International Journal of Research in Engineering and Applied Sciences

Vol. 6 Issue 6, June - 2016

ISSN: 2249-3905 | Impact Factor: 6.573 |Thomson Reuters Researcher ID: L-5236-2015

INSECT EPIGENETICS: GENOMIC CONTROL AND PHENOTYPIC VARIATION

Dr.Rama Kant

Associate Professor-Agricultural Entomology

R.S.M.College Dhampur (Bijnor) UP

ABSTRACT

Inheritance of epigenetic traits is critical to cell differentiation and development. An overview of insect molecular epigenetics is presented in this paper.. For example, epigenetic information is transferred from generation to generation by DNA alteration, histone protein modification, and noncoding RNA activity. The majority of our understanding of insect epigenetics has come from a few model species. Gene order in insect genomes is more stable than would be predicted if chromosomes evolved by random breaking. We show evidence that microsynteny has been preserved in order to preserve huge arrays of highly conserved noncoding elements (HCNEs). Genomic regulatory blocks are formed by spanning critical developmental regulating genes on these arrays (GRBs). GRBs in vertebrates, where most HCNEs operate as enhancers and HCNE arrays dictate intricate expression programs of their target genes, have recently been identified.

Keywords: Insect, epigenetics, genomic, phenotypic

INTRODUCTION

To begin, we'll go through the fundamentals of genome assembly, annotation, and criteria for evaluating the quality of draft assemblies. Afterwards, we offer an overview of genomic data for various insect species, highlighting instances from important model organisms, agricultural pests, and disease transmission vectors. Several significant insect genome databases are described as well. Development of new pest management approaches will benefit from the increased availability of insect genetic data. Data-mining algorithms that take full advantage of the insect genome resources are still possible, though.[1]The study of insect genomics has made great strides, yet there are still numerous issues to be resolved. Insect parasitism and polyphagy, for example, have been related to gene family growth or contraction utilizing comparative genomics research.. An overview of insect genome sequencing and pest control research is provided in this section. When cells divide, epigenetic information is handed down through the DNA methylation, histone protein modification, and noncoding RNA activity. We've learned a lot about insect epigenetics by studying only a few model species. However, additional studies of epigenetic information in non-model taxa will help us better grasp the evolutionary and developmental importance of epigenetic inheritance in insects. [3]The epigenome may be studied in many different ways, and we'll go through

some of the most common here. Generating novel forms and functions in a wide range of organisms across a wide range of cellular, developmental, and taxonomic levels will help us better understand how genome control leads to evolution of insect variety. HCNE density peaks centrally in large synteny blocks with many genes, as seen by this comparison of five Drosophila genomes. HCNE arrays can cover unrelated adjacent genes in addition to developmental regulators that are likely targets of HCNE enhancers. Differences in core promoters between target genes and unrelated genes are described, and this explains why enhancers have a different effect on the two groups of genes. Here, we demonstrate a significant similarity between synteny block borders, HCNE arrays and Polycomb binding areas to demonstrate that synteny blocks are regulatory domains. Drosophila and Anopheles gambiae share few non-coding elements, [5] however we found that areas orthologous to Drosophila GRBs in the yellow fever mosquito A. aegypti have an equal pattern of non-coding elements that are substantially conserved in A. aegypti. There is a structural and functional similarity between the GRBs of insects and vertebrates, making them an ancient element of the genomes of metazoans.

Insect Genome Sequencing in Context

Every order of insects is represented by the 1219 insect BioProjects at NCBI i.e. project type: primary submission. Insect genome initiatives are expected to be many because many sequencing projects are not registered with the NCBI. Diptera and Lepidoptera are in a close second and third place, respectively, in terms of the number of sequencing initiatives. Archaeognaths, Blattodea, Coleopters, Collembolas, Dipterans, Ephemerans, Hemipters, Hymenopters, Lepidans, Odonates, Orthopterans, Phthyrapterans, Plecopterans, Siphonapteras and Trichopterans are among the 18 orders of insects whose genomes have been assembled. Diptera has the most sequenced genomes, [2]followed by Lepidoptera, Hymenoptera, and Hemiptera in order of total number of assembled genomes. It's possible that the added challenges of reliably annotating assembled genomes explain why just 10% of insect genome-sequencing efforts have annotated genomes despite the wide range of assembled genomes. Insect genome annotation is currently hindered by low assembly quality, generally due to the effect of high heterozygosity. One hundred and twenty-nine NCBI BioProjects on insects i.e., type: main submission cover nearly all of the insect kingdom's classifications. Insect genome initiatives are expected to be many because many sequencing projects are not registered with the NCBI. Diptera and Lepidoptera follow Phthiraptera in terms of number of sequencing studies. A total of 18 orders of insects have built genomes, [5]including Archaeognatha, Blattodes, Coleoptera, Collembola, Diptera, Hymenoptera, Lepidoptera, Phasmatodea, Phthirapterodea, Siphonaptera, Strepsiptera and Trichoptera. ' Diptera, Lepidoptera, Hymenoptera, and Hemiptera have the most constructed genomes. It's possible that the added challenges of reliably annotating completed genomes explain why just 10 insect orders have annotated genomes, or 12.7% of the total number of insect genome-

sequencing efforts. A common and significant problem for genome annotation in insect genomes is low assembly quality, which is often owing to the effect of high heterozygosity. [2]

