



HAL
open science

A Combinatorial Model of Phyllotaxis Perturbations in *Arabidopsis thaliana*

Yassin Refahi, Etienne Farcot, Yann Guédon, Fabrice Besnard, Teva Vernoux,
Christophe Godin

► **To cite this version:**

Yassin Refahi, Etienne Farcot, Yann Guédon, Fabrice Besnard, Teva Vernoux, et al.. A Combinatorial Model of Phyllotaxis Perturbations in *Arabidopsis thaliana*. 22nd Annual Symposium on Combinatorial Pattern Matching, Palermo, Italy, 2011, Palermo, Italy. pp.323–335. hal-00828852

HAL Id: hal-00828852

<https://inria.hal.science/hal-00828852>

Submitted on 31 May 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Combinatorial Model of Phyllotaxis Perturbations in *Arabidopsis thaliana*

Yassin Refahi^{1*}, Etienne Farcot¹, Yann Guédon¹, Fabrice Besnard²,
Teva Vernoux², and Christophe Godin¹

¹ CIRAD/INRA/INRIA, Virtual Plants INRIA team, UMR AGAP, TA A-108/02,
34398 Montpellier Cedex 5, France

² RDP, ENS/CNRS/INRA/Univ. Lyon, 46, allée d'Italie 69364 LYON cedex 07,
France

Abstract. Phyllotaxis is the geometric arrangement of organs in plants, and is known to be highly regular. However, experimental data (from *Arabidopsis thaliana*) show that this regularity is in fact subject to specific patterns of permutations. In this paper we introduce a model for these patterns, as well as algorithms designed to identify these patterns in noisy experimental data. These algorithms thus incorporate a denoising step which is based on Gaussian-like distributions for circular data for which a common dispersion parameter has been previously estimated. The application of the proposed algorithms allows us to confirm the plausibility of the proposed model, and to characterize the patterns observed in a specific mutant. The algorithms are available in the OpenAlea software platform for plant modelling [10].

1 Introduction

Vascular plants produce new organs at the tip of the stem in a highly organised fashion. This patterning process occurs in small groups of stem cells, the so-called shoot apical meristem (SAM), and generates regular patterns called phyllotaxis [6]. The phyllotaxis of the model plant *Arabidopsis thaliana* follows a spiral, where single organs are initiated successively at an approximately constant divergence angle from the previous organ. The most frequent angle found in nature is the golden angle, close to 137.5° , and leads to the so-called Fibonacci phyllotaxis.

The geometric regularity of this phenomenon has impelled scientists to use mathematical approaches since early studies, two centuries ago. However a complete understanding of the biological processes that drive phyllotaxis is still far from complete. Most models are mechanistic, and allow for an explanation of the occurrence of a limited number of theoretical divergence

* yassin.refahi@inria.fr

angles (including 137.5°), as well as constrained transitions between successive angles in a given plant, see e.g. [1,3]. One leading principle of these models is based on the SAM functioning, where the appearance of new organs – called *primordia* – is supposed to be precluded both in the center of the SAM and in the vicinity of previously formed primordia. This is explained in terms of an inhibitory field surrounding existing primordia. In this paper, we are interested in the variations of angles between consecutive organs in real plants. These angles may be subject to noise and perturbations. Only few studies have been devoted to this problem. For instance, statistical tests have been proposed to distinguish between random and regular phyllotactic patterns, or combinations thereof [4,5]. Perturbations in the phyllotactic patterns have also been observed in a study about transitions between different phyllotactic modes in real plants [2]. It was suggested that these perturbations might result from permutations in the order of appearance of organs along the phyllotactic spiral. In this paper, we build up further on this initial idea. We first considered both reference (wild-type) plants with spiral phyllotaxis (model plant *Arabidopsis thaliana*) and mutant plants that were markedly perturbed in their phyllotaxy. We developed a combinatorial model for the type of perturbations observed in spiral phyllotaxis. Uncertainty is taken into account by assuming that each measured angle can correspond to several theoretical angles among those predicted by the model. Algorithms are proposed to detect such patterns in sequences of angles, and generate all candidate sequences from noisy data. For a given theoretical angle, the corresponding observed angles are modeled by a Gaussian-like distribution for circular data. For each candidate theoretical angle, the posterior probability of the measured angle is computed and compared to a threshold. This allows to reduce the set of candidate sequences.

2 Model Formulation

The exploratory analysis of our measured angles highlighted two characteristics of the measured divergence angle sequences. For a given plant, let α denote the canonical divergence angle:

- The measured divergence angles covered almost all the possible values (between 0 and 360°) with highest frequencies around the canonical Fibonacci angle of 137.5° . At least four classes of divergence angles were apparent but they were not unambiguously separated.

