

Chapter 1

Recent Development in Omics Studies



Wan Mohd Aizat, Ismanizan Ismail, and Normah Mohd Noor

Abstract The central dogma of molecular biology (DNA, RNA, protein and metabolite) has engraved our understanding of genetics in all living organisms. While the concept has been embraced for many decades, the development of high-throughput technologies particularly omics (genomics, transcriptomics, proteomics and metabolomics) has revolutionised the field to incorporate big data analysis including bioinformatics and systems biology as well as synthetic biology area. These omics approaches as well as systems and synthetic biology areas are now increasingly popular as seen by the growing numbers of publication throughout the years. Several journals which have published most of these related fields are also listed in this chapter to overview their impact and target journals.

Keywords Genomics · Metabolomics · Molecular biology · Proteomics · Systems biology · Transcriptomics

1.1 The Central Dogma of Molecular Biology and Beyond

The central dogma of molecular biology states that the genetic materials of all living being are encoded by their unique DNA sequences (also called gene), transcribed to RNA (also called transcript) and subsequently translated to proteins as the main catalytic entities (Fig. 1.1). Such concept introduced in the early 1950s certainly has been the main limelight for various research in the world, spanning all three domains of life: eukaryotes, bacteria and archea [1]. Watson and Crick were awarded the Nobel Prize in 1962 for such revolutionary insights (not to be forgotten the contribution by Rosalind Franklin who produced the X-ray of DNA), and later, Holley, Khorana and Nirenberg also won the prize in 1968 for completing the genetic coding of protein synthesis [1]. This traditional concept of molecular biology has since then been improved with the addition of metabolite at the end of the workflow

W. M. Aizat (✉) · I. Ismail · N. M. Noor
Institute of Systems Biology, Universiti Kebangsaan Malaysia (UKM), Bangi, Malaysia
e-mail: wma@ukm.edu.my; maniz@ukm.edu.my; normah@ukm.edu.my

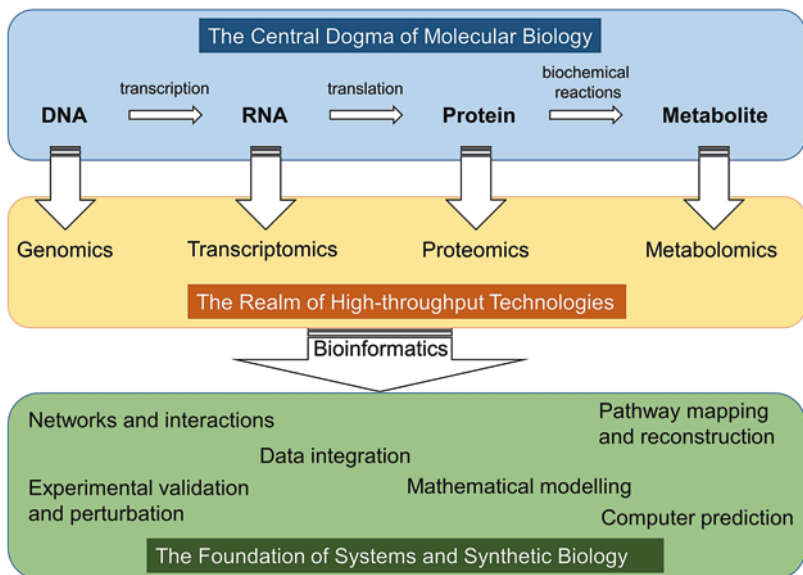


Fig. 1.1 The central dogma of molecular biology comprising of DNA, RNA, protein and metabolite has been greatly investigated by the invention of high-throughput technologies in respective omics (genomics, transcriptomics, proteomics and metabolomics). Such massive data generation requires bioinformatics to analyse and integrate them, ultimately leading to the foundation of systems and synthetic biology

(Fig. 1.1), signifying the importance of the products of biochemical activities catalysed by proteins to influence phenotypic characteristics. One obvious example is how *capsanthin-capsorubin synthase* gene encodes the final enzyme in the biosynthesis of pigmented compounds that gives rise to the red colour of chilies [2, 3]. Without a complete workflow of central dogma, encompassing the downstream metabolites, one would not be able to comprehend the observable characteristics of organisms.

