SCIENTIFIC REPORT

TRAINING WORKSHOP INTERDISCIPLINARY LIFE SCIENCES

21−25 October, 2013

Lorentz Center, Leiden, the Netherlands

Gordon Akudibillah[1], Sonja E. M. Boas[2], Benoit Carreres[3], Marchien G. Dallinga[4], Aalt-Jan van Dijk[5], Shishir K. Gupta[6], Alexander Hunt[7], Annika Jacobsen[8], Kasbawati[9], Roeland M. H. Merks[2,10], Jaap Molenaar[5], Gosse Overal[11], Margriet M. Palm[2,*], Thijs Perenboom[12], Mridula Prasad[13], Maarten Reijnders[3], Elisabeth G. Rens[2], Brendan Ryback[14], Linda Scheider[15], Mugdha Srivastava[6], Marcel Willemsen[16], Wu Xunxun[17], Antonios Zagaris[18]

1. Department of Mathematics, Oregon State University, Corvallis, USA; 2. Life Sciences Group, Centrum Wiskunde & Informatica, Amsterdam, The Netherlands; 3. Department of Agrotechnology and Food Sciences, Wageningen University, Wageningen, the Netherlands; 4. Ocular Angiogenesis Group, department of ophtalmology, Academic Medical Center, Utrecht, The Netherlands; 5. Biometris, Department for Mathematical and Statistical Methods, Wageningen University, Wageningen, The Netherlands; 6. Department of Bioinformatics, Biocenter, University of Würzburg, Würzburg, Germany; 7. Department of Mathematics, University of Kaiserslautern, Kaiserslautern, Germany; 8. Department of Bioinformatics, Free University, Amsterdam, The Netherlands; 9. Department of Mathematics, Bandung Institute of Technology, Bandung, Indonesia; 10. Mathematical Institute, Leiden University, Leiden, The Netherlands; 11. Department of Mathematics, Free University, Amsterdam, The Netherlands; 12. Department of Neurology, Leiden University Medical Center, Leiden, The Netherlands 13. Biostatistics and Data Management group, Fondazione Edmund Fmach, San michele alládige, Italy; 14. Institute of Molecular Systems Biology, ETH Zurich, Zurich, Switzerland; 15. Faculty of Mathematics and Natural Sciences, Behavioural Ecology and Self-organization, Rijksuniversiteit Groningen, Groningen, The Netherlands; 16. Department of Clinical Epidemiology, Academic Medical Center, Amsterdam, The Netherlands; 17. Department of Mechanical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands; 18. Department of Mathematics, University of Twente, Enschede, The

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.654v1 | CC-BY 4.0 Open Access | rec: 2 Dec 2014, publ: 2 Dec 2014

Netherlands; * Present Adress: Multicellular systems biology group, Institut National de Recherche en Informatique et en Automatique, Rocquencourt/Paris, France

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.654v1 | CC-BY 4.0 Open Access | rec: 2 Dec 2014, publ: 2 Dec 2014

Introduction

This preprint is the outcome of the "Training Workshop Interdisciplinary Life Sciences", held in October 2013 in the Lorentz Center, Leiden, The Netherlands[1]. The aim of the workshop was to bring together young Life Science researchers and to train them in modelling of biological systems. The purpose was to have participants with various backgrounds. As can be seen from the affiliations of the authors of this preprint, this goal has certainly be achieved. The didactic philosophy of the workshop was based on 'learning by doing'. This has been achieved by form multidisciplinary groups which tackled carefully selected open problems from modern Life Sciences. Each team worked together during one week and was supervised by an experienced senior scientist. The groups started with identifying the relevant system parameters and continued with putting the problem under consideration in quantitative and predictive models. Important aspect of the workshop was that the participant were not only involved in the problem of his/her own group. During daily sessions, in which the teams reported their progress and strategies to each other, the participants learned from all modelling cases. This setup of actively involving young researchers in problem solving was supported by lectures in which techniques and concepts necessary to address modelling problems are presented.

The motivation to organize this event stems from the following considerations, taken from the workshop proposal: "The 21st century is called the 'Century of Life Sciences', since it is expected that great breakthroughs in this field will be achieved in the coming decades, such as personalized medicine, strongly enhanced food production, and sustainable energy. This positive attitude towards the Life Sciences is inspired by the huge increase in measuring power in the past decade. The enormous progress in laboratory techniques and facilities leads to the availability of huge amounts of data at all levels of complexity (molecules, cells, tissues, organs, organisms, populations, ecosystems). Especially data at the cellular level reveal details of life processes we were unconscious of until recently. However, it becomes clear that huge amounts of data alone do not automatically lead to understanding. The data explosion in Life Sciences teaches one lesson: life processes are of a highly intricate and integrative nature. To really understand the dynamic processes in living organisms one must integrate experimental data sets in quantitative and predictive models. Only then one may hope to grasp the functioning of these complex systems and be able to convert information in understanding. In the field of physics, for instance, this strong interaction between experiment and theory is already common practice since centuries, culminating in the 20th century being called the 'Century of Physics'. In contrast to physics, the complex nature of the Life Sciences

[1] https://www.lorentzcenter.nl/lc/web/2013/598/info.php3?wsid=598&venue=Snellius

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.654v1 | CC-BY 4.0 Open Access | rec: 2 Dec 2014, publ: 2 Dec 2014

forces us to work in an interdisciplinary fashion. The necessary expertise is available, but scattered over many scientific disciplines. Only the combined efforts of biologists, chemists, mathematicians, physicists, engineers, and informaticians will lead to progress in tackling the huge challenge of understanding the complexity of life. Researchers in the Life Sciences often focus their research on a rather narrow research field. However, the majority of the upcoming generation of researchers in the Life Sciences should be trained to expand their skills, becoming able to tackle complex, multi-dimensional systems. The knowledge they have to incorporate in their research will stem from a diverse range of disciplines, So, they should be trained to integrate a broad range of modelling approaches in order to deduce quantitative, predictive and often multi-scale models from highly diverse data sets. Present curricula in the Life Sciences hardly offer this kind of training yet. This workshop intends to start filling this gap. The Workshop will be considered a success if the participants will have shown to be able to set up a first mathematical model of an open problem in the Life sciences, to obtain some first model results and to interpret them biologically."

The present preprint is proof that the result we aimed at has been achieved. Three teams worked on the following open problems:

- Modeling the influence of temperature on the Regulation of flowering time in *Arabidopsis thaliana*,
- Validation of computational models of angiogenesis to experimental data,
- Reconstructing the gene network that regulates branching in Tomato.

This preprint bundles the reports of the three teams. The title page of each report lists the participants and supervisors of the team.

Modeling the influence of temperature on the regulation of flowering time in *Arabidopsis thaliana*

Scientific Team
Gordon Akudibillah[1], Thijs Perenboom[2], Annika Jacobsen[3], Shishir K Gupta[4], Mridula Prasad[5], Maarten Reijnders[6] and Marcel Willemsen[7]

Supervisor
Aalt-Jan van Dijk[5]

1. Department of Mathematics, Oregon State University, Corvallis, USA
2. Department of Neurology, Leiden University Medical Center, Leiden, The Netherlands
3. Department of Bioinformatics, Free University, Amsterdam, The Netherlands
4. Department of Bioinformatics, University of Würzburg, Würzburg, Germany
5. Biometris, Department for Mathematical and Statistical Methods, Wageningen UR, Wageningen, The Netherlands
6. Department of Agrotechnology and Food Sciences, Wageningen UR, Wageningen, The Netherlands
7. Department of Clinical Epidemiology, Academic Medical Center, Amsterdam, The Netherlands

Introduction

Plants need to be highly adjustable to environmental factors due to their static nature. The time of flowering is regulated on a genetic level and is influenced by environmental factors (Figure 1). In *Arabidopsis thaliana*, one of these environmental factors is temperature [Pose13].

Various genes are known to be involved in this regulation of flowering time. A mathematical model of this network, specific to *A. thaliana*, was proposed recently [Leal14]. The network depicted in Figure 2 shows all seven genes and the type of regulation they exercise in the current model. The model is a set of five Ordinary Differential Equations (ODEs) describing how gene concentrations change with time. Flowering is induced by elevated concentration levels of the AP1 gene. Five of the genes, namely AGL24, SOC1, FT, FD and LFY, have an activating effect on AP1, whereas the protein complex formed by FLC and SVP has an inhibitory effect on SOC1 and FT (see, again, Figure 2).

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.654v1 | CC-BY 4.0 Open Access | rec: 2 Dec 2014, publ: 2 Dec 2014

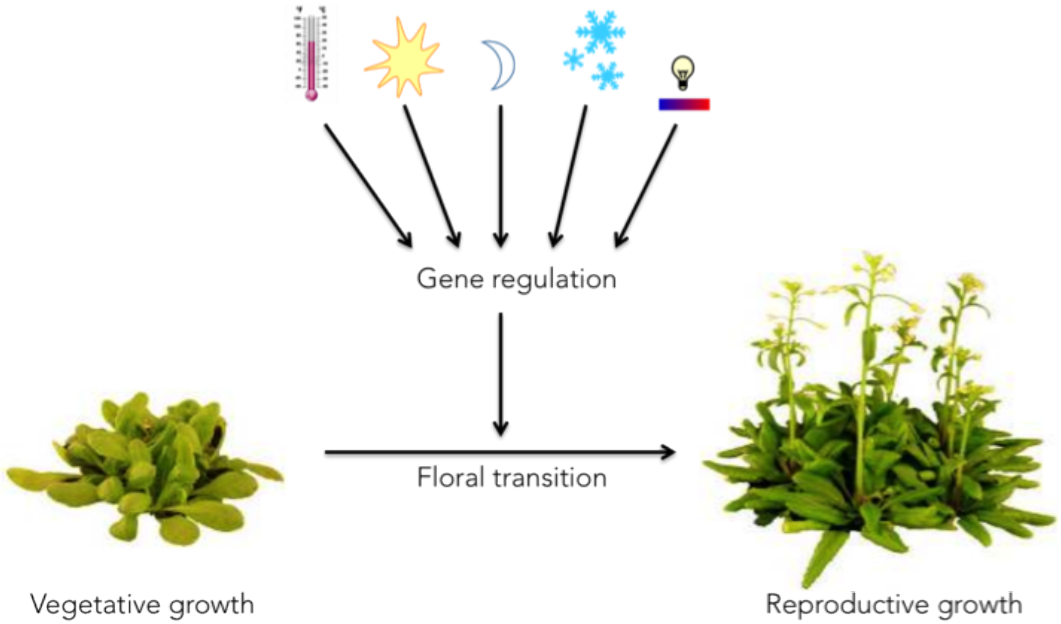

Figure 1: External factors, such as temperature and night cycle, influence the gene regulation of floral transition.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.654v1 | CC-BY 4.0 Open Access | rec: 2 Dec 2014, publ: 2 Dec 2014

The influence of temperature on the regulation of flowering time was recently shown to be directly linked to changes in amounts of two splice variants of the flowering locus M (FLM) gene, namely to FLM-β and FLM-δ [Pose13]. At low temperatures, there is more FLM-β and at high temperatures there is more FLM-δ. Both splice variants compete with FLC to bind SVP. However, only the SVP-SVP, SVP-FLM-β and SVP-FLC complexes inhibit SOC1 expression, whereas the SVP-FLM-δ complex has a counteracting effect on SOC1 repression. The relative abundance of FLM-β at low temperatures leads to a later flowering time. The situation is opposite at high temperatures, with elevated relative concentrations of FLM-δ leading to early flowering.

