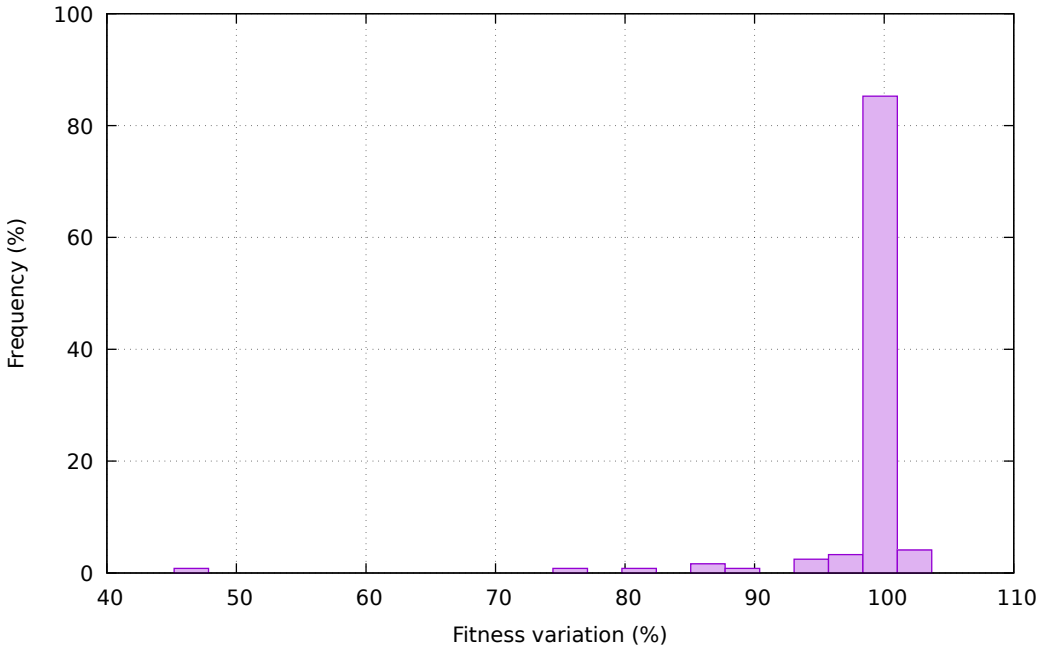


Figure D.1: backpropagations

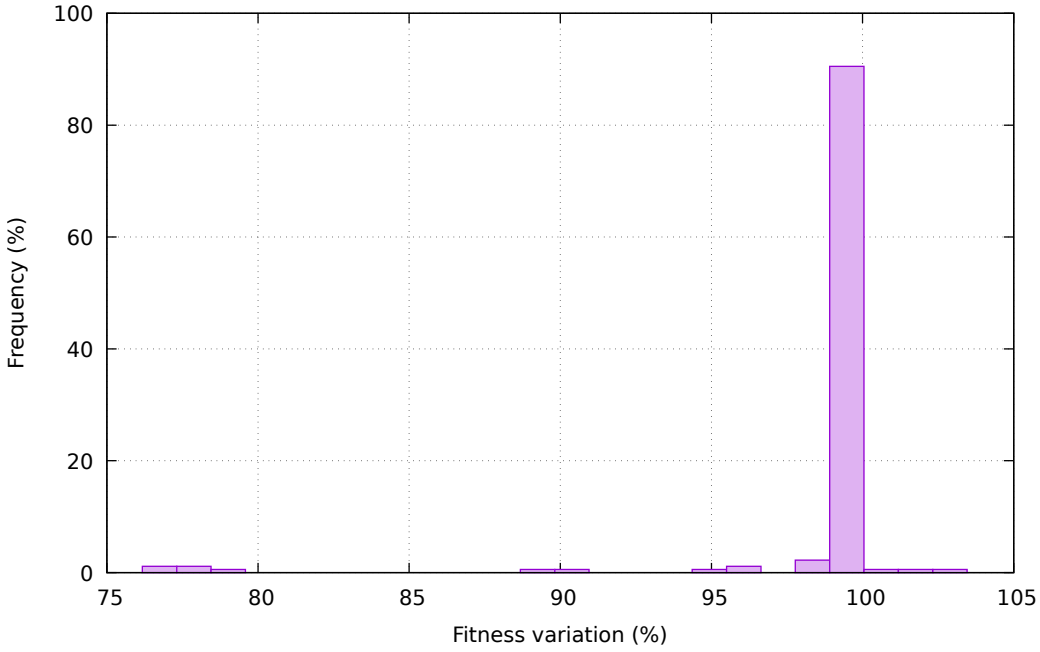


Figure D.2: densitropisms

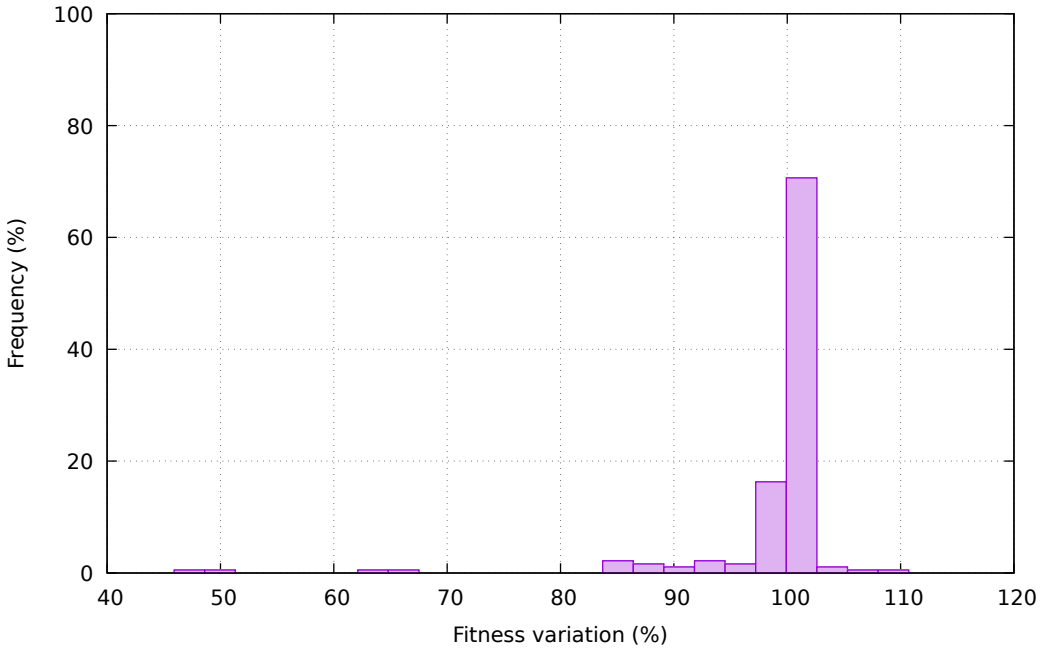


Figure D.3: gravitropisms

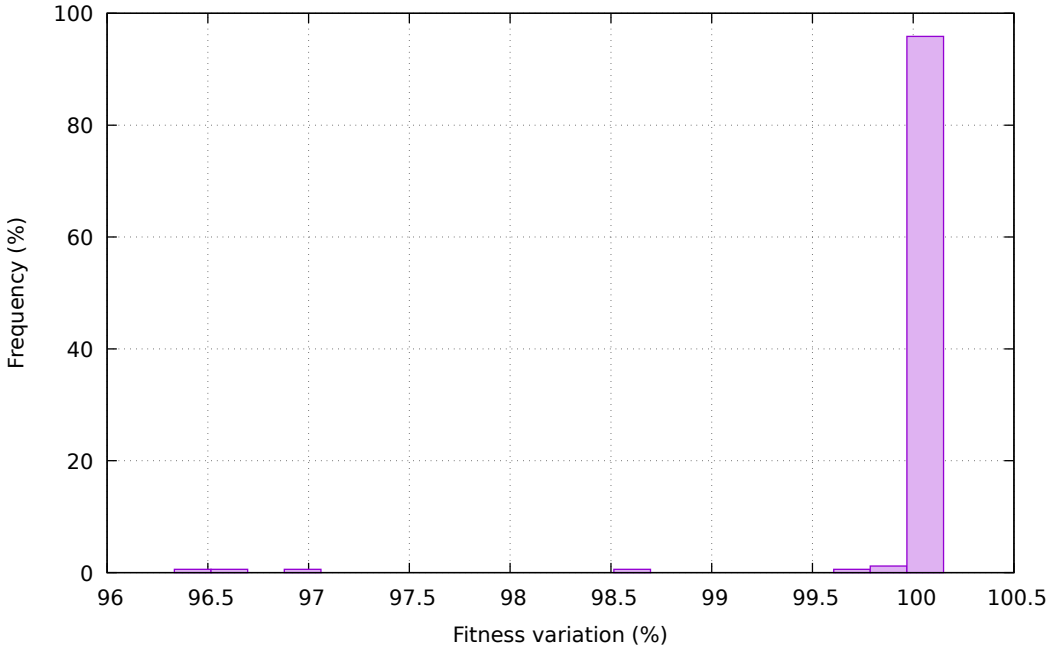


Figure D.4: phototropisms

Appendix E

Species Test-bed configuration files

Listing E.1: Configuration file the speciation test-bed's Simulation

```
assimilationRate: 0.01
baselineLight: 1
baselineShallowWater: 0.3
daysPerYear: 100
floweringCost: 3
heightPenaltyStddev: 0.5
initSeeds: 100
killSeedsEarly: true
lifeCost: 0.05
logGlobalStats: true
logIndividualStats: false
maxPlantDensity: 10
resourceCost: 0.5
saturationRate: 1
saveEvery: 1
stepsPerDay: 10
stopAtMaxGen: 4294967295
stopAtMinGen: 4294967295
stopAtYear: 10
taurusWorld: false
temperatureMaxRange: 10
temperatureRangePenalty: 10
trimmingProba: 0
updateTopographyEvery: 1
```

Listing E.2: Configuration file the speciation test-bed's PTree

```
avgCompatibilityThreshold: 0.25
    compatibilityThreshold: 0.1
             enveloppeSize: 15
          minNodeEnveloppe: 0
           minNodeSurvival: 0
             showNodeNames: true
       similarityThreshold: 0.5
          simpleNewSpecies: true
    speciesDetailVerbosity: 1
    stillbornTrimmingDelay: 2
stillbornTrimmingMinDelay: 0
   stillbornTrimmingPeriod: 1000
stillbornTrimmingThreshold: 1
         survivorNodesOnly: false
```

Listing E.3: Configuration file the speciation test-bed's Plant

```
          dethklokBounds: (10 10 10 1000)
    fruitOvershootBounds: (1 1.1 1.1 2)
     seedsPerFruitBounds: (1 3 3 100)
temperatureOptimalBounds: (-20 10 10 40)
  temperatureRangeBounds: (0.001 10 10 30)

                ls_axiom: S
       ls_maxNonTerminal: F
          ls_maxRuleSize: 4
    ls_recursivityBounds: (1 2 2 5)
        ls_rootInitRule: "S -> [+h][-h]"
       ls_rotationAngle: 0.523599
       ls_shootInitRule: "S -> s[+l][-l]f"

         distanceWeights: map(string, float) {
                 "cdata": 0
              "dethklok": 0.5
        "fruitOvershoot": 0.5
            "metabolism": 1
                  "root": 1.5
         "seedsPerFruit": 0.5
                 "shoot": 1.5
    "temperatureOptimal": 0.5
      "temperatureRange": 0.5
}
        ls_mutationRates: map(string, float) {
    "recursivity": 1
         "rules": 7
```

```
}
    ls_ruleMutationRates: map(string, float) {
    "brkSymb": 0.5
    "delSymb": 0.5
    "dupSymb": 1
    "mutSymb": 4
    "swpSymb": 3
}
 ls_ruleSetMutationRates: map(string, float) {
    "addRule": 2
    "delRule": 1
    "mutRule": 7
}
       ls_terminalsSizes: map(char, TerminalSize) {
    f: 0.01x0.05
    g: 0.02x0.02
    h: 0.01x0.1
    l: 0.01x0.1
    s: 0.01x0.1
    t: 0.01x0.1
}
           mutationRates: map(string, float) {
                 "cdata": 4
              "dethklok": 1
         "fruitOvershoot": 0.5
             "metabolism": 4
                  "root": 8
          "seedsPerFruit": 0.5
                 "shoot": 8
    "temperatureOptimal": 0.5
      "temperatureRange": 0.5
}
```

Listing E.4: Configuration file the speciation test-bed's BOCData

```
  optimalDistanceBounds: (0 0 4 100)
outbreedToleranceBounds: (0 2 2 10)
 inbreedToleranceBounds: (0 2 2 10)
              sexBounds: (F F M M)
            mutateChild: 0.5
        distanceWeights: map(string, float) {
     "inbreedTolerance": 1
      "optimalDistance": 1
    "outbreedTolerance": 1
                 "sex": 1
}
```

```
          mutationRates: map(string, float) {
     "inbreedTolerance": 1
      "optimalDistance": 1
    "outbreedTolerance": 1
                 "sex": 1
}
```

