Rna Seq De Novo Assembly Training Day 2

In this authoritative guide, expert investigators provide cutting-edge chapters dealing with modern plant systems biology approaches. This work provides the kind of detailed description and implementation advice that is crucial for getting optimal results.

The intent of this study is to further the understanding of polyphenol oxidase's (PPO) functional characteristics in walnut and the plant kingdom through the use of RNA-seq de novo assembly and differential expression analysis. We begin the study with three wild-type (natural unchanged PPO activity) and three PPO-silenced (> 95% reduction in PPO activity) samples from the walnut species Juglans regia, containing a total of approximately 106 million and 122 million paired-end reads, respectively. Since no reference genome exists, we utilized Trinity, an RNA-seq de novo transcriptome assembler, in order to reconstruct full-length transcripts and alternatively spliced isoforms from our RNA-seq data. A total of 110,068 transcripts were assembled with a N50 contig length of 1,776. We then aligned our RNA-seq data to the assembled transcriptome using Bowtie to assess the quality of the assembly. Approximately 92% of the aligned RNA-seq reads were proper pairs (i.e., each end of a paired-end read successfully mapped to the same contig). Next, we used RNA-seq by Expectation-Maximization (RSEM), which is a software tool that estimates gene and isoform expression levels from our paired-end RNA-seq data. The expression estimates generated by RSEM from both our assembled transcriptome and a reference transcriptome were used with two differential expression analysis tools, edgeR and DESeq, in order to determine which genes are being differentially expressed in PPO-silenced plants as compared to wild-type. Using edgeR and a false discovery rate of 0.001, we found 91 genes from our assembled transcriptome and 130 genes from the reference transcriptome to be differentially expressed. Using DESeq, we discovered 69 genes from our assembled transcriptome and 46 genes from the reference transcriptome to be differentially expressed based on an adjusted p-value of 0.05 or lower. We then employed BLAST (BLASTn), which compares a nucleotide query sequence against a nucleotide sequence database, to conduct various comparisons between our Trinity assembly, the reference assembly, and our differentially expressed sets discovered using edgeR and DESeq. One of the results determined that 8 of the 91 differentially expressed genes found using edgeR, and 7 of the 69 differentially expressed genes found using DESeq were absent from the reference transcriptome.

Marine mammals face a large array of stressors, including loss of habitat, chemical and noise pollution, and bycatch in fishing, which alone kills hundreds of thousands of marine mammals per year globally. To discern the factors contributing to population trends, scientists must consider the full complement of threats faced by marine mammals. Once populations or ecosystems are found to be at risk of adverse impacts, it is critical to decide which combination of stressors to reduce to bring the population or ecosystem into a more favorable state. Assessing all stressors facing a marine mammal population also provides the environmental context for evaluating whether an additional activity could threaten it. Approaches to Understanding the Cumulative Effects of Stressors on Marine Mammals builds upon previous reports to assess current methodologies used for evaluating cumulative effects and identify new approaches that could improve these assessments. This review focuses on ways to quantify exposure-related changes in the behavior, health, or body condition of individual marine mammals and makes recommendations for future research initiatives.

Transcriptomics studies often rely on partial reference transcriptomes that fail to capture the full catalog of transcripts and their variations. Recent advances in sequencing technologies and assembly algorithms have facilitated the reconstruction of the entire transcriptome by deep RNA sequencing (RNA-seq), even without a reference genome. However, transcriptome assembly from billions of RNA-seq reads, which are often very short, poses a significant informatics challenge. This Review summarizes the recent developments in transcriptome assembly approaches - reference-based, de novo and combined strategies-along with some perspectives on transcriptome assembly in the near future.