- Short segments (i.e. sub-sequences) of non-canonical divergence angles were identified along measured sequences and were more frequent in the mutant. They seemed to follow constrained patterns, or motifs.

In particular, a motif corresponding approximately to $(2\alpha, -\alpha, 2\alpha)$ was frequently observed in wild-type and even more often in mutants (see Figure 1). This motif, which was already observed in [2], can be simply explained by a permutation of two consecutive organs on the stem, without changing their angular positions. This led us to hypothesize that the segments of non-canonical angles could be explained by permutations involving 2 or even 3 organs (the most realistic numbers given the structure of the SAM). Let us now formulate this idea in more precise terms.

An ideal sequence would simply be a repetition of the canonical angle,

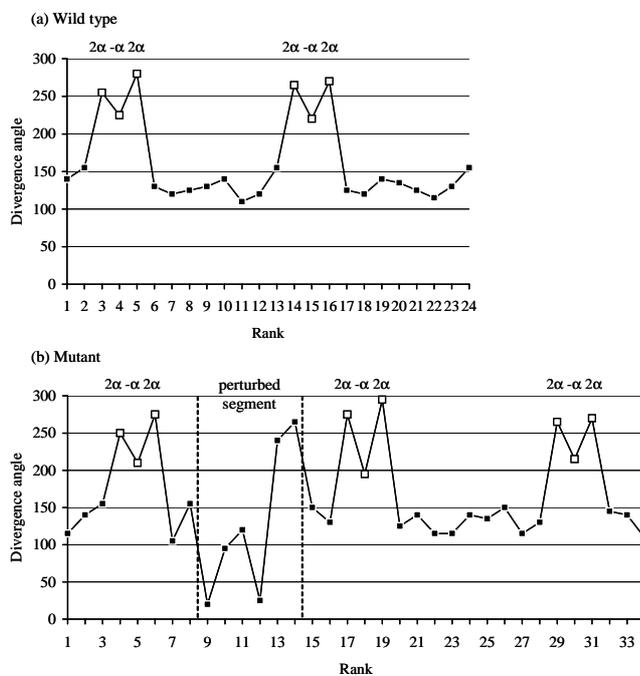


Fig. 1. Identification of M-shaped motifs corresponding to isolated 2-permutations. The perturbed segments cannot be easily explained on the mutant individual.

of the form $(\alpha, \alpha, \dots, \alpha)$. Since we assume that permutations occur, all terms in a sequence $S = (\mu_1, \dots, \mu_\ell)$ of divergence angle will in fact verify

$\mu_j \in \alpha\mathbb{Z}^* = \{i\alpha \mid i \in \mathbb{Z}, i \neq 0\}$. We define the corresponding absolute angles as follows:

$$v_0 = 0, \quad v_i = \sum_{j=1}^i \mu_j, \quad V(S) = (v_0, \dots, v_\ell), \quad i \in \{0, \dots, \ell\}. \quad (1)$$

From V we define a series containing the order of appearance of organs if the first is 0. We name it *order index series* of S , denoted $U(S)$, or simply $U = (u_0, u_1, \dots, u_\ell)$ when S is clear from the context:

$$u_i = \frac{1}{\alpha}(v_i - v_J), \quad 0 \leq i \leq \ell \quad \text{where} \quad J = \arg \min_{j \in \{0 \dots \ell\}} v_j. \quad (2)$$

From the definition it is clear that if $J > 0$ then $v_J < 0$ since $v_0 = 0$. This may occur when the sequence starts with permuted angles, a fact related to the left truncation of observed sequence with respect to complete sequences. If the sequence S follows a spiral phyllotaxis we have

$$u_i = i, \quad \forall i \in \{0, \dots, \ell\}.$$

We call the sequence S n -admissible if a finite number of permutations, applied to disjoint blocks of at most n successive organs, results in an ordered sequence.

Definition 1. A sequence $S = (\mu_1, \dots, \mu_\ell) \in (\alpha\mathbb{Z}^*)^\ell$ is n -admissible, for some $n \in \{1, \dots, \ell\}$, if and only if its associated order index series U satisfies:

$\forall i \in \{1, 2, \dots, \ell\}$, $u_i \neq i \Rightarrow \exists j, k \in \{0, 1, \dots, \ell\}$, $j \leq i \leq k$, $k - j + 1 \leq n$, (u_j, \dots, u_k) is a permutation of (j, \dots, k) , i.e. their underlying sets are equal: $\{u_j, \dots, u_k\} = \{j, \dots, k\}$.

