Systems and synthetic biology perspective of the versatile plant-pathogenic and polysaccharide-producing bacterium Xanthomonas campestris

Sarah Schatschneider, †‡ Jessica Schneider, †‡ Jochen Blom, ‡ Fabien Létisse, ‡ Karsten Niehaus, †
Alexander Goesmann †‡ and Frank-Jörg Vorhölter ‡§

Abstract

Bacteria of the genus Xanthomonas are a major group of plant pathogens. They are hazardous to important crops and closely related to human pathogens. Being collectively a major focus of molecular phytopathology, an increasing number of diverse and intricate mechanisms are emerging by which they communicate, interfere with host signaling and keep competition at bay. Interestingly, they are also biotechnologically relevant polysaccharide producers. Systems biotechnology techniques have revealed their central metabolism and a growing number of remarkable features. Traditional analyses of Xanthomonas metabolism missed the Embden–Meyerhof–Parnas pathway (glycolysis) as being a route by which energy and molecular building blocks are derived from glucose. As a consequence of the emerging full picture of their metabolism process, xanthomonads were discovered to have three alternative catabolic pathways and they use an unusual and reversible phosphofructokinase as a key enzyme. In this review, we summarize the synthetic and systems biology methods and the bioinformatics tools applied to reconstruct their metabolic network and reveal the dynamic fluxes within their complex carbohydrate metabolism. This is based on insights from omics disciplines; in particular, genomics, transcriptomics, proteomics and metabolomics. Analysis of high-throughput omics data facilitates the reconstruction of organism-specific large- and genome-scale metabolic networks. Reconstructed metabolic networks are fundamental to the formulation of metabolic models that facilitate the simulation of actual metabolic activities under specific environmental conditions.

SYSTEMS BIOLOGY

In recent years, systems and synthetic biology have substantially increased our understanding of many processes in life sciences at molecular and cellular levels [1–3]. Systems biology intends to understand the cell from a system-wide perspective. For over 100 years, life sciences have aimed at elucidating the details of molecular processes in organisms. In systems biology, scientists have started to inter-relate this data on a large scale in order to analyse the interactions of all elements [4, 5], and thus initiate the decoding of entire systems, aiming at their quantitative understanding. This became possible with the development of high-throughput techniques associated with diverse computational methods and the availability of faster computing. Fundamental high-throughput methods are now routinely available in the fields of genomics, transcriptomics, metabolomics and proteomics, while additional specialized branches of omics disciplines have been developed, such as lipidomics [6–8], glycomics [9–11] and ¹³C fluxomics [12, 13].
In quantitative biology, functional modules are reconstructed in order to understand biology at a systems level, thereby transcending the boundaries between individual omics disciplines. Omics data are used to generate metabolic models that quantitatively capture the cellular metabolism and predict phenotypes. Such model-based predictions can be validated by in vivo measurements. Based on in vivo data, metabolic models can be further expanded or enhanced to increase the precision and scope of predictions. This can result in an iterative process of experimental validation and model enhancement [2, 14], wherein data from different omics disciplines can be progressively integrated into predictive models [15, 16].

The application of systems biology methods has been particularly successful in the field of metabolic engineering and biotechnology [17–21], also called systems biotechnology or industrial systems biology [22] where it has reached industrial production [23]. When cell components are fully described by omics analyses, processes, dynamics and interactions, biotechnological processes can be optimized and the enhanced understanding of the system can be applied to manipulate organisms for targeted optimization and increased product yield. Examples of the application of systems biotechnology include the production of biofuels [24], polymers [25, 26] and succinic acid [27–29], an important building block for bio-based production [30].

However, systems-based modelling approaches are not confined to biotechnology. There are ongoing initiatives to exploit the benefits of systems biology for a deeper understanding of pathogenicity [31–33] and host–pathogen interaction at a systems level has been reviewed [34, 35]. Understandably, initial research focused on human pathogens. Salmonella, a major human pathogen, has been used as a model for bacterial pathogenicity [36]. A community of 20 experts collaborated to merge and expand a metabolic reconstruction of two Salmonella typhimurium networks and compared it to a third published network of Salmonella typhimurium [37]. The resulting metabolic network was validated using data of growth on different nutrients and applied to identify potential multi-target drug therapy approaches. As a next step for host–pathogen modelling, a review was published stating the elements of a systems biology approach and discussing applications for gaining insight into the processes involved in the pathogenesis by Salmonella and Yersinia in host–pathogen interactions [38]. The authors reviewed systems biology approaches to analyse host–pathogen interactions of the pathogens Salmonella and Yersinia under a variety of conditions mimicking pathogenesis. They discuss omics dataset integration to improve genome annotation, highlighting the importance of more sophisticated computational approaches and the integration of multidimensional high-throughput data into inference and knowledge-based models. A general platform for the organization and exchange of models and systems biology data is the SEEK database. This web-based, open-source database was originally developed for the SysMO Consortium [39].

**INITIAL SYSTEMS ANALYSIS OF PLANT-PATHOGENIC XANTHOMONAS CAMPESTRIS PV. CAMPESTRIS**

Working with human pathogens is risky and host–pathogen interactions are difficult to study in situ; sometimes it is only possible to study artificial representations of host–pathogen systems. In contrast, studying plant pathogenic bacteria has the advantage of being harmless to humans and hardly any ethical issues arise when their infection mechanisms are analysed in vivo. Moreover, some infection mechanisms and virulence factors are surprisingly similar in animal and plant pathogenic bacteria [40, 41].

Systems-based analysis of plant pathogens has been initiated in recent years. Pioneering studies relating to bacteria of the genus Xanthomonas were among the first in this research field [42, 43]. Xanthomonads are hazardous to a huge number of plants worldwide and cause substantial losses among a wide variety of agricultural crops. Xanthomonas campestris pv. campestris (Xcc) is the causal agent of black rot disease [44], which affects plants of the family Brassicaceae including the model plant Arabidopsis thaliana [45]. In a survey of phytopathologists, X. campestris pathovars were regarded as being among the top five important phytopathogenic bacteria [46]. Xcc enters its host plants via wounds or hydathodes and spreads in the xylem before systemically infecting other plant tissues. Other xanthomonads invade intercellular spaces of the mesophyll parenchyma tissue of the host. Xanthomonads are γ-proteobacteria like Escherichia coli. This phylogenetic class includes other important plant pathogens like Pseudomonas syringae [47, 48]. The family Xanthomonadaceae also includes the human pathogen Stenotrophomonas maltophilia [49, 50].

Xcc produces large quantities of the biopolymer xanthan gum, which is used in industry as a thickening agent [51]. Xanthan is a heteropolysaccharide, with a main chain that consists of a cellulose-like β-1,4-glucan with mannose-glucuronate-mannose trisaccharide side chains attached to every second glucose moiety [52]. The mannose residues of the side chains can carry acetyl [53] or pyruvyl group [54] decorations at varying degrees. Important commercial applications are in the food, cosmetics and pharmaceutical industries, and in drilling for oil and natural gas.