The Barley Genome

Corset: Enabling Differential Gene Expression Analysis for de Novo Assembled Transcriptomes

Next-generation Transcriptome Assembly

The Soybean Genome

Plant Systems Biology

Virus Bioinformatics

Accurate and comprehensive transcriptome assemblies lay the foundation for a range of analyses, such as differential gene expression analysis, metabolic pathway reconstruction, novel gene discovery, or metabolic flux analysis. With the arrival of next-generation sequencing technologies it has become possible to acquire the whole transcriptome data rapidly even from non-model organisms. However, the problem of accurately assembling the transcriptome for any given sample remains extremely challenging, especially in species with a high prevalence of recent gene or genome duplications, those with alternative splicing of transcripts, or those whose genomes are not well studied. This thesis provides a detailed overview of the strategies used for transcriptome assembly, including a review of the different statistics available for measuring the quality of transcriptome assemblies with the emphasis on the types of errors each statistic does and does not detect and simulation protocols to computationally generate RNAseq data that present biologically realistic problems such as gene expression bias and alternative splicing. Using such simulated RNAseq data, a comparison of the accuracy, strengths, and weaknesses of seven representative assemblers including de novo, genome-guided methods shows that all of the assemblers individually struggle to accurately reconstruct the expressed transcriptome, especially for alternative splice forms. Using a consensus of several de novo assemblers can overcome many of the weaknesses of individual assemblers, generating an ensemble assembly with higher accuracy than any individual assembler.

Transcriptome assays are increasingly being performed by high-throughput RNA sequencing (RNA-seq). For organisms whose genomes have not been sequenced and annotated, transcriptomes must be assembled de novo from the RNA-seq data. Here, we present novel algorithms, specific to bacterial gene structures and transcriptomes, for analysis of bacterial RNA-seq data and de novo transcriptome assembly. The algorithms are implemented in an open source software system called Rockhopper 2. We find that Rockhopper 2 outperforms other de novo transcriptome assemblers and offers accurate and efficient analysis of bacterial RNA-seq data. Rockhopper 2 is available at http://cs.wellesley.edu/~btjaden/Rockhopper.

The 14 contributed chapters in this book survey the most recent developments in high-performance algorithms for NGS data, offering fundamental insights and technical information specifically on indexing, compression and storage; error correction; alignment; and assembly. The book will be of value to researchers, practitioners and students engaged with bioinformatics, computer science, mathematics, statistics and life sciences.

Recently, the coverage of non-protein-coding RNA in the scientific literature has expanded dramatically. While the functions for many are unknown, strong interest in this aspect of cellular biology is driving development of methods for detecting non-coding genes and transcripts. During the same period, RNA sequencings high throughput and high spatial resolution have established it as the preferred method for characterising transcriptomes.

Many groups are now sequencing transcriptomes. De novo transcriptome assembly methods are being developed to address issues for which no reference genome is available. We propose a methodology that is compatible with de novo transcriptome assembly, that uses sequence, structural and genomic features to classify transcripts as non-coding vs. protein-coding RNA, and to classify different non-coding RNA types. We have applied our technique on a variety of known RNA sequences and have explored its use on contigs from the Trans-ABySS assembly pipeline for RNA-Seq data from normal mouse tissues.

An Active Learning Approach

Next Generation Sequencing

Gene Prediction

RNA-SEQ DE NOVO ASSEMBLY AND DIFFERENTIAL EXPRESSION ANALYSIS OF WILD-TYPE AND POLYPHENOL OXIDASE-SILENCED WALNUT (JUGLANS REGIA)

Computational Analysis of RNA-Seq Data in the Absence of a Known Genome

The Amaranth Genome

In his lectures my teacher Karl Mägdefrau used to say that one only becomes a real plant scientist when one enters a tropical rainforest. For me this initiation occurred in 1969 in northern Queensland, Australia, and was associated with the greatest excitement. On another level it received confirmation when I set out in 1983 together with some friends and colleagues for the first detailed ecophysiological studies of epiphytes in the wet tropics in situ in the island of Trinidad and later for similar work in Venezuela. This then promoted the idea of organizing a special symposium on "The evolution and ecophysiology of vascular plants as epiphytes" during the XIV International Botanical Congress in luly 1987 in Berlin, and to ask some of the speakers to produce chapters for a small monograph on the interesting ecologically defined group of plants "epiphytes" as presented in this volume of "Ecological Studies". The enthusiasm of the participants of the symposium giving reports and adding to the discussion was most stimulating, and it appears that epiphytes might gain well-deserved, wider consideration in the future. The cooperation with the authors of this book was very pleasant and I appreciated the new contacts established with adepts of the "epiphyte community". The chapters were organized and arranged covering first more gen eral aspects with setting the scene in Chapter 1, the evolution of epi phytism in Chapter 2 and the role of CO -concentrating mechanisms in 2 Chapter 3.

