

Bizarre movements of Earth cause deadliest tragedies and Climate Change

Received for publication, December 10, 2019, and in revised form, January 15, 2020

Nagendra Nath Mondal¹

¹Department of Physics, Techno International, B7-360 / New, Ward No. 30, Maheshtala, South 24 Parganas Pincode- 700141 West Bengal, India

Abstract

A few deadliest catastrophes in the Asia-Pacific regions are Tsunami on 22.12.2004; Aila on 25.05.2009; earthquakes on 11.03.2011 and 25.04.2015; Tornados on 26.04.1989 and Cyclones on 12.11.1970. Very recently Cyclone Fani has smashed parts of Odisha and West Bengal in May 3 – 5, 2019, which has befallen due to climate change. The monitoring of rotational and orbital motions of the Earth is one of the key objectives of this research to observe whether those have any role behind these calamities. Since April 2005 data has been recorded, and analyzed data infers the climatic change. Comparative studies between the experiment and theory show that the earth was walking away during the period of tsunami 2011 to Kathmandu earthquake 2015. A bizarre etiquette of the earth has increased *zenith* to 0.3686° and *rotational angle* to 5.58° annually. If this state of affairs continues the local temperature will be 50°C in 2050.

Keywords: Earthquake, Environmental temperature, planetary motion, solar spectrum, Tsunami

Introduction

The most deadliest Tsunami was in 2004, Aila in 2009 south Asian countries; Earthquakes in 2011 in Sendai, Japan (resulted tsunami), and in 2015 in Kathmandu, Nepal; Tornados (The world's deadliest tornado struck in the Manikganj district of Bangladesh in 1989) and Cyclones (the *Bhola cyclone* was a devastating tropical cyclone that had struck Bangladesh and India's West Bengal in November 1970). Very recently Cyclone Fani has smashed parts of Odisha and West Bengal in May 3 – 5, 2019, which has befallen due to climatic change. That's only probable to acquire worse as climate change increase the surface temperatures of the sea and triggers more intense tropical cyclones (1). Main objective of this research is to keep eyes on the motions of the Earth and its impact on environmental issues, especially on climatic change. Data have been taken since April 2005 after the Tsunami in 2004. Subsequently analyzed Zenith angle (θ), rotational angle (φ) and the temperature (T) of the Earth's atmosphere show that the Earth's orbital plane has been changing. Maximum and minimum of θ , φ and T have, also, been changed. Recently, we have seen that origins of variation of the solar luminosity (L_s), what the Earth is getting are principally the absorption by the Earth's atmosphere. It is due to sundry θ , and φ . Another important finding of this measurement was the shape of the movement which is like an Ink-pot. Significant variations of T and angular momentum were observed during monsoon (2) The chromaticity of solar radiation spectrum is studied with the help of a Solar Spectrum Mon2itor (SSM) system that can detect individual color of the spectrum and it suggests that the sun's visible spectral irradiance changes significantly from May 2009 to September 2012 (3).

Corresponding author

Email address: nn.mondal2011@gmail.com

We have developed another SSM with different colored LEDs (Victoria Lighting Pvt. Ltd.) and applied in reverse biased mode. Experimental and theoretical analyzed data had confirmed the attachment of *Celestial Positronium* (Ps) in the sun and its impact on solar radiation spectrum (4). The origin of the solar spectrum and its associated colors are the fusion reactions, thermal radiation and fundamental interactions of particles-antiparticles. We are worried about the global warming, earthquake, volcanic eruption, tornado and other relevant issues of natural causalities and some of them are directly influenced by the solar irradiation and motion of the earth. A detector based on a diffraction grating is an excellent and inexpensive tool for studying the chromaticity of solar spectrum. The sun is one of the second largest G2-type stars in the Milky Way which is going to grow old, since its birth about 4.5 billion years ago. Bose-Einstein Condensate (BEC) is the fifth state of matter and its studies infer that Ps-BEC is the densest ingredient of our early celestial bodies (5, 6).

A bizarre etiquette of the earth is studied and it is found that its motion had made the tsunami in 2011. The report is framed on the measurement of θ , φ , T and solar spectrum. Some inconsistencies among the parameters have been found, after excavating those data from the month of February to April at around 11:00 hrs. The average rate of change of θ in 2007, 2011 and 2016 respectively are 25'/day, 23'42''/day and 17'24''/day, and average changes of φ are found to be 65°15', 71°11' and 63°46' respectively. Remarkable variations of those astronomical parameters show offbeat motions of the earth (7). Recent experimental evidence shows that the deficit of higher frequency photons in solar spectrum transforms greenery to desertification (8). In a recent study we find that mortality of trees of Amazonian rainforest has been driven by climate and functional traits. This report is based on the monthly monitoring of over 1,000 survival rainforest canopy trees from over 200 species over five decades in the Central Amazon (9).

Experiments

In order to observe the motion of the earth and temperature of the local area a solar spectrum monitor (SSM) was developed using a spike on a protractor based table, thermometer, color filters and photodiodes. Details of this SSM and data taking system can be found elsewhere [1, 4].

Measurement of Earth's Motions

Earth's motions are of two kinds: (a) Rotational motion around its own axis. It takes 24 hrs. to complete a single rotation. (b) Orbital motion around the sun which takes 365 days or 8760 hrs. Both data were recorded manually by measuring the shadow length and its position respectively during time 7:00 - 16:00 in an hour interval.

Measurement of Temperature

In each position of the earth T of the local area was recorded with a laboratory thermometer. In this report only higher T at sunny days around noon are taken into account. In Fig. 1 a typical spectrum of T is displayed. Data are plotted as a histogram and fitted it with Gaussian. The mean T is noted to be 38.40° C.

