Speciation under changing environments

Kevin Godin-Dubois, Sylvain Cussat-Blanc and Yves Duthen

University of Toulouse, IRIT - CNRS UMR 5505, 2 rue du Doyen Gabriel Marty, 31042 Toulouse, France
{kevin.dubois, sylvain.cussat-blanc, yves.duthen}@irit.fr

Abstract

Progress in molecular genetics allowed taxonomists to better understand the relationships between species without the bias of morphological similarities. However, access to data from times past is limited to the fossil archives which, being far from complete, can only provide limited information. To address this problem through the field of Artificial Life, we devised a polyvalent sexual reproduction scheme and an automated phylogenetic tool capable of producing, from a stream of genomes, hierarchical species trees with relatively low memory footprint. We assert that these apparatus perform well under reasonable stress by embedding them into 2D simulations of unsupervised plant evolution in textbook cases of geographical speciation. After thousands of generations and millions of plants, the extracted phylogenetic data not only showed the expected results in terms of branching pattern (anagenesis, cladogenesis) but also exhibited complex interactions between species both in space and time.

Introduction

Phylogenetic trees of our world's species display the overwhelming amount of variations, adaptations and bifurcations generated by unbridled evolution over the course of a few billion years. Despite being visualization tools, designed to classify a continuum into more easily manageable chunks, they can provide further insight into the underlying mechanics of natural selection be it in biological or artificial systems.

Before getting to point where phylogeny is relevant, one first needs the basic block of any evolutionary process: individuals. Models of genotype-phenotype mapping abound in the literature with a specific emphasis on the class of morphologies they can generate. One of the first contributions is the Lindenmayer Systems (Prusinkiewicz et al., 1995), heavily inspired by the branching patterns exhibited by plants, which encode, in a very compact form, recursive derivation and can reproduce life-like instances in both 2D and 3D (Bornhofen, 2008). The directed graphs designed in (Sims, 1994) follow a similar approach by defining body segments and the, potentially recurrent, relationships between them and, though originally designed to model motile creatures, were successfully applied to plant morphologies (Dubois et al., 2017), as well. Further generalization led to the biologically inspired Genetic Regulatory Networks which, by defining the cell as the elementary unit, emulate its internal chemistry through self-interacting 'proteins' controlling its life-cycle. Using such a low-resolution building block allowed for the generation of specific shapes (Joachimczak and Wróbel, 2008) and organ emergence for creatures embedded in a virtual environment (Disset et al., 2016).

Natural selection, however, is not a genetic algorithm and the metaphor fails as soon as one tailors a fitness function to drive evolution into solving an optimization problem. Indeed, when left unchecked, even the simplest of rule set such as (Gardner, 1970) can create such diversity that they are still investigated almost fifty years later. Thus, it ensues that simulations in which individuals roam free have been designed: ranging from (Adami and Brown, 1994), with its computer programs fighting one-another for memory space, to (Metivier et al., 2002; Ventrella, 2005) where motile creatures are required to actively look for mating partners.

Despite the preponderant place given to the living parts of an ecosystem, its abiotic component is equally important given that models of the biosphere using only water and temperature as variables were found to account for most of the vegetal biodiversity observed in Nature (Woodward and Williams, 1987). The same holds true for artificial simulations as disruptiveness, whether sudden or diffused, promotes different strategies and leads to diversity (Bornhofen et al., 2011). Unforeseeable environmental dynamics add another layer of complexity in the generated individuals by selecting those that exhibit better adaptability (Caninokoning et al., 2016).

Circling back to the biotic component, we see its self-driving force in numerous examples of co-evolution, co-adaptation, competition, whether in the natural world or artificial systems (Miconi, 2008): arming race is a striking case of an inter-species conflict able to quickly promote divergence of character or optimization. But the concept of species is a blurred one: though one can argue that for individuals there is no such thing as a species, only mates and non-mates, we refer to the definition of *biolog-*

ical species given in (Singh, 2012) as a "group of potentially interbreeding natural population reproductively isolated from other such groups". Similarly, the process of speciation, by which species are created, has been described in a number of ways without the emergence of global consensus. (Butlin et al., 2008) argues that attempts at categorizing a continuous multi-dimensional phenomenon by discrete topology-centered methods makes one lose sight of the adjacent equally important factors.

