Bacterial transcription factor binding site inference as the basis for the study of plant-microbial interactions

Nikolaichik Y.A.

Belarusian State University, Minsk, Belarus

Email: nikolaichik@bsu.by

Key message. *Genome-wide in silico analysis of bacterial regulatory sequences reveals new regulatory interactions, is generally accessible and should become the standard in the study of plant-microbial interactions.*

Keywords: *transcriptional regulation, operator motif, genome annotation, plant-microbial interactions*

'Omic' technologies that are becoming ubiquitous generate large amounts of data that are still only partially used. Genomic projects usually end with the automatic assembly of the "draft" genome and the automatic annotation of reading frames, rRNA and tRNA genes. However, genomic sequences record all regulatory information, which can be easily extracted using generally available methods. The presentation will demonstrate the approach to deciphering regulatory information that is based on a genome-wide search for regulatory elements: transcription factor binding sites (operators), promoters and terminators. Our method is based on the analysis of 3D structures of complexes of transcription factors with their own operators and allows to semi-automatically find a significant part of the operator sequences in previously unexplored genomes. Of particular value is total genome analysis, which finds the operator and promoter sequences for most transcription units, shows multiplex regulation and allows one to predict complex transcriptional cascades. The quality of such analysis can be significantly improved with the availability of transcriptomic data for the studied bacterium. Transcriptomic data analysis facilitates the identification of promoters, terminators, and whole transcriptional units, and also helps to determine the most likely regulatory mode (activation or repression) for each TF.

The combination of a genome-wide search for operators (and other regulatory sequences) with an analysis of open RNA-seq data will be demonstrated by identifying new regulators of the virulent properties of plant pathogenic bacteria from the *Pectobacteriaceae* family.

The approach to the analysis of regulatory information described here is universal (suitable for any bacterium) and relatively easily accessible, since complex algorithms for identifying regulatory sequences are implemented as a user-friendly program available at github.com/nikolaichik/SigmoID.

Идентификация сайтов связывания транскрипционных факторов бактерий как основа для исследования растительно-микробных взаимодействий

Николайчик Е.А.

Белорусский государственный университет, Минск, Республика Беларусь

Аннотация. *Полногеномный анализ бактериальных регуляторных последовательностей in silico обнаруживает новые регуляторные взаимодействия, общедоступен и должен стать стандартом при исследовании растительно-микробных взаимодействий.*

Ключевые слова: *транскрипционная регуляция, операторный мотив, аннотация генома, растительно-микробные взаимодействия*

Становящиеся общедоступными "омиксные" технологии генерируют большие массивы данных, до сих пор используемых лишь частично. Геномные проекты обычно заканчиваются автоматической сборкой "черновика" генома и автоматической аннотацией рамок считывания, генов рРНК и тРНК. Однако в геномных последовательностях записана вся регуляторная информация, которую вполне можно извлечь общедоступными методами. В докладе будет продемонстрирован подход к расшифровке регуляторной информации, опирающийся на полногеномный поиск регуляторных элементов: сайтов связывания транскрипционных факторов (операторов), промоторов и терминаторов. Наш метод основан на анализе 3D-структур комплексов транскрипционных факторов со своими операторами, который способен в полуавтоматическом режиме находить значительную часть операторных последовательностей в неисследованных ранее геномах. Особую ценность имеет максимально полный анализ генома, который находит операторные и промоторные последовательности для большинства транскрипционных единиц, показывает множественную регуляцию и позволяет предсказывать сложные транскрипционные каскады. Качество анализа может быть существенно повышено при наличии транскриптомных данных для исследуемой бактерии, которые облегчают идентификацию промоторов, терминаторов и целых транскрипционных единиц, а также позволяют определять наиболее вероятный режим регуляции (активацию или репрессию).

Сочетание полногеномного поиска операторов (и прочих регуляторных последовательностей) с анализом открытых транскриптомных (RNA-seq) данных будет продемонстрировано на примерах идентификации новых регуляторов вирулентных свойств фитопатогенных бактерий из семейства *Pectobacteriaceae*.

Описанный здесь подход к анализу регуляторной информации является универсальным (подходит для любой бактерии) и относительно легкодоступным, поскольку сложные алгоритмы идентификации регуляторных последовательностей реализованы в виде программы с удобным графическим интерфейсом, доступной по адресу github.com/nikolaichik/SigmoID.