Comparative Genomics Study in Bottle Gourd- a Potential Medicinal Vegetable

Manjusha Verma, Bhawna, Lalit Arya
Division of Genomic Resources, NBPGR, Pusa Campus, New Delhi-110012
manjusha@nbpgr.ernet.in

ABSTRACT

In the field of comparative genomics, genomic landmarks viz genes, gene orders, DNA sequences and whole genomes are compared with the genomic features of other organisms to study the basic similarities and differences as well as evolutionary relationships. Here, we have utilized bottle gourd [Lagenaria siceraria (Mol.) Standl.] (2n=2x=22)—a non model vegetable crop, well known for its medical properties cultivated all over the tropics for comparative genomics study with cucumber, musk melon and watermelon. SSR marker characterization with recently available NGS data of bottle gourd identified 44,823 perfect microsatellite repeat-motifs spanning ~334 Mb of bottle gourd genome. These markers have been mapped on to linkage groups for bottle gourd and their synteny with cucumber, musk melon and watermelon has also been compared. In addition to implication of these numerous SSR markers in facilitating the genetic improvement of bottle gourd, the marker synteny may well lead to understanding the subtle differences among cucurbits.

Keywords: comparative genomics, bottle gourd, cucurbits

1. INTRODUCTION

The advent of next generation sequencing methodologies has provided a window to inundate genomic information for the non-model crops, in addition to completed reference genomes of the model crops. This deluge of genomics data permits an array of computational analyses to decipher crop domestication, evolution and aid in the realization crop improvement. In the field of comparative genomics, genomic landmarks viz genes, gene orders, DNA sequences and whole genomes are compared with the genomic features of other organisms to study the basic similarities and differences as well as evolutionary relationships. These emerging approaches are being used to compare individual plant genomes to get the essence of evolutionary relationships, phylogenetic trees, genetic mapping, gene function and regulation. Molecular markers have been surveyed across crop genomes for more than half a century now, but the density of these markers has augmented exponentially by employing the next generation sequencing (NGS) technologies.

Rapidly increasing reference genomes, re-sequencing of model crops and draft sequences of potential crops allot the assessment of single nucleotide polymorphisms, indels and copy number variants through new computational approaches. The burgeoning human population presses an urgent need to apply these state of the art technologies for enhancing crop production and improvement.

Bottle gourd, [*Lagenaria siceraria* (Mol.) Standl.] is a cucurbit vegetable crop with a worldwide presence. It has multitude of uses, including medicinal, ornamental and a cooked nutritious vegetable. Despite its uses, it remains a potential crop with meager genomics resources. In this chapter, we brief the current crop comparative genomics, and present the outcome of the experiments with the bottle gourd NGS data.

2. CROP COMPARATIVE GENOMICS—CHALLENGES AND PROSPECTS

Polyploidy, copious multi-gene families and widespread copy number variations complicate comparative genomics studies in even the model crops with good reference (Lockton and Gaut 2005, Haun et al 2011, Velasco et al 2007). Higher nucleotide diversity in some out-crossed crops, eg. maize (Rafalski et al 2004) or even self-fertilized crops, eg. barley (Caldwell et al 2006) and clonally propagated crops, eg. grapes (Lijavetzkyet al 2007) presents another challenge in many crop species comparative analyses. The genome sizes of some other crops, eg wheat, further sets hurdles even for the NGS sequencing, in addition to hindering the progress of comparative genomics studies. The interpretation of crop comparative genomics is also impeded by the duplications and the subsequent gene loss, rearrangements, tandem gene or segmental duplications, and divergence of duplicated gene sequences.

The use of species specific prototypes of linkage disequilibrium, sequencing diploid crop relatives and double haploid lines may reduce some the complex issues pertaining to crop comparative genomics. Despite the plentiful complexities, crop plants also offer certain advantages— long term seed storage, availability of inbred and double haploid lines, regeneration of the lines at different environmental locales permitting replications of experiments.

3. APPLICATIONS

Crop comparative studies authorize understanding about the duplications in certain plant genomes, for example, soyabean was identified as an ancient polyploid with putative homoeologous chromosomal regions through genetic mapping (Lee et al 2001). Wet lab experiments with molecular markers and *in-silico* homologous sequence mapping among related or divergent species reveals macrosynteny (conserved gene order between species). Microsynteny (conserved gene order at sequence level over a short DNA contig) reveals genome structure at specific regions offering insights into tandem duplication events occurring independently after the divergence of two species (Choi et al., 2004b). Crop comparisons have legitimated cross-species gene predictions and isolations. Map based gene cloning utilizing conserved gene architecture has aided numerous gene cloning. Crop comparisons may provide evidence in the future to critical questions about genome functions in relation to genome structure.

