Quantification of variability in trichome patterns

Bettina Greese¹, Martin Hülskamp² and Christian Fleck³*

¹ Computational Biology and Biological Physics, Faculty for Theoretical Physics and Astronomy, Lund University, Lund, Sweden
² Molecular Cell Biology and Developmental Genetics, Biocenter, Botanical Institute, Cologne University, Cologne, Germany
³ Laboratory for Systems and Synthetic Biology, Wageningen University, Wageningen, Netherlands

While pattern formation is studied in various areas of biology, little is known about the noise leading to variations between individual realizations of the pattern. One prominent example for de novo pattern formation in plants is the patterning of trichomes on Arabidopsis leaves, which involves genetic regulation and cell-to-cell communication. These processes are potentially variable due to, e.g., the abundance of cell components or environmental conditions. To elevate the understanding of regulatory processes underlying the pattern formation it is crucial to quantitatively analyze the variability in naturally occurring patterns. Here, we review recent approaches toward characterization of noise on trichome initiation. We present methods for the quantification of spatial patterns, which are the basis for data-driven mathematical modeling and enable the analysis of noise from different sources. Besides the insight gained on trichome formation, the examination of observed trichome patterns also shows that highly regulated biological processes can be substantially affected by variability.

Keywords: noise, cell-to-cell variability, pattern formation, spatial data analysis, plant development, trichome patterning

1. INTRODUCTION

Mathematical modeling has been used to study various biological patterning processes, such as trichomes and root hairs (Savage et al., 2008; Benítez et al., 2011), cell sizes in sepals (Roeder et al., 2010), hair follicles (Sick et al., 2006), fruit fly development (Reeves et al., 2006), and other systems (Othmer et al., 2009; Peltier and Schaffer, 2010). It has only recently become more popular to investigate the variance or variability within a system and to discuss the consequences of noise (see Box 1) (Kern et al., 2005; Swain and Longtin, 2006; Maheshri and O’Shea, 2007; Wilkinson, 2009; Sánchez et al., 2013). Moreover, an evaluation of the robustness (see Box 1) of a patterning system requires a quantification of the variations in its inputs and outputs (Reeves et al., 2006). Some studies have been published that focus on models with a stochastic (see Box 1) component, e.g., the stochastic Boolean network (see Box 1) model for root hairs (Savage et al., 2008) or floral morphogenesis (Alvarez-Buylla et al., 2008) or noise in the initiation of new organs in phyllotaxis (Mirabet et al., 2012). Others examine the effect of noise on patterning using stochastic differential equations (see Box 1) (Sagués et al., 2007). However, although a rich tradition exists in studying the effect of noise on pattern formation using abstract sets of equations, only few studies from developmental biology can be found where the effect of intracellular noise and/or cell-to-cell variability on a developing pattern or structure was systematically taken into account (Little et al., 2013). While advances in data acquisition and experimental manipulations increase the feasibility and popularity of noise-related studies in single cell organisms (Paldi, 2003; Kern et al., 2005; Swain and Longtin, 2006; Sánchez et al., 2013), quantitative comparisons of spatial patterns and testable predictions from mathematical models are needed in order to assess the influence of various types of noise on a developing organism (Lander, 2011). In particular, it is desirable not only to qualitatively study simulation results that arise from various perturbations of the model, but also to quantitively compare these with experimentally observed patterns. As far as we are aware, the latter aspect has rarely been studied so far. It is important to note that the existence of cell-to-cell variability is not necessarily an outcome of stochasticity, but may be due to deterministic (see Box 1) regulatory processes upstream of the observed process (Snijder and Pelkmans, 2011). Whatever the source of the variability is, the pattern will be affected by it. In many studies, reaction-diffusion systems (see Box 1) are used to describe the pattern formation process (Gierer and Meinhardt, 1972; Meinhardt and Gierer, 1974; Koch and Meinhardt, 1994). These models require some stochasticity in the initial values to start the patterning. It is thought that this initial variability among cells in a tissue stems from a spontaneous fluctuation of the abundance of the proteins involved in the process. However, apart from this, variability is neglected and the equations themselves are deterministic. To explicitly study noise in patterning, it is necessary to not only consider stochastic initial conditions but also to include some other type of stochasticity such as spatially or temporally varying parameters (Page et al., 2005; Woolley et al., 2011).

