IJSRM
INTERNATIONAL JOURNAL OF SCIENCE AND RESEARCH METHODOLOGY
An Official Publication of Human Journals

Human Journals
Research Article
September 2016 Vol.:4, Issue:3

CGPSA- A Methodology for the Comparative Genomics for Phylogenetic and Synteny Analysis of Genes and Gene-families with Available Sequence

IJSRM
INTERNATIONAL JOURNAL OF SCIENCE AND RESEARCH METHODOLOGY
An Official Publication of Human Journals

Hukam C. Rawal and T.R. Sharma

ICAR-National Research Centre on Plant Biotechnology, New Delhi, INDIA.

Submission: 10 September 2016
Accepted: 15 September 2016
Published: 25 September 2016

Keywords: Comparative genomics, methodology, synteny, phylogeny, representation

ABSTRACT

With the emergence of new NGS technologies, the genomic research is running at gigabyte speed with newly published completely sequenced genomes every day or once in a week. But what to do with the sequenced genomes? How to proceed further with the available genes or genomic sequences? To compete and utilize this explosion in the number of completely sequenced genomes for the mankind there is an urgent need of a well-established computational methodology for the analysis of different genes or gene families in these newly sequenced genomes to investigate evolutionary or syntenic relations, if any and correlate for any conclusions. While dealing with a comparison of different gene families among various plant species with complete genomes, we developed a computational strategy will be known as CGPSA- "A Methodolgy for the Comparative Genomics for Phylogenetic and Synteny Analysis" which can be used for any organism with available genome or genes. CGPSA pipeline comprised of 5 main steps including In-silico identification of genes, In-silico expression analysis, physical mapping, phylogenetic and syntenic analysis of the identified genes. Using CGPSA methodology we performed the comparativeanalysis of the seven defense response gene families such as Beta-glucanase and Chitinase in 23 plant species representing two major classes- dicots and monocots along with very diverse groups like legumes, fruits, vegetables, spices, flowers, medicinal and aromatic plants , grains, and trees. With this strategy, a total of 6864 defense response genes were identified and analysed comparatively in the 23 plant species. The developed strategy for comparative genomics of any gene or gene family with computational methods can be used effectively across the genomes.

OVERVIEW

Over the past decade, we have seen an explosion in the number of completely sequenced genomes.It brought us to the urgency to develop a well established computational strategy or methodology for comparative analysis. We developed a methodology named as CGPSA - 'Comparative Genomics for Phylogenetic and Synteny Analysis', to analysegenes or gene families in different species with complete genomes to conclude any phylogenetic and syntenic relations between these genes and hence between the investigated genomes.Using CGPSA methodology basedstrategy, a total of 6864 genes were identified for seven defense response gene families such as Beta-glucanase and Chitinaseand analyzed comparatively in the 23 plant species (Table 1).

Table 1. Data analyzed by CGPSA Methodology.

Species analysed	23 Plant species
Gene families analysed	7 Defense Response gene families
Genes identified and analysed	6864 Genes
Phylogenetic analysis	7 families and 6864 genes

INTRODUCTION

The advancement in the scientific genomic research brought us with the number of completely sequenced genomes in short time. It gave birth to the comparative genomics which includes the study of different genes and genomes to find any evolutionary or syntenic relations between them. In future, we have to comparatively analyse different sets of genes and genomes, so a well established computational strategy or methodology for the same would be quite helpful. With our needs, we developed a methodology for comparative study of defense response gene families in completed plant genomes which includes very diverse groups like legumes, fruits, vegetables, spices, flowers, medicinal and aromatic plants , grains, and trees.This methodology, CGPSA can be used effectively across any genome for any gene or gene family.

MATERIALS AND METHODS

Once the gene family to be analyse is known, we can start with any known gene of that family present in a well-known or model genome(s) related to the genomes of analysis. This

known gene can be taken as query gene and CGPSA methodology can be used for further analysis (Fig.1). This methodology is comprised of 5 main steps including In-silico identification of genes using BLAST search, In-silico expression analysis against EST database, physical mapping to show the positions of genes in the genomes, phylogenetic and syntenic analysis of the identified genes and their orthologs. BLAST based strategy was used to search and identify genes of interest (Altschul *et al.* (1990)) and their physical positions were represented by using MapChart tool (Voorrips (2002). The identified genes were aligned and compared for evolutionary relations by generating phylogenetic trees using ClustalX and ClustalW (Larkin *et al.* (2007)). Best-bidirectional hit based strategy was used to detect orthologs (Hulsen *et al.* (2006); Rawal et al. (2013)). Softwares like iTOL (Letunic& Bork (2006)) and Circos (Martin *et al*. (2009))were used to interactively show the phylogenetic and syntenic relations, respectively.