4. GENOME WIDE SSR DEVELOPMENT IN BOTTLE GOURD

Genome wide marker development was done utilizing restriction site digestion of the target genome, bottle gourd for SSR marker survey and characterization. Microsatellites are preferred genetic markers due to their high level of polymorphism, co dominant character and ease of use. In view of this, we conducted a genome wide analysis using recently available restriction site-associated DNA sequencing (RAD-Seq) scaffolds/contigs sequences data and identified 45,066 perfect microsatellite repeat-motifs spanning ~334 Mb of bottle gourd genome. Tetranucleotide repeats (34.3%) were the most prevalent, followed by trinucleotide repeats (30.73%), followed by di, penta and hexa nucleotide repeats at 21.03%, 9.6% and 4.3%, respectively in the bottle gourd genome. Of the total microsatellites, about 103 (~0.2%) markers were successfully validated in 5 accessions of *Lagenaria siceraria*.

5. BOTTLE GOURD GENETIC LINKAGE MAP AND SYNTENY WITH OTHER CUCURBITS

The physical location of the SSR markers was mapped on to the eleven bottle gourd linkage groups using the flanking genomic sequences of identified SSR motifs against the whole genome sequence of bottle gourd (http://www.ncbi.nlm.nih.gov/nuccore/542029074). The markers were located for all of the eleven bottle gourd linkage group based on their ascending order, physical position (bp), from the start position to end position of linkage group. The genome sequences of other cucurbit species like cucumber (Huang et al., 2009), melon (Garcia-Mas et al., 2012) and watermelon (Guo et al., 2013) were compared with the SSR markers to derive syntenic relationships with the linkage group of bottle gourd. In total, 207 SSRs were mapped on to the bottle gourd genome. Out of these, 193 mapped to the cucumber, 188 to the watermelon and 181 to the musk melon genome. Overall bottle gourd revealed 93.2% synteny with the 7 chromosomes of cucumber, 87.4% synteny with the 12 chromosomes of melon and 90.8% synteny with the 11 watermelon chromosomes.

6. CONCLUSION

The progress of genomic resource generation in model and non-model crop species will contribute to understanding genome function, organization and mechanism of critical gene functions. This structural and functional information emerging from various crops may then render support for crop improvement. The immense applicability of this study in germplasm characterization using the SSR markers, cucurbit comparative genomics and gene discovery in cucurbit species can be well anticipated.

REFERENCES

[1] Caldwell, K. S., Russell, J., Langridge, P. & Powell, W. (2006) Extreme population-dependent linkage disequilibrium detected in an inbreeding plant species, *Hordeum vulgare*. *Genetics* **172**, 557–567.

[2] Choi HK, Mun JH, Kim DJ, Zhu H, Baek JM, Mudge J, Roe B, Ellis N, Doyle J, KissGB, et al (2004b) Estimating genome conservation between crop and model legume species. Proc Natl Acad Sci USA 101: 15289– 15294.

[3] Garcia-Mas et al. (2012) The genome of melon (Cucumis melo L.) Proc Natl Acad Sci U S A. 109(29): 11872–11877.

[4] Haun, W. J. *et al.* (2011) The composition and origins of genomic variation among individuals of the soybean reference cultivar Williams 82. *Plant Physiol.* **155**, 645–655.

[5] Huang et al (2009) The genome of the cucumber, *Cucumis sativus* L. *Nature Genetics* **41**, 1275 – 1281.

[6] Lee JM, Grant D, Vallejos CE, Shoemaker RC (2001) Genome organization in dicots. II. Arabidopsis as a bridging species to resolve genome duplication events among legumes. Theor Appl Genet 103: 765–773.

[7] Lijavetzky, D., Cabezas, J. A., Ibanez, A., Rodriguez, V. & Martinez-Zapater, J. M. (2007) High throughput SNP discovery and genotyping in grapevine (*Vitis vinifera* L.) by combining a re-sequencing approach and SNPlex technology. *BMC Genomics* **8**, 424.

[8] Lockton, S. & Gaut, B. S. (2005) Plant conserved non-coding sequences and paralogue evolution. *Trends Genet.* **21**, 60–65.

[9] Rafalski, A. & Morgante, M. (2004) Corn and humans: recombination and linkage disequilibrium in two genomes of similar size. *Trends Genet.* **20**, 103–111.

[10] Shaogui Guo et al (2013) The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions. Nature Genetics 45, 51–58.

[11] Velasco, R. *et al.* (2007) A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. *PLoS ONE* **2**, e1326.

[12]