Xanthan is assumed to have a function in pathogenicity, although details are still subject to discussion. General functions attributed to exopolysaccharides like xanthan are protection against external stress factors such as radiation [55], varying pH values, high salt concentrations and drought [56, 57]. Xanthan pyruvylation is essential for the virulence of Xcc [58]. Xanthan is not essential for the infection of citrus plants but it contributes to the epiphytic survival of the pathogen [59]. Xanthan secreted by Xcc suppresses the deposition of callose, a plant polysaccharide, and induces susceptibility to pathogenesis [60]. The signal transduction of plants, which is initiated in response to the specific detection of bacterial pathogens based on microbe-associated molecular patterns
For xanthomonads, systems-based approaches have mainly been conducted for the strain Xcc B100. Initial steps were taken in the pre-genome era, when the analysis of metabolic key reactions facilitated the first flux predictions in X. campestris NRRL B-1459 [71]. This involved the generation of an unstructured kinetic model wherein xanthan production was described and compared to experimental data obtained from batch cultivation in a bioreactor. This model was suitable for describing metabolic phenomena such as sequential consumption of nitrogen sources, the consumption of inorganic phosphate and carbon, the evolution of biomass, and the production of xanthan. The model was used to predict parameters favourable for enhanced xanthan production by limiting phosphate concentrations in the medium. The model predictions were verified by comparison to measurements from fermentation in a bioreactor [42]. Subsequently, oxygen transport rates were included in kinetic modelling [72]. The kinetic studies were carried out with X. campestris strain NRRL B-1459 (ATCC 13951) long before the recently published draft genome [73] became available. Both of these kinetic models are pioneering studies in Xanthomonas research. Yet the kinetic models are based on a handful of reactions and on parameter estimations. Additionally, regulatory steps are not considered. Therefore, an extended model was required that included all reactions and regulatory steps for a systems-wide analysis of xanthomonads, reflecting xanthan production plant infection. To understand the whole system and to build a systems-wide model, the first crucial step was obtaining complete and meticulously annotated genomic data.

GENOMIC BASICS

Genome sequencing projects laid the essential foundations for systems-based analyses as they are performed today. Among xanthomonads, there was a focus on the Xcc strain B100 for systems biology-related studies. The complete genome sequence is available for this strain [74, 75] with the genome assembly validated using an end-sequence fosmid library [74, 76]. The functional genome annotation of Xcc B100 was carried out using the integrated genome annotation system GenDB [77] to obtain automated annotations before subsequent checks and enhancements were carried out by specialists. While annotations of pathogenicity-related genome features were transferred from the genome of Xanthomonas euvesicatoria 85-10, which is often addressed by its traditional name X. campestris pv. vesicatoria [78], analysis of the Xcc B100 genome had a clear focus on metabolism, in particular on the processes relevant for xanthan biosynthesis. This included, for example, carbohydrate uptake systems [79-82], gluconeogenesis [83] and the Entner–Doudoroff (ED) and pentose-phosphate (PP) pathways [42, 84, 85], information on the biosynthesis of sugar nucleotide biosynthetic intermediates [86-88] or lipopolysaccharide (LPS) [89-91], and information on the genes directly involved in xanthan polymerization and export [92-94]. Analysis of the Xcc B100 genome revealed an unexpected wealth of uptake systems for mono- and disaccharides as well as several genes for nucleotide sugar biosynthetic enzymes that had not been expected based on the previous experimental characterization of xanthomonads [74].

In addition to Xcc B100, further Xanthomonas genomes have been sequenced: the early availability of two other complete X. campestris pv. campestris genomes for strains 8004 [95] and ATCC 33913 [96], like the publication of the complete X. campestris pv. raphani strain 75C genome [97] and the draft genomes of X. campestris strain JX [98], X. campestris pv. campestris strain Xca5 [99] and X. campestris strains CN14, CN15, CN16 [100], LMG 8031 [101] and NRRL B-1459 [73], provided further data from phylogenetically close related strains. Software-based comparative analysis of such genome data is greatly facilitated by using the EDGAR software tool (Fig. 1, b) [102]. A new version of EDGAR was published recently [103]. The main focus of the analysis of Xanthomonas genomes is usually on virulence factors [66]. Initial characterization of the genomes of 45 additional X. campestris strains might open the door to an even wider data set for this purpose [104]. In addition, more and more genome data are becoming available for other Xanthomonas species [66]. While high-quality draft genome data are required for many analyses, in particular when the focus is on protein gene products, the availability of complete genome data is advantageous for reference strains of more general interest as it facilitates in-depth post-genomic analyses in fields such as transcriptomics, which became obvious with the emergence of RNA-Seq technologies [105, 106].

To annotate genomes, their constituent genes need to be identified before functions can be assigned. The quality of the annotation process is crucial, as missed or wrongly assigned properties affect subsequent functional analysis. This concerns, in particular, the reconstruction of metabolic networks – usually the first step required to build a metabolic model. Metabolic models often include hundreds [43] to thousands [107] of gene products. Erroneous or missing functional assignments result in misleading gaps or in dead ends in metabolic pathways, which substantially impede the scope and precision of the predictions generated. Fortunately, there are specialized bioinformatics tools for predicting the coding sequences (CDs) of protein-coding genes that usually provide satisfactory results for prokaryotes. In the past, tools like GeneMark [108, 109], Glimmer [110, 111], Critica [112], Reganor [113], Gismo [114] and Prodigal [115] were used for CDS prediction, and up-to-date versions of Glimmer, GeneMark and Prodigal are still popular.
A critical aspect is often the correct identification of translational start sites for CDSs. Considering appropriate bioinformatics tools may again help in finding such start codons in an optimal way. The wider application of RNA-Seq is starting to increase the availability of experimentally determined transcriptional start sites. This is likely to result in higher quality start codon predictions, as recently indicated for Xcc B100 [116]. In contrast to protein-coding genes, the identification of RNA genes is still poorly conceived. There are reliable results for ribosomal RNAs and tRNAs, while experimental evidence is usually required to identify small non-coding RNAs [117]. For known RNA genes, the Rfam database is a valuable tool [118]. Once genes have been reliably predicted, bioinformatics tools can help to elucidate their functions. Again, there is a wealth of software tools available for the functional annotation of CDSs while providing comparable functionality for RNA genes remains in development. The intelligent combination of multiple tools into rule-based pipelines has the potential of providing automated and expedient functional annotations (Table 1). Launching novel workflow engines [119] could further enhance such analysis pipelines; however, proofreading and curation by human experts still is essential to avoid mistakes.
For metabolic analyses, the manually curated genome annotation of Xcc B100 was enhanced in several iterations. For this purpose, published metabolic data provided over subsequent years were manually integrated using the genome annotation software GenDB ([77], Fig. 1, c). This included information relating to the uptake and utilization of nutrients from the environment [79, 120–124]. Further important metabolic studies considered in the iterative enhancement of functional genome annotation [74] focused on the biosynthesis of amino acids [107, 125] and a novel, yeast-like phosphatidylcholine synthesis route [126]. A common theme of these studies was the discovery of novel biosynthetic routes that differ from the respective metabolic pathways of other bacterial model organisms.

**VALIDATION AND AUGMENTATION OF GENOME ANNOTATIONS BASED ON TRANSCRIPTOMICS**

Besides the outcomes of traditional functional studies, results from post-genomic approaches added to the data available for genome annotation (Fig. 2). Our knowledge of transcriptomic regulation and metabolic-related genes benefited from this input. Genomic data enabled the design of genome-wide microarrays to globally explore the expression of genes and thereby complement static genomic data with a dynamic dimension to facilitate novel functional insights [127, 128]. For Xcc B100, the manually curated genome annotation was used to obtain a complete set of unambiguous probes individually representing all predicted genes. Obviously, such an experimental concept benefits from a high-quality, comprehensive genome annotation. However, there are conflicts between the aims of designing unique probes for all genes including those with sequence similarities and the necessity of obtaining oligonucleotides with physical properties useful for hybridization procedures. Unfavourable trade-offs in these fields are often unavoidable [129]. The underlying drawbacks of microarray design have been reduced by technological progress but not fundamentally eliminated, despite advancements in design and data analysis. The disadvantages associated with microarrays have led, in part, to the increasing application of RNA-Seq.

Genome-wide Xcc microarray analysis was applied to investigate the impact of galactose versus glucose carbon sources on transcription levels [123]. The software tools ArrayLIMS and EMMA [129, 130] (Fig. 1, d) were used to identify genes related to galactose utilization. Recursion to the functional annotation of operon-like organized neighbour genes and validation by quantitative real-time PCR (qRT-PCR) revealed a metabolic module involved in scavenging galactose-containing carbohydrates. The results were used to further enhance the genome annotation by using the software tool GenDB [77], as mentioned above. Subsequently, further studies confirmed the value of this genome-wide microarray as an experimental tool, also for analyses with other Xcc strains [131, 132]. In addition, the microarray was applied as a useful tool for the initial characterization of novel Xanthomonas genomes [133], but this was not pursued further due to rapid progress in high-throughput sequencing. For X. campestris strains with weaknesses in gene prediction, microarrays provided experimental evidence to identify missed CDSs [134]. This approach was subsequently renewed with RNA-Seq experiments [135].