While such a simple relationship between nucleic acids (DNA and RNA) and proteins as well as metabolites has been the mainstream of science research for decades, new formulations of big data science are now taking the stage. Specifically, the introduction of suffixes -omics and -omes has expanded our view in molecular biology. Rather than reductionist approach of focusing on only one or a few genes/proteins, omics approach allows thorough investigation of each facet of molecular biology [4], be it gene (genomics), transcript (transcriptomics), protein (proteomics) and metabolites (metabolomics). While many other omics exist [5], these four omics have been the main themes in molecular biology research, perhaps due to their close relationship with the central dogma itself. Furthermore, these omics approaches are the very foundation of the systems and synthetic biology fields (Fig. 1.1).

1.2 Genomics and Transcriptomics

Investigation of nucleic acids' composition comprising of four bases (adenine, cytosine, guanine and thymine (DNA) or uracil (RNA)) in either genome or transcriptome was greatly facilitated by the development of sequencing platforms. Since the introduction of automated DNA sequencing (1986) as well as the more recent Next - Generation Sequencing (2000) [1, 6], both genomics and transcriptomics areas have exploded with the increasing number of published articles (Figs. 1.2a and 1.3a). Some of the top journals in these areas are *PLOS ONE*, *Nucleic Acids Research* and *BMC Genomics* (Figs. 1.2b and 1.3b). We have combined the great details of the applications and concepts of both omics in Chap. 2 “Functional Genomics”.

1.3 Proteomics

Proteins are the workhorses of cellular and biochemical processes. Understanding organism at the level of functional genomics alone may not tell the whole story of how organism functions and responds to environment [7]. As such, proteomics is a vital area to bridge the gap between the more static genome and observable phenotypic characteristics. Evidently, the number of publications related to this area has increased exponentially since the late 1990s (Fig. 1.4a). This is perhaps also due to

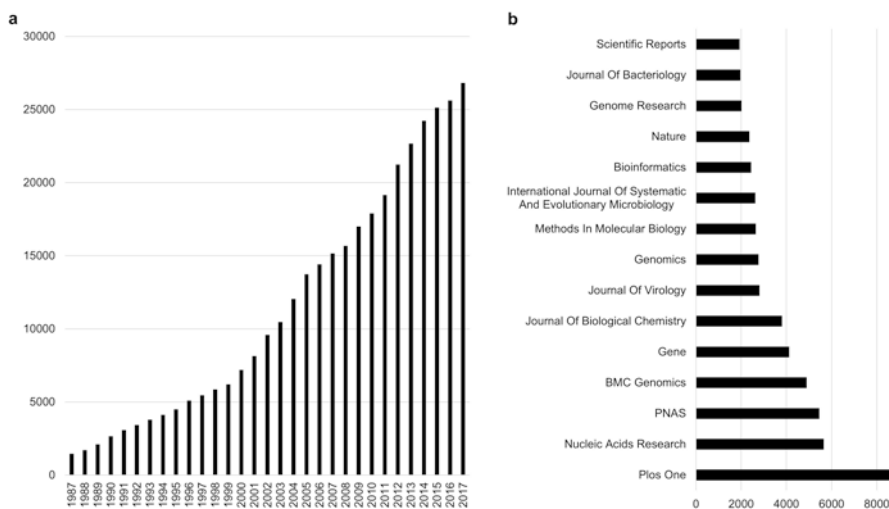


Fig. 1.2 Genomics has become an increasingly attractive research area as shown by the surge of published articles over the years (a). The top 15 journals that published the most genomics-related papers of all time are also shown in (b). Statistics were obtained from SCOPUS database in July 2018 by searching “Genomic*” in the “Article title, Abstract and Keywords” search field. *PNAS*, *Proceedings of the National Academy of Sciences of the United States of America*