De novo RNA-Seq assembly facilitates the study of transcriptomes for species without sequenced genomes, but it is challenging to select the most accurate assembly in this context. To address this challenge, we developed a model-based score, RSEM-EVAL, for evaluating assemblies when the ground truth is unknown. We show that RSEM-EVAL correctly reflects assembly accuracy, as measured by REF-EVAL, a refined set of ground-truth-based scores that we also developed. Guided by RSEM-EVAL, we assembled the transcriptome of the regenerating axolotl limb; this assembly compares favorably to a previous assembly. A software package implementing our methods, DETONATE, is freely available at http://deweylab.biostat.wisc.edu/detonate.

This volume focuses on various approaches to studying long non-coding RNAs (lncRNAs), including techniques for finding lncRNAs, localization, and observing their functions. The chapters in this book cover how to catalog lncRNAs in various plant species; determining subcellular localization; protein interactions; structures; and RNA modifications. Written in the highly successful Methods in Molecular Biology series format, chapters include introductions to their respective topics, lists of the necessary materials and reagents, step-by-step, readily reproducible laboratory protocols, and tips on troubleshooting and avoiding known pitfalls. Cutting-edge and innovative, Plant Long Non-Coding RNAs: Methods and Protocols is a valuable resource that aids researchers in understanding the functions of lncRNAs in different plant species, and helps them explore currently uncharted facets of plant biology.

The large potential of RNA sequencing and other "omics" techniques has contributed to the production of a huge amount of data pursuing to answer many different questions that surround the science's great unknowns. This book presents an overview about powerful and cost-efficient methods for a comprehensive analysis of RNA-Seq data, introducing and revising advanced concepts in data analysis using the most current algorithms. A holistic view about the entire context where transcriptome is inserted is also discussed here encompassing biological areas with remarkable technological advances in the study of systems biology, from microorganisms to precision medicine.

Rnnotator Assembly Pipeline

Oat

Classification of Coding and Non-coding RNA in RNA-Seq Data

Rnnotator

The Tomato Genome

Defining the Maize Transcriptome de Novo Using Deep RNA-Seq

De Novo Assembly of Bacterial Transcriptomes from RNA-Seq Data

Background: Comprehensive annotation and quantification of transcriptomes are outstanding problems in functional genomics. While high throughput mRNA sequencing (RNA-Seq) has emerged as a powerful tool for addressing these problems, its success is dependent upon the availability and quality of reference genome sequences, thus limiting the organisms to which it can be applied. Results: Here, we describe Rnnotator, an automated software pipeline that generates transcript models by de novo assembly of RNA-Seq data without the need for a reference genome. We have applied the Rnnotator assembly pipeline to two yeast transcriptomes and compared the results to the reference gene catalogs of these organisms. The contigs produced by Rnnotator are highly accurate (95percent) and reconstruct full-length genes for the majority of the existing gene models (54.3percent). Furthermore, our analyses revealed many novel transcribed regions that are absent from well annotated genomes, suggesting Rnnotator serves as a complementary approach to analysis based on a reference genome for comprehensive transcriptomics. Conclusions: These results demonstrate that the Rnnotator pipeline is able to reconstruct full-length transcripts in the absence of a complete reference genome.