Such a (u_j, \dots, u_k) , of length in $\{1, \dots, n\}$, is called a *shuffled block*.

Property 1. If S is n -admissible, then U is a permutation of $(0, \dots, \ell)$. In general, the converse holds only for a certain $n \in \{2, \dots, \ell\}$.

U is a permutation of $(0, \dots, \ell) \iff \exists n \in \{1, \dots, \ell\}$ s.t. S is n -admissible.

Example 1. The sequence $S = (-\alpha, 2\alpha, 3\alpha, -\alpha, -\alpha, 3\alpha)$ is 3-admissible, but not 2-admissible. Indeed, its absolute angles are $V = (0, -\alpha, \alpha, 4\alpha, 3\alpha, 2\alpha, 5\alpha)$. Hence $v_J = v_1 = -\alpha$, and $U = (1, 0, 2, 5, 4, 3, 6)$. Then, $(u_0, u_1) = (1, 0)$ and $(u_3, u_4, u_5) = (5, 4, 3)$ are shuffled blocks of length at most 3, and suffice to reconstruct the canonical sequence $(0, 1, 2, 3, 4, 5, 6)$.

For n -admissible sequences, the μ_i only belong to a finite subset of $\alpha\mathbb{Z}$:

Property 2. The divergence angles of an n -admissible sequence, take values in:

$$D_n = \{i\alpha \mid (1-n) \leq i \leq (2n-1), i \neq 0\}.$$

Proof. Let μ_i be a divergence angle in an n -admissible sequence, , we know from Eq. (1)-(2) that $\mu_i = (u_i - u_{i-1})\alpha$. In other words, note that, up to the multiplicative constant α , S is the first-order differenced sequence of U . There are four possible cases for u_{i-1} and u_i :

1. Neither u_{i-1} nor u_i are in any shuffled block, so $u_{i-1} = (i-1)$, $u_i = i$ and $\mu_i = \alpha$.
2. u_{i-1} is in a shuffled block but u_i is not in a shuffled block, so $u_i = i$, and $u_{i-1} \in \{(i-n), \dots, (i-2)\}$ then $\mu_i \in \{2\alpha, \dots, n\alpha\}$.
3. u_{i-1} is not in a shuffled block but u_i is in a shuffled block, so $u_{i-1} = (i-1)$ and $u_i \in \{i+1, \dots, (i+n-1)\}$ then $\mu_i \in \{2\alpha, \dots, n\alpha\}$.
4. Both u_{i-1} and u_i are in a shuffled block.
 - u_{i-1} and u_i are in the same shuffled block so $u_i, u_{i-1} \in \{j, \dots, (n+j-1)\}$, for some $j < i-1$. Hence $\mu_i \in \{(1-n)\alpha, \dots, (n-1)\alpha\} \setminus \{0\}$.
 - u_{i-1} and u_i are in two different but chained shuffled blocks so $u_{i-1} \in \{(i-n), \dots, (i-2)\}$, $u_i \in \{i+1, \dots, (i+n-1)\}$ then $\mu_i \in \{3\alpha, \dots, (2n-1)\alpha\}$. \square

In general, the concatenation of two n -admissible sequences is not n -admissible. However this can be true after translating only the first angle of the second sequence. As we show now after two preliminary observations.

Proposition 1. *Let $S = (\mu_1, \dots, \mu_\ell)$ be n -admissible and $V(S)$ and $U(S)$ be the sequence of absolute angles and the order index series respectively. Then $J < n$ and $0 \leq -v_J < n\alpha$, where J is defined as in (1).*

Proof. First we prove that $J < n$. By construction of order index series we know that $u_J = (v_J - v_J)/\alpha = 0$.

If $J = 0$ then $J < n$. We suppose that $J \neq 0$, therefore $u_J \neq J$ and by Definition 1, $\exists j, k \in \{0, 1, \dots, \ell\}$, $j \leq J \leq k$, $k - j + 1 \leq n$, $u_J = 0 \in \{u_j \dots u_k\} = \{j \dots k\}$. Thus $j = 0$, and $k - j + 1 = k + 1 \leq n$. Hence $J \leq k < n$.

Now we prove $0 \leq -v_J < n\alpha$. Since $u_0 = -v_J/\alpha \geq 0$ by definition, this amounts to $0 \leq u_0 < n$, in which only the second part remains to be proved. It holds obviously for $u_0 = 0$. Otherwise, u_0 is part of a shuffled block $\{u_j \dots u_k\} = \{j \dots k\}$, where $j = 0$, and $k < n$, whence $u_0 < n$. \square

The order index series of two concatenated sequences does not always begin with the order index series of the first sequence. However, this is true if the first sequence is long enough.