Listing E.5: Configuration file the speciation test-bed's Metabolism

```
conversionRatesBounds: ([ 0.001 0.001 ]
                        [ 0.001 0.001 ] [ 1 1 ] [ 1 1 ])
     deltaWidthBounds: (0 0.2 0.2 1)
      resistorsBounds: ([ 0 0 ] [ 1 1 ] [ 1 1 ] [ 10 10 ])
    growthSpeedBounds: (0 1 1 10)
      distanceWeights: map(string, float) {
    "conversionRates": 1
         "deltaWidth": 1
        "growthSpeed": 1
          "resistors": 1
}
        mutationRates: map(string, float) {
    "conversionRates": 2
         "deltaWidth": 1
        "growthSpeed": 1
          "resistors": 2
}
```

Listing E.6: Configuration file the speciation test-bed's Environment

```
  depthBounds: (50 50 50 50)
   maxTBounds: (40 40 40 40)
   minTBounds: (-20 -20 -20 -20)
rngSeedBounds: (0 0 4294967295 4294967295)
 voxelsBounds: (100 100 100 100)
  widthBounds: (10 1000 1000 10000)
mutationRates: map(string, float) {
      "depth": 0.1
    "envCtrl": 1
       "maxT": 0.1
       "minT": 0.1
    "rngSeed": 0.1
     "voxels": 0.1
      "width": 0.1
}
```

Listing E.7: Configuration file the speciation test-bed's EnvCTRL

```
        c0Bounds: (0 1 1 1)
        c1Bounds: (0 1 1 1)
distanceWeights: map(string, float) {
    "_expression": 0.1
                "c0": 0.1
                "c1": 0.1
}
  mutationRates: map(string, float) {
    "_expression": 0
                "c0": 0.1
                "c1": 0.1
}
```