In general, transcriptome analysis is expected to deliver even more useful information by means of RNA-Seq. RNA-Seq data cover the entire genome irrespective of weaknesses in gene prediction or limitations in unique probe design, providing a high dynamic range in resolving quantitative differences [136–138]. So far, RNA-Seq experiments with xanthomonads (Table 2) have focused on obtaining superior gene expression data and on finding novel genes, including small non-coding RNA (ncRNA) genes. In a basic RNA-Seq analysis, the gene expression reproducibility was initially demonstrated between Xcc 8004 replicates [139]. When the bacteria were grown in two different media, namely minimal medium MMX and rich medium NYG [140], 629 differentially expressed CDS were identified. Of these, 495 MMX-induced genes were found to be mainly involved in amino acid metabolism, transport systems, atypical condition adaptation and pathogenicity, whereas 134 MMX-repressed genes were involved in chemotaxis and degradation of small molecules [139].

For Xcc, cell-to-cell signalling, signal transduction and transcriptional regulation have been revealed to have a substantial influence on various virulence factors including xanthan.

---

**Table 1. Bioinformatic tools and databases for the curation of functional annotations**

| Name                | Brief description                                      | Website                                               | Reference    |
|---------------------|--------------------------------------------------------|                                                      |             |
| BLAST               | Algorithm for biological sequence comparison           | http://blast.ncbi.nlm.nih.gov/Blast.cgi              | [289, 290]   |
| UniProtKB/ SwissProt| Annotated and non-redundant protein database            | www.uniprot.org/                                     | [204]        |
| KEGG                | Metabolic role, gene names frequent in use              | www.genome.jp/kegg                                   | [183]        |
| BRENDA              | Comprehensive enzyme database including kinetics and molecular properties | www.brenda-enzymes.org/                              | [291, 292]   |
| TCDB                | Transporter classification database                     | www.tcdb.org                                         | [293]        |
| PRIAM               | Enzyme specificity                                     | http://priam.prabi.fr/                               | [294]        |
| ENZYMEDb            | Enzyme/gene product nomenclature                        | http://www.ebi.ac.uk/enzymeportal/                    | [295]        |
| EzCatDB             | Enzyme catalytic-mechanism database                     | http://mbs.cbr.jp/EzCatDB/                           | [296]        |
| CAZy                | Database for carbohydrate-active enzymes                | http://www.cazy.org/                                 | [297]        |
Fig. 2. Reconstructed metabolic network as a basis to display omics data. The network reconstruction originates from the functional annotation of the genome of *X. campestris* pv. *campestris* B100. The functions of gene products, in particular of enzymes, reveal the substrates, products, and RNA-Seq of metabolic interconversions, which when combined form metabolic pathways that can be
Protein interaction is the most important and most studied subject relating to the field of pathogen–host interaction [34]. Proteomic analyses attempt to reveal and quantify all proteins in a culture at a given moment under specific conditions. Traditionally, gel-based techniques have been employed to separate the distinct proteins of a sample, in particular two-dimensional gel electrophoresis. In order to obtain an idea of the metabolic status of the cell [153, 154], the separated proteins can be visualized by staining techniques and, following specific fragmentation with enzymes like trypsin, identified by mass spectrometry, usually by matrix-assisted laser desorption/ionization time-of-flight (MALDI-ToF) detection using peptide mass fingerprints (PMFs) or tandem mass spectrometry (MS/MS). In this way, protein identification is facilitated by comparison to in silico proteome data derived from genome annotation [155, 156]. Hence, an important limitation of successful protein identification is the quality of the available genome data, resembling the situation for microarray-based transcriptomics. For Xcc, proteomics techniques have been initially applied to obtain an insight into the cytoplasmic proteome [157]. Subsequent studies analysed the secretome [158], characterized outer-membrane vesicles [159], and explored the cytoplasmic proteome more comprehensively [160]. Specific applications facilitated the characterization of the MAMPs involved in plant interactions [161]. A promising issue is the analysis of Xcc proteomes derived from interaction with host or non-host plants. Analysis of proteomes in xylem sap is a good example of such approaches [162]. Specific bioinformatics tools emerged to handle the complex data resulting from such proteomics approaches [163]. Other studies of xanthomonad proteomics focus on subjects such as Hrp proteins [164], secretome analysis of Xanthomonas oryzae [165] and analysis of biofilm formation of Xanthomonas campestris pv. citri [166] (Table 3). Multiple regulatory mechanisms influence protein functions: regulation of expression, which is addressed in transcriptomics, post-translational modifications such as phosphorylation or acetylation and allosteric regulation. Furthermore, the kinetic properties and concentrations of the enzymes affect the metabolic pools and metabolic fluxes in the cell [167]. Recently, protein phosphorylation has been identified as a mechanism of post-translational modification (PTM) for some key Xcc
<table>
<thead>
<tr>
<th>Xanthomonas species:pathovar</th>
<th>Subject of RNA-Seq analysis</th>
<th>Pathogen reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>X. campestris pv. campestris</td>
<td>Functional characterization and transcriptional analysis reveal multiple roles for prc in the pathogenicity of the black rot pathogen Xanthomonas campestris pv. campestris</td>
<td>[273]</td>
</tr>
<tr>
<td>X. oryzae pv. oryzae strains</td>
<td>Transcriptome-based identification of differently expressed genes from Xanthomonas oryzae pv. oryzae strains</td>
<td>[274]</td>
</tr>
<tr>
<td>X. perforans race T3</td>
<td>Comparative transcriptome analysis of resistant and susceptible tomato lines in response to infection by Xanthomonas perforans Race T3</td>
<td>[275]</td>
</tr>
<tr>
<td>X. oryzae pv. oryzicola</td>
<td>Comparative RNA-seq analysis of early-injected pathovars reveals universal importance of the TAL effector gene cluster in Xanthomonas lacticarida</td>
<td>[276]</td>
</tr>
<tr>
<td>X. arboricola pv. pruni</td>
<td>Comparative RNA-seq analysis of early-infected peach leaves by the invasive phytopathogen Xanthomonas arboricola pv. pruni</td>
<td>[277]</td>
</tr>
<tr>
<td>X. citri subsp. citri strain 306, X. citri subsp. citri strain AvrL2879</td>
<td>Comparative genomic and transcriptome analyses of pathotypes of Xanthomonas citri subsp. citri provide insights into mechanisms of bacterial virulence and host range</td>
<td>[278]</td>
</tr>
<tr>
<td>X. campestris pv. vesicatoria</td>
<td>RNA-seq analysis of cassava reveals similar plant responses upon infection with pathogenic and non-pathogenic strains of Xanthomonas axonopodis pv. manihotis</td>
<td>[279]</td>
</tr>
<tr>
<td>X. citri subsp. citri</td>
<td>Comparative transcriptome profiling reveals different expression patterns in Xanthomonas oryzae pv. oryzae strains with putative virulence-relevant genes</td>
<td>[280]</td>
</tr>
<tr>
<td>X. oryzae pv. oryzicola</td>
<td>Comparative transcriptome profiling reveals different expression patterns in Xanthomonas oryzae pv. oryzicola relative to a flagellin-deficient mutant</td>
<td>[281]</td>
</tr>
<tr>
<td>X. campestris pv. vesicatoria</td>
<td>Genome-wide transcriptome analysis of the plant pathogen Xanthomonas identifies sRNAs with putative virulence functions</td>
<td>[282]</td>
</tr>
</tbody>
</table>
genes (Fig. 2c) [168]. PTM is increasingly regarded as the regulatory layer that directly takes control of metabolic decisions [169], in particular in the central metabolism. Despite its merits, gel-based proteome analysis has substantial limitations; in particular, analysis of insoluble proteins in aqueous solutions is problematic, which results in a blind-spot rendering many membrane proteins inaccessible by this technique. Moreover, gel-free approaches allow the simultaneous identification and quantification of hundreds of proteins in a single experiment with less need for the time-consuming, tedious and (to some extent) error-prone experimental steps required in gel electrophoresis. The current technological advances have thereby enabled a straightforward investigation of complete proteomes. Mass spectrometry-based proteomics turns quantitative with the employment of stable isotopes or chemical tags as internal standards [170]; that said, xanthomonads have not yet been studied in this way [171]. Algorithmic and experimental foundations are readily available, however, and will facilitate such experiments in the near future (Fig. 1, e) [172].