Proposition 2. *Let $S = (\mu_1, \dots, \mu_i)$, $P = (\mu_{i+1}, \dots, \mu_{i+k})$. If $i \geq n$, then $U(S)$ is subsequence of $U(S \cdot P)$ where $S \cdot P = (\mu_1, \dots, \mu_{i+k})$ denotes the concatenation of S and P .*

Proof. $V(S)$ is subsequence of $V(S \cdot P)$. Therefore $U(S)$ is a subsequence of $U(S \cdot P)$ iff $v_J = v_{J'}$ where J and J' are defined as in Eq. (2) for S and $S \cdot P$ respectively. Since $J < n, J' < n$ from Proposition 1, the minimal element of V and V' appears among their first $n - 1$ elements, which they share if $i \geq n$. \square

Proposition 3. *Let $S = (\mu_1, \dots, \mu_i)$, $i \geq n$, and $P = (\mu_{i+1}, \dots, \mu_{i+k})$. Let $U(S) = (u_0, \dots, u_i)$ and $U(P) = (u'_0, u'_1, \dots, u'_k)$ be the order index series of S and P respectively. Suppose that S is n -admissible. Then, the concatenated sequence $S \cdot P = (\mu_1, \dots, \mu_{i+k})$ is n -admissible iff*

$$P|_{u_i} \doteq \left(\mu_{i+1+(u_i-i)\alpha}, \mu_{i+2}, \dots, \mu_{i+k} \right) \text{ is } n\text{-admissible and } u'_0 = 0. \quad (3)$$

Proof. We use again the identity $\mu_i = \alpha(u_i - u_{i-1})$. Let $U = U(S \cdot P) = (u_0, \dots, u_{i+k})$ denote the order index series $S \cdot P$. Since from Proposition 2 we know that $U(S)$ is a subsequence of $U(S \cdot P)$, we can easily show that $U(S \cdot P) = U(S) \cdot (u'_0 + i, u'_1 + i, \dots, u'_k + i)$. From Property 1 we know that $\{u_0, \dots, u_i\} = \{0, \dots, i\}$. Since S is n -admissible and $u'_0 = 0$ then it is clear that $S \cdot P$ is n -admissible if and only if $(u_{i+1}, \dots, u_{i+k})$ is a permutation of $(i + 1, \dots, i + k)$ that can be decomposed into disjoint shuffled blocks of length $\leq n$, or equivalently for $(u_{i+1} - i, \dots, u_{i+k} - i)$ and $(1, \dots, k)$. In other words, $S \cdot P$ is n -admissible iff $(u_{i+1} - i, \dots, u_{i+k} - i)$ is the order index series of an n -admissible sequence. From the initial remark, the divergence angle sequence leading to this order index series can be written as

$$\begin{aligned} & \alpha \left(u_{i+1} - i, (u_{i+2} - i) - (u_{i+1} - i), \dots, (u_{i+k} - i) - (u_{i+k-1} - i) \right) \\ & = \alpha \left(u_{i+1} - i, u_{i+2} - u_{i+1}, \dots, u_{i+k} - u_{i+k-1} \right), \end{aligned}$$

where the multiplication by α is applied to each component. Then, the same remark again shows that this sequence is exactly $P|_{u_i}$. \square

It will be useful in the last section to scan sequences backwards. One shall then rely on reversibility of the n -admissible property:

Property 3. Let $S = (\mu_1, \dots, \mu_\ell)$ be a sequence of divergence angles. S is n -admissible iff the reversed sequence $S' = (\mu_\ell, \dots, \mu_1)$ is n -admissible.

Proof. Let $U(S) = (u_0, u_1, \dots, u_\ell)$ be the order index series of S , we know that $S = ((u_1 - u_0)\alpha, \dots, (u_\ell - u_{\ell-1})\alpha)$ and $S' = ((u_\ell - u_{\ell-1})\alpha, \dots, (u_1 - u_0)\alpha)$. Moreover, $V(S)$ and $V(S')$ obviously have the same minimum, say v_J . Then,

$$\begin{aligned} U(S') &= (-v_J, (u_\ell - u_{\ell-1}) - v_J, (u_\ell - u_{\ell-2}) - v_J, \dots, (u_\ell - u_0) - v_J) \\ &= u_\ell - v_J - (U(S))', \end{aligned}$$

where $(U(S))' = (u_\ell, \dots, u_0)$ is the reversed order index sequence of S . It is clear that the latter can be decomposed into shuffled blocks of length $\leq n$ iff $U(S)$ itself can. Since $U(S')$ is seen above to be a translation of this reversed sequence it also shares this property. \square

Property 2 defines the theoretical angles that may occur in an n -admissible sequence, but the measured angles are never exactly in D_n , and could correspond to two or more of these theoretical angles. This may lead to several n -admissible sequences. This will later be stored as a suffix tree.