Bioinformatics Algorithms: an Active Learning Approach is one of the first textbooks to emerge from the recent Massive Online Open Course (MOOC) revolution. A light-hearted and analogy-filled companion to the authors' acclaimed online course (http://coursera.org/course/bioinformatics), this book presents students with a dynamic approach to learning bioinformatics. It strikes a unique balance between practical challenges in modern biology and fundamental algorithmic ideas, thus capturing the interest of students of biology and computer science students alike.Each chapter begins with a central biological question, such as "Are There Fragile Regions in the Human Genome?" or "Which DNA Patterns Play the Role of Molecular Clocks?" and then steadily develops the algorithmic sophistication required to answer this question. Hundreds of exercises are incorporated directly into the text as soon as they are needed; readers can test their knowledge through automated coding challenges on Rosalind (http://rosalind.info), an online platform for learning bioinformatics.The textbook website (http://bioinformaticsalgorithms.org) directs readers toward additional educational materials, including video lectures and PowerPoint slides.

We present a new de novo transcriptome assembler, Bridger, which takes advantage of techniques employed in Cufflinks to overcome limitations of the existing de novo assemblers. When tested on dog, human, and mouse RNA-seq data, Bridger assembled more full-length reference transcripts while reporting considerably fewer candidate transcripts, hence greatly reducing false positive transcripts in comparison with the state-of-the-art assemblers. It runs substantially faster and requires much less memory space than most assemblers. More interestingly, Bridger reaches a comparable level of sensitivity and accuracy with Cufflinks. Bridger is available at https://sourceforge.net/projects/rnaseqassembly/files/?source=navbar.

Evolution and Ecophysiology

Techniques, Approaches, and Applications

RNA Bioinformatics

The RNA-Seq Analysis Pipeline on Galaxy

Advances, Applications and Challenges

The genus Thymus consists of about 350 species of perennial, aromatic herbs and subshrubs native to Europe and North Africa. Various types of thyme are used all over the globe as condiments, ornamentals and sources of essential oil. Thyme oil (distilled from its leaves) is among the world's top ten essential oils, displaying antibacterial, antimyco

This book collates the most up to date information on Fragaria, and Rubus genomes. It focuses on the latest advances in the model system Fragaria vesca, along with the allied advances in economically important crops. Covering both basic and applied aspects of crop genomics, it illustrates strategies and resources for the study and utilization of genome sequences and aligned functional genomics resources. Rosaceous berries are collectively an increasingly important set of high-value global crops, with a trade value of over £2 billion dollars per annum. The rosaceous berries strawberry, raspberry and blackberry share some common features at the genome scale, namely a range of ploidy levels in each genus and high levels of heterozygosity (and associated inbreeding depression) due to self-incompatibility systems, dioecy, or multispecies hybridization events. Taken together, although the genomes are relatively compact, these biological features lead to significant challenges in the assembly and analysis of berry genomes, which until very recently have hampered the progress of genome-level studies. The genome of the woodland strawberry, Fragaria vesca, a self-compatible species with a homozygous genome was first sequenced in 2011 and has served as a foundation for most genomics work in Fragaria and to some extent Rubus. Since that time, building upon this resource, there have been significant advances in the development of genome sequences for related crop species. This, coupled with the revolution in affordable sequencing technology, has led to a suite of genomics studies on Fragaria and more recently Rubus, which undoubtedly aid crop breeding and production in future years.

This book describes the development of genetic resources in amaranths, with a major focus on genomics, reverse, and forward genetics tools and strategies that have been developed for crop improvement. Amaranth is an ancient crop native to the New World. Interest in amaranths is being renewed, due to their adaptability, stress tolerance, and nutritional value. There are about 65 species in the genus, including Amaranthus caudatus L., A. cruentus L., and A. hypochondriacus L., which are primarily grown as protein-rich grains or pseudocereals. The genus also includes major noxious weeds (e.g., A. palmeri). The amaranths are within the Caryophyllales order and thus many species (e.g., A. tricolor) produce red (betacyanin) or yellow (betaxanthin) betalain pigments, which are chemically distinct from the anthocyanins responsible for red pigmentation in other plants. A. hypochondriacus, which shows disomic inheritance (2n = 32; n= 466 Mb), has been sequenced and annotated with 23,059 protein-coding genes. Additional members of the genus are now also been sequenced including weedy amaranths, other grain amaranths, and their putative progenitors.