Appendix F

Distribution of population-level variables

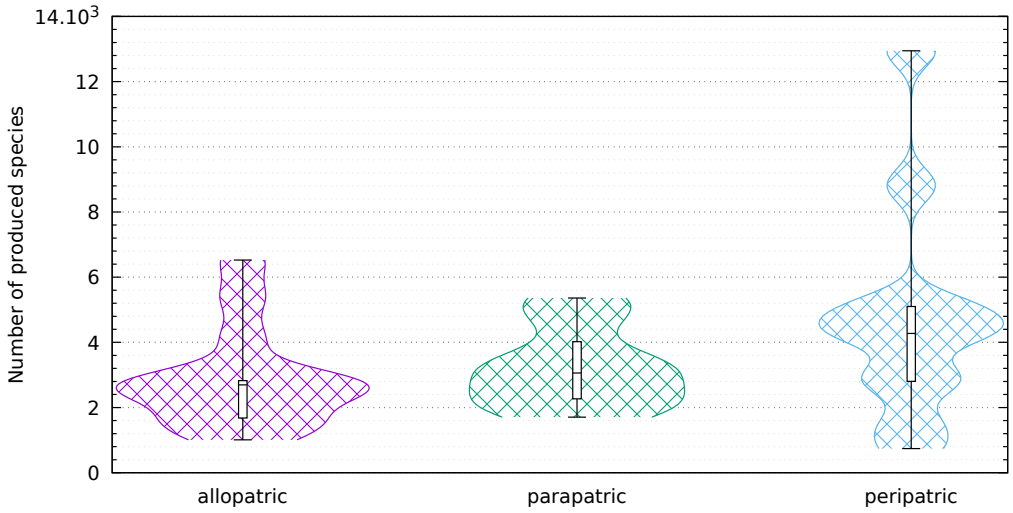


Figure F.1: Number of observed species per experiment

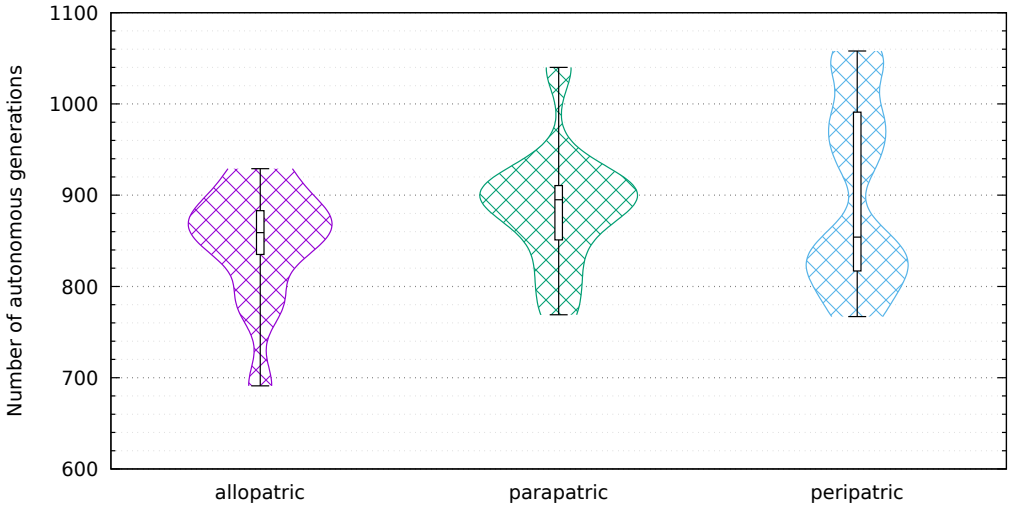


Figure F.2: Number of observed generations per experiment

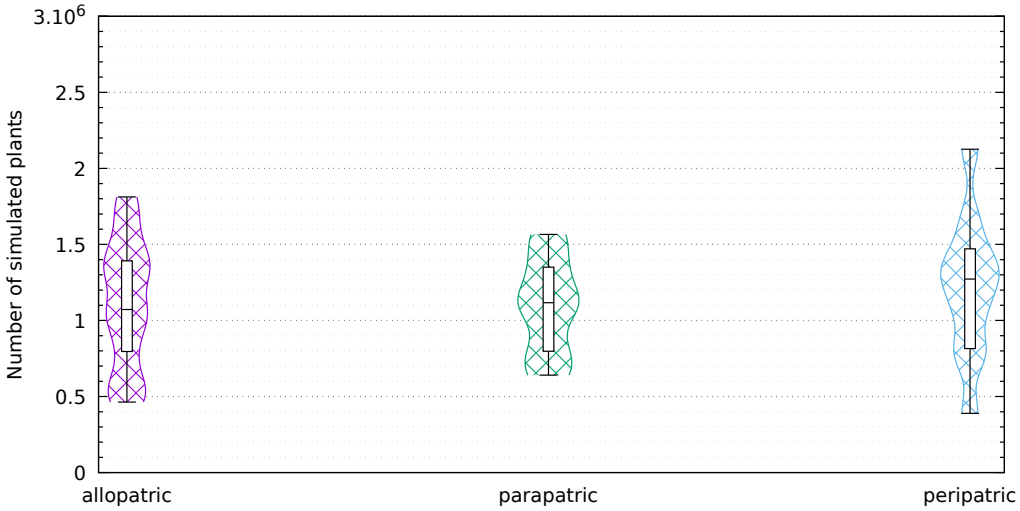


Figure F.3: Number of observed plants per experiment

Appendix G

Additional species dynamics

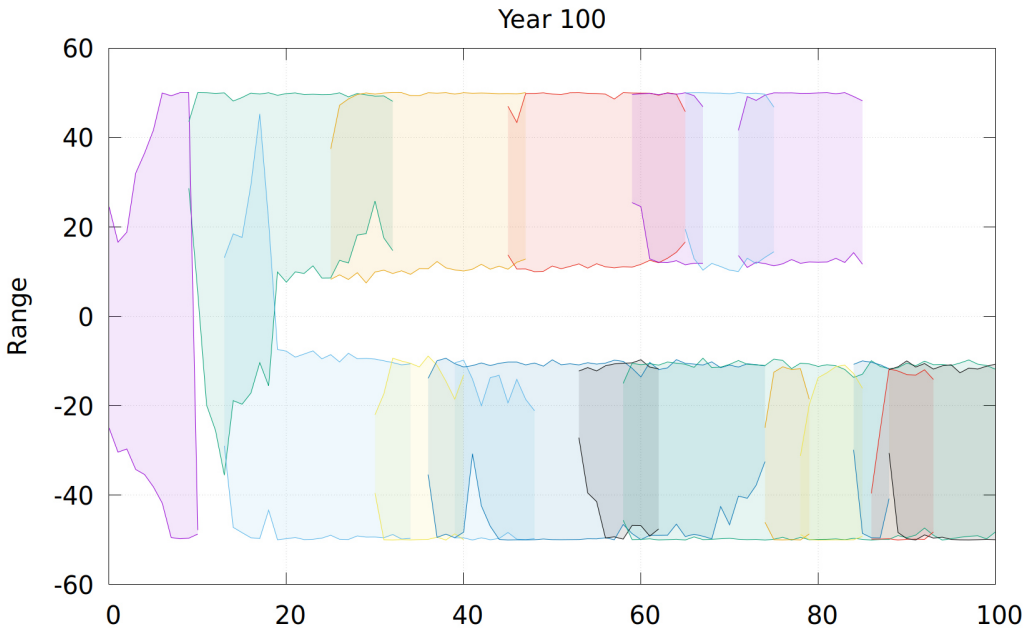


Figure G.1: Example of dynamics in the allopatric scenario

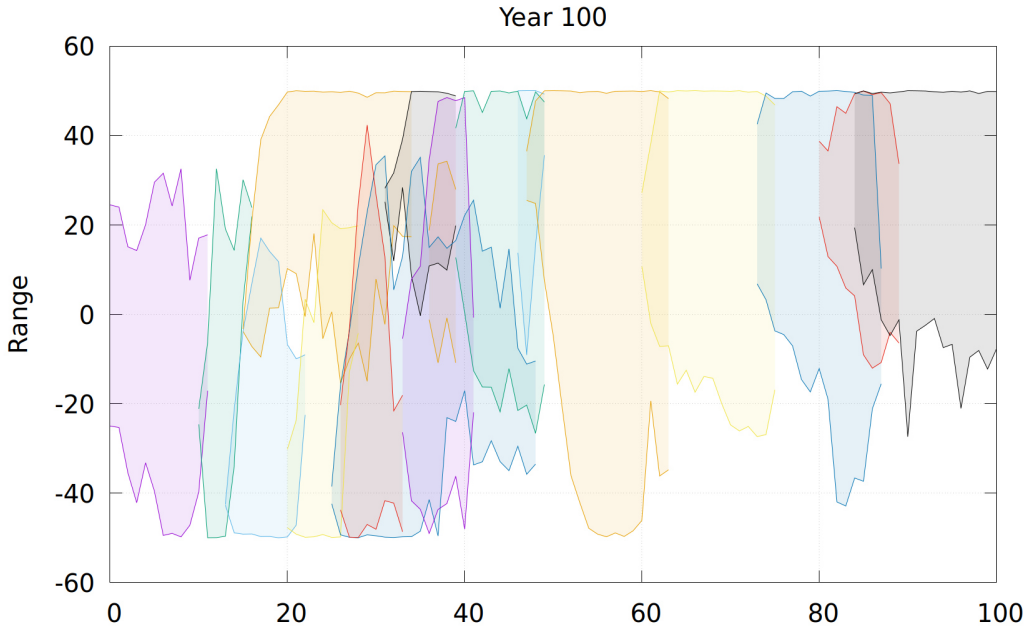


Figure G.2: Dynamics in the 'best' run of the parapatric scenario

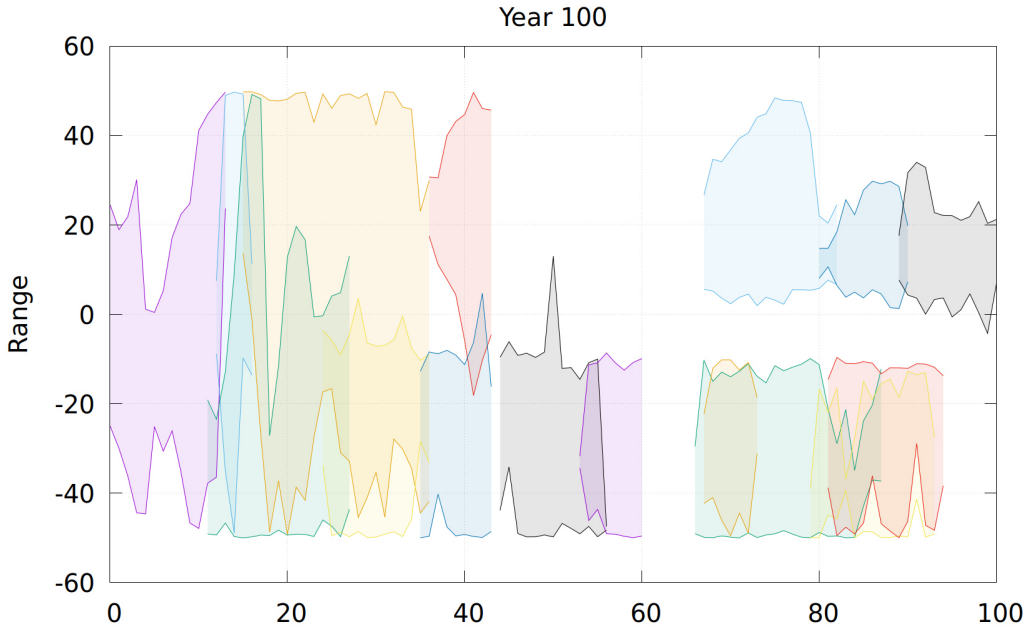


Figure G.3: Dynamics in the 'worse' run of the peripatric scenario

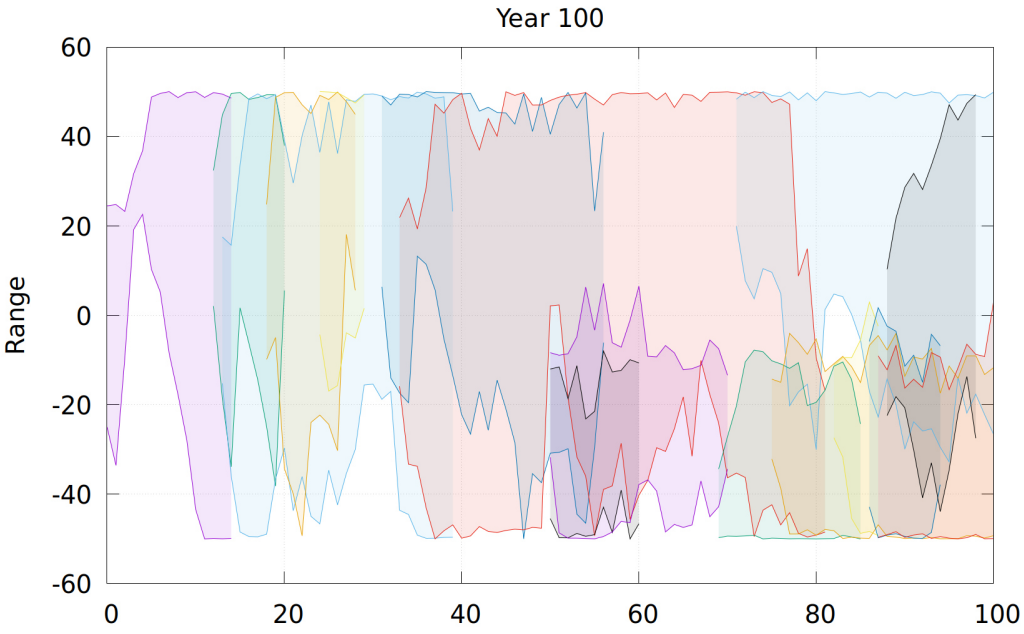


Figure G.4: Dynamics in the ‘best’ run of the peripatric scenario

Appendix H

Function set for the environmental controller

	Function	Range
mone =	-1	
zero =	0	
one =	1	
rand =	$\emptyset$	$[-1;1]$
id(X) =	X	$[-1;1]$
abs(X) =	$\lvert X\rvert$	$[0;1]$
sq(X) =	X^2	$[0;1]$
sqrt(X) =	$\sqrt{\lvert X\rvert}$	$[0;1]$
exp(X) =	$\frac{e^X-1}{e-1}$	$[-e^{-1};1]$
sin(X) =	$sin(X)$	$[-1;1]$
cos(X) =	$cos(X)$	$[-1;1]$
tanh(X) =	$tanh(X)$	$[-1;1]$
asin(X) =	$\frac{2sin^{-1}(X)}{\pi}$	$[-1;1]$
acos(X) =	$\frac{cos^{-1}(X)}{\pi}$	$[-1;1]$
step(X) =	$\begin{cases}-1 & X<0\\ 1 & X>0\\ 0 & otherwise\end{cases}$	$\{-1,0,1\}$
inv(X) =	$\begin{cases}0 & X=0\\ max(-1,min(\frac{1}{X},1)) & \end{cases}$	$[-1;1]$
rond(X) =	$\lfloor X\rceil$	$\{-1,0,1\}$
flor(X) =	$\lfloor X\rfloor$	$\{-1,0,1\}$
ceil(X) =	$\lceil X\rceil$	$\{-1,0,1\}$
del(X,Y) =	$\frac{X-Y}{2}$	$[-1;1]$
add(X,Y) =	$\frac{X+Y}{2}$	$[-1;1]$
mult(X,Y) =	XY	$[-1;1]$
gt(X,Y) =	$(X>Y)$	$\{0,1\}$
lt(X,Y) =	$(X<Y)$	$\{0,1\}$
eq(X,Y) =	$(X==Y)$	$\{0,1\}$
max(X,Y) =	$max(X,Y)$	$[-1;1]$
min(X,Y) =	$min(X,Y)$	$[-1;1]$
hgss(X,Y) =	$\begin{cases}1 & Y=0\wedge X=0\\ 0 & Y=0\wedge X\neq 0\\ e^{-\frac{X^2}{2Y^2}} & \end{cases}$	$[-1;1]$

Table H.1: Function set

Appendix I

Details of pair-wise evalution scores

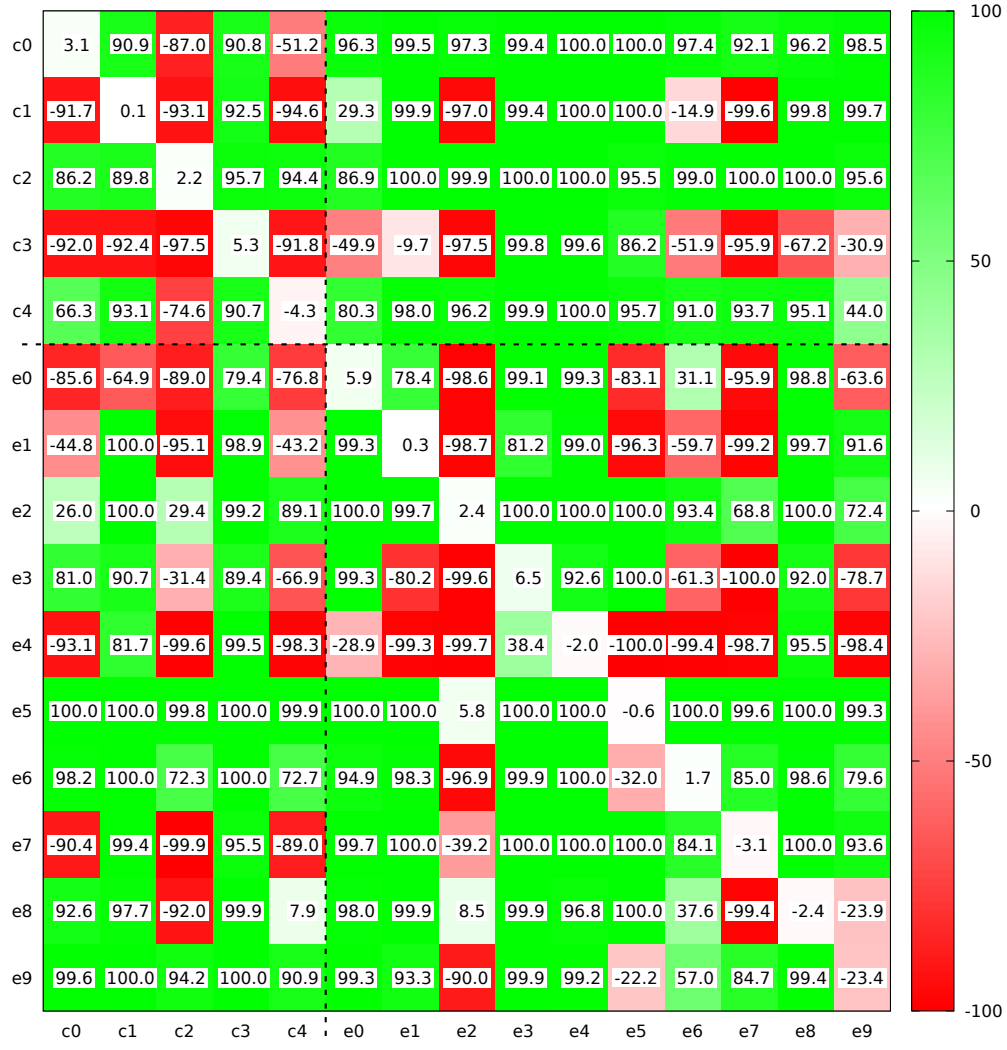


Figure I.1: Pair-wise scores in the Merged naturalisation scenario

Figure I.2: Ordered pair-wise scores in the Merged naturalisation scenario showing the three major groups: c (left), e^+ (center), e^- (right)