Results & Discussion

In this report the highest T per week is considered, i.e., every months 4 – 5 data of each parameter are taken which resulted in 48 – 50 data (January – December) in a year except 2015 (August – December). Data are analyzed statistically in order to find out the average of

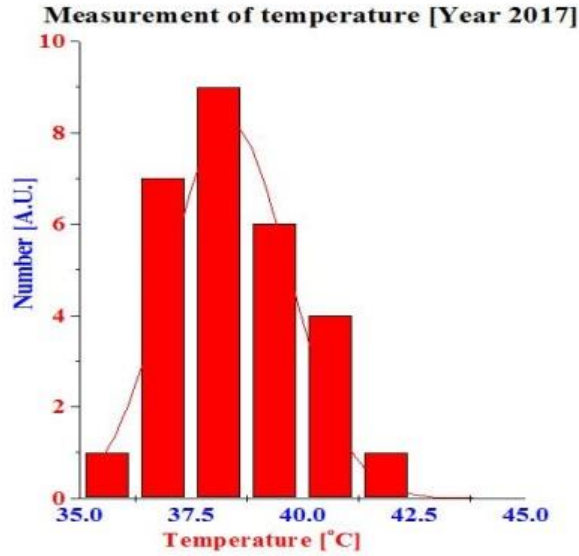


Fig 1. A histogram is depicted of measured Temperature.

each parameter and those averages are plotted along the vertical axis with respect to year along the horizontal scale. Spectra of each parameter are depicted in Fig. 2 (a-d).

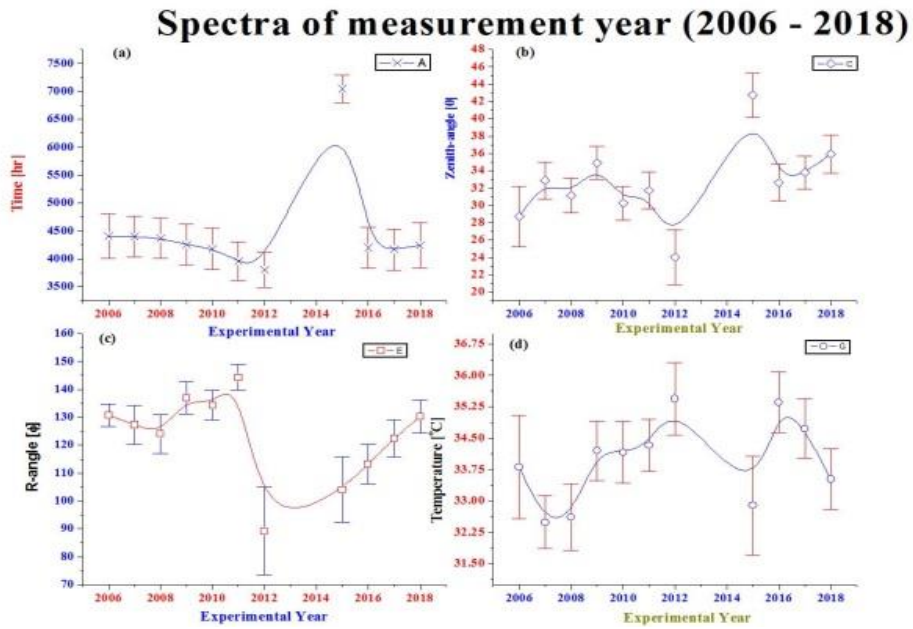


Fig 2. Experimental results of different astronomy and environmental parameters

The statistical and observational errors were estimated using the following formula.

$$\hat{\sigma}_{err}^2(x) = \frac{\sum_{i=1}^n [x_i - \bar{x}]^2}{n-1} \quad (1)$$

$$\text{Or, } \sigma_{err}(x) = \frac{\hat{\sigma}_{err}}{\sqrt{n}} \quad (2)$$

Where n is the number of recorded data and x is the experimental parameters. The vertical error bars show the uncertainty of each measurement. In order to minimize the experimental and observational errors a huge number of data were collected throughout the year. There is no recorded data of the year 2013 – 2014. In **Fig.2** (a) average time in hour is illustrated and it declines from 2006 to 2012. During this period T goes up (see spectrum d). The θ goes up till Tsunami 2011 and goes down to devastating earthquake in Kathmandu, Nepal 2015 (see (b)). After then it rises sharply to get back its track in 2018. This kind of bizarre movement of the earth is also reflected by the spectrum-(c) of φ , where φ increases from 2006 to 2011 and falls till 2015 and then gradually rises again till 2018.

Spectra of **Fig.2** are linearly fitted with $Y=A+B \times Year$ and the results are displayed in Table 1. A is the intercept and B is the slope of the straight line equation. **Table 1** (a) and (d) infer that T gradually increases with the onset of March instead of June during 2006 – 2012, i.e., a seasonal transformation - summer sets in earlier than before, which results in climatic changes. Again it rises from 2015 after the Kathmandu devastation. Spectral analysis of (b) concludes that the θ is increasing with 0.37° per year although sharp fall and rise are seen between the years 2011 and 2015 during the deadliest tragedies of nature.

