ECRONICON OPEN ACCESS

EC AGRICULTURE
Research Article

Core Genome of *Poales*, An Economically Important Order of Monocotyledons

Zhi Jue Kuan and Maurice HT Ling*

School of Applied Sciences, Temasek Polytechnic, Singapore

***Corresponding Author:** Maurice HT Ling, School of Applied Sciences, Temasek Polytechnic and HOHY PTE LTD, Singapore.

Received: January 02, 2021; **Published:** January 30, 2021

Abstract

The importance of *Poales* species; which includes rice, wheat, and maize; has led to various studies on its tolerance and evolution. Evolutionary studies are largely dependent on the presence of orthologs. A recent study suggests that the complete set of orthologs is required to reflect actual evolutionary history; thereby, underpinning the need to identify the core genome of *Poales* representing the set of orthologs across *Poales* species. Here, we identified a 6,122 gene core genome of *Poales* and functional analysis suggests that a strong role of interspecies interactions within *Poales* core genome.

Keywords: *Core Genome; Poales; Rice; Maize; Pineapple; Sorghum; Millet*

Introduction

The taxonomic order of *Poales* comprises of monocotyledon plants with total economic value of nearly USD 1 trillion [1] as it consists of staple crops such as rice, millet, sorghum, wheat and maize. Its implications on social and politics, in event of food shortage, cannot be overemphasized. This leads to the study of parasite tolerance [2], environmental tolerance [3-5], colonization [6] and evolution [7] of *Poales* species.

Phylogeny is an important tool to study the evolution of species [8-11] and its success is largely dependent on the presence of orthologs (which may be genes or peptides) across the species in question [12-14]. A recent study suggests that phylogenetic analysis requires the complete set of orthologs as phylogeny from single ortholog or multiple single orthologs is not likely to reflect actual evolutionary history [15]. This suggests that the core genome, which is the set of orthologous genes within a set of related genomes [16], which may be from different strains of a species [17] or different species of a genus [18]. In this case, a phylogenetic analysis of *Poales* will require the core genome of *Poales*, representing orthologous genes in various *Poales* genus and species.

However, the core genome of *Poales* has not been identified. In this study, we identified a 6,122 gene core genome of *Poales* from all 10 available RefSeq genome assemblies, represents between 7.89% and 20.72% of *Poales*' genome. Functional analysis of the core genome using its mapping to *Zea mays* (maize) suggests that a strong role of interspecies interactions within *Poales* core genome.

Materials and Methods

Sequences: RNA sequences from all 10 available RefSeq genome assemblies (as of 20 November 2020) from Order *Poales* (NCBI:txid38820) were obtained from NCBI; namely, (a) *Aegilops tauschii* AL8/78 (P1; NCBI RefSeq Assembly Accession GCF_001957025.1; hereinafter,

known as Accession), (b) *Ananas comosus* F153 (P2; Accession GCF_001540865.1), (c) *Brachypodium distachyon* Bd21 (P3; Accession GCF_000005505.3), (d) *Oryza brachyantha* (P4; Accession GCF_000231095.1), (e) *Oryza sativa Japonica Group* Nipponbare (P5; Accession GCF_001433935.1), (f) *Panicum hallii* FIL2 (P6; Accession GCF_002211085.1), (g) *Setaria italica* Yugu1 (P7; Accession GCF_000263155.2), (h) *Setaria viridis* A10 (P8; Accession GCF_005286985.1), (i) *Sorghum bicolor* BTx623 (P9; Accession GCF_000003195.3), and (j) *Zea mays* B37 (P10; Accession GCF_902167145.1).

Determining core genome by intersecting genomes: The procedure of identifying core genome of Order *Poales* by genome intersection, which was based on that of previous study [19] using NCBI BLAST [20] version 2.11.0. Briefly, the intersection of *A. tauschii* AL8/78 (P1) and *A. comosus* F153 (P2) were determined by constructing a BLAST database out of the RNA sequences of P2 and the RNA sequences of P1 were used as query. The RNA sequences of P1 lesser than the E-value threshold (average E-value using a random set of 9 pairwise BLAST comparisons, representing 20% of the total combinatorial pairwise BLASTs) when blast with P2 were the genome intersection representing the core genome between P1 and P2; thereby, denoted as P1P2 and extracted from P1 using SeqProperties [21]. This process was repeated until all 10 genomes were intersected, which represented the core genome and was denoted as P1P2P3P4P5P6P7P8P9P10.