**METABOLOMICS**

To understand the physiological status of an organism, cellular metabolites are analysed and quantified using a set of analysis methods collectively termed ‘metabolomics’. For the measurement of intracellular metabolites, this requires specific additional steps, namely: quenching of the cells to stop metabolism; sampling of cells and media; extraction; concentration; and analysis of metabolites. There are a couple of methods and challenges for each step [173, 174]. Numerous methods, such as nuclear magnetic resonance (NMR), GC–MS, LC–MS and capillary electrophoresis-MS (CE–MS), applied in conjunction with tandem MS, have been developed to detect metabolites [20, 175]. The enormous diversity of metabolite structures in comparison to the rather uniformly composed macromolecules analysed in the other omics disciplines justifies this wide range of approaches. Technological advances in recent years have increased the number of detectable metabolites and similar advances in methods for the quantification of metabolites may follow [176].

A specific challenge exists for the analysis of metabolites of X. campestris. The bacterium excretes the polysaccharide xanthan, which can viscosify the cultivation broth substantially. This viscosity necessitates an immediate quenching of the cells; however, fast separation of the cells from the surrounding medium is challenging. To address this issue, a method with simultaneous quenching and metabolites extraction using cold ethanol [177] and a rapid centrifugation method [178] were established. For NMR measurements ultracentrifugation was suggested as an appropriate pre-treatment step for efficient removal of the xanthan gum [179]. However, regarding obtaining realistic in vivo concentrations, there is still a need to demonstrate that the suggested procedure is advantageous for the quantification of unstable metabolites. For their sensitivity, versatility and ease-of-use, GC–MS (Fig. 2d) and LC–MS still represent the most commonly used methods for metabolite analysis. A variety of methods exists to further analyse metabolomics data. The web-based software platform MeltDB (Fig. 1, f) has been evaluated for the analysis of Xcc metabolomics data by means of GC–MS [180–182]. This software tool supports storage, sharing, analysis and integration of metabolomics experiments and mapping of the measured metabolites onto the metabolic pathway maps provided by an integrated KEGG [183] database. Recently, a complementary web-platform called ALlocator was published for the identification and quantitation of metabolites using LC–MS data [184] (Fig. 1, g and Table 4).

In summary, all main omics methods and appropriate bioinformatics tools (Fig. 1) are readily available for xanthomonads and examples have shown the applicability of xanthomonads for systems biology. The data analysis in the individual omics disciplines already results in new findings as outlined above. However, a more systematic interpretation of this data by means of data integration into metabolic networks will elucidate biological meaning and regulatory networks and deliver new perspectives that would otherwise remain unintuitive. The reconstruction of metabolic networks comprises several steps, as discussed in the next section.

**RECONSTRUCTION OF THE METABOLIC NETWORK**

A graphical representation of the analysed organisms’ metabolism is helpful in validating the interpretation of analytical data (Fig. 2). This is particularly important when substantial quantities of high-throughput omics data are concerned. Correspondingly, systems-level analyses of metabolism often depend on reconstructions of microbial metabolic networks. The availability of high-throughput genome sequencing enables the reconstruction of genome-wide metabolic networks (Fig. 3a) [185, 186], and the number of genome-scale models increases with the growing number of sequenced genomes. Likewise, the variety of established modelling applications based on reconstructed metabolic networks is expanding.

A metabolic network reconstruction is a representation of an organism’s constituent metabolic biochemical reactions. Such representations comprise the enzymes and metabolites involved in biochemical reactions and the corresponding enzyme-encoding genes (Fig. 2a, b). In general, network reconstruction and metabolic modelling together form an iterative process. Steps may be repeated until predictions meet phenotypic data or are falsified by measurements. The initial step is obtaining a network of interrelated biochemical pathways from functional genome annotations (Fig. 3a). There are two fundamental approaches to creating a metabolic network: (i) automatic reconstruction based on genome data and hence a top-down approach [187] and (ii) manual reconstruction of individual metabolic pathways.
<table>
<thead>
<tr>
<th>Xanthomonad</th>
<th>Host</th>
<th>Subject of analysis</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>X. campestris pv. campestris</td>
<td>X. campestris pv. campestris</td>
<td>In vivo proteomic analysis of Xanthomonas campestris pv. campestris in the interaction with the host plant Brassica oleracea</td>
<td>Pathogen [157]</td>
</tr>
<tr>
<td>X. campestris pv. campestris</td>
<td>X. campestris pv. campestris</td>
<td>Comparative proteome analysis of Xanthomonas campestris pv. campestris in the interaction with the susceptible and the resistant cultivars of Brassica oleracea</td>
<td>Pathogen [158]</td>
</tr>
<tr>
<td>X. campestris pv. campestris</td>
<td>X. campestris pv. campestris</td>
<td>Comparative proteome analysis of Xanthomonas campestris pv. campestris in the interaction with the host plant Brassica oleracea</td>
<td>Pathogen [159]</td>
</tr>
<tr>
<td>X. campestris pv. campestris</td>
<td>X. campestris pv. campestris</td>
<td>Comparative analysis of the extracellular proteins from Xanthomonas campestris pv. campestris B100</td>
<td>Pathogen [160]</td>
</tr>
<tr>
<td>X. campestris pv. campestris</td>
<td>X. campestris pv. campestris</td>
<td>Identification of proteins in susceptible and resistant Brassica oleracea responsive to Xanthomonas campestris pv. campestris infection</td>
<td>Host [161]</td>
</tr>
<tr>
<td>X. campestris pv. campestris</td>
<td>X. campestris pv. campestris</td>
<td>Comparative proteomics reveals differential induction of both biotic and abiotic stress response associated proteins in rice during Xanthomonas oryzae pv. oryzae infection</td>
<td>Pathogen [162]</td>
</tr>
<tr>
<td>X. campestris pv. campestris</td>
<td>X. campestris pv. campestris</td>
<td>Genetic and proteomic analyses of a novel porin-like protein in Xanthomonas oryzae pv. oryzae</td>
<td>Pathogen [163]</td>
</tr>
<tr>
<td>X. campestris pv. campestris</td>
<td>X. campestris pv. campestris</td>
<td>Proteomic analysis of rice plasma membrane reveals proteins involved in early defense response to bacterial infection</td>
<td>Host [164]</td>
</tr>
<tr>
<td>X. campestris pv. campestris</td>
<td>X. campestris pv. campestris</td>
<td>Secretome analysis of the rice bacterium Xanthomonas oryzae pv. oryzae</td>
<td>Pathogen [165]</td>
</tr>
<tr>
<td>X. campestris pv. campestris</td>
<td>X. campestris pv. campestris</td>
<td>Identification and functional characterization of small non-coding RNAs in Xanthomonas oryzae pv. oryzae</td>
<td>Pathogen [166]</td>
</tr>
<tr>
<td>X. campestris pv. campestris</td>
<td>X. campestris pv. campestris</td>
<td>Proteomic analysis of the regulatory function of DSF-dependent quorum sensing in Xanthomonas oryzae pv. oryzae</td>
<td>Pathogen [167]</td>
</tr>
<tr>
<td>X. campestris pv. campestris</td>
<td>X. campestris pv. campestris</td>
<td>A plant natriuretic peptide-like molecule of the pathogen Xanthomonas axonopodis pv. citri causes rapid changes in the proteome of its citrus host</td>
<td>Pathogen [168]</td>
</tr>
</tbody>
</table>
which are progressively combined as a bottom-up approach [188]. For the top-down approach, annotated genome data and information from databases that link the genomic information with functional data are used as the origin of an initial metabolic reconstruction of genome-scale networks [186, 189, 190]. Genome annotations that are useful as references can be found in organism-specific databases, such as EcoCyc [191] or AraCyc [192]. Software tools like Pathway Tools [193, 194] (Fig. 1, q) and the commercially available SimPheny (Genomatica) are applied for automated draft reconstructions of metabolic networks. The SuBLiMi-nal Toolbox [195], the SEED model [196], and the RAVEN Toolbox [197] are three additional software platforms which support steps of the reconstruction process. A review giving an overview of those three software platforms and their functionalities was published recently [189]. Different programs are available for the visual representation of metabolic networks, such as CellDesigner [198] (Fig. 1, j), CARMEN [199] (Fig. 1, i) and Cytoscape [200] (Fig. 1, o).