This paper comes as a proof-of-concept of how environmental conditions can be used as the sole driving force of an evolutionary process. To this end, we hereby describe the model for autonomously reproducing individuals and the phylogenetic tool used to monitor species dynamics.

Self-reproducing vegetals

Here, individuals are not only expected to adapt to an unknown, and potentially precarious, environment but also to thrive by self-reproducing to the utmost limit. As our objective is to study evolutionary dynamics instead of individual development, we decided upon using L-Systems as our morphological controller, as opposed to more complex plant growth models, thanks to their intelligibility and computational lightness.

L-System

Each plant comes with a pair of L-System: the shoot and root. These are deterministic, context-free and share the same set of control instruction: $+/-$ for left/right rotations, $[...]$ branching, $A...F$ non-terminals and S the initial non-terminal. 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).

Individual rules can mutate through duplication, replacement or suppression of an existing symbol, extraction into its own branch (e.g. slf becoming $s[l]f$) and swapping adjacents (e.g. slf giving sfl). In addition, rules can be added (e.g. $S \to slf$ giving $S \to sAf; A \to l$) or removed (e.g. $S \to sAf; A \to l$ reducing to $S \to sf$).

Some restrictions apply to these operators. There is always at least one rule, in which case it must be the one derivating the initial non-terminal (S which can be seen as a seed) so that the plant can germinate. As these L-Systems are deterministic the maximal number of rules is the size of the non-terminals set. Rules cannot be longer than $M = 4$ non-control ($\notin \{+, -, [,]\}$) characters long so that derivations must occur when aiming for complex morphologies. Finally, the number of replacements a plant can perform for a given compartment is limited to a small value $D \leq 5$, itself subject to mutations, which bounds the number of symbols in the derived phenotype to M^D.

Constants	
k	assimilation rate
J	saturation rate
f	resource cost
l	life cost
m_{Tr}, s_{Tr}	temperature range regulation
Genetic fields	
g_s	Growth speed
m_T, s_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 $T < m_T$, 0 otherwise
T^+	1 if $T > m_T$, 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 1: Metabolic variables

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 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 1. Given the bell curve function of mean m and standard deviation s

$$gauss(x, m, s) = exp^{-\frac{(x-m)}{2s^2}} \quad (1)$$

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

$$h_{eff}(T) = gauss(T, m_T, s_T)\ gauss(s_T, m_{Tr}, s_{Tr}) \quad (2)$$

The left-hand part of the equation impedes the metabolism as T goes further from the plant's optimal temperature m_T while the right-hand part regulates the tolerance range s_T so that it cannot grow unchecked. Indeed, 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). This impacts water uptake as, the lower the temperature is below m_T, the less a plant can absorb water through its root hairs:

$$U_W(T) = \frac{T^-(h_{eff}(T) - 1) + 1}{1 + C_w^{rt} J} \sum_{h, \text{root hair}} k w_h s_h \quad (3)$$

Nonetheless, the root compartment shares a portion of its water reserve to the shoot, according to the relative concen-

trations and transport resistors:

$$T_W = \frac{C_W^{rt} - C_W^{sh}}{\frac{R_W}{X^{rt}} + \frac{R_W}{X^{sh}}} \tag{4}$$

Leaves in the upper layer with direct access to sunlight then produce glucose and similarly to (4) transports part of it to the lower layer.

$$U_G = \frac{1}{1 + C_G^{sh} J} \sum_{l,leaf} k l_l \tag{5}$$

$$T_G = \frac{C_G^{sh} - C_G^{rt}}{\frac{R_G}{X^{rt}} + \frac{R_G}{X^{sh}}} \tag{6}$$

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

$$R_W^{sh}(T) = (1 - T^{+} h_{eff}(T)) R_W^{sh} \tag{7}$$

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:

$$W^L(T) = l(2 - h_{eff}(T)) X^L \tag{8}$$

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.

$$\dot{X}^L(T) = g_s X^L C_W^L C_G^L - W^L(T) \tag{9}$$

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, thus preventing immortal phenotypes from monopolizing the environment.

Self-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 schemes have been developed. We chose to focus, in this work, on sexual reproduction because of its greater degree of interactions and inter-species diversity.