In plants, the question whether the spatial distribution is random or ordered was first investigated for developing stomata (Sachs, 1974; Rasmussen, 1986; Croxdale, 2000). Stomata patterns are well suited for investigation because the patterns are two-dimensional and occur on the organ surface, which makes them readily accessible. Stomata are an example of a biological
realization of a planar point pattern (see Box 1) (Larkin et al., 1997; Torii, 2012). Another prominent example for such a pattern from the plant kingdom are epidermal hairs, called trichomes, for which the regularity of the patterning process has been studied (Greese et al., 2012). The spatial distribution of trichomes is regulated by a genetic network and involves cell-to-cell communication. Trichome formation is promoted by three proteins that form an activating protein complex which can be inhibited by a fourth protein (Hülskamp, 2004; Digiuni et al., 2008). Because the inhibitor is mobile (i.e., non-cell autonomous), it effectively coordinates the patterning process between cells. In order to enable data-driven modeling for pattern formation, it is necessary to derive and evaluate models based on experimental data. To this end, statistical methods are needed which are suitable for the available type and amount of data and the studied system.

### 2. QUANTITATIVE CHARACTERIZATION OF NOISY POINT PATTERNS

The quantification of spatial variability is tightly related to the determination of the degree of regularity in a specific pattern. In other words, to be able to describe any kind of variability between two patterns, a suitable method to describe each pattern by itself is needed. Understanding the geometrical properties of a biological pattern helps to explore its functional role and its development (Galli-Resta et al., 1999). Moreover, appropriate statistical measures will be needed to analyze the effect of system perturbations (e.g., mutations), which will help together with mathematical models to elucidate the mechanistic role of the different components in a regulatory network. In the following, we first outline how planar point patterns arising in biology have been analyzed and then focus on the methods applied to trichome patterns.

One statistical method frequently used to assess a point pattern is the mean neighbor distance, which is compared to the mean neighbor distance of a completely random distribution (Clark and Evans, 1954). This was applied to stomata distribution where typically an ordered distribution was found (Miskin and Rasmussen, 1970; Croxdale, 2000). Because the next neighbor method is simple, it is easy to apply. However, all detailed information or spatial aspects of the pattern are lost. Therefore, more advanced methods were applied to analyze the stomata pattern (Martins et al., 2011). Trichome patterns have been examined through tests for deviation from randomness using quadrant counts (see Box 1) and their ratio of variance to mean (Smith and Watt, 1986), through analogous tests using nearest neighbor distances (Larkin et al., 1996), and through classification of mutant patterns based on cluster frequency and distance between trichomes (Schnittger et al., 1999). A variety of more advanced methods have been discussed and compared for spatial point patterns in general, e.g., Boots (1986); Legendre and Fortin (1989);
Different indices of dispersion that are based on distances and counts in quadrats have been used to compare plant patterns (Pielou, 1960; Goodall and West, 1979). Measures based on various graphs such as the Voronoi diagram and the minimal spanning tree have been used to analyze biological point patterns (Dussert et al., 1987; Wallet and Dussert, 1997) and geographical settlement patterns (Boots, 1986), and different structure indices as well as correlation functions have been applied to quantify forest structures (Pommerening, 2002). An extensive discussion of methods to analyze the spatial structure of ecological populations, illustrated on vegetation data, is given by Legendre and Fortin (1989), which includes correlograms, spectral analysis, periodograms, variograms, clustering, mapping, and testing for autocorrelation. Alternative approaches that characterize the geometry and topology of spatial patterns are Minkowski functionals, which have been applied to chemical reaction-diffusion systems (Mecke, 1996) and galaxy clusters (Kerscher et al., 1997).