After automated reconstruction, the network is checked for consistency as well as for gaps, and redundancies must be resolved (Fig. 3b). The refinement of the metabolic network for the specific organism includes verification of metabolic functions, substrate and cofactor usage, reaction stoichiometry, reaction directionality, reaction localization, and information regarding the genes, gene–protein–reaction (GPR) associations, transport reactions, mass- and charge-balance reactions, and the biomass composition [190]. For this purpose, different resources and databases can be used, such as KEGG [183], BRENDA [201], the SEED database [202, 203] and SwissProt [204], as well as organism-specific literature, publications and experimental data, biochemical studies or protein localization analyses. Notes are inserted for each individual part of the metabolic reaction to record confidence scores for each reaction part. This is important for the evaluation of information in the model to quantify the reconstruction confidence for each reaction.

When available, reconstructed metabolic networks that truly represent the specific organism are a valuable resource for transcriptomics, proteomics and, in particular, metabolomics. Gaps and inconsistencies in the network prompt experimental clarification where appropriate methods are available (Fig. 3c). At the same time, our understanding of measurement data from post-genomic disciplines benefits from visualization in the context of metabolic pathways, which facilitates their biological interpretation. The reconstruction of a metabolic network of *X. campestris* involved several successive steps and a variety of data sources (Fig. 3). PathwayTools software (Fig. 1, q) was used to obtain an automated genome-based metabolic reconstruction for *Xanthomonas* (Fig. 3a). The resulting pathway-genome database (PGDB), termed XccCyc [107] (Fig. 1, p), was used in an iterative debugging process to resolve gaps and ambiguities in important pathways and, subsequently, to improve the functional annotation of corresponding genes (Fig. 3b). As mentioned above, post-genomics experiments were carried out to address weaknesses
of the metabolic reconstruction (Fig. 3c). An initial approach focused on amino acid biosynthesis. Following its automated generation, the XccCyc database indicated unclear amino acid biosynthetic pathways. While the biosynthetic pathway for arginine was available from a previous experimental study [125], biosynthetic pathways for 16 proteinogenic amino acids were determined by employing systematic and thorough interpretation of results from the bioinformatic tools (Table 1). For three remaining amino acids, $^{13}$C NMR measurements were required to resolve those metabolic routes actually used for their synthesis (Fig. 3c) [107]. Thus, in the end, biosynthetic pathways were determined for all proteinogenic amino acids, thereby closing gaps in the initial metabolic reconstruction (Figs 3d and 4c). The experimental results indicated an unusual variant of an isoleucine biosynthetic pathway in Xanthomonas that had not been observed for other organisms before [107], thereby exceeding the limits of insight possible through current bioinformatics-assisted analyses.

Despite advantages such as its comprehensive scale, the XccCyc-based metabolic network had drawbacks as it retained several gaps in metabolic pathways. This restricted its usefulness as a foundation for the mathematical models that underlie computational simulations of metabolic flux. Hence, data incorporated in this network were re-used to generate an independent large-scale metabolic network [43] using a bottom-up approach with CARMEN software [199]. Following manual revision, this large-scale metabolic network was suitable for obtaining metabolic models for the simulation of metabolic fluxes by flux balance analysis (FBA) or $^{13}$C metabolic flux analysis (see below) (Fig. 3e). The reconstructed network included 352 genes, 437 biochemical reactions, 10 transport reactions and 338 internal metabolites. Graphical displays of such large networks are often highly informative (Fig. S2 of [43]), but usually exceed the scope of regular scientific illustrations. However, excerpts or simplified representations of these comprehensive diagrams can be valuable visualizations of the properties related to the metabolic network Fig. 2(a, b).

The large-scale metabolic network provided a new view on the Xcc central carbohydrate metabolism. Xanthomonas was known to derive energy from sugars like glucose mainly by means of the ED pathway, with a minor role of the PP pathway [205], while the Embden–Meyerhof–Parnas (EMP) pathway, usually termed glycolysis, was concluded to be absent due to missing activity of the key enzyme 6-phosphofructokinase PFK [85, 206, 207]. In contrast to these findings, a well-conserved pfkA gene had been identified in the Xcc B100 genome and included in the reconstructed metabolic network. This stimulated the hypothesis that xanthomonads might have a complex carbohydrate catabolism, with the ED, PP and EMP pathways available to break down glucose.

After building a graphical representation of metabolism, the metabolic network can be converted to a mathematical model for detailed analysis of metabolic network structure and capacity, and to predict metabolic states and phenotypes using various modelling tools (Fig. 3e).

### Table 4. Studies of xanthomonads metabolism by means of metabolomics and metabolic network analysis

<table>
<thead>
<tr>
<th>Xanthomonad</th>
<th>Title</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>X. oryzae pv. oryzae</td>
<td>Metabolite analysis</td>
<td>Metabolomic and transcriptomic analysis of the rice response to the bacterial blight pathogen Xanthomonas oryzae pv. oryzae [326]</td>
</tr>
<tr>
<td>X. axonopodis pv. citri 306</td>
<td>Metabolite analysis</td>
<td>Xanthan gum removal for $^{1}H$-NMR analysis of the intracellular metabolome of the bacteria Xanthomonas axonopodis pv. citri 306 [179]</td>
</tr>
<tr>
<td>X. oryzae pv. oryzae</td>
<td>Metabolite analysis</td>
<td>Metabolomic profiling of bacterial leaf blight in rice (Application note) [327]</td>
</tr>
<tr>
<td>X. campestris pv. campestris</td>
<td>Metabolite analysis</td>
<td>Involvement of bacterial TonB-dependent signaling in the generation of an oligogalacturonide damage-associated molecular pattern from plant cell walls exposed to Xanthomonas campestris pv. campestris pectate lyases [124]</td>
</tr>
<tr>
<td>X. campestris pv. campestris</td>
<td>Metabolite analysis</td>
<td>An intracellular metabolite quantification technique applicable to polysaccharide-producing bacteria [177]</td>
</tr>
<tr>
<td>X. campestris pv. campestris</td>
<td>Metabolite analysis</td>
<td>The influence of metabolic network structures and energy requirements on xanthan gum yields [71]</td>
</tr>
<tr>
<td>X. campestris pv. campestris</td>
<td>Metabolite analysis</td>
<td>Development of a phenomenological modeling approach for prediction of growth and xanthan gum production using Xanthomonas campestris [42]</td>
</tr>
<tr>
<td>X. campestris pv. campestris strain B100</td>
<td>Model</td>
<td>Establishment, <em>in silico</em> analysis, and experimental verification of a large-scale metabolic network of the xanthan producing Xanthomonas campestris pv. campestris strain B100 [43]</td>
</tr>
<tr>
<td>X. campestris</td>
<td>Model</td>
<td>Characterization of xanthan gum biosynthesis in a centrifugal, packed-bed reactor using metabolic flux analysis [328]</td>
</tr>
<tr>
<td>X. campestris pv. campestris</td>
<td>Metabolite analysis</td>
<td>Genome-enabled determination of amino acid biosynthesis in Xanthomonas campestris pv. campestris and identification of biosynthetic pathways for alanine, glycine, and isoleucine by $^{13}$C-isotopologue profiling [107]</td>
</tr>
<tr>
<td>X. campestris pv. campestris</td>
<td>Metabolite analysis</td>
<td>Metabolic flux pattern of glucose utilization by Xanthomonas campestris pv. campestris prevalent role of the Entner-Doudoroff pathway and minor fluxes through the pentose phosphate pathway and glycolysis [182]</td>
</tr>
<tr>
<td>X. fragariae</td>
<td>Metabolite analysis</td>
<td>Metabolic response of strawberry (Fragaria ananassa) leaves exposed to the angular leaf spot bacterium (Xanthomonas fragariae) [329]</td>
</tr>
</tbody>
</table>
Fig. 3. Reconstruction of the *X. campestris* metabolic networks and its use related to omics data and flux analysis. (a) Large or genome-scale metabolic networks can be deduced from genome data as outlined in the text. Availability of high-quality genome annotation is a prerequisite for obtaining a metabolic reconstruction. Gaps and dead ends in an initial version of the reconstructed metabolic network could be resolved by re-assessing gene functions, which leads to an enhanced annotation of the respective genes (b), a process that involves multiple iterations. The reconstructed metabolic network facilitated high-throughput experiments in post-
genomics disciplines (c) and visualization of results related to metabolic phenomena. Such experimental results further enhanced the metabolic reconstruction (d). The reconstructed metabolic network is an essential requirement for obtaining metabolic models that facilitate both predictions and measurements of the metabolic flux (e). The resulting dynamic perspective of the Xanthomonas metabolism provided a quantitative approximation to the flow of metabolites under the respective study’s conditions. The predicted or measured flux distribution resulted in hypotheses that can lead to experiments (f) where molecular biology or post-genomics approaches are used for validation (g). Biochemical experiments such as enzyme tests of complex metabolic key elements (h) may further enhance models for flux analysis (i) or the underlying metabolic network (j).