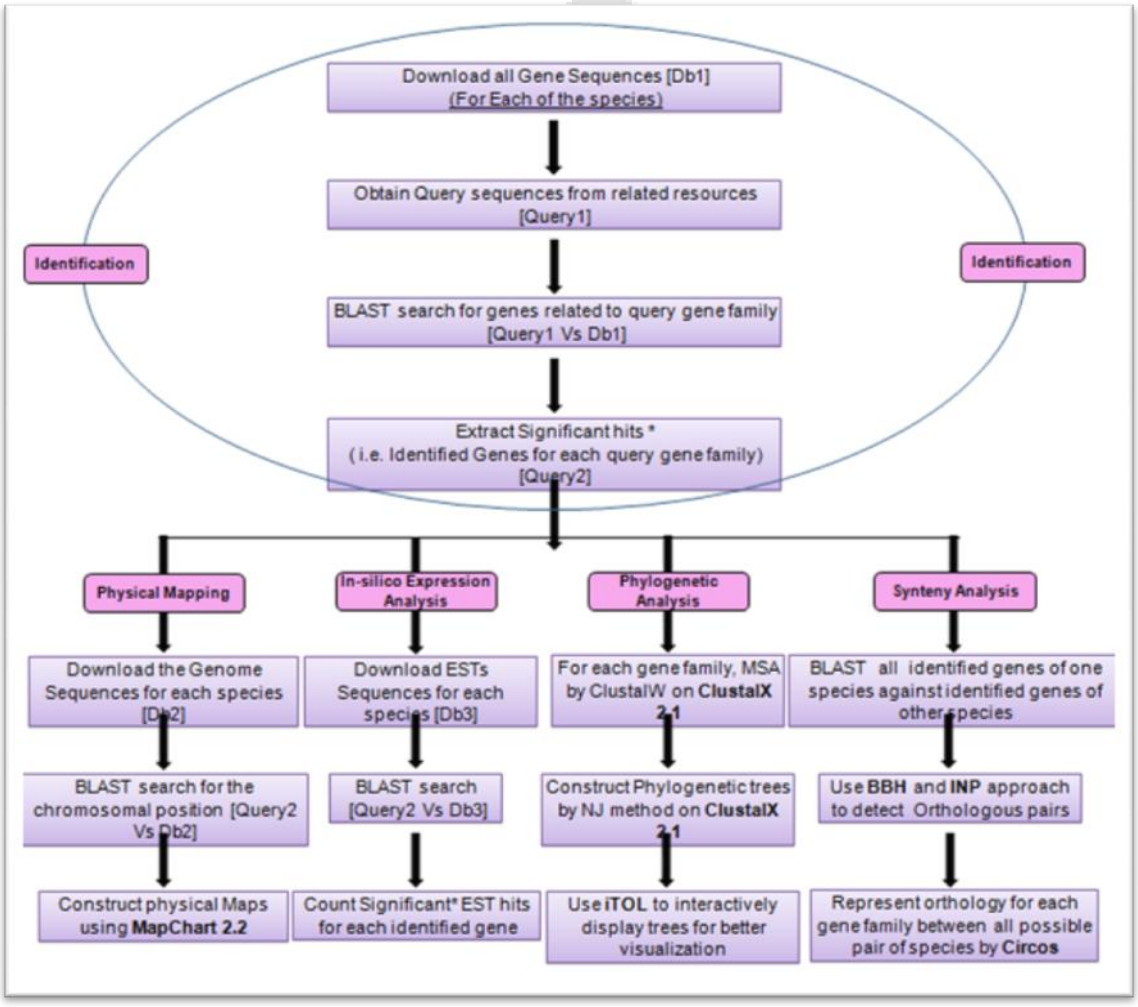

Figure 1. CGPSA Methodology for comparative genomics. *BLAST hits with bit score >= 100 and E- Value =< e^{-20}

RESULTS

Using CGPSA methodology we identified:

→6864 defense response genes, representing 7 gene families, in 23 plant species, contributing even more than 1 % of total gene contents in some species (Table 1).

→*In-silico* physical mapping reveal some gene families as telomere or subtelomeric specific (Fig. 2).

→In phylogeny analysis, some gene families found to form species-specific groups/sub-groups (Fig. 3).

→Synteny analysis brought us with orthology percent for each gene family between each possible pair of species and also the number of non-orthologous genes in each gene family (Fig. 4).