The subset of a plant's genotype devoted to reproduction includes its gender, compatibility metrics $CM = \{\mu, \sigma_i, \sigma_o\}$ and sexual organs. These interact with one another according to the algorithm defined in previous work (Godin-Dubois et al., 2019).

The genotypic distance is defined recursively: given A,B two genomes, e an elementary field with range $[e_{min}, e_{max}]$ (e.g. the growth speed g_s) and r a compound field composed of subfields $f^1 \dots f^n$ with weights $w^1 \dots w^n$

$$dist(e_A, e_B) = \frac{|e_A - e_B|}{e_{max} - e_{min}} \tag{10a}$$

$$dist(r_A, r_B) = \sum_{i=1}^{n} w^i dist(f_A^i, f_B^i) \tag{10b}$$

This metric is *objective* in the sense that it provides information on the amount of genetic divergence between a pair of individuals but makes no hypothesis as to their capability to mate. The compatibility value, on the other hand, is *subjective* and asymmetric as different sets of the reproduction parameters CM may give very different results for an identical genetic distance d.

This crossover operator differs from those commonly found in the literature (Sims, 1994; Bornhofen, 2008; 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.

Embedding the compatibility function into the genome allows for the emergence of species-specific segregation schemes which is of utmost importance as our interest lies in obtaining speciation as a by-product of reproduction at the individual level. Furthermore having both in-/out-breeding coefficients makes specification of the search spaces possible, with adaptive plants accepting a broader range of incoming genetic material while more conservative ones could instead focus on controlled inbreeding to solidify their alleles.

Automated phylogeny tool (APOGeT)

Studying long term evolutionary dynamics generates a massive amount of data which precludes observation at an individual level. To this end, we devised a tool for automated phylogeny which only relies on genomes possessing both a distance metric and a compatibility function (e.g. as defined above).

In APOGeT, species are modeled by a fixed-size collection of "representative" points that form an envelope in the

genetic space. This allows for a compact, yet diverse, description of a species without resorting to centroids which, when applicable, would shrink individual differences.

The procedure for inserting a genome g in the tree is twofold: first, determine the correct species for g and update the envelope if need. Given S, the species of g's parents and $env(S)$ the collection of representatives, we test whether $match(g, S) \geq T$, given that:

$$xcompat(g, e) = min(g.compat(d), e.compat(d)) \quad (11)$$
$$\text{with } d = dist(g, e)$$

$$match(g, S) = \frac{1}{|env(S)|} \sum_{e \in env(S)} xcompat(g, e) \quad (12)$$

If the result is positive we can assign g to S. Otherwise, the procedure is performed for each direct subspecies of S until either a match is found or a new species is created with g as its sole representative.

Then the envelope E of the modified species' is checked for update. If it is not yet full (less than K are recorded) then the genome is simply appended. Otherwise, g's contribution is confronted against that of the $e_i \in E$ according to:

$$C(E, g) = \max_{e_i \in E}(- \min_{e_j \neq e_i} dist(e_i, e_j) + \min_{e_j \neq e_i} dist(g, e_j)) \quad (13)$$

If $C(E, g) > 0$, then g is more different than a current envelope point and will be inserted in its place. In this manner, the envelope for a given species is a set of those individuals, while still capable of inter-crossing, that are the most different.

Hybridization

Earlier work on this tool faced us with the problem of hybridization between species, whether occasional or deliberate. Indeed the algorithm presented above makes the assumption that both parents belong to the same species which, under unrestricted genome flow conditions, is not necessarily the case. This implies that for a given species I individuals may come from a, potentially large, set of candidate parent species $H_1 \dots H_n$ thus changing this tool from a tree to a graph and losing much intelligibility in the process.

To solve this issue we introduced the concept of *major contributor* as follow: Given g with parents p_m, p_f belonging to different species S_f,S_m, the algorithm presented earlier makes the assumption that $S_f = S_m$. The sufficient extension managing multiple parent species is to check against both and resume the rest of the procedure for:

$$S = \arg\max_{S_f, S_m} match(g, S_i) \quad (14)$$

In order to keep track of these hybridizations, each species S maintains a list of contributions $C^S = \{\{S_i, c_i\} \dots\}$ which records for any species S_i how many times c_i it provided genetic material. This allowed us to the redefine the notion of parent species as the *major contributor* i.e. $S_i \in C^S / S_i \neq S, \forall_j c_i > c_j$. It also implies that whole subtrees can be reparented to reflect the change in genetic material source.