**MODELLING**

Reconstructed metabolic networks can be converted to a mathematical model which represents the basis for network simulations (Fig. 3e). Usually, this requires extensive beforehand checking and manual refinement of the network. A kinetic model would comprise all the reactions, metabolites and cofactors included in the curated network. The kinetic model is usually represented by ordinary or partial differential equations [208, 209]. Because many enzymes have not been characterized biochemically (and many factors are necessary for biochemical characterization), the application of kinetic models is restricted to small biochemical networks in most organisms. A kinetic model of limited size describes the biosynthesis of sugar nucleotides, metabolic precursors of xanthan in Xcc [210]. The model makes use of the Mod- elica modelling environment and implements Petri Nets, while parameters are estimated by applying the MATLAB Optimization Toolbox (MathWorks) (Fig. 1, k). In contrast, constraint-based modelling, also termed structural or static modelling, can be applied to genome-scale systems where only the stoichiometry of the enzyme catalysed reactions, the reaction direction and the mass balance are known [211].

FBA is the basis for many applications in systems biology. A genome-scale model of *E. coli* was used to provide the organism with direct biocatalytic routes to 1,4-butanediol [212] and to identify better drug targets or biosynthetic pathways for antibiotics [213–216].

For Xcc, a large-scale metabolic network was reconstructed as described above and used to obtain a model for FBA (Fig. 3e, [43]). The model provided detailed predictions of metabolic features of Xcc (Fig. 4a, b). Initially it was utilized for *in silico* analyses to predict biomass generation and gene essentiality (Fig. 3f). These predictions were validated by analysing single gene deletion mutants causing deficiencies in the central carbohydrate metabolism [43]. The predicted effects were widely confirmed by the experimental results (Fig. 3g). Regarding the central metabolism, predictions of culture growth and xanthan production for the wild-type yielded realistic results when flux along the EMP pathway (glycolysis) was limited to 1% of total glucose catabolization. Mutant strains were constructed to assess the roles of the central metabolic pathways. The ED, PP and EMP pathways were blocked by specific inactivation of key enzymes. The growth of the mutant with blocked ED pathway was dramatically reduced, but the cells remained viable, while blocking the EMP pathway had only minimal effects [43].

These observations were in accordance with glucose catabolization mainly but not entirely via the ED pathway, with a minor flux through the EMP pathway, if at all. Hence, it remained unclear whether the EMP pathway and the PfkA enzyme had a metabolic role in Xcc. In addition, a xanthan-deficient mutant was studied *in vivo* and *in silico*. An FBA analysis of the xanthan-deficient mutant in comparison to Xcc wild-type shows differences in flux distribution during the growth phase in which xanthan production was observed in the wild-type. Here, a substantial redistribution of resources from xanthan production to culture growth and thus biomass generation was predicted and observed, rather than a decrease in carbon uptake [43]. This correlated with a phenomenon in terms of carbohydrate conversion that was observed earlier: over 70% of carbohydrate uptake [217] can be converted to the polysaccharide xanthan instead of biomass.

Despite its successful application in analysing the metabolism of Xcc, FBA faces limitations. The calculated optimal flux distribution is often not unique; hence there may be more than one possible optimal flux distribution. Reversible reactions, parallel paths or cyclic reactions complicate an accurate flux determination [218]. Therefore, with FBA alone, it is usually difficult to predict the absolute flux rates in the cell and it can be advisable to complement predictions based on stoichiometric networks with further measurements. An important approach for such measurements is 13C metabolic flux analysis (13C MFA) (Fig. 3e).

**13C MFA**

In 13C MFA, metabolites are measured that were labelled with the heavy carbon isotope 13C [218–222]. The method is useful for resolving metabolic flux when there are two or more pathways in parallel that facilitate metabolic conversions of the same substrate or generate the same product. An example relevant for xanthomonads was the possible parallel occurrence of EMP (glycolysis), ED and PP pathways. In each of these pathways, the intermediate metabolites are converted distinctly, resulting in a characteristic pattern of carbon atoms. Based on the observed labelling pattern it can be inferred through which pathway the substrate was catabolized. To facilitate such measurements, the bacteria are fed with a labelled substrate, usually glucose. Glucose is taken up by the cells and converted to biomass.

For biotechnology-related analyses, the generation of specific products is usually considered as well. On the basis of the metabolite labelling, the pathway within the metabolic
Fig. 4. Prediction and measurement of the metabolic flux of the *X. campestris* central metabolism. A large-scale metabolic model established with the software tool CellNetAnalyzer facilitated a detailed prediction of the metabolic fluxes for *X. campestris* B100 [43]. A subset of the results with reactions of the central carbohydrate metabolism has been visualized in this figure using ProMeTra software [288] for the wild-type (a) and for a mutant strain where the xanthan biosynthesis had been specifically
network mainly responsible for the degradation of the imported sugar can be identified. Herewith, the measurement of the $^{13}$C labelling pattern of the metabolites is crucial for calculating flux through the network [218, 220]. A robust method for flux determination is the measurement of the labelling pattern of amino acids that are available from hydrolysation of cell proteins. This analysis requires knowledge of the amino acid biosynthetic pathways for the respective organism (Fig. 4c). For this so-called stationary flux analysis, the cell pellet is hydrolysed and the labelling pattern is measured in the amino acids. This allows the determination of metabolic fluxes during steady-state growth [223, 224]. In addition, quantitative physiological data such as uptake rates, production rates and biomass composition can be determined and considered for flux calculation. There are open source software tools for $^{13}$C MFA such as OpenFlux [225] (Fig. 1, m), OpenMebius [226], influx_s [227] and $^{13}$CFLUX2 [228]. $^{13}$C MFA differs fundamentally from FBA in being a method to measure metabolic flux, albeit relying on heavy computation and on assumptions on possible metabolic routes, while FBA is a method to simulate metabolic fluxes and thereby delivers predictions, not measurements.