Table 1. Results of the fitted parameters of **Fig.2**

Spectrum	Duration	$A \pm \text{error}$	$B \pm \text{error}$	Remarks
(a)	2006 – 2012	219978.84 ± 30468.45	-107.41 ± 15.16	Decline
	2015 – 2018	-38626.24 ± 62359.98	21.23 ± 30.92	Incline
(b)	2006 – 2018	708.60 ± 568.56	00.39 ± 00.28	Incline
(c)	2006 – 2011	-4933.84 ± 2076.08	02.53 ± 01.03	Incline
	2015 – 2018	17309.84 ± 579.33	08.64 ± 00.28	Incline
(d)	2006 – 2018	209.52 ± 144.06	00.12 ± 00.07	Incline

A sharply increase of φ is 2.53° per year is to be found from 2006 – 2011 and after Kathmandu disaster it rises more rapidly to 8.64° per year (2015 – 2018). As a result bizarre movement of the earth plays crucial role to increase T about 0.12° C per year from 2006. Local temperature increases to 1.56° C after 13 years.

In order to confirm the bizarre drive of the earth a comparative studies of declination between experiment and theory is executed. The theoretical declination of the earth is given by

$$\delta_{th} = 23.45 \times \sin \left\{ \frac{360}{365} \times (d + 284) \right\} \quad (3)$$

Where, variable d is the day of the experimental year. Similarly experimental declination is determined by the following equation,

$$\delta_{ex} = 90^\circ - (22^\circ 59' + 23.45^\circ + \theta) \quad (4)$$

Where 90° , $22^\circ 59'$ and 23.45° are normal to the earth surface, position of the experiment (Kalyani, $88^\circ 27'$ E and $22^\circ 59'$ N) and inclination of the earth respectively. Results of equations (3) (Red) and (4) (various colors) are depicted in **Fig. 3**. The theoretical and experimental spectra attribute resemblance of declination of every experimental year except 2011. The displacement of Tectonic plates inside the globe is one of the major causes behind it. The line of tsunami 2011

is noted in the spectrum of 2011 after which the bizarre nature of the earth is clearly visible. This nature of movements continue until 2015 (see **Fig.2 (b) and (c)**).

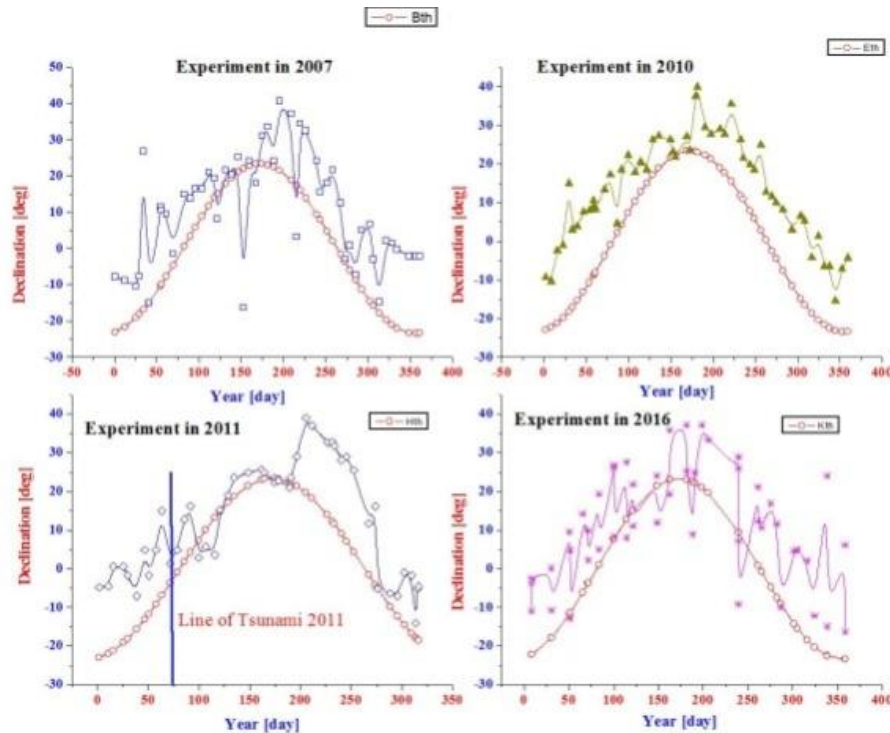


Fig 3. Theoretical and experimental declinations

Hence, it is not only an alarming situation but also a Global threat for many Asiatic countries like India, Bangladesh, Maldives, Sri Lanka, Indonesia, Japan, Philippines etc.; the countries which are surrounded by many Seas. More than 5 sq. km Island of Sunderban is sinking every year because the sediment has been inflating the basement and the Global warming has been melting the large amount of Arctic Sea ices annually. Not only have that, but a seasonal transformation (climatic change) forced extinction of many species of insects, grass, bush, birds, animals, seeds and fruits in our local area. In a recent research it is found that ghost forests created by the submergence of low-lying land are one of the most striking indicators of climate change along the Atlantic coast of North America. It has led to new recognition that the submergence of earthly land is geographically prevalent, ecologically and economically important, and globally relevant to the survival of coastal wetlands in the face of rapid sea level rise [10]. British Broad Casting (BBC) news in April 30, 2019 mentioned that Indonesia is shifting its capital city Jakarta as it is sinking into the Sea. News also dubbed that Jakarta is one of the fastest- sinking cities in the world [11]. According to Reuter many parts of the city has sunk 13 ft in the past 30 years and it will be fully flooded by 2050 expected by several researchers. Many countries are suffering from heat waves. In a news of Japan Times we have found that millions dead fish causes environmental stink in Australia (fishes died in the Darling River, New South Wales). Not only in Asia but also in European countries like UK heat-wave of 2018 led to changes in consumer behavior in summer, including large increases in electricity demand due to increased use and intensity of refrigeration and air-conditioning devices [12].

Conclusion

A bizarre movement of the earth is studied by measuring the rotational, orbital motions and local temperature. The significant results of every measurement infer that those parameters play crucial roles behind the deadliest tragedies in the past, present and will act in future too. In order to save living creatures and our beloved planet we should develop scientific awareness to control the vagaries of nature.

