

Experimental ‘omics’ data in tree research: facing complexity

Wolfgang zu Castell · Dieter Ernst

Received: 2 April 2012 / Revised: 26 July 2012 / Accepted: 12 September 2012
© Springer-Verlag Berlin Heidelberg 2012

Abstract High-throughput experimental technology has provided insight into the inner functioning of plants. The current experimental technology facilitates the study of plant systems in a holistic manner, measuring observables from the genome, proteome, and metabolome up to the level of the ecosystem. The call for a systemic view in plant research is being made from multiple research fields. Although not yet fully developed for tree research, data sources are also rapidly growing in this area. Nevertheless, there are challenges and pitfalls in dealing with such increases in data. Some of these difficulties are deeply rooted in the complexity of the evolutionary systems. The lessons from complexity theory are rooted in studies performed several decades ago. Honouring principles that were formulated before bioinformatics and systems biology had been introduced facilitates the derivation of analytical methods with the potential to overcome these challenges in several ways.

Keywords ‘Omics’ data · Plant systems biology · Systems theory · Complexity · Large-scale modelling

Communicated by R. Matussek.

Special topic: Integrating Modeling and Experiment.

W. zu Castell (✉)
Scientific Computing Research Unit, Helmholtz Zentrum
München, Ingolstädter Landstraße 1,
85764 Neuherberg, Germany
e-mail: castell@helmholtz-muenchen.de

D. Ernst
Institute of Plant Pathology, Helmholtz Zentrum München,
Ingolstädter Landstraße 1, 85764 Neuherberg, Germany
e-mail: ernst@helmholtz-muenchen.de

‘Omics’ research: where are we today?

Despite the unquestionable success of modern high-throughput technology in deciphering regulatory mechanisms in plants (see e.g. Ahuja et al. 2010; Urano et al. 2010), the advances in so-called ‘omics’ technology continues to provide insight into an ever-growing complexity of life. Starting with an analysis of plant genomes, the technology moved on to transcriptomes, proteomes, and metabolomes and then to modern research on plant phenomes and interactomes (Mochida and Shinozaki 2011; Weckwerth 2011). There are few examples linking such analyses in forest trees, including climate change conditions (Matussek et al. 2007; Bylesjö et al. 2009; Hall et al. 2011; Ernst et al. 2012a, b). At the same time a growing number of authors reiterate the promises and hopes put into holistic views of plants and the possibility to answer all questions using in silico modelling (Shulaev et al. 2008; Yuan et al. 2008).

Thus, what Brenner wrote over a decade ago still remains true: “There seems to be no limit to the amount of information that we can accumulate, and today, at the end of the millennium, we face the question of what is to be done with this information. ...[I]t is clear that the prime intellectual task of the future lies in constructing an appropriate theoretical framework for biology” (Brenner 1999).

Systems biology seems to provide the realisation of this task (Kitano 2002; Ng et al. 2006) as it picks up complexity, shifting the focus in molecular biology from studying single genes to complex interaction networks of genes, proteins and metabolites (see also Fox Keller 2005a for theoretical comments).

The initial enthusiasm is waning, however, as time passes. Experimentalists and theorists are encountering the limits of their approaches. Raising doubts culminate in

provocative titles, such as “Can biological phenomena be understood by humans?” (*Nature* 403:345; 27 January 2000). One obstacle is the challenge of processing, integrating and standardising the large amounts of data created from high-throughput experiments (see e.g. Cassman 2005; Joyce and Palsson 2006; Fukushima et al. 2009). We will not dwell on these issues here, but rather, we will focus on some theoretical aspects, i.e. analysing and modelling experimental data. Thus, the aim of this review is to show the common pitfalls of data analysis and modelling in ‘omics’ experiments (**‘Omics’ research: pitfalls and challenges**). We will closely examine the origins of these drawbacks, i.e. the complexity of the phenomena being studied (**Systems biology: what is its intention?**). By reviewing the evidence obtained from the theoretical research in **“Systems theory: where are we coming from?”**, we can see that the call for holistic approaches is not new. We will therefore consider lessons learned from cybernetics and systems theory during the first half of the twentieth century. After focusing on examples of ‘omics’ experiments in tree research, the last section will present several approaches based on these lessons that might therefore provide an alternative path for future research.

‘Omics’ research: pitfalls and challenges

Lay et al. (2006) outlined the challenges of state-of-the-art ‘omics’ research, grouping them into four categories: bias, statistics, methodology, and method misuse.

As with bias, the problem partly lies in the size of the available data sets. With such large sets, correlations apart from any cause–effect relationship are likely to occur. Thus, the risk of false positives is significantly higher. At the same time, the increasing pressure to publish quickly introduces a bias towards positive findings (see also Ransohoff 2005; Pfeiffer and Hoffmann 2009). There is an analogous risk of false negative findings, but due to the common practice of publishing, the risk of false-positive findings is higher.

The statistical challenge is rooted in the high number of observables compared with the commonly low numbers of replicates. Wacholder et al. (2004) examined this problem in more detail, investigating the factors that determine the so-called false-positive-report-probability (FPRP), including the level of significance for a given p value, the statistical power of the test, and the prior probability of the hypothesis to be true.

The challenge of methodology comes from the huge number of methods needed to properly perform ‘omics’ experiments and analysis (Galbraith 2006). It can hardly be expected that a researcher is able to deal with all the necessary methods stemming from various areas of research at the same competence level.

The fourth class, called ‘fitness of use’, comprises multiple problems rooted in different laboratories performing the various steps without standardisation, particularly in emerging areas of research, and a lack of proper validation methods. The latter is accompanied by the reliance on validated databases and access to meta-data to properly evaluate the appropriateness of the data used for comparisons (see e.g. Cassman 2005).

Apart from these challenges and difficulties, there is an underlying, deeper problem that stems from the complexity of life itself. The fallacies of purely gene-oriented research in plant sciences have been previously noted (see e.g. Sinclair and Purcell 2005; Lüttge 2012). Post-genomic research requires new strategies and methods which even might not have been developed yet (Brenner 1999; Castriello and Oliver 2004). “To pretend that technological innovations will enable the understanding of these complex phenomena is just wishful thinking. Instead we need a novel way of thinking about these problems” (Soto and Sonnenschein 2006).

With respect to this ‘novel way of thinking’, let us mention two aspects of major importance. First, we are interested in understanding the functioning of regulatory mechanisms in living organisms. Apart from the pure presence of constituents, there are multiple interactions among them. Regulation in plants is often redundant and realised through the interactions of components on various biological scales. Thus, we need a shift from presence/absence correlation, which is the main tool in large-scale ‘omics’ analysis, to functioning, i.e. interaction networks (Sweetlove and Fernie 2005; Sauer et al. 2007; Mazzochi 2008).

Second, there is a common understanding that environmental adaptation plays an important role in this process. However, the identification of this role and the characterisation of its mechanisms of action remain unknown. For example, considering the question of whether changes in gene expression are fixed through evolution, there is support for the hypothesis that “the majority of expression differences observed between species are selectively neutral or nearly neutral and likely to be of little or no functional significance” (Khaitovich et al. 2004). Without doubt, the interaction always occurs within an environment. Thus, we have to consider that environment. Understanding natural variation as a response to evolution within a certain environment has consequently been proposed as one of the “grand challenges in plant systems biology” (Fernie 2012; see also Baker 2012).

A consequence of the precise definition of boundaries for the system under scope is the possibility of considering hierarchies of systems. However, as such, the hierarchies also become contingent on the perspective of the observer. What might be diversity at one level will be a collection of

constituents on the higher level. Thus, diversity on the ecosystem level can be both a means of internal variability of the system and pleiotropy of the environmental interactions driving adoption on the species level. In the first interpretation, it is a way to maintain internal stability, whereas the second point of view stresses the amount of pressure for adoption. Regardless of the choice, natural diversity should be considered as a chance to address environmental influence on organismal systems (for plant systems, see Kliebenstein 2010; Fernie 2012).

Systems biology: what is its intention?

The call for holistic approaches and system perspectives is not new. We next trace the roots of this call.