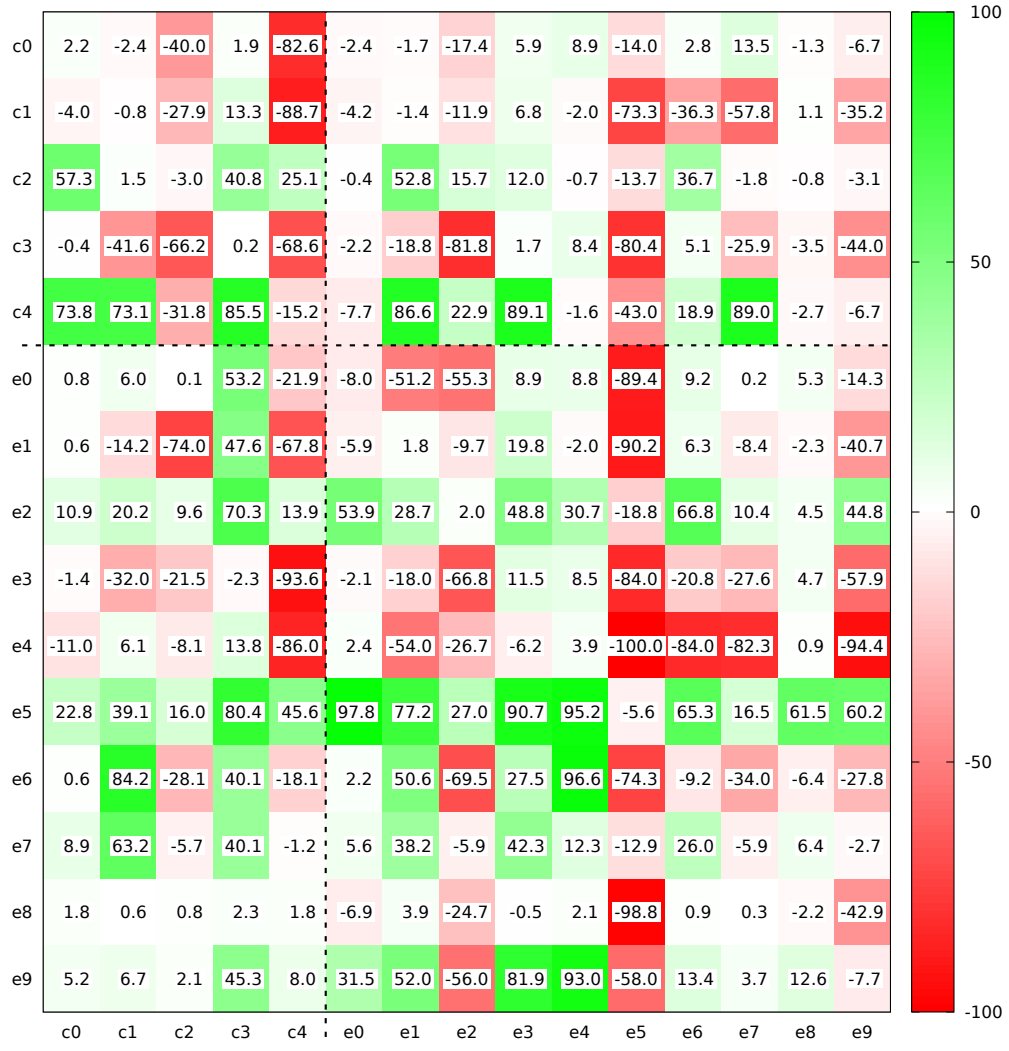


Figure I.3: Pair-wise scores in the Contiguous naturalisation scenario

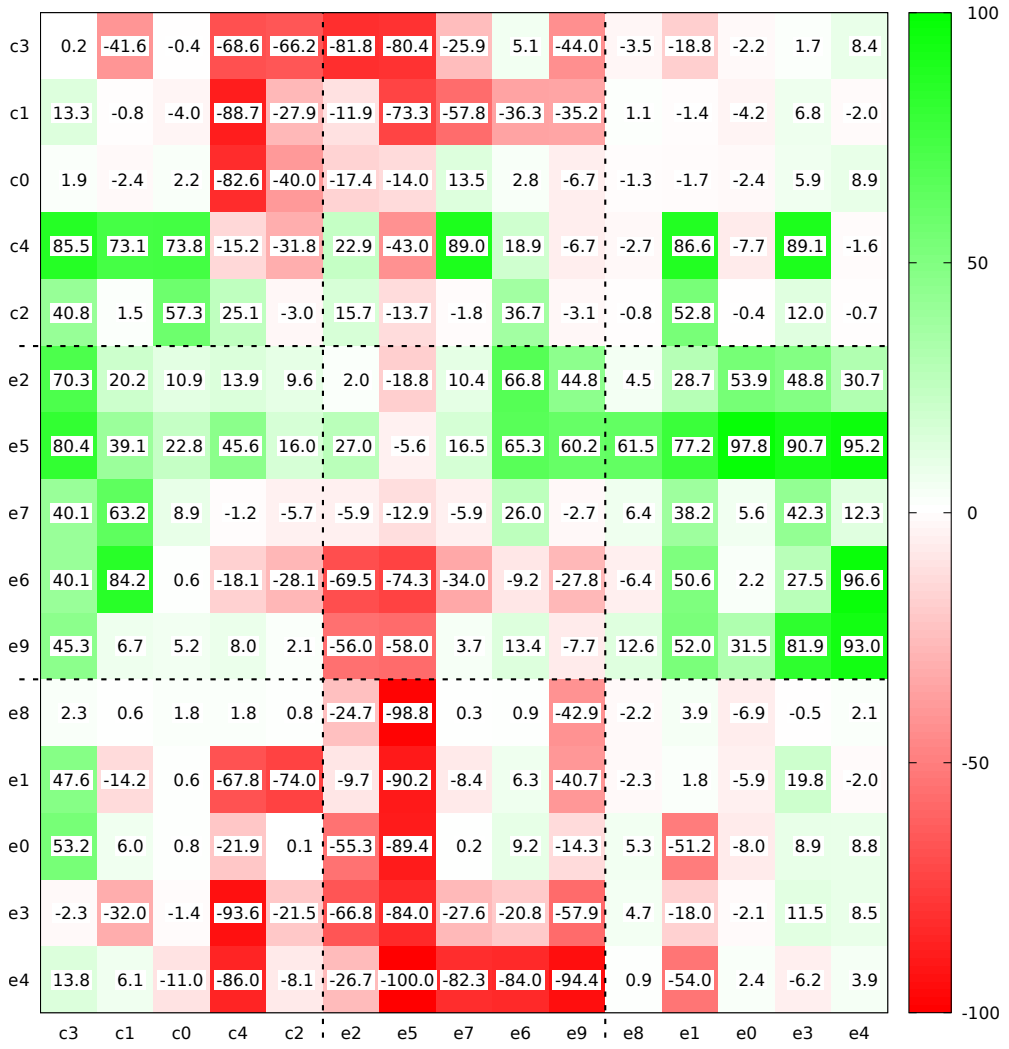
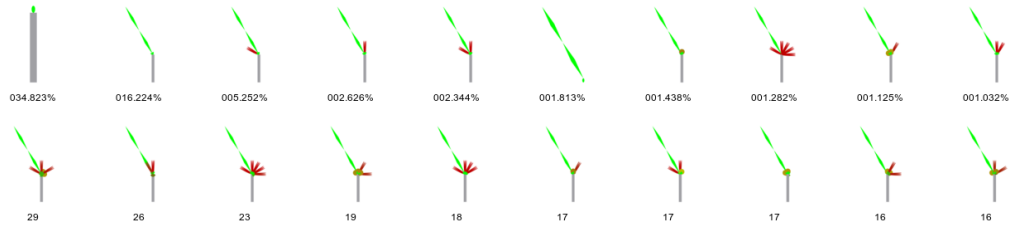


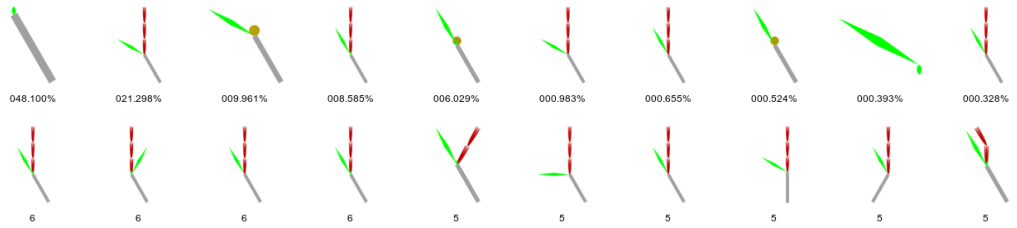
Figure I.4: Ordered pair-wise scores in the Contiguous naturalisation scenario showing the three major groups

Appendix J

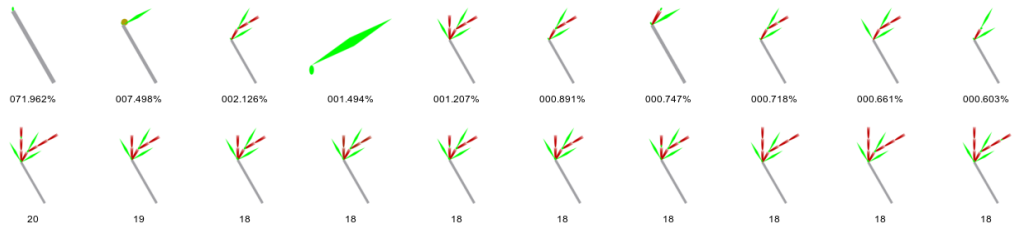
Intra-population morphological diversity



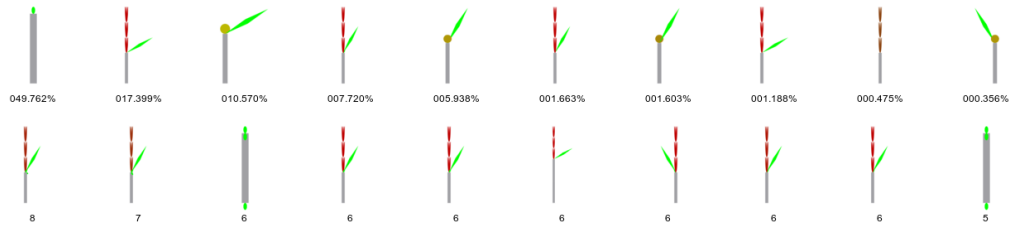
(c_0)



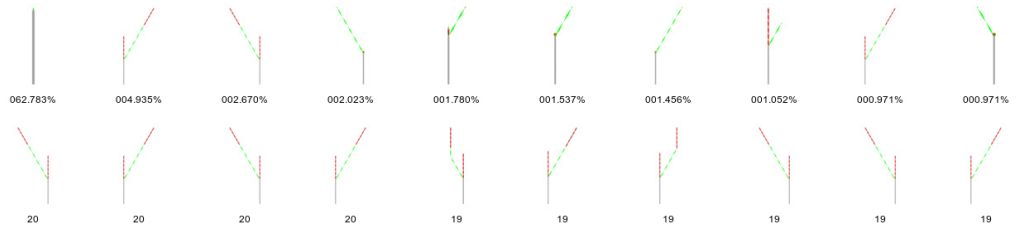
(c_1)



(c_2)

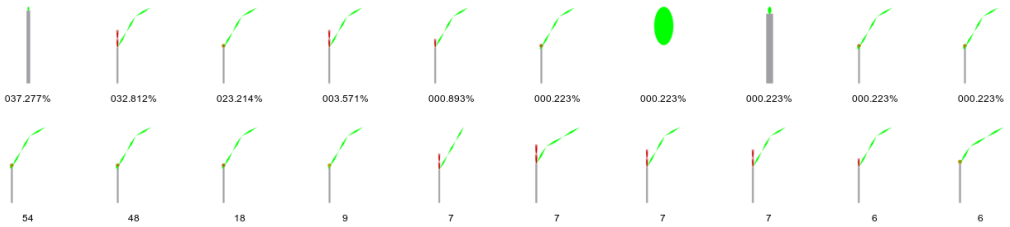


(c_3)

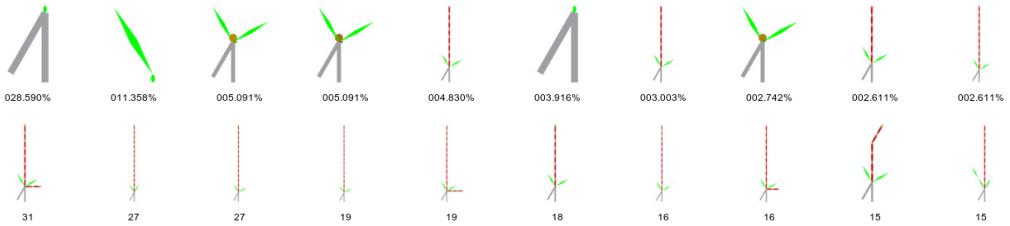


(c_4)

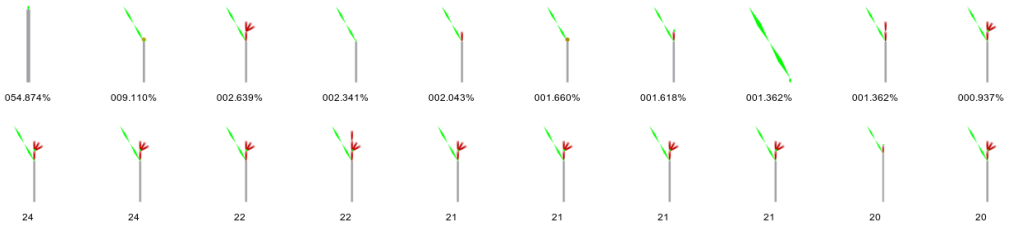
Figure J.1: Details of morphological variation. First row is sorted by decreasing order of frequency. Second row is by organ count.



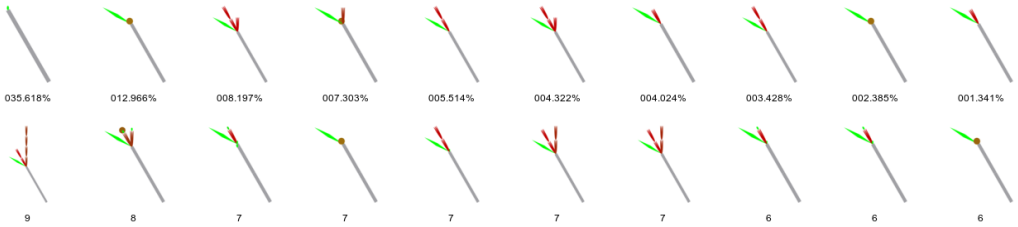
(e_0)



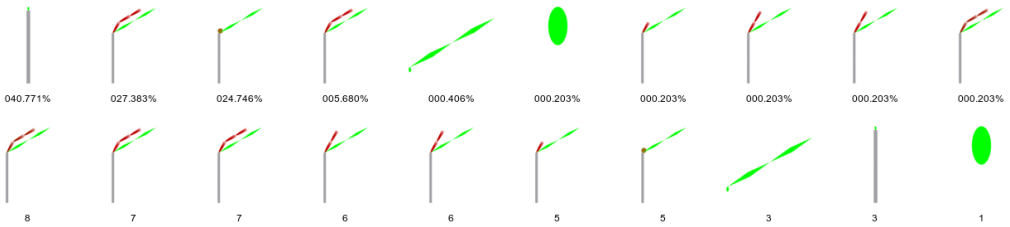
(e_1)



(e_2)



(e_3)



(e_4)

Figure J.1: Details of morphological variation (cont.)

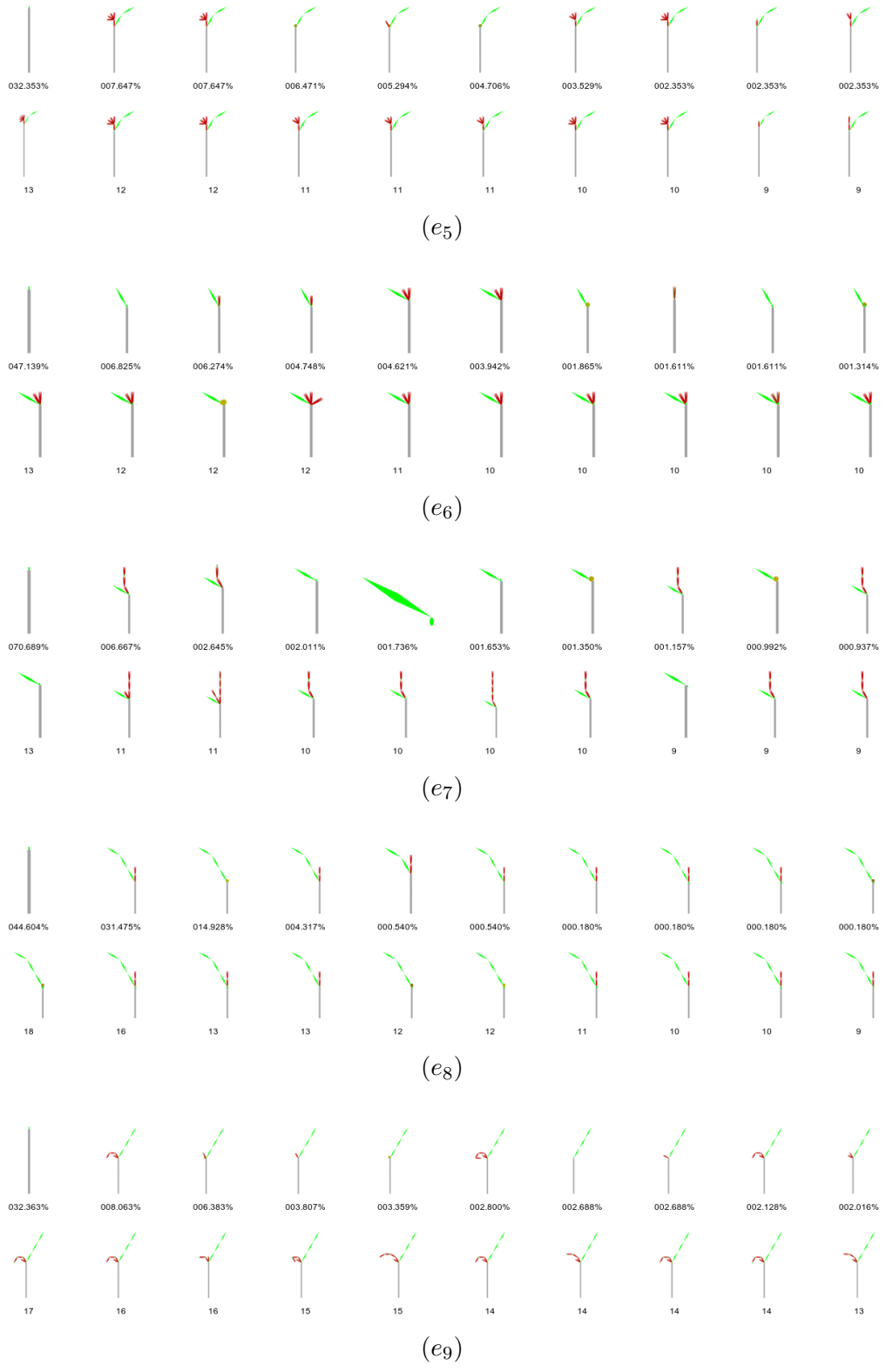


Figure J.1: Details of morphological variation (cont.)