In this study, we aim to extend the existing model by allowing for temperature dependence. In this manner, we hope to reproduce quantitatively the effects that FLM and temperature jointly have on flowering. In addition, we employ this model to obtain insights into the regulatory mechanisms involved in flowering.

Materials and Methods

Data

Expression time series (qPCR) of FLM-β and FLM-δ at 23°C, as well as relative expression of these splice variants after 10 days at 16°C, 23°C and 27°C were obtained from [Pose13]. The equation describing how FLM-β and FLM-δ concentrations change with temperature was approximated by a linear fit between these data points.

Ordinary Differential Equations

The current model is a system of five ODEs [Leal14], of which the equation for SOC1 is as follows:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}x_{SOC1} &= \left[\frac{\beta_3 x_{AGL24}}{K_4 + x_{AGL24}} + \frac{\beta_4 x_{SOC1}}{K_5 + x_{SOC1}} + \frac{\beta_5 x_{FT,t} - \Delta}{K_6 + x_{FT,t} - \Delta}\frac{x_{FD}}{K_7 + x_{FD}}\right] \\ &\cdot \frac{K_8}{K_8 + x_{SVP,m}}\frac{K_9}{K_9 + x_{FLC,m}} - d_3 x_{SOC1}.\end{aligned}$$

We modified this equation to incorporate the effect of FLM, thus creating three models. In model 1, we replaced $x_{FLC,m}$ by $x_{FLM\beta}$. Model 2 extends model 1 by adding the activating effect of $x_{FLM\delta}$. Finally, model 3 extends model 2 by adding the repressing effect of $x_{FLC,m}$.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.654v1 | CC-BY 4.0 Open Access | rec: 2 Dec 2014, publ: 2 Dec 2014

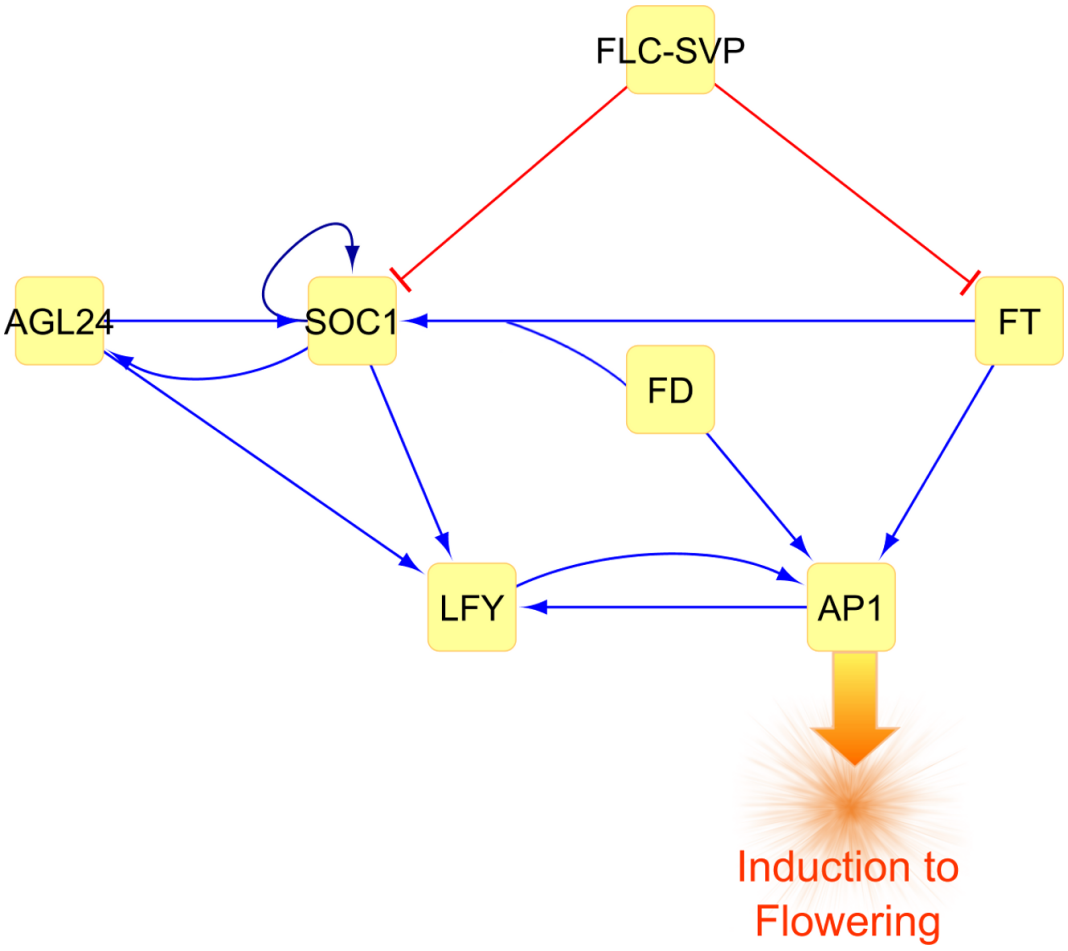

Figure 2: Current model of the gene regulatory network influencing the flowering time.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.654v1 | CC-BY 4.0 Open Access | rec: 2 Dec 2014, publ: 2 Dec 2014

$$
\begin{aligned}
\text{Model 1} \quad \frac{\mathrm{d}}{\mathrm{d}t}x_{SOC1} &= [\dots]\frac{K_8}{K_8+x_{SVP,m}}\frac{K_{16}}{K_{16}+x_{FLM\beta}} - d_3 x_{SOC1} \\
\text{Model 2} \quad \frac{\mathrm{d}}{\mathrm{d}t}x_{SOC1} &= [\dots]\frac{K_8}{K_8+x_{SVP,m}}\frac{K_{16}}{K_{16}+x_{FLM\beta}}\frac{\beta_4 x_{FLM\delta}}{K_{17}+x_{FLM\delta}} - d_3 x_{SOC1} \\
\text{Model 3} \quad \frac{\mathrm{d}}{\mathrm{d}t}x_{SOC1} &= [\dots]\frac{K_8}{K_8+x_{SVP,m}}\frac{K_{16}}{K_{16}+x_{FLM\beta}}\frac{\beta_4 x_{FLM\delta}}{K_{17}+x_{FLM\delta}} \\
&\quad + \frac{K_8}{K_8+x_{SVP,m}}\frac{K_9}{K_9+x_{FLC,m}} - d_3 x_{SOC1}
\end{aligned}
$$

Calibration

We calibrated the model with the expression time series (qPCR) of FLM-β and FLM-δ at 23 °C. Flowering was predicted to occur at the time that AP1 expression reached a concentration of 1.3, as previously determined [Leal14]. We optimized the parameters in the FLM terms to fit the SOC1 expression previously obtained at 23 °C [Leal14].

Validation

We predicted the flowering times for a temperature range from 13°C to 30°C by incorporating temperature dependence into the concentrations of FLM-β and FLM-δ, respectively. The concentration of FLM-β and FLM-δ for each temperature (T) were calculated by multiplying the concentration at 23°C with the ratio between the concentration at T to the concentration at 23°C by using the above-mentioned linear fit for the dependency of FLM splice variant levels on temperature:

$$
x_\beta(T) = x_\beta(23)\cdot\frac{-14.59\cdot T + 469.7}{134.13} \quad \text{and} \quad x_\delta(T) = x_\delta(23)\cdot\frac{7.66\cdot T + 76.27}{99.91}.
$$

Results

Initially three different models of gene regulation of flowering time with temperature dependence were created, see Figure 3. The new parameters in each model were calibrated to fit the SOC1 data; Figure 4 shows this fit for each model. To select the model that best predicts flowering time, we compared each of them to experimental data. When comparing the predicted flowering times for all three temperatures 16°C, 23°C and 27°C with experimental flowering times, we found that there was virtually no difference between the models (Figure 5). However, by predicting flowering times from 13°C to 30°C for each model (Figure 6), we saw that model 3 captured best temperature-variations in flowering time. We therefore propose that model as the best one.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.654v1 | CC-BY 4.0 Open Access | rec: 2 Dec 2014, publ: 2 Dec 2014

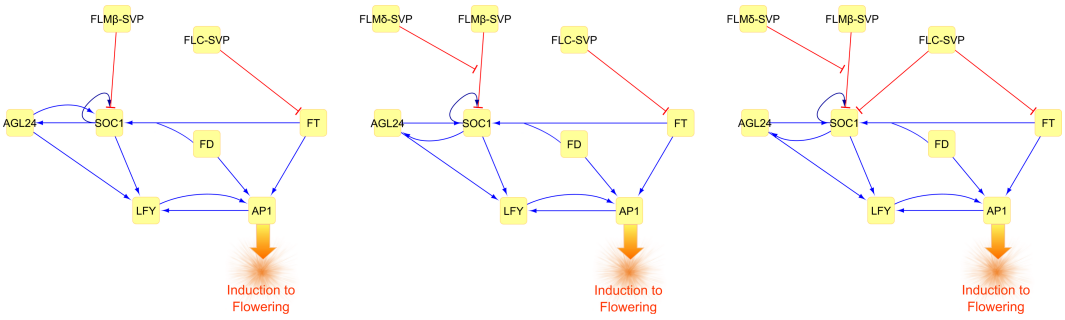

Figure 3: Three models of gene regulation of flowering time with temperature dependence.