Pauline Hogeweg and Ben Hesper are commonly mentioned as having coined the term ‘bioinformatics’ (see Hogeweg 2011 for some historical remarks). It is one of these surprising arabesques of history where the original intention for introducing a new discipline seems to match the modern day agreement on the future need for systems biology, rather than the dominant understanding of bioinformatics as a discipline of data collection, database search and data organisation. Clearly, systems biology might have as many meanings as there are authors writing on subjects from the field, but there appears to be two major interpretations (see also Westerhoff and Palsson 2004). One of them might be called the ‘systemic approach’: applying all types of ‘omics’ technology to obtain a holistic view of the living system on all scales of biological organisation, from genomes, proteomes and metabolomes to ecosystems and even biotechnology (Fukushima et al. 2009; Sauer et al. 2007; Weckwerth 2011). Regarding plants such a program has been mainly executed for the model system *Arabidopsis thaliana*. Next-generation sequencing also focuses on the study of important crop plants. Regarding forest trees, apart from poplar, not much information is available (Wullschleger et al. 2009). The establishment of tree databases and transcriptomics (Pavy et al. 2007; Yang et al. 2009; Parchman et al. 2010; Ueno et al. 2010; Fernández-Pozo et al. 2011), proteomics (Abril et al. 2011) and metabolomics (Ossipov et al. 2008; Businge et al. 2012) will support comparative ‘omics’ for forest trees.

With a slightly different focus, the second interpretation stresses *in silico* modelling of large-scale interaction networks as a tool to unravel the fundamental mechanisms of life. As Denis Noble, one of the pioneers of systems biology, states: “There is therefore no alternative to copying nature and computing these interactions to determine the logic of healthy and diseased states” (Noble 2002). In terms of bioinformatics, this is indeed what Hogeweg and Hesper aimed: “first, to analyze patterns of

variation at multiple levels in organisms; second, to detect emergent phenomena in models; third, to compare the outcome of such models with ‘real’ data; and finally, and most profoundly, because the relationship between genotype, phenotype, behavior, and environment itself can be seen as a type of pattern recognition or pattern transformation ... understanding these processes” (Hogeweg 2011). Thus, it surprised the inventors of bioinformatics that with increasing experimental technology, bioinformatics became the field of computational methods for data analysis and data management (see Hogeweg 2011).

Why did the field develop this way? Interestingly, both interpretations of systems biology and bioinformatics in its original sense address the limits of the reductionist approach. It is a common understanding in system-centric approaches that we have to move beyond these limits of a purely reductionist view. “The reductionist approach has successfully identified most components and many interactions but, unfortunately, offers no convincing concepts or methods to understand how system properties emerge” (Sauer et al. 2007). Soto and Sonnenschein (2006) discuss what they propose as the shortcomings of reductionism.

The key argument against reductionism is that emergent properties of complex living systems cannot be explained through the system’s parts. Although there is a philosophical debate concerning the proper foundation of the concept of emergence (see e.g. Kim 1999), there is no doubt that certain properties of complex systems, e.g. emergence and contingency, challenge the traditional way natural sciences proceed, particularly when moving from a mainly descriptive perspective towards a functional understanding and controllability of living systems (Mazzochi 2008).

The current claims of authors today have already been formulated in the first half of the twentieth century. The study of the properties of complex systems and the deduction of principles of complexity has a long tradition, in particular, in theoretical biology. Norbert Wiener is considered as one of the founders of the field, coining the notion of cybernetics (Wiener 1948), a topic that has been discussed from the point of view of theoretical biology by Ashby (1957). Later, von Bertalanffy (1969) formulated his ‘General System Theory’ with the aim of developing a unified, multi-disciplinary approach to general systems.

We might only speculate about the reasons why these approaches have been lost attention in many publications in large-scale biology during the last decades. One reason might reflect the rapid advances in computing technology nourishing the hope to overcome the limits of complexity through intensive computing. Interestingly, living systems, evolutionary development and self-organisation, through the interaction of biological agents, have always been and still are an important source of inspiration in artificial intelligence and technical cognition.

Systems theory: where are we coming from?

Considering the ideas provided by these early pioneers of cybernetics and systems theory, we propose that certain aspects of the difficulties mentioned above are rooted in the complexity of the systems being investigated. Let us briefly recall some of the basic ideas in systems theory, that have been exposed in much greater detail in the works mentioned.

There is no common definition of a system. The discussion concerning such a precise definition is a rather philosophical one. One of the difficulties lies in the fact that a system cannot be defined without the role of the observer, leading into the dispute over constructivism. However, for our purpose here, providing a practical notion is sufficient. Therefore, we propose that a system is defined through individual agents and their interactions. The aspect that the agents interact should be stressed. Through interaction, the system is defined as a whole and cannot be taken apart (Ashby 1962; Gershenson 2011b).

Notably, the choice of agents and the interactions considered remains contingent on the perspective of the observer. Thus, biases are unavoidable not only on the methodological side. “Ontological statements, i.e. about things as they ‘really’ are, independent of an observer, are fictitious. Once a statement—ontological or of any kind—is made, this has been made by an observer, turning it into epistemology, i.e. things as they are perceived and described” (Gershenson 2011b). Therefore, the bias challenge mentioned by Lay et al. (2006) partly results from the subconscious influence of the observer. This role naturally resurfaces in statistical methodology (cf. FPRP factors mentioned above).

Referring a system back to interactions implies shifting the focus onto the latter. Systems theory can be considered as the theory of the interaction of individual agents. Because interaction is a concept involving two partners, there is no interaction without communication in a general sense of the word. Thus, we can use communication as a surrogate to experimentally study biotic/abiotic interactions (Ashby 1962). Notably, such an approach provides a suitable method to experimentally study interactions. Communication can be measured through information (see Shannon and Weaver 1949). Moreover, information can also be used to describe the organisational status of a system (Gershenson 2007).

The concept of communication as a way to address organisational complexity is rooted in the theory of thermodynamics. Prigogine developed a theory of dissipative structures (see e.g. Nicolis and Prigogine 1977) pushing Boltzmann’s work on statistical mechanics beyond the limits of closed systems in equilibrium. Within this framework, structure emerges from conditions that keep the system away from equilibrium through the flow of

energy. In this way, a complex system counterbalances instability resulting from fluctuations within its environment. “There is competition between stabilization through communication and instability through fluctuations.” (Prigogine and Stengers 1984). The parameters of the driving systems of differential equations can be used to study the passage from instability to stability. Dissipative stability therefore expresses itself through bifurcations, breaking of symmetry, or phase transitions.

As the system also interacts with its environment, determining a system implicitly fixes the boundary between the system’s interior and exterior. We have previously mentioned above that the boundary is contingent on the perspective of the observer. Once fixed, adaption can be considered as an emergent characteristic which rises from the interaction of the system with its environment. Similarly, fitness can be interpreted as the property of a system to persist for the long term.

Considering three agents, we can address the question of whether the interaction of one agent with another is conditioned in some way through an interaction with the third. In accordance with Ashby (1962), as soon as conditioning is considered, we can think about organisation. Organisation describes the conditioned interaction of agents. As such, organisation can be considered as a universal phenomenon underlying complex systems. Haken (1983) established the field of synergetics, studying the principles of ordered and collective interaction. Both, Haken and Prigogine provided a basis to study organisation in dissipative systems within a formal theory. Since then, their ideas have provided rich inspiration for theoretical work.

Because all of these ideas are based on the interaction of agents, communication provides the clue to analyse complex phenomena. As previously discussed, there is a well-established theory of communication (Shannon and Weaver 1949). Von Neumann (Morgenstern and von Neumann 1944) studied the concepts of interaction strategies and resulting patterns, providing models of interaction systems. Again, there is a wealth of theoretical work to be utilised to analyse and unravel mechanisms of communication and interaction in living systems.

Another aspect which has been mentioned previously, is hierarchy. Although we have not properly defined what is meant by ‘complex’, there is an intuitive understanding of some systems being more complex than others. Thus, one might utilise the characteristic of possessing parts that to some extent share the properties of the whole as a working substitute for complexity. In this sense, there is a hierarchy of complex systems. In addition, the properties of the system might not be observable at all scales. Turing’s work on reaction–diffusion models (1952) is one of the first formal analyses of emergent properties, arising from the interaction of subsystems on a lower scale.

It is widely accepted that living systems are complex. However, is there any difference between a living and a complex system?

