

# Morphometrics of complex cell shapes: Lobe Contribution Elliptic Fourier Analysis (LOCO-EFA)

Yara E. Sánchez-Corrales<sup>1,3</sup>, Matthew Hartley<sup>1</sup>, Jop van Rooij<sup>1,2</sup>,  
Athanasius F. M. Marée<sup>1,4</sup>, Verônica A. Grieneisen<sup>1,4</sup>

29th June 2017

<sup>1</sup>Computational and Systems Biology, John Innes Centre, Norwich NR4 7UH, UK.

<sup>2</sup>Theoretical Biology/Bioinformatics, Dept. of Biology, Utrecht University,  
Padualaan 8, 3584 CH Utrecht, The Netherlands.

<sup>3</sup>Current address: MRC-Laboratory of Molecular Biology, Cambridge Biomedical  
Campus, Francis Crick Avenue, Cambridge CB2 0QH, UK

<sup>4</sup>To whom correspondence should be addressed.

E-mail: Stan.Maree@jic.ac.uk, Veronica.Grieneisen@jic.ac.uk

Short title: Morphometrics of complex cell shapes

Keywords: Cell shape | Cellular Potts Model | Image analysis | Pavement cells |  
*Arabidopsis thaliana* | *Drosophila* |

## Abstract

Quantifying cell morphology is fundamental to the statistical study of cell populations, and can help us unravel mechanisms underlying cell and tissue morphogenesis. Current methods, however, require extensive human intervention, are highly sensitive to parameter choice, or produce metrics that are difficult to interpret biologically. We therefore developed a novel method, Lobe Contribution Elliptical Fourier Analysis (LOCO-EFA), which generates from digitalised cell outlines meaningful descriptors that can be directly matched to morphological features. We show that LOCO-EFA provides a tool to phenotype efficiently and objectively populations of cells by applying it to the complex shaped pavement cells of *Arabidopsis thaliana* wild type and *speechless* leaves. To further validate our method, we analysed computer-generated tissues, where cell shape can be specified in a controlled manner. LOCO-EFA quantifies deviations between the specified shape that an individual *in silico* cell takes up when in isolation and the resultant shape when they are allowed to interact within a confluent tissue, thereby assessing the role of cell-cell interactions on population cell shape distributions.

## Summary statement

Novel method (LOCO-EFA) quantifies complex cell shapes, extracting meaningful biological features such as protrusion number and amplitude; here shown for plant pavement cells and validated on *in silico* tissues.

## Introduction

Cell geometry has long fascinated biologists (Thompson, 1917). This interest is driven by a wide range of underlying scientific questions. For instance, cell shape changes can be linked to physiological responses of cells, such as membrane protrusions during apoptosis and migration (Charras and Paluch, 2008); they are involved in controlling tissue homeostasis (Marinari et al., 2012; Veeman and Smith, 2013) and underlie cell behaviour such as chemotaxis to ensure survival (Keren et al., 2008; Driscoll et al., 2012). The cell shape itself also influences intracellular processes such as microtubule organisation (Gomez et al., 2016) and the stress patterns in plant epithelia (Sampathkumar et al., 2014); it indirectly positions the plane of cell division (Minc et al., 2011) and can even determine the way a flower attracts pollinators (Noda et al., 1994). Given the rich diversity of processes in which cell shape plays a decisive role, either actively or passively, it is not surprising that the qualitative and quantitative study of cell shape characteristics, cell morphometrics, has become important in developmental biology, enriching molecular and genetic techniques. In parallel, recent advances in imaging technology, including microscopy, usage of fluorescent proteins and imaging software, allow us to collect and access remarkable amounts of cell morphological data, which in turn urgently calls for analytic tools for the extraction of meaningful cell shape information (Zhong et al., 2012). In stark contrast to the technological advances on the imaging front, there are relatively few automatic and quantitative tools available to analyse complex cell shapes (Rajaram et al., 2012; Ljosa et al., 2012; Ivakov and Persson, 2013). This gap reflects the non-trivial nature of this task: cell shape is often irregular and variable, making it very difficult to establish universal criteria encompassing cell geometry.

To illustrate the issues involved in capturing quantitative aspects of complex cell shapes, consider the case of pavement cells in the plant epidermis (Figure 1A, B) or amnioserosa cells in the *Drosophila* embryo (Figure 1C). Pavement cells present a striking development, requiring multiple locally divergent growth fronts within each cell, which are coordinated amongst neighbouring cells; amnioserosa cells present dynamical cell shape fluctuations within a confluent tissue. The resultant jigsaw-puzzle-like features of such cells illustrate the challenges in quantitatively characterising cell shape: 1) their complex non-holomorphic geometry cannot be captured in a meaningful way with traditional shape metrics (see below); and 2) the lack of recognisable landmarks (neither these plant nor animal cells have distinct features to serve as reference points), renders a myriad of shape statistics that have been developed in other fields (such as principal component and Procrustes analysis) effectively useless (Klingenberg, 2010).

Traditional metrics for cell morphology include cell area, cell perimeter, aspect ratio and form factor. Useful as general descriptors, the shape information that can be extracted from them is limited. Very different shapes may yield a similar aspect ratio or form factor (Figure 1D–H). Besides not being unique, such descriptors tend to omit information regarding biologically relevant shape features (Fu et al., 2005). Several approaches classically used to quantify cells, including pavement cells, are summarised in Table 1. Some of those methods, such as the skeleton method, are highly sensitive to image noise as well as to the precise choice of parameters (for an example, see Le et al., 2006). Other metrics, such as lobe length and neck width assessments (Fu et al., 2005), are entirely dependent on humans to judge what a lobe is. Such decisions then strongly impact the average lobe length and neck width measurements (Figure 1 and Figure S1). It renders these metrics highly variable from cell to cell, from phenotype to phenotype and from human to human. In this respect, recently an automatic method was developed to count lobes and indentations (Wu et al., 2016), making this process more reliable. As we will discuss further on, such a method, although already very useful, finds its limitations when the characteristics of the shape reside in the distribution and amplitude of lobes, not only in their number. For instance, some *Arabidopsis* mutants can present pavement cells with lobes that look more elongated or shallower, but which occur at a similar spatial frequency (Lin et al., 2013).

Promising alternatives are methods that take into account the full cell outline information and then simplify the contour of a cell into a series of coefficients that can be employed as shape descriptors in a multivariate study (Ivakov and Persson, 2013; Pincus and Theriot, 2007). One example of these methods is Elliptical Fourier Analysis (EFA), that has been used to quantify two-dimensional complex shapes (Kuhl and Giardina, 1982; Diaz et al., 1989; Schmittbuhl et al., 2003). It takes the coordinates of the outline and decomposes them into a series of related ellipses (described by EFA coefficients) that can be combined to reconstitute that given shape. Despite its wide usage in morphometric studies, EFA cannot retrieve information that directly relates to morphological features of a cell, obstructing biological interpretation. This is because the EFA coefficients are highly dependent on how the cell outline is approximated (a detailed explanation is given in the Supplementary Materials and Methods).

Here we present a new method based on EFA, termed Lobe-Contribution Elliptic Fourier Analysis (LOCO-EFA), that overcomes the common obstacles described above. Our method also uses the information of the whole cell contour but, unlike EFA, provides a set of metrics that directly relate to morphological features, permitting the assessment of cell shape complexity in an objective and automatic manner. Importantly, it is not

sensitive to cell orientation or imaging resolution.

To validate the usage of our method on larger cellular datasets, we analyse pavement cell populations from confocal images of *Arabidopsis thaliana*. We then complement this study with the analysis of synthetic tissues generated using the Cellular Potts model (Graner and Glazier, 1992; Glazier and Graner, 1993), in which pavement-like cells have a parametrised specified shape. The latter approach also allows us to touch on a fundamental question in developmental biology, by asking to what degree the resultant cell shape within a confluent tissue context is shaped by cell-to-cell interactions and to what degree it can be explained by intracellular shape control mechanisms. Specifically, here we explore the influence of cell neighbours on the morphogenesis of individual cells by measuring the divergence of their specified cell shape within an isolated context to the shape taken up when being immersed within a tissue. This analysis, performed on *in silico* tissues, allows us to quantify such differences for a generic situation by combining CPM and LOCO-EFA .

## Results

### Quantitative characterisation of cell shape using LOCO-EFA

Applying EFA to quantify cell shapes, we came across a number of specific shortcomings. We first explain those issues to highlight our motivation and choices that led to the development of LOCO-EFA. Detailed mathematical implementation is provided in the Supplementary Materials and Methods. Here, we focus on explaining the analysis in terms of its biological relevance, emphasising how it can be applied and interpreted.

The shape analysis proposed here is linked to harmonic compositions of digitalised cell shape outlines. We find it useful to compare the decomposition of a complex cell shape, such as that of a pavement cell, to the way one decomposes the sounds of musical instruments. This analogy will be used throughout the manuscript for purposes of illustrating the different traits that our morphometric analysis is capturing. For example, when listening to a note generated by an instrument, a quantifiable observable is the pitch. Within the context of pavement cell shape this corresponds to the observed number of lobes or, as we will explain in detail, to the dominant spatial frequency of the pavement cell outline. Another quantifiable property of musical note is its volume, the amplitude. In terms of cell shape this corresponds to the extent to which lobes and indentations are protruded and retracted. Finally, the timbre of musical instruments is what essentially distinguishes, for example, a clarinet from an oboe, as both play the same note (same pitch) at the same volume. An analogous notion in cell shape studies would be the ability to capture additional aspects of cell shape morphology, as well as to recognise subtle differences between cells, even when the number of lobes and their level of protrusion is the same.

As a starting point, Elliptical Fourier Analysis (EFA, Kuhl and Giardina, 1982) can be used to describe the contour of any complex two-dimensional shape, including non-holomorphic shapes such as pavement cells, which most other methods are unable to capture (see Figure S2 and Supplementary Materials and Methods). Using the coordinates of the two-dimensional (cell) outline (Figure S3A), EFA decomposes the shape into an infinite series of ellipses (also referred to as “modes” or “harmonics”). This series of ellipses,  $n = 1 \dots \infty$ , can then be combined in the following way to exactly retrieve the original shape: each  $n$ th elliptic harmonic traces  $n$  revolutions around the first ellipse

while orbiting around the previous ( $n - 1$ ) harmonic ellipse, which in its turn is orbiting around its previous one ( $n - 2$ ), and so forth (Figure S3B). This summation results in an outline being ‘drawn’, as is visualised in Movie S1. Obviously, a cut-off has to be chosen for the number of modes that are taken into account. In general this will be at a value for which the cell contour description is reasonably close to the original cell outline. The fact that each ellipse represents a certain harmonic suggests that it captures dominant spatial frequencies within the original shape. This is why it has been considered to be a reasonable descriptor for cell outline properties (Schmittbuhl et al., 2003).

When shapes are quantified using the EFA method, however, the pitch, i.e., the most basic cellular feature to quantify, is actually not directly retrieved, not even for the most simple shapes. For instance, one might expect a six-sided shape to present a strong contribution from the 6th mode. Instead, the EFA method represents such a shape as a mixed contribution from the two adjacent modes, the 5th and 7th mode (Figure 2D and Figure S3C). This mismatch is a consequence of how the individual ellipses obtained from decomposing the outline contribute to the outline. Each ellipse does so by rotating either clockwise or counterclockwise. The direction of this rotation in respect to the direction of rotation of the first mode causes either an increase or a decrease in the actual number of features being drawn, always one off from the actual mode (Figure S3D, E, Movie S2 and Movie S3). As a consequence, the ‘pitch’ obtained using EFA does not correspond to actual cell features, which hinders interpreting the resultant values of this method. Moreover, EFA coefficients on their own are redundant, i.e., there are more parameters than needed to specify the same specific shape (Haines and Crampton, 2000). Therefore, comparison of cell shapes on the basis of their EFA coefficients (for example, by means of principle component analysis) becomes nonsensical. Together, these traits make the EFA method unsuitable for reliable quantification of cell morphology and renders prohibitive any meaningful comparisons between multiple cell outlines.

Diaz et al. (1990) proposed a heuristic solution for the mismatch between actual shape features and EFA’s dominant harmonics, using the fact that the relative direction of rotation is a main determinant of the generated pitch (see Supplementary Materials and Methods for details). It turns out, however, that each ellipse is simultaneously contributing to two different spatial frequencies, something this heuristic can not solve (see Figure S3F and Movie S4). As a consequence, although their method is often (but not always) able to correctly recapitulate the ‘pitch’, it is never able to correctly capture the amplitude or timbre of the cell shape. Further details regarding the range of issues when using EFA are discussed in depth in the Supplementary Materials and Methods.

To overcome these limitations we essentially propose a new basis for the outline reconstruction, which we have coined  $\mathcal{L}_n$ , after lobe number. In a similar manner as EFA, the modes can be summed up to recreate the original shape and each individual mode is captured by a set of four unique parameters. However, there are two important distinctions. Firstly, any possible cell outline is now only represented by one unique combination of  $\mathcal{L}_n$  coefficients. In the Supplementary Materials and Methods we discuss in detail the steps required to eliminate all levels of redundancy for shape analysis. Secondly, cell shape features such as the protrusion number (the “pitch”), their amplitude (the “volume”) and the characteristic lobe distributions (“timbre”) are directly represented by the  $\mathcal{L}_n$  coefficients. The  $\mathcal{L}_n$  coefficients are obtained after decomposing each EFA harmonic into its exact, specific contributions to two separate  $\mathcal{L}_n$  modes (Figure 2A–C). In general, the EFA modes  $n - 1$  and  $n + 1$  both partly contribute to mode  $\mathcal{L}_n$ , although some specific exceptions apply (see Figure S4). The resulting method, which we coin Lobe Contribution-EFA or



LOCO-EFA, thus consists of: eliminating multiple representations of a given outline; decomposing each  $n$ th elliptical EFA mode into two separate lobe contributions; and finally integrating those separate modes into single LOCO-EFA modes. Every  $\mathcal{L}_n$  mode can then be regarded to represent two oppositely rotating circles, each with its own starting point for the rotation, the four  $\mathcal{L}_n$  coefficients providing the radii and starting angles of both circles. We next assign a scalar  $L_n$  value, to specifically capture the amplitude contribution (“volume”) of each mode. We found that in order to correctly capture this, both the radii of and the angular distance between the starting points of the two contributing circles, as well as the starting point of the main circle,  $\mathcal{L}_1$ , have to be taken into account (see Figure 2C, and Supplementary Materials and Methods for further details regarding its derivation). Calculating the  $L_n$  values results in a spectrum representing the relative contribution of each individual mode to the cell shape (Figure S3C). For instance, the spectrum of the six-lobed test shape used for Figure S3C indeed contains a pronounced peak at mode six, as well as a peak at mode one, which represents the overall circular shape of that outline. The original shape can be reconstructed using all modes up to a specific mode number. This allows us to visually appreciate the specific contribution of mode six to the six-lobed test shape (compare Figure 2E with Figure 2D).

To further illustrate how LOCO-EFA characterises shapes, we apply it to simple geometrical shapes with variable numbers of protrusions (Figure 3A–I). Indeed, LOCO-EFA robustly determines the main  $L_n$ -mode of each shape corresponding to the correct number of lobes (Figure 3J). Each  $L_n$  profile provides information about the composition of all morphological periodicities contributing to the shape, reflecting the geometry being considered. We next tested if LOCO-EFA also correctly captures the “volume”, by applying the method to generated shapes of the same “pitch” (i.e., having a same number of lobes), but with variable amplitudes (Figure 3N–Q). Indeed, the magnitude of the  $L_n$  components changes accordingly (Figure 3T), and their absolute value quantitatively corresponds to the size of the extensions. Thus, LOCO-EFA not only retrieves the main number of morphological features of hypothetical cell shapes, but also a range of additional characteristics.

Following the analogy of sound decomposition, a more nuanced perception for quantification is timbre. Timbre can be understood as residing in the additional amplitude spectrum. It is determined by which overtones are emphasised in relation to one another. For cell shape studies, we consider “timbre” analysis the ability to capture additional aspects of the cell shape complexity, besides the main number of protrusions/lobes which a cell presents. This additional information should enable distinction between different cellular phenotypes, such as between wild type and mutants (Lin et al., 2013). To illustrate, Figure 3R, S show two additional six-sided shapes which differ in “timbre” from Figure 3Q, with their accompanying  $L_n$  values (Figure 3T). For both shapes, there is a clear peak at  $L_6$ , reflecting their six-lobedness, and additional peaks or overtones, such as  $L_2$ , capturing the elongated nature of these shapes, and so forth. Thus, LOCO-EFA retrieves not only the main number of morphological features of an hypothetical cell (i.e., lobes), but also important fine-grained characteristics.

From the full characterisation of cell shape by  $\mathcal{L}_n$  modes, additional objective metrics can be derived to help quantify different aspects of “cell shape complexity”. We here define four metrics: XOR difference; marginal difference; cumulative difference; and entropy.

Firstly, complementary to the magnitude of each individual  $L_n$  mode, the cell shape complexity can be estimated using the information of the approximation to the original shape by the first  $N$  LOCO-EFA modes only. It addresses the question how relevant

each subsequent  $L_n$  mode is for explaining that specific shape. Figure 2E already illustrated the importance of a specific mode for reconstructing the original shape (in that case, mode six). One can straightforwardly quantify the relative contribution of that mode for explaining the shape by calculating the total difference (in area, expressed as either number of grid points or  $\mu\text{m}^2$ ) between the original shape and the superimposed reconstructed shape when the first  $N$  LOCO-EFA modes are used. To do so, we take the *XOR* (exclusive or) between the original and reconstructed cells (see Figure S5). In this context, a more “complex” shape is one in which more LOCO-EFA modes are needed to obtain a good match between the reconstruction and the real shape (i.e., when the *XOR* approaches zero only at higher  $N$  numbers). For this criterion, note that a circular cell can be perfectly reconstituted using only the contribution of the first LOCO-EFA mode ( $N = 1$ ), and is therefore the least complex shape with an XOR value of zero for the whole spectrum. On the other hand, cells presenting a high number of heterogeneous lobes require a high number of modes for *XOR* to approach zero (Figure 3). This can be observed in Figure 3K, U, which show the *XOR* difference profiles for the series of test shapes.

This manner of defining cellular complexity can be further compressed by taking the total area under the *XOR* difference curve, a scalar quantity which we call the cumulative difference (cd). It derives a single value from the XOR spectrum, with higher values for more complex shaped cells. The more similar the cell shape is to a circle (which can be perfectly described using the first LOCO-EFA mode only), the more the cd value approximates zero. Figure 3M, W shows the cd values for the series of test shapes. From these, it becomes clear that cd becomes high for shapes that strongly deviate from circular, it is therefore not just correlated with having many lobes. In general, however, when morphological protrusions increase in number or become larger in amplitude, the cumulative difference also increases since higher order modes are required to approximate the shape.

*XOR* profiles typically do not change smoothly. Instead, some modes peak as they strongly contribute in capturing the main features of the shape. Hence, the marginal decrease in the *XOR* value when an extra mode is added, coined *marginal difference*, provides additional information about the shape’s dominant modes (Figure 3L, V). This measure is comparable to the  $L_n$  values, also determining something akin to “pitch” and “amplitude”. We found, however, that it bears a higher discriminatory power for more complex and irregular cell shapes. Moreover, in the case that a cell shape has significant contributions from a combination of different modes, then high marginal difference levels can be directly linked to specific cellular features (see Figure S5). Thus, the marginal difference can help identifying which modes are the most relevant for specific shape aspects.

Finally, one can argue that shape complexity is not solely about high numbers and large amplitudes of the protrusions, but instead reflects overall irregularity of the protrusions. For example, with the previous measures a highly regular star-shaped cell with five outspoken lobes could be quantified as being as complex as a highly distorted cell with different amplitudes and distributions of five lobes, albeit less pronounced than the star-shaped case. One might therefore prefer to define cell shape complexity as the tendency of a cell to deviate from geometrically well-defined periodic outlines. A useful measure for this alternative definition of “cell shape complexity” is calculating the Shannon entropy  $E$  of the  $L_n$  spectrum. The *entropy* measure is based upon the information content contained within the distribution over the whole  $L_n$  profile (see Equation 47 in the Sup-

plementary Materials and Methods). For many shapes, the entropy measure yields very similar results to the cumulative difference. They diverge for cell outlines having a strong contribution from the lower modes. We observe that in these cases the entropy values deliver more meaningful results in regard to what one considers being more “complex”. This is due to the fact that lower modes can have a large contribution to the cumulative difference. For example, there can be a high contribution from  $L_2$  in the case that a cell is very elongated. Simply being elongated, however, does not so much represent shape complexity in the way defined above. For such a simple elongated shape, the cumulative difference (i.e., the integral or area under the curve of the *XOR* between the original and reconstructed profile) can be very similar to a cell with contributions distributed among many modes. We would consider the latter outline, however, to be much more ‘complex’. The entropy measure is able to correctly capture this form of complexity.

Next, we measured the  $L_n$  profiles and corresponding cell shape complexity measures for a population of real cell shapes, which are much less geometric than the idealised cases treated so far.

## LOCO-EFA applied to plant pavement cells

To validate our method, and its derived measures introduced above, we analysed *Arabidopsis thaliana* leaf epidermal pavement cells, individual cells that were tracked *in vivo* over time as well as whole populations of cells at a single time point. Actual biological cells, such as pavement cells, can be highly asymmetrical, with their  $L_n$  landscape characterised by multiple peaks (Figure 3S, T). This is because the outline of an asymmetrical cell with a given number of protrusions placed quasi-periodically along its edge can be interpreted as multiple protrusion frequencies superimposed. In general, the total number of hand-counted lobes in a cell image matches to a peak at the corresponding  $L_n$  value (but note here that hand-counting is very subjective). For instance, for nine lobes a peak will be observed at  $L_9$ . However, if the lobes are distributed in a more or less pentagonal clustered fashion, this would lead to an additional peak at  $L_5$ , whilst superimposed on a triangular shaped cell basis a  $L_3$  contribution would be found, and so forth.