Appendix K

Strategies (complete)

Run	*Rep*	Merg.
c_2	1	63.56
e_2	2	56.24
c_4	3	40.88
c_0	4	28.93
e_5	5	25.35
e_7	6	23.96
e_6	7	18.9
e_9	8	17.55
e_8	9	−19.48
e_1	10	−24.84
c_1	11	−25.21
e_0	12	−39.19
e_3	13	−39.67
c_3	14	−60.75
e_4	15	−66.23

Table K.1: Reproduction (aggregate score in Merged evaluation)

Run	*Def*	Score
c_2	1	44.84
e_5	2	43.46
e_2	3	39.34
c_4	4	35.5
c_0	5	34.11
e_9	6	32.73
e_6	7	32.41
e_7	8	21.69
e_8	9	20.71
c_1	10	7.66
e_3	11	4.45
e_1	12	4.44
e_0	13	−5.51
c_3	14	−16.19
e_4	15	−20.08

(a) Defender (rows)

Run	*Atk*	Score
c_2	1	18.71
e_2	2	16.9
c_4	3	5.38
e_7	4	2.27
c_0	5	−5.18
e_6	6	−13.5
e_9	7	−15.18
e_5	8	−18.1
e_1	9	−29.27
c_1	10	−32.87
e_0	11	−33.68
e_8	12	−40.19
e_3	13	−44.12
c_3	14	−44.56
e_4	15	−46.15

(b) Aggressor (columns)

Table K.2: Substrategies from Merged evaluation obtained by partial aggregation over the corresponding row or column from the matrices in annex I

Run	*Col*	Cont.
e_5	1	54.87
c_4	2	29.31
e_2	3	25.17
e_9	4	17.1
c_2	5	16.53
e_7	6	10.62
e_6	7	−2.23
e_8	8	−7.98
e_0	9	−10.03
c_0	10	−10.05
e_1	11	−16.19
c_1	12	−17.73
e_3	13	−28.12
e_4	14	−29.58
c_3	15	−31.68

Table K.3: Colonizer (aggregate score in Contiguous evaluation)

Run	On left side		On right side		Difference		Symmetry
	Rank	Score	Rank	Score	Raw	Norm.	Rank
e_1	11	−7.97	12	−8.22	−0.25	1.54	1
e_0	9	−4.91	9	−5.12	−0.21	2.09	2
c_4	2	14.34	2	14.97	0.63	2.15	3
e_5	1	26.32	1	28.55	2.23	4.06	4
e_3	13	−13.45	14	−14.67	−1.22	4.34	5
e_2	3	13.23	3	11.94	−1.29	−5.13	6
e_9	4	7.79	4	9.31	1.52	8.89	7
c_0	8	−4.45	10	−5.6	−1.15	11.44	8
c_2	5	7.28	5	9.26	1.98	11.98	9
c_3	14	−13.94	15	−17.75	−3.81	12.03	10
e_4	15	−17.52	13	−12.07	−5.45	18.42	11
c_1	12	−10.75	11	−6.99	3.76	−21.21	12
e_7	6	6.96	6	3.67	−3.29	−30.98	13
e_8	10	−5.39	7	−2.59	2.80	−35.09	14
e_6	7	1.15	8	−3.38	−4.53	203.14	15

Table K.4: Symmetry. Left and right side scores are obtained by aggregating rows and columns, respectively. Normalized difference (Norm.) is obtained by dividing Raw difference by CScore (above).

Run	Hyb_M	Score
c_4	1	83.88
e_6	2	73.06
c_0	3	71.02
c_3	4	69.32
e_9	5	69.29
e_0	6	67.25
e_2	7	64.96
e_8	8	62.58
e_7	9	61.3
e_1	10	60.76
c_1	11	57.42
c_2	12	54.12
e_3	13	48.79
e_4	14	44.31
e_5	15	32.34

(a) Merged

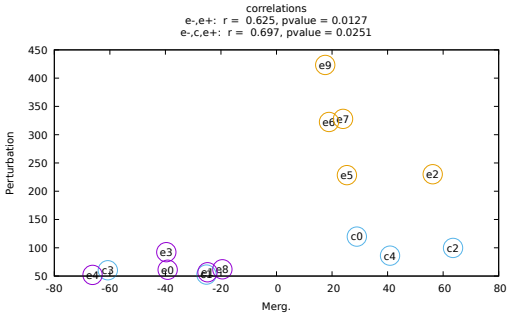
Run	Hyb_C	Score
e_9	1	81.95
e_6	2	80.00
c_4	3	78.72
e_5	4	68.16
c_3	5	66.71
e_1	6	65.24
c_2	7	59.31
e_2	8	59.13
c_1	9	52.18
e_7	10	50.45
c_0	11	48.81
e_3	12	38.31
e_0	13	33.89
e_4	14	27.2
e_8	15	7.8

(b) Contiguous

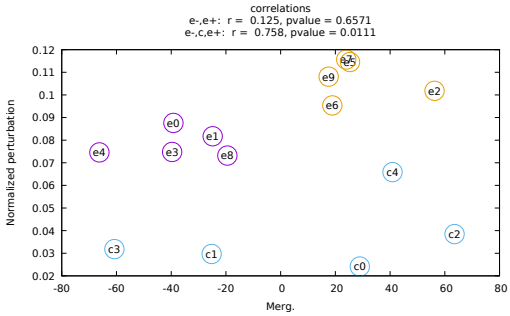
Table K.5: Hybridation. Average proportion of the post-evaluation populations with genetic material from both source populations.

Appendix L

Perturbation versus Score scatter plots

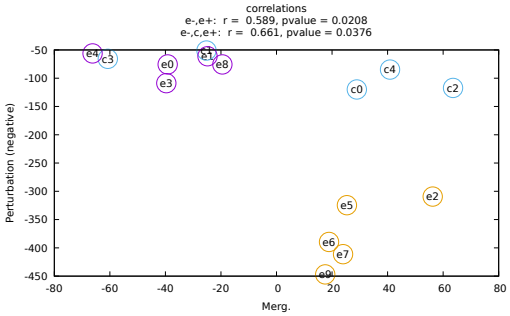


(a) P_r

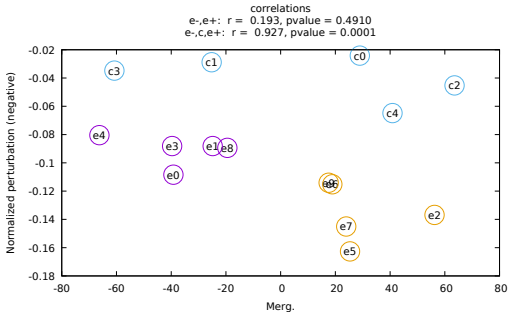


(b) $\bar{P}_r$

Figure L.1: Scores in Merged evaluation against perturbation

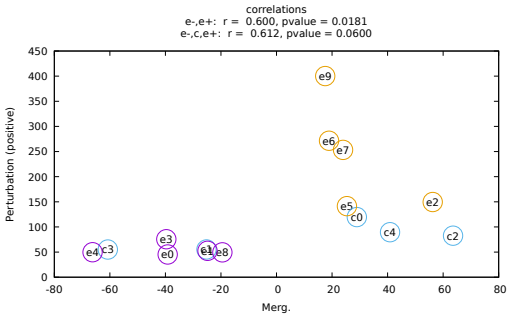


(a) P_r^-

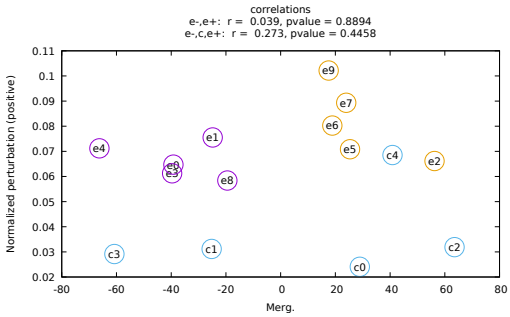


(b) $\bar{P}_r^-$

Figure L.2: Scores in Merged evaluation against negative perturbation

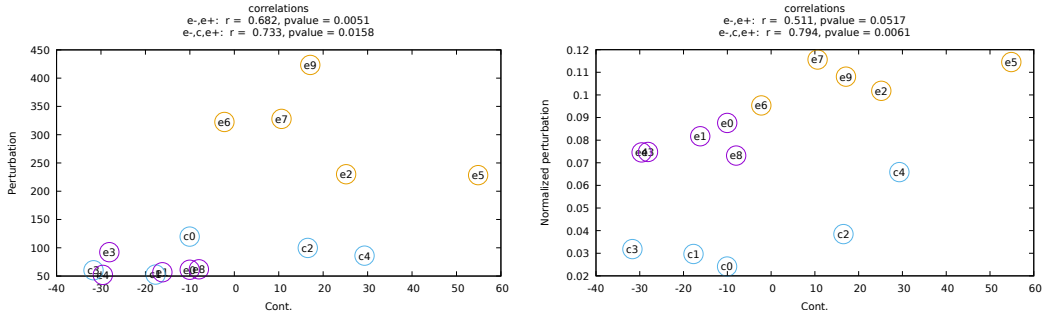


(a) P_r^+



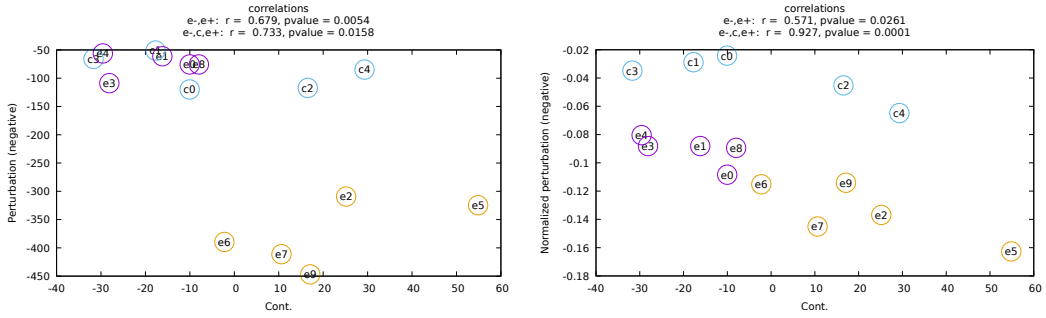
(b) $\bar{P}_r^+$

Figure L.3: Scores in Merged evaluation against positive perturbation



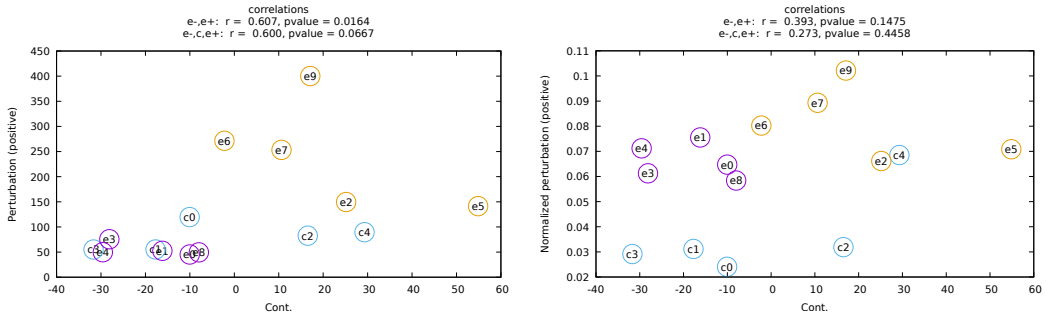
(a) P_r (b) $\bar{P}_r$

Figure L.4: Scores in Contiguous evaluation against perturbation



(a) P_r^- (b) $\bar{P}_r^-$

Figure L.5: Scores in Contiguous evaluation against negative perturbation



(a) P_r^+ (b) $\bar{P}_r^+$

Figure L.6: Scores in Contiguous evaluation against positive perturbation

Bibliography

[Adami et al. 1994] Adami C. and C. T. Brown (May 1994). "Evolutionary Learning in the 2D Artificial Life System "Avida"". In: *Artificial Life IV* 1194, pp. 1–5. ISSN: 00142999. arXiv: 9405003 [adap-org] (cit. on pp. 31, 32).