Definition 2. A labelled tree $T = (V, E, L)$, where $L : V \rightarrow D_n$, is called an n -admissible tree if all leaves have a common depth $\ell \in \mathbb{N}$, and every path from the root to a leaf is labelled by an n -admissible sequence.

Let Γ be a mapping that for each measured angle proposes candidate theoretical angles among those in D_n

$$\begin{aligned} \Gamma : [0, 360^\circ) &\longrightarrow 2^{D_n} \\ x_i &\longmapsto C_i = \{\mu_{i1}, \mu_{i2}, \dots, \mu_{ik}\} \subset D_n \end{aligned} \quad (4)$$

We also consider a function $\omega : [0, 360^\circ) \times D_n \rightarrow [0, 1]$ that returns a confidence level $\omega(x_i, \mu_q)$ – typically a probability – for each (x_i, μ_q) .

3 Detecting n -admissibility in Noisy Sequences

3.1 Problems

Given the Γ function above, a set of measured angles will generate a possibly high number of candidate sequences.

Problem 1. Let $x = (x_1, \dots, x_\ell) \in [0, 360^\circ)^\ell$ be measured angles, and $C = \prod_{i=1}^\ell C_i \subset D_n^\ell$ where $C_i = \Gamma(x_i)$. The task is to find all n -admissible $S = (\mu_1, \mu_2, \dots, \mu_\ell)$ in C .

In order to deal with this problem we first need to know whether a given sequence of divergence angles is n -admissible.

Problem 2. Given a sequence S of divergence angles, the task is to determine whether S is n -admissible.

The following is a straightforward observation which is a special case of lemma 10.3 in [8]. It will be used to recognise n -admissible sequences.

Lemma 1. *Let U be a permutation of $\{1, \dots, \ell\}$. Then for all $1 \leq i < j \leq \ell$, $\{u_j, \dots, u_k\} = \{j, \dots, k\}$ iff $\min\{u_j, \dots, u_k\} = j$ and $\max\{u_j, \dots, u_k\} = k$.*

3.2 Algorithms

First we propose an algorithm to solve Problem 2. In order to use Lemma 1, it first checks whether $\{u_0, \dots, u_\ell\} \neq \{0, 1, \dots, \ell\}$ in linear time, using a Parikh mapping [8]. Then, it determines whether an input sequence is n -admissible by a single scan of the sequence from left to right. The time complexity of the algorithm is $O(\ell)$ where ℓ is length of the input sequence.

n-admissible algorithm:

input: $n, S \# S$: a sequence of length ℓ a priori composed of theoretical divergence angles
output: Boolean (true or false)

Begin

if S_0 not in D_n :

 return False # since $(\mu_{i+1} + (u_i - i)\alpha)$ in $P|_{u_i}$ in proposition 3 could be not in D_n

Construct the order index series $U(S)$

if $\{u_0, \dots, u_\ell\} \neq \{0, 1, \dots, \ell\}$: return False # using Parikh mapping

$i:=0$

while $i \leq \ell$:

 if $u_i \neq i$: lo:= u_i ; up:= u_i ; j:= $i+1$

 while true:

 if $j > \ell$: return false

 if $j - i > n - 1$: return false

 lo:= $\min(u_j, lo)$; up:= $\max(u_j, up)$

 if $(lo = i) \ \& \ (up = j)$: $i:=j+1$; break # (u_i, \dots, u_j) shuffled block

 j:= $j+1$

 else: $i:=i+1$

return true # the order index series $U(S)$ can also be returned if needed

End

Remark 1. The notion of shuffled block can be seen as a special case of interval [8], Ch. 10. However, because it is much more specific, existing interval extraction algorithms would return invalid subsequences, whence the need for a new algorithm as above.