Pavement cells acquire their characteristic jigsaw puzzle-like shape through multipolar growth patterns, such that relative simple shaped cells become highly complex over time (Figure 4A–G). Notably, the smooth shape changes over time are clearly reflected in the  $L_n$  profiles over time (Figure 4I). Its initial squarish shape and later nine- and thirteen-lobeness are well captured by the LOCO-EFA method, through peaks at mode  $L_4$ ,  $L_9$ , and  $L_{13}$ , and corresponding peaks in the marginal difference profile. In contrast, when the EFA method is used (Kuhl and Giardina, 1982), the third and fifth mode are erroneously indicated to represent shape features, as well as other mismatches (Figure 4H). Importantly, the smooth transition in cell shape development over time is captured by a corresponding smooth change in the LOCO-EFA  $L_n$  profile for the different time points, but by highly irregular changes in the EFA profile. This illustrates that very comparable shaped cells have very different EFA profiles, making EFA further unsuitable to analyse real pavement cell populations. Figure S6 presents the robust time dynamics analysis for yet another pavement cell using LOCO-EFA.

To visualise the shape characteristics of populations of pavement cells, we applied the analysis to leaves of the *speechless* mutant (MacAlister et al., 2007), which does not generate during the leaf development any other cell types such as meristemoids or stomata (Figure 1B, Figure 5A), as well as to wild type leaf epidermis, consisting of pavement cells,



stomata and other cells from the stomatal lineage (Figure 1A, Figure 5B).

Using LOCO-EFA, it is straightforwardly possible to dissect the precise contribution of each mode for each cell in the population. Figure 5A, B show the spatial distribution of cells within a tissue which are predominantly 4-, 5-, 6-, or 7-lobed, by colour-coding the cells according to those specific  $L_n$  values. Very few cell shapes present only a single high  $L_n$  value. Instead the majority of shapes are composed of significant contributions stemming from multiple modes. Consequently, simply manually counting, or applying automatic counting algorithms, for such biological cells the number of lobes would lead to incomplete information regarding their shape. It would be difficult to compare mutant phenotypes. Moreover, our data shows pavement cells do not have a population-wide preferential  $L_n$  with a highest contribution to the shape (Figure 5A, B).

The heterogeneity in modes that composes real populations of pavement cells suggests that their resultant cell shapes cannot be explained solely by currently proposed intracellular molecular mechanisms underlying lobe and indentation patterning. This is because those mechanisms, based on the existence of two counteracting pathways (one for lobe formation and another for indentation formation, see details in Xu et al., 2010) should give rise to Turing-like instabilities which are therefore expected to generate symmetrical shapes (Vanag and Epstein, 2009). Moreover, these models would predict that equally sized cells exhibit equal lobe numbers. However, the cell shape patterning takes place within a confluent tissue, which obviously complicates how individual cells generate their shape. In the experimental setting it is very hard to distinguish between the preferred shape of each individual cell due to its intracellular patterning, and the constraints imposed by the tissue. It is well-known that if cells prefer to be round, they will take up an hexagonal shape within a tissue context (Thompson, 1917), but it is unclear what to expect for multi-lobed shapes. Therefore, to both better understand what should be our expectation when a population of cells with complex shape preferences form a confluent tissue alike the leaf, and to further validate the LOCO-EFA method on populations of cells, we performed cell-based simulations of interacting cells with prespecified cell shape preferences, and employed LOCO-EFA on the resulting *in silico* tissue.

## Applying LOCO-EFA to *in silico* populations and the effect of interactions between preferred cell shapes

We create *in silico* cells using the Cellular Potts Model (CPM), an energy-based framework that allows us to represent cells and their dynamics through small membrane extensions and retractions (see Methods section). In its basic form, CPM cell shapes emerge due to the interaction between interfacial tension (the strength of which can depend on the cell types involved), internal cellular pressure and cortical tension (Magno et al., 2015). Here we used an extension of the CPM which predefines intrinsic forces causing elongation and lobeness, resulting in more complex cell shapes. This extension consists in applying additional, cell-specific forces each time a cell extension or retraction is considered, giving rise to elongated and/or multilobed preferred cell shapes (Equation 3). Three additional forces are used that capture (i) an intrinsic tendency to elongate; (ii) a tendency to form a specified number of lobes; and (iii) an additional force for the cell to round up (Figure 5E–G, van Rooij et al. (2017a) and Movie S5). The latter term robustly prevents cells from falling apart, especially within a confluent tissue with many conflicting preferred cell shapes. A population of cells, all with a same preferred shape, were then allowed to interact with each other within a tissue context. In this way we can compare

the shape of a single cell in isolation with the shape cells attain within a tissue.

We here present the analysis on two distinct specified shapes (Figure 5C, D; see Table S1 for the specific parameters used). Both preferred shapes have six lobes, the cells shown in Figure 5D moreover tend to be elongated. Above the panels we plot the shape that the cells attain in isolation. Although the same cell shape is specified for all cells within the population, cells acquire very different shapes due to their local interactions. We quantified this diversity using LOCO-EFA. This was done by colour-coding the contribution of  $L_4$ – $L_7$  (as indicated for each panel), for both the isolated cells and for the resultant shapes of all the cells within the simulated tissue. As expected, the isolated cells present a very high contribution by  $L_6$ , with marginal contributions from the other modes. In contrast, due to the cellular interactions within the tissue, other modes can also become prominent, highly varying from cell to cell, even though each cell has exactly the same specified shape. Thus, although a cell in isolation would be able to generate regular protrusions with specific amplitudes, the periodical lobe formation becomes inhibited in a packed tissue environment, with symmetry and shape distortions being a direct consequence of tissue packing (Figure 5C, D). Such a tendency was observed for all simulations performed, irrespective of the specified cell shape, i.e., the number of lobes, their amplitude, and the level of overall cell elongation, as well as over a wide range of parameter values for the cell-cell interactions (Figure S7, Figure S8). Radially, well-spaced periodically lobed cell shapes are not likely space filling, hence resolving the competition between preferred shape and confluency seems to be a universal driving force to complex cell shapes.

## Discussion

The progress in microscopy and imaging techniques within the biological sciences generates a need for adequate analytic tools to capture relevant information efficiently and objectively (Zhong et al., 2012). Image acquisition through high-throughput microscopy approaches generates amounts of data that are beyond the human ability (or patience) to be analysed manually, demanding computational automatic tools. We have developed a new analytic tool which takes as the input the contour of a 2D cell projection, and extracts from it, in an efficient and parameter-independent manner, quantitative meaningful cell shape information. Importantly, the pipeline can be integrated within other recent advances in segmentation techniques (Fernandez et al., 2010; van Rooij et al., 2017b), to fully automate shape analysis of series of images.

Our method can intuitively be understood using the analogy of music perception. To quantify an instrument playing a certain note, say a violin playing the note A, one would firstly wish to have a device that can capture which note is played. We have indeed shown that LOCO-EFA, unlike EFA, is able to correctly determine the analogous feature for shapes, which is the number of protrusions.

A recent method developed by Wu et al. (2016) can also be employed to count the number of protrusions. When the biological question asked requires not only the “pitch” to be measured, but also the “volume” and “timbre”, corresponding to lobe amplitude and other irregularities, such a method will be insufficient. Indeed, LOCO-EFA provides a broader spectrum of the shape properties, a holistic set of measurements that allows complex morphologies to be quantified in a reproducible manner.

We illustrated how the measurements obtained with LOCO-EFA can be interpreted,

first using simple shapes (geometrical or symmetrical forms), then extending the analysis to highly asymmetric shapes, for which the pavement cells of Arabidopsis are a good paradigm. To assess the performance of our technique, we extended our study to confocal microscopy images of populations of pavement cells. Such a shape analysis is biologically relevant, because many of the players accounting for the lobe and indentation patterning are known (Xu et al., 2010; Grieneisen et al., 2017), enabling one to extend the study of cell shape control to mutants and experimental interferences in the future. We found that very few cells have a symmetrical shape, i.e., they can not be represented well by a single high  $L_n$  value. Such composition in real cell shapes of several  $L_n$  values is not likely to be explained only by the existence of two counteracting pathways specifying lobe and indentation identities. Our *in silico* approach instead suggests that the dynamics of space-filling complex shapes can dramatically increase the overall irregularity: even when the CPM cells are specifically programmed to develop well-defined characteristic shapes, the interactions between them trigger the formation of highly deviating cell shape characteristics. As a result of being within the tissue, the main, specified mode decreases in strength while other modes become relevant. Altogether it leads to very asymmetrical resultant shapes, even though very symmetrical shapes were specified.

Although our synthetic data is but a phenomenological reconstruction of real shapes, our results suggest that the local influence of neighbours during pavement cell development should be important for the acquisition of their final shape. To assess this hypothesis in further studies it will be crucial to perform quantitative shape analysis on *in vivo* cell populations over time, combined with growth tensor analysis (i.e., determination of the spatial pattern in growth rate and its anisotropy). Such studies, combined with genetic or physical perturbations in cell growth and deformation, and with *in silico* cell growth models, should allow untangling of the relationship between the cell shape as specified at the cellular level and the resultant shape arising from the complex interactions between the cells at the tissue level.

When LOCO-EFA was applied to cell tracking data, we observed that the profiles of those dynamically changing cells varied smoothly over time. Such trajectories are distinct from cell-to-cell and provide unique fingerprints of each individual developing cell. This opens the possibility of using the  $L_n$  modes as cell identifiers within a temporal sequence of images, to help track populations of cells automatically.

Although we have here applied LOCO-EFA to objectively measure cell shape differences in pavement cells, providing us a powerful tool to evaluate pavement cell shape dynamics and compare plant phenotypes, it can also be applied to different cell types with complex cell shapes of any species, as well as to any other layouts, such as entire leaves, within or outside the realm of biology. Recently, Fourier analysis was applied to quantify myosin polarity (asymmetrical distribution of molecules) in animal epithelia (Tetley et al., 2016). In a similar way, the LOCO-EFA method could be applied to quantify the cellular distribution of molecules within complex cell shapes. Finally, our method could be integrated within recent image analysis pipelines, allowing one to extract and analyse cell shape information in a high throughput manner (van Rooij et al., 2017b; Heller et al., 2016; Stegmaier et al., 2016).

# Material and Methods

## Confocal images and image processing

Columbia wild type or *speechless* mutant (MacAlister et al., 2007) leaves expressing pmCherry-Aquaporin (Nelson et al., 2007) were imaged using a confocal microscope Leica SP5 at comparable stages and in comparable regions. Further image processing was done using ImageJ and images were segmented using in-house software (Segmentation Potts Model (van Rooij et al., 2017b)). Cells changing over time were imaged using a custom-made perfusion chamber (Robinson et al., 2011; Sauret-Güeto et al., 2012; Kuchen et al., 2012).

## Shape descriptors

Average lobe lengths and neck widths were calculated using ImageJ (Analyse  $\rightarrow$  Measure). The skeleton was calculated using Better Skeletonization by Nicholas Howe, available through MATLAB File Exchange.

## Geometric shapes

All geometric shapes were generated by the “superformula” described in Gielis (2003), and were analysed in the same manner as the confocal images.

## XOR and other measurements

All the grid points belonging to each individual real or synthetic pavement cell were compared to all the grid points captured by the subsequent series of LOCO-EFA reconstructions. A reconstruction of level  $N$  takes into account the first  $N$   $\mathcal{L}_n$  modes. The *in silico* cells were generated using the Cellular Potts Model, which is a grid-based formalism, while for the experimental data the grid points were directly defined by the imaging resolution. The scripts used to calculate the XOR and to colour-code the real and synthetic cells were written in the coding language C. Shape approximations, cumulative difference and entropy were calculated using the first 50  $L_n$  modes. To capture cell shape complexity linked to protrusions rather than mere anisotropy, cumulative difference is calculated from the second  $L_n$  mode onwards. This value turned out to be more than sufficient to capture any cell shape given the grid point resolution used for all cases here. Note, however, that very high resolution images might require additional modes to fully capture the shape.

## Cellular Potts Model generating complex cell shapes

The Cellular Potts Model (CPM) is an energy-based model formalism used to model cellular dynamics in terms of Cell Surface Mechanics (Magno et al., 2015). Individual cells are described by a set of grid points on a lattice. In this manuscript we used the CPM to generate *in silico* cells with relatively complex shape preferences that are allowed to interact within a confluent setting. During each simulation step a grid point is chosen in a random fashion to evaluate whether its state changes into one of its neighbouring states, effectively corresponding to a small cell shape modification at that point. To evaluate if such state change will occur, the energy change is calculated that such a copy would

cause. This is done by calculating the change in the configurational energy as defined by the following Hamiltonian, which sums up the energy contribution of each pixel within the entire field as well as of all cells:

$$\mathcal{H} = \sum_{ij} \sum_{i'j'} J \left( 1 - \delta_{c_{i,j}, c_{i',j'}} \right) + \sum_c \lambda_a (a_c - A)^2 + \lambda_p \sum_c (p_c - P)^2, \quad (1)$$

$J$  refers to the coupling energy, summed over all grid points  $(i, j)$  and their eight (2nd order) neighbours  $(i', j')$ . The Kronecker delta term  $\left( 1 - \delta_{c_{i,j}, c_{i',j'}} \right)$  simply assures that neighbouring lattice sites of the same state (i.e., belonging to the same cell) do not contribute to the total energy of the system. The variables  $a_c$  and  $p_c$  denote respectively the actual cell area and the actual cell perimeter for each cell ( $c$ ); the parameters  $A$  and  $P$  denote the target cell area and perimeter. The parameters  $\lambda_a$  and  $\lambda_p$  describe the resistance to deviate from the target area and perimeter, respectively. The probability a copying event is accepted depends on the change in the Hamiltonian,  $\Delta\mathcal{H} = \mathcal{H}_{after} - \mathcal{H}_{before}$ , in the following way:

$$P = \begin{cases} 1 & \text{if } \Delta\mathcal{H} < -Y, \\ e^{(-\frac{\Delta\mathcal{H}+Y}{T})} & \text{if } \Delta\mathcal{H} \geq -Y, \end{cases} \quad (2)$$

where  $Y$  corresponds to the yield or ability of a membrane to resist a force and  $T$  (simulation temperature) captures additional stochastic fluctuations. Copying events which decrease  $\mathcal{H}$  by at least  $Y$  will always be accepted, otherwise acceptance follows a Boltzmann probability distribution (Equation 2).

To generate cells with a particular number of preferred protrusions, we modify the change in the Hamiltonian as is calculated for every evaluated copying event, effectively shortcutting intracellular biochemistry and biophysics, in the following way. Simulated cells are attributed with a specified preferred number of lobes, amplitude of lobes, overall elongation and roundness, implemented by modifying the change in the Hamiltonian for every evaluated copy event as follows (van Rooij et al., 2017a):

$$\Delta\mathcal{H}' = \Delta\mathcal{H} - \nu \cos(n\theta) - \chi \cos(2\alpha) - \mu(\sqrt{A/\pi} - r). \quad (3)$$

Those three additional terms are evaluated for both cells involved in the copying event, so there are effectively six additional terms. The first term captures the tendency to form  $n$  lobes, with  $\nu$  capturing the propensity to extend to form a lobe or to retract to form an indentation, thus giving rise to the amplitude or pointedness of the lobes.  $\theta$  describes the angle between any of the  $n$  equally spread out target directions for outgrowth and the vector determined by the coordinates of the grid point under evaluation and the centre-of-mass of the cell (hereafter called the copy vector) (Figure 5E). To clarify, when a cell extension is considered right on top of one of the target directions, then  $n\theta = 0$ ,  $\cos(n\theta) = 1$ , and tendency to extend is maximally increased, while halfway two target directions,  $n\theta = \pi$ ,  $\cos(n\theta) = -1$ , and the tendency to extend is maximally suppressed.

The second term in Equation 3 captures an overall elongation, implemented in a similar fashion. The parameter  $\chi$  corresponds to the propensity to elongate and  $\alpha$  is the angle between the elongation vector and the copy vector.

If only these two terms are used, cells within tissue simulations can easily lose coherence, i.e., fall apart. Therefore a third term was added, capturing a propensity to roundness. The parameter  $\mu$  captures the resistance of a cell to deviate from a circle,



with  $r$  being the length of the copy vector, and  $\sqrt{A/\pi}$  being the preferred radius of cell, given its target area. For further details, see van Rooij et al. (2017a).

Importantly, the target lobe and elongation vectors are not fixed during the simulation. At intervals of 100 simulation time steps they are dynamically updated, in order to attain the most favourable position, effectively “accommodating” its lobe positions with respect to its neighbours. During a vector update step, the preferred directions of extension are matched to the set of directions the current shape of cells presents the strongest level of extension.

The initial cell positions within the field were randomly chosen. Simulations were run for 10000 time steps (see an example in Movie S5). Parameters used for each used specified cell shape are given in Table S1.

## Acknowledgements

We thank Enrico Coen for stimulating discussions, and John Fozard for critical reading and helpful comments. This work has been supported by Consejo Nacional de Ciencia y Tecnología (CONACYT) and by the UK Biological and Biotechnology Research Council (BBSRC) via grant BB/J004553/1 to the John Innes Centre. VAG acknowledges support from the Royal Society Dorothy Hodgkin fellowship. JAvR acknowledges support from the Netherlands Consortium for Systems Biology (NCSB), which is part of the Netherlands Genomics Initiative/Netherlands Organization for Scientific Research (NGI/NWO).

## References

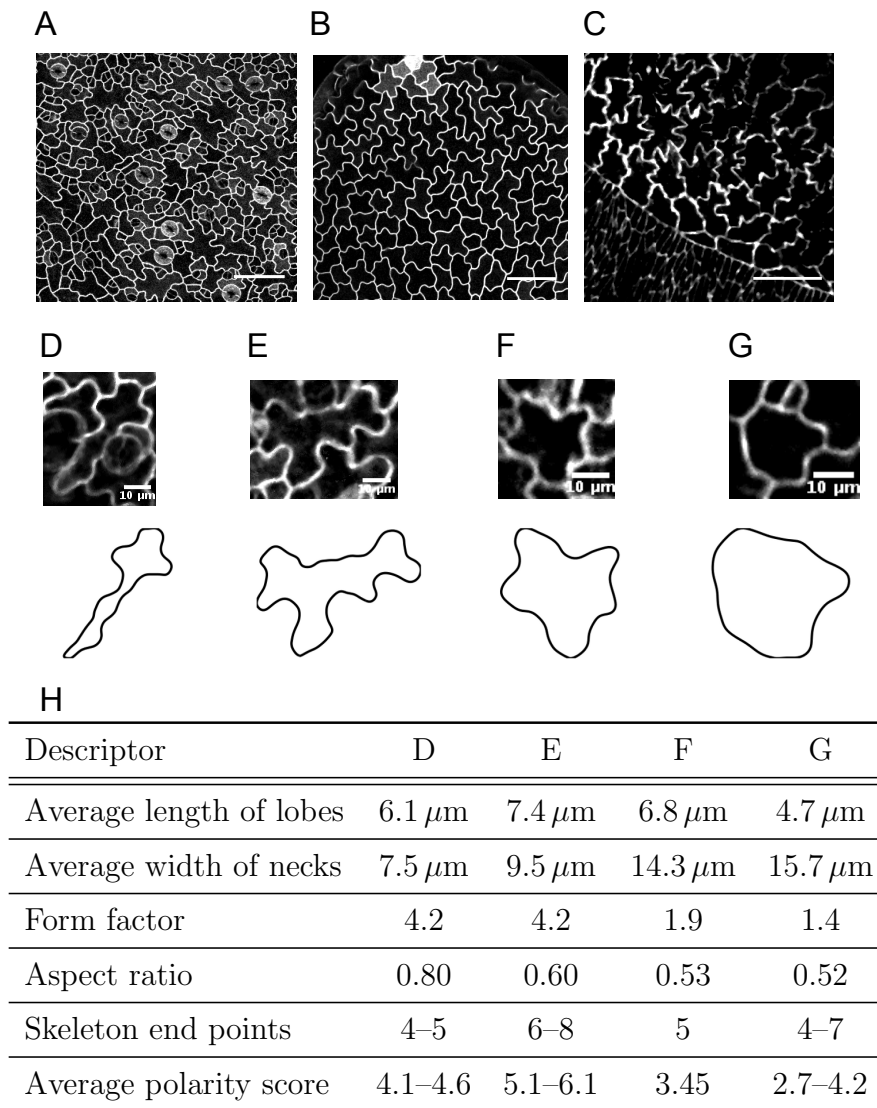
- Thompson DW (1917) On growth and form. Cambridge: Cambridge University Press.
- Charras G, Paluch E (2008) Blebs lead the way: how to migrate without lamellipodia. *Nat Rev Mol Cell Biol* 9:730–736. doi:10.1038/nrm2453.
- Marinari E, Mehonic A, Curran S, Gale J, Duke T, et al. (2012) Live-cell delamination counterbalances epithelial growth to limit tissue overcrowding. *Nature* 484:542–545. doi:10.1038/nature10984.
- Veeman MT, Smith WC (2013) Whole-organ cell shape analysis reveals the developmental basis of ascidian notochord taper. *Dev Biol* 373:281–289. doi:10.1016/j.ydbio.2012.11.009.
- Keren K, Pincus Z, Allen GM, Barnhart EL, Marriott G, et al. (2008) Mechanism of shape determination in motile cells. *Nature* 453:475–480. doi:10.1038/nature06952.
- Driscoll MK, McCann C, Kopace R, Homan T, Fourkas JT, et al. (2012) Cell shape dynamics: from waves to migration. *PLoS Comput Biol* 8:e1002392. doi:10.1371/journal.pcbi.1002392.
- Gomez JM, Chumakova L, Bulgakova NA, Brown NH (2016) Microtubule organization is determined by the shape of epithelial cells. *Nat Commun* 7:13172. doi:10.1038/ncomms13172.
- Sampathkumar A, Krupinski P, Wightman R, Milani P, Berquand A, et al. (2014) Sub-cellular and supracellular mechanical stress prescribes cytoskeleton behavior in *Ara-bidopsis* cotyledon pavement cells. *Elife* 3:e01967. doi:10.7554/eLife.01967.