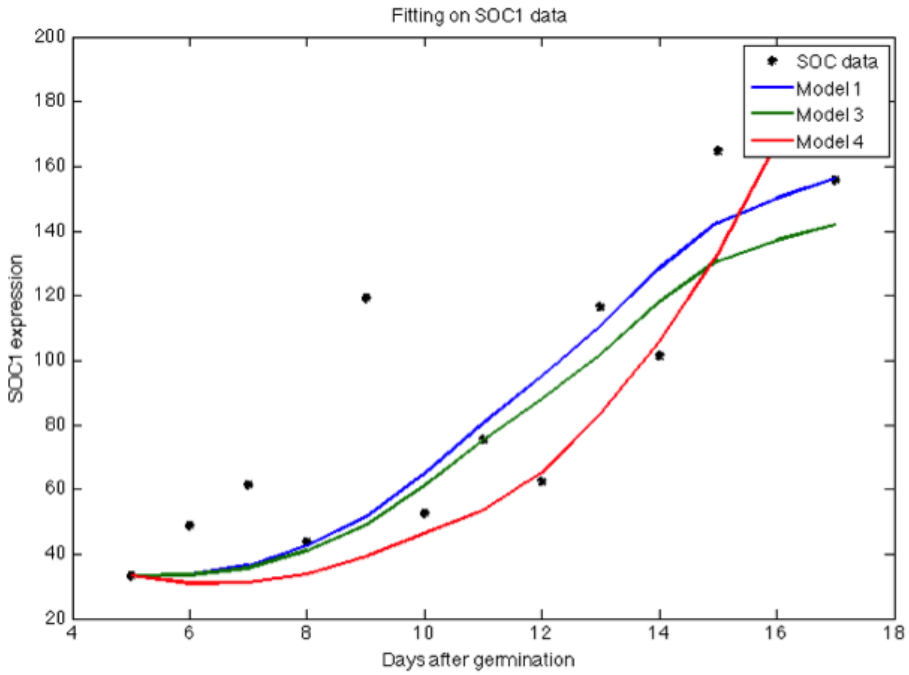

Figure 4: Parameters for different models for regulation of SOC1 by FLM are fitted using SOC1 expression data.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.654v1 | CC-BY 4.0 Open Access | rec: 2 Dec 2014, publ: 2 Dec 2014

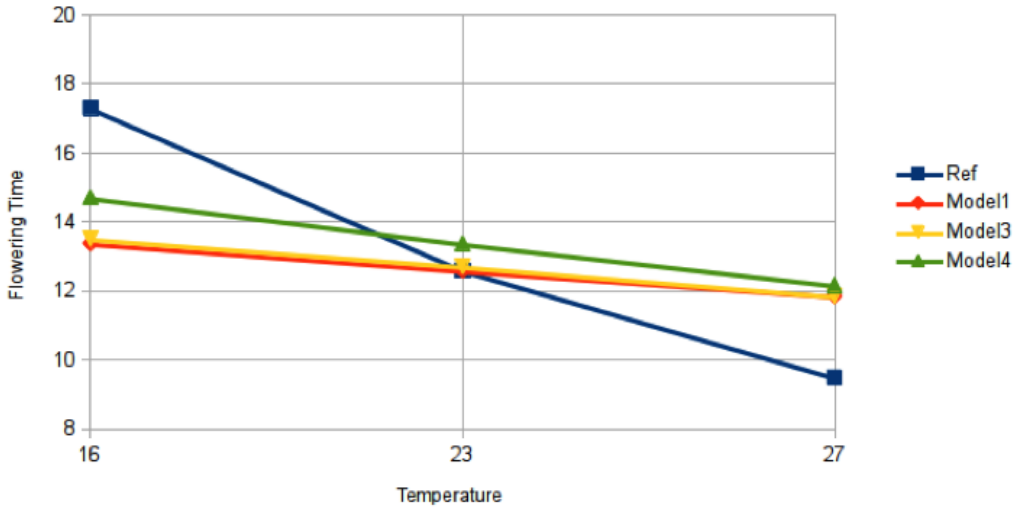

Figure 5: Flowering time at three temperatures for the three models and the experimental data.

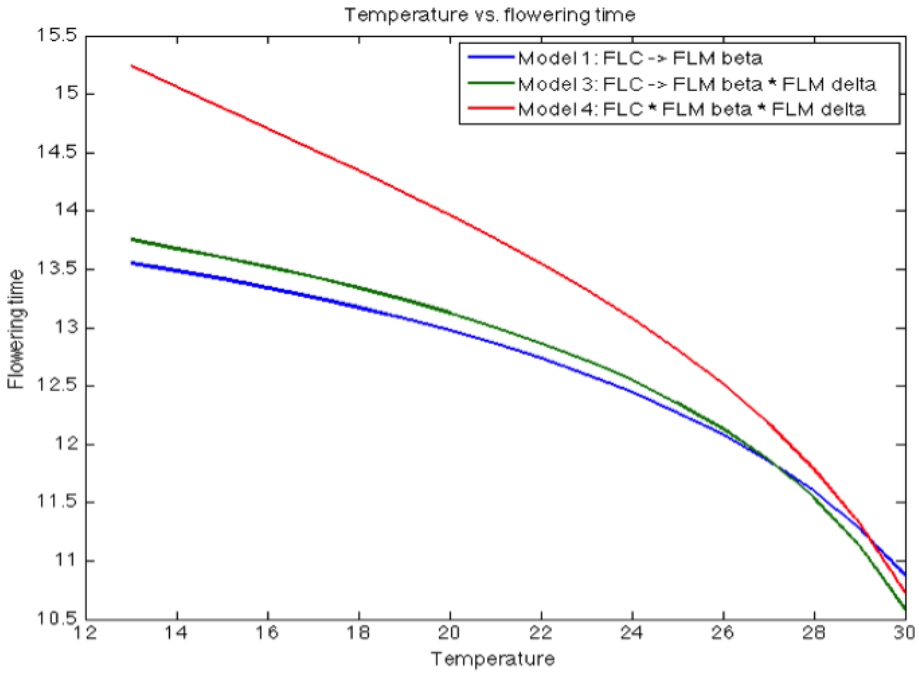

Figure 6: Flowering time versus temperature for each model.

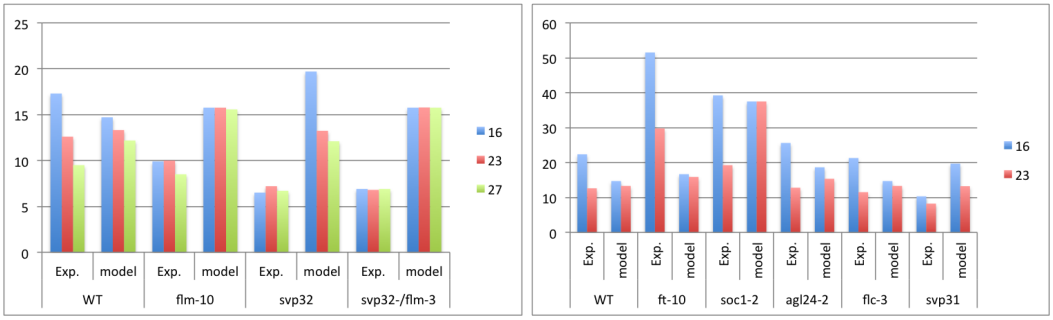

Figure 7: Data plots of two independent experimental sets. Each panel shows both the experimentally measured and the predicted flowering times for various temperatures: 16°C, 23°C and 27°C, for the left panel, and 16°C and 23°C for the right one.

Application: Modeling Mutations

We used model 3 to predict the flowering times for various mutations at 16°C, 23°C and 27°C and compared these with available experimental data [Pose13]. Loss of temperature dependence is seen both in experiment and in prediction of the FLM knockout. However, the experimental data showed much earlier flowering than our prediction. One way to obtain this behaviour in our model would be to have a stronger activation of AP1 by FT.

Conclusion

The first goal of this study was to develop a model that reproduces quantitatively the joint effects of FLM and temperature on flowering. We created three models that captured the effect of changes in temperature. However, we suggest to obtain more expression time series data for FLMβ, FLMδ and SOC1 at different temperatures to better enable fitting model parameters, and possibly discriminate between the different models. The second goal of this study was to obtain insight into the model mechanism. Our analysis shows that activation of FT on AP1 should be much stronger to correctly predict mutation effects.

Validation of computational angiogenesis models to experimental data

Scientific Team
Marchien Dallinga[1], Alexander Hunt[2], Elisabeth G. Rens[3], Brendan Ryback[4], Linda Scheider[5] and Wu Xunxun[6]

Supervisors
Sonja E. M. Boas[3] Roeland M.H. Merks[3], Margriet M. Palm[3,*]

1. Ocular Angiogenesis Group, Academic Medical Center, Utrecht, the Netherlands
2. Department of Mathematics, University of Kaiserslautern, Kaiserslautern, Germany
3. Life Sciences, Centrum Wiskunde & Informatica, Amsterdam, the Netherlands
4. Department of Agrotechnology and Food Sciences, Wageningen UR, Wageningen, the Netherlands
5. Department of Psychology, Free University of Berlin, Berlin, Germany
6. Department of Mechanical Engineering, Eindhoven University of Technology, Eindhoven, the Netherlands

∗ Present Adress: Multicellular systems biology group, Institut National de Recherche en Informatique et en Automatique, Rocquencourt/Paris, France

Neovascularization

The formation of new blood vessels, called neovascularization, is a fundamental developmental process that occurs throughout the life cycle of all vertebrates. Angiogenesis and vasculogenesis are terms referring to the formation of new blood vessels from pre-existing blood vessels or endothelial cells, respectively. As both processes are involved in organismic homeostasis and developmental diseases, a great deal of research is aimed at an improved understanding of their underlying mechanisms [Bau04]. Recently, the use of reliably reproducible *in vitro* model systems applied in conjunction with automated cell-tracking technology have facilitated the quantitative analysis of the cellular interactions giving rise to network patterns [Par11].

