BACTERIAL IMAGE ANALYSIS AND SINGLE-CELL ANALYTICS TO DECIPHER THE BEHAVIOR OF LARGE MICROBIAL COMMUNITIES

Athanasios D. Balomenos,   Victoria Stefanou,   Elias S. Manolakos

Dept. of Informatics and Telecommunications, National and Kapodistrian University of Athens, Greece

Graduate Program "Information Technologies in Medicine and Biology"

{abalomenos} eliasm@di.uoa.gr

Introduction - Motivation

❖ Understanding the role of single-cell heterogeneity on the collective behavior of cell populations will uncover mechanisms leading to phenotypes of interest for human health, such as persister (dormant) cells, or biofilms, that promote pathogen's virulence.

❖ Time-lapse microscopy can provide answers by generating "big data" at multiple scales of the community organization, from cell populations down to intracellular molecules.

❖ Accurate and fully automated image analysis and single-cell analytics pipelines are needed to exploit such "big data" for systems microbiology and drug discovery.

❖ Well-known software packages for analyzing bacterial cell-movies are, among others, Oufti, Schnitzcells and SuperSegger.

❖ The inability of state-of-the-art software to process **complex cell-movies** with many colonies and thousands of cells in the field of view led us to start the *Bacterial Single-cell Analytics (BaSCA) project* [1] [2] [3].

Our new well-documented R-package:

➤ builds lineage trees and allows users to visualize cell attribute trends as well as identify and correct inevitable segmentation and tracking errors

➤ perform analytics on single-cell data extracted from cell-movies processed with different image analysis software in different granularity levels (per frame, per subpopulation, per colony, per generation), aggregating cell's attributes

Results and Discussion

➤ Analyzed data: Time-lapse movies with more than 16 S. Typhimurium single-cells forming overcrowded microbial community with more than 1500 cells in the last frame.

Overview of the BaSCA computational pipeline
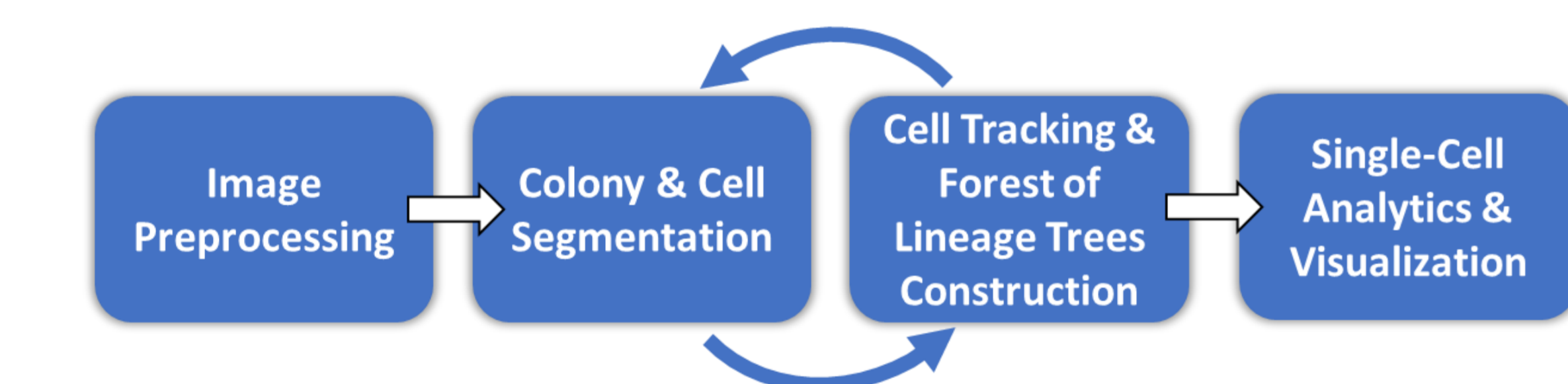

Overview of the newly added analytics capabilities

We developed a domain-specific R-package which enables:

➤ Visualization colony and single-cell attributes in multiple ways

➤ Performing statistical analysis at different levels of community organization, i.e. whole community, individual colonies, subtrees of colonies, cell generations across colonies, etc.