- Minc N, Burgess D, Chang F (2011) Influence of cell geometry on division-plane positioning. *Cell* 144:414–426. doi:10.1016/j.cell.2011.01.016.
- Noda K, Glover BJ, Linstead P, Martin C (1994) Flower colour intensity depends on specialized cell shape controlled by a Myb-related transcription factor. *Nature* 369:661–664. doi:10.1038/369661a0.
- Zhong Q, Busetto AG, Fededa JP, Buhmann JM, Gerlich DW (2012) Unsupervised modeling of cell morphology dynamics for time-lapse microscopy. *Nat Methods* 9:711–713. doi:10.1038/nmeth.2046.
- Rajaram S, Pavie B, Wu LF, Altschuler SJ (2012) PhenoRipper: software for rapidly profiling microscopy images. *Nat Methods* 9:635–637. doi:10.1038/nmeth.2097.
- Ljosa V, Sokolnicki KL, Carpenter AE (2012) Annotated high-throughput microscopy image sets for validation. *Nat Methods* 9:637. doi:10.1038/nmeth.2083.
- Ivakov A, Persson S (2013) Plant cell shape: modulators and measurements. *Front Plant Sci* 4:439. doi:10.3389/fpls.2013.00439.
- Klingenberg CP (2010) Evolution and development of shape: integrating quantitative approaches. *Nat Rev Genet* 11:623–635. doi:10.1038/nrg2829.
- Fu Y, Gu Y, Zheng Z, Wasteneys G, Yang Z (2005) *Arabidopsis* interdigitating cell growth requires two antagonistic pathways with opposing action on cell morphogenesis. *Cell* 120:687–700. doi:10.1016/j.cell.2004.12.026.
- Le J, Mallery EL, Zhang C, Brankle S, Szymanski DB (2006) *Arabidopsis* BRICK1/HSPC300 is an essential WAVE-complex subunit that selectively stabilizes the Arp2/3 activator SCAR2. *Curr Biol* 16:895–901. doi:10.1016/j.cub.2006.03.061.
- Wu TC, Belteton SA, Pack J, Szymanski DB, Umulis DM (2016) Lobefinder: A Convex Hull-Based Method for Quantitative Boundary Analyses of Lobed Plant Cells. *Plant Physiol* 171:2331–2342. doi:10.1104/pp.15.00972.
- Lin D, Cao L, Zhou Z, Zhu L, Ehrhardt D, et al. (2013) Rho GTPase signaling activates microtubule severing to promote microtubule ordering in *Arabidopsis*. *Curr Biol* 23:290–297. doi:10.1016/j.cub.2013.01.022.
- Pincus Z, Theriot JA (2007) Comparison of quantitative methods for cell-shape analysis. *J Microsc* 227:140–156. doi:10.1111/j.1365-2818.2007.01799.x.
- Kuhl FP, Giardina CR (1982) Elliptic Fourier features of a closed contour. *Comput Gr Image Process* 18:236–258.
- Diaz G, Zuccarelli A, Pelligra I, Ghiani A (1989) Elliptic Fourier analysis of cell and nuclear shapes. *Comput Biomed Res* 22:405–414.
- Schmittbuhl M, Allenbach B, Le Minor JM, Schaaf A (2003) Elliptical descriptors: some simplified morphometric parameters for the quantification of complex outlines. *Math Geol* 35:853–871.

- Graner F, Glazier JA (1992) Simulation of biological cell sorting using a two-dimensional extended Potts model. *Phys Rev Lett* 69:2013–2016.
- Glazier JA, Graner F (1993) Simulation of the differential adhesion driven rearrangement of biological cells. *Phys Rev E* 47:2128–2154.
- Haines AJ, Crampton JS (2000) Improvements to the method of Fourier shape analysis as applied in morphometric studies. *Palaeontology* 43:765–783.
- Diaz G, Quacci D, Dell’Orbo C (1990) Recognition of cell surface modulation by elliptic Fourier analysis. *Comput Methods Programs Biomed* 31:57–62.
- MacAlister CA, Ohashi-Ito K, Bergmann DC (2007) Transcription factor control of asymmetric cell divisions that establish the stomatal lineage. *Nature* 445:537–540. doi:10.1038/nature05491.
- Xu T, Wen M, Nagawa S, Fu Y, Chen JG, et al. (2010) Cell surface- and rho GTPase-based auxin signaling controls cellular interdigitation in *Arabidopsis*. *Cell* 143:99–110. doi:10.1016/j.cell.2010.09.003.
- Vanag VK, Epstein IR (2009) Pattern formation mechanisms in reaction-diffusion systems. *Int J Dev Biol* 53:673–681. doi:10.1387/ijdb.072484vv.
- Magno R, Grieneisen VA, Marée AFM (2015) The biophysical nature of cells: potential cell behaviours revealed by analytical and computational studies of cell surface mechanics. *BMC Biophys* 8:8. doi:10.1186/s13628-015-0022-x.
- van Rooij JA, Magno R, Marée AFM (2017a) Capturing complex cell shapes in the Cellular Potts Model framework. *manuscript in preparation*.
- Fernandez R, Das P, Mirabet V, Moscardi E, Traas J, et al. (2010) Imaging plant growth in 4D: robust tissue reconstruction and lineaging at cell resolution. *Nat Methods* 7:547–553. doi:10.1038/nmeth.1472.
- van Rooij JA, Fozard JA, Carter R, Hartley M, Olsson TSG, et al. (2017b) The Segmentation Potts Model: extracting quantitative cell information from 2d and 3d confocal microscopy images. *manuscript in preparation*.
- Grieneisen VA, van Berkel K, Xu T, Scheres B, Hogeweg P, et al. (2017) Intracellular ROP partitioning and auxin-mediated indirect cell-cell communication can explain interdigitation of pavement cells. *under revision*.
- Tetley RJ, Blanchard GB, Fletcher AG, Adams RJ, Sanson B (2016) Unipolar distributions of junctional Myosin II identify cell stripe boundaries that drive cell intercalation throughout *Drosophila* axis extension. *Elife* 5:2331–2342. doi:10.7554/eLife.12094.
- Heller D, Hoppe A, Restrepo S, Gatti L, Tournier AL, et al. (2016) Epitools: An Open-Source Image Analysis Toolkit for Quantifying Epithelial Growth Dynamics. *Dev Cell* 36:103–116. doi:10.1016/j.devcel.2015.12.012.
- Stegmaier J, Amat F, Lemon WC, McDole K, Wan Y, et al. (2016) Real-Time Three-Dimensional Cell Segmentation in Large-Scale Microscopy Data of Developing Embryos. *Dev Cell* 36:225–240. doi:10.1016/j.devcel.2015.12.028.

- Nelson BK, Cai X, Nebenführ A (2007) A multicolored set of *in vivo* organelle markers for co-localization studies in *Arabidopsis* and other plants. *Plant J* 51:1126–1136. doi:10.1111/j.1365-313X.2007.03212.x.
- Robinson S, Barbier de Reuille P, Chan J, Bergmann D, Prusinkiewicz P, et al. (2011) Generation of spatial patterns through cell polarity switching. *Science* 333:1436–1440. doi:10.1126/science.1202185.
- Sauret-Güeto S, Calder G, Harberd NP (2012) Transient gibberellin application promotes *Arabidopsis thaliana* hypocotyl cell elongation without maintaining transverse orientation of microtubules on the outer tangential wall of epidermal cells. *Plant J* 69:628–639. doi:10.1111/j.1365-313X.2011.04817.x.
- Kuchen EE, Fox S, Barbier de Reuille P, Kennaway R, Bensmihen S, et al. (2012) Generation of leaf shape through early patterns of growth and tissue polarity. *Science* 335:1092–1096. doi:10.1126/science.1214678.
- Gielis J (2003) A generic geometric transformation that unifies a wide range of natural and abstract shapes. *Am J Bot* 90:333–338. doi:10.3732/ajb.90.3.333.
- Bai Y, Falk S, Schnittger A, Jakoby MJ, Hülskamp M (2010) Tissue layer specific regulation of leaf length and width in *Arabidopsis* as revealed by the cell autonomous action of ANGUSTIFOLIA. *Plant J* 61:191–199. doi:10.1111/j.1365-313X.2009.04050.x.
- Russ JC (2011) *The Image Processing Handbook*. Boca Raton: CRC Press, 6 edition.
- Andrianakaja M, Dhondt S, De Bodt S, Vanhaeren H, Coppens F, et al. (2012) Exit from proliferation during leaf development in *Arabidopsis thaliana*: a not-so-gradual process. *Dev Cell* 22:64–78. doi:10.1016/j.devcel.2011.11.011.
- Sorek N, Gutman O, Bar E, Abu-Abied M, Feng X, et al. (2011) Differential effects of prenylation and s-acylation on type I and II ROPs membrane interaction and function. *Plant Physiol* 155:706–720. doi:10.1104/pp.110.166850.

## Figures



**Figure 1: Complex cell shapes and the shortcomings of traditional shape quantifiers.** (A–C) Complex cell shapes in both plant (A, B) and animal (C) tissues. (A, B) Pavement cells of *Arabidopsis thaliana* wild type (A) and *speechless* mutant (B), characterised by jigsaw-like shapes consisting of multiple alternating protrusions (lobes) and indentations. (C) Amnioserosa cells in the *Drosophila* embryo also present cell shapes with similar complexity. Scale bars 50 μm. (D–G) Individual cells from the imaged tissues (upper panel), and the corresponding segmented cell outlines (lower panel). Scale bars 10 μm. (H) Traditional metrics to quantify cell shape can lead to similar values for very different shapes and are sensitive to parameter choices and imaging conditions. Here we compare the cells D–G. See also Figure S1.



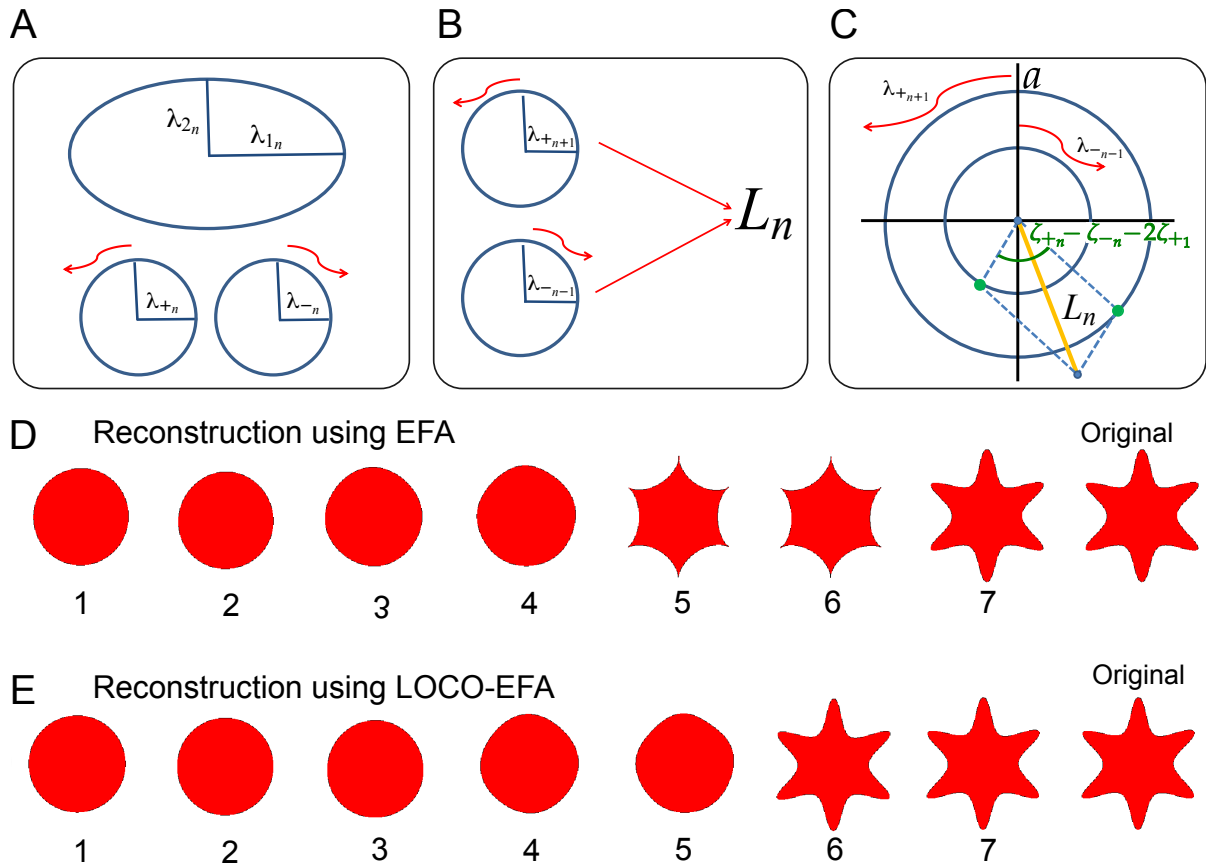
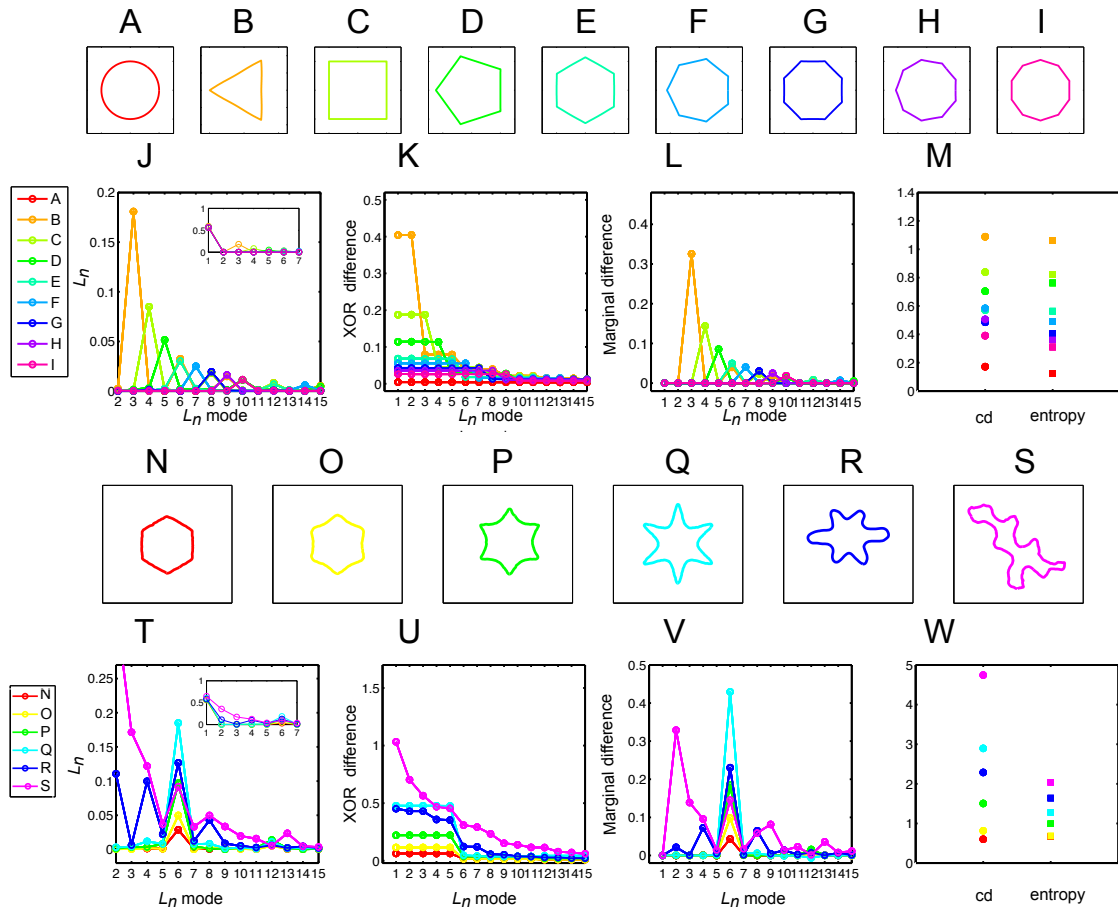


Figure 2: **To get LOCO.** (A) Each EFA elliptic harmonic is decomposed into two counter-rotating circles. (B) Mode  $L_n$  is composed of the counter-clockwise rotating  $n + 1$ th EFA harmonic circle and the clockwise rotating  $n - 1$ th circle. (C) The combined amplitude contribution to  $L_n$  (yellow line) of the two counter-rotating circles with radii  $\lambda_{+n+1}$  and  $\lambda_{-n-1}$  is also dependent on the offset in their starting points and the offset of the overall (mode 1) starting point, which together determine the initial phase shift (green dots) in the amplitude contribution of each rotor. (D, E) Comparison of closed contour reconstruction through either EFA (D) or LOCO-EFA (E). Although both approximations eventually converge to the original 6-lobed star shape (labelled ‘Original’), the reconstruction using EFA harmonics (D) generates a spurious shape after addition of the 5th harmonic and only recovers the original shape after addition of the 7th harmonic, whereas the reconstruction using LOCO-EFA (E) reconstitutes the original shape only at the 6th mode, matching the number of protrusions. The number of modes used for each sequential reconstruction is indicated below each shape.



**Figure 3: Interpreting LOCO-EFA-derived measures for geometrical and asymmetric shapes.** (A–I) Symmetrical and well-defined geometrical shapes with normalised area. (J–L)  $L_n$ , XOR, and marginal difference profiles for shapes (A–I). (J, L) For each geometric shape, a clear peak appears in the profiles, this main contributor to the shape always coinciding with the number of protrusions. (M) Cumulative difference (cd) and entropy for shapes (A–I). (N–Q) Symmetrical shapes with increasing protrusion amplitude. (R, S) Asymmetrical shapes. (T–V)  $L_n$ , XOR, and marginal difference profiles for shapes (N–S). Increasing protrusion amplitude leads to increasing peak levels in the profiles. Asymmetric shapes present multiple peaks, indicating that multiple modes are needed to recapitulate the original shape. (W) Cumulative difference (cd) and entropy for shapes (N–S).

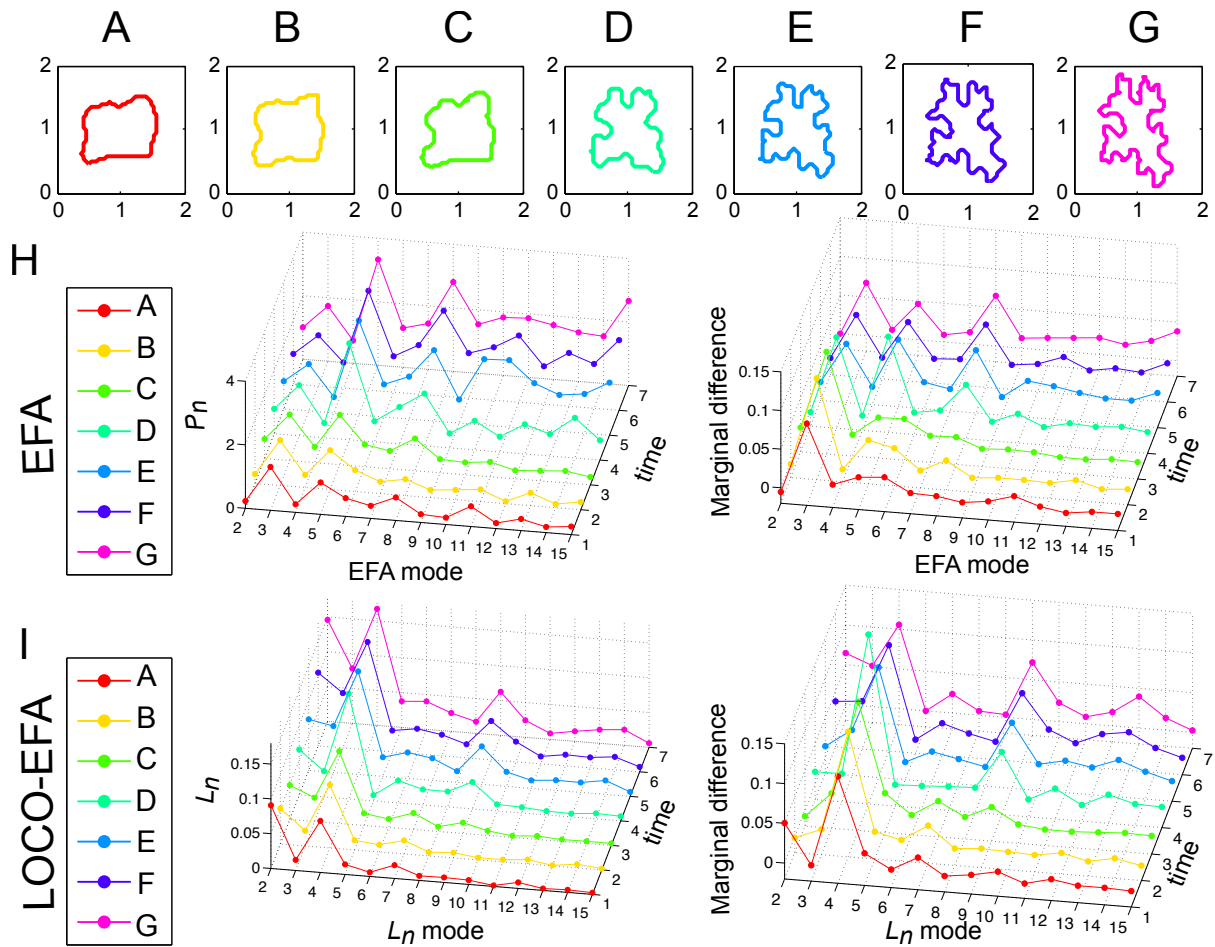


Figure 4: **LOCO-EFA metrics on a cell changing its shape over time.** (A–G) Sequence of a tracked pavement cell growing over time with normalised area. (H)  $P_n$  and marginal difference profiles using EFA. Applying EFA modes to approximate the cell shapes leads to erratic profiles that fail to recover the biological sequence of development, as can be observed in the  $P_n$  profile and as spurious peaks at the third and fifth harmonics in the marginal difference profile. (I)  $L_n$  and marginal difference profiles using LOCO-EFA. The LOCO-EFA measurements recover the smooth transitions during the cell morphogenesis. The overall square symmetry of the cell is captured by a peak at  $L_4$ , the formation of lobes by a smooth increase in  $L_9$ , and later  $L_{13}$ .