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.654v1 | CC-BY 4.0 Open Access | rec: 2 Dec 2014, publ: 2 Dec 2014

Cellular modeling

Novel experimental, analytical and computational methods are enabling the formulation of integrative, systems-level descriptions of biological processes across a range of scales and levels of detail. In particular, meso-scale cell-based *in silico* models of self-organizing cellular behavior have been shown to allow effective compromises to be made between mechanistic accuracy, phenomenological predictiveness and computational tractability [Mer06, Mer08]. The Glazier–Graner–Hogeweg (GGH) method (also known as "Cellular Potts Model") [Gra92, Gla93] has been implemented in a feature-rich and flexible multi-cellular simulation environment named CompuCell3D (CC3D) [Tre05, Gla13]. This tool allows a large range of cellular properties (e.g. surface area, cohesiveness, flexibility) and behaviors (e.g. chemotaxis, contact-inhibition, secretion of signaling molecules) to be translated into simple mathematical abstractions that allow complex cellular patterns to emerge.

In short, the Cellular Potts Model [Gra92, Gla93] represents cells as patches of grid sites with identical cell identifiers (σ) and cells can be differentiated by types (τ) with specific properties and behaviors. Cells can migrate by the extension or retraction of pseudopodia. Such movements at the cell membrane are mimicked by copying the state of a neighboring grid site ($\vec{x}$') to grid site ($\vec{x}$). Whether such a copy is allowed depends on a Hamiltonian energy function, which summarizes the balance of forces resulting from cell behaviors such as adhesion, repulsion, cell size, cell length and chemotaxis. For example, to approximately preserve a preferred cell length (L_{target}) for each cell, deviation of the actual cell length (L) from this target length increases the total energy of the system by addition of the following term to the energy function: $\sum_{c \in cells} \lambda_{\text{length}} \big(L(c) - L_{\text{target}}(c)\big)^2$, with λ_{length} describing the weight of the constraint. Copies that decrease the total energy of the system (E) will always be accepted; while a copy that increases the total energy will be accepted according to a Boltzmann probability function: $\text{P}_{\text{accept}}(\Delta\text{E}) = e^{\frac{-\Delta E}{\mu}}$, with the parameter μ a cell motility parameter.

Identification of current limitations

Using observations from a recently published experimental study [Par11], we aimed to modify, expand and consolidate features from two existing GGH models to more accurately replicate the reported pattern formation observed during *in vitro* vasculogenesis. Both models have been shown to reproduce vascular network formation [Rei13]. The first model, referred to as the "elongation model" [Mer06], consists of cells that elongate unconditionally as a result of an intrinsic propensity to maintain a pre-defined length. In this first model, cells also produce growth factors, which attract other cells via chemotaxis. This chemotaxis approach is the basis for the second model, known as the "contact inhibition model" [Mer08]. Here, cells also move in the direction of the growth factor gradient of growth factors produced by other cells. However, in this model, the cells only respond to this growth factor in parts of the cell its membrane that is not in contact with other cells. In other words, the chemotaxis is inhibited by cell–cell contact.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.654v1 | CC-BY 4.0 Open Access | rec: 2 Dec 2014, publ: 2 Dec 2014

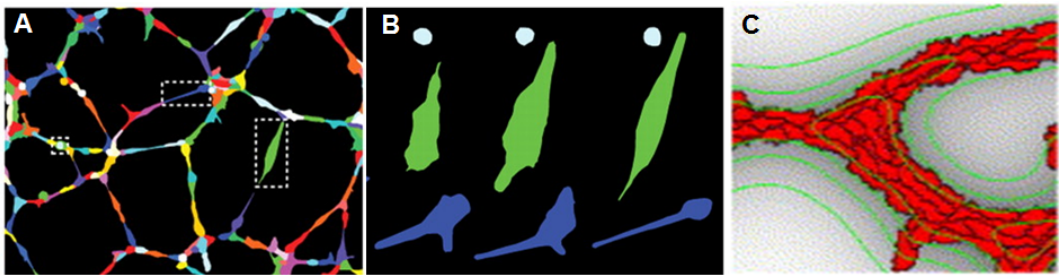

Figure 8: Comparison between in vitro and in silico pattern formation. (A) Color-enhanced vascular network and (B) detailed view of differentiated cells published in [Par11]. (C) Simulation based on the elongation model with chemotaxis [Mer08].

By comparing the *in vitro* and *in silico* cellular morphologies and emergent patterns, we identified three significant limitations of the current models:

- In the *in vitro* studies, cells undergo drastic morphological changes, while the cell size and shape in the computer model does not change substantially (Figure 8B).
- *In vitro* observed sprouts consist of one to a few cells (Figure 8A), while *in silico* sprouts comprise many (Figure 8C).
- The direction of the protrusion of pseudopods in the *in vitro* studies is more continuous than in the computer models (Figure 8B).

Model adaptations

In order to overcome the aforementioned limitations, we implemented a Cellular Potts Model with three distinct features: *Conditional elongation*, *persistence of motion*, and *3-dimensionality*. We also adjusted the field size, cell size, and cell density to approximate the reported values used in the experimental study.

Conditional elongation. The target (i.e. energetically optimal) cell length was made to be a function of the cell's exposure to the surrounding medium—essentially an adapted version of contact inhibition. Using a custom Python plug-in for CC3D, we tested both a logarithmic function and a sigmoidal function, which yielded similar network patterns (Figure 9).

Persistence of motion. Cells were biased to extend in the direction of their previous extension, which is consistent with the biological mechanism behind pseudopod formation [Bos09]. This was accomplished by calculating a motion vector derived from the difference between the center of mass of a cell before and after a stochastic step. To this end, we used the CompuCell3D built-in plugins *CenterOfMass* and *PolarizationVector*. Testing the effect of the persistence function showed less stochasticity in the movement pattern of a single cell's center of mass (Figure 10).

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.654v1 | CC-BY 4.0 Open Access | rec: 2 Dec 2014, publ: 2 Dec 2014

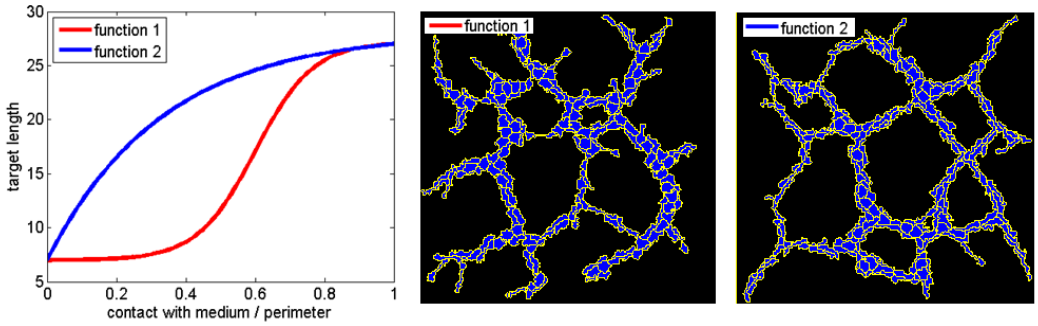

Figure 9: Comparison of two conditional elongation functions.

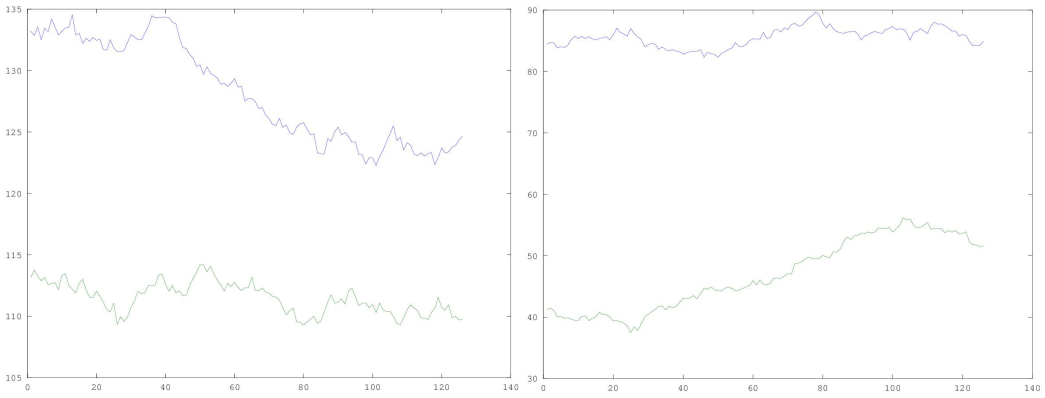

Figure 10: Change in $x-$ and $y-$coordinates of a cell's center of mass over time without (left) and with (right) the persistence function.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.654v1 | CC-BY 4.0 Open Access | rec: 2 Dec 2014, publ: 2 Dec 2014

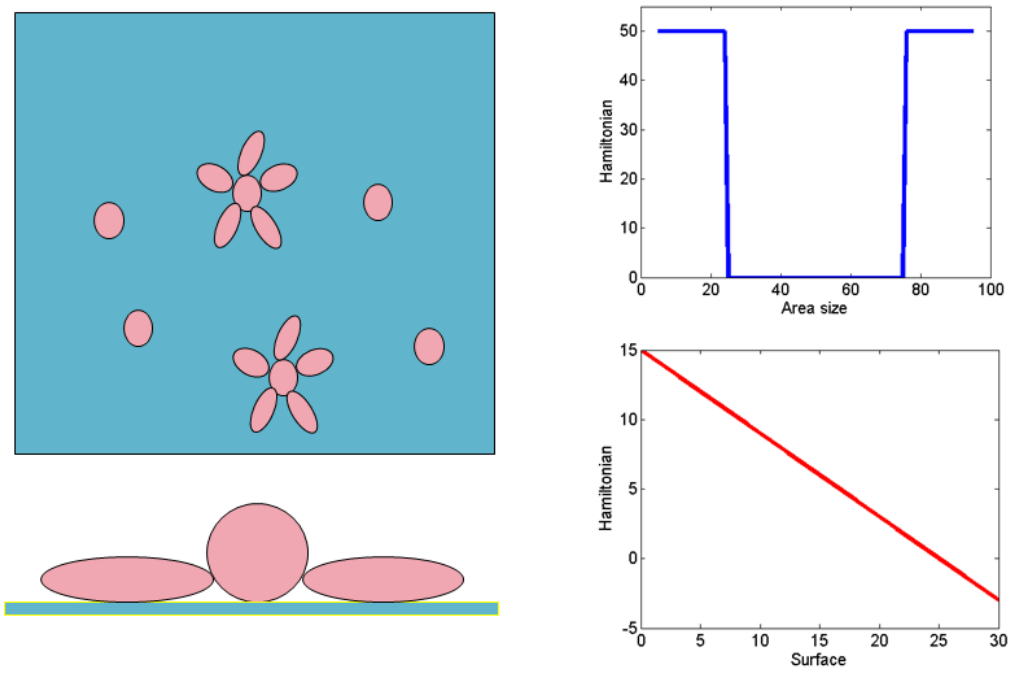

Figure 11: Introduction of a third spatial dimension. Left: schematic mock-up of cells taking on different shapes while maintaining a constant volume. Right: The calculation of the Hamiltonian is dependent on the cell area and surface (circumference).

