Impact of Genome Analyses on Phytoplasma Research

Kube M, Siewert C, Holz S, Büttner C
Humboldt-Universität zu Berlin, Faculty of Agriculture and Horticulture, Department of Crop and Animal Sciences, Division Phytomedicine, Lentzeallee 55/57, D-14195 Berlin
Email: Michael.Kube@agrar.hu-berlin.de

1. **INTRODUCTION**

Phytoplasmas are associated with diseases of several hundred-plant species including many important crops. These phytopathogenic bacteria are grouped in the provisional genus '*Candidatus* Phytoplasma'. Today, the majority of our knowledge on phytoplasmas is directly connected or derived from genome research, as a cell-free cultivation of phytoplasmas was unavailable for a long time. Only four genomes were completely determined so far (reviewed in Kube *et al.*, 2012) as there are '*Ca*. P. asteris' strains OY-M and AY-WB, '*Ca*. P. australiense,' and '*Ca*. P. mali'. Chromosome condensation and decreased G + C content are characteristical for these genomes. Sequences provided information for design of diagnostic markers as well as subsequent experiments on key proteins interacting with the plant host or insect vector.

Several key questions have to be answered in the future. They aim to extend the knowledge of so far uncovered '*Ca*. Phytoplasma species' providing information on their metabolism, membrane proteins, effectors and virulence genes. A promising starting point is the comparative analysis of complete genome sequences or phytoplasma draft sequences obtained from metagenomic data. Hence, several studies are in progress worldwide. An overview is provided on the core metabolism of phytoplasmas and the strategies currently used in our own studies. Benefits and limits of draft sequences are discussed.

2. **MATERIAL AND METHODS**

The genetic core of the four complete phytoplasma genomes was compared (Kube *et al.*, 2012) with the results of phytoplasma draft sequences recently published (Saccarrdo *et al.*, 2012).

Limits of phytoplasma drafts obtained from short reads were estimated by *in silico* experiments. Complete phytoplasma chromosomes were used to calculate Illumina read data sets of various read length and number using the ARTtool (Huang *et al.*, 2012). Reads were *de novo* assembled using the CLC Genomic Workbench and compared to published results.

3. **RESULTS AND DISCUSSION**

Data analyses indicate an evolutionary adaptation resulting in obligate parasitism of phytoplasmas. Effectors and prominent membrane proteins manipulating vector and host are evolved in '*Ca*. Phytoplasma species'. This stands in contrast to the shared evolutionary adaption on the nutrient-rich environments corresponding to a common repertoire of metabolic features, which may undergo further condensation in some phylogenetic groups. Results of genomic drafts support this estimation. Different strategies for extraction of phytoplasma drafts from metagenomic data can be summarized as positive selection or initial negative selection approaches. Positive selection approaches uses reads or contigs assigned to known phytoplasma sequences for subsequent processing. Initial negative selection approach starts by the identification and removal of the contigs assigned to host/vector background. Remaining reads were assembled and resulting contigs undergo a subsequent positive selection of contigs after assignment to taxonomical groups.

An initial enrichment of phytoplasma DNA is not performed for most sequencing projects today. Deep sequencing compensates this step. Assemblies obtained from simulated single or paired-end reads with a length of 36-120 b frequently result in misassemblies of repeat regions and/or rejection of reads during the assembly because of conflicts. In consequence, phytoplasma drafts cannot cover the complete genome. Analysis of the shared gene content should be limited to identified features. Statements on the absence of genes in comparative analysis or genome size need additional verification.

3. **ACKNOWLEDGEMENTS**

The Deutsche Forschungsgemeinschaft supports this work (KU 2679/2-1 and BU 890/21-1).

4. **REFERENCES**

Huang W; Li L; Myers JR; Marth GT (2012). ART: a next-generation sequencing read simulator. *Bioinformatics* 28, 593-594.

Kube M; Mitrovic J; Duduk B; Rabus R; Seemüller E (2012). Current view on phytoplasma genomes and encoded metabolism. *ScientificWorldJournal* 2012, 185942.

Saccardo F; Martini M; Palmano S; Ermacora P; Scortichini M; Loi N; Firrao G (2012). Genome drafts of four phytoplasma strains of the ribosomal group 16SrIII. *Microbiology* 158, 2805-2814.