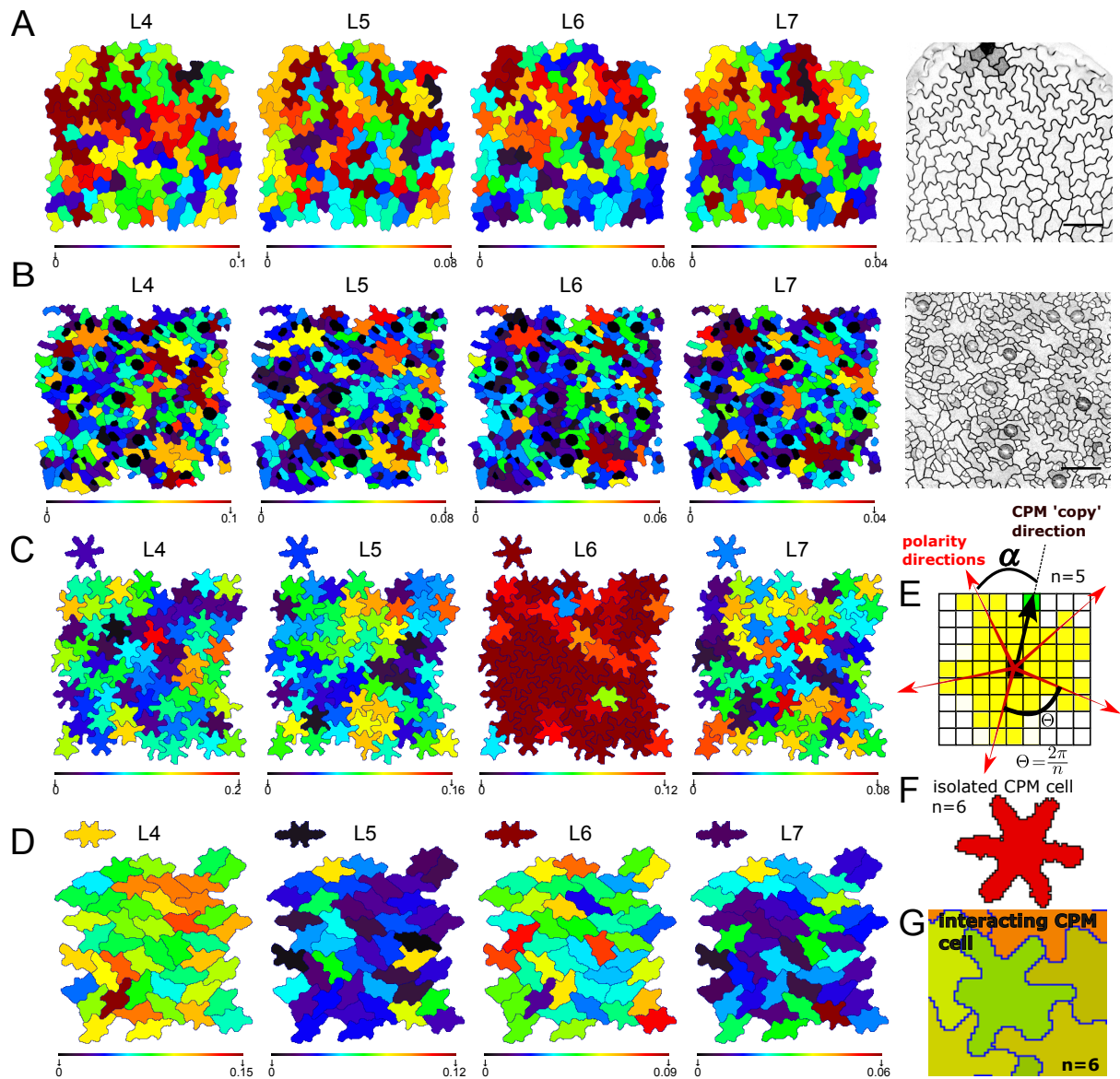


Figure 5: **LOCO-EFA analysis on *in vivo* and *in silico* pavement cells.** (A, B) LOCO-EFA applied to *speechless* mutant (A) and wild type (B) leaf tissue. Colour coding depicts the  $L_n$  values for 4 different  $L_n$  modes ( $n = 4 \dots 7$ ), as indicated above each panel, with the scale shown below each panel. It shows that very few cell shapes can be reasonably captured through a single  $L_n$  mode, revealing the complexity of the cell shapes. (C, D) LOCO-EFA applied to *in silico* pavement cells reveals the degree of divergence from their specified shape that interacting cells within a tissue experience. Two different specified cell shape populations are shown (SCS1 and SCS3, each with six lobes, see Table S1). The specified shapes are depicted above each panel. Colour coding within the panels and of the specified shapes above each panel again depicts the  $L_n$  values, with the scale shown below each panel. Because cells interact with each other within the tissue, strong deviations in  $L_n$  contributions can be observed. (E–G) Modelling framework to generate the *in silico* tissues. (E) Standard Cellular Potts Model is modified to allow for a specified number of lobes (here,  $n = 5$ ) to be formed at a regular radial spacing ( $\alpha$ ). (F) This gives rise to a symmetric, multi-lobed specified cell shape, shown in red. (G) Within a tissue context, however, the same specified shape deforms as to accompany the neighbouring cells.

## Table

Table 1: **Distinct shape descriptors have been used to quantify pavement cells.**

Measure	Description	Reference
Average lobe length and neck width	Length of each lobe and the distance between opposite indentations within a cell (called necks) are shown in Figure S1. The final measure for a cell is the average of all lobe lengths and the average of all the neck widths. These measurements depend on human criteria to identify lobes and necks, and are given in absolute length units (thus, incomparable throughout growth stages).	(Fu et al., 2005)
Form factor (or circularity)	Defined as $\frac{P^2}{4\pi A}$ , where $p$ is the perimeter and $A$ is cell area. A circle corresponds to a form factor 1, the lowest value possible.	(Bai et al., 2010; Russ, 2011; Andriankaja et al., 2012)
Skeleton	This metric relies on the number of end points of a skeleton representation of the cell shape. The skeleton is formed by iteratively removing pixels from a grid-based cell shape representation, such that eventually a branched one-dimensional graph remains. There are different variants of this algorithm to skeletonise shapes; the resulting branch patterns and length of branches highly depends on the parameters used and are very sensitive to the image resolution.	(Le et al., 2006; Russ, 2011)
Average polarity score	Defined as $\frac{c+s}{2}$ , where $c$ is the circularity and $s$ the number of skeleton end points.	(Sorek et al., 2011)



## Supplementary Materials and Methods

### Decomposing shape: Lobe Contribution Elliptical Fourier Analysis (LOCO-EFA)

In this section, we first summarise previous efforts to make EFA coefficients interpretable within a morphometrics perspective and explain why matching EFA coefficients with shape features generally does not hold. We describe in detail our new method, Lobe Contribution Elliptical Fourier Analysis (LOCO-EFA). We show how it provides quantitative and biologically interpretable measurements that are unique for a given shape, overcoming the shortfalls of the previous methods.

#### Standard Fourier Analysis cannot be used to quantify complex cell shapes

Standard Fourier Analysis has been widely used to analyse cell morphology. It can, however, only be applied when cells present simple holomorphic shapes, i.e., when the radii emanating from the centroid of a cell intersect its outline only once (Figure S2A and Pincus and Theriot (2007)). When the geometry of a cell is more complex, as in the case of pavement cells, and radii emanating from the centroid can intersect the outline more than once, the shape cannot be decomposed using a Fourier expansion based on polar coordinates (Figure S2B and Schmittbuhl et al. (2003)).

#### Elliptical Fourier Analysis fails to align mode frequency with morphological features

In 1982 Kuhl and Giardina proposed the Elliptical Fourier Analysis (EFA) to describe the contour of any two-dimensional shape (both holomorphic and non-holomorphic), derived from the coordinates of all the points along its outline.

In short, EFA takes the  $x$  and  $y$  coordinates of a closed contour and decomposes it into an infinite summation of related ellipses:

$$x(t) = \alpha_0 + \sum_{n=1}^{\infty} \left( \alpha_n \cos \left( \frac{2n\pi t}{T} \right) + \beta_n \sin \left( \frac{2n\pi t}{T} \right) \right), \quad (4a)$$

$$y(t) = \gamma_0 + \sum_{n=1}^{\infty} \left( \gamma_n \cos \left( \frac{2n\pi t}{T} \right) + \delta_n \sin \left( \frac{2n\pi t}{T} \right) \right), \quad (4b)$$

where  $\alpha_n$ ,  $\beta_n$ ,  $\gamma_n$  and  $\delta_n$  are the so-called EFA coefficients and  $\alpha_0$  and  $\gamma_0$  are the  $x$ - and  $y$ -offset of the initial contour. The detailed derivation of the formulae for  $\alpha_0$ ,  $\gamma_0$ ,  $\alpha_n$ ,  $\beta_n$ ,  $\gamma_n$  and  $\delta_n$  can be found in Kuhl and Giardina (1982). They are calculated from a discrete chain of contour points  $(x_i, y_i)$  with  $i = 1, \dots, K$  (see Figure S3A),  $K$  being the total number of points along the closed contour. We define  $(x_0, y_0) \equiv (x_K, y_K)$ , given that cell contours are closed. Now imagine drawing the contour of the cell, then  $\Delta t_i$  is the time spent drawing the line segment of the contour that links  $(x_{i-1}, y_{i-1})$  to  $(x_i, y_i)$ , i.e.,  $\Delta t_i = \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2} = \sqrt{\Delta x_i^2 + \Delta y_i^2}$ . Note that  $\Delta t_i$  is not fixed but can vary for each interval. Define  $T$  as the total time spent to draw the whole contour, i.e.,  $T = \sum_{i=1}^K \Delta t_i$ . The “time” passed while drawing the contour, starting from contour point  $(x_0, y_0)$ , or, equivalently, the distance passed along the contour to reach each contour

point  $(x_i, y_i)$ , is referred to as  $t_i$ , i.e.,  $t_i = \sum_{p=1}^i \Delta t_p$ , with  $t_0 = 0$ , and  $t_K = T$ , the “total drawing time” or total perimeter length (see Figure 1 in the main text). Given that no equal spacing between the points is required, it is straightforward to define  $K$  observation points from any kind of cell contour. The only requirements are that the contour is closed and the coordinates form an ordered list that follows the contour. The EFA coefficients are then given by:

$$\alpha_n = \frac{T}{2n^2\pi^2} \sum_{i=1}^K \frac{\Delta x_i}{\Delta t_i} \left( \cos \frac{2n\pi t_i}{T} - \cos \frac{2n\pi t_{i-1}}{T} \right), \quad (5a)$$

$$\beta_n = \frac{T}{2n^2\pi^2} \sum_{i=1}^K \frac{\Delta x_i}{\Delta t_i} \left( \sin \frac{2n\pi t_i}{T} - \sin \frac{2n\pi t_{i-1}}{T} \right), \quad (5b)$$

$$\gamma_n = \frac{T}{2n^2\pi^2} \sum_{i=1}^K \frac{\Delta y_i}{\Delta t_i} \left( \cos \frac{2n\pi t_i}{T} - \cos \frac{2n\pi t_{i-1}}{T} \right), \quad (5c)$$

$$\delta_n = \frac{T}{2n^2\pi^2} \sum_{i=1}^K \frac{\Delta y_i}{\Delta t_i} \left( \sin \frac{2n\pi t_i}{T} - \sin \frac{2n\pi t_{i-1}}{T} \right). \quad (5d)$$

The offset to the contour is given by:

$$\alpha_0 = \frac{1}{T} \sum_{i=1}^K \left( \frac{\Delta x_i}{2\Delta t_i} (t_i^2 - t_{i-1}^2) + \xi_i (t_i - t_{i-1}) \right) + x_0, \quad (6a)$$

$$\gamma_0 = \frac{1}{T} \sum_{i=1}^K \left( \frac{\Delta y_i}{2\Delta t_i} (t_i^2 - t_{i-1}^2) + \varepsilon_i (t_i - t_{i-1}) \right) + y_0. \quad (6b)$$

where  $\xi_i = \sum_{j=1}^{i-1} \Delta x_j - \frac{\Delta x_i}{\Delta t_i} \sum_{j=1}^{i-1} \Delta t_j$ ,  $\varepsilon_i = \sum_{j=1}^{i-1} \Delta y_j - \frac{\Delta y_i}{\Delta t_i} \sum_{j=1}^{i-1} \Delta t_j$ , and  $\xi_1 = \varepsilon_1 = 0$ . Further details and full derivation can be found in Kuhl and Giardina (1982).

Each set of four coefficients yields an ellipse (also referred to as the “ $n$ th mode” or “ $n$ th elliptic harmonic”), with a certain orientation and a certain starting point. The original cell outline can thus be expressed as an infinite summation of ellipses. Note that  $x(t)$  and  $y(t)$  are periodic functions with period equal to  $T$ .

A visual way to understand how the set of ellipses gives rise to the final shape is as follows: the second elliptic harmonic traces two clockwise or counter-clockwise revolutions around the first harmonic; the third harmonic traces three revolutions around the path drawn by the second harmonic; and the  $n$ th harmonic traces  $n$  revolutions around the path drawn by the previous harmonic (see Figure S3 and Movie S1).

Diaz et al. (1990) proposed a heuristic measure regarding the contribution of each harmonic to the shape through an approximation of the perimeter of each ellipse multiplied by its harmonic number  $n$ :

$$P_n = 2\pi n \sqrt{\frac{\lambda_{1n}^2 + \lambda_{2n}^2}{2}}, \quad (7)$$

where  $\lambda_{1n}$  and  $\lambda_{2n}$  are the major and minor axis of the  $n$ th ellipse. Moreover, Diaz et al. (1990) introduced an additional correction to capture the complex relationship between EFA modes and shape feature periodicity.

The direction of rotation of the  $n$ th harmonic ellipse is given by the determinant of the EFA coefficients matrix,  $\det \begin{bmatrix} \alpha_n & \beta_n \\ \gamma_n & \delta_n \end{bmatrix}$ , i.e., the direction of rotation is given by the sign of

$$r_n = \alpha_n \delta_n - \beta_n \gamma_n. \quad (8)$$

If  $r_n < 0$ , the elliptic harmonic is rotating clockwise; if  $r_n > 0$  the elliptic harmonic is rotating counter-clockwise. When EFA is used for shape approximation, mode  $n$  contributes to shape features with an  $n + 1$  or  $n - 1$  periodicity. This is in contrast to standard Fourier Analysis, in which mode  $n$  contributes to shape features with an  $n$  periodicity. Standard Fourier Analysis, however, is only possible for holomorphic shapes, and hence cannot be applied to, for example, pavement cells. Diaz et al. (1990) observed that whether mode  $n$  predominantly contributes to shape features with an  $n + 1$  or with an  $n - 1$  periodicity strongly depends on whether the  $n$ th harmonic rotates together with or against the direction of the first harmonic (see Movie S2 and Movie S3). This effect of presenting contributions to the  $n + 1$ 'th and  $n - 1$ 'th mode depending on the rotation direction of the first and  $n$ th harmonic is a common phenomenon observed for objects orbiting around others (hereafter referred to as the relative direction effect). A well-known example of the relative direction effect is the rotation of the Earth and its movement around the sun. The actual number of rotations our planet makes per year (as observed from "star-rise to star-rise", the so-called sidereal days) is one off from the number of days we perceive in a year (from "sunrise to sunrise", the so-called solar days). Because our planet rotates around its axis in the same direction as it moves around the Sun, the number of solar days per year is 365, one less than the number of sidereal days per year, which is 366. If the rotation of Earth would have been in the opposite direction as its movement around the sun, the number of solar days per year would instead have been 367. In light of exactly the same principle, Diaz et al. (1990) introduced that when the  $n$ th elliptic harmonic is moving in the same direction as the first harmonic, its shape contribution  $P_n$  should be assigned to  $n - 1$ ; inversely, when the direction of a given mode is opposite to the first harmonic, its shape contribution  $P_n$  should be assigned to  $n + 1$ .

We will show below that this simple heuristic is reasonable as long as the ellipse marginally deviates from a circle, but is not valid in general. When the aspect ratio of the ellipse ( $\lambda_{1n}/\lambda_{2n}$ ) is large (i.e., the elliptical harmonic is very flat, deviating significantly from a circular shape), the proposed rule fails to apply. Figure S3F illustrates a situation when the rotation direction of the first and third harmonic are opposite (and no other modes are used), yet instead of generating a contour with  $n - 1 = 2$  protrusions, as expected from the heuristic rule, a four-sided outline is generated, clearly illustrating that this method of  $P_n$  shifting does not work in general (see also Movie S4). Moreover, it is not possible to reconstruct the original shape using the  $P_n$  values, and therefore cannot be used for additional analysis based on shape reconstruction as presented in the main paper. This strongly limits usage of EFA for biological shape interpretations and statistical population analysis. Surprisingly, although EFA has been used to quantify morphology at the organ level, relative direction effect has typically been ignored altogether (Yoshioka et al., 2005; Frieß and Baylac, 2003; Neto et al., 2006; Iwata et al., 1998, 2010; Chitwood et al., 2013).

Realising that the source of the problem is linked to the eccentricity of the ellipses, it became clear to us that we could overcome this issue by essentially decomposing each ellipse into two counter-rotating circles. All circles can then be redistributed, forming a

new base. The details of how to do so are discussed below. We call the new base  $\mathcal{L}_n$ , which, when summed up, can also reconstitute the original shape.

## Contouring the limitations: Lobe Contribution Elliptical Fourier Analysis (LOCO-EFA)

To capture the biologically relevant cell shape features, overcoming the limitations of using  $P_n$  and rotation-dependent  $n + 1$ ,  $n - 1$  adjustments, we have developed an alternative method coined Lobe Contribution Elliptical Fourier Analysis (LOCO-EFA). As the name indicates, it correctly maps the contribution of each mode/harmonic to the corresponding morphological features. This is done by separating each elliptic harmonic into two circular harmonics, each rotating in an opposite direction.

First we rewrite the EFA (Equation 4) in matrix form:

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} \alpha_0 \\ \gamma_0 \end{bmatrix} + \sum_{n=1}^N \begin{bmatrix} \alpha_n & \beta_n \\ \gamma_n & \delta_n \end{bmatrix} \begin{bmatrix} \cos\left(\frac{2n\pi t}{T}\right) \\ \sin\left(\frac{2n\pi t}{T}\right) \end{bmatrix}, \quad (9)$$

with the infinite sum being truncated at the  $N$ th order harmonic.

Equation 9 can concisely be expressed as

$$[X(t)] = [A_0] + \sum_{n=1}^N [A_n] [M_n(t)], \quad (10)$$

in which  $[X(t)]$  corresponds to the drawn cell outline  $\begin{bmatrix} x(t) \\ y(t) \end{bmatrix}$ ;  $[A_0]$  represents the spatial offset  $\begin{bmatrix} \alpha_0 \\ \gamma_0 \end{bmatrix}$ ;  $[A_n]$  corresponds to the EFA coefficients matrix  $\begin{bmatrix} \alpha_n & \beta_n \\ \gamma_n & \delta_n \end{bmatrix}$ ; and  $[M_n(t)]$  refers to the rotor  $\begin{bmatrix} \cos\left(\frac{2n\pi t}{T}\right) \\ \sin\left(\frac{2n\pi t}{T}\right) \end{bmatrix}$ . (For clarity, we will use the notation  $[..]$  throughout to emphasise we are dealing with matrices; not to be confused with  $|\cdot|$  that represents determinant, which we here only refer to as  $\det[..]$ .)

The LOCO-EFA method consists of three steps: 1) eliminate multiple representations of the same outline; 2) decompose each  $n$ th elliptic harmonic into two circular harmonics, each rotating in an opposite direction; and 3) determine  $\mathcal{L}_n$  and  $L_n$  for all  $N$  modes. Below we describe these steps in detail.

### (1) Eliminate multiple representations of the same outline

It had already been noted that EFA coefficients are redundant and therefore compromise statistical analysis and shape comparisons (Haines and Crampton, 2000). We found that there are three sources of degeneracy in the EFA coefficients that therefore have to be eliminated. First, a contour can be drawn starting from any arbitrary initial point along the contour. While exactly the same outline is drawn, each starting point is represented by a completely different set of EFA coefficients for all modes  $[A_n]$ . Basically, whenever the starting point is changed, all elliptic harmonics take a different orientation (Kuhl and Giardina, 1982). The first step is therefore to transform the EFA coefficients such that the starting point of the first harmonic is always positioned at, for standardisation, the extreme of the semi-major axis (see further below). The second source of degeneracy, however, is that such a normalisation still allows for two possible representations of the

outline, since each of the two extremes along the semi-major axis can be chosen as the starting point. Moreover, a third source of degeneracy is due to the fact that the outline can be drawn clockwise or counter-clockwise. Clearly, all three sources of degeneracy have to be removed to make any comparison between cells sensible.

