Preface

The Tomato Genome Sequence: How Did It Happen and Why Does It Matter?

The tomato genome sequencing project was initiated as part of the International Solanaceae Project (SOL) by a large international consortium of 10 countries (Korea, China, UK, India, The Netherlands, France, Japan, Spain, Italy and the United States). The tomato was chosen as reference species for the Solanaceae due to the high level of macro and micro-synteny within this plant family which comprises more than 3000 species among which some are important crops such as the fruit-bearing vegetables tomato, eggplant, and pepper, and the tuber-bearing potato, in addition to a number of medicinal and ornamental plants. The goal of the tomato genome sequencing project was to generate new information and resources allowing to shed light on how a common set of genes can give rise to a wide range of morphologically and ecologically distinct organisms, and how a better understanding of the genetic basis of plant diversity can be harnessed to meet the needs of the fast growing world population for a sustainable food crop production. It is important to mention that the launching of the tomato genome sequencing project would have not been possible without the use of the rich resources previously generated using this plant species. Undoubtedly, the project took advantage of the large collection of EST sequences, the high number of genetic markers, the dense and saturated genetic maps, and the well-characterized genomic libraries already available (http://sgn.cornell.edu/).

In many ways, the project represented a unique scientific and human adventure where the participants shared the scientific effort and the financial outlay and worked in close collaboration. Starting with conventional sequencing technologies the project shifted to the new high-throughput sequencing technologies, just emerging at the time. In this regard, the tomato genome sequencing project accompanied the transition from the old to the new sequencing era. Indeed, the Sanger sequencing method was initially used, but the advent of next-generation (NextGen) sequencing technologies has prompted the consortium to adopt these promising techniques. In retrospect, we can now say that the choice of these pioneering technologies was a wise decision, although it posed a risk at the time because there was no prior experience where the NextGen sequencing technologies have been applied de novo to sequence a large and complex eukaryotic genome. The consortium

had to overcome the difficulties of high-throughput data processing and assembly of "reads" without any possibility to rely on past experience in this area. An important challenge was the buildup of a pipeline for the genome sequence assembly, and in this respect, one of the most striking aspects of the project's success had been to produce finally a high-quality assembled tomato genome sequence using for the first time the new sequencing technologies.

Due to the estimated elevated cost of producing a high-quality sequence of the complete tomato genome, the initial strategy was the preferred sequencing of the euchromatin region where the majority of genes reside. This approach presents the advantage to target only 25 % of the total tomato genome thus allowing to significantly reduce the sequencing effort. The BAC-by-BAC sequencing strategy built on the existing saturated tomato genetic map, and made use of the genetic markers to select seed BACs within the gene-rich part of the tomato genome. The starting point for sequencing the genome was BACs anchored to the genetic map, and this minimal tiling path then extends from seed BACs to cover the whole genome. Once completed, the BAC-by-BAC tomato genome sequence was anticipated to provide a framework for shotgun sequencing of other Solanaceae species. While this approach enabled a rapid progress at the early phases of the project, it struck quickly with the difficulty of selecting BACs to power the sequencing pipeline. Finally, the slowness of this process became a serious obstacle pushing the consortium to seek other alternatives to reinvigorate the project. The advent of next-generation sequencing technologies offered an attractive option despite the lack of experience in applying these techniques to complex genomes. Switching to high-throughput sequencing launched the project into a new and original adventure where you have to discover simultaneously both the problems and their solutions. In particular, the consortium realized that these approaches require massive use of bioinformatics tools that had to be acquired and implemented in a short period of time.

The switch to a whole genome sequencing approach that combines both next-generation sequencing and Sanger sequencing boosted the project leading to a high-quality assembled tomato genome sequence within a relatively short period of time. The present book tells the tale of the tomato genome sequencing adventure with the various chapters describing in great detail every step of the sequencing project. Chapters 1 and 2 provides a brief review of the birth of the tomatoes in the Andean regions of South America, the history of their botanical classification along with other wild and cultivated Solanaceae as well as information about the main production areas. The following chapters deal with gene and QTL mapping in tomato with a particular emphasis on the new opportunities that the tomato genome sequences are providing for the genetic and molecular dissection of complex traits and how it helps breeders to shape new and better tomato varieties. The chapter on tomato resources for functional genomics describes the main resources, strategies, and tools currently available for linking genes to phenotypes in tomato. The chapters devoted to the generation of the tomato genome sequence per se emphasize the sequencing and assembling strategies

used in the project and the genome quality evaluation and the finishing methods. A separate chapter is dedicated to the annotation of the tomato genome with the aim to provide the best gene structures, a high-quality functional description for the protein-coding genes. The sequencing of the chloroplast and mitochondrial genomes, described in a specific chapter, adds to the understanding of the plant evolutionary history of tomato based on the phylogenetic position inferred from the organelles sequences information. The following two chapters review recent research on the timing and formation of ancient genome duplications and their evolutionary effects on the shaping of modern Solanaceae genomes. They also address the synteny among Solanaceae genomes providing insight into the modes and tempo of plant genome evolution and illustrating how a better knowledge of genome synteny and colinearity can facilitate the mobilization of resources from one species to other in this agronomically important family. The last chapter describes the tomato-centric databases and other generic resources freely accessible to Solanaceae community.

While the effort to produce an improved assembly with a larger coverage of the tomato genome is ongoing, the present version of the tomato genome (The Tomato Genome Consortium, Nature 2012) is among, if not the best quality of, all dicot genomes published to date, excluding Arabidopsis. Producing a reference tomato genome sequence represented a major breakthrough and has provided invaluable resource that has opened new avenues for research. Building on this resource enabled the development of a variety of genome-wide approaches like whole genome transcriptomic profiling that is nowadays becoming a routine method for expression studies. Likewise, genotyping-by-sequencing is currently spreading as a method of choice and mapping by sequencing is being increasingly used. The access to a complete genome sequence also fostered epigenetics studies allowing to establish a genome-wide mapping of various epigenetic marks. More recently, genome editing is experiencing a rapid growth to address the functional significance of candidate genes in the tomato model. These are some of the main areas that have been impacted by the acquisition of a high-quality reference genome for tomatoes, but most likely, we are only at the dawn of these dramatic developments and more unexpected ones will break out in the future.

Castanet Tolosan, France

Mondher Bouzayen
Mohamed Zouine

http://www.springer.com/978-3-662-53387-1

The Tomato Genome

Causse, M.; Giovannoni, J.; Bouzayen, M.; Zouine, M. (Eds.)

2016, XII, 259 p. 48 illus., 40 illus. in color., Hardcover
ISBN: 978-3-662-53387-1