Acknowledgements

Author is pleased to acknowledge to his family members for allowing him to perform this prolong experiments in the house premises and having financial support from them.

References

- [1] Dasgupta A. and Priyadarshini S., Why Odisha is a sitting duck for extreme cyclones doi:10.1038/nindia.2019.69 Published online 29 May 2019.
- [2] Mondal N. N., Versatile nature of the earth causes disasters. *Journal of Environmental Research and Development* Vol. 9 No. 01, 75–82, 2014
- [3] Mondal N. N., Study chromaticity of Solar spectrum, *International Journal of Astronomy and Astrophysics*, Vol. 4, 510-518 (2014).
- [4] Mondal N. N., Discovered solar Positronium, *International Journal of Astronomy and Astrophysics*, Vol. 4, 620-627, 2014.
- [5] Mondal N. N., Development of solar spectrum monitor based on diffraction grating. *IOSR Journal of Applied Physics (IOSR-JAP)* e-ISSN: 2278-4861. Volume 8, Issue 6 Ver. II 01-03 2016.
- [6] Mondal N. N., Does the fifth state of matter originate the early universe? *Journal of Physical Science and Application* Vol. 5 Issue 6, 407-414 2015.
- [7] Mondal N. N., Studies atypical behavior of earth during Tsunami 2011 in Japan, *Journal of Environmental Research and Development* Vol. 12 No. 01, 8 –14, 2017.
- [8] Mondal N. N., Deficit of higher frequency photons in solar spectrum transforms greenery to desertification, *Journal of Environmental Research and Development* Vol. 13 No. 02, 181 –185 2018.
- [9] Izabela Aleixo, Darren Norris, Lia Hemerik, Antenor Barbosa, Eduardor Prata, Flavia Costa and Lourens Poorter, Amazonian rainforest tree mortality driven by climate and functional traits, *Nature Climate Change* 9, 384 – 388, 2019.
- [10] Katthaw L. Kiran and Keryn B. Gedan, Sea-level driven land conversion and the formation of ghost forests. *Nature Climate Change* 9, 450–457, 2019.
- [11] <https://earth.gizmodo.com/indonesia-is-moving-its-capital-city-as-the-current-one-1834420363>.
- [12] Shaun Larcom, Po-Wen She and Terry van Gevelt, The UK summer heatwave of 2018 and public concern over energy security, *Nature Climate Change* 9, 370–373, 2019.

A review article on ChIP-Seq tools: MACS2, HOMER, SICER, PEAKANNOTATOR AND MEME

Received for publication, December 15, 2019, and in revised form, January 8, 2020

Shrestha Bhowmik¹, Anish Ganguly^{2*}

¹Department of Microbiology, University of Calcutta

²Department of Environmental Science, University of Calcutta

Abstract:

Sequencing techniques are improving day by day and along with its improvement Chromatin Immuno-Precipitation followed by high-throughput sequencing (ChIP-Seq) for genome-wide analysis of protein-DNA interaction is getting popular (1). But due to lack of proper powerful ChIP-Seq analysis method several tools with different functions have been developed to identify different transcription factor binding sites, histone modifications, etc., that influences genome complexity along with other post-transcriptional changes (2). MACS2, HOMER, SICER, PEAK ANNOTATOR, MEME, etc., are some of the tools that help to identify different types of genome modifications due to protein-DNA interaction (3) (4).

Keywords: ChIP-Seq, transcription factor binding sites, histone modification, MACS2, HOMER, SICER, PEAK ANNOTATOR, MEME, post-transcriptional changes.

Introduction

Immunoprecipitation (IP) is the technique of precipitating a protein antigen out of a blend with the administration of an antibody that specifically binds to that particular protein. This process can be used to seclude and concentrate a specific protein from a test containing many thousands of different proteins. Immunoprecipitation requires an antibody to be conjoin to a solid substrate at some point in the procedure (5) Chromatin Immunoprecipitation (ChIP) uses the same technique of immunoprecipitation to precipitate the particular protein that interacts with the DNA *in vivo* (Fig 1).

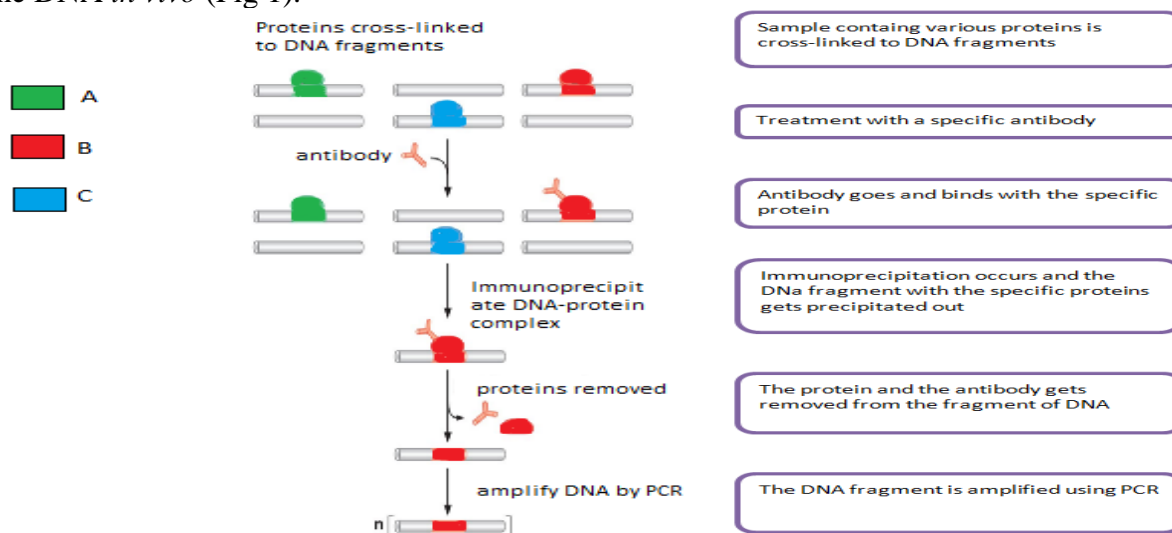


Figure 1:-Workflow of Chromatin Immunoprecipitation(Retrieved from <https://patents.google.com/patent/US20160140289A1/en>)