Gene regulatory modifications that are heritable are the focus of epigenetics, which focuses on environmental influences on gene expression. Since various variables might have a long-term effect on gene regulation, epigenetic information can be broadly defined. In contrast, the discipline of molecular epigenetics is focused on mechanisms that directly change, alter, or interact with the chromatin of a genome. [3] Mitotic cell division inside people and meiotic cell division resulting in progeny can both convey epigenetic information. The process of development and the conundrum of how a single-celled egg may evolve into a multicellular creature composed of several tissues are at the heart of intragenerational epigenetic inheritance. To the contrary, the intergenerational transmission of epigenetic information to children is the focus of intergenerational epigenetic inheritance. Even though it is supposed to be rare, intergenerational epigenetics influences evolution directly, therefore it is significant. Despite this, much of the research on molecular epigenetics in insects has been done at the intragenerational level. [6]

Methylation of DNA

In terms of epigenetic information, DNA methylation is perhaps the most reliably heritable. DNA methylation, in contrast to other kinds of epigenetic information, is mostly lacking from the genomes of many insect species. [5]There is no indication of considerable DNA methylation in the genomes of any of the six analyzed insect orders, including Drosophila, that have been examined. Even yet, entomologists have recently become interested in studying DNA methylation in an attempt to better understand how it affects embryonic plasticity.

There are two types of enzymes in animals that are involved in DNA methylation: de novo DNA methyltransferases and maintenance DNA methyltransferases (DNMT3 proteins) (DNMT1 proteins). CpG contexts in animals are the most common place to find CpGs, and the addition of methyl groups to them is the most common method of increasing DNA methylation in animals. As CpGs are symmetrical, they provide a method for reestablishing methylation patterns following semiconservative DNA replication; DNMT1 restores symmetry by targeting hemimethylatedCpGs. However, methylation of cytosine has also been found in insect genomes, albeit at a considerably lower degree than in CpG contexts. Hence, in a CpG context, we refer to DNA methylation as the same as cytosine methylation. [7]

It's become popular to study long-range cis-regulation in vertebrates, because to the recent finding of genome-wide HCNEs (highly conserved noncoding elements), which span the loci

of developmental regulating genes. Using whole-genome comparisons, researchers discovered an abundance of HCNEs that tend to cluster throughout the chromosomes after seeing significant conservation of individual developmental enhancers. [8] Clusters of genes coding for developmental and differentiation-related transcription factors are most frequently found in the vicinity of the clusters. Individual gene analyses have revealed that several HCNEs serve as long-range enhancers. Several overrepresented motifs in HCNE sequences have been discovered by genome-wide analysis, and they are thought to be linked to context-specific enhancer activation. As a result of this developing concept, the actual expression pattern of the gene is defined by the total number and location of HCNEs in its target gene(s). Genes with the most complicated spatiotemporal expression should have more complex regulatory inputs, according to this theory. Axonal guidance and associated central nervous system functions have been shown to be the primary targets of the most extensive HCNE arrays, which is in keeping with the findings of this study. [1]

DNA assembly in insects

Currently, most insect genome efforts use a whole genome shotgun method (WGS), which yields a large number of short sequence fragments that must be assembled. De novo assembly or mapping assembly can be used for this purpose, depending on whether or not a reference genome is available. Genome assemblies that are de novo are built from scratch, while those that are mapping are built from a reference genome and then assembled from the overlapping information between the two assemblies. [3]

Algorithms for de novo assembly may be divided into three groups. Cabog, newbler, shorty, edena and many more fall into the first type of assemblers that are based on the overlap/layout/consensus between lengthy sequences. For assembling medium-length reads obtained by the Sanger sequencing method, these software packages can be used; however, they are often less effective for'short-read' sequencing methods, such as PacBio and Nanopore sequencing. [4]

Repetitive sequences and heterozygosity are the two key variables that influence the quality of insect genome assembly. The assembly of contigs and scaffolds might be complicated by the presence of many repeated sequences in the genome. Genome assembly is further complicated by the sequenced individual's heterozygosity, [5]or allelic variance. As a result of heterozygosity, genome assembly data should be gathered from inbred homozygous individuals or haploid males for the Hymenoptera species. There have been recent efforts to create techniques specifically tailored for the assembly of heterozygous genomes, however, because collecting such materials is typically difficult or impossible. The assembly quality of heterozygous genomes has been claimed to be improved by two programs at the moment: platanus and redundans.