➤ Cell attributes: area, major/minor cell axis length, cell distance from the colony's centroid, etc. that may change in every frame.

➤ Cell life attributes: division time, length at division, elongation rate, statistics (e.g. min, max, mean, standard deviation) on a cell's attribute, e.g. average cell length (from birth to cell division), that characterize a cell's life span.

➤ All these single-cell properties are estimated and exported (stored) into a bio-database for each cell and time frame of the movie.

Tree representation of complex cell-movies

❖ **Forest of Lineage Trees (FLT)** of an analyzed time-lapse movie (16 colonies, 16 colored lineage trees co-evolving):
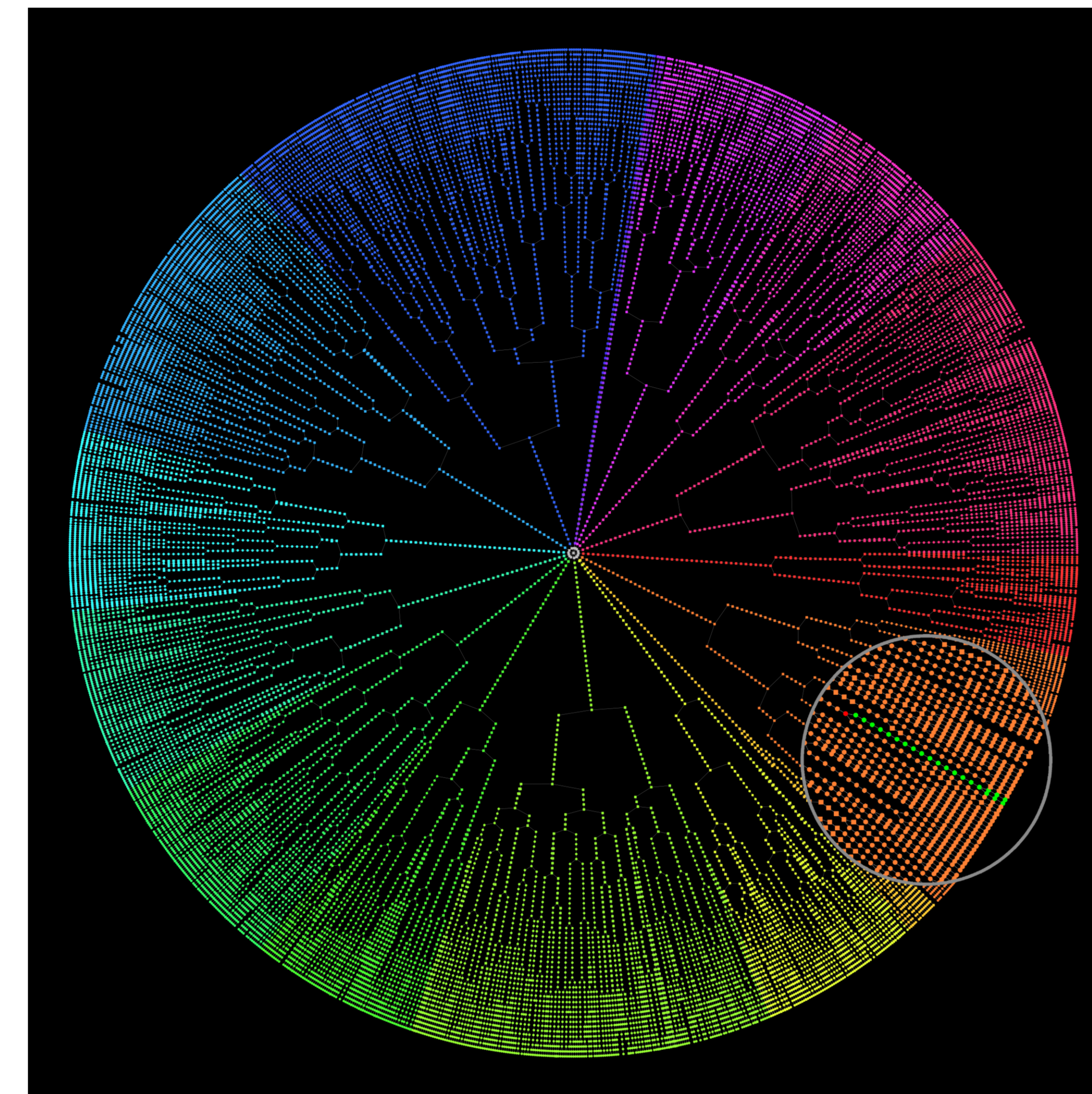