ChIP-Seq is used primarily to determine how transcription factors and other chromatin-associated proteins (histones, RNA polymerase, etc.) influence phenotype-affecting (morphology, biochemical or physiological properties and behaviours, etc) mechanisms[6]. It is essential to understand how proteins interconnect with DNA for authorisation of gene expressions which in turn is important for understanding the various biological processes and disease states. ChIP-Seq technology allows:

1. Combination of chromatin immune-precipitation (ChIP) with ultra-high-throughput massively parallel sequencing.(7)

2. Portraying of protein–DNA interactions *in-vivo* on a genome-scale.(8,9)

Sequencers are used to perform the DNA sequencing process. In **ChIP-Seq** many types of sequencing platforms are used which are listed below:-

Table1. Different types of Sequencers

NAME OF SEQUENCERS	SIZE	BASE LENGTH
Solexa (Illumina)[7]	1 GB of sequences in a single run	35 bases in length
454 Life Sciences (Roche Diagnostics)[8]	25-50 MB of sequences in a single run	Up to 500 bases in length
SOLiD (Applied Biosystems)[9]	6 GB of sequences in a single run	35 bases in length

Quality control of given datasets are done using **FASTQC** software where we get the idea of the read sequence length, number of tags, GC content, adapter content etc(10). Thereafter, adapter trimming is done and reads are aligned to the reference genome using **BOWTIE** software and the aligned reads are further used for peak calling(11).

Peak Calling is a computational protocol which is utilised to detect sectors in a genome that have been augmented with aligned reads as a result of performing a ChIP-sequencing (12). These zones are those where a protein links with DNA. When the protein is a transcription factor, the enhanced area is its transcription factor binding site (TFBS) (12). Popular software programs we use for transcription factors containing data sets include **MACS2** (13). When the protein is a histone modification the software we generally use is **HOMER** or **SICER** (14).

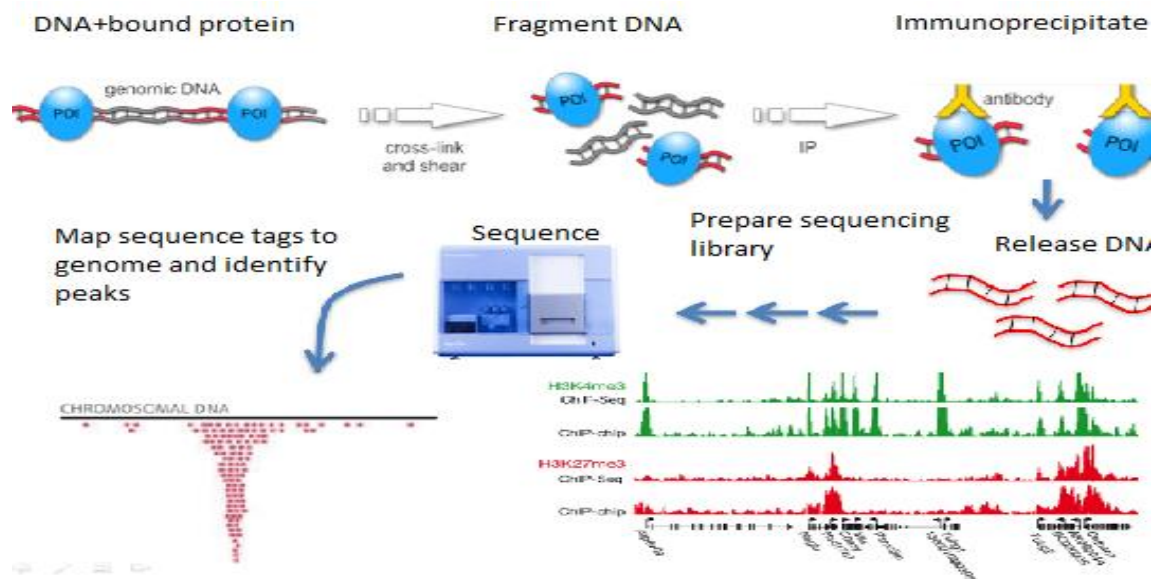


Figure 2:- Workflow for ChIP-Seq

The workflow of ChIP-Seq data generation and analysis is given below:

Different Tools Used for ChIP-Seq Analysis:

FastQC is a software executed for quality control.[14] The protocol in FastQC is employed by a succession of analysis modules like basic statistics, adapter content, sequence duplication levels, individual sequence quality score, individual sequence GC content, sequence length distribution, over expressed sequences, etc. The results of the module give a normal score when it shows green tick, slightly abnormal if it shows orange triangle or very unusual if it shows red cross[10].