This book describes the basic botanical features of kiwifruit and its wild relatives, reports on the steps that led to its genome sequencing, and discusses the results obtained with the assembly and annotation. The core chapters provide essential insights into the main gene families that characterize this species as a crop, including the genes controlling sugar and starch metabolism, pigment biosynthesis and degradation, the ascorbic-acid pathway, fruit softening and postharvest metabolism, allergens, and resistance to pests and diseases. The book offers a valuable reference guide for taxonomists, geneticists and horticulturists. Further, since information gained from the genome sequence is extraordinarily useful in assessing the breeding value of individuals based on whole-genome scans, it will especially benefit plant breeders. Accordingly, chapters are included that focus on gene introgression from wild relatives and genome-based breeding.

Thyme

A Practical Approach

Transcriptome De Novo Assembly, Clustering, and Annotation of Novel Transcripts

Consensus Ensemble Approaches Improve De Novo Transcriptome Assemblies

Bioinformatics Research and Applications

Applications of RNA-Seq and Omics Strategies

Virus bioinformatics is evolving and succeeding as an area of research in its own right, representing the interface of virology and computer science. Bioinformatic approaches to investigate viral infections and outbreaks have become central to virology research, and have been successfully used to detect, control, and treat infections of humans and animals. As part of the Third Annual Meeting of the European Virus Bioinformatics Center (EVBC), we have published this Special Issue on Virus Bioinformatics.

De novo assembly of the transcriptome is crucial for functional genomics studies in bioenergy research, since many of the organisms lack high quality reference genomes. In a previous study we successfully de novo assembled simple eukaryote transcriptomes exclusively from short Illumina RNA-Seq reads [1]. However, extensive alternative splicing, present in most of the higher eukaryotes, poses a significant challenge for current short read assembly processes. Furthermore, the size of next-generation datasets, often large for plant genomes, presents an informatics challenge. To tackle these challenges we present a combined experimental and informatics strategy for de novo assembly in higher eukaryotes. Using maize as a test case, preliminary results suggest our approach can resolve transcript variants and improve gene annotations.

The State of the Art in Transcriptome AnalysisRNA sequencing (RNA-seq) data offers unprecedented information about the transcriptome, but harnessing this information with bioinformatics tools is typically a bottleneck. RNA-seq Data Analysis: A Practical Approach enables researchers to examine differential expression at gene, exon, and transcript le

Transcriptome analysis is the study of the transcriptome, of the complete set of RNA transcripts that are produced under specific circumstances, using high-throughput methods. Transcription profiling, which follows total changes in the behavior of a cell, is used throughout diverse areas of biomedical research, including diagnosis of disease, biomarker discovery, risk assessment of new drugs or environmental chemicals, etc. Transcriptome analysis is most commonly used to compare specific pairs of samples, for example, tumor tissue versus its healthy counterpart. In this volume, Dr. Pyo Hong discusses the role of long RNA sequences in transcriptome analysis, Dr. Shinichi describes the next-generation single-cell sequencing technology developed by his team, Dr. Prasanta presents transcriptome analysis applied to rice under various environmental factors, Dr. Xiangyuan addresses the reproductive systems of flowering plants and Dr. Sadovsky compares codon usage in conifers.

Transcriptome Analysis

From Microorganisms to Human Health

An Automated de Novo Transcriptome Assembly Pipeline from Stranded RNA-Seq Reads

De Novo - Assembly and Annotation of RNA-Seq Data from the Visual Cortex of Cystophora Cristata

The Kiwifruit Genome

Plant Long Non-Coding RNAs

This book describes the strategy used for sequencing, assembling and annotating the tomato genome and presents the main characteristics of this sequence with a special focus on repeated sequences and the ancestral polyploidy events. It also includes the chloroplast and mitochondrial genomes. Tomato (Solanum lycopersicum) is a major crop plant as well as a model for fruit development, and the availability of the genome sequence has completely changed the paradigm of the species' genetics and genomics. The book describes the numerous genetic and genomic resources available, the identified genes and quantitative trait locus (QTL) identified, as well as the strong synteny across Solanaceae species. Lastly, it discusses the consequences of the availability of a high-quality genome sequence of the cultivated species for the research community. It is a valuable resource for students and researchers interested in the genetics and genomics of tomato and Solanaceae.