The environment

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

Topological y, with seeds being much harder to disseminate onto higher ground

Hygrometric w, water availability has a direct impact on the plants' ability to thrive

Temperature t, equations (2-8) show how deleterious this can be on the metabolism

The system is designed as a closed one, 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 ($\in [0 : 1]$) and Y the relative time in the planned simulation duration (same range) and x the position in the environment. All other values have range $[-1 : 1]$. For this article, we resorted to a simple expression parser to easily define straightforward experimental validations.

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.

$$C0 : \hat{t} = -c_0 \mathbb{1}_{\mathbb{R}_{\geq 0}} y + (1 - c_0)\dot{t} \quad (15)$$

$$C1 : \hat{w} = -c_1 \mathbb{1}_{\mathbb{R}_{\leq 0}} t + (1 - c_1)\dot{w} \quad (16)$$

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

Additionally, a bare-bone physics engine is embedded in the system to prevent plant-plant collisions, manage light availability and perform mates detection. All of our simulations generate the initial population from a single primordial genome, which is cloned 100 times and disseminated regularly around the center of the environment. These are then left to their own devices for a number of years where days and years have durations of 10 ticks and 100 days, respectively.

Experiments

In order to validate both our autonomous reproduction scheme and phylogeny extraction tool we devised simple scenarios to test our system on. Namely, we explore allopatric, parapatric and a form of peripatric speciations.

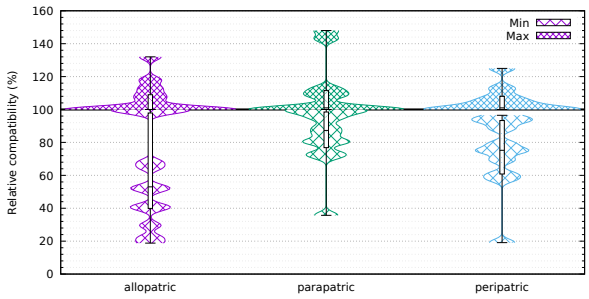


Figure 1: Speciation results for the three experiments

Parameters not subjected to variation are the environments' width (100m) and height (50m), simulation time of 100 years and identical primordial plant genome.

In all of these experiments, 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 the following metrics:

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:

$$c_a(A,B) = \frac{1}{|P^A||P^B|} \sum_{f \in P^A} \sum_{m \in P^B} f.compat(m) \quad (17)$$

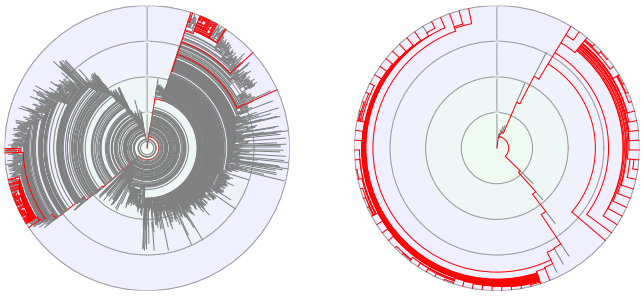
That is the average compatibility between possible mating pairs of each considered species. We then derived from $c_a(A,B)$ the relative compatibility as follow:

$$c_r(A,B) = \frac{c_a(A,B)}{c_a(A,A)} \quad (18)$$

which provides a normalized metric whose comparison between different reproductive trends or even simulations is more straightforward. Results across all three experiments are summarized in figure 1 with an uneven number of repeats: 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 is not the case, we can safely conclude that *some* did occur, which will be explored in the following sections.