Table 2: Summary of the modules with quick evaluation(Leggett et. al, 2013)

!	Basic Statistics
+	Per base sequence quality
+	Per tile sequence quality
+	Per sequence quality scores
+	Per base sequence content
!	Per sequence GC content
+	Per base N content
!	Sequence Length Distribution
!	Sequence Duplication Levels
+	Overrepresented sequences
+	Adapter Content

BOWTIE is the alignment software which is utilized to align the files.[11]This software is an meteoric, memory-efficient short read aligner. It orients short DNA sequences (reads) to the human genome at a rate of over 25 million 35-bp reads per hour. Bowtie gives its result in SAM format[11]’

MACS2 or Model based Analysis on ChIP-Seq is used on short read sequencers such as Genome Analyzer (Illumina / Solexa). This is a peak calling software which usually operates for transcription factors that gives narrow peaks although it has option for analyzing for both narrow and broad peaks[13].

Usage

macs2 [-h] [--version]

{callpeak,bdgpeakcall,bdgbroadcall,bdgcmp,bdgopt,cmbreps,bdgdiff,filterdup,predictd,pileup,randsample,refinepeak}

Example for regular peak calling: **macs2 callpeak -t ChIP.bam -c Control.bam -f BAM -g hs -n test -B -q 0.01**

Example for broad peak calling: **macs2 callpeak -t ChIP.bam -c Control.bam --broad -g hs --broad-cutoff 0.1**

There are twelve functions available in MAC2S serving as sub-commands.

Table3. Twelve Sub-commands of MACS2 (Retrieved from: <https://taoliu.github.io/MACS/>)

Subcommand	Description
callpeak	Main MACS2 Function to call peaks from alignment results.
bdgpeakcall	Call peaks from bedGraph output.
bdgbroadcall	Call broad peaks from bedGraph output.
bdgcmp	Comparing two signal tracks in bedGraph format.
bdgopt	Operate the score column of bedGraph file.
cmbreps	Combine BEDGraphs of score from replicates.
bdgdiff	Differential peak detection based on paired four bedGraph files.
filterdup	Remove duplicate reads, then save in BED/BEDPE format.
predictd	Predict d or fragment size from alignment results.
pileup	Pileup aligned reads (single-end) or fragments (paired-end)
randsample	Randomly choose a number/percentage of total reads.
refinepeak	Take raw reads alignments, refine peak summits.

HOMER or Hyper-geometric Optimization of Motif Enrichment (<http://homer.ucsd.edu/>) is a software whose operation is based on the tracing the location of histone modification that gives broad peaks[4].

Make Tag Directory: It facilitates the analysis of ChIP-Seq, by transforming the sequence alignment into platform independent data structure representing the experiment which is analogous to loading the data into a database. HOMER does this by creating a "Tag Directory", which is essentially a directory on your computer that contains several files describing your experiment where all the relevant data of the experiment is placed[4][15]

makeTagDirectory also performs several quality control steps shown below.

tagInfo.txt – this file contains all basic configuration information, such as the total number of reads, the total number of unique positions with aligned reads (by genome and chromosome), and various other statistics.

tagLengthDistribution.txt – this file contains a histogram of read lengths used for alignment.

tagCountDistribution.txt – this file contains the histogram of clonal read depth, that shows the number of reads per unique position. [4][15]

tagAutocorrelation.txt- The autocorrelation routine creates a distribution of distances between two adjacent reads in the genome. If reads are mapped to the same strand, they are added to the first column. If adjacent reads map to different strands, they are added to the second column. The results from autocorrelation analysis are very useful for troubleshooting problems with the experiment, and are used to estimate the fragment length for ChIP-Seq. The fragment length is estimated by finding the position where the "opposite strand" distribution is maximum. (<http://homer.ucsd.edu/>)[16]

SICER or Spatial clustering for Identification of ChIP-Enriched Regions is a clustering approach for identification of enriched domains from histone modification ChIP-Seq data[17]. The key concept that SICER uses to capture spatial clustering of reads is island[17]. To delineate the islands and assess the statistical significance of ChIP enrichment on them, SICER carry out the following steps:

It partitions the genome into non-overlapping windows of size w .

It identifies windows with enrichment (i.e., "eligible" windows). A window is deemed "eligible" or "ineligible" if the number of ChIP-Seq reads in this window is equal to or above/below a read count threshold l_0 .

It identifies islands as clusters of "eligible" windows separated by gaps of size no larger than a predetermined value g . A gap is a contiguous stretch of "ineligible" windows between two neighboring "eligible" windows. When $g = 0$, islands are uninterrupted clusters of "eligible" windows. [17].

It identifies "candidate" islands that exhibit significant clustering of "eligible" windows that are unlikely to appear by chance. If a control library be found, SICER will further filter the "candidate" islands using that control library, releasing all excluding those that exhibit significant enrichment of ChIP signal compared to control on the islands[17]. The statistical significance versus control library is characterized by a p -value based on Poisson distribution. A false discovery rate (FDR) is also reported using p -value adjusted for multiple testing. Because of the presence of systematic biases in a typical ChIP-Seq library, it is highly desirable to have a matching control library[17].

PEAK ANNOTATOR[18] is a software that helps us to find the nearest downstream gene (NDG) from the peak that we found by MACS2, HOMER and SICER as well as the transcription start sites (TSS). It has three utility:-

NDG - For each locus, search for its Nearest Downstream Genes on both the forward and reverse strand. If the position of the locus is within a gene, the program describes in which part of that gene the locus is located. It generates the types of file[18].

NDG.TEST;-This file identifies the closest downstream genes for each locus, and contains the following fields[18]:

Chromosome

Start

End - These first three columns describe the genomic location of the peak.

Overlapped Genes - Number of transcripts overlapping the genomic loci. Details about these genes are reported in the second output file described below.

Downstream_FW_Gene- ID of the closest downstream gene on the forward strand.

Symbol- Symbol of the closest downstream gene on the forward strand.