The volume provides detailed protocols that have been developed or modified exclusively for the study of oat. The topics discussed in this book are a selection of various molecular biology and biotechnology methods, such as the application of molecular markers for polymorphism analyses and cytological manipulations, the production of synthetic polyploids, and in vitro cultures and genetic modifications. Written in the highly successful Methods in Molecular Biology series format, chapters include introductions to their respective topics, lists of the necessary materials and reagents, step-by-step, readily reproducible laboratory protocols, and tips on troubleshooting and avoiding known pitfalls. Cutting-edge and comprehensive, Oat: Methods and Protocols is a useful resource in the development of new research approaches toward organizing the oat genome and the identification of new and useful traits for further improvements of this exceptional crop.

This volume introduces software used for gene prediction with focus on eukaryotic genomes. The chapters in this book describe software and web server usage as applied in common use-cases, and explain ways to simplify re-annotation of long available genome assemblies. Written in the highly successful Methods in Molecular Biology series format, chapters include introductions to their respective topics, lists of the necessary computational requirements, step-by-step, readily reproducible computational protocols, and tips on troubleshooting and avoiding known pitfalls. Cutting-edge and thorough, Gene Prediction: Methods and Protocols is a valuable resource for researchers and research groups working on the assembly and annotation of single species or small groups of species. Chapter 3 is available open access under a CC BY 4.0 license via link.springer.com.

Bioinformatics has evolved significantly in the era of post genomics and big data. Huge advancements were made toward storing, handling, mining, comparing, extracting, clustering and analysis as well as visualization of big macromolecular data using novel computational approaches, machine and deep learning methods, and web-based server tools. There are extensively ongoing world-wide efforts to build the resources for regional hosting, organized and structured access and improving the pre-existing bioinformatics tools to efficiently and meaningfully analyze day-to-day increasing big data. This book intends to provide the reader with updates and progress on genomic data analysis, data modeling and network-based system tools.

Evaluation of de Novo Transcriptome Assemblies from RNA-Seq Data

RNA-seq Data Analysis

Bridger: a New Framework for de Novo Transcriptome Assembly Using RNA-Seq Data

Vascular Plants as Epiphytes

The Genomes of Rosaceous Berries and Their Wild Relatives

Approaches to Understanding the Cumulative Effects of Stressors on Marine Mammals

RNA03Seq technology has revolutionized the way we study transcriptomes. In particular, it has enabled us to investigate the transcriptomes of species that have not yet had their genomes sequenced. This thesis focuses on two computational tasks that are crucial to analyzing RNA03Seq data in the absence of a sequenced genome: transcript quantification and de novo transcriptome assembly evaluation. For transcript quantification, RNA03Seq is considered a more accurate replacement for microarrays. However, to allow for the highest accuracy, methods for analyzing RNA03Seq data must address the challenge of handling reads that map to multiple genes or isoforms. We present RSEM, a generative statistical model of the sequencing process and associated inference methods, which tackles this challenge in a principled manner. Our results on both simulated and real data sets suggest that RSEM has superior or comparable performance to other quantification methods developed at the same time. To facilitate the usage of our method, we implement RSEM as a robust and user03friendly software package for quantifying gene and isoform abundances from single03end or paired03end RNA03Seq data. RSEM outputs abundance estimates, 95% credibility intervals, and visualization files and can also simulate RNA03Seq data. In contrast to other existing tools, the software does not require a reference genome. Thus, in combination with a de novo transcriptome assembler, RSEM enables accurate transcript quantification for species without sequenced genomes. Building off of RSEM, we have developed a novel probabilistic model based method, RSEM03EVAL, for evaluating de novo transcriptome assemblies from RNA03Seq data without the ground truth. Our RSEM03EVAL score has a broad range of potential applications, such as selecting assemblers, optimizing parameters for an assembler and guiding new assembler design. Results on both simulated and real data sets show that the RSEM03EVAL score correctly reflects the accuracies of the assemblies. To demonstrate its usage, we assembled the transcriptome of the regenerating axolotl limb by selecting among over 100 candidate assemblies based on their RSEM03EVAL scores.