3-dimensionality. Cells were given a height property which allowed their areas to change dynamically while maintaining a preferred cell size or target volume. The aim was to provide cells with more flexibility and facilitate the emergence of morphological diversity within populations. Rather than using an actual 3-dimensional model, which would be slower and would require tuning of the parameters, we adapted the volume constraint to mimic 3-dimensionality. In the adapted volume constraint, the energy of this term is only positive for cell sizes smaller than 25 and larger than 75 (Figure 11). It also proved necessary to add an energy term that is linearly decreasing with the surface (i.e. circumference) of the cell (Figure 11), as the repulsive forces from the extracellular space would otherwise prevent flattening from occurring, forcing cells to remain spherical.

Benchmarking results

In order to investigate whether the model adaptations resulted in more realistic network patterns, we calculated the means and standard deviations of the cell parameters *area*, *perimeter* and *box ratio* and compared them to the values reported in [Par11].

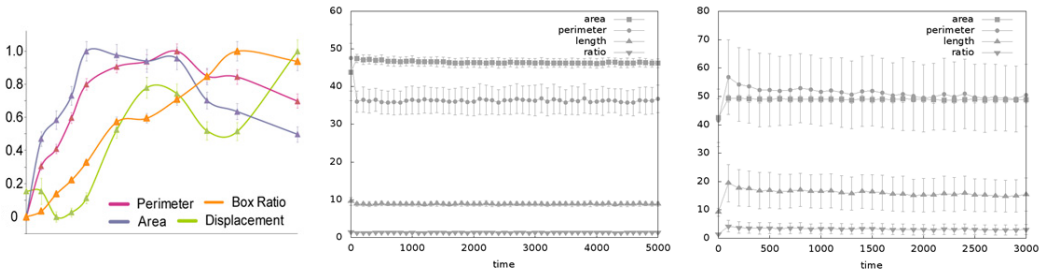

Figure 12: Measurement of cellular parameters. Left: experimental data show the evolution population-averaged parameters over the course of an experiment [Par11]. Center: dimensions of simulated cells with fixed length and proportions. Right: dimensions of simulated cells using the expanded model with dynamic cell size.

In Figure 12, the area, perimeter and box ratio are plotted for the experimental data [Par11], for dispersed cells that adhere and chemotact to a chemoattractant they secrete themselves (*de novo* model), and for the *de novo* model including the three model adaptations described above. Our adaptations to the model dramatically increase the variability of the cellular dimensions (see error bars in 12). This indicates the emergence of a heterogeneous population of cells, a fact becoming even more evident when cell area is represented as a histogram (Figure 13). Visual comparison of the *de novo* model, the elongation model [Mer06], and the adapted model with the experimental findings reveal dramatic differences (Figure 14). To illustrate the heterogeneity in cell length in the model with the combined modifications, the cells are color-coded based on their cell length.

The *de novo* model does not result in network formation at all, but forms blobs as expected. The model with long cells reported in [Mer06] produces a realistic network at first glance, but closer inspection reveals that most sprouts consist of bundles of cells. Experimental findings clearly show, nevertheless, that most sprouts consist of one to a few cells that do not form bundles. Viewed in these terms, the new model, comprising 3-dimensionality, persistence of motion and conditional elongation rests somewhere in between. While the overall network formed is clearly not as well-connected, it contains a more realistic distribution of cell shapes and most sprouts do not consist of bundles of cells. We therefore conclude that the new model represents a minor but not trivial improvement over the other models.

Limitations and recommendations

Due to the limited duration of the workshop, it was not possible to rigorously test the model's parameters to determine whether better network structures could be produced. One of our objectives was to fit our model's parameters to the *in vitro* data based on the area, perimeter, and box ratio, and then to use the measurement of cell displacement as a benchmark to validate the model. Unfortunately, this was not possible during the workshop, due to the amount of time required to code the new features.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.654v1 | CC-BY 4.0 Open Access | rec: 2 Dec 2014, publ: 2 Dec 2014

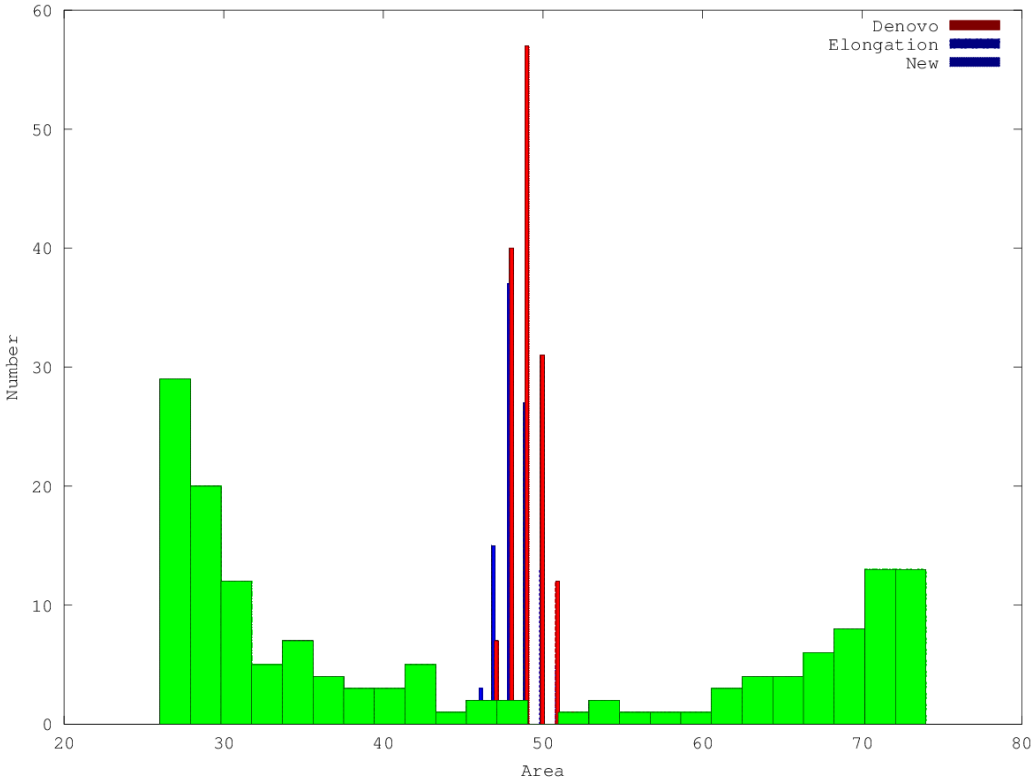

Figure 13: Histogram comparing distribution of cellular area between models.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.654v1 | CC-BY 4.0 Open Access | rec: 2 Dec 2014, publ: 2 Dec 2014

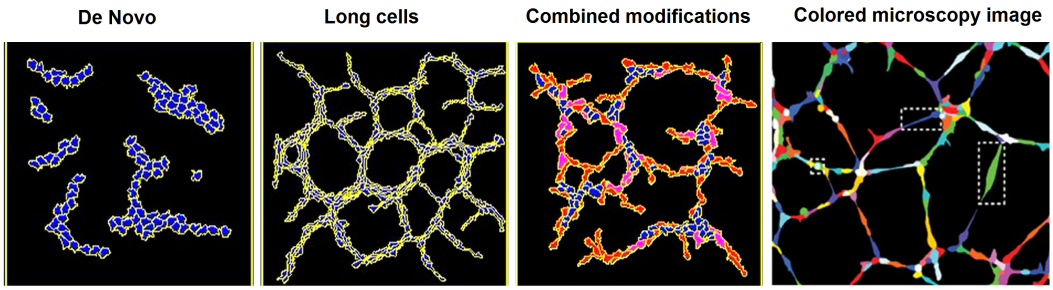

Figure 14: Comparison of simulations based on different models (far left to second from right) with experimental findings (far right). Note that the coloring of the cells in the *Combined modifications* image corresponds to cell length, to emphasize that long cells (red) form sprouts whereas round cells are mostly surrounded by others (blue). The colors in the far right image are merely used to differentiate individual cells, not cell types.

Overall, our implementation of new model features represents an advancement in terms of the approximation of dynamic variability in cellular morphology that clearly plays an important role in vasculogenic pattern formation. A common criticism directed to such models is that they emphasize phenomenology over fundamental mechanisms. While our model is not exempt from similar criticism, we feel that the implementation of 3-dimensionality and persistence of motion are largely consistent with physical realities and known biological mechanisms, respectively. The mechanistic validity of our conditional elongation function requires closer scrutiny. This feature may, indeed, be somewhat oversimplified. However, as it is conceptually derived from an established model ("contact inhibition"), we felt it to be an acceptable solution within the context of this course. Visual inspection of the simulated patterns suggests that conditional elongation is the most significant contributor to network formation. However, this assessment is inherently speculative. Quantitatively determining the impact of the three adaptations would require a rigorous sensitivity analysis to be performed.

It is clear that future research on cellular Potts models will not be able to ignore 3-dimensionality, as it provides cells with a level of morphological flexibility that is necessary to emulate the wide range of cellular shapes observed *in vitro*. Currently, cells with multiple pseudopods and "pointy" pseudopods do not emerge *in silico*. Achieving this will most likely require more sophisticated models of chemotaxis and cellular elongation.