The abundance of noncoding components that are well conserved is higher in biga block in synteny. According to Methods, we selected synteny blocks that were shared by all fly genomes to analyze the distribution of HCNEs. The genomes of the four species we compared to Dmel are all incomplete. [6]

Although the sequence is mostly broken up into extremely big scaffolds, our findings show that it is possible to assemble trustworthy synteny blocks. Although the synteny blocks had only a few scaffolds, they covered 76% of the euchromatic sequence of Dmel. We distinguish between the span and coverage of a synteny block, the latter of which refers to the mutually aligned syntenic bases inside the block that lie between the block's extreme boundaries. A total of 94% of the HCNEs were completely spanned by synteny blocks, while an additional 86% had at least a 98% synteny block coverage. It was our goal to compare synteny block coverage of the HCNE sequence with CDS coverage while taking into account the fact that CDS is less conserved overall.. Since the Dmel sequence was aligned in a reciprocal-best way in all four pairwise genome comparisons (reciprocally best aligned [RA] sequence), [9] we were able to determine the percentage of these bases covered by synteny blocks. While only 75% of the R-HCNE sequence was covered by synteny blocks, 90% of the R-HCNE sequence was covered by synteny blocks. Large synteny blocks in the RA-HCNE sequence were more prevalent than in the RA-CDS sequence.

Changes in Proteins Associated With Chromatin

The vast bulk of DNA in metazoan nuclei is found in nucleosomes, which are protein complexes made up of eight histone proteins and contain 147 bp of DNA. Histones H2A, H2B, H3, and H4 are found in each of the two tetramers of these proteins. In eukaryotes, Histone proteins are among the most conserved and fundamentally significant proteins in the organism. Nucleosome-bound DNA is more difficult for other proteins to access. When it comes to transcription factors, for example, nucleosome-free regions are preferred. [3]As a result, the modification of the histone–DNA association has been related to several regulatory mechanisms in eukaryotes, and many additional proteins bind to histones. Nucleosomes can be manipulated in a variety of ways to influence gene expression. Hetone proteins may be altered during translation. Second, a different sequence variation of a histone can be used in place of the core histone. For the last step, a nucleosome may be physically shifted to expose key underlying binding sequences. Nucleosomes and their changes appear to be heritable throughout cell divisions. [4] Therefore, epigenetic alterations such as histone modification and replacement are critical in eukaryotes. Even though histones have long been researched in Drosophila as a model organism, it is only lately that histones have been examined in non-model insect species. Numerous, functionally conserved, and firmly connected to different kinds of gene regulation have been found in a variety of organisms. Histone proteins are more directly involved in the modulation of phenotypic flexibility than DNA methylation.

Vol. 6 Issue 6, June - 2016

ISSN: 2249-3905 | Impact Factor: 6.573 |Thomson Reuters Researcher ID: L-5236-2015

Annotation of a genome

Genome annotation is essential for determining the functions of the genome's structural components. Annotating a structure and annotating a function are two distinct processes. In order to identify which portions of the assembly correspond to certain characteristics (such as genes and transposable elements), structural annotation is the first step. Annotating genes and other elements based on sequence similarities is a common method for inferring their functions and identities once the structural characteristics have been defined. [8]

Identifying recurring patterns

There are two types of methods for detecting repeated sequences: homology searching and a priori prediction. Through the use of sequence similarity, a method known as homology searching may locate homologous repeat sequences. The program repeatmasker and the RepBase collection of TEs are commonly used for this job. Ab initio prediction employs structural aspects of the repeating sequence to find new repeating patterns. [9]If you're trying to forecast repeating structures like small inverted-repeat TEs or lengthy terminal repeats, this technique is ideal. Recon, piler, repeatscout, ltr-finder, and repeatmodeler are some of the most often used software packages for this purpose. Homology searches and ab initio methods are commonly employed to identify repetitive sequences in insect genomes.