[Adami et al. 2000] Adami C., C. Ofria, and T. C. Collier (Apr. 2000). "Evolution of biological complexity". In: *Proceedings of the National Academy of Sciences* 97.9, pp. 4463–4468. ISSN: 0027-8424. DOI: 10.1073/pnas.97.9.4463 (cit. on p. 33).

[Aubert-Kato et al. 2015] Aubert-Kato N., O. Witkowski, and T. Ikegami (July 2015). "The Hunger Games: Embodied agents evolving foraging strategies on the frugal-greedy spectrum". In: *07/20/2015-07/24/2015*. 1968. The MIT Press, pp. 357–364. ISBN: 9780262330275. DOI: 10.7551/978-0-262-33027-5-ch065 (cit. on pp. 36, 43).

[Auerbach et al. 2010] Auerbach J. E. and J. C. Bongard (2010). "Evolving CPPNs to grow three-dimensional physical structures". In: *Proceedings of the 12th annual conference on Genetic and evolutionary computation - GECCO '10*. New York, New York, USA: ACM Press, p. 627. ISBN: 9781450300728. DOI: 10.1145/1830483.1830597 (cit. on pp. 23, 24).

[Banzhaf 2003] Banzhaf W. (2003). "Artificial Regulatory Networks and Genetic Programming". In: *Genetic Programming Theory and Practice*. Ed. by R. L. Riolo and B. Worzel. Boston, MA: Springer US. Chap. 4, pp. 43–61. ISBN: 1-4020-7581-2. DOI: 10.1007/978-1-4419-8983-3_4 (cit. on pp. 17, 18).

[Bonfim et al. 2005] Bonfim D. M. and L. N. de Castro (2005). "FranksTree: A Genetic Programming Approach to Evolve Derived Bracketed L-Systems". In: pp. 1275–1278. DOI: 10.1007/11539087_168 (cit. on p. 81).

[Bongard et al. 2003] Bongard J. C. and R. Pfeifer (2003). "Evolving Complete Agents using Artificial Ontogeny". In: *Morpho-functional Machines: The New Species*. Tokyo: Springer Japan, pp. 237–258. DOI: 10.1007/978-4-431-67869-4_12 (cit. on p. 19).

[Bornhofen et al. 2011] Bornhofen S., S. Barot, and C. Lattaud (Jan. 2011). "The evolution of CSR life-history strategies in a plant model with explicit physiology and architecture". In: *Ecological Modelling* 222.1, pp. 1–10. ISSN: 03043800. DOI: 10.1016/j.ecolmodel.2010.09.014 (cit. on pp. 2, 37, 38, 43, 44, 86, 114, 117).

[Bornhofen et al. 2009] Bornhofen S. and C. Lattaud (June 2009). "Competition and evolution in virtual plant communities: a new modeling approach". In: *Natural Computing* 8.2, pp. 349–385. ISSN: 1567-7818. DOI: 10.1007/s11047-008-9089-5 (cit. on pp. 10, 48, 59, 61).

[Butlin et al. 2008] Butlin R. K., J. Galindo, and J. W. Grahame (Sept. 2008). "Sympatric, parapatric or allopatric: the most important way to classify speciation?" In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 363.1506, pp. 2997–3007. ISSN: 0962-8436. DOI: 10.1098/rstb.2008.0076 (cit. on p. 96).

[Canino-Koning et al. 2016] Canino-Koning R., M. J. Wiser, and C. Ofria (2016). "The Evolution of Evolvability : Changing Environments Promote Rapid Adaptation in Digital Organisms". In: *Proceedings of the European Conference on Artificial Life*, pp. 268–275 (cit. on pp. 1, 33, 43).

[Chaumont et al. 2007] Chaumont N., R. Egli, and C. Adami (Apr. 2007). "Evolving virtual creatures and catapults". In: *Artificial Life* 13.2, pp. 139–157. ISSN: 10645462. DOI: 10.1162/artl.2007.13.2.139 (cit. on pp. 15, 16).

[Chavoya et al. 2006] Chavoya A. and Y. Duthen (2006). "Using a genetic algorithm to evolve cellular automata for 2D/3D computational development". In: *Proceedings of the 8th annual conference on Genetic and evolutionary computation - GECCO '06*. Vol. 1. New York, New York, USA: ACM Press, p. 231. ISBN: 1595931864. DOI: 10.1145/1143997.1144036 (cit. on p. 25).

[Cheney et al. 2013] Cheney N. et al. (2013). "Unshackling Evolution: Evolving Soft Robots with Multiple Materials and a Powerful Generative Encoding". In: *Proceeding of the Fifteenth Annual Conference on Genetic and Evolutionary Computation - GECCO '13*, p. 167. ISSN: 19318499. DOI: 10.1145/2463372.2463404 (cit. on p. 24).

[Chiba et al. 2017] Chiba N., R. Suzuki, and T. Arita (Sept. 2017). "Evolution of a complex anti-predator niche construction in a physical 2D predator-prey simulation and a feature analysis of defensive structures using a deep auto-encoder". In: *Proceedings of the 14th European Conference on Artificial Life ECAL 2017*. Cambridge, MA: MIT Press, pp. 96–97. ISBN: 978-0-262-34633-7. DOI: 10.7551/ecal_a_019 (cit. on pp. 2, 27, 28, 43, 44).

[Clement 2006] Clement R. (Jan. 2006). "Visualizing Speciation in Artificial Cichlid Fish". In: *Artificial Life* 12.2, pp. 243–257. ISSN: 1064-5462. DOI: 10.1162/artl.2006.12.2.243 (cit. on p. 97).

[Corchado et al. 2009] Corchado M. A. R. et al. (Nov. 2009). "Growing plants for virtual 3D environments". In: *2009 6th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE)*. IEEE, pp. 1–6. ISBN: 978-1-4244-4688-9. DOI: 10.1109/ICEEE.2009.5393469 (cit. on p. 9).

[Coumans et al. 2013] Coumans E. et al. (2013). "Bullet physics library". In: *Open source: bulletphysics. org* 15.49, p. 5 (cit. on pp. 22, 57).

[Cussat-Blanc et al. 2012] Cussat-Blanc S. and J. Pollack (2012). "Using Pictures to Visualize the Complexity of Gene Regulatory Networks". In: *Artificial Life*, pp. 491–498 (cit. on p. 21).

[Cussat-Blanc et al. 2015] Cussat-Blanc S., K. Harrington, and J. Pollack (Dec. 2015). "Gene regulatory network evolution through augmenting topologies". In: *IEEE Transactions on Evolutionary Computation* 19.6, pp. 823–837. ISSN: 1089778X. DOI: 10.1109/TEVC.2015.2396199 (cit. on pp. 96, 98, 112).

[Cussat-Blanc et al. 2014] Cussat-Blanc S. and J. Pollack (2014). "Cracking the egg: virtual embryogenesis of real robots." In: *Artificial life* 20.3, pp. 361–83 (cit. on p. 22).

[Damer et al. 1998] Damer B., K. Marcelo, and F. Revi (1998). "Nerve Garden: A Virtual Terrarium In Cyberspace". In: *American Association for Artificial Intelligence*. ISSN: 16113349 (cit. on p. 29).

[Darwin 1859] Darwin C. (Dec. 1859). *On the origin of species by means of natural selection, or, The preservation of favoured races in the struggle for life /*. London : John Murray, pp. 113–113. DOI: 10.5962/bhl.title.68064 (cit. on pp. 98, 164).

[Dawkins 1982] Dawkins R. (1982). *The extended phenotype*. Oxford University Press (cit. on p. 28).

[Dawkins 1986] — (1986). *The blind watchmaker: Why the evidence of evolution reveals a universe without design*. Ed. by W. N. &. Company. ISBN: 0-393-31570-3 (cit. on p. 8).

[Deussen et al. 1998] Deussen O. et al. (1998). "Realistic modeling and rendering of plant ecosystems". In: *Proceedings of the 25th annual conference on Computer graphics and interactive techniques - SIGGRAPH '98*. New York, New York, USA: ACM Press, pp. 275–286. ISBN: 0897919998. DOI: 10.1145/280814.280898 (cit. on p. 29).

[Disset et al. 2014] Disset J., S. Cussat-Blanc, and Y. Duthen (July 2014). "Self-Organization of Symbiotic Multicellular Structures". In: *Artificial Life 14: Proceedings of the Fourteenth International Conference on the Synthesis and Simulation of Living Systems*. The MIT Press, pp. 541–548. ISBN: 9780262326216. DOI: 10.7551/978-0-262-32621-6-ch087 (cit. on pp. 22, 23, 28, 43, 44).

[Disset et al. 2016] — (2016). *Evolved Development Strategies of Artificial Multicellular Organisms*. Cancun, Mexico (cit. on pp. 22, 23, 81, 85, 114).

[Djezzar et al. 2019] Djezzar N. et al. (Jan. 2019). "Quorum Sensing Digital Simulations for the Emergence of Scalable and Cooperative Artificial Networks". In: *International Journal of Artificial Intelligence and Machine Learning* 9.1, pp. 13–34. ISSN: 2642-1577. DOI: 10.4018/IJAIML.2019010102 (cit. on p. 36).

[Dorin 2005] Dorin A. (2005). "A Co-evolutionary Epidemiological Model for Artificial Life and Death". In: pp. 775–784. DOI: 10.1007/11553090_78 (cit. on pp. 35, 36).

[Doursat 2009] Doursat R. (2009). "Facilitating evolutionary innovation by developmental modularity and variability". In: *Proceedings of the 11th Annual conference on Genetic and evolutionary computation - GECCO '09*. New York, New York, USA: ACM Press, p. 683. ISBN: 9781605583259. DOI: 10.1145/1569901.1569996 (cit. on pp. 19, 20).

[Doursat et al. 2014] Doursat R. and C. Sánchez (June 2014). "Growing Fine-Grained Multicellular Robots". In: *Soft Robotics* 1.2, pp. 110–121. ISSN: 2169-5172. DOI: 10.1089/soro.2014.0014 (cit. on pp. 20, 43, 44).

[Dubois et al. 2017] Dubois K., S. Cussat-Blanc, and Y. Duthen (2017). "Towards an Artificial Polytrophic Ecosystem". In: *Morphogenetic Engineering Workshop, at the European Conference on Artificial Life (ECAL) 2017 September 4* (cit. on pp. 2, 69).

[Ebner 2003] Ebner M. (2003). "Evolution and Growth of Virtual Plants". In: *Advances in Artificial Life. 7th European Conference on Artificial Life*. Vol. 2801, pp. 228–237. ISBN: 3-540-20057-6. DOI: DOI:10.1007/b12035 (cit. on pp. 37, 38, 43, 44).

[Eggenberger 1997] Eggenberger P. (1997). "Evolving Morphologies of Simulated 3d Organisms Based on Differential Gene Expression". In: *Fourth European Conference on Artificial Life*. MIT Press, p. 205 (cit. on pp. 18, 19).

[Eloy et al. 2017] Eloy C. et al. (Dec. 2017). "Wind loads and competition for light sculpt trees into self-similar structures". In: *Nature Communications* 8.1, p. 1014. ISSN: 2041-1723. DOI: 10.1038/s41467-017-00995-6 (cit. on pp. 2, 7, 38, 39, 43).

[Erdei et al. 2013] Erdei J. et al. (Sept. 2013). "Evolving gene regulatory networks controlling foraging strategies of prey and predators in an artificial ecosystem". In: *Advances in Artificial Life, ECAL*. Vol. 12. MIT Press, pp. 531–537. ISBN: 9780262317092. DOI: 10.7551/978-0-262-31709-2-ch077 (cit. on p. 35).

[Fernández et al. 2012a] Fernández J. D. et al. (Apr. 2012a). "Emergent Diversity in an Open-Ended Evolving Virtual Community". In: *Artificial Life* 18.2, pp. 199–222. ISSN: 1064-5462. DOI: 10.1162/artl_a_00059 (cit. on pp. 37, 39, 48).

[Fernández et al. 2012b] Fernández J. D., F. Vico, and R. Doursat (2012b). "Complex and diverse morphologies can develop from a minimal genomic model". In: *Proceedings of the fourteenth international conference on Genetic and evolutionary computation conference - GECCO '12*. New York, New York, USA: ACM Press, p. 553. ISBN: 9781450311779. DOI: 10.1145/2330163.2330242 (cit. on p. 26).

[Fortuna et al. 2013] Fortuna M. A. et al. (Mar. 2013). "Evolving Digital Ecological Networks". In: *PLoS Computational Biology* 9.3. Ed. by S. Wodak, e1002928. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1002928 (cit. on pp. 32, 33).