The first step is to determine where the new starting point should be positioned, as well as the scaled amount of time or temporal angle ( $\tau_1 = \frac{2\pi t_1}{T}$ ) required to reach the starting point (see Kuhl and Giardina, 1982). As stated above, we wish the starting point to coincide with one of the extremes of the semi-major axis of the first harmonic. Points along the first harmonic  $(x_1, y_1)$  can be described as :

$$x_1(\tau) = \alpha_1 \cos \tau + \beta_1 \sin \tau, \quad (11a)$$

$$y_1(\tau) = \gamma_1 \cos \tau + \delta_1 \sin \tau, \quad (11b)$$

with  $\tau = \frac{2\pi t}{T}$  being the scaled time or temporal angle. By differentiating the magnitude of the first harmonic ellipse  $E(\tau) = \sqrt{x_1(\tau)^2 + y_1(\tau)^2}$  and setting its derivative to zero ( $\frac{dE(\tau)}{d\tau} = 0$ ), the temporal angles can be found at which the extremes along the semi-major and semi-minor axes of the first harmonic are reached (Kuhl and Giardina, 1982):

$$\tau_1(\nu) = \frac{1}{2} \arctan \left( \frac{2(\alpha_1\beta_1 + \gamma_1\delta_1)}{\alpha_1^2 + \gamma_1^2 - \beta_1^2 - \delta_1^2} \right) + \frac{\nu}{2}\pi. \quad (12)$$

The values  $\nu = 0, 1, 2, 3$  give the four possible solutions along both the axes, after which the same points get repeated. For LOCO-EFA, it is required (see further below) to limit the starting point to the semi-major axis only. To satisfy this condition, the second derivative of  $E(\tau)$ , evaluated at the temporal angle, should be negative, i.e.,  $\left. \frac{d^2E(\tau)}{d\tau^2} \right|_{\tau_1} < 0$ . Substituting the found solutions into the second derivative results in  $\nu = 0$  and  $\nu = 2$  belonging to the points along the semi-major axis whenever the denominator of the arctan term is positive, and the solutions  $\nu = 1$  and  $\nu = 3$  belonging to the points along the semi-major axis whenever the denominator of the arctan term is negative. A very straightforward computational implementation of this result is to make use of the four-quadrant inverse tangent function (`atan2`) as provided by most programming languages (i.e., such that `atan2(1, 1) =  $\pi/4$`  is different from `atan2(-1, -1) =  $-3\pi/4$` ). Then, using

$$\tau_1 = 0.5 \text{atan2} \left( 2(\alpha_1\beta_1 + \gamma_1\delta_1), \alpha_1^2 + \gamma_1^2 - \beta_1^2 - \delta_1^2 \right) \quad (13)$$

automatically and unambiguously ensures that the temporal angle  $\tau_1$  is located at one of the extremes of the semi-major axis.

This still leaves two ways to position the starting point (one for each of the extremes of the semi-major axis) and thereby two distinct representations of a same outline. We therefore further restrict  $\tau_1$  to always lie within the first or second quadrant (I or II in Figure S9A). This is achieved by testing if the obtained  $\tau_1$  that shifts the starting point of the first harmonic to the semi-major axis indeed positions it within quadrant I or II. To shift the starting point we first introduce a time shift  $\tau' = \tau - \tau_1$  such that at  $\tau' = 0$  the first harmonic is positioned along its semi-major axis:



$$[A_1] [M_1(t)] = [A_1] [M_1(\tau' + \tau_1)] , \quad (14a)$$

$$= [A_1] \begin{bmatrix} \cos(\tau' + \tau_1) \\ \sin(\tau' + \tau_1) \end{bmatrix} , \quad (14b)$$

$$= [A_1] \begin{bmatrix} \cos(\tau_1) \cos(\tau') - \sin(\tau_1) \sin(\tau') \\ \sin(\tau_1) \cos(\tau') + \cos(\tau_1) \sin(\tau') \end{bmatrix} , \quad (14c)$$

$$= [A_1] \begin{bmatrix} \cos(\tau_1) - \sin(\tau_1) \\ \sin(\tau_1) + \cos(\tau_1) \end{bmatrix} \begin{bmatrix} \cos(\tau') \\ \sin(\tau') \end{bmatrix} , \quad (14d)$$

$$= [A_1] [\psi_{\tau_1}] [M_1(\tau')] , \quad (14e)$$

$$= [A'_1] [M_1(\tau')] , \quad (14f)$$

were  $[\psi_{\tau_1}]$  is the rotation operator, rotating by an angle  $\tau_1$  and  $[A'_1] = [A_1] [\psi_{\tau_1}]$ . The spatial angle at the shifted starting point  $\varrho$  is given by

$$\varrho(\nu) = \arctan\left(\frac{\gamma'_1}{\alpha'_1}\right) + \nu\pi . \quad (15)$$

Again, a single, unique and correct solution for  $\varrho$  can be obtained by using  $\varrho = \text{atan2}(\gamma'_1, \alpha'_1)$  instead. The starting point lies in quadrant III or IV when  $\varrho < 0$ . In that case,  $\tau_1$  is modified as follows:

$$\tau_1^* = (\tau_1 + \pi) . \quad (16)$$

Otherwise (when the starting point is already in quadrant I or II),  $\tau_1^* = \tau_1$ .

The new EFA coefficients corrected for the starting point then become:

$$\begin{bmatrix} \alpha_n^* & \beta_n^* \\ \gamma_n^* & \delta_n^* \end{bmatrix} = \begin{bmatrix} \alpha_n & \beta_n \\ \gamma_n & \delta_n \end{bmatrix} \begin{bmatrix} \cos(n\tau_1^*) & -\sin(n\tau_1^*) \\ \sin(n\tau_1^*) & \cos(n\tau_1^*) \end{bmatrix} . \quad (17)$$

Finally, we ensure that the direction of contour approximation of the first harmonic is always counter-clockwise (i.e., that  $r_1 \geq 0$ , Equation 8). Besides removing redundancy by restricting the freedom of choice regarding the overall direction of contour approximation, this transformation also guarantees a unique correspondence between the properties of each subsequent harmonic and its contribution to the morphological features. When the direction of the first harmonic is clockwise ( $r_1 < 0$ ), we therefore invert the direction of motion of all ellipses, maintaining thereby their inter-relationships. This can be done by running “time” backwards:

$$\begin{aligned} \begin{bmatrix} x(-t) \\ y(-t) \end{bmatrix} &= \begin{bmatrix} \alpha_n^* & \beta_n^* \\ \gamma_n^* & \delta_n^* \end{bmatrix} \begin{bmatrix} \cos\left(\frac{2n\pi(-t)}{T}\right) \\ \sin\left(\frac{2n\pi(-t)}{T}\right) \end{bmatrix} \\ &= \begin{bmatrix} \alpha_n^* & \beta_n^* \\ \gamma_n^* & \delta_n^* \end{bmatrix} \begin{bmatrix} \cos\left(\frac{2n\pi t}{T}\right) \\ -\sin\left(\frac{2n\pi t}{T}\right) \end{bmatrix} \\ &= \begin{bmatrix} \alpha_n^* & -\beta_n^* \\ \gamma_n^* & -\delta_n^* \end{bmatrix} \begin{bmatrix} \cos\left(\frac{2n\pi t}{T}\right) \\ \sin\left(\frac{2n\pi t}{T}\right) \end{bmatrix} . \end{aligned} \quad (18)$$

In short, whenever  $r_1 < 0$ , all indices  $\beta_n^*$  and  $\delta_n^*$  should be negated. After these steps, each unique cell contour is represented by a unique set of EFA coefficients. Note that the steps above do not alter the layout, nor do they rotate the shape. In certain study contexts, however, it might be desirable to rotate the contour itself, positioning

the semi-major axis, for example, to be parallel to the x-axis (or in any other preferred orientation). The details on how to perform those rotations can be found in Kuhl and Giardina (1982). Please note that unlike in their study, our subsequent analysis does not require such a cell contour realignment.

For simplicity of notation in the rest of the Supplementary Materials and Methods we refer to the  $[A_n]$  matrix, which elements have been normalised regarding the starting point and direction of reconstruction of the first harmonic:

$$[A_n] \equiv \begin{bmatrix} a_n & b_n \\ c_n & d_n \end{bmatrix} \equiv \begin{bmatrix} \alpha_n^* & \beta_n^* \\ \gamma_n^* & \delta_n^* \end{bmatrix}. \quad (19)$$

After all possible sources of redundancy have been removed, the next step of the LOCO-EFA method is to split each elliptic harmonic into two counter-rotating circles.

## (2) Decompose each $n$ th elliptic harmonic into two circles with opposite direction of rotation

In order to find the contribution of  $n$ th harmonic to a given morphological feature, we rewrite the  $[A_n]$  matrices in Equation 9 such as to explicitly introduce the length of the semi-major and semi-minor axis of the  $n$ th ellipse ( $\lambda_{1n}$  and  $\lambda_{2n}$ ).

For this purpose, it is necessary to introduce both the temporal and the spatial rotation operator of each elliptic harmonic, given by  $[\psi_{T_n}]$  and  $[\psi_{S_n}]$ , respectively. The temporal operator is defined as

$$[\psi_{T_n}] = \begin{bmatrix} \cos \phi_n & -\sin \phi_n \\ \sin \phi_n & \cos \phi_n \end{bmatrix}, \quad (20)$$

and the spatial operator is defined as

$$[\psi_{S_n}] = \begin{bmatrix} \cos \theta_n & -\sin \theta_n \\ \sin \theta_n & \cos \theta_n \end{bmatrix}, \quad (21)$$

where  $\phi_n$  is the temporal angle (i.e., the time  $\tau_n$  required to rotate to the semi-major axis) and  $\theta_n$  the spatial angle (i.e., the angle of this position along the semi-major axis with the positive x-axis) (Figure S9A).

Equation 10 can be written as:

$$[X(t)] = [A_0] + \sum_{n=1}^N [\psi_{S_n}] [\psi_{S_n}]^{-1} [A_n] [\psi_{T_n}] [\psi_{T_n}]^{-1} [M_n(t)], \quad (22)$$

given that  $[\psi_{S_n}] [\psi_{S_n}]^{-1}$  and  $[\psi_{T_n}] [\psi_{T_n}]^{-1}$  correspond to the identity matrix  $[I]$ .

Equation 10 can also be written in a form which directly highlights the contribution of the semi-major and semi-minor axis:

$$[X(t)] = [A_0] + \sum_{n=1}^N [\psi_{S_n}] [\Lambda_n] [\psi_{T_n}]^{-1} [M_n(t)]. \quad (23)$$

This equation can be understood as follows: for each mode, correctly position the starting point relative to the semi-major axis, transform the original circle into an ellipse, its semi-major axis along the x-axis and semi-minor axis along the y-axis, and finally rotate the ellipse to its correct position (see Figure S9B–D). Since both descriptions should

be equivalent, by combining Equation 22 and Equation 23, the  $[\Lambda_n]$  matrix can be identified, with  $\lambda_{1n}$  corresponding to the length of the semi-major axis, and the modulus of  $\lambda_{2n}$  to the length of the semi-minor axis of the  $n$ th ellipse:

$$[\Lambda_n] = [\psi_{S_n}]^{-1} [A_n] [\psi_{T_n}] = \begin{bmatrix} \lambda_{1n} & 0 \\ 0 & \lambda_{2n} \end{bmatrix}. \quad (24)$$

Note that this process is similar to a singular value decomposition of  $[A_n]$ , with the difference that here  $\lambda_{2n}$  (but not  $\lambda_{1n}$ ) can be negative. The temporal angle  $\phi_n$  corresponds, for each elliptic harmonic, to the scaled time  $\tau_n$  to reach an extreme along the semi-major axis (Figure S9). Similarly to Equation 12, this corresponds to

$$\phi_n(\nu) = \frac{1}{2} \arctan \left( \frac{2(a_n b_n + c_n d_n)}{a_n^2 + c_n^2 - b_n^2 - d_n^2} \right) + \frac{\nu}{2} \pi. \quad (25)$$

Once again, using atan2 ensures that  $\phi_n$  corresponds to the semi-major axis (see above). For the next steps of the analysis (see below) it is essential that  $\phi_n$  corresponds to the temporal angle to reach the semi-major axis, not the semi-minor axis, hence usage of atan2 or any equivalent function which determines the quadrant of the return value is essential.

The spatial angle  $\theta_n$  (see Figure S9) can be calculated after applying the temporal modification:

$$\theta_n(\nu) = \arctan \frac{c'_n}{a'_n} + \nu \pi, \quad (26)$$

where  $c'_n$  and  $a'_n$ , are the new coefficients after the temporal transformation (i.e., after applying  $[A_n] [\psi_{T_n}]$ ). Again, it is essential to use atan2 to ensure that  $\phi_n$  and  $\theta_n$  are both relative to the same extreme of the semi-major axis. It does not matter, however, which of the two extremes is being used, hence for this step no check is required regarding the quadrants. Applying Equation 24 then provides  $\lambda_{1n}$  and  $\lambda_{2n}$ . Deriving a temporal and spatial angle relative to the semi-major axis and both being related to the same extreme guarantees that  $\lambda_{1n} \geq 0$  (while  $\lambda_{2n}$  can be positive or negative, depending on the rotation direction of the rotor) and  $|\lambda_{1n}| \geq |\lambda_{2n}|$ .

Using the above, the EFA (Equation 10) can be rewritten as:

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = [A_0] + \sum_{n=1}^N \begin{bmatrix} \cos \theta_n & -\sin \theta_n \\ \sin \theta_n & \cos \theta_n \end{bmatrix} \begin{bmatrix} \lambda_{1n} & 0 \\ 0 & \lambda_{2n} \end{bmatrix} \begin{bmatrix} \cos \phi_n & \sin \phi_n \\ -\sin \phi_n & \cos \phi_n \end{bmatrix} \begin{bmatrix} \cos \left( \frac{2n\pi t}{T} \right) \\ \sin \left( \frac{2n\pi t}{T} \right) \end{bmatrix}. \quad (27)$$

Written in this form, the contribution of each harmonic can easily be separated to correctly map to morphological feature number. The diagonal matrix containing the length of the semi-major and semi-minor axis of each  $n$ th mode can be decomposed into two diagonal matrices, each corresponding to circular orbits moving in opposite directions:

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = [A_0] + \sum_{n=1}^N \begin{bmatrix} \cos \theta_n & -\sin \theta_n \\ \sin \theta_n & \cos \theta_n \end{bmatrix} \left( \begin{bmatrix} \lambda_{+n} & 0 \\ 0 & \lambda_{+n} \end{bmatrix} + \begin{bmatrix} \lambda_{-n} & 0 \\ 0 & -\lambda_{-n} \end{bmatrix} \right) \begin{bmatrix} \cos \phi_n & \sin \phi_n \\ -\sin \phi_n & \cos \phi_n \end{bmatrix} \begin{bmatrix} \cos \left( \frac{2n\pi t}{T} \right) \\ \sin \left( \frac{2n\pi t}{T} \right) \end{bmatrix}, \quad (28)$$

where  $\lambda_{+n}$  and  $\lambda_{-n}$  are the radii of those circles (Figure 2).

Summing up the diagonal matrices in Equation 28 yields

$$\begin{aligned} & \begin{bmatrix} x(t) \\ y(t) \end{bmatrix} \\ &= [A_0] + \sum_{n=1}^N \begin{bmatrix} \cos \theta_n & -\sin \theta_n \\ \sin \theta_n & \cos \theta_n \end{bmatrix} \begin{bmatrix} \lambda_{+n} + \lambda_{-n} & 0 \\ 0 & \lambda_{+n} - \lambda_{-n} \end{bmatrix} \begin{bmatrix} \cos \phi_n & \sin \phi_n \\ -\sin \phi_n & \cos \phi_n \end{bmatrix} \begin{bmatrix} \cos \left( \frac{2n\pi t}{T} \right) \\ \sin \left( \frac{2n\pi t}{T} \right) \end{bmatrix}, \end{aligned} \quad (29)$$

in which the major and minor axes of each elliptic harmonic are

$$\lambda_{1n} = \lambda_{+n} + \lambda_{-n}, \quad (30a)$$

$$\lambda_{2n} = \lambda_{+n} - \lambda_{-n}, \quad (30b)$$

and hence the radii of each oppositely-rotating circle is given by

$$\lambda_{+n} = (\lambda_{1n} + \lambda_{2n})/2, \quad (31a)$$

$$\lambda_{-n} = (\lambda_{1n} - \lambda_{2n})/2. \quad (31b)$$

Given that  $\lambda_{1n} \geq 0$  and  $|\lambda_{1n}| \geq |\lambda_{2n}|$ ,  $\lambda_{+n}$  and  $\lambda_{-n}$  are always positive. To approximate the cell contour  $(x(t), y(t))$  using the circles  $\lambda_{+n}$  and  $\lambda_{-n}$  requires completing the transformations using the spatial ( $\theta_n$ ) and temporal angle ( $\phi_n$ ) as calculated before, most clearly seen through the expression

$$[X(t)] = [A_0] + \sum_{n=1}^N [\psi_{S_n}] [\Lambda_{+n} + \Lambda_{-n}] [\psi_{T_n}]^{-1} [M_n(t)]. \quad (32)$$

The term  $\Lambda_{+n}$  presents the subset of the  $n$ th elliptic harmonic which is moving in the same direction as the first harmonic, therefore purely contributing to  $n - 1$  “lobes”, i.e., shape features with periodicity  $n - 1$ . In contrast,  $\Lambda_{-n}$  presents the subset moving in the opposite direction, purely contributing to  $n + 1$  “lobes” (shape features) only. Their contributions can be separated by writing:

$$[X(t)] = [A_0] + \sum_{n=1}^N [A_{+n}] [M_n(t)] + \sum_{n=1}^N [A_{-n}] [M_n(t)], \quad (33)$$

where

$$[A_{j_n}] = [\psi_{S_n}] [\Lambda_{j_n}] [\psi_{T_n}]^{-1}, \quad \text{for } j = +, -. \quad (34)$$

This can be further simplified. Straightforwardly,  $[\psi_{S_n}] [\Lambda_{+n}] = [\psi_{S_n}] \lambda_{+n} [I] = [\Lambda_{+n}] [\psi_{S_n}]$ . Regarding  $[\Lambda_{+n}]$ ,

$$[\psi_{S_n}] [\Lambda_{-n}] = \begin{bmatrix} \cos \theta_n & -\sin \theta_n \\ \sin \theta_n & \cos \theta_n \end{bmatrix} \begin{bmatrix} \lambda_{-n} & 0 \\ 0 & -\lambda_{-n} \end{bmatrix}, \quad (35a)$$

$$= \begin{bmatrix} \lambda_{-n} \cos \theta_n & +\lambda_{-n} \sin \theta_n \\ \lambda_{-n} \sin \theta_n & -\lambda_{-n} \cos \theta_n \end{bmatrix}, \quad (35b)$$

$$= \begin{bmatrix} \lambda_{-n} & 0 \\ 0 & -\lambda_{-n} \end{bmatrix} \begin{bmatrix} \cos \theta_n & \sin \theta_n \\ -\sin \theta_n & \cos \theta_n \end{bmatrix}, \quad (35c)$$

$$= \begin{bmatrix} \lambda_{-n} & 0 \\ 0 & -\lambda_{-n} \end{bmatrix} \begin{bmatrix} \cos(-\theta_n) & -\sin(-\theta_n) \\ \sin(-\theta_n) & \cos(-\theta_n) \end{bmatrix}, \quad (35d)$$

$$= [\Lambda_{-n}] [\psi_{S_n}]^{-1}. \quad (35e)$$

Thus,

$$[A_{+n}] = [\Lambda_{+n}] [\psi_{S_n}] [\psi_{T_n}]^{-1}, \quad (36a)$$

$$[A_{-n}] = [\Lambda_{-n}] [\psi_{S_n}]^{-1} [\psi_{T_n}]^{-1}. \quad (36b)$$

Hence, introducing  $\zeta_{+n} = \theta_n - \phi_n$  and  $\zeta_{-n} = -\theta_n - \phi_n$ , Equation 34 can be written as

$$[A_{j_n}] = [\Lambda_{j_n}] [\psi_{\zeta_{j_n}}], \quad \text{for } j = +, -, \quad (37)$$

which corresponds to

$$\begin{aligned} [A_{j_n}] &= \begin{bmatrix} a_{j_n} & b_{j_n} \\ c_{j_n} & d_{j_n} \end{bmatrix} \\ &= \begin{bmatrix} \lambda_{j_n} & 0 \\ 0 & j\lambda_{j_n} \end{bmatrix} \begin{bmatrix} \cos \zeta_{j_n} & -\sin \zeta_{j_n} \\ \sin \zeta_{j_n} & \cos \zeta_{j_n} \end{bmatrix}, \quad \text{for } j = +, -. \end{aligned} \quad (38)$$

We next label those matrices with respect to their lobe contribution instead of their EFA mode. To make the distinction, we here use subscript  $n$  to indicate the EFA mode, and subscript  $l$  to indicate the LOCO-EFA mode. In general,  $[A_{+l}] = [A_{+n+1}]$  and  $[A_{-l}] = [A_{-n-1}]$ . There are, however, a few exceptions, see Figure S4:

1) The overall offset of the contour is not solely given by  $[A_0]$ . An additional contribution to the offset is coming from  $[A_{+n=2}]$ . Note, however, that the contribution from  $[A_{+n=2}]$  is in fact not a perfect offset to the contour, but also causes a kidney bean-shaped distortion to the contour, its deviation from a pure offset becoming more pronounced when the contribution of this mode relative to the overall contour size is larger (see Figure S4B).

2) The overall circular shape of the contour (i.e., LOCO-mode 1) receives solely a contribution from  $[A_{+n=1}]$  itself, i.e.,  $[A_{+l=1}] = [A_{+n=1}]$ .

3) When  $N$  EFA modes are taken into account, then LOCO-mode  $N$  only receives a contribution from  $[A_{-n=N-1}]$ .

4) Likewise, when taking  $N$  EFA modes into account, there is still a contribution to LOCO-mode  $L = N + 1$ , solely coming from  $[A_{-n=N}]$ .