Distance- Peak distance to its closest downstream gene on the forward strand.

Downstream_REV_gene- ID of the closest downstream gene on the reverse strand.

Symbol- Symbol of the closest downstream gene on the reverse strand.

Distance- Peak distance to its closest downstream gene on the reverse strand.

OVERLAP.TEST:- This file describes the transcripts overlapping the peaks, if any such are found[18].

Chromosome

Start

End - These first three columns describe the genomic location of the peak

OverlapGene - Overlapping gene ID

Symbol- Overlapping gene symbol

Overlap_Begin- In which part of the gene does the peak's start position overlap

Overlap_Center- In which part of the gene does the peak's central position overlap

Overlap_End - In which part of the gene does the peak's end position overlap

SUMMARY.TEST:- This file contains the following fields[18]

Chromosome

Start

End - These first three columns describe the genomic location of the peak.

OverlapGene - Overlapping gene Symbol.

Downstream Gene- Nearest downstream gene.

Distance- Peak distance to its nearest downstream gene.

TSS - For each locus, find its closest TSS (transcription start site). In order to do this, the program searches both upstream and downstream for the closest genes to the genomic coordinate. [18]

TSS.TEST:- This file contains the following fields[18]:

Chromosome

Start

End- These first three columns describe the genomic location of the peak.

Distance- The distance from the peak to its closest TSS.

GeneStart - The start location of the closest gene on the genome.

GeneEnd- The end location of the closest gene on the genome.

ClosestTSS_ID- ID of the closest gene.

Symbol- Symbol of the closest gene.

Strand- Strand of closest gene.

ODS - Compare between two position files, to identify overlapping and unique genomic locations.

MEME-ChIP performs comprehensive motif analysis (including motif discovery) on large sets of (typically nucleotide) sequences such as those identified by ChIP-seq or CLIP-seq experiments (sample output from sequences)[19].

Usage:

meme-chip [options] [-db<motif file>]* <primary sequence file>

Input

[-db<motif file>]*

(Optional but recommended) The names of one or more files containing MEME formatted motifs. Outputs from MEME and DREME are supported, as well as Minimal MEME Format. You can convert many other motif formats to MEME Motif format using conversion scripts available with the MEME Suite. These motif file(s) will be used by Tomtom and CentriMo.

<primary sequence file>

The name of a file of sequences in FASTA format. Ideally the sequences should be all the same length, between 100 and 500 base-pairs long and centrally enriched for motifs. The immediate regions around individual ChIP-seq "peaks" from a transcription factor (TF) ChIP-seq experiment are ideal. The suggested 100 base-pair minimum size is based on the typical resolution of ChIP-seq peaks but it is useful to have more of the surrounding sequence to give CentriMo the power to tell if a motif is centrally enriched. We recommend that you "repeat mask" your sequences, replacing repeat regions to the "N" character. [19].

Output

MEME-ChIP writes its output to files in a directory named `memechip_out`, which it creates if necessary[19]. You can change the output directory using the `-o` or `-oc` options. The directory will contain the following files:

`meme-chip.html` - an HTML file that provides the results in an interactive, human-readable format that contains links to the other files produced by the analyses performed by MEME-ChIP

`summary.tsv` - a TSV (tab-separated values) file that provides a summary of the results in a format suitable for parsing by scripts and viewing with Excel

`combined.meme` - a text file that contains all the motifs identified by MEME-ChIP in MEME Motif Format

In addition, the MEME-ChIP output directory will contain sub-directories with the results of each of the individual analyses it performed. The results in these directories are all linked to from the MEME-ChIP HTML output file.

MEME-ChIP outputs a tab-separated values (TSV) file ('summary.tsv') that contains one line for each motif found by MEME-ChIP. The lines are sorted in order of decreasing statistical significance. The first line in the file contains the (tab-separated) names of the fields[20]. Your command line is given at the end of the file in a comment line starting with the character '#' [21]. The names and meanings of the fields in the output are described in the table below.

Table4. The names and meanings of the fields in the output.

(Retrieved from <https://academic.oup.com/nar/article/40/4/e31/2411061>)

Field	Name	Contents
1	MOTIF_INDEX	The index of the motif in the "Motifs in MEME text format" file('combined meme')output by MEME-ChIP.
2	MOTIF_SOURCE	The name of the program that found the <i>de novo</i> motif, or the name of the motif file containing the known motif.
3	MOTIF_ID	The name of the motif, which is unique database file.
4	ALT_ID	An alternate name for the motif, which is unique in the motif database file .
5	CONSENSUS	The ID of the <i>de novo</i> motif, or a consensus sequence computed from the letter frequencies in the known motif (as described above)
6	WIDTH	The width of the motif.
7	SITES	The number of sites reported by the <i>de novo</i> program , or the number of "Total Matches " reported by CentriMo.
8	E-VALUE	The statistical significance of the motif.
9	E-VALUE_SOURCE	The program that reported the <i>E</i> -value
10	MOST_SIMILAR_MOTIF	The known motif most similar to this motif according to Tomtom.
11	URL	A link to a description of the most similar motif, or to the known motif.

A consensus sequence is constructed from each column in a motif's frequency matrix using the "50% rule" as follows:

The letter frequencies in the column are sorted in decreasing order.

Letters with frequency less 50% of the maximum are discarded.

The letter used in this position in the consensus sequence is determined by the first rule below that applies:

If there is only one letter left, or if the remaining letters exactly match an ambiguous symbol in the alphabet, the letter or ambiguous symbol, respectively, is used.

Otherwise, if the remaining set contains at least 50% of the core symbols in the alphabet, the alphabet's wildcard (e.g., "N" for DNA or RNA, and "X" for protein) is used.