❖ **Forest of Division Trees (FDT)** of the cells existed in a movie:
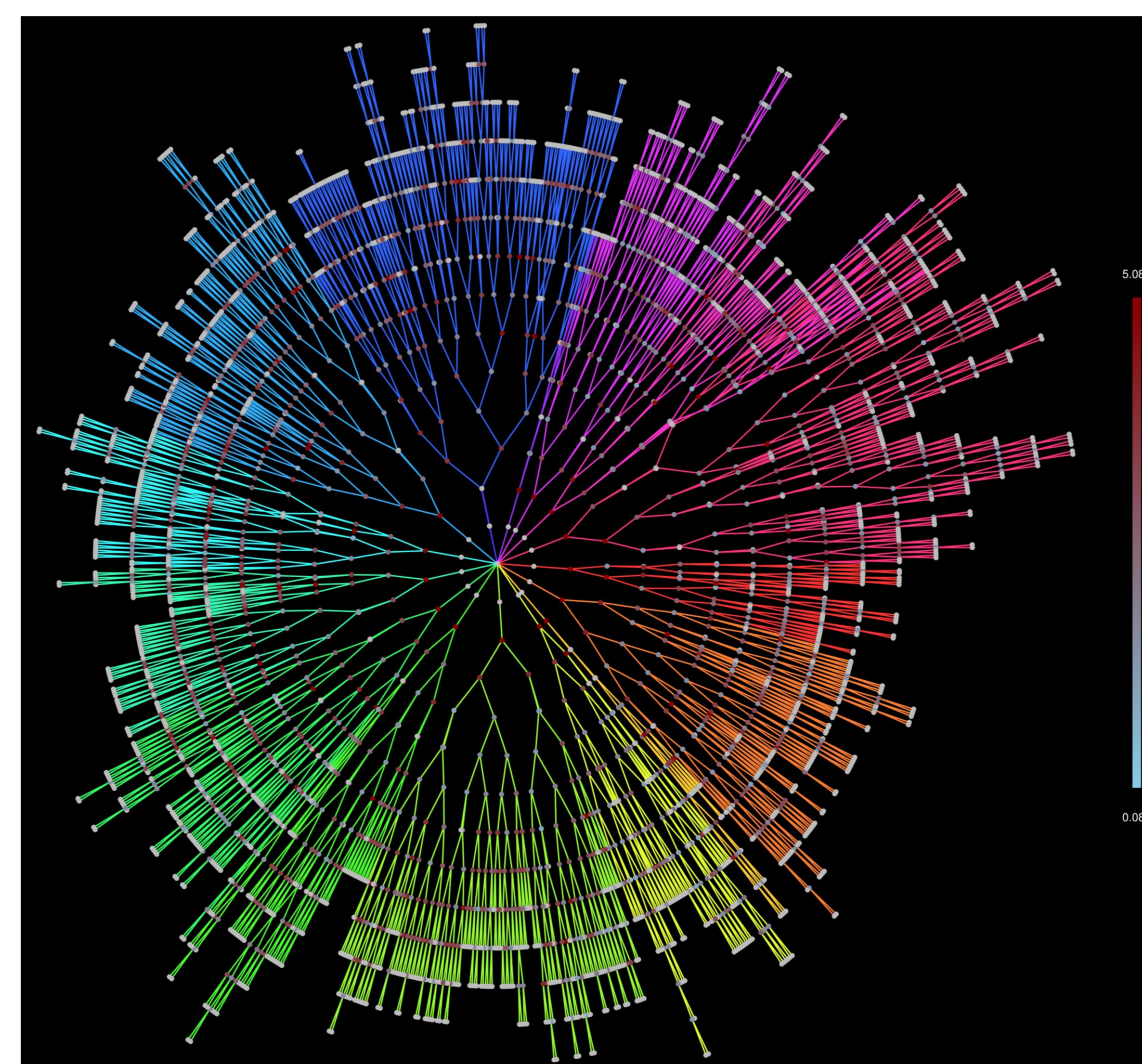
➤ Visual analytics on LTs and DTs enable quick assessment of how cell (life) attributes evolve across generations and "spotting" possible epigenetic effects from generation to generation.