For Xcc, a $^{13}$C metabolic flux analysis based on GC-MS and NMR has been performed [182]. Metabolic fluxes in the central carbon metabolism were determined using [1-$^{13}$C]-glucose as the sole carbon source for growing Xcc. Based on the large-scale metabolic network reconstructed before (see above) and initially used to generate a model for simulations of metabolic fluxes by means of FBA [43], an additional metabolic model that included carbon atom transitions was established for $^{13}$C MFA with Xcc, consisting of 79 reactions including 42 balanced reactions. As the large-scale metabolic network reconstructed for the genome of Xcc B100 had indicated a putative EMP pathway for glucose utilization, which was suggested by FBA simulation results to have a minor role for the cultivation conditions in the laboratory, it was tempting to also analyse the Xcc glucose catabolism by $^{13}$C MFA. The $^{13}$C MFA confirmed the prevalent catabolic role of the ED pathway in Xcc and a minor flux via the PP pathway. In addition, $^{13}$C MFA indicated a minor flux via the EMP pathway (glycolysis) (Fig. 4d). This finding was verified experimentally by comparative NMR-based isotope profiling of a mutant deficient in glycolysis by means of pfkA mutation (Fig. 3h), which provided evidence for a moderate flux through glycolysis in the wild-type [182] and thus confirmed the $^{13}$C MFA results in line with previous findings from FBA. Hence, Xcc obviously had three parallel pathways available to generate energy and biosynthetic building blocks from glucose, the ED, PP and EMP pathways. Some bacteria are known to use the ED, PP and EMP pathways in parallel [182]. To shed more light on the EMP pathway used in Xcc, the 6-phosphofructokinase PFK was biochemically characterized (Fig. 3h) [229]. The identification of its pfkA gene had triggered the reconstruction of the EMP pathway (see above). The Xcc PfkA turned out to have no canonical PFK activity. Instead of ATP it used pyrophosphate as phosphate donor, and was not subject to allosteric regulation, while it unusually catalysed the generation of fructose 1,6-bisphosphate in a reversible reaction [229]. Experimental results from other studies clearly indicated the remaining enzymes that constitute the EMP pathway to be active in xanthomonads [229]. These non-canonical PFK properties are consistent with the flux determinations based on FBA and $^{13}$C MFA (Fig. 3i). At the same time they explain the misleading original EMP assessments for xanthomonads and the recent discovery that Xcc has a complete set of enzymes of the EMP pathway (glycolysis) available for metabolism of glucose in addition to the ED and PP pathways [229] – knowledge that facilitates an enhanced metabolic reconstruction (Fig. 3j). Moreover, comparative genomics revealed the presence of highly similar pfkA genes in other xanthomonads, indicating the availability of EMP activity throughout the Xanthomonadaceae [229].

ENVIRONMENTAL SYSTEMS BIOLOGY AS A PERSPECTIVE FOR ANALYSIS OF XANTHOMONADS

Xanthomonads encounter diverse influences in various environments and are subject to different biotic and abiotic factors. Biotic (stress) factors include other xanthomonad cells, other competing bacteria and bacteriophages, and plant cell defence mechanisms (Fig. 5). Regarding a deeper understanding of xanthomonads as pathogens, a next stage towards analysing host interactions may be the integrated modelling of both host and pathogen metabolic networks. Pioneering studies for this approach have been conducted with human pathogens. The cell-specific human alveolar...
macrophage model iAB-AMΦ-1410 was reconstructed and integrated with a Mycobacterium tuberculosis H37Rv model, iNJ661, to build a combined host–pathogen genome-scale reconstruction, iAB-AMΦ-1410-Mt-661 [230]. The integrated host–pathogen network was used to simulate metabolic changes during infection. Subsequently, the metabolic flux was analysed using $^{13}$C flux analysis [231] to determine metabolic interactions between M. tuberculosis and its macrophage host cell. The results indicate that M. tuberculosis acquires a mixture of amino acids, C1 and C2 substrates, and it consumes substrates which generate $^{13}$C labelled 3-carbon glycolytic intermediates from its host cell. Other approaches targeted the metabolic interactions of gut symbionts with their hosts [232–235]. Two recent reviews focus comprehensively on the systems biology of host (gut)–microbe interactions [34, 236].

So far, no host–pathogen models are available that feature plants as hosts. One might expect that such models ideally include representations of the specific host plant, the respective pathogen and relevant environmental parameters. In the longer term, metabolic models that reflect the whole environmental system would be favourable.

Metabolic models are available for several plants, among them hosts of pathogen interactions. A genome-scale metabolic reconstruction has been established for the model...
<table>
<thead>
<tr>
<th>Database</th>
<th>Address</th>
<th>Type</th>
<th>No. of PHI</th>
<th>Pathogens</th>
<th>Special features</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHI-base</td>
<td>phi-base.org</td>
<td>Pathogenicity, virulence and effector genes database</td>
<td>5017</td>
<td>55 prokaryotes: 27 plant-infecting pathogens, 28 animal-infecting pathogens. 105 eukaryotes: 76 plant-infecting pathogens and pests, four plant-infecting pests, three fungal-infecting pathogens, 21 animal- and other host-infecting pathogens, five animal-infecting parasites</td>
<td>Host species type(s) in PHI-base: diverse dicot, monocot, insect, bird, nematode, animal, fish</td>
<td>[330]</td>
</tr>
<tr>
<td>PHISTO</td>
<td>phisto.org</td>
<td>Protein–protein interaction (PPI) database</td>
<td>39,675</td>
<td>420 pathogen species: 58 bacteria, three fungi, four protozoa, 355 viruses</td>
<td>Specialized PPI, pathogen–human protein–protein interaction</td>
<td>[331]</td>
</tr>
<tr>
<td>PATRIC</td>
<td>patricbrc.org</td>
<td>Data integration database/website</td>
<td>1,300</td>
<td>Actinobacteria, 813 genomes. Proteobacteria, 211 genomes. Chlamydiae, 96 genomes. Firmicutes, 2,589 genomes. Spirochaetes, 386 genomes.</td>
<td>Web site that stores and integrates a variety of data types like genomics, transcriptomics, PPIs, three-dimensional protein structures</td>
<td>[333, 334]</td>
</tr>
<tr>
<td>Phytopath</td>
<td>phytopathdb.org</td>
<td>Genomic and phenotypic data from plant pathogen species</td>
<td>Linked to PHI-base</td>
<td>135 genomic sequences from 87 plant pathogen species, 22 host plants</td>
<td>Website specializing in plant pathogenic species, integration of phenotypic data for genes from the PHI-base, and an curated catalogue of experimentally verified genes</td>
<td>[335]</td>
</tr>
<tr>
<td>HPIDB</td>
<td><a href="http://agbase.mstate.edu/phi/phi/">http://agbase.mstate.edu/phi/phi/</a> main.html</td>
<td>Host–pathogen PPI database</td>
<td>45,238 PPI</td>
<td>45,238 unique protein interactions between 70 host and 594 pathogen species</td>
<td>Database can be searched with a variety of options such as sequence identifiers, symbol, taxonomy, publication, author, or interaction type, network visualization with Cytoscape</td>
<td>[336]</td>
</tr>
<tr>
<td>ENHanCED Infectious Diseases (EID2)</td>
<td><a href="http://www.zoonosis.ac.uk/EID2/">http://www.zoonosis.ac.uk/EID2/</a></td>
<td>22,515 PHI</td>
<td>22,515 unique species interactions between 6314 'carrier species' (often host) and 8905 'cargo species' (often pathogens)</td>
<td>Database focussing on species interactions and geographical distribution of a species, ability to search all the pathogens of a host, all the hosts of a pathogen.</td>
<td>[337]</td>
<td></td>
</tr>
</tbody>
</table>

*As of March 2015.
plant Arabidopsis thaliana [237], besides a compartmentalized model of the A. thaliana metabolism [238]. Being a member of the genus Brassicaceae, A. thaliana is a host of Xcc. Other genome-scale reconstructions are available for maize [239] and barley [240], a host of X. translucens pv. translucens [241, 242], and it is likely that more plant models will become available with the increasing number of genomes being sequenced. On the pathogen side, besides the Xcc models mentioned above, there are genome-scale metabolic reconstructions for other plant pathogenic bacteria [243–245].