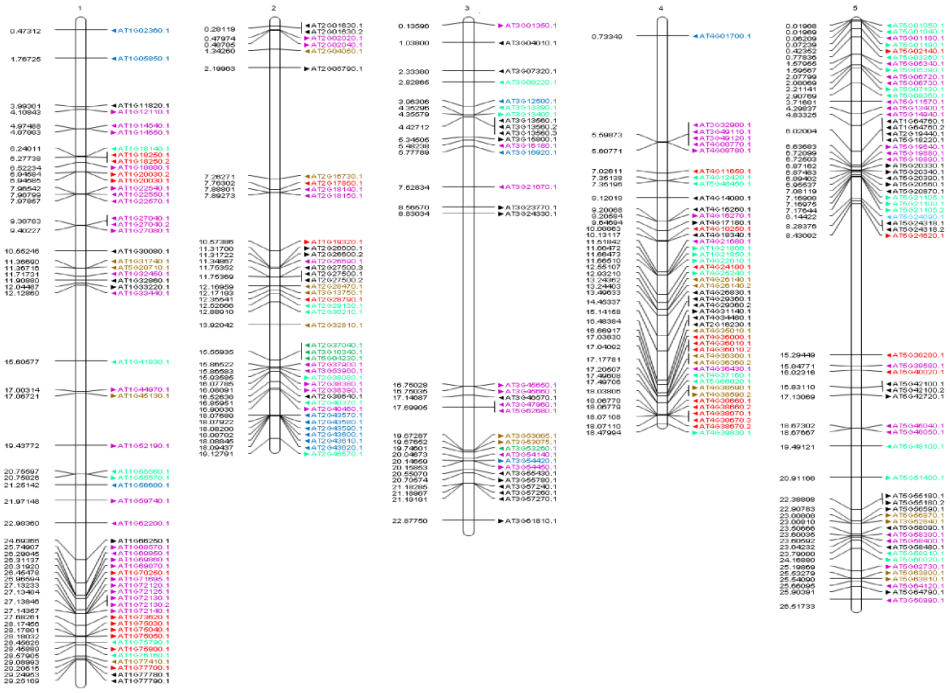

Figure 2. Physical Mapping using MapChart

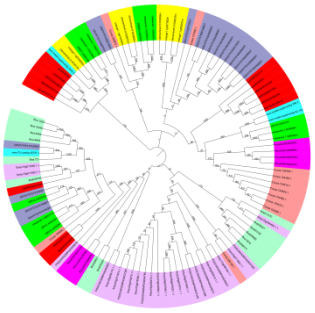

Figure 3. Interactive Phylogenetic tree by iTOL for a gene family with different color for each species

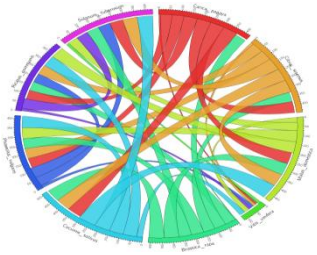

Figure 4. Synetney analysis of a gene family showing orthology percent between different species

CONCLUSIONS

CGPSA methodology brought us with all the important analysis steps needed for comparative genomics analysis of gene families in different genomes. The included publicly available softwares not only makes it useful for everyone but also work well to represent large data in better way. Once well established and automated, it will definitely fasten the speed of comparative genomics, especially when we are sequencing genomes rapidly these days.

ACKNOWLEDGEMENT

Financial assistance received from ICAR under NPTC project is thankfully acknowledged. We thank each and every researcher for making the sequence data of different species available in public domain.

REFERENCES

1. Altschul*et al*. (1990). "Basic local alignment search tool." *J. Mol. Biol*. 215:403-410.
2. Voorrips, R.E., (2002). MapChart: Software for the graphical presentation of linkage maps and QTLs. *The Journal of Heredity* 93 (1): 77-78.
3. Larkin *et al*. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*23:2947–2948.
4. Hulsen *et al*.(2006). Benchmarking ortholog identification methods using functional genomics data. *Genome Biol.* 7: R31. 10.1186/gb-2006-7-4-r31.
5. Rawal HC, Singh NK, Sharma TR. (2013) Conservation, divergence and genome-wide distribution of PAL and POX A gene families in plants. *Int. Journal of Genomics*.doi:/10.1155/2013/678969.
6. Letunic I and Bork P (2006). Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation.*Bioinformatics* **23(1)**:127-8 .
7. Martin *et al*. (2009). Circos: An information aesthetic for comparative genomics. *Genome Res.*, doi:10.1101/gr.092759.109.