A popular method to analyze stochastic patterns is the structure function, which is the Fourier transform of the density correlation function, the spatial pendant of the power spectrum (Torquato, 2002). Peaks in the structure function denote regularity of the stochastic pattern. However, many of the well-known methods used in physics and astronomy require large domains and/or large sample sizes, which make them less suitable for the data typically available for biological systems. For instance, the sample size needed to obtain smooth statistics for the structure function grows exponentially with the noise in the pattern (Bastian, 2013). An estimate reveals that for the noise present in trichome patterns in a given sample size of the order of $10^4$ would be necessary (Bastian, 2013). Because one leaf with its trichome distribution represents one realization of the noisy patterning process, this is beyond the currently available amount of data. In general, the variety of strategies applied to investigate spatial structure illustrates that the choice of method is not straightforward and depends on the data and the question to be examined.

A suitable characterization of spatial trichome patterns is built on a tessellation (see Box 1) of the trichome positions, which splits the domain of the leaf into polygons that do not overlap or intersect, i.e., together they exactly cover the domain. A commonly used tessellation is the Voronoi diagram (Okabe et al., 2000), in which each point is assigned a polygon that contains that part of the domain that is closer to its defining point than to any other point (Figure 1A, left). Hence, the Voronoi diagram can be interpreted as an assignment of an influence area around each trichome that results from the inhibitory signal. Notably, the inverse of its area can be taken as a local density at its defining point (Duyckaerts et al., 1994). When pairs of trichomes whose Voronoi polygons share a common edge are connected by a straight line, the result is a Delaunay triangulation of the leaf (Figure 1A, left). The agglomerate of all triangles involving a selected trichome is called the contiguous Voronoi polygon, and it can also be used to calculate local density (Schaap and van de Weygaert, 2000). Various modifications have been proposed to adapt Delaunay triangulations to specific biological systems, resulting in different neighborhood graphs (Jaromczyk and Toussaint, 1992; Raymond et al., 1993). Pairs of neighbors can be defined by the edges present in the modified triangulation, such that each trichome is assigned a set of (mostly six) neighbors (Figure 1A, right).

A good visual impression of the order and symmetry inherent in a given point pattern can be obtained from a spatial autocorrelogram (Galli-Resta et al., 1999; Raven and Reese, 2002). This graphical representation of a given set of points is constructed by superimposing one copy of the pattern per point whereby the point is placed in the origin of the coordinate system (Figure 1B). For increasingly noisy patterns, the autocorrelogram becomes less distinct because the correlation is lost (first long-range, then short-range). If the region around the origin is devoid of points, i.e., an exclusion zone exists (Galli-Resta et al., 1999; Raven and Reese, 2002), the pattern exhibits a minimal distance between points, which can be seen as the simplest possible type of order (Larkin et al., 1996). The autocorrelogram can be used to extract the density recovery profile (Galli-Resta et al., 1999; Raven and Reese, 2002), which is an approximation of the autocorrelation function. In general, the central part of the autocorrelogram is most important for its interpretation, which allows a truncation of the plot to a chosen radius (compare Raven and Reese, 2002). The spatial autocorrelogram can be further adapted for the local analysis of trichome patterns and to avoid various artifacts (Figures 1C–E) (Greese, 2011). This modified autocorrelogram shows the distribution of neighbor distances and angles and hence gives a first impression about the regularity in a given point pattern. A more rigorous quantification that also allows easy comparison of different patterns requires the derivation of appropriate mathematical functions.

Measures suitable for the characterization of the regularity of the trichome pattern are local measures based on the neighborhood of each trichome. For each individual trichome, the distance to its neighbors, the angle between pairs of adjacent neighbors, and the anisotropy of the neighbors’ distribution can be measured (Greese et al., 2012). The local anisotropy can be described using the eigenvalues of the inertia tensor (Goldstein et al., 2002). Their inverse values correspond to the length of the principal axes of an ellipse (Figure 1A, right), such that their ratio determines the deviation from isotropy (i.e., the case where the ellipse is a circle). For all measures it is advantageous to use the variation coefficient, i.e., the ratio of the standard deviation to the mean, in order to obtain measures independent of scale or density.