Reconstructing the gene network that regulates branching in Tomato

Scientific Team
Antonios Zagaris[1], Benoit Carreres[2], Gosse Overal[3], Kasbawati[4] and Mugdha Srivastava[5]

Supervisor
Jaap Molenaar[6]

1. Department of Mathematics, University of Twente, Enschede, The Netherlands
2. Department of Agrotechnology and Food Sciences, Wageningen UR, Wageningen, The Netherlands
3. Department of Mathematics, Free University, Amsterdam, The Netherlands
4. Department of Mathematics, Bandung Institute of Technology, Bandung, Indonesia
5. Department of Bioinformatics, University of Würzburg, Würzburg, Germany
6. Biometris, Department for Mathematical and Statistical Methods, Wageningen University, Wageningen, The Netherlands

Introduction

It is known that the architecture of inflorescence meristems highly influences flower production and crop yields. In this case study on the tomato plant Solanum lycopersicum, we will focus on a specific type of branching called *sympodial*. Little is yet known of the mechanisms underlying sympodial growth and branching diversity.

To better understand these mechanisms, we make use of the data from the experimental study described in [Par12]. the authors of this study performed deep transcriptome sequencing (*Illumina RNAseq*). Their objective was to capture gene expression dynamics from individual shoot meristems during their transition from the vegetative state to the final flower state. Measurements were taken in five instants: middle vegetative meristem (MVM), late vegetative meristem (LVM), transition meristem (TM), sympodial inflorescence (SIM) and finally flower meristem (FM). These measure points correspond to days 10, 13, 15, 16 and 17 in the developmental process. Three crops were chosen to perform the experiments: first, *Solanum lycopersicum*, also called *money maker*, a commonly used tomato variant, widely grown for

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.654v1 | CC-BY 4.0 Open Access | rec: 2 Dec 2014, publ: 2 Dec 2014

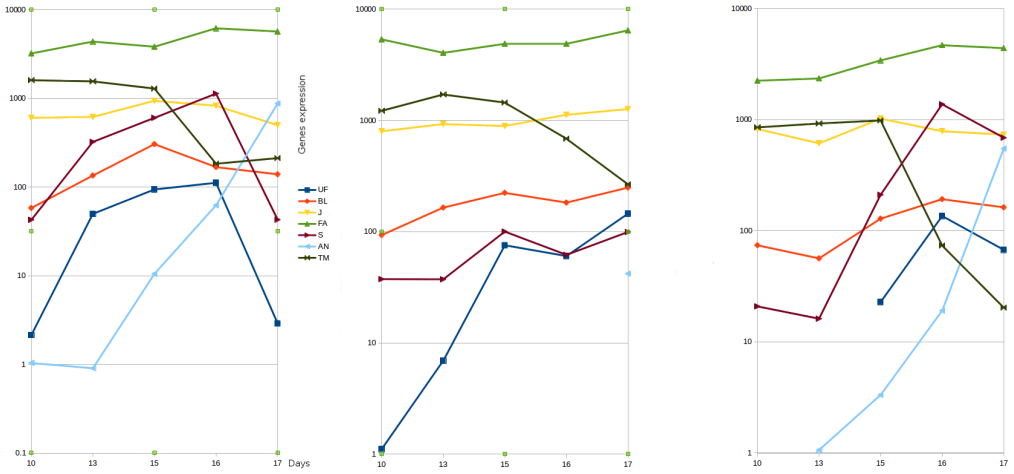

Figure 15: Plot of all seven gene expression levels versus time for the three tomato variants.

commercial purposes; second, a mutant, whose inflorescence branching gene S is mutated and its activity reduced; and third, *Solanum peruvianum*, originating in Peru and one of the first wild tomatoes to have been cultivated in European botanical gardens.

In the literature a lot of discussion is going which genes are involved in the branching process, but expert knowledge assured us that at least the following seven candidates are known to be involved in the inflorescence development of tomato: FALSIFLORA (FA), ANANTHA (AN), the inflorescence branching gene S, BLIND (BL), UNIFLORA (UF), JOINTLESS (J), and TMF. The measured expression dynamics of these seven genes are taken from [Par12] and presented in Figure 15. These data were the starting point for further analysis.

Research Question

Starting from the seven genes that are believed to form the network regulating branching in tomato, the aim of this study is to find the potential interactions between these seven candidates. These regulations may be positive (up regulation) but also negative (down regulation). In terms of network theory: we have a network with seven nodes and our aim is to find the edges. So, the question is whether there is an edge between any pair of nodes, and if there is one, to establish whether it is a promoting or an inhibiting interaction. We do not aim at finding other details of the interactions. In view of the poor data, the present aim is already hard enough.

Having three data sets available, we decided to use two of them, namely Solanum lycopersicum and its S-mutant, for calibration (estimation of the model parameters), and the third one, namely Solanum peruvianum, for validation. The choice of Solanum peruvianum was the best choice to validate the model, as it can be considered as the *wild type*. Furthermore, building up a mathematical model using two data sets that are very close biologically (only

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.654v1 | CC-BY 4.0 Open Access | rec: 2 Dec 2014, publ: 2 Dec 2014

one mutant gene) would make the setup more logical and allow to finally estimate how powerful the model in predicting a less related specimen.

Methodology and modeling choices

We decided to follow two fundamentally different approaches in reconstructing the gene regulatory network (GRN), namely to use either knowledge available in biological databases and/or the literature, on one hand, and mathematical modeling on the other. We spent most of the time on the latter, since the first approach yielded scarce information that did not suffice to fully reconstruct the GRN.

Database exploration

The interaction information from the different databases was explored, in the hope that some biological hints and useful insights could be generated for a modeling approach. We employed the protein interaction databases (DBs) *String v9.1* [Fra13] and *iHop* [Hof04]. The *String* DB quantitatively integrates interaction data from genomiccontext, high-throughput experiments, conserved co-expression and previous knowledge from *Pubmed* and other resources. The *iHop* DB provides a natural way of accessing millions of *Pubmed* abstracts. The genes and proteins appearing in *iHop* act as hyperlinks between sentences and abstracts, and the relevant information can be readily fetched from the literature. Disappointingly, we found no interaction information for the seven genes in tomato.

In the absence of such appropriate direct interaction information, wefollowed another approach called *Signalogs identification* [Kor11] based on *pathway annotation transfer*. In short, this approach advocates transferring interaction information from a certain pair of interacting genes/proteins in some species to some other, homologous pair in another species. The Plant Transcription Factor Database (*PlantTFDB*) [Jin14] was investigated to identify homologues for genes of tomato (*Lycopersicum esculentum*) present in other organisms. The best 'hits' (homologues) were found in *Arabidopsis thaliana*. The functional partners for the identified homologues were predicted using the *String* DB. The single interacting pair of genes in *Arabidopsis thaliana* was identified using the *iHop* DB and mapped to tomato, thus generating a biological hint for the mathematical modeling described below. See also Figure 16

Mathematical modeling

To capture the temporal variability of gene expression levels, we postulated a model of linear ordinary differential equations (ODEs). That is, we assumed that every gene A affects linearly the expression of every gene B, including itself. In this context, we look for connectivity in a linear ODE model $\dot{x}(t) = Ax(t)$, where the 7−D vector x collects gene expression levels and A is the connectivity matrix. This approach does not suffice for describing interaction effects in detail. We do believe it, however, sufficient in view of our restricted purpose: finding out the pairs of interacting genes and the nature of interaction (promoting/inhibiting). Hence,

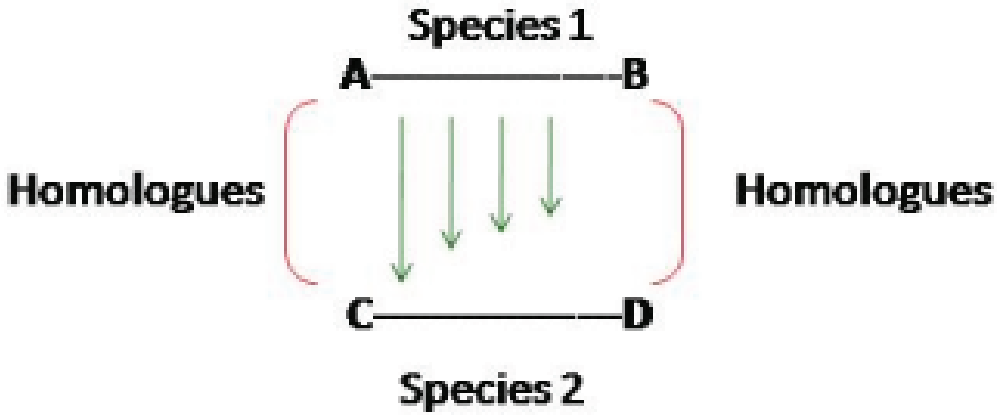

Figure 16: Pathway annotation transfer from species 1 to species 2.

we do not assume the simplifying assumption to be correct but, instead, to be accurate enough for revealing the desired connectivity characteristics.

A set of linear ODEs is described by the matrix A modeling interactions in the system.The problem, then, becomes to estimate the matrix elements from the data. In our optimization-based approach, this is done by iterating on an initial guess for these matrix elements so as to minimize a certain *cost* (or *objective*) *function*. In particular, starting from the the data points at the first time instant, the ODEs are integrated in forward time. The resulting (*predicted*) time series for the gene expressions are compared with the measured ones, and the difference between predicted and measured values is fed to standard MATLAB packages to re-estimate (*update*) the matrix elements. As an extra optimization criterion, we also attempt to introduce sparsity into the network. Indeed, it is expected that a given node has at most one or two incoming and one or two outgoing edges since, biologically speaking, the transcription is only influenced by one or two enzymes. Since zero matrix elements imply that there is no interaction between the corresponding nodes, the degree of sparsity is measured by the number of zero entries in the matrix A.

In the above description, the matrix A represents the connectivity of the gene network for the *Lycopersicum* plant. Its S-mutant has a very similar network, as it differs only in how the branching gene S affects the rest. To exploit this structure, we assumed the same matrix A with the exception of the 5^{th} column, which describes precisely the effect of S on the GRN. This column was allowed to differ from its counterpart in A by a multiplicative constant $0 < c < 1$, corresponding to a mutation which transcribes an enzyme of decreased efficacy.