Functional classification of core genome: The identified core genome was functionally classified into molecular functions and biological processes with mapped *Z. mays* transcriptome (P10) using PANTHER [22] (http://pantherdb.org/).

Results and Discussion

The number of RNA sequences ranges from to 40,869 in *S. italica* to 76,669 in *Z. mays* (Table 1). Using 9 out of 45 (20%) possible pairwise comparisons, our results suggest that the average E-values range from 8.97E-08 to 7.08E-05, with average percent identities range from 81.26% to 94.01% (Table 2). The grand mean of E-value is 6.89E-06, which is used as E-value threshold for this study. This is consistent with that of recent studies on core genomics such as Costa., *et al.* [23] and Barajas., *et al*. [16] whom use E-values of 1E-05 and 1E-06 as thresholds respectively. Guimaraes., *et al*. [24] review that PanFunPro [25] uses E-value less than 1E-03 to create functional profiles and protein grouping. The grand mean percent identity of 87.56% is not used as threshold as it is substantially higher than that of Costa., *et al*. [23] whom use 50% identity with 60% coverage.

ID	Organism	Strain/Cultivar	Common Name	Number of RNA Sequence
P1	*A. tauschii* AL8/78	AL8/78	--	69,124
P2	*A. comosus* F153	F153	Pineapple	42,940
P3	*B. distachyon* Bd21	Bd21	Stiff Brome	47,462
P4	*O. brachyantha*	--	Malo Sina	29,549
P5	*O. sativa Japonica Group*	Nipponbare	Japanese Rice	53,404
P6	*P. hallii* FIL2	FIL2	--	43,787
P7	*S. italica* Yugu1	Yugu1	Foxtail Millet	40,869
P8	*S. viridis* A10	A10	--	46,950
P9	*S. bicolor*	BTx623	Sorghum	48,195
P10	*Z. mays*	B73	Maize	76,669

***Table 1**: Number of RNA sequences in each organism.*

Comparison	Count	E-value		% Identity	
		Average	Standard Deviation	Average	Standard Deviation
P1/P6	93,081	9.41E-08	1.42E-06	82.66	5.17
P3/P1	129,350	7.08E-05	1.47E-02	84.70	5.68
P3/P8	101,647	1.08E-07	1.92E-06	83.87	5.94
P7/P3	82,005	9.83E-08	1.56E-06	82.72	5.11
P7/P8	344,427	8.97E-08	1.85E-06	94.01	7.16
P9/P7	121,468	1.36E-07	1.70E-06	85.44	5.68
P10/P2	35,162	6.49E-07	9.10E-06	81.26	6.30
P10/P3	235,754	5.92E-08	1.40E-06	88.27	7.37
P10/P7	205,545	1.09E-07	2.54E-06	86.04	5.49
Total	1,348,439	6.89E-06	4.54E-03	87.56	7.59

Table 2: *E-values and % identity in 9 random comparisons.*

Using the threshold of E-value less than 6.89E-06, a 6,122 gene core genome of Order *Poales* is identified from 10 available RefSeq genome assemblies (Table 3). This represents between 7.89% (using *Z. mays* as reference) and 20.72% (using *O. brachyantha* as reference) of *Poales*' genome, which is significantly larger (Chi-Square = 168.57, df = 1, p-value = 1.52E-38) than the core genome of prokaryotes [19]. This may be indicative of fundamental differences between prokaryotes and eukaryotes. Our results show higher percent identity among the intersected transcripts (Table 4) as compared to the threshold proposed by Costa., *et al*. [23] suggesting that using E-value threshold is sufficient in this case.