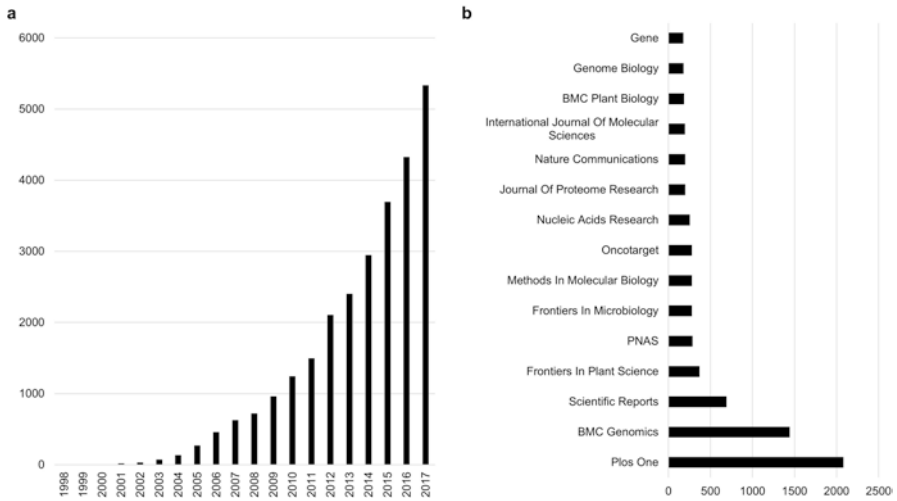


Fig. 1.3 Published articles for transcriptomics studies have steeply increased over the past 20 years (a). The top 15 journals that published the most transcriptomics-related papers of all time are also depicted in (b). Statistics were obtained from SCOPUS database in July 2018 by searching “Transcriptomic*” in the “Article title, Abstract and Keywords” search field. *PNAS*, *Proceedings of the National Academy of Sciences of the United States of America*

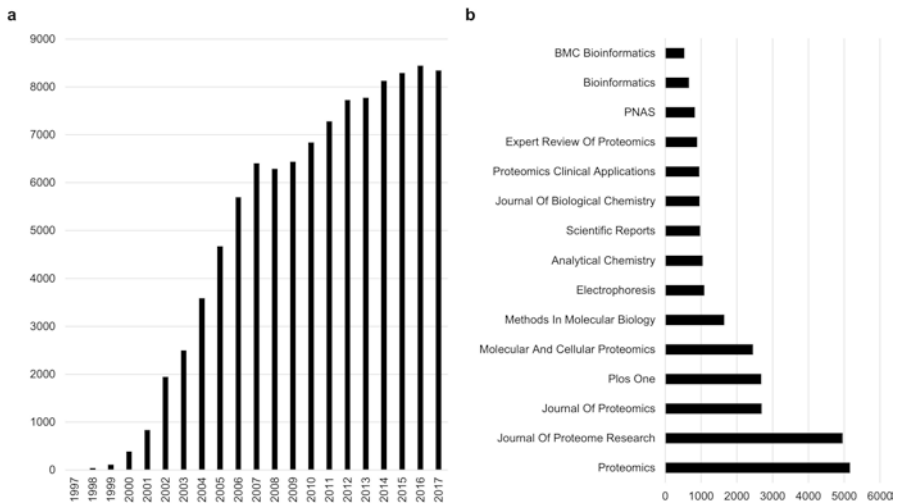


Fig. 1.4 The surge of published articles related to proteomics over the years (a) may have been contributed by mass spectrometry development. The top 15 journals that published the most proteomics-related papers of all time are also depicted in (b). Statistics were obtained from SCOPUS database in July 2018 by searching “Proteomic*” in the “Article title, Abstract and Keywords” search field. *PNAS*, *Proceedings of the National Academy of Sciences of the United States of America*

the advent and development in mass spectrometry technologies that offer accurate, fast and sensitive methods for protein identification and quantitation [8–10]. The inventors of soft ionisation method in mass spectrometry, John Fenn and Koichi Tanaka, had also been awarded with a Nobel Prize in 2002, suggesting the importance of such system in biomolecular discovery. Some of the key journals that published proteomics papers are *Proteomics*, *Journal of Proteome Research*, *Journal of Proteomics* and *Molecular and Cellular Proteomics* (Fig. 1.4b). We have detailed the applications of proteomics in Chap. 3 “Proteomics in Systems Biology”.

1.4 Metabolomics

While proteins may be the workhorses in cells, metabolites also serve as important molecules to be studied in living organism. Metabolomics is a relatively new area compared to the earlier omics, yet its significance cannot be denied. Similar to proteomics, metabolomics also relies on the development of reliable and accurate mass spectrometry systems [11–13]. Such development catalysed the remarkable growth in the paper published in this area since the early 2000s (Fig. 1.5a). Several journals cover this research area extensively such as *Metabolomics*, *PLOS ONE* and *Analytical Chemistry* (Fig. 1.5b). In Chap. 4 “Metabolomics in Systems Biology”, this research field is described with emphasis to its applications and current workflow.