Defining  $[M_{+l}] = [M_{n+1}]$  and  $[M_{-l}] = [M_{n-1}]$ , with the exceptions  $[M_{+l=0}] = [M_{n=2}]$  and  $[M_{+l=1}] = [M_{n=1}]$ , then the same shape approximation can be achieved through LOCO-EFA as through EFA:

$$\begin{aligned} [X(t)] &= [A_0] + [A_{+l=0}] [M_{+l=0}(t)] + [A_{+l=1}] [M_{+l=1}(t)] \\ &\quad + \sum_{l=2}^{L-2} [A_{+l}] [M_{+l}(t)] + \sum_{l=2}^L [A_{-l}] [M_{-l}(t)]. \end{aligned} \quad (39)$$

Equation 39 can be used to reconstruct the shape up to a certain  $L$  number, requiring EFA coefficients up to mode  $L + 1$ :

$$[X_{L=1}(t)] = [A_0] + [A_{+l=0}] [M_{+l=0}(t)] + [A_{+l=1}] [M_{+l=1}(t)], \quad (40a)$$

$$[X_{L \geq 2}(t)] = [A_0] + [A_{+l=0}] [M_{+l=0}(t)] + [A_{+l=1}] [M_{+l=1}(t)] + \sum_{l=2}^L [A_{+l}] [M_{+l}(t)] + \sum_{l=2}^L [A_{-l}] [M_{-l}(t)]. \quad (40b)$$



Each LOCO-EFA mode can be fully described by the combination of the radii and the starting positions of both circles. The radii are, as previously defined,  $\lambda_{+l} = \lambda_{+n+1}$  and  $\lambda_{-l} = \lambda_{-n-1}$ , with the exceptions  $\lambda_{+l=0} = \lambda_{+n=2}$  and  $\lambda_{-l=0} = 0$ , and  $\lambda_{+l=1} = \lambda_{+n=1}$  and  $\lambda_{-l=1} = 0$ . The starting points of both circles are defined as  $\zeta_{+l} = \zeta_{+n+1}$  and  $\zeta_{-l} = \zeta_{-n-1}$ , again with the exceptions  $\zeta_{+l=0} = \zeta_{+n=2}$  and  $\zeta_{-l=0} = \zeta_{-n=2}$ . Together this gives a set of four coefficients  $\mathcal{L}_l = (\lambda_{+l}, \lambda_{-l}, \zeta_{+l}, \zeta_{-l})$  that fully capture each LOCO-EFA mode and allow for a full reconstitution of the original shape:

$$\begin{aligned} [X(t)] = & [A_0] + [\Lambda_{+l=0}] [\psi_{\zeta_{+l=0}}] [M_{+l=0}(t)] + [\Lambda_{+l=1}] [\psi_{\zeta_{+l=1}}] [M_{+l=1}(t)] \\ & + \sum_{l=2}^{L-2} [\Lambda_{+l}] [\psi_{\zeta_{+l}}] [M_{+l}(t)] + \sum_{l=2}^L [\Lambda_{-l}] [\psi_{\zeta_{-l}}] [M_{-l}(t)]. \end{aligned} \quad (41)$$

In the next section we derive how from those values a single amplitude value can be found, the  $L_l$  contribution.

### (3) $L_l$ amplitude contributions

In this section we first show that the contribution to the shape of the two counter-rotating circles of a specific LOCO-EFA mode is not simply determined by the radii of those circles, but also depends on the relative position of the starting points. We then derive a heuristic which optimally captures its contribution, determining as well the limits of this approximation. To illustrate how a difference between the starting points  $\zeta_{+l}$  and  $\zeta_{-l}$  of mode  $l$  can affect the generated amplitude, we first plot the contribution of  $\mathcal{L}_3 = (\lambda_{+3}, \lambda_{-3}, \zeta_{+3}, \zeta_{-3})$  superimposed on  $\mathcal{L}_1 = (1, 0, 0, 0)$ , i.e., superimposed on the unit circle starting at zero degrees. As long as  $\lambda_{+3}, \lambda_{-3}$  are not too large, a single amplitude value  $a$  (which is the amplitude of deviation from the unit circle) can be observed for any phase  $\omega$ ; peak amplitude values  $A$  occur for specific phases  $\Omega$  (as illustrated in FigureS10I). FigureS10A–C depict the shape contribution of the negative rotor (green), positive rotor (red) and their summed contribution (orange), as a function of the phase  $\omega$ , for  $\mathcal{L}_3 = (0.15, 0.15, 0, 0)$ . This scenario illustrates that even though the negative rotor makes two sidereal rotations while the positive rotor makes four sidereal rotations, the amplitude pattern that they generate not only has a period three in both cases, as argued throughout the paper, but also perfectly match one another regarding the phases at which the peaks and troughs in the amplitude pattern are reached. The summed contribution (FigureS10C) therefore indeed yields a peak amplitude exactly equal to  $\lambda_{+3} + \lambda_{-3}$ . In contrast, when both rotors are exactly out-of-phase ( $\mathcal{L}_3 = (0.15, 0.15, 0, \pi)$ , FigureS10D–F), the peaks and troughs are exactly out-of-phase as well, and the patterns almost (but not totally) cancel each other out (FigureS10F). To determine the effective contribution of a LOCO-EFA mode we therefore have to determine the phase at which each rotor reaches its peak amplitude.

The angles  $\Omega_+, \Omega_-$  at which the positive and negative rotor reach their peak amplitude can be calculated straightforwardly. They occur when the phase of the rotor itself is equivalent to the overall phase generated by  $\mathcal{L}_1$  (as illustrated in FigureS10G). The phase of  $\mathcal{L}_1$  starts at  $\zeta_{+1}$ , while the phase of the positive rotor starts at  $\zeta_{+l}$  (illustrated in FigureS10K). When the pattern is laid down, the phase of the positive rotor changes  $(l + 1)$  times faster than the phase of  $\mathcal{L}_1$  (Equation 41). Thus, regarding the phase at peak amplitude it holds that

$$\Omega_{+l} = (l + 1) (\Omega_{+l} - \zeta_{+1}) + \zeta_{+1} + 2\pi\nu, \quad (42a)$$

$$-l\Omega_{+l} = \zeta_{+1} - (l + 1)\zeta_{+1} + 2\pi\nu, \quad (42b)$$

$$\Omega_{+l} = \zeta_{+1} + \frac{\zeta_{+1} - \zeta_{+l}}{l} + \frac{2\pi}{l}\nu. \quad (42c)$$

where  $\nu$  is any integer, and the values  $\nu = 0, 1, \dots, l - 1$  provide the complete set of phases at which peak amplitude is reached. Likewise, the negative rotor rotates  $(l - 1)$  times faster than  $\mathcal{L}_1$  and in the opposite direction, hence starting at  $-\zeta_{-l}$  (see Figure S10J). Regarding the phase at peak amplitude it therefore holds that

$$\Omega_{-l} = -(l - 1) (\Omega_{+l} - \zeta_{+1}) - \zeta_{-l} + 2\pi\nu, \quad (43a)$$

$$l\Omega_{-l} = -\zeta_{-l} + (l - 1)\zeta_{+1} + 2\pi\nu, \quad (43b)$$

$$\Omega_{-l} = \zeta_{+1} - \frac{\zeta_{+1} + \zeta_{-l}}{l} + \frac{2\pi}{l}\nu. \quad (43c)$$

To assess the summed amplitude contribution we next fit the amplitude pattern laid down by each rotor to a sine wave, exactly matching both the peak amplitude and the phase at which the peak amplitude is reached, and then sum those two contributions:

$$a_{+l} = \lambda_{+l} \cos(l\omega + \zeta_{+1} - (l + 1)\zeta_{+1}), \quad (44a)$$

$$a_{-l} = \lambda_{-l} \cos(l\omega + \zeta_{-l} - (l - 1)\zeta_{+1}), \quad (44b)$$

$$a_{+l} + a_{-l} = \lambda_{+l} \cos(l\omega + \zeta_{+1} - (l + 1)\zeta_{+1}) + \lambda_{-l} \cos(l\omega + \zeta_{-l} - (l - 1)\zeta_{+1}). \quad (44c)$$

Equation 44a exactly matches the peak amplitude and its phase of the positive rotor, as derived in Equation 42, while Equation 44b exactly matches the peak amplitude and its phase of the negative rotor, as derived in Equation 43. While peak amplitude and phase match perfectly, Figure S10G, H illustrate that for not too large values of  $\lambda_{+l}$ ,  $\lambda_{-l}$  also the rest of the pattern presents a close match.

Using standard trigonometry, the summed amplitude can be written as

$$a_{+l} + a_{-l} = a_l = \sqrt{\lambda_{+l}^2 + \lambda_{-l}^2 + 2\lambda_{+l}\lambda_{-l} \cos(\zeta_{+1} - \zeta_{-l} - 2\zeta_{+1})} \cos(l\omega + \zeta_l), \quad (45a)$$

where

$$\zeta_l = \text{atan2}(\lambda_{+l} \sin(\zeta_{+1} - (l + 1)\zeta_{+1}) + \lambda_{-l} \sin(\zeta_{-l} - (l - 1)\zeta_{+1}), \lambda_{+l} \cos(\zeta_{+1} - (l + 1)\zeta_{+1}) + \lambda_{-l} \cos(\zeta_{-l} - (l - 1)\zeta_{+1})). \quad (45b)$$

Figure S10I illustrates that for not too large values of  $\lambda_{+l}$ ,  $\lambda_{-l}$  this expression provides a close match to the pattern generated by the  $\mathcal{L}_l$  mode, here illustrated using both a different phase and a different amplitude for both rotors. Using the equation above, the amplitude or contribution of mode  $l$  can hence be defined as

$$L_l = \sqrt{\lambda_{+l}^2 + \lambda_{-l}^2 + 2\lambda_{+l}\lambda_{-l} \cos(\zeta_{+1} - \zeta_{-l} - 2\zeta_{+1})}. \quad (46)$$

While above we have performed a more formal derivation, a more intuitive way to understand the amplitude contribution of the positive and negative rotor combined, is in terms of how much the amplitude contributions of both rotors are out-of-phase with each other. The most straightforward moment to determine their phase shift is at the initial point of drawing the cell's outline. As illustrated in Figure S10J, K, the phase shift regarding the amplitude contribution of the negative rotor is given by  $\zeta_{+1} + \zeta_{-1}$ , and of the positive rotor by  $\zeta_{+1} - \zeta_{-1}$ . To understand the graphical argument for the negative rotor case, one has to realise that the negative rotor rotates clockwise, and hence its starting angle is  $-\zeta_{-1}$  (for the positive rotor it is simply  $\zeta_{+1}$ ), while for the same reason the phase shift between initial amplitude and maximum amplitude has to be calculated clockwise (instead of standard counterclockwise, as done for the positive rotor), again as shown in Figure S10J, K. Plotting the phase shift and strength of both amplitude contributions as vectors and then adding them up then provides the above equation for  $L_l$  (as is depicted in Figure 2C).

Note that the pattern can never be perfectly described by a sine wave, which is why four parameters are needed to describe each mode, rather than the two required by Equation 45. Even when following Equation 45 the two waves cancel each other out, this is not exactly the case, as seen in Figure S10F. There is therefore no additional level of redundancy, as the usage of the  $L_l$  numbers might suggest, even when the value of  $L_l = 0$ . When the amplitude of a rotor becomes larger, the generated pattern starts to deviate from a sine wave (Figure S10J, K), and their summed pattern from Equation 45 (Figure S10L). For simple closed contours (which cell outlines should always be), the deviations cannot be very large, since otherwise the contour becomes non-simple (i.e., crosses itself). This confinement in deviations for cell outlines allows us to base a significant part of our analysis on  $L_l$  values.

Moreover, we observe that the positive rotor generates protrusions that are flatter than a sine wave (Figure S10B, J), whereas the negative rotor generates lobes that are more pointy (Figure S10A, K). The extent of "flatness" or "pointiness" of each mode can be determined by the proportion  $w_{+l} = \lambda_{+l}/(\lambda_{+l} + \lambda_{-l})$  and  $w_{-l} = 1 - w_{+l}$ , respectively; a useful additional measure if needed.

## Entropy measurement

The entropy measure is defined as

$$E = - \sum_{l=1}^L f_l \ln f_l, \quad (47)$$

where  $f_l$  refers to the relative proportion of each  $L_l$  for a given  $L$  number of modes analysed, i.e.,  $f_l = \frac{L_l}{\sum_{l=1}^L L_l}$ .

## Final remarks

To summarise, the LOCO-EFA method consists in: 1) eliminating degeneracies in the EFA coefficients; 2) decomposing each elliptic harmonic into two circles rotating in opposite directions ( $\lambda_{+n}$  and  $\lambda_{-n}$ ) and therefore contributing to  $n - 1$  and  $n + 1$  number of lobes (more generally, morphological features); and 3) calculating the offset between

starting points of these two circles derived from each ellipse to estimate the amplitude of the  $L_l$ th lobe contribution.

To eliminate cell area effects (for example, when looking at shape diversity within a cell populations), it might also be desirable to normalise for cell size by dividing each  $L_n$  value by the square root of the cell area. This then provides a complete description of the number of lobes and their amplitude, which can be used to characterise and quantify intrinsic cell shape properties, irrespective of cell area, spacing between sampling points, and rotations or inversions of the cell (Figure S11A, B). Changes in the image resolution, however can of course affect the fine-grained information retrieval, but only if the resolution becomes very low (Figure S11C, D).

## Supporting Figures

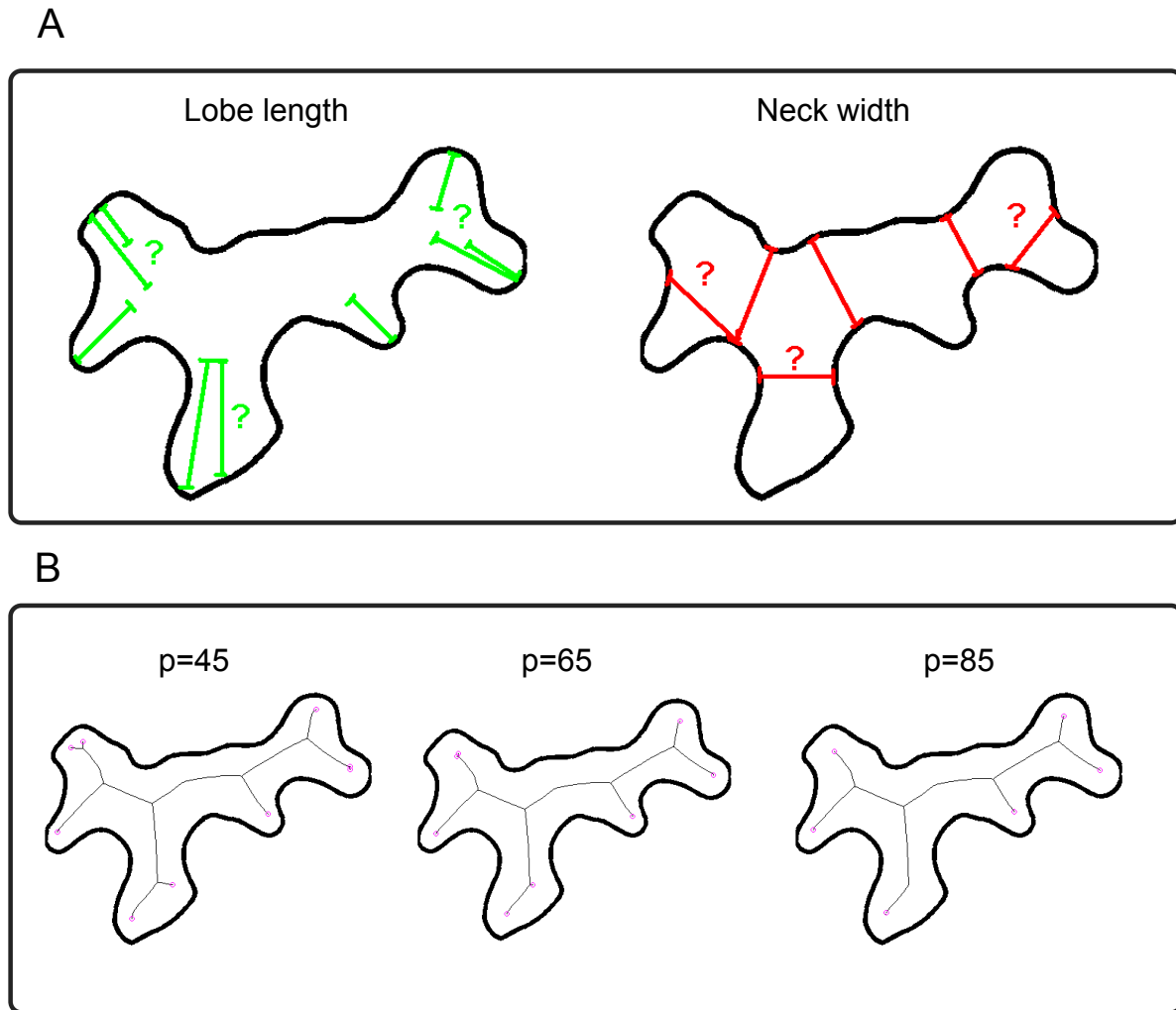


Figure S1: **Subjective choices involved in neck width and lobe length determination and Skeletonisation.** (A) Neck width and lobe length depend on human criteria for both identifying and quantifying such structures, as indicated by the question-marks. (B) Both the number of skeleton end-points and the length of the branches strongly depend on the parameter settings used for the skeletonisation algorithm. Here, parameter  $p$  (see material and Methods) was set to 45 (left), 65 (middle), or 85 (right). For those values, the number of predicted lobes varies between 6 and 8.



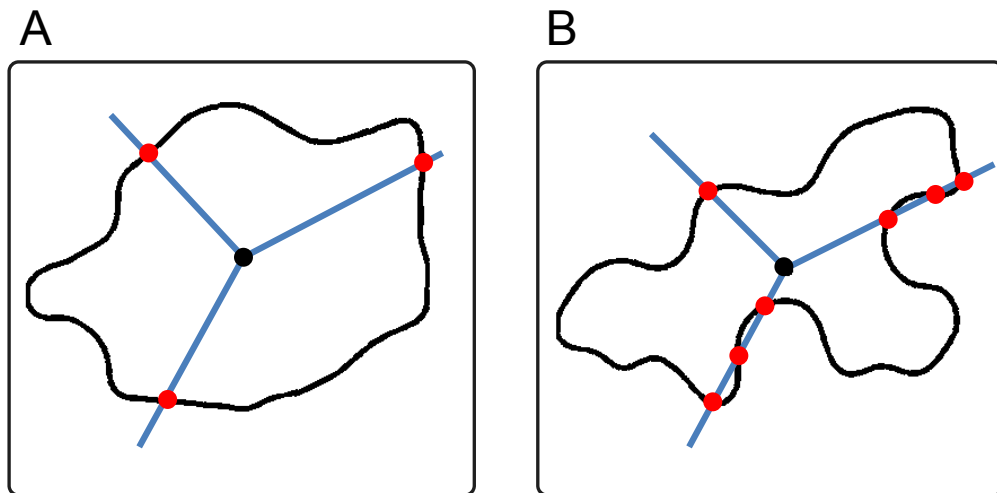


Figure S2: **Holomorphic and non-holomorphic shapes.** (A) In holomorphic shapes all radii starting from the centroid intersect the outline only once. (B) In non-holomorphic shapes some radii intersect more than once. Very few pavement cells have a holomorphic shape, the majority presenting highly complex non-holomorphic outlines. The outline of non-holomorphic shapes cannot be represented as a function of the angle, precluding, for example, standard Fourier analysis.