RNA Sequence Set	Number of RNA	Percentage
P1	69,124	100.00%
P2	42,940	62.12%
P1P2	7,743	11.20%
P1P2P3	7,478	10.82%
P1P2P3P4	6,855	9.92%
P1P2P3P4P5	6,727	9.73%
P1P2P3P4P5P6	6,546	9.47%
P1P2P3P4P5P6P7	6,439	9.32%
P1P2P3P4P5P6P7P8	6,426	9.30%
P1P2P3P4P5P6P7P8P9	6,290	9.10%
P1P2P3P4P5P6P7P8P9P10	6,122	8.86%

Table 3: *Progressive reduction of number of RNA sequences.*

	E-value		% Identity	
RNA Sequence Set	Average for all E-values (a)	Average for E-value < 6.89E-06 (b)	Average for all E-values (a)	Average for E-value < 6.89E-06 (b)
P1P2	3.54E-07	5.17E-08	81.18	80.57
P1P2P3	9.57E-08	2.08E-08	85.06	85.00
P1P2P3P4	8.74E-08	2.56E-08	83.81	83.75
P1P2P3P4P5	1.53E-07	1.87E-08	83.60	83.50
P1P2P3P4P5P6	1.09E-07	1.69E-08	83.42	83.36
P1P2P3P4P5P6P7	6.50E-08	1.49E-08	83.47	83.43
P1P2P3P4P5P6P7P8	8.81E-08	2.14E-08	83.59	83.54
P1P2P3P4P5P6P7P8P9	1.16E-07	1.86E-08	83.49	83.43
P1P2P3P4P5P6P7P8P9P10	2.65E-07	1.17E-08	83.59	83.52

Table 4: *E-value and % identity of intersections. (a) refers to average of all the values from BLAST output without filtering. (b) refer to the average of all the values from BLAST output after filtering for E-value of less than 6.89E-06 (the E-value threshold in this study).*

The 6,122 gene core genome maps to 14,524 transcripts in *Z. mays*, representing 18.95% of *Z. mays*' transcriptome. Of which, 1,884 transcripts are mapped to 8 molecular functions (Figure 1); namely, (i) catalytic activity (GO:0003824), (ii) binding (GO:0005488), (iii) molecular function regulator (GO:0098772), (iv) transporter activity (GO:0005215), (v) structural molecule activity (GO:0005198), (vi) molecular transducer activity (GO:0060089), (vii) translation regulator activity (GO:0045182) and (viii) molecular adaptor activity (GO:0060090).

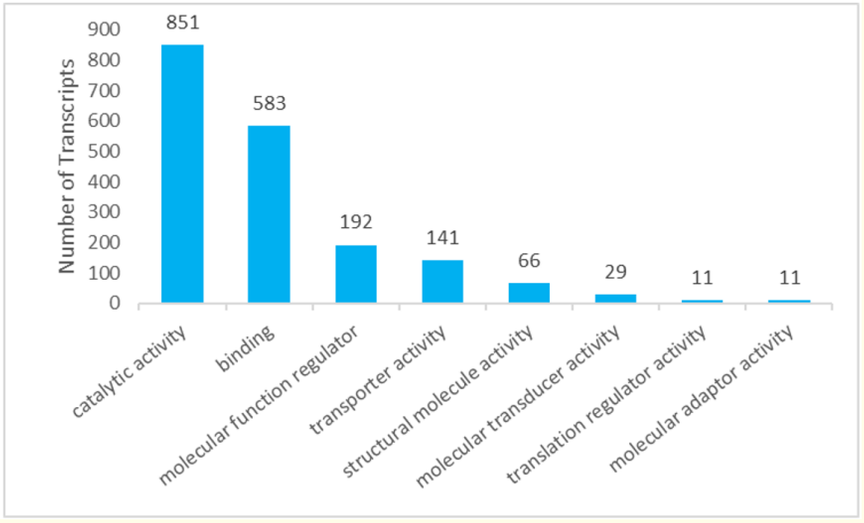

Figure 1: *Molecular function classification of core genome.*

2,937 transcripts are mapped to 15 biological processes; of which, the top 8 biological processes (Figure 2) are (i) cellular process (GO:0009987), (ii) metabolic process (GO:0008152), (iii) biological regulation (GO:0065007), (iv) response to stimulus (GO:0050896), (v) localization (GO:0051179), (vi) signaling (GO:0023052), (vii) developmental process (GO:0032502), and (viii) interspecies interaction between organisms (GO:0044419). Seven of the 8 biological processes identified are common and fundamental biological processes expected of plant physiology. Only interspecies interaction (GO:0044419) is notable, which highlights the importance of plant-plant [26-28], plant-animal [29] and multispecies [30] interactions in the agroecosystem.