Even under this simplifying assumption, there is much freedom in the model. The values of the 49 matrix entries and the factor c must be estimated, but only 56 data points are available; this approach is thus prone to overfitting. This proneness is, however, counterbalanced by the sparsity of the matrix. Due to time restrictions, we did not exploit this idea to its full extent. Instead, we restricted ourselves to penalizing non-zero matrix entries by modifying

accordingly the objective function.

Technical approaches

Approximating the ODE with forward Euler

A possible technical approach we discussed reads as follows. We start from our central model equation,

$$\dot{x}(t) = Ax(t), \tag{0.1}$$

with A a 7×7 matrix; we discretize using forward Euler:

$$\frac{x(t+\Delta t) - x(t)}{\Delta t} = Ax(t). \tag{0.2}$$

Set $A = (a)_{ij}$ for the matrix corresponding to data set 1 and $\bar{A} = (\bar{a})_{ij}$ to data set #2. Since, in data set 2 only the S gene is mutated, this mutation only affects the fifth column of $\bar{A}$ leaving the others unaffected. Thus, we may use of that A and $\bar{A}$ are identical except for the fifth column.

Next, we supplement A with the fifth column of $\bar{A}$ to form the 7×8 matrix $(A, \bar{a}_5^{\cdot})$. Then, arrange the 7 rows of this matrix into a $56-$D row vector a, starting from the top one and proceeding down.

• Define, also, the 7×1 row vector of the discretized derivative corresponding to the first time instant (t_1) of data set I:

$$\dot{x}_{I1} = \begin{bmatrix} \frac{x_I^1(t_1) - x_I^1(t_0)}{t_1 - t_0} & \cdots & \frac{x_I^7(t_1) - x_I^7(t_0)}{t_1 - t_0} \end{bmatrix}$$

• Finally define the 56×7 matrix L_I^1 as

$$L_I^1 = \begin{bmatrix} \overbrace{\begin{bmatrix} x_I^1(t_0) \\ \vdots \\ x_I^7(t_0) \\ 0 \end{bmatrix}}^{8\times 1} & & \overbrace{0}^{8\times 6} \\ & \ddots & \\ 0 & & \begin{bmatrix} x_I^1(t_0) \\ \vdots \\ x_I^7(t_0) \\ 0 \end{bmatrix} \end{bmatrix}.$$

Define, also, the 56×7 matrix L^1_{II} similarly as

$$L^1_{II} = \begin{bmatrix} \overbrace{\begin{bmatrix} x^1_{II}(t_0) \\ \vdots \\ x^4_{II}(t_0) \\ 0 \\ x^6_{II}(t_0) \\ x^7_{II}(t_0) \\ x^5_{II}(t_0) \end{bmatrix}}^{8\times 1} & & \overbrace{0}^{8\times 6} \\ & \ddots & \\ 0 & & \begin{bmatrix} x^1_{II}(t_0) \\ \vdots \\ x^4_{II}(t_0) \\ 0 \\ x^6_{II}(t_0) \\ x^7_{II}(t_0) \\ x^5_{II}(t_0) \end{bmatrix} \end{bmatrix}.$$

• The conditions obtained from discretizing the derivatives at time t_1 and for for either of the two data sets become, then,

$$a\, L^1_I = \dot{x}^1_I \quad \text{and} \quad a\, L^1_{II} = \dot{x}^1_{II}.$$

• This same process must be repeated for all 4 time instants. This result in a linear system of 56 equations in as many unknowns:

$$a\,[L^1_I \;\; L^1_{II} \;\; \ldots \;\; L^4_I \;\; L^4_{II}] = [\dot{x}^1_I \;\; \dot{x}^1_{II} \;\; \ldots \;\; \dot{x}^4_I \;\; \dot{x}^4_{II}].$$

This can now be solved. Due to lack of time we did not yet investigate the power of this idea in practice.

Fitting the ODE solutions to the data

We mainly worked directly with solutions to the ODE system (0.1). In particular, initializing the ODE system with the available data x_0 at $t = 10$, we used MATLAB's built-in integrator `ode45` to obtain numerically the corresponding solution $x(t)$. The difference between this computed solution and the available data $x_{\text{ref}}(t)$ could be computed to obtain the total square deviation (TSD). We subsequently optimized the $A-$entries to maximize agreement with data while, concurrently, penalizing each connection to avoid overpopulating our network. The technical implementation of this plan proceeded through the minimization of the *objective function*

$$O(A, \alpha, c) = \sum_{t\in\{13,15,16,17\}} ||x(t) - x_{\text{ref}}(t)||^2 + \alpha \sum_{i=1}^{7}\sum_{j=1}^{7} ||A_{ij}||^2 + p(c). \tag{0.3}$$

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.654v1 | CC-BY 4.0 Open Access | rec: 2 Dec 2014, publ: 2 Dec 2014

Here again, A generates the solution $x(t)$ corresponding to the initial condition $x_0 = x_{\text{ref}}(10)$ through (0.1). Plainly, the first term in (0.3) measures the aforementioned TSD, while the second one penalizes (strong) network connections; the relative importance of these terms is weighed by the scalar α. The third and final term relates to the multiplicative constant c modeling the effect of the mutation of S on the dynamic model. It is meant to penalize values of c that are either larger than one (gene functionality enhanced by the mutation) or negative (character of the gene functionality reversed by the mutation), as both are improbable from a physiological perspective. For that reason, we set its value to some large $p_{\max}$ for these values and to zero for $0 < c < 1$.

Scaling

It was noticed that the expressions of the individual genes attained different orders of magnitude. Since this may lead to unbalanced terms in the object function, two solutions were proposed:

- Use scaled data, by dividing the data points by the maximum value of this specific gene.
- Scale the residuals by the data-point, by replacing $x^i(t_j) - \tilde{x}^i(t_j)$ by $\frac{x^i(t_j) - \tilde{x}^i(t_j)}{x^i(t_j)}$.

The second method proved ineffective: the residual function might be scaled. The difference of scale in the data still makes the entries in the matrix A account for differences in scales. Hence the choice was made to scale the data, not the residuals.

Results

Database exploration

For four of the seven genes, homologues were found in the plant Arabidopsis thaliana. See the Table underneath.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.654v1 | CC-BY 4.0 Open Access | rec: 2 Dec 2014, publ: 2 Dec 2014

Gene Name	Transcription Factor ID	Homologue in Arabidopsis	Functional Partner (StringDatabase)
Uniflora (UF)	Solyc09g005070.1.1	AT5G01310 (BHLH Protein Family)	AT1G08125 AT5G25475 AT1G10610 WRKY36 WRKY9 WRKY35 WRKY19 POLGAMMA1 AtTLP8 AT1G804020
Blind(BL)	Solyc11g069030.1.1	AT5G57620 Myb-DNA-binding	AP4.3A, BPC4 OPT6,WRKY69 AT4G10160 WRKY39, FLA8 AT2G15580
Falsiflora(FA)	Solyc03g118160.1.1	AT5G61850 LEAFY (LFY)	AP1,FLC,GI FT,CO,ATJ3,CAL FLR1, AGL20, AGL6
S	Solyc02g077390.1.1	AT2G33880 (homeobox-3 (WUS))	TOE2, NS2 AT5G08330 KAN,WRKY2 AT5G65683 AT5G62090 AT3G54826 AT3G18380 NDF4

However, homologues were found only for the interacting pair LFY and WUS of Arabidopsis thaliana, and the interaction was transferred from it to Lycopersicum esculentum. Since LFY activates WUS, the homologue pair Falsiflora should activate S in tomato. This hint generated from databases was suggested for the further mathematical modeling paths.

ODE Modeling

The least squares problem was solved using MATLAB's `lsqnonlin` optimizer. In the absence of an appropriate estimate for a 'reasonable' $\alpha-$value, we researched the effect that α has on the *structure* of the (minimizer) A (*i.e.*, on the position of zero entries and signs of nonzero ones). For each $\alpha-$value we considered, we optimized for a multitude of pseudo-random initial choices of A, generated with MATLAB's built-in random matrix generator `rand`. In that way, we sampled the high-dimensional graph of A (*landscape*) for local minima. We found that values of α of the order of unity or higher tended to produce minimizers with

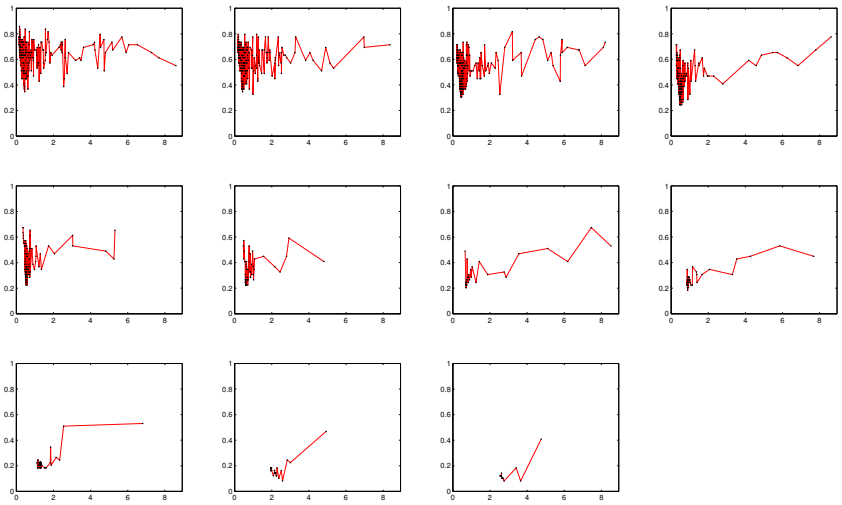

Figure 17: Value of the object function as a function of the total number of connections in the network for several $\alpha-$values. $\alpha-$values increase across panels from left to right and then from top to bottom. The $\alpha-$values are logarithmically equidistant: 0.01, 0.016, 0.025, 0.04, 0.063, 0.1, 0.158, 0.251, 0.4, 0.63 and 1. What is reported on the horizontal axis is their index, not the alpha values themselves.

all entries close to zero. The $\alpha-$penalty term dominated $O(A, \alpha, c)$ and $O(A, \alpha, c)$ was minimized close to $A \approx 0$. For such small $A-$values, $x(t) \approx 0$ and $O(A, \alpha, c)$ approximately equals the least square deviation between initial conditions and data at later times. On the contrary, $\alpha-$values approximately below 0.05 led to $O(A, \alpha, c)$ being dominated by the first term. In that case, the minimizer A was fully populated and with large entries, as the minimizer strove to fit the data; also, there appeared to be many local minima present in the landscape. We found that $\alpha-$values between 0.05 and 0.5 *reliably* produced minimizers with similar structure ($A-$entries below 1/10 in absolute value were considered to be zero). These observations are summarized in Figure 17.