Bacterial Single-Cell Analytics capabilities – Some examples

a) Micro-colony level growth curves.

b) Single-cell growth curves derived by fitting an exponential model to the length time-series data of each single-cell of a population.

c) Cell life attributes variability per generation (Best fit distributions of cell division time (gamma) and division length (log-normal)).

d) Division time scatter-plots: Inter-generation exploration of cell life attribute correlation trends

e) Cell (life) attributes correlation.
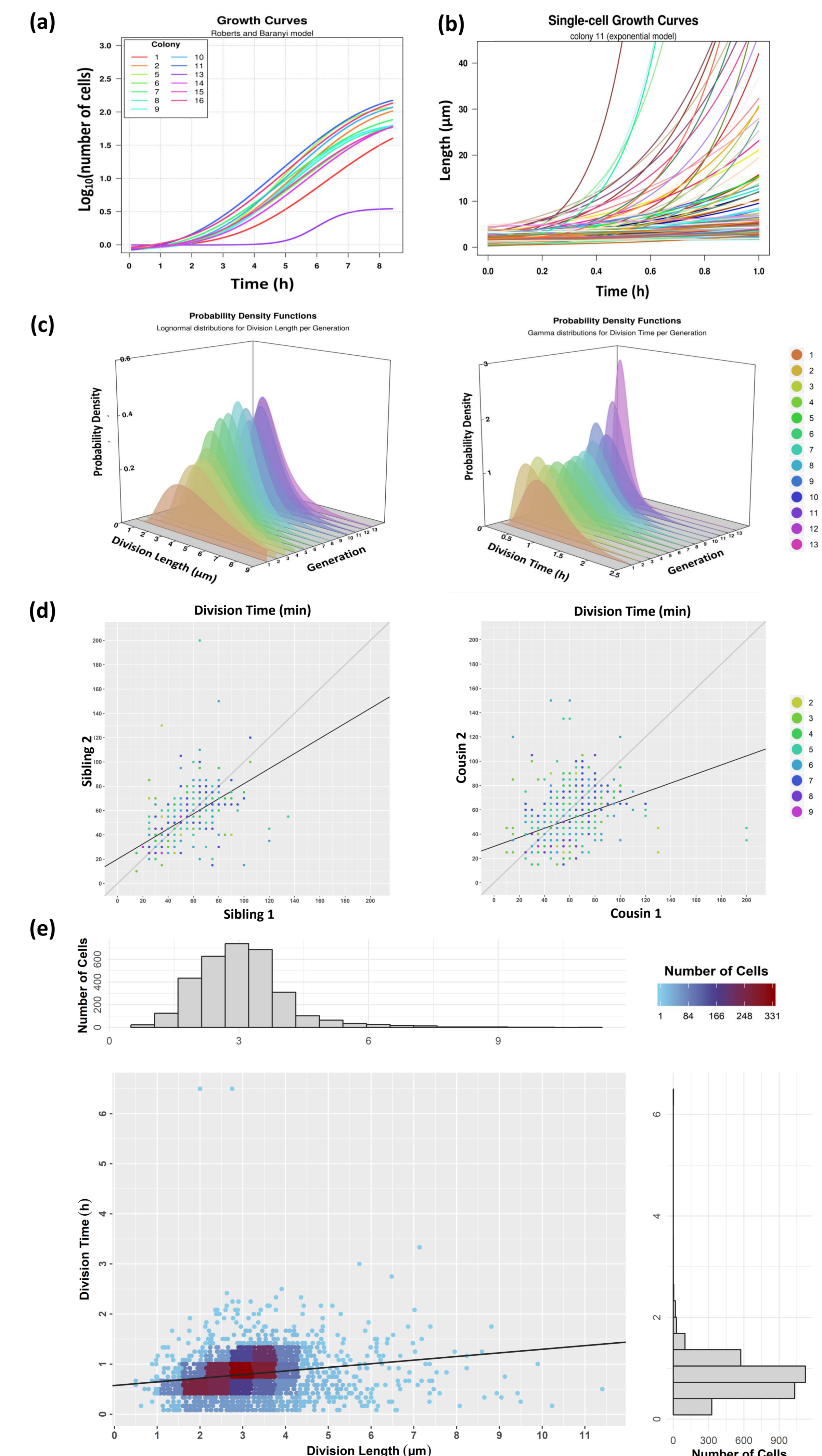
Correction of segmentation and tracking errors