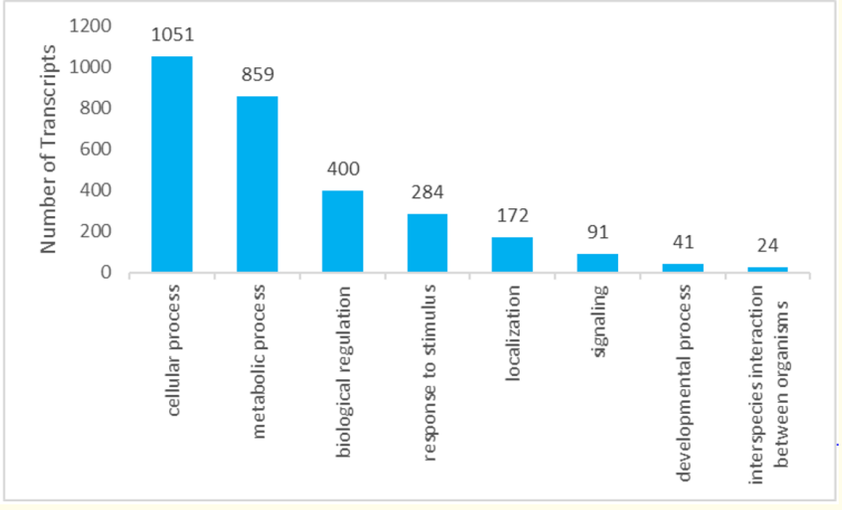

Figure 2: *Top 8 biological process classification of core genome.*

Conclusion

The core genome of Order *Poales* consisting 6,122 gene is identified.

Data Availability

The data files for this study can be downloaded at https://bit.ly/Core*Poales*Genome.

Conflict of Interest

The authors declare no conflict of interest.

Bibliography

1. Vallée GC., *et al.* "Economic Importance, Taxonomic Representation and Scientific Priority as Drivers of Genome Sequencing Projects". *BMC Genomics* 17.10 (2016): 782.

2. Higginbotham RW., *et al.* "Tolerance of Wheat (Poales: Poaceae) Seedlings to Wireworm (Coleoptera: Elateridae)". *Journal of Economic Entomology* 107.2 (2014): 833-837.

3. Ganie SA., *et al.* "Advances in Understanding Salt Tolerance in Rice". *Theoretical and Applied Genetics* 132.4 (2019): 851-870.

4. Wang X., *et al.* "Genetic Variation in ZmVPP1 Contributes to Drought Tolerance in Maize Seedlings". *Nature Genetics* 48.10 (2016): 1233-1241.

5. Mammadov J., *et al.* "Wild Relatives of Maize, Rice, Cotton, and Soybean: Treasure Troves for Tolerance to Biotic and Abiotic Stresses". *Frontiers in Plant Science* 9 (2018): 886.

6. Linder HP., *et al.* "Global Grass (Poaceae) Success Underpinned by Traits Facilitating Colonization, Persistence and Habitat Transformation". *Biological reviews of the Cambridge Philosophical Society* 93.2 (2018): 1125-1144.

7. Darshetkar AM., *et al.* "Understanding Evolution in Poales: Insights from Eriocaulaceae Plastome". *PLOS ONE* 14.8 (2019): e0221423.

8. Adamek M., *et al.* "Applied Evolution: Phylogeny-Based Approaches in Natural Products Research". *Natural Product Report* 36.9 (2019): 1295-1312.

9. Yermanos AD., *et al.* "Tracing Antibody Repertoire Evolution by Systems Phylogeny". *Frontiers in Immunology* 9 (2018): 2149.