This volume provides an overview of RNA bioinformatics methodologies, including basic strategies to predict secondary and tertiary structures, and novel algorithms based on massive RNA sequencing. Interest in RNA bioinformatics has rapidly increased thanks to the recent high-throughput sequencing technologies allowing scientists to investigate complete transcriptomes at single nucleotide resolution. Adopting advanced computational technics, scientists are now able to conduct more in-depth studies and present them to you in this book. Written in the highly successful Methods of Molecular Biology series format, chapters include introductions to their respective topics, lists of the necessary materials and equipment, step-by-step, readily reproducible bioinformatics protocols, and key tips to avoid known pitfalls. Authoritative and practical, RNA Bioinformatics seeks to aid scientists in the further study of bioinformatics and computational biology of RNA.

Next generation sequencing (NGS) has surpassed the traditional Sanger sequencing method to become the main choice for large-scale, genome-wide sequencing studies with ultra-high-throughput production and a huge reduction in costs. The NGS technologies have had enormous impact on the studies of structural and functional genomics in all the life sciences. In this book, Next Generation Sequencing Advances, Applications and Challenges, the sixteen chapters written by experts cover various aspects of NGS including genomics, transcriptomics and methylomics, the sequencing platforms, and the bioinformatics challenges in processing and analysing huge amounts of sequencing data. Following an overview of the evolution of NGS in the brave new world of omics, the book examines the advances and challenges of NGS applications in basic and applied research on microorganisms, agricultural plants and humans. This book is of value to all who are interested in DNA sequencing and bioinformatics across all fields of the life sciences.

This book examines the application of soybean genome sequences to comparative, structural, and functional genomics. Since the availability of the soybean genome sequence has revolutionized molecular research on this important crop species, the book also describes how the genome sequence has shaped research on transposon biology and applications for gene identification, tilling and positional gene cloning. Further, the book shows how the genome sequence influences research in the areas of genetic mapping, marker development, and genome-wide association mapping for identifying important trait genes and soybean breeding. In closing, the economic and botanical aspects of the soybean are also addressed.

Bioinformatics in the Era of Post Genomics and Big Data

The Genus Thymus

RNA-Seq Analysis: Methods, Applications and Challenges

Bioinformatics Algorithms

Computational Methods for Next Generation Sequencing Data Analysis

Algorithms for Next-Generation Sequencing Data

This book constitutes the proceedings of the 13th International Symposium on Bioinformatics Research and Applications, ISBRA 2017, held in Honolulu, HI, USA, in May/June 2017. The 27 full papers presented together with 18 short papers and 24 invited abstracts were carefully reviewed and selected from 131 submissions. They cover topics such as: biomarker discovery; biomedical databases and data integration; biomedical text mining and ortologies; biomolecular imaging; comparative genomics; computational genetic epidemiology; computational proteomics; data mining and visualization; gene expression analysis; genome analysis; high-performance bio-computing; metagenomics; molecular evolution; molecular modelling and simulation; next-generation sequencing data analysis; pattern discovery and classification; population genetics; software tools and applications; structural biology; and systems biology.

Next generation sequencing has made it possible to perform differential gene expression studies in non-model organisms. For these studies, the need for a reference genome is circumvented by performing de novo assembly on the RNA-seq data. However, transcriptome assembly produces a multitude of contigs, which must be clustered into genes prior to differential gene expression detection. Here we present Corset, a method that hierarchically clusters contigs using shared reads and expression, then summarizes read counts to clusters, ready for statistical testing. Using a range of metrics, we demonstrate that Corset out-performs alternative methods. Corset is available from https://code.google.com/p/corset-project/.

Jeff Martin of the DOE Joint Genome Institute discusses a de novo transcriptome assembly pipeline from short RNA-Seq reads on June 3, 2010 at the "Sequencing, Finishing, Analysis in the Future" meeting in Santa Fe, NM.