In biology, we deal with living systems. Again, it is a matter of philosophical debate what living means precisely. Within a slightly more general context, we can talk about viable systems. We will come back to this subtle difference later on.

Ashby (1957) defined viability as an observable outcome of a system that enables the system to return to internal stability upon disturbance rooted in its environment. Therefore, it is important to stress that stability is considered to be internal. In this context, an uprooted tree with some external reinforcing structure will not be considered as stable. The ability of the system to reorganise itself towards internal stability is called self-organisation. Prigogine and Haken (Prigogine and Stengers 1984; Haken 1983) have given a formal definition to this concept. Again, there is a philosophical peculiarity. Self-organisation implies the availability of a suitable definition of 'self'. One way to escape the need for any teleological explanation is to again consider the role of the observer. It is the observer who assigns states, properties, characteristics, etc., to a system that he expects to be preserved during reorganisation (see also Ashby 1962; Gershenson and Heylighen 2003). The debate is much less theoretical than it might seem from an experimentalists' perspective. As previously discussed, bias is immanent in all experimental work. Lay's call for fitness of use will not suffice. Can scientists actually distinguish the tree from its rhizosphere in field experiments? Because we will never be able to observe an organism as a living entity without its microbiome, does our organismic notion of a tree make sense from a functional perspective? There is increasing support for the idea to consider the plant and its associated microorganisms as one system (Zilber-Rosenberg and Rosenberg 2008). In a similar way, other levels of biological organisation have to be considered within a larger scope, leading to the concepts of the metagenome, metatranscriptome, and metaproteome.

With consideration for viability, we have to study the mechanisms, allowing a system to return to its stable internal state. How can such a goal be achieved within a highly complex and unpredictable environment? Non-predictability, at least in any practical terms, is an immediate consequence of open systems and non-linear dynamics. Ashby introduced the concept of variety to address such questions (Ashby 1957). In short, *variety* is a measure of complexity quantifying the number of states a system can take. Thus, a simple binary switch being on or off has a variety value of two. Combining several binary switches, the number of possible states quickly increases. For example, consider a computer, which internally consists of multiple binary switches linked together in multiple ways.

Being able to react to perturbations of the environment in a certain way is then a matter of *controllability*. To return to a given initial state, the system must be able to return from—theoretically—any state back to its initial state, which, clearly, is impossible. Thus, the weaker formulation of the problem requires for the ability of the system to return from a perturbation of bounded intensity back into one stable state—called an *attractor*—which is acceptable in the sense of viability, i.e. retains the ability to self-organise. Notably, such a state might not be a healthy one. As such, nothing is said about the long-term consequences of choosing such a state. A pathological state, for example, might lead to lower living expectance, although for the moment, it might be stable. Furthermore, stability alone is not enough. Clearly, death is also a quite stable state, but not one which allows the system to self-organise any longer. Thus, distinguishing between viability and living becomes crucial at this point. The notion of life carries some additional meanings such as, e.g. reproducibility, which we do not need to consider at this point. Therefore, it makes sense here to distinguish living from viable.

One fundamental result from cybernetics is Ashby's Law of Requisite Variety (Ashby 1957), which states that the variety of a controller must be at least as large as the variety of the system to be controlled. In other words, if a system is able to balance the much larger variety of its environment, it needs variety attenuators to cope with the complexity of its environment. We note in passing that there is a related statement in communication theory (Ashby 1958).

In the same way, we as an observer with limited variety have to cope with the much larger variety of the system that we are studying. This can already be seen on the methodological side. To perform proper experiments requires diverse skills, ranging from chemistry to bioinformatics and statistics. The three challenges mentioned by Lay et al. (2006), i.e. bias, the challenge of methodology, and the challenge of fitness of use are necessary variety attenuators. In addition, our experimental setup should carefully select for parameters that can be manipulated utilising variety attenuators. Otherwise, we might not be able to observe any effect, as the system will be able to buffer our external disturbance.

Experimental research: approaching the complex!

Let us now turn our attention to approaches honouring the insights from systems theory and cybernetics.

The first important chance lies in the diversity of approaches. Recall the role of the observer when studying complex systems, and consider how reliable scientific

insight could ever be achieved. History has provided us many examples, showing that although specialised experiments will almost always be valid only within the precise scope of the experimental setup, comparing various related experiments will provide insight concerning the mechanisms behind the observations. Let us consider a famous example. How much insight do we need to understand that the principle behind an apple falling down a tree is the same as the principle by which the earth rotates around the sun? Without abstraction, we would never be able to deduce such a principle. It is the mental capability of abstraction that allows us to compare seemingly different situations and infer the common principles underlying them. Thus, a diversity of approaches is not only a consequence of the impossibility of different scientists to address the same question in exactly the same way, but also a chance to overcome the individual limitations of each experiment. In this way, we respect the role of the observer and use it to our advantage. Moreover, we should expect that evolutionary systems have to come with a certain amount of complexity because otherwise they might most likely not have persisted within the previous process of selection. Therefore, it seems quite likely that nature will have varying answers to solve one problem, a phenomenon that is widely observed in epigenetics. Heylighen formulated the principle of selective variety: “The larger the variety of configurations a system undergoes, the larger the probability that at least one of these configurations will be selectively resistant” (Heylighen 1992; see also von Neumann 1966). Evidence for this principle comes from experimental work. The functioning of evolutionary systems critically depends on a certain amount of diversity. For example, Isbell et al. (2011) studied plant ecosystems. Another interesting study reports on agents in driving divergence (Godsoe et al. 2009).

Many mathematical models use absence/presence data for modelling biological systems and infer dynamic behaviour from estimating fluxes. Although precise hypotheses can be tested, there is a comparably large amount of information required in advance to set up a model. In contrast to these traditional models, other approaches are aimed directly at interactions. Thus, the focus of the model shifts from the agents towards their interactions. In a mostly qualitative way, Kauffmann (1969) used Boolean networks to study the properties of self-organisation. Their studies led Kauffman and his co-authors to formulate five minimal conditions for biological agency (Kauffman and Clayton 2006). Such conditions, e.g. self-propagating work and constraint construction, could lead to precise hypotheses that can be addressed through experimental studies. We can hypothesise how a certain function—which needs to be present to guarantee viability—might be realised in a concrete system. Large-scale ‘omics’ data provide a valuable basis to deduce

information about the functional relations which might be relevant in interaction modelling. Clearly, data integration is a crucial issue. For a recent example in *A. thaliana*, see Pop et al. (2010). Large-scale ‘omics’ initiatives have been undertaken in trees. For example, a microarray analysis identified candidate regulators in the drought-stressed roots of loblolly pine (Lorenz et al. 2011), and a proteome analysis showed changes in carbon metabolism in ozone-treated poplar leaves (Bohler et al. 2007). Comparative transcriptomics and proteomics in poplar discovered drought-stress-responsive transcripts and proteins that can be grouped into regulatory networks (Plomion et al. 2006; Cohen et al. 2010), and a systems biology model of the regulatory network of poplar leaves has recently been outlined (Street et al. 2011). In European beech, birch and *Populus* transcriptomics, proteomics and metabolomics have been linked to climate change in free-air ozone and CO₂ fumigation sites (Gupta et al. 2005; Taylor et al. 2005; Kontunen-Soppela et al. 2007, 2010; Ossipov et al. 2008; Olbrich et al. 2009; Calfapietra et al. 2010; Kerner et al. 2010; Ernst et al. 2012b).

Gershenson provided a modelling methodology to directly address questions of interaction principles across multiple scales (Gershenson 2011a). In particular, such tools are suitable to shed light on the emergence of properties on higher scales. For example, Gershenson discusses the conflict between ‘selfishness’ on the lower scale and the need for mutualistic behaviour on a higher level. Experimentally, such a methodology can be applied based on information flow. Gershenson’s approach uses a ‘satisfaction’ function which quantifies the extent to which an agent can maintain its own stability. Within cellular networks, it can be hard to directly define satisfaction. Communication might then be considered as a surrogate. In many cases, we can measure metabolites or peptides that can be associated with certain agents, e.g. through pathway information. Thus, we consider information flow, a claim that has also been made by Noble laureate Paul Nurse (2008).

The key shift in paradigm of experimental setup means to consider directly measuring interaction rather than deducing interaction indirectly from state type observables of the agents.