Other studies have used various related tessellation-based measures to characterize spatial point patterns (Boots, 1986; Marcellopol and Usson, 1992; Duyckaerts et al., 1994; Croxdale, 2000; Schaap and van de Weygaert, 2000; Chiu, 2003), mostly focusing on the area of the Voronoi polygons or Delaunay triangles. Depending on whether one wants to detect differences in point density or measure spatial arrangement independent of point density, different measurements are appropriate. In order to estimate the overall amount of noise present in the observed
Greese et al. Quantification of variability in trichome patterns

FIGURE 1 | Quantification of trichome patterns. (A) Neighborhood measures for trichomes (black dots) on a single leaf. The left panel shows the Voronoi diagram (light red lines) and the modified Delaunay triangulation (light green lines) as well as the Voronoi polygon (dark red line) and the contiguous Voronoi polygon (dark green line) for a selected trichome. The right panel shows a magnification of one trichome (in the center) with its six neighbors and the neighbor distances and angles (blue lines and arcs). The anisotropy is related to the ratio of the principal axes of the ellipse (red lines). Reproduced with permission from Greese et al. (2012) ©The Institution of Engineering and Technology. (B) Construction of an autocorrelogram for a simple pattern containing three points. Three copies of the original pattern are superimposed such that each time one point lies in the origin of the coordinate system. (C) Truncated autocorrelogram for a data set with real trichome data. (D) Additionally rotated autocorrelogram. (E) Further reduced autocorrelogram where the mean and the standard deviation of the neighbor distances and angles are highlighted.

trichome pattern, one can compare the values of the neighborhood measures obtained from experiments with the corresponding values for hexagonal point patterns with increasing noise level. Figures 2A–C show hexagon patterns for an increasing amount of irregularity (see text box for details). Aggregating the differences between the experimentally obtained values and the values for a noisy hexagonal point pattern into an objective function allows for estimation of the noise level which best reflects the noise in trichome patterning (Figure 2D, see Shapiro et al., 1985; Kinney et al., 2001 for similar calibration methods). This approach shows that trichomes show about 44% noise in relation to hexagonal patterns (see text box Figure 2 or Greese et al., 2012 for details), which is considerable for a tightly regulated patterning system. What does this mean for the patterning process? As it appears the patterning mechanism is important, as contrasted to a purely random process, to achieve a more or less homogeneous trichome density. The exact spatial distribution seems to be of less importance.

3. SOURCES OF NOISE IN TRICHOME PATTERNING

The trichome pattern is noisy, somewhat midway between a regular and a random pattern (Greese et al., 2012). But what are the sources of this observed irregularity? A clear distinction between different sources of noise (e.g., molecular processes, environmental conditions) is a challenge in any experimental or modeling study (Swain and Longtin, 2006). Because cells in a tissue will slightly vary in their protein content at a given time point (Kim and Price, 2010; Snijder and Pelkmans, 2011; Jeschke et al., 2013; Little et al., 2013), one first step is to investigate the effect of spatially varying initial states or reaction rates on the simulated trichome pattern (Page et al., 2005; Greese et al., 2012).