Noncoding RNAs have been identified

Genes that do not encode proteins are known as noncoding RNA genes. Examples include transfer RNA, small nucleolar or repeat-associated small interfering RNA, small nucleolar RNA, piwi-interacting RNA, microRNA, and so on. Many biological activities are controlled by noncoding RNAs. As a result, the process of identifying noncoding RNA in the genome is critical. Noncoding RNA may be identified using a variety of software tools, including mirdeep, trnascan, infernal, and rnastructure. This may be done with the use of noncoding RNA databases. RNAdb, NONCODE, Rfam, miRBase, and snoRNABase are examples of databases that provide information about RNA. [4]

Prediction of genes that code for proteins

The most critical component of structural annotation is the discovery of genes that code for proteins. Protein-coding genes can be predicted from the genome by identifying homologues of known protein-coding genes through sequence similarity; de novo predicting the protein-coding genes with software developed via machine learning of protein-coding gene structures; and determining exonic regions by direct transcriptome sequencing. Although the protein-coding genes uncovered by software produced by machine learning protein-coding gene structures to find homologues of existing protein-coding genes; [6] and software developed by machine learning protein-coding gene structures to predict new protein-coding

genes. For the most definitive proof, RNA-seq data is often the best option, but it relies greatly on how much transcriptome data is available and which samples are used for RNA-seq. It is already normal practice to combine all three levels of evidence into a single estimate of gene structure in order to increase protein-coding gene accuracy. [3] There are a variety of software tools available to help with this integrated approach. Some examples are augustus/evidencemodeler and glean/evigan/maker/jigsaw. omiga, a genome annotation pipeline for insects, was created in 2014. For the de novo prediction program, omiga not only integrates the three sources of evidence, but it also detects CDS from the assembled transcriptome. Prediction accuracy can be greatly enhanced by completing this step. The NCBI's genome annotation process for eukaryotic genomes is also frequently utilized.

Food-producing invertebrates

The term "resource insects" refers to insects whose products, or perhaps the insects themselves, have a high value to the economy. Currently, only Dactylopiuscoccus, three silkworm moths, and four honey bees have had their genomes sequenced for use as a resource. Among the insects introduced here are B. mori and A. mellifera. [9]

CONCLUSION

The B. mori genome was sequenced and put together in 2004. A. gambiae's genome was the first to be sequenced in a domesticated mammal, making it a valuable resource for research into the effects of domestication on gene evolution over lengthy periods of time. Using a WGS approach, the B. moriDazao strain was sequenced Genome size is 428.7 Mb, contig N50 is 12.9 kb, scaffold N50 is 26.9 kb, and 18 510 genes have been found by annotating the assembled genome data. Contig N50 was extended to 15.5 kb in 2008 by the International Silkworm Genome Consortium, while scaffold N50 was increased from 3.7 M to 3.7 M. An estimated 14 623 genes, including 3223 that are unique to B. mori, made up 87% of the scaffold and were anticipated to be located on 28 chromosomes in total. Researchers have found a link between B. mori's high silk production and an increase in the number of copies of a tRNA gene cluster. In order to encode the silk gum components, [6]Ser1, Ser2, and Ser3 genes were employed. Transported amino acids glycine, alanine, and serine have much more tRNA gene numbers than other amino acid tRNAs The fructofuranosidase gene, which allows B. mori to breakdown alkaloids in its mulberry host plant Folium mori that are harmful to other insects like 1,4-dideoxy-1,4-imino-d-arabinitol and 1-deoxynojirimycin, was acquired by B. mori from bacteria by horizontal gene transfer.

REFERENCES

[1] Alabert C, Barth TK, Reveron-G ´ omez N, Sidoli S, Schmidt A, et al. Two distinct modes for ´ propagation of histone PTMs across the cell cycle. (2015)

[2] Allen, J.E., Majoros, W.H., Pertea, M. and Salzberg, S.L. jigsaw, genezilla, and glimmerhmm: puzzling out the features of human genes in the ENCODE regions.(2006)

[3] Bassett AR, Tibbit C, Ponting CP, Liu J-L. Highly efficient targeted mutagenesis of Drosophila with the CRISPR/Cas9 system. (2013)

[4] Bose P, Dai Y, Grant S. Histone deacetylase inhibitor (HDACI) mechanisms of action: emerging insights. (2014)

[5] Chen, X.G., Jiang, X.T., Gu, J.B., Xu, M., Wu, Y., Deng, Y.H.et al. Genome sequence of the Asian tiger mosquito, Aedesalbopictus, reveals insights into its biology, genetics, and evolution. (2015)

[6] Davies, N.J. and Tauber, E. (2015) WaspAtlas: a Nasoniavitripennis gene database and analysis platform. Database: The Journal of Biological Databases and Curation. (2015)

[7] Davie K, Jacobs J, Atkins M, Potier D, Christiaens V, et al. Discovery of transcription factors and regulatory regions driving in vivo tumor development by ATAC-seq and FAIRE-seq open chromatin profiling. (2015)

[8] Kung JTY, Colognori D, Lee JT. Long noncoding RNAs: past, present, and future. (2013)

[9] Lyko F, Foret S, Kucharski R, Wolf S, Falckenhayn C, Maleszka R. The honey bee epigenomes: differential methylation of brain DNA in queens and workers. (2010)