Once the interesting interactions have been identified, we have to infer patterns of interaction which might be relevant within a certain situation. A prominent approach to analysing interaction networks is based on the observation that patterns of dynamic structures on networks, such as, e.g. forward or backward feedback, can be aligned using small topological structures, called motifs (Milo et al. 2002; Doyle and Csete 2005). On the larger scale of the network as a whole, there are several characteristics that can be measured and related to network models. Among others, scale-free networks have received much attention. Barabási and co-workers have analysed many real-world

networks and found stunning similarities. The most important observation showed that many of the networks in biology, social sciences, informatics, etc., show a so-called power law degree distribution (Barabási and Albert 1999; Jeong et al. 2000). With such a distribution, there is a positive probability for few nodes to have extremely high connectivity within the network. Such a type of node is relatively unlikely in Erdős–Rényi networks, a model type which has been widely used before in network analysis (see Newman 2003 for a review). Notably, apart from the hype created by Barabási's work, we should not over-interpret the importance of degree distribution (see e.g. Fox Keller 2005b). Several authors have stressed the importance of network analysis in understanding functioning of complex organisation and ecology. Barabási and Oltavi (2004) have written a review on the cellular level (see also Luscombe et al. 2004; Sweetlove and Fernie 2005 for other examples). There is a wide range of applications of network analysis on the ecological level (see Proulx et al. 2005; Bascompte 2007 for reviews).

Some authors argued that modern biological research should be closer to engineering than to physics or chemistry (see e.g. Csete and Doyle 2002; Fox Keller 2005a). The idea is appealing because ultimately, we want to control complex biological systems. Due to the much larger variety of the system to be controlled, the controller needs to use variety amplifiers and variety attenuators (Ashby 1957). This so-called variety engineering is a natural concept in the construction of multi-functional technical systems. Robustness and modularity are two principles being crucial for the successful design. While the first guarantees the long-term functionality of the system within a dynamically changing environment, the latter is a design principle that breaks down complexity, allowing the encapsulation of vital functions. In particular, modular design leads to a decoupling of low-level variation and functionality on the higher level. Ideas of this type are close to the concept of niching in ecology (Maynard Smith and Szathmáry 1995).

Robustness and modularity are fundamental prerequisites for the evolution of a system in a dynamic environment (cf. Kitano 2004). Uncovering such vital functions through experimentation and observation is a problem of reverse engineering. Csete and Doyle (2002) report on several parallel concepts between technical and living systems, demonstrating that an engineering perspective might be very useful in trying to understand the functioning of complex biological systems (see also Fox Keller 2005a).

Shifting the focus from copying to controlling also leads to new insights in modelling. In doing so, control theory provides a rich toolbox of methods (Liu et al. 2011b). However, this route is not just one-way. In addition, systems biology poses new theoretical challenges in control theory (Sontag 2004).

The lessons learned from cybernetics also need to be respected when dealing with modelling.

Considering contingency of the system of scope, we should carefully consider what to model. Accepting models as epistemological, we might model our expectation of what should be observed. Alternatively, we could model a framework of rules within which self-organisation takes place. The latter is the approach followed in most work dealing with cellular automata (cf. Kauffman 1996). Learning algorithms and other data-driven methods in statistics also fall into this category (zu Castell 2012). As a disadvantage, one might object that these approaches do not provide much explanatory insight.

Following the direct approach, i.e. modelling of the actual interactions, there are also limitations that have to be considered. A principle, which is commonly referred to as Occam's razor (see e.g. Sober 1981 including a critical discussion), postulates that the simplest explanation for a phenomenon will also be the most likely to be confirmed. This principle has important implications for model selection. Following Occam's razor, models should be as simple as possible. There are further practical arguments supporting such a principle. Indeed, the complexity of a model is typically associated with a large number of parameters that have to be estimated from the data. Given a model of large enough complexity, in the sense of the number of equations or parameters, almost any data can be fitted (see e.g. Smale 1976; Hornik et al. 1989). Thus, the predictive capacity of the model decreases with increasing numbers of parameters, both in statistical and deterministic modelling.

This well-known trade-off in theoretical modelling might reflect the consequences of a similar question of complexity in general. The founder of cybernetics John von Neumann postulated that there is a certain minimal amount of variety required in order to allow for complexity at the scale of a living system (von Neumann 1966). May (2001) later has proven that for random systems, there is also an upper bound, above of which stability will not be possible. Interestingly, this upper bound does not seem to hold if preferential attachment is present, i.e. mechanisms like mutualism or competition (Allesina and Tang 2012). Thus, it remains a challenging problem to explore the chances and limits of interaction networks and their capabilities to generate stable patterns through self-organisation.

Large-scale 'omics': recent progress in tree research

Over the last 10–20 years, little research has focused on tree genomics being far away from model and crop plants. This situation changed with the genome sequencing of *Populus*, revealing more protein-coding sequences than

A. thaliana. The whole sequence genome of *Populus trichocarpa* was published in 2006 (Tuskan et al. 2006). Using next-generation platforms, other additional *Populus* species were sequenced, and *Populus* is currently the model system for trees (Jansson and Douglas 2009; Yang et al. 2009). In addition, several other tree species have been sequenced, including conifers (Eckert et al. 2009; Liu et al. 2011a; Rigault et al. 2011), *Eucalyptus* (Grattapaglia et al. 2012) and Fagaceae (Kremer et al. 2012). Moreover, the genomic data for fruit trees such as citrus or apple are available (Gmitter et al. 2012; Troggio et al. 2012). A “walk through the forest of tree genomes” has recently become a focus in a special issue in *Tree Genetics & Genomes* (Abbott 2012).

Regarding transcriptomics, microarray analyses and next-generation sequencing led to a breakthrough in the identification of thousands of abiotic/biotic-, developmental- or tissue-specific transcripts. Forest FACE experiments contributed to a more molecular ecophysiological understanding of trees upon environmental impact. Transcriptome analyses of *Populus* and birch in these studies indicated long-term genetic expression changes upon elevated ozone and/or CO₂ (Gupta et al. 2005; Taylor et al. 2005; Kontunen-Soppela et al. 2010). Similarly, transcriptional changes could be demonstrated in European beech (Olbrich et al. 2009, 2010). Comparative transcriptomics of drought responses in *Populus* or loblolly pine resulted in the identification of central nodes and regulators that are important for trees in regulating this environmental stress (Cohen et al. 2010; Lorenz et al. 2011). Similarly, salt stress has been analysed in desert poplar, being important for the molecular adaptation of this tree to stress (Qiu et al. 2011). Bioinformatic analysis of 454-sequenced oak and pine ESTs will help to study the functional diversity in the natural population of these important forest trees (Parchman et al. 2010; Ueno et al. 2010). Regarding tree regeneration and transformation, the large-scale transcriptomics of *Populus* identified important events during the early stages of dedifferentiation, and only 25 % of these events were shared with *A. thaliana* (Bao et al. 2009). Global transcript profiling of poplar and rice identified identical promoter motifs conserved in circadian-controlled networks that will be important in understanding similarities and differences between dicot and monocot plants (Filichkin et al. 2011). Interestingly, the sequencing of high-quality *Picea* ESTs demonstrated a substantial number of transcripts with no sequence similarity to angiosperm sequences (Ralph et al. 2008). Thus, the analyses of transcriptome data will contribute not only to an enhanced understanding of the response of trees to the environment and to a comparison of different plant species, but also to a systems biology model of trees for predicting expression levels in new data (Street et al. 2011).

Within the ‘omics’ technologies, proteomics becomes increasingly fundamental in the post-genomic area. Until recently, only a few plant species have been analysed. However, the variety of plant species, including trees, used for proteomics is increasing (Kosová et al. 2011). Proteomic approaches have been conducted to understand the break of dormancy in tree seeds (Pawlowski 2010). Drought-responsive proteins were discovered in poplar and oak (Plomion et al. 2006; Durand et al. 2011; Sergeant et al. 2011). Regarding air pollution, a different proteomic profile has been shown in the leaves of poplar and European beech upon ozone fumigation (Bohler et al. 2007; Kerner et al. 2010). Oak cork formation and population variabilities were studied using proteomic analyses (Ricardo et al. 2011; Valero Galván et al. 2011). Targeted proteomics were used to understand insect attack in Norway spruce (Zulak et al. 2009). A recent review on proteomic research on forest trees summarises the state of the art for the most important species (Abril et al. 2011).