Now we can deal with Problem 1. A naive algorithm would construct all candidate sequences $S = \mu_1, \mu_2, \dots, \mu_\ell$, and then apply the **n-admissible algorithm**. The number of candidate sequences equals $\prod_{i=1}^{\ell} |C_i|$, where the $|C_i|$ is the cardinality of C_i . Since C_i are typically not singletons, $|C|$ increases exponentially with ℓ . Therefore we propose a lookahead algorithm to explore the search space avoiding non necessary paths, as sketched below. The source code is available for more details [10].

n-admissible tree algorithm:
input: $\Gamma, n, (x_1, \dots, x_\ell)$ # a sequence of measured divergence angles of length ℓ
output: **n-admissible tree**
with nodes labelled by both divergence angles μ_i and order indices u_i
Begin
 $T := \{\text{root}\}$; $\mu(\text{root}) := 0$; $u(\text{root}) := 0$
find:= false
while True:
 nLeaves:= *nonterminal_leaves*(T) # leaves of depth $< \ell$
 if nLeaves is empty: return T
 for leaf in nLeaves:
 $d := \text{depth}(\text{leaf})$; $m := \min(n, \ell - d)$;
 for $k \in \{1, \dots, m\}$:
 for $P \in \Gamma(x_{d+1}) \times \dots \times \Gamma(x_{d+k})$:
 Compute $P|_{u(\text{leaf})}$ # cf. (3), Proposition 3
 if **n-admissible**($P|_{u(\text{leaf})}$): # also returns μ and u for nodes in P
 append all nodes on P to leaf
End

Thanks to the use of Proposition 3, *n*-admissibility can be tested on subsequences only. The time complexity of the *n*-admissible tree algorithm increases exponentially with the lookahead limit *n*. When the returned tree contains only one *n*-admissible sequence (as was generally the case in practice), it is more precisely $O(\ell \times (k)^n)$ where $k = \max(|C_i|)$.

Proposition 4. *Let us call $\mathcal{A}_n(C)$ the set of n-admissible sequences in C, and $\pi(T)$ the set of all (labels of) paths in T, from the root to the leaves. Then $\pi(T) = \mathcal{A}_n(C)$, i.e. the tree built in the n-admissible tree algorithm contains exactly the n-admissible sequences in C.*

Proof. The inclusion $\pi(T) \subset \mathcal{A}_n(C)$ is clear, since in the 2nd for loop only paths which are *n*-admissible can be added, thanks to Proposition 3. To see that the converse holds, it suffices to remark that given an *n*-admissible sequence (μ_1, \dots, μ_ℓ) , one of the *n* subsequences

$$(\mu_1, \dots, \mu_{\ell-1}), \quad (\mu_1, \dots, \mu_{\ell-2}) \quad \dots \quad (\mu_1, \dots, \mu_{\ell-n})$$

must be n -admissible as well, as follows from the definition. Because all these subsequences are tested in the `for` loop, there cannot be an n -admissible sequence that is not detected by the algorithm, and thus $\pi(T) \supset \mathcal{A}_n(C)$. \square

Further Pruning.

In the case where $\mathcal{A}_n(C)$ is large, one may use the weights $\omega(x_i, \mu_j)$ to sort the sequences according to a confidence level. Actually, to a given path with labels μ_1, \dots, μ_ℓ in the computed n -admissible tree, one may naturally assign the weight $\prod_j \omega(x_j, \mu_j)$. Then, all paths in the tree can be ordered according to their weight.

These weights can also be used to limit the size of the constructed tree, by ruling out all candidate paths whose weight is below a certain threshold. Since the weight of each node is lower than 1, the weight of a path can only be lower than that of any of its subpaths. Hence, it is possible in the `for` loop of the algorithm to prune not only the non admissible paths, but also those having a weight below a threshold. This is how we have actually implemented the algorithm, using posterior probabilities for weights, and adding a threshold as an input to the algorithm, as explained in the next section.

4 Results

4.1 Assignment of Measured Angles to Theoretical Angles

We have used the proposed algorithms to analyse the sequences of measured angles. The Γ function was parametrised using a statistical model. In a first step, a hidden Markov chain was estimated on the basis of the pooled wild-type + mutant measured divergence angle sequences in order to estimate an angle measurement uncertainty parameter; see more details in [9]. In this hidden Markov chain, the states of the non-observable Markov chain represents “theoretical” divergence angles while the von Mises observation distributions attached to each state of the non-observable Markov chain represents measurement uncertainty. The von Mises distribution [7], also known as the circular Gaussian distribution, is a univariate Gaussian-like periodic distribution for a variable $x \in [0, 360^\circ)$. Let $g(x; \mu_q, \kappa)$ denote the probability density function of the von Mises distribution, with parameters μ_q (mean direction) and κ (concentration parameter). The main output of this first step of analysis was the estimated common concentration parameter (inverse variance) κ . This parameter corresponds to a standard deviation of approximately $\sigma = 18^\circ$ (for our