Correspondingly, the first steps towards the study of groups of interaction partners relating to soil bacteria have been published. The competition between two Fe-reducing bacteria, namely Geobacter sulfurreducens and Rhodotherax ferrireducens, has been modelled in a soil community [246], and it was predicted that under natural conditions Rhodoferax will outcompete Geobacter at the site as long as sufficient ammonium is available. In contrast, if high concentrations of acetate are added during bioremediation, Geobacter species will predominate. This prediction was verified with field-scale observations. The results show that genome-scale modelling is a useful tool for predicting microbial interactions in environmental systems. Databases exist for metabolic networks concerning host–pathogen interactions (Table 5) with models available for well-studied organisms.

**METABOLISM AS AN INTEGRATIVE CONSTITUENT OF A PATHOGENICITY-DRIVEN LIFESTYLE**

Plant pathologists have identified substantial sets of virulence factors that xanthomonads have available to successfully exploit resources provided by their hosts. Most of these host resources are metabolites, valuable inorganic ions or organic plant material that can be converted to useful metabolites by appropriate enzyme-catalysed conversions. *Xanthomonas* genomes often encode dozens of exoenzymes [96] that are to be exported to the bacterial environment by a type II secretion system [247, 248] where they come into contact with the host plant. A substantial part of these exoenzymes are glycosidases and other polysaccharide-degrading enzymes that breach plant cell walls, thereby generating smaller carbohydrate compounds. Plants have evolved mechanisms to recognize such damage-associated microbial patterns (DAMPs) and subsequently initiate defence against the microbial hazard [124, 249]. The carbohydrate compounds arising as debris from plant cell wall penetration can be scavenged by xanthomonads and are assumed to be used as a valuable resource for energy and building blocks in the metabolism. This is reflected in the genome by the clustering of genes that encode degradative exoenzymes, transporters to facilitate import across the outer and inner membrane, and intracellular enzymes to convert the imported compounds into metabolites that can be easily utilized in the *Xanthomonas* metabolism when appropriate [79]. An unusually elaborate TonB system of outer membrane importers reflects the important role of these molecular machineries for the scavenging of obviously plant-derived materials and the relevance of resource uptake for xanthomonads [79, 82, 121–124, 250, 251]. The metabolic interconversion reactions discovered in these studies have hardly been included in metabolic models, nor has the metabolic flux through these pathways been simulated or measured so far.

Besides type II secretion, xanthomonads have further protein secretion systems available that are relevant for its plant pathogenicity [252]. There is a particular focus on the type III secretion system [40] that is encoded by *hrp* (hypersensitive response and pathogenicity) genes. Type III secretion systems occur as detrimental virulence factors in several human pathogens too [40, 253]. This system constitutes a huge molecular machinery that spans both cell membranes and the periplasmic space to form a syringe-shaped structure with a needle-like extracellular extension called the Hrp pilus that facilitates the transfer of bacterial effector proteins into the cytoplasm of plant host cells. Several of these effector proteins contribute to plant defence mechanisms otherwise triggered by the perception of conserved bacterial molecular structures called pathogen-associated molecular patterns (PAMPs) [254]. In several *Xanthomonas* strains there are proteins among these type III secreted effectors that are translocated to the nucleus of the host cell where they interfere as transcriptional regulators into gene expression [255, 256]. They were termed transcription activator-like (TAL) effectors (see above). Some TAL effectors turned out to activate the expression of sucrose exporters of the host plant cells. Hence, there is a link of this intricate virulence mechanism to the scavenging and metabolic utilization of this sugar by *Xanthomonas*. Again, this close interlocking of microbial and plant metabolism emphasizes the need for integrated host–pathogen models.

**DISCUSSION**

All relevant omics technologies are well developed for xanthomonads and in particular for Xcc to make the data available that are required to create hypotheses, enhance metabolic reconstructions and falsify simulations. At the same time, fundamental modelling approaches have already been applied successfully.

As a consequence, there have been first insights into the structure and dynamics of the *Xanthomonas* central metabolism. However, there are still many open questions. What if other carbon sources are available that Xcc prefers to take up when available [178]? How are the Xcc metabolic fluxes in planta, where initial data has just became available for the xylem [162], an important environment of Xcc? Will there be more flux along the EMP pathway under such conditions? How will metabolic fluxes be modified when Xcc lives in biofilms? Can findings from Xcc be transferred to other xanthomonads? How will metabolic activity be influenced by stress? Stress occurring in the natural environment includes abiotic factors...
like temperature changes or starvation, but also biotic stress provoked by plant defence [257, 258], competing microorganisms [259, 260] or phages [261–265]. Even effects arising from regulation or cell-to-cell communication [266] have not been considered in metabolic models so far. Furthermore, interesting secondary metabolism features have been identified in some xanthomonads, such as non-ribosomal peptide synthesis gene clusters [267]. To date it is completely unclear whether and how such secondary metabolic processes react to changes in the central metabolism. In summary, meaningful questions still need to be addressed, often those that are relevant for the interactions of xanthomonads with their environment including their hosts.

Further advancing and broadening this approach not only is promising for gaining deeper insights into the fundamental concepts of pathogenicity, in an inspiring way it also brings together data and expertise from diverging disciplines of life sciences. Areas of particular interest for future research might be the analysis of metabolism as an integrative constituent of a pathogenicity-driven lifestyle and environmental systems biology as pointed out in the preceding paragraphs.

The systems biology approaches and methods described for xanthomonads in this review can be adopted to characterize biological networks of other bacteria. Obviously, the omics technologies that provide the basis for modelling approaches are – sometimes with adaptations – applicable to almost all prokaryotes and even far beyond. Likewise, all referenced modelling techniques are now frequently used for bacteria applied in biotechnology. There is no reason to expect particular problems from adapting approaches based upon combinations of omics and modelling techniques to other plant pathogenic bacteria. Actually, the first such studies have already been done. For the β-proteobacterium *Ralstonia solanacearum* for example, which causes bacterial wilt on a wide range of important host plants, a resource allocation trade-off between virulence and proliferation has been recently studied [244].

Funding information
This work was funded by the German Federal Ministry of Education and Research (BMBF) (FKZ 031A553) within the de.NBI network.

Acknowledgements
The authors thank Rupert Norman for proofreading the manuscript.

Conflicts of interest
The authors declare that there are no conflicts of interest.

References
15. Ishii N, Nakahigashi K, Baba T, Robert M, Soga T et al. Multiple high-throughput analyses monitor the response of E. coli to perturbations. Science 2007;316:593–597.


42. Vicente JG, Holub EB. Xanthomonas campestris pv. campestris (cause of black rot of crucifers) in the genomic era is still a worldwide threat to brassica crops. Mol Plant Pathol 2013;14:2–18.


Wiggerich HG, Pühler A. The exbD2 gene as well as the iron uptake genes tonB, exbB and exbD1 of Xanthomonas campestris pv. campestris are essential for the induction of a hypersensitive response on pepper (Capsicum annuum). Microbiology 2000;146:1053–1060.


144. Chin KH, Lee YC, Tu ZL, Chen CH, Tseng YH et al. The CAMP receptor-like protein CLP is a novel c-di-GMP receptor linking...


150. Chen XL, Tang DJ, Jiang RP, He YQ, Jiang BL et al. sRNA-Xcc1, an integron-encoded transposon- and plasmid-transferred trans-acting sRNA, is under the positive control of the key virulence regulators HrpG and HrpX of Xanthomonas campestris pathovar campestris. RNA Biol 2011;8:947–953.