❖ Lineage trees enable identification of: (i) over-segmentation of a cell in previous image frames, (ii) under-segmentation of a cell in the current frame, (iii) failure to match cells in consecutive frames.

❖ Such errors lead to unbalanced LTs because whole subtrees are not attached to any colony's LT.

❖ Using the R-package functions allows identification and "gluing" of such "lonely subtrees" to the appropriate position, thus correcting and extending the FLT of community.

❖ Practically unavoidable segmentation and tracking errors can be corrected

❖ BaSCA supports identification of segmentation and/or tracking errors through its powerful FLT representation of cell movies and associated analytics and visualization capabilities.
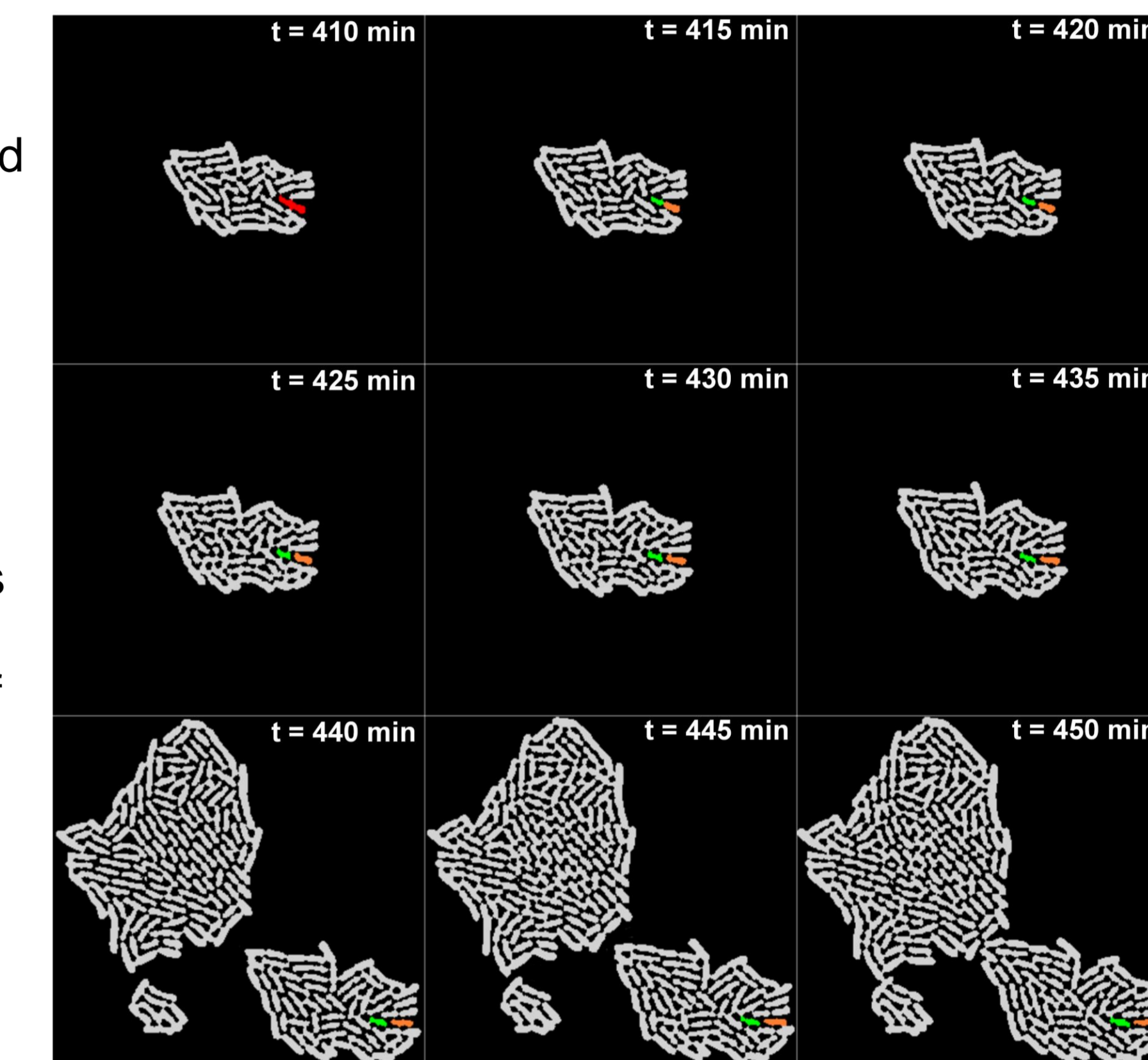
Conclusions – Research directions of the BaSCA project

❖ BaSCA extracts the forest of lineage and division trees of large complex bacterial communities and maps and visualizes the distribution of single-cell attributes over them.

❖ BaSCA helps scientists to see the bigpicture (community state) and understand trends and possible epigenetic effects along clones and across several cell generations.

❖ Our longer-term goal is to be able to analyze efficiently and without human intervention stacks of cell-movies in a high throughput mode.

❖ BaSCA: a tool to construct repositories of FLT/FDT representations of cell movies, capturing and characterizing adequately the "logic" of bacterial complex communities' behavior and interaction (e.g. pathogens vs. benign microbiome competition) under different stress conditions.

❖ We plan to release the BaSCA R package to the community once it is fully developed.

Related BaSCA References

[1] A. D. Balomenos, P. Tsakanikas, Z. Aspridou, A. Tampakaki, K. Koutsoumanis and E. S. Manolakos, "Image analysis driven single-cell analytics for systems microbiology," *BMC Systems Biology*, vol. 11, no. 1, 2017.

[2] A. D. Balomenos, P. Tsakanikas and E. S. Manolakos, "Tracking single-cells in overcrowded bacterial colonies," 2015 37th Annual Int'l. Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, 2015, pp. 6473-6476.

[3] A. D. Balomenos and E. S. Manolakos, "Reconstructing the forest of lineage trees of diverse bacterial communities using bio-inspired image analysis," 2017 25th European Signal Processing Conference (EUSIPCO), Kos, 2017, pp. 1887-1891.