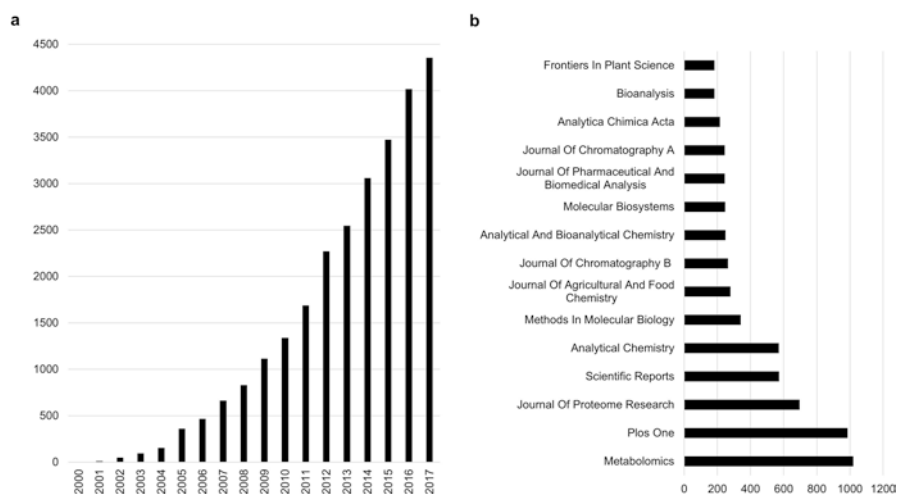


Fig. 1.5 Metabolomics-related articles have been exponentially growing over the years (a). The top 15 journals that published the most metabolomics-related papers of all time are shown in (b). Statistics were obtained from SCOPUS database in July 2018 by searching “Metabolomic*” in the “Article title, Abstract and Keywords” search field

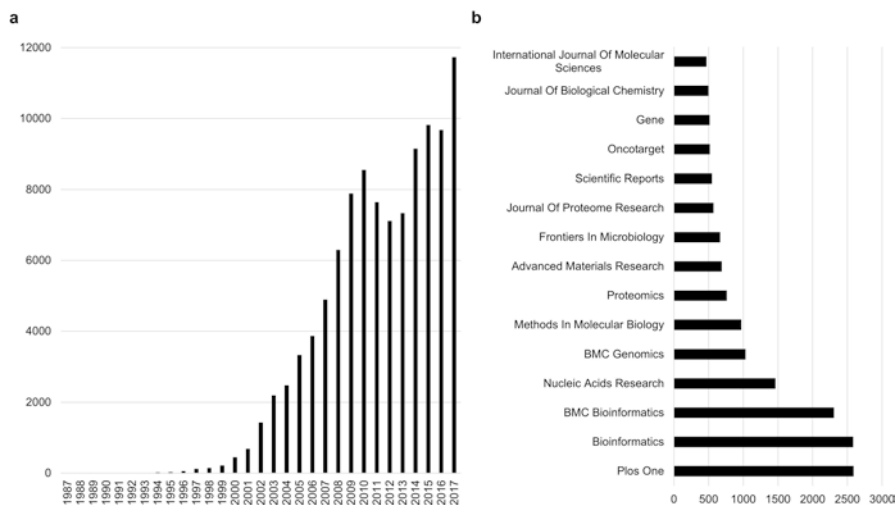


Fig. 1.6 The number of published articles in the area of bioinformatics has surged exponentially over the years (a). The top 15 journals that published the most bioinformatics-related papers of all time are indicated in (b). Statistics were obtained from SCOPUS database in July 2018 by searching “Bioinformatic*” in the “Article title, Abstract and Keywords” search field

1.5 Bioinformatics

Despite the high-throughput technologies that each omics approach offers, a significant portion of the analysis rely on bioinformatics analysis. Such massively generated data often hamper scientists with “too much” information that may need some sort of computer prediction and analysis to speed up the process [14, 15] (Fig. 1.1). Bioinformatics can be done at each level of omics or in an effort to combine a few omics’ results together. This is to “make sense” of the data for scientists to decipher the research question at hand. Furthermore, this research area has certainly benefited with the increasingly powerful computer systems available. Papers related to bioinformatics have reached nearly 12,000 marks in the year 2017 alone, a stark contrast compared to the number of publication in the 1990s (Fig. 1.6a). Some of the journals that actively publish bioinformatics papers include *PLOS ONE*, *Bioinformatics* and *BMC Bioinformatics* (Fig. 1.6b). The concept and applications of bioinformatics in understanding the various omics are detailed in Chap. 5 “Integrative Multi-omics Through Bioinformatics”.