[Frénoy et al. 2012] Frénoy A., F. Taddei, and D. Misevic (July 2012). "Robustness and evolvability of cooperation". In: *Artificial Life 13*. MIT Press, pp. 53–58. ISBN: 9780262310505. DOI: 10.7551/978-0-262-31050-5-ch008 (cit. on p. 35).

[Frénoy et al. 2013] — (Nov. 2013). "Genetic Architecture Promotes the Evolution and Maintenance of Cooperation". In: *PLoS Computational Biology* 9.11. Ed. by C. O. Wilke, e1003339. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1003339 (cit. on p. 35).

[Gardner 1970] Gardner M. (1970). "Mathematical games: The fantastic combinations of John Conway's new solitaire game "life"". In: *Scientific American*. ISSN: 0036-8733. DOI: 10.1038/scientificamerican0169-116. arXiv: arXiv:1011.1669v3 (cit. on pp. 7, 29).

[Godin-Dubois 2020] Godin-Dubois K. (2020). "Retina-based foraging strategies in an open-ended evolution of Splinoids". In: *ALIFE2020*, in preparation (cit. on p. 178).

[Godin-Dubois et al. 2019a] Godin-Dubois K., S. Cussat-Blanc, and Y. Duthen (2019a). "APOGeT: Automated Phylogeny Over Geological Timescales". In: *ALIFE 2019 (MethAL workshop)*, to appear. DOI: 10.13140/RG.2.2.33781.93921 (cit. on p. 3).

[Godin-Dubois et al. 2019b] — (2019b). "Self-sustainability Challenges of Plants Colonization Strategies in Virtual 3D Environments". In: *Applications of Evolutionary Computation*. Ed. by P. Kaufmann and P. A. Castillo. Cham: Springer International Publishing, pp. 377–392. ISBN: 978-3-030-16692-2. DOI: 10.1007/978-3-030-16692-2_25 (cit. on pp. 2, 92).

[Godin-Dubois et al. 2019c] — (2019c). "Speciation under Changing Environments". In: *ALIFE 19*. Vol. 31. Cambridge, MA: MIT Press, pp. 349–356. ISBN: 978-0-262-35844-6. DOI: 10.1162/isal_a_00186 (cit. on pp. 2, 135).

[Godin-Dubois et al. 2020] — (2020). "Beneficial catastrophes: dynamical and heterogeneous environments promote population robustness in EDEA". In: *ALIFE 2020*, in preparation (cit. on p. 169).

[Goldman et al. 2013] Goldman B. W. and W. F. Punch (2013). "Reducing Wasted Evaluations in Cartesian Genetic Programming". In: pp. 61–72. DOI: 10.1007/978-3-642-37207-0_6 (cit. on p. 139).

[Gracias et al. 1997] Gracias N., J. Lima, and A. Rosa (1997). "Gaia: An Artificial Life Environment for Ecological Systems Simulation". In: (cit. on p. 39).

[Graham et al. 2007] Graham L. and F. Oppacher (Apr. 2007). "Speciation through exaptation". In: *Proceedings of the 2007 IEEE Symposium on Artificial Life, CI-ALife 2007*. IEEE, pp. 433–439. ISBN: 142440701X. DOI: 10.1109/ALIFE.2007.367827 (cit. on p. 96).

[Grime 1977] Grime J. P. (Nov. 1977). "Evidence for the Existence of Three Primary Strategies in Plants and Its Relevance to Ecological and Evolutionary Theory". In: *The American Naturalist* 111.982, pp. 1169–1194. ISSN:

0003-0147. DOI: 10.1086/283244. arXiv: arXiv:1011.1669v3 (cit. on pp. 38, 87).

[Hamann 2013] Hamann H. (Sept. 2013). "Speciation Dynamics: Generating Selective Pressure Towards Diversity". In: *Advances in Artificial Life, ECAL 2013*. MIT Press, pp. 947–954. ISBN: 9780262317092. DOI: 10.7551/978-0-262-31709-2-ch141 (cit. on p. 97).

[Hamann 2015] — (Nov. 2015). "Lessons from Speciation Dynamics: How to Generate Selective Pressure Towards Diversity". In: *Artificial Life* 21.4, pp. 464–480. ISSN: 1064-5462. DOI: 10.1162/ARTL_a_00186 (cit. on p. 97).

[Harrington et al. 2017] Harrington K. and L. Magbunduku (Sept. 2017). "Competitive dynamics in eco-evolutionary genetically-regulated swarms". In: *Proceedings of the 14th European Conference on Artificial Life ECAL 2017*. Cambridge, MA: MIT Press, pp. 190–197. ISBN: 978-0-262-34633-7. DOI: 10.7551/ecal_a_034 (cit. on p. 42).

[Hornby et al. 2001] Hornby G. S. and J. B. Pollack (2001). "Evolving L-systems to generate virtual creatures". In: *Computers and Graphics (Pergamon)* 25.6, pp. 1041–1048. ISSN: 00978493. DOI: 10.1016/S0097-8493(01)00157-1. arXiv: arXiv:1011.1669v3 (cit. on pp. 10, 11, 48).

[Ito et al. 2013] Ito T. et al. (Sept. 2013). "Coevolutionary Dynamics Caused by Asymmetries in Predator-Prey and Morphology-Behavior Relationships". In: *Ecal*. MIT Press, pp. 439–445. ISBN: 9780262317092. DOI: 10.7551/978-0-262-31709-2-ch063 (cit. on p. 48).

[Joachimczak et al. 2013] Joachimczak M. et al. (Sept. 2013). "Controlling development and chemotaxis of soft-bodied multicellular animats with the same gene regulatory network". In: *Advances in Artificial Life, ECAL 2013*. MIT Press, pp. 454–461. ISBN: 9780262317092. DOI: 10.7551/978-0-262-31709-2-ch065 (cit. on p. 20).

[Joachimczak et al. 2011] Joachimczak M. and B. Wróbel (2011). "Evolution of the morphology and patterning of artificial embryos: Scaling the tricolour problem to the third dimension". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol. 5777 LNAI. PART 1, pp. 35–43. ISBN: 9783642212826. DOI: 10.1007/978-3-642-21283-3_5 (cit. on p. 20).

[Kadish et al. 2019] Kadish D., S. Risi, and L. Beloff (2019). "An artificial life approach to studying niche differentiation in soundscape ecology". In: *The 2019 Conference on Artificial Life*. Cambridge, MA: MIT Press, pp. 52–59. ISBN: 978-0-262-35844-6. DOI: 10.1162/isal_a_00140 (cit. on p. 28).

[Knibbe et al. 2005] Knibbe C. et al. (2005). "Self-adaptation of Genome Size in Artificial Organisms". In: pp. 423–432. DOI: 10.1007/11553090_43 (cit. on pp. 33, 34).

[Knibbe et al. 2007] Knibbe C. et al. (Oct. 2007). "A Long-Term Evolutionary Pressure on the Amount of Noncoding DNA". In: *Molecular Biology and Evolution* 24.10, pp. 2344–2353. ISSN: 1537-1719. DOI: 10.1093/molbev/msm165 (cit. on p. 34).

[Komosinski 2003] Komosinski M. (2003). "The Framsticks system: versatile simulator of 3D agents and their evolution". In: *Kybernetes: The International Journal of Systems & Cybernetics* 8, pp. 156–173. ISSN: 0368-492X. DOI: 10.1108/03684920310452382 (cit. on p. 25).

[Lalejini et al. 2017] Lalejini A., M. J. Wiser, and C. Ofria (Sept. 2017). "Gene duplications drive the evolution of complex traits and regulation". In: *Proceedings of the 14th European Conference on Artificial Life ECAL 2017*. Cambridge, MA: MIT Press, pp. 257–264. ISBN: 978-0-262-34633-7. DOI: 10.7551/ecal_a_045 (cit. on pp. 33, 43).

[Langton 1986] Langton C. G. (Oct. 1986). "Studying artificial life with cellular automata". In: *Physica D: Nonlinear Phenomena* 22.1-3, pp. 120–149. ISSN: 01672789. DOI: 10.1016/0167-2789(86)90237-X (cit. on p. 6).

[Langton et al. 1989] Langton C. G. and S. Wilson (1989). "Artificial Life". In: *Artificial Life: Proceedings of an Interdisciplinary Workshop on The Synthesis and Simulation of Living Systems, Los Alamos, 1987* 24.2, p. 244. ISSN: 0024094X. DOI: 10.2307/1575317 (cit. on p. 6).

[Lassabe et al. 2007] Lassabe N., H. Luga, and Y. Duthen (Apr. 2007). "A New Step for Artificial Creatures". In: *2007 IEEE Symposium on Artificial Life*. IEEE, pp. 243–250. ISBN: 1-4244-0701-X. DOI: 10.1109/ALIFE.2007.367803 (cit. on pp. 16, 43, 48).

[Lehman et al. 2012] Lehman J. and K. O. Stanley (2012). "Beyond open-endedness: Quantifying impressiveness." In: *Proc. of ALIFE*, pp. 75–82 (cit. on p. 23).

[Lehman et al. 2008] Lehman J. and K. O. Stanley (2008). "Exploiting Open-Endedness to Solve Problems Through the Search for Novelty". In: *Artificial Life XI* Alife Xi, pp. 329–336 (cit. on pp. 10, 85).

[Lenski et al. 2003] Lenski R. E. et al. (May 2003). "The evolutionary origin of complex features". In: *Nature* 423.6936, pp. 139–144. ISSN: 0028-0836. DOI: 10.1038/nature01568 (cit. on p. 32).

[Lessin et al. 2015] Lessin D. and S. Risi (2015). "Evolved Virtual Creatures with Soft-Body Muscles". In: *Proceedings of the Companion Publication of the 2015 on Genetic and Evolutionary Computation Conference - GECCO Companion '15*. New York, New York, USA: ACM Press, pp. 761–762. ISBN: 9781450334884. DOI: 10.1145/2739482.2764897 (cit. on pp. 17, 48).

[Lindenmayer 1968] Lindenmayer A. (Mar. 1968). "Mathematical models for cellular interactions in development I. Filaments with one-sided inputs". In: *Journal of Theoretical Biology* 18.3, pp. 280–299. ISSN: 00225193. DOI: 10.1016/0022-5193(68)90079-9 (cit. on pp. 7–9).

[Luo et al. 2019] Luo T.-t. et al. (Mar. 2019). "Examining Community Stability in the Face of Mass Extinction in Communities of Digital Organisms". In: *Artificial Life* 24.4, pp. 250–276. ISSN: 1064-5462. DOI: 10.1162/artl_a_00272 (cit. on pp. 2, 33, 43).

[Metivier et al. 2002] Metivier M. et al. (2002). "A Stress-based Speciation Model in LifeDrop characters". In: *Artificial Life*, pp. 121–126 (cit. on pp. 40, 43, 97, 112).

[Miconi 2008a] Miconi T. (June 2008a). "Evosphere: Evolutionary dynamics in a population of fighting virtual creatures". In: *2008 IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence)*. IEEE, pp. 3066–3073. ISBN: 9781424418237. DOI: 10.1109/CEC.2008.4631212 (cit. on pp. 41, 42, 96).

[Miconi 2008b] — (2008b). "In Silicon No One Can Hear You Scream: Evolving Fighting Creatures". In: *Proceedings of the 11th European Conference on Genetic Programming, EuroGP 2008*. Vol. 4971 LNCS, pp. 25–36. ISBN: 3540786708. DOI: 10.1007/978-3-540-78671-9_3 (cit. on pp. 16, 17, 41, 48, 177).

[Miller et al. 2000] Miller J. F. and P. Thomson (2000). "Cartesian Genetic Programming". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol. 1802. Springer Verlag, pp. 121–132. ISBN: 3540673393. DOI: 10.1007/978-3-540-46239-2_9 (cit. on p. 138).

[Miras et al. 2018] Miras K. et al. (2018). "Effects of Selection Preferences on Evolved Robot Morphologies and Behaviors". In: *The 2018 Conference on Artificial Life*. Cambridge, MA: MIT Press, pp. 224–231. DOI: 10.1162/isal_a_00047 (cit. on pp. 10, 11).

[Misevic et al. 2012] Misevic D. et al. (July 2012). "Effects of public good properties on the evolution of cooperation". In: *Artificial Life 13*. MIT Press, pp. 218–225. ISBN: 9780262310505. DOI: 10.7551/978-0-262-31050-5-ch030 (cit. on p. 35).

[Nahum et al. 2017] Nahum J. R. et al. (Sept. 2017). "Improved adaptation in exogenously and endogenously changing environments". In: *Proceedings of the 14th European Conference on Artificial Life ECAL 2017*. Cambridge, MA: MIT Press, pp. 306–313. ISBN: 978-0-262-34633-7. DOI: 10.7551/ecal_a_052 (cit. on pp. 2, 33, 43, 44).

[Oei et al. 1991] Oei C., D. Goldberg, and S.-J. Chang (1991). "Tournament selection, niching, and the preservation of diversity". In: *Illigal report* 51, p. 61801 (cit. on p. 96).