Allopatric speciation

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



(a) Complete (b) Simplified

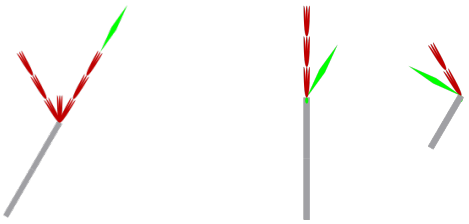
Figure 2: Phylogenic tree for the lowest c_r at the 100th year

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 previously. This gradual process produces, at the end of the simulation, a topological 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 1, speciation did occur in this experiment, however aggregated data can only show a coarse picture. To this end, we extracted the phylogenetic tree produced during the most successful run (minimal $c_r = 18.8\%$, maximal = 100%) which can be seen in figure 2a.

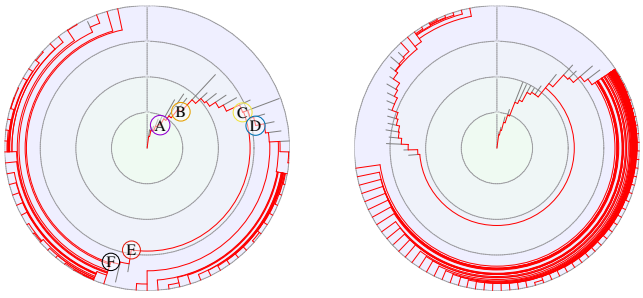
For a given species the number of available information is limited to the minimum of what can be easily processed at a glance: an arced path connects it to its parent species with the distance to center providing the date of the first appearance of this species. The timeline pointing outward stops as soon as no more individuals can be found in the simulation. Additionally, paths in red denotes species on a 'survivor path' i.e. those that left living descendants at the end of the simulation.

One can clearly distinguish the two species clusters stemming from the geographical separation with the lower part of the leftmost 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



(a) Most complex (b) Divergent

Figure 3: Morphologies show limited complexity



(a) 'Worst' (b) Best

Figure 4: Phylogenic trees for the parapatric runs with the most extreme speciations at y=100.

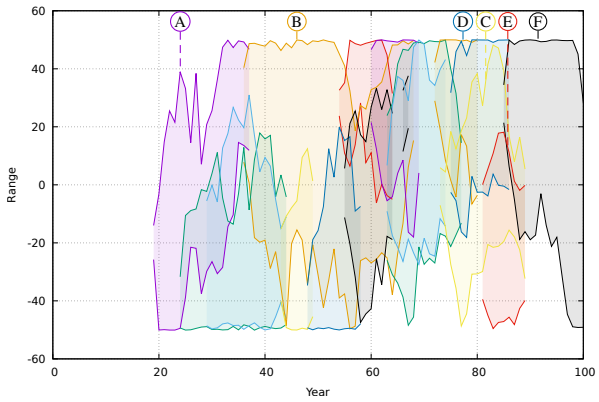


Figure 5: Colonization ranges for the 16 most populated species in a 'negative' run

much insight on when speciation did occur.

To this end we use, instead, the simplified version in figure 2b which only shows the species on the survivor paths. 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 topological 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 3a, 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 3b 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, ...) were direct observation is much less straightforward and is left to future work.

Parapatric speciation

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

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

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 1 to assert that this experiment also produces divergences and clustering though more limited in range due to cross-breeding being expected but not enforced. One should also note that some simulations fail 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 4b 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 compatibility indicates that for at least one species it is 1.5 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. Summarized in figure 5 are the dynamics of the 16 more populated species generated by this 'worse' simulation. 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, as time goes by and temperature diverges in the desert (lower part of the graph) and temperate regions (upper part) various dynamics emerge. 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 4a).

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.

Peripatric speciation

$$
\begin{aligned}
c0 &= 1 \\
\dot{y} &= .4sin(.5Y\pi)(.5(tanh(8(x-.5))+1) \\
&\quad + .5gauss(x, .5, .05)) \\
c1 &= \dot{t} = \dot{w} = 0
\end{aligned}
$$

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 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 c0, 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 topological 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 does not come as a surprise. The topological asymmetry induces a slightly more dispersed distribution of relative compatibilities than in the parapatric case, as seen in figure 1. 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: 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 (3228/1.09×10^6/887) and stays comparable with the third one (4813/1.2×10^6/898). 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.