10. Mitter C., *et al.* "Phylogeny and Evolution of Lepidoptera". *Annual Review of Entomology* 62 (2017): 265-283.

11. Teixidor-Toneu I., *et al.* "Comparative Phylogenetic Methods and the Cultural Evolution of Medicinal Plant Use". *Nature Plants* 4.10 (2018): 754-761.

12. Sawa G., *et al.* "Current Approaches to Whole Genome Phylogenetic Analysis". *Briefings in Bioinformatics* 4.1 (2003): 63-74.

13. Liu C., *et al.* "Phylogenetic Clustering of Genes Reveals Shared Evolutionary Trajectories and Putative Gene Functions". *Genome Biology and Evolution* 10.9 (2018): 2255-2265.

14. Coimbra NDR., *et al.* "Reconstructing the Phylogeny of *Corynebacteriales* while Accounting for Horizontal Gene Transfer". *Genome Biology and Evolution* 12.4 (2020): 381-395.

15. Wang VC., *et al.* "A Case Study Using Mitochondrial Genomes of the Order Diprotodontia (Australasian Marsupials) Suggests that Single Ortholog is Not Sufficient for Phylogeny". *EC Clinical and Medical Case Reports* 3.9 (2020): 93-114.

16. Barajas HR., *et al.* "Global Genomic Similarity and Core Genome Sequence Diversity of the Streptococcus Genus as a Toolkit to Identify Closely Related Bacterial Species in Complex Environments". *Peer Journal* 6 (2019): e6233.

17. Goodall ECA., *et al.* "The Essential Genome of Escherichia coli K-12". *mBio* 9.1 (2018): e02096-e02017.

18. Alcaraz LD., *et al.* "Understanding the Evolutionary Relationships and Major Traits of Bacillus through Comparative Genomics". *BMC Genomics* 11 (2010): 332.

19. Tan XT., *et al.* "Core Pseudomonas Genome from 10 Pseudomonas Species". *MOJ Proteomics and Bioinformatics* 9.3 (2020): 68-71.

20. Altschul SF., *et al.* "Basic Local Alignment Search Tool". *Journal of Molecular Biology* 215.3 (1990): 403-410.

21. Ling MHT. "SeqProperties: A Python Command-Line Tool for Basic Sequence Analysis". *Acta Scientific Microbiology* 3.6 (2020): 103-106.

22. Mi H., *et al.* "PANTHER Version 14: More Genomes, A New PANTHER GO-Slim and Improvements in Enrichment Analysis Tools". *Nucleic Acids Research* 47.D1 (2019): D419-D426.

23. Costa SS., *et al.* "First Steps in the Analysis of Prokaryotic Pan-Genomes". *Bioinformatics and Biology Insights* 14 (2020): 1177932220938064.

24. Guimarães LC., *et al.* "Inside the Pan-Genome - Methods and Software Overview". *Current Genomics* 16.4 (2015): 245-252.

25. Lukjancenko O., *et al.* "PanFunPro: PAN-genome analysis based on FUNctional PROfiles". *F1000 Research* 2 (2013): 265.

26. Raath-Krüger MJ., *et al.* "Positive Plant-Plant Interactions Expand the Upper Distributional Limits of Some Vascular Plant Species". *Ecosphere* 10.8 (2019): e02820.

27. Sanjerehei MM., *et al.* "Facilitative and Competitive Interactions between Plant Species (An Example from Nodushan Rangelands, Iran)". *Flora: Morphology, Distribution, Functional Ecology of Plants* 206.7 (2011): 631-637.

28. Wang C-H and Li B. "Salinity and Disturbance Mediate Direct and Indirect Plant-Plant Interactions in an Assembled Marsh Community". *Oecologia* 182.1 (2016): 139-152.

29. Strauss SY and Irwin RE. "Ecological and Evolutionary Consequences of Multispecies Plant-Animal Interactions". *Annual Review of Ecology, Evolution, and Systematics* 35.1 (2004): 435-466.

30. Silva RF., *et al.* "The Ecology of Plant Chemistry and Multi-Species Interactions in Diversified Agroecosystems". *Frontiers in Plant Science* 9 (2018): 1713.

Volume 7 Issue 2 February 2021