1.6 Systems and Synthetic Biology

The different omics concepts have certainly revolutionised the understanding of our modern concept of biology, sparking various technological improvements and breakthrough in science. Systems biology is one area that was initiated by the efforts in

understanding cell as a whole rather than its parts [16–18]. This area mainly integrates the different omics to understand the whole process of biochemistry in living being. Systems understanding of biology can be a daunting task, requiring experts in various fields including computer science and wet lab scientists to interrogate and integrate the plethora of information generated (Fig. 1.1). Systems biology ultimately aims to develop mathematical models that explain a biological system by systematically perturbing them in series of experiments [17, 18]. The different approaches in omics will be fed into making the working model to be tested in the experiments. Interestingly, while the papers in systems biology area actively increased from the late 1990s to 2012, a declining trend is seen in recent years (Fig. 1.7a). This may be due to the complexity of the whole topic itself, requiring a more specialised skills to develop the model further [19]. The top journals publishing in this research endeavour are *BMC Systems Biology*, *PLOS ONE*, *Bioinformatics* and *Biosystems* (Fig. 1.7b).

Omics research is also important for the synthetic biology area where biological systems can be designed, interrogated and tested using genetic engineering and genome editing techniques [20]. This research area has actively grown over the last 10 years (Fig. 1.8a), and some of the related journals are *ACS Synthetic Biology*, *Methods in Molecular Biology* and *Proceedings of the National Academy of Sciences of the United States of America (PNAS)* (Fig. 1.8b). This topic is detailed in Chap. 6 “Metabolic Engineering and Synthetic Biology”.

In this book, each of the chapters is comprised of detailed description of the applications of each omics and its general methodology. We have also expanded the view by incorporating how such omics contribute to the systems biology perspective. While more details on such an infant area of science is duly needed, we reserved the details for now to give broad understanding of this new topic. This is to ensure read-

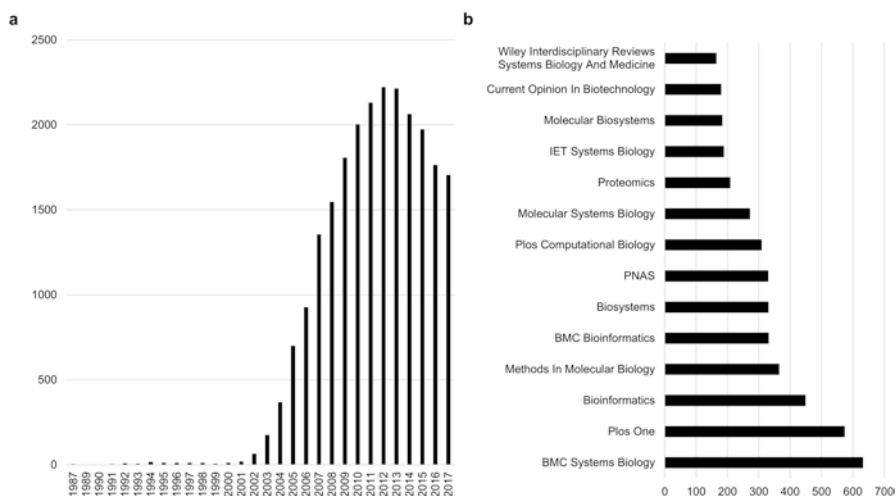


Fig. 1.7 Systems biology has become an increasingly attractive research area as shown by the surge of published articles over the 30-year period (a). The top 15 journals that published the most systems biology-related papers of all time are indicated in (b). Statistics were obtained from SCOPUS database in July 2018 by searching “Systems biology” in the “Article title, Abstract and Keywords” search field. *PNAS*, *Proceedings of the National Academy of Sciences of the United States of America*