The trichome initiation process resembles an activator-inhibitor system with an immobile activator (Gierer and Meinhardt, 1972; Meinhardt and Gierer, 1974; Koch and Meinhardt, 1994). If both, the activator and the inhibitor, are mobile, the resulting pattern depends only weakly on the initial conditions (Maini et al., 1997; Page et al., 2005). In a fast
initial phase the early activator peaks are formed. These are usually not very pronounced. On a longer time-scale the activator peaks align and grow. Biologically, only the peaks at the later stage lead to an observable result, unless it becomes feasible to track the protein content in single cells in a tissue. The mobility of the activator allows the activator peaks to move slightly for optimal alignment (Holloway and Harrison, 1995; Ward and Wei, 2002). However, in the singular limit of vanishing activator mobility the optimal alignment of the peaks is impaired, and noise from the initial conditions remains in the final pattern. This can be seen in Figures 3A,B, where we show examples of simulation results for increasing mobility of the activator (see Figure 3, text box for further explanation). In Figure 3C the local irregularity of the simulated trichome pattern is plotted against the mobility of the activator (which is a complex consisting of GL1 and GL3 in case of the simulated trichome system). The pattern becomes more irregular with decreasing activator mobility, which is a known effect in reaction-diffusion systems (Holloway and Harrison, 1995). In other words, the cell autonomy of the activator in trichome patterning restricts the degree of regularity (see Greese et al., 2012 for details).

In general, if the parameters of a pattern formation network vary slightly from cell-to-cell, the resulting pattern will have a lower degree of regularity. How much the spatial variability of a single parameter affects the pattern depends on the details of the reaction network. Figure 3D shows as examples the dependence of the local irregularity on the activation of GLABRA3 (one of the patterning proteins) by the activator (solid lines), degradation of the activator (dashed lines), synthesis rate of GLABRA3 (dashed-dotted lines) (see Greese et al., 2012 for details). One interesting aspect of the effect of cell-to-cell variability on patterning is that protein binding rates strongly vary under conditions of macromolecular crowding (Minton, 2005, 2006; Grima, 2010). Because the abundance of molecules will, in general, slightly differ from cell to cell, the macromolecular crowding will vary and as a consequence also the protein-binding rates and in turn the resulting pattern will be less regular (Greese et al., 2012). The specific effects of crowding on gene expression have also been addressed in pure modeling and simulation studies, which were able to separate the effects related to binding and diffusion rates (Morelli et al., 2011) and to show the dependency on transcript levels on the volume fraction (Matsuda et al., 2014).
4. PERSPECTIVE

To analyze spatial patterns in case of small sample sizes and few repetitions, it is useful to focus on methods that are sufficiently local and sensitive to subtle differences. Local measures which quantify the regularity of the local environment of each trichome as defined by tessellations can successfully be applied to compare experimental observation and results from computer simulations (Greese et al., 2012).

Any data analysis task poses several challenges, such as obtaining enough meaningful data, selecting appropriate methods, and linking observations to causes. The comparison of spatial patterns has often been done by simply looking at them and deciding whether they agree or not (sometimes referred to as “eyeballing”), which can be problematic as two realizations that look alike do not necessarily have to arise from the same mechanism. The use of simplistic measures may not be helpful as two very different patterns can lead to the same measured value, e.g., when the nearest neighbor distance is compared. Hence, it is important to carefully select the method(s) for analysis and make sure that the system under study can be adequately described and different situations can be discriminated. Regarding the analysis of noise, determining its overall magnitude is only the first step, the next—more interesting and challenging—part is then to trace it to various aspects of the system, in other words, find the sources of noise.

Variability is generally present in any natural process and hence introduces an additional level of complexity in the investigation of the process. A biological process may be strongly influenced or almost unchanged by noise, depending on the specific kind of variation and the properties of the particular system. Hence, the variability should—whenever possible—be treated as part of the process and not merely seen as a nuisance to be avoided. With the current trend of supplementing qualitative descriptions by quantitative measurements, it becomes not only feasible to estimate parameters for models but also necessary to compare results in more detail. In addition to a sound analysis of experimental data, it is very instructive to generate predictions with the help of a mathematical model because this approach allows the evaluation of different assumptions (including unrealistic situations like a noise-free system) and the separation of tightly coupled effects (like crowding on different reaction rates). All these efforts together will advance the understanding of the biological process under study.

ACKNOWLEDGMENTS

Bettina Greese was supported by the Knut and Alice Wallenberg Foundation under the project grant “ShapeSystems.” Christian Fleck received support from the Human Frontier Science Program (HFSP Research Grant RGP00025/2013).