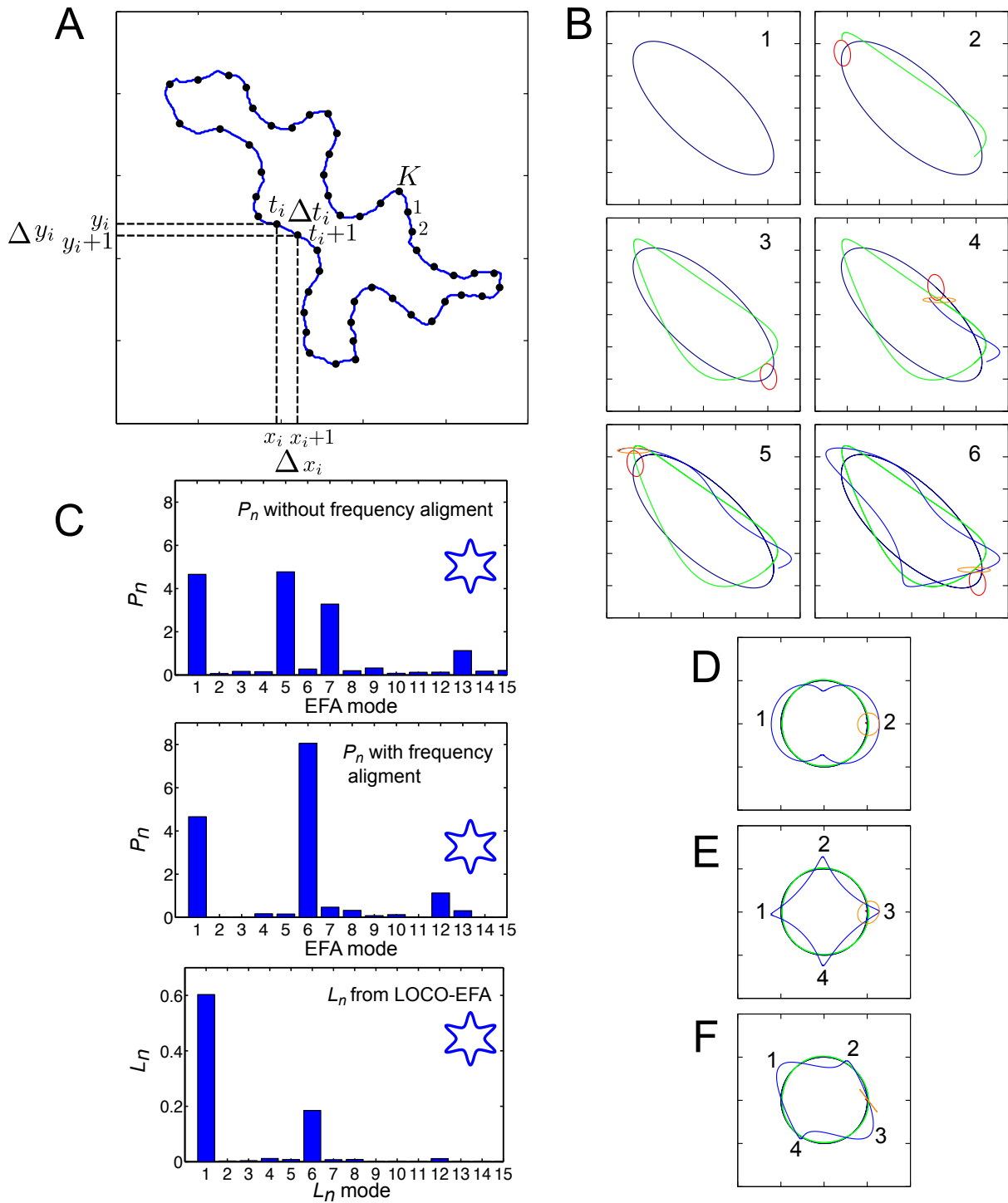


Figure S3

Figure S3: **Cell outline reconstruction using EFA and LOCO-EFA.** (A) Calculating the EFA coefficients from a cell outline. The discrete chain of contour points can be positioned arbitrarily (for example, the points do not have to be associated to an underlying grid). Also the distances between points can be arbitrarily long. (B) Sequential approximation of the cell's contour. The first harmonic forms an elliptic shape (1, blue). The second harmonic describes an elliptic orbit (2–6, red), orbiting twice while moving around the first harmonic (2, 3). Their summed trajectories are shown in green. The third elliptic harmonic (4–6, orange) orbits thrice while moving around this summed trajectory (4–6). The summation of the first three elliptic harmonics is shown in blue. See Movie S1 for this dynamical reconstitution of the cell contour. (C) The EFA coefficients cannot be directly linked to shape features, here shown through the power contribution of each harmonic ( $P_n$ ) (upper panel): the main  $P_n$  contributions of a six-pointed star-like shape (shown as an inset) come from the 5th and 7th harmonic. To align the coefficients to actual shape features, Diaz et al. (1990) proposed to shift the contribution of each harmonic to either  $n + 1$  or  $n - 1$ , depending on the rotation direction of each individual harmonic (middle panel). This brings the main shape contributor and the actual number of shape features in alignment to each other. This method, however, does not always hold (see F), and moreover generates a range of spurious contributions from a large set of different modes, which is a notorious issue that hampers analysis of complex shapes using standard EFA (Haines and Crampton, 2000). The LOCO-EFA method correctly aligns the shape assessment and the real shape features (lower panel), without generating any additional spurious contributions. Moreover, unlike the other methods, the values correctly represent the amplitude of the shape features. (D–F) Contours (shown in blue), generated from the first (green) and third EFA harmonic (orange) only. The number of morphological protrusions (lobes) specified by the  $n$ th EFA mode is affected, but not fully determined, by its rotation direction. The heuristic rule proposed by Diaz et al. (1990) states that if the first harmonic and the  $n$ th harmonic rotate in the same direction, a contour is generated with  $n - 1$  protrusions, while if their rotation direction is opposite, the contour will contain  $n + 1$  lobes. This rule, however, is only correct when the elliptic harmonic has a circular shape (D, E), in which case indeed overall shapes are generated with either 2 (D) or 4 lobes (E), simply dependent on the rotation direction with respect to the first mode. When, however, the elliptic harmonic has a higher eccentricity (F), the final shape can have  $n + 1$  protrusions (4 lobes) even though the rotation direction of the first and third harmonic are opposite. This is a consequence of each single EFA mode actually contributing to two different spatial modes. See also Movie S2–S4.

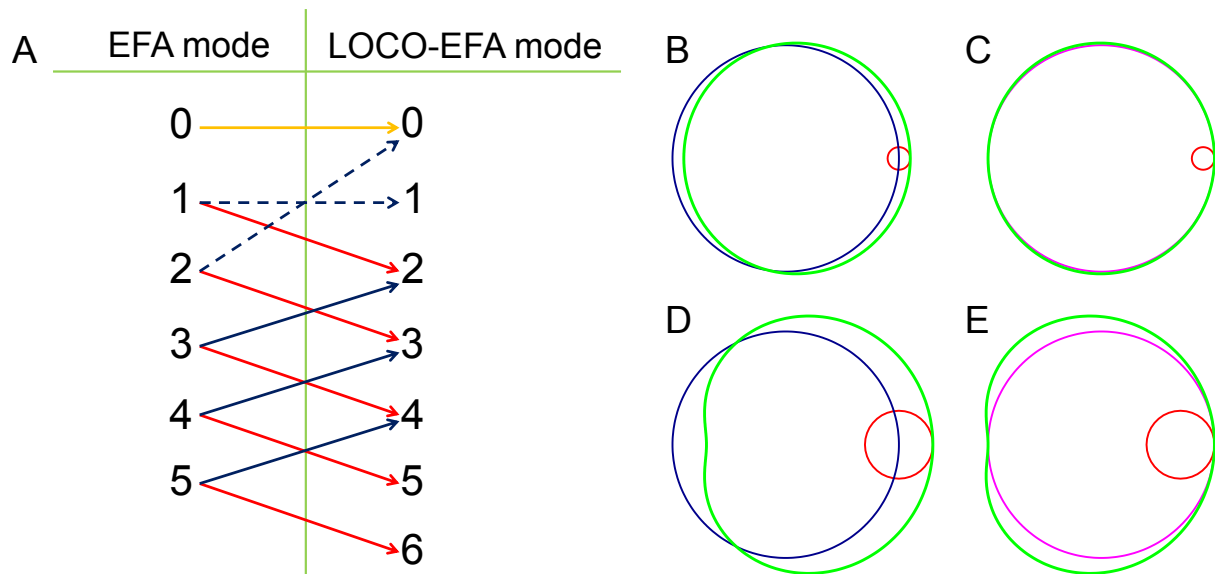


Figure S4: **Schematic mapping between EFA modes and LOCO-EFA modes.** (A) Each  $n$ th EFA mode contributes to both  $n + 1$  and  $n - 1$  morphological periodicities. The red arrows represent the contributions of the  $n$ th EFA mode to  $n + 1$  protrusions, due to the clockwise rotations of the circular harmonics  $\lambda_{-n+1}$ . The blue arrows indicate the contributions to  $n - 1$  protrusions, due to the counter-clockwise rotations of the circular harmonics  $\lambda_{+n-1}$ . A few exceptions apply: The second EFA mode contributes to a shift in the positioning of the layout, i.e., to LOCO-EFA mode 0 ( $\lambda_{+0}$ , blue and dashed line), rather than to the overall size of the layout, as might have been expected. The first EFA mode contributes to the overall circular size of the layout ( $\lambda_{+1}$ , blue and dashed line), rather than to a shift in the positioning of the layout, as might have been expected. The zeroth EFA mode only contributes to a shift in the positioning of the layout (yellow line). Finally, the two highest LOCO-EFA modes have incomplete contributions, given any cutoff in the number of EFA modes. (B–E) The contribution  $\lambda_{+0}$  is not simply an offset of the contour, but also involves a kidney bean-shaped distortion, more pronounced for larger contributions. Mode  $\lambda_{+1}$  (the circular mode) is shown in blue; mode  $\lambda_{+0}$  in red; their summation in green; and the predicted shape if the contribution were purely an offset in magenta. (B)  $\lambda_{+0}$  being 10% of  $\lambda_{+1}$ , effectively resulting in just an offset to the shape outline (C). (D, E)  $\lambda_{+0}$  being 30% of  $\lambda_{+1}$ , resulting in a clear kidney bean-shaped distortion of the shape.

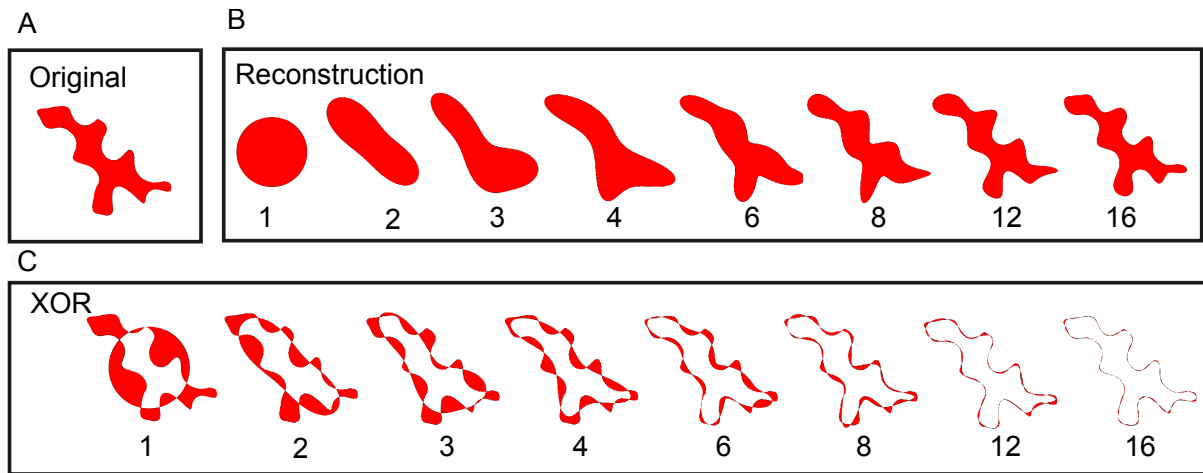


Figure S5: XOR measure extracted from the LOCO-EFA coefficients to provide additional information regarding the cell shape complexity. (A) Original cell contour. (B) LOCO-EFA reconstruction taking the first  $n$   $\mathcal{L}_n$  modes into account, as indicated below the panels. (C) Determination of the level of mismatch between the original cell shape and its  $n$ th order truncated LOCO-EFA approximation, by applying an XOR (Exclusive OR) function.

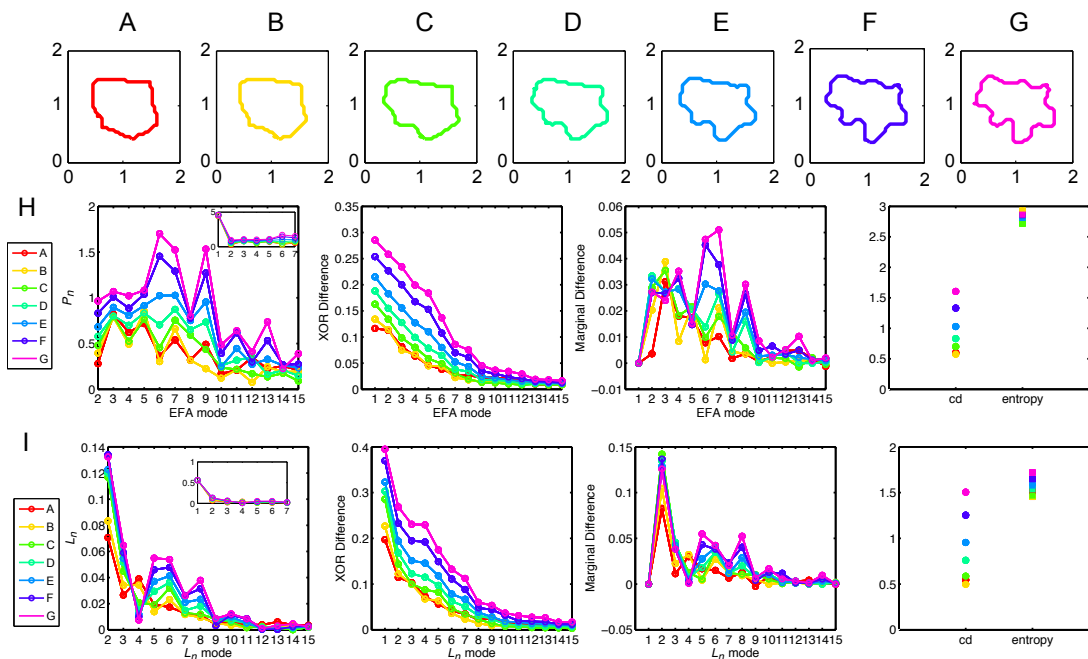


Figure S6: Additional example of LOCO-EFA metrics on a cell changing its shape over time. (A–G) Sequence of a tracked pavement cell growing over time with normalised area. (H)  $P_n$ , XOR difference and marginal difference profiles, cumulative difference and entropy using EFA. (I)  $L_n$ , XOR difference and marginal difference profiles, cumulative difference and entropy using LOCO-EFA. See also the first example presented in Figure 4 in the main text.



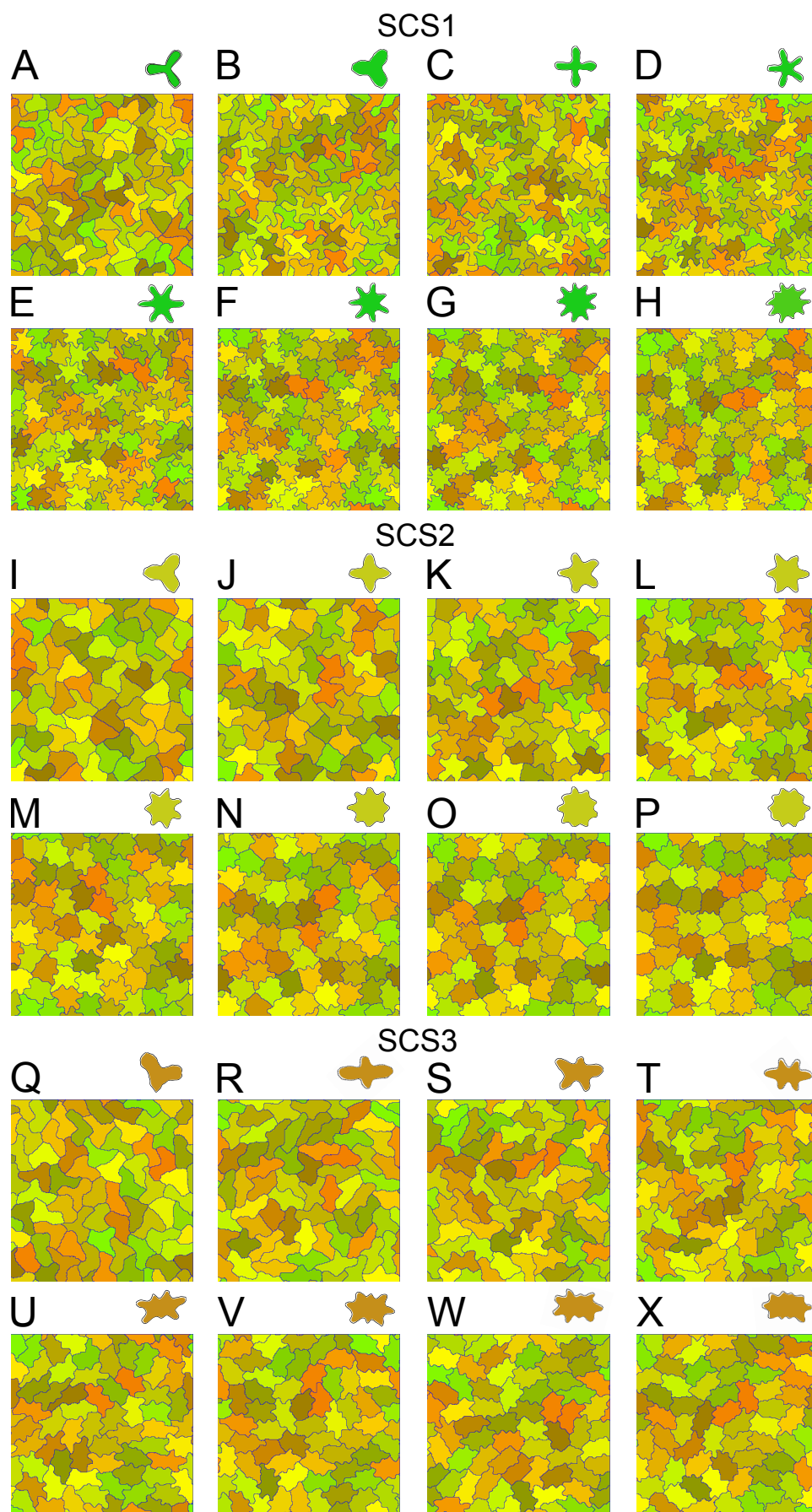
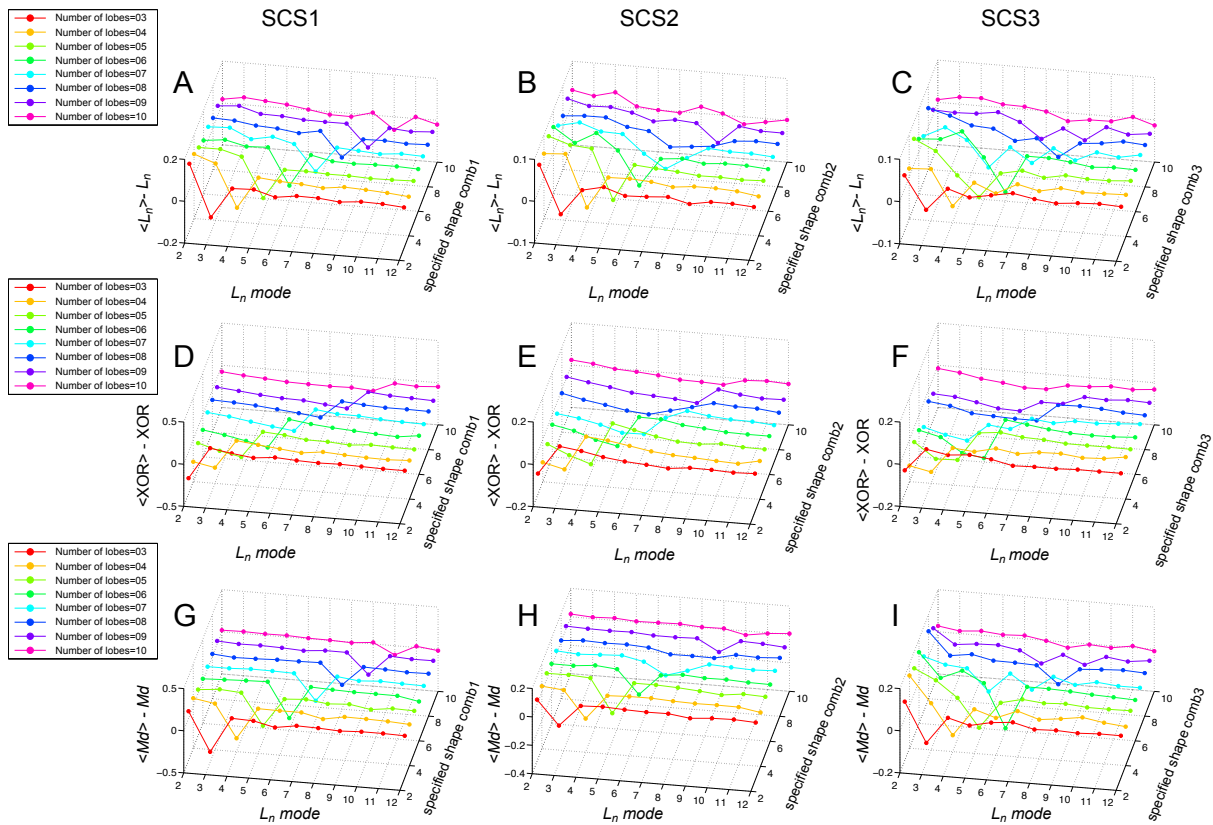


Figure S7

Figure S7: **Confluent *in silico* cell populations, simulated with three types of specified cell shapes (SCSs) and different number of specified lobes.** Parameters for the different specified cell shapes are given in Table S1. (A–H) Cells with large protrusions (Specified Cell Shapes 1, SCS1), with specified lobe number increasing from 3 (A) to 10 (H); (I–P) Cells with small protrusions (SCS2), with specified lobe number increasing from 3 (I) to 10 (P); (Q–X) Elongated cells (SCS3), with specified lobe number increasing from 3 (Q) to 10 (X). Specified cell shapes, as resulting from single-cell simulations, are shown above each panel. Cells are randomly coloured. See also Figure 5 and Figure S8.



**Figure S8: Divergence between specified cell shapes and resultant population-level cell shape diversity within confluent *in silico* cell populations.** The mismatch is visualised as the difference in LOCO-EFA-derived measures between the average for confluent population simulations and its value for single cell simulations. Three types of specified cell shapes (SCSs) were simulated, as indicated for each column, and parametrised in Table S1. Lobe numbers vary from 3 to 10, as indicated for each row and depicted in Figure S7. (A–C) Difference between the average  $L_n$  ( $\langle L_n \rangle$ ) values of the cells of the simulated populations and that of a single simulated cell. (D–F) Difference between average  $XOR$  ( $\langle XOR \rangle$ ) of the cells of the simulated population and that of a single simulated cell. (G–I) Same as in (A–C, D–F), but now for marginal difference (Md). In all cases the specified cell shape becomes less pronounced in the confluent population simulations (indicated by negative values that all measures yield at the given specified lobe number), while the cells present an increased shape diversity and complexity (seen by the broad flanking regions with positive values in the profiles, indicating a large range of modes that contribute to the shapes). These effects are present regardless of the number of specified lobes or the combination of parameters used, but becomes less pronounced at higher lobe numbers. See also Figure 5.