[Ofria et al. 1999] Ofria C. et al. (1999). "Evolution of Differentiated Expression Patterns in Digital Organisms". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol. 1674, pp. 129–138. DOI: 10.1007/3-540-48304-7_19 (cit. on p. 32).

[Olson et al. 2016] Olson R. et al. (2016). "Exploring the coevolution of predator and prey morphology and behavior". In: *Proceedings of the Artificial Life Conference 2016*. Cambridge, MA: MIT Press, pp. 250–257. ISBN: 978-0-262-33936-0. DOI: 10.7551/978-0-262-33936-0-ch045. arXiv: 1602.08802 (cit. on p. 25).

[Ouannes et al. 2014] Ouannes N. et al. (Dec. 2014). "Modeling a bacterial ecosystem through chemotaxis simulation of a single cell". In: *Artificial Life and Robotics* 19.4, pp. 382–387. ISSN: 1433-5298. DOI: 10.1007/s10015-014-0187-4 (cit. on pp. 36, 43).

[Payne et al. 2007] Payne J. L., M. J. Eppstein, and C. C. Goodnight (Apr. 2007). "Sensitivity of Self-Organized Speciation to Long-Disctance Dispersal". In: *2007 IEEE Symposium on Artificial Life*. IEEE, pp. 1–7. ISBN: 1-4244-0701-X. DOI: 10.1109/ALIFE.2007.367651 (cit. on p. 97).

[Pichler et al. 2008] Pichler P. P. and L. Cañamero (2008). "Evolving morphological and behavioral diversitywithout predefined behavior primitives". In: *Artificial Life XI: Proceedings of the 11th International Conference on the Simulation and Synthesis of Living Systems, ALIFE 2008*, pp. 474–481 (cit. on p. 41).

[Prusinkiewicz et al. 2013] Prusinkiewicz P. and B. Lane (2013). "Modeling Morphogenesis in Multicellular Structures with Cell Complexes and L-systems". In: pp. 137–151. DOI: 10.1007/978-3-642-20164-6_12 (cit. on pp. 11, 12).

[Prusinkiewicz et al. 1988] Prusinkiewicz P., A. Lindenmayer, and J. Hanan (1988). "Development models of herbaceous plants for computer imagery purposes". In: *ACM SIGGRAPH Computer Graphics* 22.4, pp. 141–150. ISSN: 00978930. DOI: 10.1145/378456.378503 (cit. on pp. 9, 48).

[Rasmussen et al. 1990] Rasmussen S. et al. (June 1990). "The coreworld: Emergence and evolution of cooperative structures in a computational chemistry". In: *Physica D: Nonlinear Phenomena* 42.1-3, pp. 111–134. ISSN: 01672789. DOI: 10.1016/0167-2789(90)90070-6 (cit. on p. 29).

[Ray 1991] Ray T. S. (1991). "An approach to the synthesis of life". In: *SFI Studies in the Sciences of Complexity* X (cit. on pp. 7, 29, 31, 183).

[Ray 1993] — (Oct. 1993). "An Evolutionary Approach to Synthetic Biology: Zen and the Art of Creating Life". In: *Artificial Life* 1.1_2, pp. 179–209. ISSN: 1064-5462. DOI: 10.1162/artl.1993.1.1_2.179 (cit. on p. 6).

[Ray 2001] — (Aug. 2001). "Aesthetically Evolved Virtual Pets". In: *Leonardo* 34.4, pp. 313–316. ISSN: 0024-094X. DOI: 10.1162/00240940152549230 (cit. on pp. 13, 15).

[Reil 1999] Reil T. (1999). "Dynamics of Gene Expression in an Artificial Genome — Implications for Biological and Artificial Ontogeny". In: pp. 457–466. DOI: 10.1007/3-540-48304-7_63 (cit. on pp. 17, 18).

[Reynolds 1987] Reynolds C. W. (1987). "Flocks, herds and schools: A distributed behavioral model". In: *Proceedings of the 14th annual conference on Computer graphics and interactive techniques - SIGGRAPH '87*. New York, New York, USA: ACM Press, pp. 25–34. ISBN: 0897912276. DOI: 10.1145/37401.37406 (cit. on p. 42).

[Rieffel 2013] Rieffel J. (2013). "Heterochronic Scaling of Developmental Durations in Evolved Soft Robots". In: *Gecco'13: Proceedings of the 2013*

Genetic and Evolutionary Computation Conference, pp. 743–750. DOI: 10.1145/2463372.2463466 (cit. on p. 26).

[Shao et al. 2010] Shao J. and T. S. Ray (2010). "Maintenance of species diversity by predation in the tierra system". In: *Artificial Life XII: Proceedings of the 12th International Conference on the Synthesis and Simulation of Living Systems, ALIFE 2010*, pp. 533–540 (cit. on p. 31).

[Shim et al. 2004] Shim Y.-S., S. J. Kim, and C. H. Kim (2004). "Evolving Flying Creatures with Path Following Behaviors". In: *9th International Conference on the Simulation and Synthesis of Living Systems* October, pp. 125–132 (cit. on p. 25).

[Sims 1994a] Sims K. (1994a). "Evolving 3D Morphology and Behavior by Competition". In: *Artificial Life* 1.4, pp. 353–372. ISSN: 1064-5462. DOI: 10.1162/artl.1994.1.4.353 (cit. on pp. 1, 11, 14, 96).

[Sims 1994b] — (1994b). "Evolving virtual creatures". In: *Proceedings of the 21st annual conference on Computer graphics and interactive techniques - SIGGRAPH '94*. New York, New York, USA: ACM Press, pp. 15–22. ISBN: 0897916670. DOI: 10.1145/192161.192167 (cit. on pp. 7, 11, 14, 25, 48, 65, 66, 81, 171).

[Singh 2012] Singh B. N. (2012). "Concepts of species and modes of speciation". In: *Current Science* 103.7 (cit. on pp. 2, 96, 98, 238).

[K. O. Stanley et al. 2005] Stanley K. O., B. D. Bryant, and R. Miikkulainen (2005). "Evolving Neural Network Agents in the NERO Video Game". In: *Proceedings of the IEEE* Essex Univ, pp. 182–189 (cit. on p. 7).

[K. Stanley et al. 2002] Stanley K. and R. Miikkulainen (2002). "Efficient evolution of neural network topologies". In: *Proceedings of the 2002 Congress on Evolutionary Computation. CEC'02 (Cat. No.02TH8600)*. Vol. 2. figure 1. IEEE, pp. 1757–1762. ISBN: 0-7803-7282-4. DOI: 10.1109/CEC.2002.1004508 (cit. on pp. 96, 112).

[K. O. Stanley 2007] Stanley K. O. (June 2007). "Compositional pattern producing networks: A novel abstraction of development". In: *Genetic Programming and Evolvable Machines* 8.2, pp. 131–162. ISSN: 1389-2576. DOI: 10.1007/s10710-007-9028-8 (cit. on pp. 23, 24).

[Steinberg et al. 1999] Steinberg D. et al. (1999). "Plant growth simulation in virtual worlds: towards online artificial ecosystems". In: *First Workshop on Artificial Life Integration in Virtual Environments (ALIVE1) at the 5th European Conference on Artificial Life (ECAL1999)*. Lausanne, Switzerland, pp. 19–24 (cit. on p. 29).

[Thearling et al. 1996] Thearling K. and T. S. Ray (1996). "Evolving Parallel Computation". In: *Complex Systems* 10.3, pp. 229–237 (cit. on p. 31).

[Turk 2010] Turk G. (2010). "Sticky Feet : Evolution in a Multi-Creature Physical Simulation". In: *Artificial Life XII. Proceedings of the 12th International Conference on the Synthesis and Simulation of Living Systems*, pp. 496–503 (cit. on pp. 41, 42).

[Ventrella 2005] Ventrella J. (2005). "GenePool: Exploring the interaction between natural selection and sexual selection". In: *Artificial Life Models in Software*, pp. 81–96. DOI: 10.1007/1-84628-214-4_4 (cit. on pp. 40, 177).

[Wilson et al. 2001] Wilson R. and F. Keil (2001). *The MIT Encyclopedia of the Cognitive Sciences*. MIT Press. ISBN: 9780262731447 (cit. on p. 6).

[Witkowski et al. 2019] Witkowski O. and T. Ikegami (May 2019). "How to make swarms open-ended? evolving collective intelligence through a constricted exploration of adjacent possibles". In: *Artificial Life* 25.2, pp. 178–197. ISSN: 15309185. DOI: 10.1162/artl_a_00288. arXiv: 1903.08228 (cit. on p. 42).

[Woehrer et al. 2012] Woehrer M., D. Hougen, and I. Schlupp (2012). "Sexual selection, resource distribution, and population size in synthetic sympatric speciation". In: *Artificial Life 13: Proceedings of the 13th International Conference on the Simulation and Synthesis of Living Systems, ALIFE 2012*, pp. 137–144. ISBN: 9780262310505 (cit. on pp. 97, 177).

[Wood et al. 2008] Wood A. J. et al. (Jan. 2008). "Daisyworld: A review". In: *Reviews of Geophysics* 46.1, RG1001. ISSN: 8755-1209. DOI: 10.1029/2006RG000217 (cit. on p. 37).

[Yaeger 1994] Yaeger L. S. (1994). "Computational Genetics, Physiology, Metabolism, Neural Systems, Learning, Vision and Behavior or Polyworld: Life in a New Context". In: 1, pp. 1–25 (cit. on pp. 39, 40).

[Zahadat et al. 2017a] Zahadat P., D. N. Hofstadler, and T. Schmickl (2017a). "Vascular morphogenesis controller: A Generative Model For Developing Morphology". In: *Proceedings of the Genetic and Evolutionary Computation Conference on - GECCO '17*. New York, New York, USA: ACM Press, pp. 163–170. ISBN: 9781450349208. DOI: 10.1145/3071178.3071247 (cit. on pp. 26, 27, 43, 44, 74, 114).

[Zahadat et al. 2017b] Zahadat P. and T. Schmickl (Oct. 2017b). "Evolving vascular morphogenesis controller to demonstrate locomotion". In: *2017 Artificial Intelligence and Signal Processing Conference (AISP)*. IEEE, pp. 352–357. ISBN: 978-1-5386-2585-9. DOI: 10.1109/AISP.2017.8515124 (cit. on pp. 26, 27).

[Zhang et al. 1996] Zhang T., R. Ramakrishnan, and M. Livny (1996). "BIRCH: an efficient data clustering method for very large databases". In: *Proceedings of the 1996 ACM SIGMOD international conference on Management of data - SIGMOD '96*. Vol. 1. 2. New York, New York, USA: ACM Press, pp. 103–114. ISBN: 0897917944. DOI: 10.1145/233269.233324 (cit. on p. 98).

Glossary

AABB Axis Aligned Bounding Box. 78, 122

Abiotic The abiotic component of an ecosystem are non-living parts of the environment that has an impact on living organisms. xiii, 32, 37, 43, 44, 121, 126, 131, 134

Abscission Spontaneous shedding of a body part such as leaves or fruits. 71

Allopatric Mode of speciation that occurs when biological populations of the same species become isolated from each other to an extent that prevents of interferes with gene flow. Genetic drift and different evolutionary pressures on each sub-population lead to cladogenesis preventing reversion to a single species should those sub-populations meet again. xv, 127, 129, 235

Anagenesis Gradual evolution of a species that continues to exist as an interbreeding population. 4, 95, 110

Biotic The biotic component of an ecosystem are the living entities of which it is composed and which have an impact on the environment and other living entities. xiii, 32, 37, 120, 126

Cladogenesis Evolutionary splitting of a parent species into two distinct (reproductively isolated) species. 4, 95, 97, 100, 105, 127, 235

LUCA The Last Universal Common Ancestor is the most recent population of organisms from which all currently observable organisms have a common descent. 99

OBB Oriented Bounding Box. 122

Parapatric Mode of speciation that occurs when two sub-populations of the same species evolve reproductive isolation from one another while continuing to exchange genes. xv, 127, 130

Peripatric Mode of speciation where an a new species is formed from an isolated peripheral population. The primary difference with allopatric speciation is that one the population is smaller than the other. xv, 127, 133

Quaternion Unit quaternions, also known as versors, provide a convenient mathematical notation for representing orientations and rotations of objects in three dimensions. Compared to Euler angles they are simpler to compose and avoid the problem of gimbal lock. Compared to rotation matrices they are more compact, more numerically stable, and more efficient. 65, 77

Quorum sensing Ability to detect and respond to cell population density through regulation of specific genes. 36

Red Queen Effect The Red Queen Effect is named in reference to a response made by the Req Queen in "Alice in Wonderland": In this place one must run constantly to stay in the same place. In biological terms this refers to an evolutionary dynamic between pairs of competing populations for which any improvement from one of the competitor is matched, at a latter date, by the other. Examples of such a dynamic include predator/prey configurations where every increase in the prey's capability to escape death is quickly counteracted by the predator, and oppositely. 12, 37, 96

Speciation Evolutionary process by which populations evolve to become distinct species. 98, 115, 126–131, 133–135, 235

Species "group of potentially interbreeding natural population reproductively isolated from other such groups" [Singh 2012]. 2–4, 98, 114, 126–135, 235, 236