Sophisticated tools are necessary to measure metabolites and the end products of gene and protein expression. The application of metabolomics in forest tree research contributes to an understanding of their interactions with abiotic and biotic challenges. In addition wood properties and thus the chemical composition of wood are of great commercial importance. Changes in the birch leaf metabolome were observed in long-term open-field ozone fumigation experiments (Kontunen-Soppela et al. 2007; Ossipov et al. 2008). Metabolomic studies demonstrated different responses of *Populus* grown on different salt-containing soils (Brosché et al. 2005). The simulation of insect attack on Norway spruce through the application of methyl jasmonate showed that the trees had an altered terpenoid response (Zulak et al. 2009). Metabolite profiles revealed changes during embryo development in spruce (Businge et al. 2012), and the circadian clock influenced the carbon metabolism in *Populus* (Hoffman et al. 2010). Metabolite phenotyping was used for an association genetic study in loblolly pine (Eckert et al. 2012). In *Populus*, metabolite profiling in combination with a quantitative trait loci analysis was used to identify loci that control metabolite abundance (Morreel et al. 2006). Metabolome analyses clearly distinguished between salt-sensitive and -tolerant responses to water stress of poplar and *Eucalyptus* species (Janz et al. 2010; Warren et al. 2012). Regarding secondary metabolism, a recent review focuses on all aspect of genomics-based investigation, including also data mining (Chen et al. 2012).

The belowground part of trees, the rhizosphere, includes the associated microorganisms forming the mycorrhiza and also endophytic bacteria. This symbiotic relationship can be considered as a holobiontic system (Zilber-Rosenberg and Rosenberg 2008). To assess this diversity, ‘omics’ technologies are a valuable tool. A breakthrough in

unravelling this complexity was the genome sequencing of *Laccaria bicolor* (Martin et al. 2008). Next-generation sequencing will contribute to the identification of genomes of microbial eucaryotic symbionts and hosts (Medina and Sachs 2010). Courty et al. (2008) described gene expression studies in ectomycorrhizas *Lactarius quietus*–*Quercus petraea* under field conditions, and insights into the genomic level of *Populus* and the associated microbiome were recently reviewed (Podila et al. 2009). Positive interactions of root endophytes and plants were analysed by metabolic profiling in poplar (Scherling et al. 2009). Pyrosequencing of forest soils showed a high fungal diversity that was not expected (Buée et al. 2009). Furthermore, transcriptome and metabolome analyses of the holobiontic mycorrhizal system contributed to a better understanding of the forest ecosystem and its sustainability (Larsen et al. 2011).

Regarding systems biology, it is important to integrate all of the ‘omics’ data up to the physiological level, which is a great challenge, especially in tree research (Wullschleger et al. 2009). There are few examples linking transcriptome and metabolome data (Brosché et al. 2005; Druart et al. 2007; Hoffman et al. 2010; Janz et al. 2010), transcriptomic and proteomic data (Galindo González et al. 2012), and proteomics with metabolites (Kieffer et al. 2009). An integration of all ‘omics’ techniques has been reported for analysing lignin biosynthesis in aspen (Bylesjö et al. 2009), for Sitka spruce upon white pine weevil attack (Hall et al. 2011) and partially for European beech upon ozone fumigation (Matyssek et al. 2007; Ernst et al. 2012a, b).

Conclusion: where are we going?

Overwhelmed by the pure amount of information hidden in high-throughput ‘omics’ data of modern molecular biology, there is a mood for change pervading all fields of biology. The call for holistic approaches and system perspectives permeates the field. We argued that besides the methodological needs for a system-wide experimental setup, data integration and new methods of data synthesis and analysis, there is a deeper demand to rethink our scientific approach and notion of scientific progress. Changing our tools will not suffice to cope with the challenges of the biological questions posed. We will further need to rethink our methods to approach complexity. Cybernetics and systems theory provide insights which might help to guide our way.

Many aspects, some yet unfamiliar to classical experimental scientists, will lead our way to proceed in the future. Emergence, hierarchy, robustness, modularity and openness are just candidates (van Regenmortel 2004; Mazzochi 2008). Theoretical modelling will provide a better understanding

of complexity. However, this will probably not be achieved in a direct, immediate way. In silico modelling faces theoretical limits of undecidability and intractability. The most efficient way to learn about future states of the system might indeed be to let natural evolution proceed (Fernández and Solé 2006; Beckage et al. 2011). Thus, we have to change the questions we are asking. If unpredictability is an immanent feature of complex evolutionary systems, controllability might be a better goal. Concerning modelling, we have to recall what models can indeed do for us: “No substantial part of the universe is so simple that it can be grasped and controlled without abstraction. Abstraction consists in replacing the part of the universe under consideration by a model of similar but simpler structure. Models, formal or intellectual on the one hand, or material on the other, are thus a central necessity of scientific procedure” (Rosenblueth and Wiener 1945).

In the future, we might have to enlarge our theory, widening it for a holistic explanation (see e.g. Corning 1997). We might also need to rethink our language or even develop a new, more adequate form (Fox Keller 2005a). Whatever the precise path might be, it will be one of interaction among various disciplines, proceeding in cycles of experimentation, theory and simulation.

References

- Abbott A (2012) The genome of giants: a walk through the forest of tree genomes. *Tree Genet Genomes* 8:443
- Abril N, Gion J-M, Kerner R, Müller-Starck G, Navarro Cerrillo RM, Plomion C, Renaut J, Valledor L, Jorrin-Novo JV (2011) Proteomics research on forest trees, the most recalcitrant and orphan plant species. *Phytochemistry* 72:1219–1242
- Ahuja I, de Vos RCH, Bones AM, Hall RD (2010) Plant molecular stress responses face climate change. *Trends Plant Sci* 15:664–674
- Allesina S, Tang S (2012) Stability criteria for complex ecosystems. *Nature* 483:205–208
- Ashby WR (1957) *An introduction to cybernetics*. Chapman & Hall, London
- Ashby WR (1958) Requisite variety and its implications for the control of complex systems. *Cybernetica* 1:83–99
- Ashby WR (1962) Principles of the self-organizing system. In: von Foerster H, Zopf GW Jr (eds) *Principles of self-organization: transactions of the University of Illinois Symposium*. Pergamon Press, London, pp 255–278
- Baker M (2012) The changes that count. *Nature* 482:257–262
- Bao Y, Dharmawardhana P, Mockler TC, Strauss SH (2009) Genome scale transcriptome analysis of shoot organogenesis in *Populus*. *BMC Plant Biol* 9:132
- Barabási A-L, Albert R (1999) Emergence of scaling in random networks. *Science* 286:509–512
- Barabási A-L, Oltavi ZN (2004) Network biology: understanding the cell’s functional organization. *Nat Rev Genet* 5:101–113
- Bascompte J (2007) Networks in ecology. *Basic Appl Ecol* 8:485–490
- Beckage B, Gross LJ, Kauffman S (2011) The limits to prediction in ecological systems. *Ecosphere* 2:125