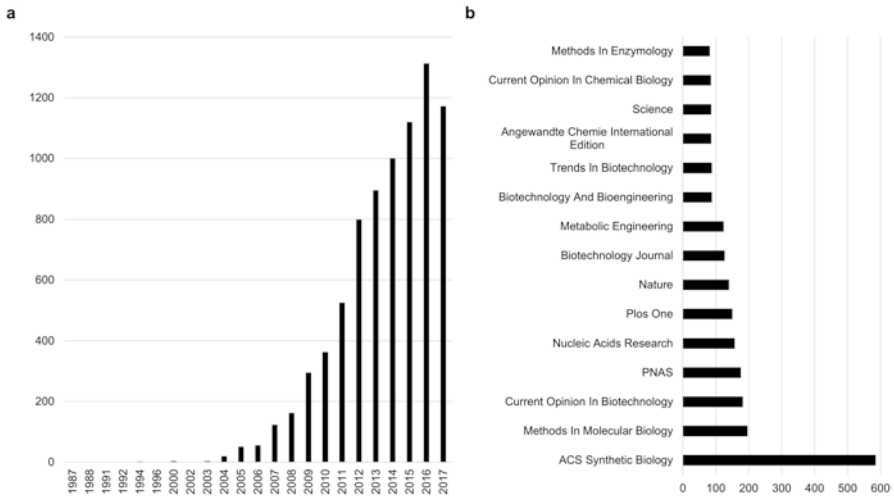


Fig. 1.8 Synthetic biology publications have grown exponentially over the last 30 years (a). The top 15 journals that published the most systems biology-related papers of all time are indicated in (b). Statistics were obtained from SCOPUS database in July 2018 by searching “Systems biology” in the “Article title, Abstract and Keywords” search field. *PNAS*, *Proceedings of the National Academy of Sciences of the United States of America*



Fig. 1.9 Kesum (*Persicaria minor*) has been utilised in various omics research for systems understanding of this tropical herb

ers grasp the omics concept well and are able to appreciate the level of information generated from these approaches to feed into the systems and synthetic biology research. Furthermore, we have also highlighted our current efforts in understanding tropical herbs such as *Persicaria minor* (syn. *Polygonum minus* or kesum) (Fig. 1.9) and how omics technologies were utilised to characterise this non-model organism.

References

1. Weaver RF (2012) Molecular biology. McGraw Hill, New York
2. Guzman I, Hamby S, Romero J, Bosland PW, O'Connell MA (2010) Variability of carotenoid biosynthesis in orange colored *Capsicum* spp. *Plant Sci* 179:49–59
3. Ha SH, Kim JB, Park JS, Lee SW, Cho KJ (2007) A comparison of the carotenoid accumulation in *Capsicum* varieties that show different ripening colours: deletion of the capsanthin-capsorubin synthase gene is not a prerequisite for the formation of a yellow pepper. *J Exp Bot* 58:3135–3144
4. Voit EO (2012) A first course in systems biology. Garland Science, New York
5. Baker M (2013) Big biology: the 'omes puzzle'. <http://www.nature.com/news/big-biology-the-omes-puzzle-1.12484>
6. Ignacimuthu S (2005) Basic bioinformatics. Alpha Science Int'l Ltd., Oxford
7. Twyman RM (2013) Principles of proteomics. Garland Science, Abingdon
8. Baginsky S (2009) Plant proteomics: concepts, applications, and novel strategies for data interpretation. *Mass Spectrom Rev* 28:93–120
9. Domon B, Aebersold R (2006) Mass spectrometry and protein analysis. *Science* 312:212–217
10. Domon B, Aebersold R (2010) Options and considerations when selecting a quantitative proteomics strategy. *Nat Biotechnol* 28:710–721
11. Dettmer K, Aronov PA, Hammock BD (2007) Mass spectrometry-based metabolomics. *Mass Spectrom Rev* 26:51–78
12. Patti GJ, Yanes O, Siuzdak G (2012) Innovation: metabolomics: the apogee of the omics trilogy. *Nature* 13:263
13. Villas-Bôas SG, Mas S, Åkesson M, Smedsgaard J, Nielsen J (2005) Mass spectrometry in metabolome analysis. *Mass Spectrom Rev* 24:613–646
14. Joyce AR, Palsson BØ (2006) The model organism as a system: integrating 'omics' data sets. *Nature* 7:198–210
15. Mochida K, Shinozaki K (2011) Advances in omics and bioinformatics tools for systems analyses of plant functions. *Plant Cell Physiol* 52:2017–2038
16. Gatherer D (2010) So what do we really mean when we say that systems biology is holistic? *BMC Syst Biol* 4:1–22
17. Ideker T, Galitski T, Hood L (2001) A new approach to decoding life: systems biology. *Annu Rev Genomics Hum Genet* 2:343–372
18. Kitano H (2002) Systems biology: a brief overview. *Science* 295:1662–1664
19. Chubukov V, Mukhopadhyay A, Petzold CJ, Keasling JD, Martín HG (2016) Synthetic and systems biology for microbial production of commodity chemicals. *Nature* 2:16009
20. Cameron DE, Bashor CJ, Collins JJ (2014) A brief history of synthetic biology. *Nat Rev Microbiol* 12:381–390