Otherwise, the letter with the maximum frequency is used.

Note: The input sequences should be centered on a 100 character region expected to contain motifs. (<http://meme-suite.org/tools/meme-chip>)

MEME-ChIP Combined Motifs Format (text) Format

MEME-ChIP outputs a text file ('combined.meme') containing all the significant motifs found by MEME-ChIP. The motifs are in Minimal MEME Motif format, and their IDs correspond to the motif indices given in the "Summary in TSV Format" file ('summary.tsv') [20].

Note: The 'nsites=' and 'E=' fields in the motif headers are only relevant for the MEME and DREME motifs. For known motifs, those values do not refer to the number of sites in the input sequences.

Conclusion

Chip-Seq is a widely applicable method which is useful for tracking various epigenetic modifications, the type of signal is diverse, and makes the detection of differential regions, a challenging task. Thus various tools are required to interpret its result. So here we have put down all the important tools that are required to obtain analysis of transcription factor-binding sites, histone modifications, etc., along with protein binding site model building tools which will make accession and usage of tools very easy.

Reference

- [1] Visel, A., Blow, M. J., Li, Z., Zhang, T., Akiyama, J. A., Holt, A., & Afzal, V. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature*, 457(7231), 854, 2009.
- [2] Bailey, T. L. DREME: motif discovery in transcription factor ChIP-seq data. *Bioinformatics*, 27(12), 1653-1659, 2011.
- [3] Schmidl, C., Rendeiro, A. F., Sheffield, N. C., & Bock, C. ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. *Nature methods*, 12(10), 963, 2015.
- [4] Barta, E. Command line analysis of ChIP-seq results. *EMBnet. journal*, 17(1), 13-17, 2011.
- [5] Kaboord, B., & Perr, M. Isolation of proteins and protein complexes by immunoprecipitation. In 2D PAGE: sample preparation and fractionation (pp. 349-364). Humana Press 2008.
- [6] Valouev, A., Johnson, D. S., Sundquist, A., Medina, C., Anton, E., Batzoglou, S., ...& Sidow, A. Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. *Nature methods*, 5(9), 829, 2008.
- [7] Whiteford, N., Skelly, T., Curtis, C., Ritchie, M. E., Löhr, A., Zaranek, A. W., ...& Brown, C. Swift: primary data analysis for the IlluminaSolexa sequencing platform. *Bioinformatics*, 25(17), 2194-2199 2009.
- [8] Niklas, N., Pröll, J., Danzer, M., Stabentheiner, S., Hofer, K., & Gabriel, C. Routine performance and errors of 454 HLA exon sequencing in diagnostics. *BMC bioinformatics*, 14(1), 176, 2013.
- [9] Ondov, B. D., Varadarajan, A., Passalacqua, K. D., & Bergman, N. H. Efficient mapping of Applied Biosystems SOLiD sequence data to a reference genome for functional genomic applications. *Bioinformatics*, 24(23), 2776-2777, 2008.

- [10] Brown, J., Pirrung, M., & McCue, L. A. FQC Dashboard: integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. *Bioinformatics*, 33(19), 3137-3139, (2017)..
- [11] Langmead, B. Aligning short sequencing reads with Bowtie. *Current protocols in bioinformatics*, 32(1), 11-7,2010.
- [12] Diaz, A., Park, K., Lim, D. A., & Song, J. S. Normalization, bias correction, and peak calling for ChIP-seq. *Statistical applications in genetics and molecular biology*, 11(3), 2012.
- [13] Liu, T. Use model-based analysis of ChIP-Seq (MACS) to analyze short reads generated by sequencing protein–DNA interactions in embryonic stem cells. In *Stem Cell Transcriptional Networks* (pp. 81-95). Humana Press, New York, NY 2014.
- [14] Leggett, R. M. Ramirez- Gonzalez RH, Clavijo BJ, et al. Sequencing quality assessment tools to enable data- driven informatics for high throughput genomics, 4, 288, 2013.
- [15] De la Rosa, J. V., Ramón-Vázquez, A., Tabraue, C., &Castrillo, A. Analysis of LXR nuclear receptor cistrome through ChIP-seq data bioinformatics. In *Lipid-Activated Nuclear Receptors* (pp. 99-109). Humana Press, New York, NY, 2019.
- [16] Homer.(n.d.)Retrieved from: <http://homer.ucsd.edu/>)
- [17] Xu, S., Grullon, S., Ge, K., &Peng, W. Spatial clustering for identification of ChIP-enriched regions (SICER) to map regions of histone methylation patterns in embryonic stem cells. In *Stem Cell Transcriptional Networks* (pp. 97-111). Humana Press, New York, NY2014.
- [18] Zhu, L. J., Gazin, C., Lawson, N. D., Pagès, H., Lin, S. M., Lapointe, D. S., & Green, M. R.ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC bioinformatics*, 11(1), 237. 2010
- [19] Machanick, P., & Bailey, T. L. MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics*, 27(12), 1696-1697,2011.
- [20] Ma, W., Noble, W. S., & Bailey, T. L. Motif-based analysis of large nucleotide data sets using MEME-ChIP. *Nature protocols*, 9(6), 1428, (2014)..
- [21] Thomas-Chollier, M., Herrmann, C., Defrance, M., Sand, O., Thieffry, D., & van Helden, J. (2011). RSAT peak-motifs: motif analysis in full-size ChIP-seq datasets. *Nucleic acids research*, 40(4), e31-e31.
- [22] Gibiansky, A. L., Haque, I. S., Maguire, J. R., & Robertson, A. D. J. (2016). U.S. Patent Application No. 14/884,656.