- Bohler S, Bagard M, Oufir M, Planchon S, Hoffman L, Jolivet Y, Hausmann J-F, Dizengremel P, Renaut J (2007) A DIGE analysis of developing poplar leaves subjected to ozone reveals major changes in carbon metabolism. *Proteomics* 7:1584–1599
- Brenner S (1999) Theoretical biology in the third millennium. *Phil Trans R Soc Lond B* 354:1963–1965
- Brosché M, Vincour B, Alatalo ER et al (2005) Gene expression and metabolite profiling of *Populus euphratica* growing in the Negev desert. *Genome Biol* 6:R101
- Buée M, Reich M, Murat C, Morin E, Nilsson RH, Uroz S, Martin F (2009) 454 pyrosequencing analyses of forest soils reveal an unexpectedly high fungal diversity. *New Phytol* 184:449–456
- Businge E, Brackman K, Moritz T, Egertsdotter U (2012) Metabolite profiling reveals clear metabolic changes during somatic development of Norway spruce (*Picea abies*). *Tree Physiol* 32:232–244
- Bylesjö M, Nilsson R, Srivastava V, Grönlund A, Johansson AI, Jansson S, Karlsson J, Moritz T, Wingsle G, Trygg J (2009) Integrated analysis of transcript, protein and metabolite data to study lignin biosynthesis in hybrid aspen. *J Prot Res* 8:199–210
- Calafapietra C, Ainsworth EA, Beier C et al (2010) Challenges in elevated CO₂ experiments on forests. *Trends Plant Sci* 15:5–10
- Cassman M (2005) Barriers to progress in systems biology. *Nature* 438:1079
- Castrillo JI, Oliver SG (2004) Yeast as a touchstone in post-genomic research: strategies for integrative analysis in functional genomics. *J Biochem Mol Biol* 37:93–106
- Chen F, Liu C-J, Tschaplinski TJ, Zhao N (2012) Genomics of secondary metabolism in *Populus*: interactions with biotic and abiotic environment. *Crit Rev Plant Sci* 28:375–392
- Cohen D, Bogeat-Triboulet M-B, Tisserant E, Balzergue S, Martin-Magniette M-L, Lelandais G, Ningre N, Renou J-P, Tamby J-P, Le Thiec D, Hummel I (2010) Comparative transcriptomics of drought responses in *Populus*: a meta-analysis of genome-wide expression profiling in mature leaves and root apices across two genotypes. *BMC Genomics* 11:630
- Corning PA (1997) Holistic Darwinism: ‘Synergistic selection’ and the evolutionary process. *J Soc Evol Syst* 20:363–400
- Courty PE, Poletto M, Duchaussoy F, Buée M, Garbaye J, Martin F (2008) Gene transcription in *Lactarius quietus-Quercus petraea* ectomycorrhizas from a forest soil. *Appl Environ Microbiol* 74:6598–6605
- Csete ME, Doyle JC (2002) Reverse engineering of biological complexity. *Science* 295:1664–1669
- Doyle JC, Csete ME (2005) Motifs, control, and stability. *PLoS Biol* 3:e392
- Druart N, Johansson A, Baba K, Schrader J, Sjödin A, Bhalerao RR, Resman L, Trygg J, Moritz T, Bhalerao RP (2007) Environmental and hormonal regulation of the activity-dormancy cycle in the cambial meristem involves stage-specific modulation of transcriptional and metabolic networks. *Plant J* 50:557–573
- Durand TC, Sergeant K, Renaut J, Planchon S, Hoffmann L, Carpin S, Label P, Morabito D, Hausman JF (2011) Poplar under drought: comparison of leaf and cambial proteomic responses. *J Proteomics* 74:1396–1410
- Eckert AJ, Pande B, Ersoz ES, Wright MH, Rashbrook VK, Nicolet CM, Neale DB (2009) High-throughput genotyping and mapping of single nucleotide polymorphisms in loblolly pine (*Pinus taeda* L.). *Tree Genet Genomes* 5:225–234
- Eckert AJ, Wegrzyn JL, Cumbie WP, Goldfarb B, Huber DA, Tolstikov V, Fiehn O, Neale DB (2012) Association genetics of the loblolly pine (*Pinus taeda*, Pinaceae) metabolome. *New Phytol* 193:890–902
- Ernst D, Bahnweg G, Heller W (2012a) Effects of abiotic and biotic stress on gene transcription in European beech: from saplings to mature trees. *Nova Acta Leopoldina* (in press)
- Ernst D, Jürgensen M, Bahnweg G, Heller W, Müller-Starck G (2012b) Common links of molecular biology with biochemistry and physiology in plants under ozone and pathogen attack. In: Matyssek R, Schnyder H, Oßwald W, Ernst D, Munch JC, Pretzsch H (eds) *Growth and defence in plants: resource allocation at multiple scales*. Ecological studies, vol 220. Springer, Heidelberg, pp 29–51
- Fernández P, Solé RV (2006) The role of computation in complex regulatory networks. In: Koonin EV, Wolf Y, Karev GP (eds) *Power laws, scale-free networks and genome biology*. Landes bioscience. Springer, New York, pp 206–225
- Fernández-Pozo N, Canales J, Guerrero-Fernández D et al (2011) EuroPineDB: a high-coverage web database for maritime pine transcriptome. *BMC Genomics* 12:366
- Fernie AR (2012) Grand challenges in plant systems biology: closing the circle(s). *Front Plant Sci* 3:35
- Filichkin SA, Breton G, Priest HD, Dharmawardhana P, Jaiswal P, Fox SE, Michael TP, Chory J, Kay S, Mockler TC (2011) Global profiling of rice and poplar transcriptomes highlights key conserved circadian-controlled pathways and cis-regulatory modules. *PLoS One* 6:e16907
- Fox Keller E (2005a) The century beyond the gene. *J Biosci* 30:3–10
- Fox Keller E (2005b) Revisiting “scale-free” networks. *Bioessays* 27:1060–1068
- Fukushima A, Kusano M, Redestig H, Arita M, Saito K (2009) Integrated omics approaches in plant systems biology. *Curr Opin Chem Biol* 13:532–538
- Galbraith DW (2006) DNA microarray analyses in higher plants. *OMICS* 10:455–473
- Galindo González LM, El Kayal W, Ju CJ-T, Allen CCG, King-Jones S, Cooke JEK (2012) Integrated transcriptomic and proteomic profiling of white spruce stems during the transition from active growth to dormancy. *Plant Cell Environ* 35:682–701
- Gershenson C (2007) The world as evolving information. In: Minai A, Braha D, Bar-Yam Y (eds) *Online Proceedings of the Seventh International Conference on Complex Systems*. Paper #17. <http://necsi.edu/events/iccs7/papers/9da71337b2793874036e781a0c6c.pdf>
- Gershenson C (2011a) The sigma profile: a formal tool to study organization and its evolution at multiple scales. *Complexity* 16:37–44
- Gershenson C (2011b) The implications of interactions for science and philosophy. Cornell University Library. arXiv:1105.2827v1. (<http://arxiv.org/pdf/1105.2827.pdf>)
- Gershenson C, Heylighen F (2003) When can we call a system self-organizing? In: Banzhaf W, Christaller T, Dittrich P, Kim JT, Ziegler J (eds) *Advances in artificial life*. 7th European Conference, ECAL 2003 LNAI 2801. Springer, Heidelberg, pp 606–614
- Gmitter FG, Chen C, Machado MA, Alves de Souza A, Ollitrault P, Froehlicher Y, Shimizu T (2012) Citrus genomics. *Tree Genet Genomes* 8:611–626
- Godsoe W, Strand E, Smith CI, Yoder JB, Esque TC, Pellmyr O (2009) Divergence in an obligate mutualism is not explained by divergent climatic factors. *New Phytol* 183:589–599
- Grattapaglia D, Vaillancourt RE, Shepard M, Thumma BR, Foley W, Külheim C, Potts BM, Myburg AA (2012) Progress in Myrtaceae genetics and genomics: *Eucalyptus* as the pivotal genus. *Tree Genet Genomes* 8:463–508
- Gupta P, Duplessis S, White H, Karnosky DF, Martin F, Podila GK (2005) Gene expression patterns of trembling aspen trees following long-term exposure to interacting elevated CO₂ and tropospheric O₃. *New Phytol* 167:129–142
- Haken H (1983) Synergetics, an introduction. In: *Nonequilibrium phase-transitions and self-organization in physics, chemistry and biology*, 3rd edn. Springer, New York