Conclusion

In this work we set out to validate both our autonomous reproduction scheme and tool for automated phylogeny. 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 ren-

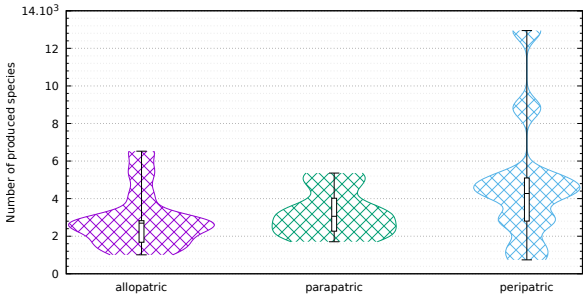


Figure 6: Number of generated species per experiment

dered into either their full or simplified forms, were instrumental in determining whether speciation emerged from the underlying plant-controlled reproductions. Though 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.

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 in this paper. 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, pending further examination of the whole data set. Furthermore, the impact of individual genetic fields was only briefly examined, mostly regarding morphologies.

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, we plan to extend the presented model by using an evolvable substrate (CGP, GRN, ANN) as the basis for the environmental controller. This would 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.

Source code

The C++ code for this project is available at `https://github.com/kgd-al` under the repositories Tools, APOGeT and ReusWorld

Acknowledgments

Computations for this work were performed using HPC resources from CALMIP (Grant P16043).

References

Adami, C. and Brown, C. T. (1994). Evolutionary Learning in the 2D Artificial Life System "Avida". *Artificial Life IV*, 1194:1–5.

Bornhofen, S. (2008). *Emergence de dynamiques évolutionnaires dans une approche multi-agents de plantes virtuelles*. PhD thesis, Paris 11.

Bornhofen, S., Barot, S., and Lattaud, C. (2011). The evolution of CSR life-history strategies in a plant model with explicit physiology and architecture. *Ecological Modelling*, 222(1):1–10.

Butlin, R. K., Galindo, J., and Grahame, J. W. (2008). Sympatric, parapatric or allopatric: the most important way to classify speciation? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1506):2997–3007.

Canino-koning, R., Wiser, M. J., and Ofria, C. (2016). The Evolution of Evolvability : Changing Environments Promote Rapid Adaptation in Digital Organisms. *Proceedings of the European Conference on Artificial Life*, pages 268–275.

Disset, J., Cussat-Blanc, S., and Duthen, Y. (2016). Evolved Development Strategies of Artificial Multicellular Organisms.

Dubois, K., Cussat-Blanc, S., and Duthen, Y. (2017). Towards an Artificial Polytrophic Ecosystem. In *Morphogenetic Engineering Workshop, at the European Conference on Artificial Life (ECAL) 2017 September 4*.

Gardner, M. (1970). Mathematical games: The fantastic combinations of John Conway's new solitaire game "life". *Scientific American*.

Godin-Dubois, K., Cussat-Blanc, S., and Duthen, Y. (2019). Self-sustainability Challenges of Plants Colonization Strategies in Virtual 3D Environments. In Kaufmann Paul, , and Castillo, P. A., editors, *Applications of Evolutionary Computation*, pages 377–392. Springer International Publishing, Cham.

Joachimczak, M. and Wróbel, B. (2008). Evo-devo in silico-a Model of a Gene Network Regulating Multicellular Development in 3D Space with Artificial Physics. *Alife*, pages 297–304.

Metivier, M., Lattaud, C., Heudin, J.-c., and Universitaire, L. (2002). A Stress-based Speciation Model in LifeDrop characters. *Artificial Life*, pages 121–126.

Miconi, T. (2008). In silicon no one can hear you scream: Evolving fighting creatures. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 4971 LNCS, pages 25–36.

Prusinkiewicz, P., Hammel, M., and Mech, R. (1995). The ArtiïňĄcial Life of Plants. *Artificial life for graphics, animation, and virtual reality*, 7:39.

Sims, K. (1994). Evolving 3D Morphology and Behavior by Competition. *Artificial Life*, 1(4):353–372.

Singh, B. N. (2012). Concepts of species and modes of speciation. *Current Science*, 103(7).

Ventrella, J. (2005). GenePool: Exploring the interaction between natural selection and sexual selection. *Artificial Life Models in Software*, pages 81–96.

Woodward, F. I. and Williams, B. G. (1987). Climate and plant distribution at global and local scales. *Vegetatio*, 69(1-3):189–197.