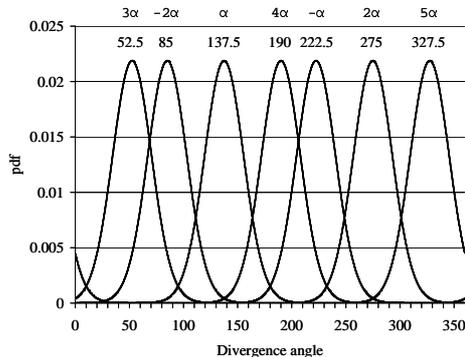


Fig. 2. The estimated von Mises distributions used for the Γ mapping.

set of measured angles). Using this standard deviation, the Γ mapping (4) for our data is defined as follows:

$$\Gamma(x_i) = \{\mu_q \in D_n \mid \mu_q - \rho\sigma \leq x_i \leq \mu_q + \rho\sigma\}$$

The intervals defined by parameters ρ and σ correspond typically to a cumulative probability of 0.9975 with respect to the angle distribution centered at μ_q . For each possible theoretical angle of index q , the posterior probability

$$\omega(x_i, \gamma) = \frac{g(x_i; \mu_q, \kappa)}{\sum_{\mu_r \in D_n} g(x_i; \mu_r, \kappa)}$$

was calculated and compared with a predefined threshold (typical value 0.05) to decide whether this angle should be kept or rejected for the labelling of the sequence in the `n-admissible tree` algorithm. The implicit underlying hypothesis was that the theoretical angles were *a priori* equally probable.

4.2 Interpretation on Our Dataset

We applied the modeling approach to our data set, see Figure 3 for an example where the predicted divergence angle sequence is in close agreement with the measured divergence angle sequence. For some sequences however, there was no *n*-admissible sequence in C as defined in Problem 1. This was often due to either some error in the measurement of divergence angles, or to too large deviations between an observed angle and any predicted angle corresponding to a valid prediction. A single measured angle

x_i is non-explained, if no theoretical angle in $\Gamma(x_i)$ leads to a non-empty output of the `n-admissible tree` algorithm. Due to the dependencies in-

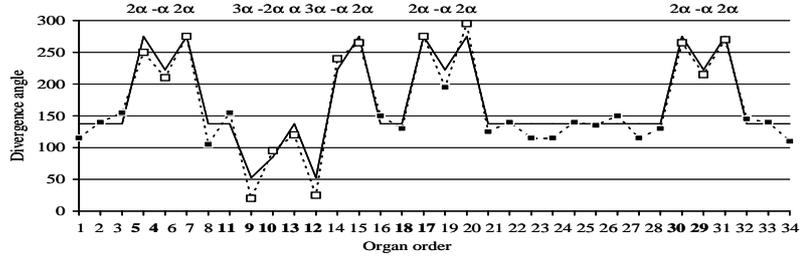


Fig. 3. Mutant individual: prediction of the of divergence angle sequence (continuous line) and labelling of the divergence angles within the permuted segments.

duced by the permutation patterns, the angle at which the algorithm fails may in fact result from an isolated error, earlier in the sequence. Hence, all shuffled blocks preceding a non-explained angle should be marked as not valid. To achieve this goal, we define splitting points. The notion of splitting point can be viewed as a deterministic analogue of a regeneration point for a stochastic process. A regeneration point is a time instant at which the future of the process depends only of its state at that instant and is thus independent of its past before that instant. The process is thus reborn at a regeneration point.

Definition 3. Let S be a sequence of divergence angles and $U(S)$ be the order index series of S , u_i is a splitting point iff $u_i = i$ and u_i is not in a shuffled block.

Using the notion of splitting point, we implemented a procedure which was applied after the `n-admissible tree` algorithm. It consisted in a backtracking starting at the non-explained angle and progressing towards a splitting point. This allowed us to automatically invalidate blocks of angles preceding a value at which the algorithm failed.

To refine this analysis, we used reversibility (Property 3), and applied the whole procedure to both measured sequences and their reverse. Then, the intersection of angles invalidated on a sequence and its reverse, was often reduced to a single angle. Moreover these angles were likely due to measurement errors, typically the omission of one angle in the series, leading to an isolated 2α in a sequence of canonical angles α , see Figure 4. An

expert investigation of these automatically detected subsequences enabled us to find with increased accuracy those angles which were not explained by our model. The proposed modeling approach allows to explain a very