This book presents an overview of the state-of-the-art in barley genome analysis, covering all aspects of sequencing the genome and translating this important information into new knowledge in basic and applied crop plant biology and new tools for research and crop improvement. Unlimited access to a high-quality reference sequence is removing one of the major constraints in basic and applied research. This book summarizes the advanced knowledge of the composition of the barley genome, its genes and the much larger non-coding part of the genome, and how this information facilitates studying the specific characteristics of barley. One of the oldest domesticated crops, barley is the small grain cereal species that is best adapted to the highest altitudes and latitudes, and it exhibits the greatest tolerance to most abiotic stresses. With comprehensive access to the genome sequence, barley's importance as a genetic model in comparative studies on crop species like wheat, rye, oats and even rice is likely to increase.

De Novo Assembly of Bacterial Transcriptomes from RNA-Seq Data

13th International Symposium, ISBRA 2017, Honolulu, HI, USA, May 29 – June 2, 2017, Proceedings

Q: How do I know my RNA-Seq experiments worked well A: RNA-Seq QC PipelineQ: How do I detect transcripts which are over expressed or under expressed in my samples A: Counting and Statistic AnalysisQ: What do I do if I don't have a reference genome A: Rnnotator de novo Assembly.

Recent advances in Next Generation Sequencing (NGS) have allowed for unparalleled access to genetic information for organisms in both the functional and phylogenetic realms of biology. Analysis of the RNA transcripts of cells of organisms using Next Generation Sequencing (called RNA-seq) has opened doors for unique insights into the genomic complexity of organisms and has provided researchers with invaluable tools for analysis of function of gene products and phylogenetic relatedness. Application of this method has moved beyond model organisms. It has provided a lot of potentials, in ecological research and comparative transcriptomics, in non-model organisms. This thesis presents an overview on existing applications of RNA-seq in non-model organisms. Furthermore, it presents a new clustering design on handling the data, which led to identification of twelve new fluorescent protein isoforms in corals. In addition, de novo assembly and annotation of the data from polychaete Hermodice carunculata made possible the identification of one new phylogenetic marker and eight bioluminescent protein isoforms. Also, twelve new bilirubin-induced fluorescent proteins were identified from false moray eel Kaupichthys hyproroides. This approach can be applied on any other data.

Introduces readers to core algorithmic techniques for next-generation sequencing (NGS) data analysis and discusses a wide range of computational techniques and applications This book provides an in-depth survey of some of the recent developments in NGS and discusses mathematical and computational challenges in various application areas of NGS technologies. The 18 chapters featured in this book have been authored by bioinformatics experts and represent the latest work in leading labs actively contributing to the fast-growing field of NGS. The book is divided into four parts: Part I focuses on computing and experimental infrastructure for NGS analysis, including chapters on cloud computing, modular pipelines for metabolic pathway reconstruction, pooling strategies for massive viral sequencing, and high-fidelity sequencing protocols. Part II concentrates on analysis of DNA sequencing data, covering the classic scaffolding problem, detection of genomic variants, including insertions and deletions, and analysis of DNA methylation sequencing data. Part III is devoted to analysis of RNA-seq data. This part discusses algorithms and compares software tools for transcriptome assembly along with methods for detection of alternative splicing and tools for transcriptome quantification and differential expression analysis. Part IV explores computational tools for NGS applications in microbiomics, including a discussion on error correction of NGS reads from viral populations, methods for viral quasispecies reconstruction, and a survey of state-of-the-art methods and future trends in microbiome analysis. Computational Methods for Next Generation Sequencing Data Analysis: Reviews computational techniques such as new combinatorial optimization methods, data structures, high performance computing, machine learning, and inference algorithms Discusses the mathematical and computational challenges in NGS technologies Covers NGS error correction, de novo genome transcriptome assembly, variant detection from NGS reads, and more This text is a reference for biomedical professionals interested in expanding their knowledge of computational techniques for NGS data analysis. The book is also useful for graduate and post-graduate students in bioinformatics.

Methods and Protocols