In this figure we plot the value of the object function as a function of the total number of connections. This is done for 11 values of α in the range $[0.01, 1]$ and for a number of optimizations corresponding to pseudo-random initializations (300 per panel). In each panel, we have discarded results with either a $c-$value outside $(0, 1)$ or a TSD greater than that corresponding to $A = 0$ (network with no dynamics). Each connection counts for 1/49, so that the trivial network is represented by zero on the vertical axis and the fully connected network by one. For each point in each panel, all other points to its left and above may be safely discarded, as they correspond to more populated networks exhibiting worse a data fit. Under this convention, we essentially only need consider the lower envelope of each graph (not sketched) and can draw certain conclusions outlined below.

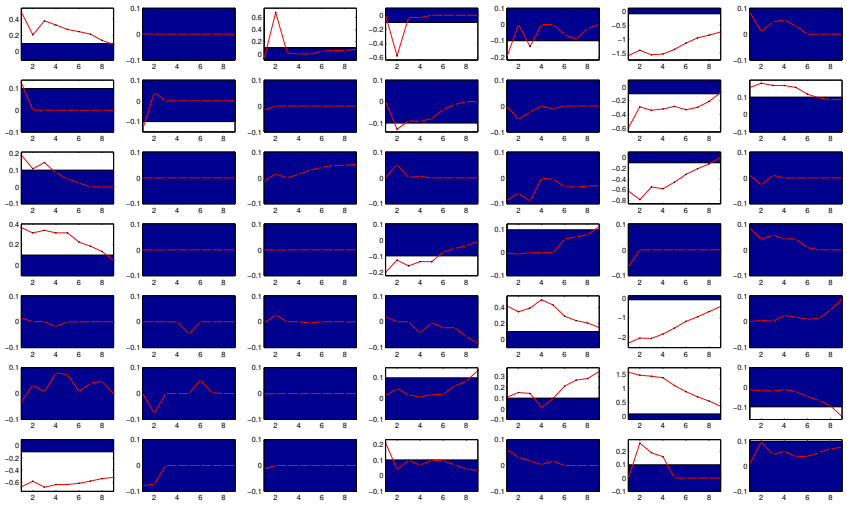

Figure 18: Entries of the matrix A for the most likely network for varying values of α.

Plainly, small $\alpha-$values lead to many local minima with a high number of connections; see also our remark above. Interestingly, a rather robust result is that, for $\alpha-$values in the aforementioned range $[0.05, 0.5]$, there exists a minimizer with connectivity approximately 1/5 (7 out of 49 possible connections) with the following property: minimizers with smaller TSD also have higher connectivity. This minimizer corresponds, in all cases, to the same $A-$structure. We consider this minimizer to optimize the trade-off between data fit and connectivity and suggest it as *the* solution to the problem. For completeness, we also plot in Figure 18 the entries of A for exactly that minimizer and for each value of α.

Note that in Figure 18 the $(i, j)-$panel corresponds to the A_{ij} and that we have excluded the two largest $\alpha-$values, as they are outside the robustness region. The blue-shaded band within each panel marks the region $[-0.1, 0.1]$; recall that we consider $A-$entries within that region to be too weak and thus absent from the network. According to those results, the matrix structure appears to be

$$A = \begin{bmatrix} + & 0 & 0 & 0 & 0 & - & 0 \\ 0 & 0 & 0 & 0 & 0 & - & + \\ 0 & 0 & 0 & 0 & 0 & - & 0 \\ + & 0 & 0 & - & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & + & - & 0 \\ 0 & 0 & 0 & 0 & + & + & 0 \\ - & 0 & 0 & + & 0 & + & 0 \end{bmatrix}.$$

Due to lack of time this network is not yet validated on the data set of Solanum peruvianum.

Discussion

Interpretation of Results

The purpose of this study was to find the interactions between the seven tomato genes that together form the network whose dynamics is responsible for the regulation of branching behavior in tomato. The reconstruction of this network was based on the information contained in time series gene expression datasets of Solanum lycopersicum tomato species and its S-mutant. We first followed a bioinformatics approach to find out how much is already known about this network in the literature and databases. It appeared that there is hardly any information available. We only found an indication that Falsiflora should activate S in tomato, so one promoting edge. This info was obtained via a homologue pair in Arabidopsis thaliana.

We spent most time to use an ODE approach to look for the connectivity of the genes. Essential reduction was applied to the model based on the assumption that linearization of the real model is reliable since we are only interested in the positions of the edges and their promoting or inhibiting character and not in the details of the interactions. This assumption allowed us to investigate the connectivity of the genes in a relatively straight forward way. We optimized to find the network that fits the data well but with the smallest number of edges. This was obtained by minimizing an objective function that consisted of the sum of the squared residuals together with a penalty function for the number of edges. This last penalty term was weighed with a penalty weight. For large values of this weight, we obtained a network with very few edges, but with a bad fit. On the other hand, for small values of this weight, we obtained a network with many edges and a very good fit. So, we conluded that these effects have to be balanced and that is why we also optimized over this penalty weight. This resulted in a network with 14 connections; 5 activating ones, 5 inhibiting ones, and 4 self-regulating ones. In this optimal network, almost all genes connect and influence each other. Among the 7 genes, ANANTHA is the gene which gives rise to most inhibition effects to the other genes. BLIND and JOINTLESS are somewhat special, since they are nor regulating any other gene and also not theirselve.

Suggestions for Further Research

This study leaves open many interesting questions related to the regulation of the branching network in tomato. First, the obtained network should be tested on the data for Solanum peruvianum. Second, this initial study hardly yields information about the interaction strength of the edges. To find the genes that are most influential to branching of tomato is still an open problem. It is important to know these genes since only then the branching could be controlled, which could have tremendous influence on the tomato phenotype. This study shows that availability of complete gene expression data sets is essential in getting a grip on tomato branching. We thus recommend obtaining these data from lab measurements.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.654v1 | CC-BY 4.0 Open Access | rec: 2 Dec 2014, publ: 2 Dec 2014

Bibliography

[Pose13] D. Pose, L. Verhage, F. Ott, L. Yant, J. Mathieu, G. C. Angenent, R. G. Immink, and M. Schmid, Temperature-dependent regulation of flowering by antagonistic FLM variants, *Nature* **503**(7476):414–417 (2013).

[Leal14] F. Leal Valentim, S. van Mourik, D. Pose, M. C. Kim3, M. Schmid, R. C. H. J. van Ham, M. Busscher, G. F. Sanchez-Perez, J. Molenaar, G. C. Angenent, R. G. H. Immink, A. D. J. van Dijk, A quantitative and dynamic model of the Arabidopsis flowering time gene regulatory network, *PLoS ONE* **in press**

[Bau04] V. L. Bautch and C. A. Ambler, Assembly and patterning of vertebrate blood vessels, *Trends Cardiovasc. Med.* **14**(4):138–143 (2004).

[Bos09] L. Bosgraaf and P. J. M. Van Haastert, Navigation of chemotactic cells by parallel signaling to pseudopod persistence and orientation, *PLoS One* **4**(8):e6842 (2009).

[Gla13] J. A. Glazier, Multi-Scale Modeling of Tissues using CompuCell3D, in *Methods In Cell Biology*, A. R. Asthagiri and A. P. Arkin (ed.),USA: Academic Press, **110**:325–366 (2012).

[Gla93] J. A. Glazier and F. Graner, Simulation of the differential adhesion driven rearrangement of biological cells , *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.* **47**:2128–2154 (1993).

[Gra92] F. Graner and J. A. Glazier, Simulation of biological cell sorting using a two-dimensional extended Potts model, *Phys. rev. Lett.* **69**:2013–2016 (1992).

[Mer06] R. M. H. Merks et al, Cell elongation is key to in silico replication of in vitro vasculogenesis and subsequent remodeling, *Dev. Biol.* **289**(1):44–54 (2006).

[Mer08] R. M. H. Merks et al, Contact-Inhbited Chemotaxis in De Novo and Sprouting Blood-Vessel Growth, *PLOS Comp. Biol.* **4**(9) (2008).

[Par11] H. Parsa, R. Upadhyay and S. K. Sia, Uncovering the behaviors of individual cells within a multicellular microvascular community, *Proc. Natl. Acad. Sci. U.S.A.* **108**(12):5133–5138 (2011).

[Rei13] C. A. Reinhart-King et al, Computational modeling of angiogenesis: Towards a multi-scale understanding of cell–cell and cell–matrix interactions, in *Mechanical and*

Chemical Signaling in Angiogenesis, C. A. Reinhart-King (ed.), Springer, pp. 161–183 (2013).

[Tre05] M. C. Trevor, A framework for three-dimensional simulation of morphogenesis, *IEEE/ACM Trans. Comp. Biol. Bioinf.* **2**(4):273–288 (2005).

[Fra13] A. Franceschini, D. Szklarczyk, S. Frankild et al, STRING v9.1: Protein–protein interaction networks, with increased coverage and integration, *Nucleic Acids Res.* **41**:D808-15 (2013).

[Hof04] R. Hoffmann and A. Valencia, A gene network for navigating the literature, *Nat. Genet.* **36**(7):664 (2004).

[Jin14] J. P. Jin, H. Zhang, L. Kong et al, PlantTFDB 3.0: A portal for the functional and evolutionary study of plant transcription factors, *Nucleic Acids Res.* **42**:D1182-7 (2014).

[Kor11] T. Korcsmáros, M. S. Szalay, P. Rovó et al, Signalogs: Orthology-based identification of novel signaling pathway components in three metazoans, *PLoS One* **6**(5):e19240 (2011).

[Par12] S. J. Park, K. Jiang, M. C. Schatz and Z. B. Lippman, Rate of meristem maturation determines inflorescence architecture in tomato, *Proc. Natl. Acad. Sci. U.S.A.* **109**(2):639–644 (2012).

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.654v1 | CC-BY 4.0 Open Access | rec: 2 Dec 2014, publ: 2 Dec 2014