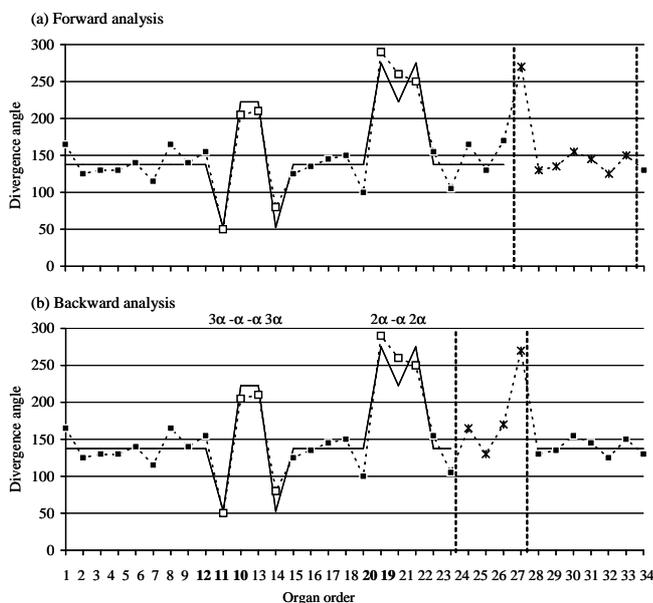


Fig. 4. Analysis of the sequence in both directions to detect segments that are invalid with respect to the permutation assumption. The invalid segments are delimited by dashed lines. The continuous line corresponds to the predicted divergence angles.

large proportion of the non-canonical divergence angles despite the relatively high standard deviation (approx. 18°) of the estimated von Mises distributions. This indicates that the proposed model correctly describes the phyllotactic patterns of *Arabidopsis thaliana*. Wild-type plants were characterized by relatively frequent occurrences of 2-permutations generally isolated while mutants were characterized by the frequent occurrences of both 2- and 3-permutations whose succession generates highly complex motifs, see Figure 3 for an example. A summary of the results is shown in Table 1, with more precise counts of patterns in Table 2.

The term Lucas phyllotaxis refers to a spiral phyllotaxis where the canonical divergence angle α is 99.5° . Although rarer than the Fibonacci spiral ($\alpha = 137.5^\circ$), it is known to occur in nature, and was able to explain two wild-type and two mutant sequences, for which 137.5° failed.

	Wild-type plant	Mutant
# sequences/# organs	82/2405	89/ 2815
% of non-canonical angles	15%	37%
% of unexplained angles	2%	5%
# individuals, Lucas phyllotaxis	2	2
# 2-permutations	123	297
# 3-permutations	3	53

Table 1. Summary of the permutation patterns observed in both wild type and mutant plants.

	organ order	divergence angles	wild-type	mutant
2-permutation	2 1 3	2 -1 2	90	193
	3 2 1 4	3 -1 -1 3	1	11
3-permutation	3 1 2 4	3 -2 1 2	1	9
	2 3 1 4	2 1 -2 3		13
total			2	33
2 2-permutations	2 1 4 3 5	2 -1 3 -1 2	16	32

Table 2. Permuted segments up to length 5. These segments are delimited by two splitting points. The divergence angle sequence is the first-order differenced organ sequence. By convention, the origin of the organ sequence is 0 (not indicated).

References

1. Adler I., Barabé D., Jean R.V., *A History of the Study of Phyllotaxis*, Annals of Botany, 80 (3):231-244 (1997).
2. Couder Y., *Initial transitions, order and disorder in phyllotactic patterns: the ontogeny of Helianthus annuus. A case study*, Acta Societatis Botanicorum Poloniae, 67:129:150 (1998).
3. Douady S, Couder Y., *Phyllotaxis as a self organizing process: I,II,III*, Journal of Theoretical Biology 178:255-312 (1996) .
4. Jeune B., Barabé J., *Statistical Recognition of Random and Regular Phyllotactic Patterns*, Annals of Botany, 94:913-917 (2004).
5. Jeune B., Barabé J., *A stochastic approach to phyllotactic pattern analysis*, Journal of Theoretical Biology 238:52-59 (2006).
6. Kuhlemeier C., *Phyllotaxis*, Trends in Plant Science, 12:143-150 (2007).
7. Mardia, K. V. and Jupp, P. E. *Directional Statistics*. John Wiley & Sons: Chichester, 2000.
8. Parida L., *Pattern Discovery in Bioinformatics: Theory & Algorithms*, Chapman & Hall/CRC Mathematical Computational Biology series, 2008.
9. Refahi, Y., Guédon, Y., Besnard, F., Farcot, E., Godin, C., Vernoux, T., *Analyzing perturbations in phyllotaxis of Arabidopsis thaliana*. In: 6th International Workshop on Functional-Structural Plant Models, Davis, California, 170-172 (2010).
10. http://openalea.gforge.inria.fr/doc/vplants/phyllotaxis_analysis/doc/_build/html/contents.html