- Hall DE, Robert JA, Keeling CI, Domanski D, Quesada AL, Jancsik S, Kuzyk MA, Hamberger B, Borchers CH, Bohlmann J (2011) An integrated genomic, proteomic and biochemical analysis of (+)-3-carene biosynthesis in Sitka spruce (*Picea sitchensis*) genotypes that are resistant or susceptible to white pine weevil. *Plant J* 65:936–948
- Heylighen F (1992) Principles of systems and cybernetics: an evolutionary perspective. In: Trappl R (ed) *Cybernetics and systems'92*. World Science, Singapore, pp 3–10. (<http://pcp.lanl.gov/papers/PrinciplesCybSys.pdf>)
- Hoffman DE, Jonsson P, Bylesjö M, Trygg J, Antti H, Eriksson ME, Moritz T (2010) Changes in diurnal patterns within the *Populus* transcriptome and metabolome in response to photoperiod variation. *Plant Cell Environ* 33:1298–1313
- Hogeweg P (2011) The roots of bioinformatics in theoretical biology. *PLoS Comput Biol* 7:e1002021
- Hornik K, Stinchcombe M, White H (1989) Multilayer feedforward networks are universal approximators. *Neural Netw* 2:359–366
- Isbell F, Calcagno V, Hector A, Connolly J, Harpole WS, Reich PB, Scherer-Lorenzen M, Schmid B, van Ruijven J, Weigelt A, Wilsey BJ, Zavalety ES, Loreau M (2011) High plant diversity is needed to maintain ecosystem services. *Nature* 477:199–202
- Jansson S, Douglas CJ (2009) *Populus*: a model system for plant biology. *Annu Rev Plant Biol* 58:435–458
- Janz D, Behnke K, Schnitzler J-P, Kanawati B, Schmitt-Kopplin P, Polle A (2010) Pathway analyses of the transcriptome and metabolome of salt sensitive and tolerant poplar species reveals evolutionary adaptation to stress tolerance mechanisms. *BMC Plant Biol* 10:150
- Jeong H, Tombor B, Albert R, Oltvai ZN, Barabási A-L (2000) The large-scale organization of metabolic networks. *Nature* 407:651–654
- Joyce AR, Palsson BØ (2006) The model organism as a system: integrating ‘omics’ data sets. *Nat Rev Mol Cell Biol* 7:198–210
- Kauffman S (1969) Metabolic stability and epigenesis in randomly constructed genetic nets. *J Theor Biol* 22:437–467
- Kauffman S (1996) *At home in the universe: the search for laws of self-organization and complexity*. Oxford University Press, New York
- Kauffman S, Clayton P (2006) On emergence, agency, and organization. *Biol Philos* 21:501–521
- Kerner R, Winkler JB, Dupuy JW, Jürgensen M, Lindermayr C, Ernst D, Müller-Starck G (2010) Changes in the proteome of juvenile European beech following three years exposure to free-air elevated ozone. *iForest* 4:69–76
- Khaitovich P, Weiss G, Lachmann M, Hellmann I, Enard W, Muetzel B, Wirkner U, Ansorge W, Pääbo S (2004) A neutral model of transcriptome evolution. *PLoS Biol* 2:0682–0689
- Kieffer P, Planchon S, Oufir M, Ziebel J, Dommès J, Hoffmann L, Hausman JF, Renault J (2009) Combining proteomics and metabolite analyses to unravel cadmium stress-response in poplar leaves. *J Prot Res* 8:400–417
- Kim J (1999) Making sense of emergence. *Philos Stud* 95:3–36
- Kitano H (2002) Systems biology: a brief overview. *Science* 295:1662–1664
- Kitano H (2004) Biological robustness. *Nat Rev Genet* 5:826–837
- Kliebenstein DJ (2010) Systems biology uncovers the foundation of natural genetic diversity. *Plant Physiol* 152:480–486
- Kontunen-Soppela S, Ossipov V, Ossipova S, Oksanen E (2007) Shift in birch leaf metabolome and carbon allocation during long-term open-field ozone exposure. *Glob Change Biol* 13:1053–1067
- Kontunen-Soppela S, Parviainen J, Ruhanen H, Brosché M, Keinänen M, Thakur RC, Kohlemainen M, Kangasjärvi J, Oksanen E, Karnosky DF, Vapaavuori E (2010) Gene expression responses of paper birch (*Betula papyrifera*) to elevated CO₂ and O₃ during leaf maturation and senescence. *Environ Pollut* 158:959–968
- Kosová K, Vítámvás P, Prášil IT, Renaut J (2011) Plant proteome changes under abiotic stress—contribution of proteomics studies to understanding plant stress responses. *J Proteomics* 74:1301–1322
- Kremer A, Abbott AG, Carlson JE, Manos PS, Plomion C, Sisco P, Staton ME, Ueno S, Vendramin GG (2012) Genomics of Fagaceae. *Tree Genet Genomes* 8:583–601
- Larsen PE, Sreedasyam A, Trivedi G, Podila GK, Cseke LJ, Collart FR (2011) Using next generation transcriptome sequencing to predict an ectomycorrhizal metabolome. *BMC Syst Biol* 5:70
- Lay JO Jr, Liyanage R, Borgmann S, Wilkins CL (2006) Problems with the “omics”. *Trends Anal Chem* 25:1046–1056
- Liu W, Thummasuwan S, Sehgal SK, Chouvarine P, Peterson DG (2011a) Characterization of the genome of bald cypress. *BMC Genomics* 12:553
- Liu Y-Y, Slotine J-J, Barabási A-L (2011b) Controllability of complex networks. *Nature* 473:167–173
- Lorenz WW, Alba R, Yu YS, Bordeaux JM, Simões M, Dean JFD (2011) Microarray analysis and scale-free gene networks identify candidate regulators in drought-stressed roots of loblolly pine (*P. taeda* L.). *BMC Genomics* 12:264
- Luscombe NM, Babu MM, Yu H, Snyder M, Teichmann SA, Gerstein M (2004) Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature* 431:308–312
- Lüttge U (2012) Whole-plant physiology: synergistic emergence rather than modularity. *Prog Bot* 75 (in press)
- Martin F, Aerts A, Ahrén D et al (2008) The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. *Nature* 452:88–93
- Matyssek R, Bahnweg G, Ceulemans R, Fabian P, Grill D, Hanke DE, Kraigher H, Obwald W, Rennenberg H, Sandermann H, Tausz M, Wieser G (2007) Synopsis of the CASIROZ case study: carbon sink strength of *Fagus sylvatica* L. in a changing environment—experimental risk assessment of mitigation by chronic ozone impact. *Plant Biol* 9:163–180
- May RM (2001) *Stability and complexity in model ecosystems*. Princeton University Press, Princeton
- Maynard Smith J, Szathmáry E (1995) *The major transitions in evolution*. WH Freeman, New York
- Mazzochi F (2008) Complexity in biology. *EMBO Rep* 9:10–14
- Medina M, Sachs JL (2010) Symbiont genomics, our new tangled bank. *Genomics* 95:129–137
- Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U (2002) Network motifs: simple building blocks of complex networks. *Science* 298:824–827
- Mochida K, Shinozaki K (2011) Advances in omics and bioinformatics tools for systems analyses of plant functions. *Plant Cell Physiol* 52:2017–2038
- Morgenstern O, von Neumann J (1944) *The theory of games and economic behavior*. Princeton University Press, Princeton
- Morreel K, Goeminne G, Storme V, Sterck L, Ralph J, Coppieters W, Breyne P, Steenackers M, Georges M, Messens E, Boerjan W (2006) Genetical metabolomics of flavonoid biosynthesis in *Populus*: a case study. *Plant J* 47:224–237
- Newman MEJ (2003) The structure and function of complex networks. *SIAM Rev* 45:167–256
- Ng A, Bursteinas B, Gao Q, Mollison E, Zvelebil M (2006) Resources for integrative systems biology: from data through databases to networks and dynamic system models. *Briefings Bioinform* 7:318–330
- Nicolis G, Prigogine I (1977) *Self-organization in non-equilibrium systems*. Wiley, New York
- Noble D (2002) Modeling the heart—from genes to cells to the whole organ. *Science* 295:1678–1682
- Nurse P (2008) Life, logic and information. *Nature* 454:424–426
- Olbrich M, Gerstner E, Welzl G, Winkler JB, Ernst D (2009) Transcript responses in leaves of ozone-treated beech saplings at