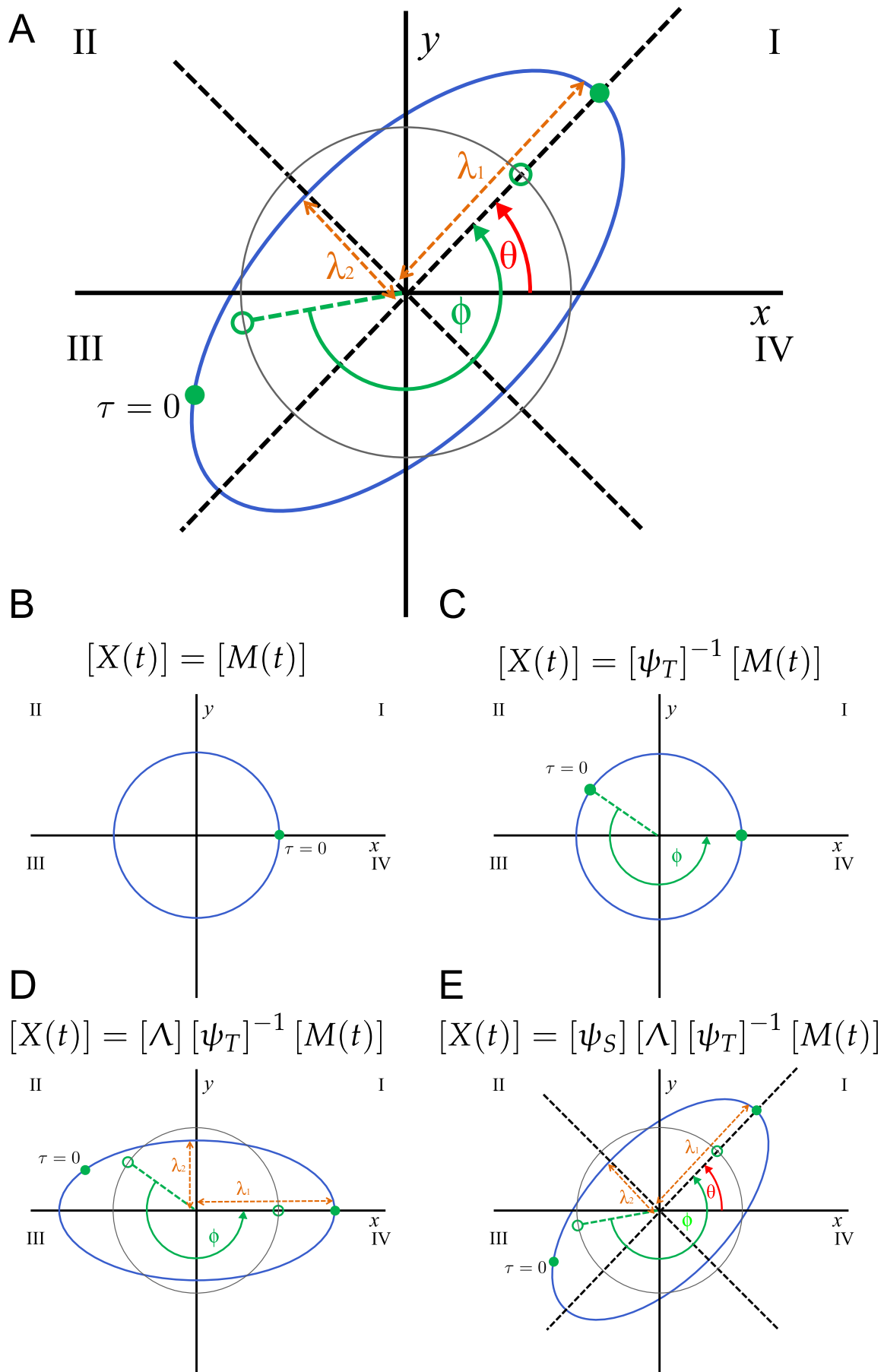


Figure S9

Figure S9: **Temporal and spatial transformations required to calculate the precise contribution of each EFA mode.** (A) The blue ellipse depicts an elliptic harmonic, given by  $[X(t)] = [A][M(t)]$ , while the grey circle depicts the unit circle, given by  $[X(t)] = [M(t)]$ . The temporal angle  $\phi$  is the scaled time required to move along an elliptic harmonic from the starting point at  $\tau = 0$  to one extreme along the semi-major axis (green filled circles). The angle  $\phi$  cannot be trivially derived from the spatial position at  $\tau = 0$ , requiring first an effective projection upon the unit circle (green open circles). The spatial angle  $\theta$  is the inclination of the elliptic harmonic. After applying the appropriate spatial and temporal transformations to the EFA coefficients using this geometrical interpretation, the semi-major and semi-minor axis,  $\lambda_1$  and  $\lambda_2$ , can be straightforwardly obtained. To eliminate multiple representations of the same outline, the starting point of the first harmonic is specifically positioned at the extreme along the semi-major axis which lies in either quadrant I or II, given by temporal angle  $\tau_1^*$  (Equation 17). In contrast, all other temporal angles  $\phi_n$  used, while also positioned along the semi-major axis, are not confined to quadrant I or II. (B–E) Visual interpretation of Equation 23 as a step-wise construction of the elliptic harmonic. The matrix  $[M(t)]$  corresponds to the unit circle (B). For each mode, first the starting point relative to the semi-major axis is correctly positioned (C); then the original circle is transformed into an ellipse, its semi-major axis along the x-axis and semi-minor axis along the y-axis (D); and finally the ellipse is rotated to its correct position (E). See further details in the Supplementary Materials and Methods.

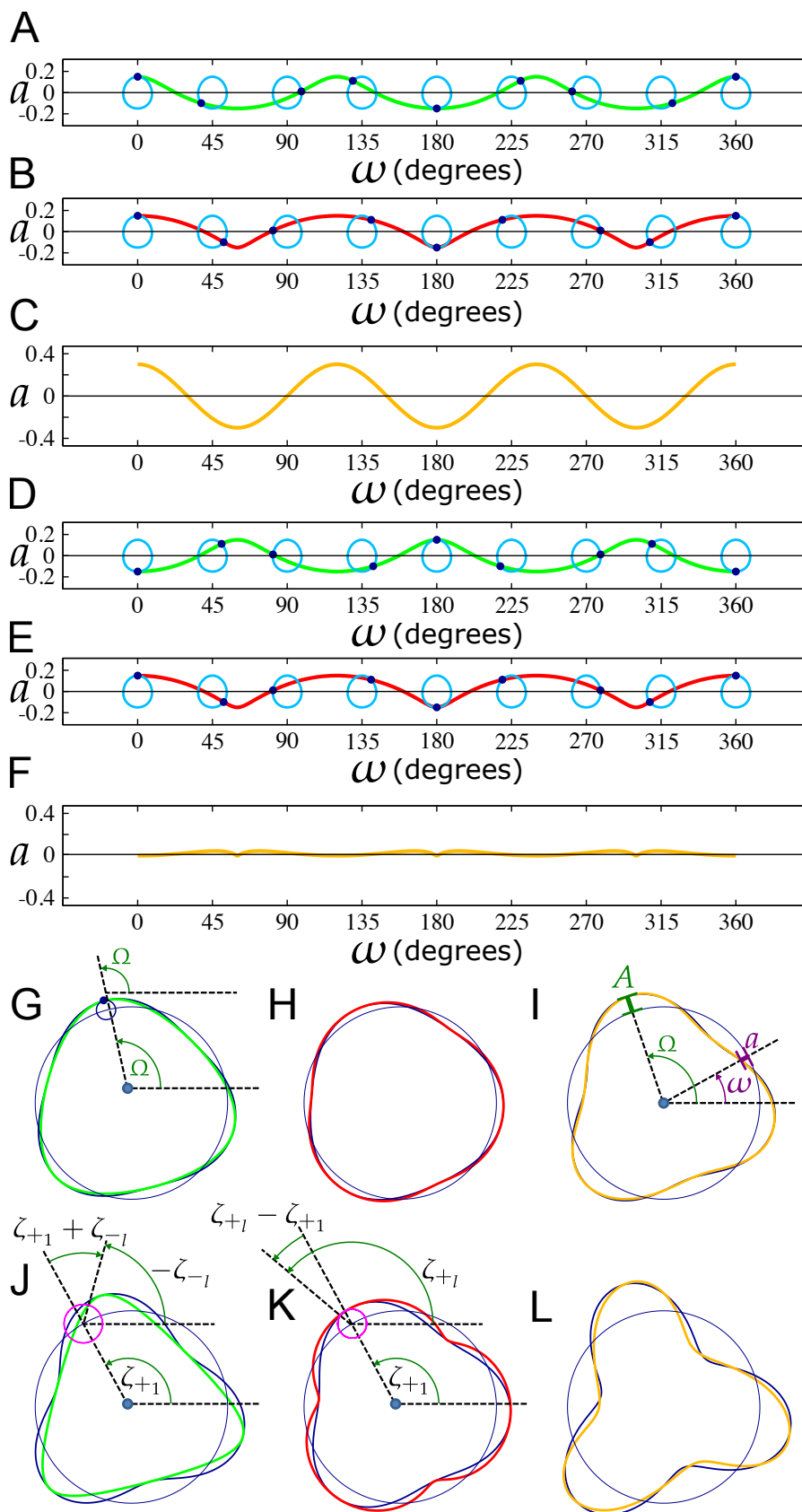


Figure S10



Figure S10: **Impact of LOCO-EFA starting points  $\zeta_{+l}$  and  $\zeta_{-l}$  on the amplitude of the reconstructed shape.** All panels show  $\mathcal{L}_3 = (\lambda_{+3}, \lambda_{-3}, \zeta_{+3}, \zeta_{-3})$  superimposed on  $\mathcal{L}_1 = (1, 0, \zeta_{+1}, 0)$ , i.e., the unit circle. In (A–F),  $\mathcal{L}_1 = (1, 0, 0, 0)$ . (A) Amplitude  $a$  as a function of phase  $\omega$  (as depicted in I) for the negative rotor only ( $\mathcal{L}_3 = (0, 0.15, 0, 0)$ ), shown in green. For several phases also the rotor (light blue) itself is drawn. Note that by plotting the rotor in the  $(\omega, a)$  plane, the circles are slightly deformed. (B) Same for the positive rotor only ( $\mathcal{L}_3 = (0.15, 0, 0, 0)$ ), in red. (C) Same for both rotors superimposed ( $\mathcal{L}_3 = (0.15, 0.15, 0, 0)$ ), in orange. (D–F) Alike (A–C), but for an out-of-phase starting angle of the negative rotor. (D)  $\mathcal{L}_3 = (0, 0.15, 0, \pi)$ , in green; (E)  $\mathcal{L}_3 = (0.15, 0, 0, 0)$ , in red; (F)  $\mathcal{L}_3 = (0.15, 0.15, 0, \pi)$ , in orange. In (G–L),  $\mathcal{L}_1 = (1, 0, \frac{2}{3}\pi, 0)$ . (G) Pattern generated by the negative rotor only ( $\mathcal{L}_3 = (0, 0.1, 0, -\frac{5}{12}\pi)$ ), shown in green, compared to the equation  $a_{-l} = \lambda_{-l} \cos(l\omega + \zeta_{-l} - (l-1)\zeta_{+1})$  (Equation 44b), in dark blue. They present a close match. (H) Same for the positive rotor only ( $\mathcal{L}_3 = (0.075, 0, \frac{3}{4}\pi, 0)$ ), shown in red, compared to  $a_{+l} = \lambda_{+l} \cos(l\omega + \zeta_{+l} - (l+1)\zeta_{+1})$  (Equation 44a), show in dark blue, again presenting a close match. (I) Same for both rotors superimposed ( $\mathcal{L}_3 = (0.075, 0.1, \frac{3}{4}\pi, -\frac{5}{12}\pi)$ ), shown in orange, and  $a_l = \sqrt{\lambda_{+l}^2 + \lambda_{-l}^2 + 2\lambda_{+l}\lambda_{-l} \cos(\zeta_{+l} - \zeta_{-l} - 2\zeta_{+1})} \cos(l\omega + \zeta_l)$  (Equation 45), show in dark blue, again presenting a close match. Panel (G) also illustrates that at peak amplitude the phase of the main and subrotor are equal; panel (I) also illustrates the concepts amplitude ( $a$ ); peak amplitude ( $A$ ); phase ( $\omega$ ); and phase at peak amplitude ( $\Omega$ ). (J–L) Alike (G–I), but for larger amplitude. (J)  $\mathcal{L}_3 = (0, 0.2, 0, -\frac{5}{12}\pi)$ , in green; (K) ( $\mathcal{L}_3 = (0.15, 0, \frac{3}{4}\pi, 0)$ ), in red; (L)  $\mathcal{L}_3 = (0.15, 0.2, \frac{3}{4}\pi, -\frac{5}{12}\pi)$ , in orange. Panel (J) and (K) also illustrate how the initial phase shifts with respect to the amplitude contribution of the negative and positive rotor depend on  $\zeta_{+1}$ , and on  $\zeta_{-l}$  and  $\zeta_{+l}$ , respectively. They are given by  $\zeta_{+1} + \zeta_{-l}$  for the negative rotor and by  $\zeta_{+l} - \zeta_{+1}$  for the positive rotor.

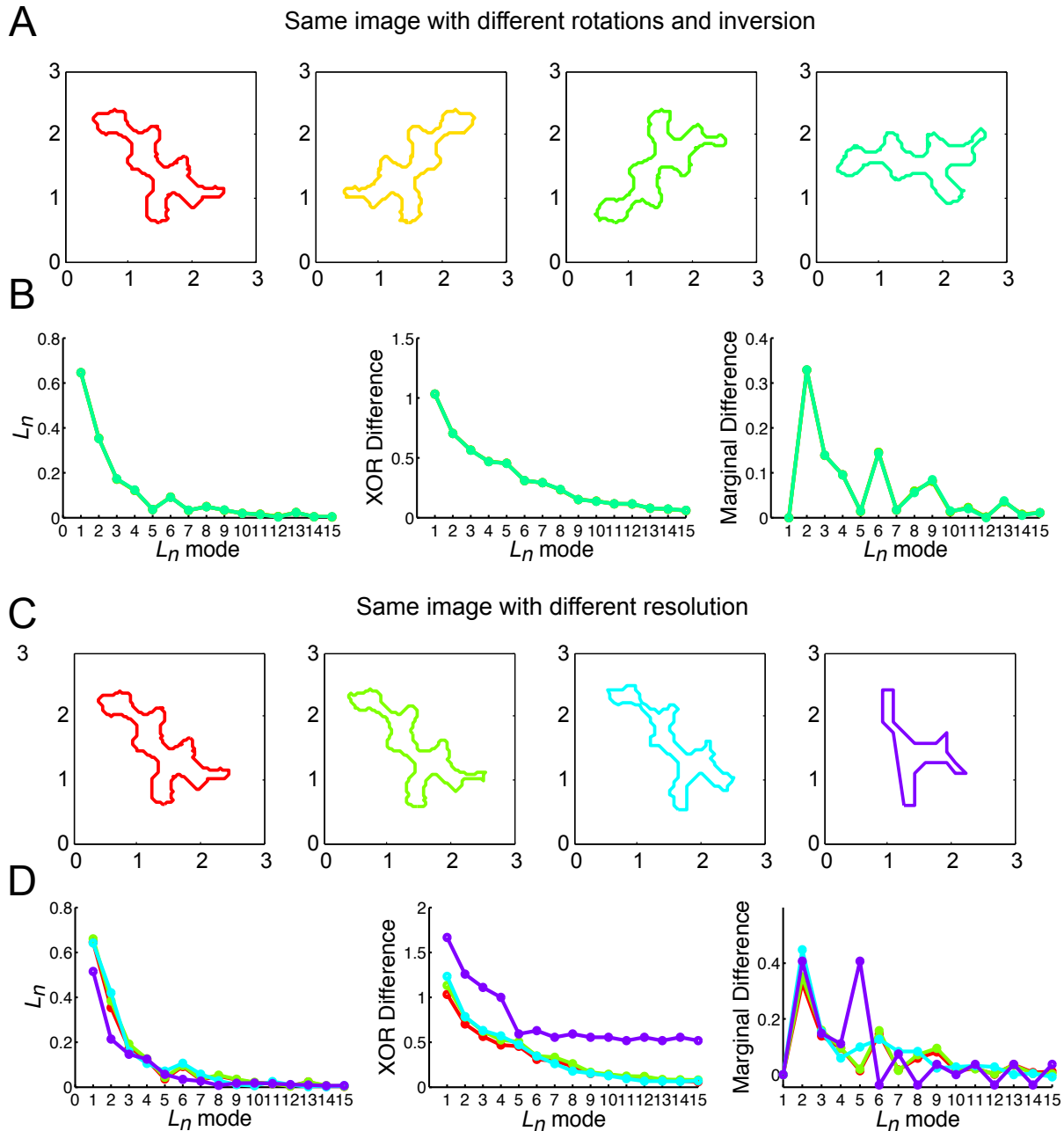


Figure S11: **LOC-EFA is invariant to image rotation or mirroring, nor sensitive to image resolution.** (A) The cell outline of an experimentally observed cell was mirrored along the y-axis and/or rotated over different angles, after which LOC-EFA was applied to each image separately. (B) The  $L_n$  values and other derived metrics were invariant to those transformations. (C) The resolution of the original image was reduced, such that the number of contour points decreased from 1104 to 253, 108 and 27 (from left to right, respectively). (D) The  $L_n$  numbers and the associated metrics only deviated from the high-resolution values when the resolution was very low and the cell outline was clearly deviating from original cell outline. This in contrast to the skeletonisation method, for which even marginal resolution reductions can cause large deviations in the outcome.

## Supporting Table

Table S1: Three types of specified cell shapes (SCS) as used in Figure S7 and Figure S8.

	SCS1	SCS2	SCS3
Target Area	858	1167	1197
Pointedness	6912	5328	5207
Number of lobes	3–10	3–10	3–10
Roundness	382	518	434
Elongation	4	28	5927

## Supporting Movies

**Movie S1: Approximation of a closed contour using Elliptic Fourier Analysis.**

A given two-dimensional shape can be approximated using EFA by summing  $n$  elliptic harmonics as follows: each  $n$ th elliptic harmonic traces  $n$  clockwise or counter-clockwise revolutions while moving around the previous elliptic harmonic.

**Movie S2: Direction of rotation opposite to the first harmonic ellipse.** EFA mode 3 generating a shape with 4 features, its rotation direction opposite to the rotation direction of the first harmonic.

**Movie S3: Direction of rotation same as the first elliptic harmonic.** EFA mode 3 generating a shape with 2 features, its rotation direction the same as the rotation direction of the first harmonic.

**Movie S4: Exception of the rule regarding rotation direction.** If the eccentricity of an elliptical harmonic is very high, the number of generated lobes does not follow the rule-of-thumb proposed by Diaz et al. (1990). Here, EFA mode 3 generates a shape with 4 features, although its rotation direction is the same as the rotation direction of the first harmonic.

**Movie S5: Example of an *in silico* simulation of a population of cells with more complex specified shapes.** CPM simulation of multi-lobed cells take up pavement-like cell shapes when they are allowed to interact with their neighbours within a confluent tissue.

## Supporting Code

**loco-efa\_example.c** Code written in C which applies the full procedure to calculate the  $\mathcal{L}_l$  and  $L_l$  values to any file containing contour coordinates. For each computational step it is indicated where the relevant mathematical details can be found in the Supplementary Materials and Methods. We have intentionally kept the code as bare as possible, without for example any graphical interface, to allow it to be trivially compiled and run on any platform. Details regarding compilation and execution can be found in the header of the file.

**cell\_outline.csv** Contour data of the cell presented in Figure 3S, Figure S3A and Figure S11. This file, or any other file containing contour data, can be used as an input for the program.

## Supporting References

- Pincus Z, Theriot JA (2007) Comparison of quantitative methods for cell-shape analysis. *J Microsc* 227:140–156. doi:10.1111/j.1365-2818.2007.01799.x.
- Schmittbuhl M, Allenbach B, Le Minor JM, Schaaf A (2003) Elliptical descriptors: some simplified morphometric parameters for the quantification of complex outlines. *Math Geol* 35:853–871.
- Kuhl FP, Giardina CR (1982) Elliptic Fourier features of a closed contour. *Comput Gr Image Process* 18:236–258.
- Diaz G, Quacci D, Dell’Orbo C (1990) Recognition of cell surface modulation by elliptic Fourier analysis. *Comput Methods Programs Biomed* 31:57–62.
- Yoshioka Y, Iwata H, Ohsawa R, Ninomiya S (2005) Quantitative evaluation of the petal shape variation in *Primula sieboldii* caused by breeding process in the last 300 years. *Heredity (Edinb)* 94:657–663. doi:10.1038/sj.hdy.6800678.
- Frieß M, Baylac M (2003) Exploring artificial cranial deformation using elliptic Fourier analysis of Procrustes aligned outlines. *Am J Phys Anthropol* 122:11–22. doi:10.1002/ajpa.10286.
- Neto JC, Meyer GE, Jones DD, Samal AK (2006) Plant species identification using Elliptic Fourier leaf shape analysis. *Comput Electron Agric* 50:121–134.
- Iwata H, Niikura S, Matsuura S, Takano Y, Ukai Y (1998) Evaluation of variation of root shape of Japanese radish (*Raphanus sativus* L.) based on image analysis using elliptic Fourier descriptors. *Euphytica* 102:143–149.
- Iwata H, Ebana K, Uga Y, Hayashi T, Jannink JL (2010) Genome-wide association study of grain shape variation among *Oryza sativa* L. germplasms based on elliptic Fourier analysis. *Mol Breeding* 25:203–215.
- Chitwood DH, Kumar R, Headland LR, Ranjan A, Covington MF, et al. (2013) A quantitative genetic basis for leaf morphology in a set of precisely defined tomato introgression lines. *Plant Cell* 25:2465–2481. doi:10.1105/tpc.113.112391.
- Haines AJ, Crampton JS (2000) Improvements to the method of Fourier shape analysis as applied in morphometric studies. *Palaeontology* 43:765–783.