- an outdoor free air model fumigation site over two growing seasons. *Plant Soil* 323:61–74
- Olbrich M, Knappe C, Wenig M et al (2010) Ozone fumigation (twice ambient) reduces leaf infestation following natural and artificial inoculation by the endophytic fungus *Apiognomonia errabunda* of adult European beech trees. *Environ Pollut* 158:1043–1050
- Ossipov V, Ossipova S, Bykov V, Oksanen E, Koricheva J, Haukioja E (2008) Application of metabolomics to genotype and phenotype discrimination of birch trees grown in a long-term field experiment. *Metabolomics* 4:39–51
- Parchman TL, Geist KS, Grahnen JA, Benkman CW, Buerkle CA (2010) Transcriptome sequencing in ecologically important tree species: assembly, annotation, and marker discovery. *BMC Genomics* 11:180
- Pavy N, Johnson JJ, Crow JA, Paule C, Kunau T, Mackay J, Retzel EF (2007) Forest TreeDB: a database dedicated to the mining of tree transcriptomes. *Nucleic Acids Res* 35:D887–D894
- Pawlowski TA (2010) Proteomic approach to analyze dormancy breaking of tree seeds. *Plant Mol Biol* 73:15–25
- Pfeiffer T, Hoffmann R (2009) Large-scale assessment of the effect of popularity on the reliability of research. *PLoS One* 4:e5996
- Plomion C, Lalanne C, Claverol S, Meddour H, Kohler A, Bogaert-Triboulet M-B, Barre A, Le Provost G, Dumazet H, Jacob D, Bastien C, Dreyer E, de Daruvar A, Guehl J-M, Schmitter J-M, Martin F, Bonneu M (2006) Mapping the proteome of poplar and application to the discovery of drought-stress responsive proteins. *Proteomics* 6:6509–6527
- Podila GK, Sreedasyam A, Muratet MA (2009) *Populus* rhizosphere and the ectomycorrhizal interactome. *Crit Rev Plant Sci* 5:359–367
- Pop A, Huttenhower C, Iyer-Pascuzzi A, Benfey PN, Troyanskaya OG (2010) Integrated functional networks of process, tissue, and development state specific interactions in *Arabidopsis thaliana*. *BMC Syst Biol* 4:180
- Prigogine I, Stengers I (1984) Order out of chaos: man's new dialogue with nature. Heinemann, London
- Proulx SR, Promislow DEL, Phillips PC (2005) Network thinking in ecology and evolution. *Trends Ecol Evol* 20:345–353
- Qiu Q, Ma T, Hu Q, Liu B, Wu Y, Zhou H, Wang Q, Wang J, Liu J (2011) Genome-scale transcriptome analysis of the desert poplar, *Populus euphratica*. *Tree Physiol* 31:452–461
- Ralph SG, Chun HJE, Kolosova N et al (2008) A conifer genomics resource of 200,000 spruce (*Picea* spp.) ESTs and 6,464 high-quality, sequence-finished full-length cDNA for Sitka spruce (*Picea sitchensis*). *BMC Genomics* 9:484
- Ransohoff DF (2005) Bias as a threat to the validity of cancer molecular-marker research. *Nat Rev Cancer* 5:142–149
- Ricardo CPP, Martins I, Francisco R, Sergeant K, Pinheiro C, Campos A, Renaut J, Fevereiro P (2011) Proteins associated with cork formation in *Quercus suber* L. stem tissue. *J Proteomics* 74:1266–1278
- Rigault P, Boyle B, Lepage P, Cooke JEK, Bousquet J, MacKay JJ (2011) A white spruce gene catalog for conifer genome analyses. *Plant Physiol* 157:14–28
- Rosenblueth A, Wiener N (1945) The role of models in science. *Philos Sci* 12:316–321
- Sauer U, Heinemann M, Zamboni N (2007) Getting closer to the whole picture. *Science* 316:550–551
- Scherling C, Ulrich K, Ewald D, Weckwerth W (2009) A metabolic signature of the beneficial interaction of the endophyte *Paenibacillus* sp. isolate and in vitro-grown poplar plants revealed by metabolomics. *Mol Plant Microb Interact* 22:1032–1037
- Sergeant K, Spieß N, Renault J, Wilhelm E, Hausman JF (2011) One dry summer: a leaf proteome study on the response of oak to drought exposure. *J Proteomics* 74:1385–1395
- Shannon CE, Weaver W (1949) The mathematical theory of communication. University of Illinois Press, Urbana
- Shulaev V, Cortes D, Miller G, Mittler R (2008) Metabolomics for plant stress response. *Physiol Plant* 132:199–208
- Sinclair TR, Purcell LC (2005) Is a physiological perspective relevant in a 'genocentric' age? *J Exp Bot* 56:2777–2782
- Smale S (1976) On the differential equations of species in competition. *J Math Biol* 3:5–7
- Sober E (1981) The principle of parsimony. *Brit J Phil Sci* 32:145–156
- Sontag ED (2004) Some new directions in control theory inspired by systems biology. *Syst Biol* 1:9–18
- Soto AM, Sonnenschein C (2006) Emergentism by default: a view from the bench. *Synthese* 151:361–376
- Street N, Jansson S, Hvidsten TR (2011) A systems biology model of the regulatory network in *Populus* leaves reveals interacting regulators and conserved regulation. *BMC Plant Biol* 11:13
- Sweetlove LJ, Fennie AR (2005) Regulation of metabolic networks: understanding metabolic complexity in the systems biology era. *New Phytol* 168:9–24
- Taylor G, Street NR, Tricker PJ, Sjödin A, Graham L, Skogström O, Calfapietra C, Scarascia-Mugnozza G, Jansson S (2005) The transcriptome of *Populus* in elevated CO₂. *New Phytol* 167:143–154
- Troggio M, Gleave A, Salvi S, Chagné D, Cestaro A, Kumar S, Crowhurst RN, Gardiner SE (2012) Apple, from genome to breeding. *Tree Genet Genomes* 8:509–529
- Turing A (1952) The chemical basis of morphogenesis. *Phil Trans Roy Soc B* 237:37–72
- Tuskan GA, DiFazio S, Jansson S et al (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313:1596–1604
- Ueno S, Le Provost G, Léger V et al (2010) Bioinformatic analysis of ESTs collected by Sanger and pyrosequencing methods for a keystone forest tree species: oak. *BMC Genomics* 1:650
- Urano K, Kurihara Y, Seki M, Shinozaki K (2010) 'Omics' analyses of regulatory networks in plant abiotic stress responses. *Curr Opin Plant Biol* 13:132–138
- Valero Galván J, Valledor L, Navarro Cerrillo RM, Pelegrín EG, Jorrín-Novo JV (2011) Studies of variability in holm oak (*Quercus ilex* subsp. *ballota* [Desf.] Samp.) through acorn protein profile analysis. *J Proteomics* 74:1244–1255
- van Regenmortel MHV (2004) Reductionism and complexity in molecular biology. *EMBO Rep* 5:1016–1020
- von Bertalanffy L (1969) General system theory: foundations, development, applications. Braziller, New York
- von Neumann J (1966) Theory of self-reproducing automata. University of Illinois Press, Urbana
- Wacholder S, Chanock S, Garcia-Closas M, Elghormli L, Rothman N (2004) Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. *J Natl Cancer Inst* 96:434–442
- Warren CR, Aranda I, Cano FJ (2012) Metabolomics demonstrates divergent responses of two *Eucalyptus* species to water stress. *Metabolomics* 8:186–200
- Weckwerth W (2011) Green systems biology—from single genomes, proteomes and metabolomes to ecosystems research and biotechnology. *J Proteom* 75:284–305
- Westerhoff HV, Palsson BO (2004) The evolution of molecular biology into systems biology. *Nat Biotech* 22:1249–1252
- Wiener N (1948) Cybernetics, or control and communication in the animal and the machine. MIT Press, Cambridge
- Wullschlegel SD, Weston DJ, Davis JM (2009) *Populus* responses to edaphic and climatic cues: emerging evidence from systems biology research. *Crit Rev Plant Sci* 28:368–374

- Yang X, Kalluri UD, DiFazio SP, Wullschleger SD, Tschaplinski TJ, Cheng Z-M, Tuskan GA (2009) Poplar genomics: state of the science. *Crit Rev Plant Sci* 28:285–308
- Yuan SJ, Galbraith DW, Dai SY, Griffin P, Stewart CN Jr (2008) Plant systems biology comes of age. *Trends Plant Sci* 13: 165–171
- Zilber-Rosenberg I, Rosenberg E (2008) Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. *FEMS Microbiol Rev* 32:723–735
- zu Castell W (2012) Complex systems: chances and risks for experimental data analysis. *Nova Acta Leopoldina* (in press)
- Zulak KG, Lippert DN, Kuzyk MA, Domanski D, Chou T, Borchers CH, Bohlmann J (2009) Targeted proteomics using selected reaction monitoring reveals the induction of specific terpene synthases in a multi-level study of methyl jasmonate-treated Norway spruce (*Picea abies*). *Plant J* 60:1015–1030