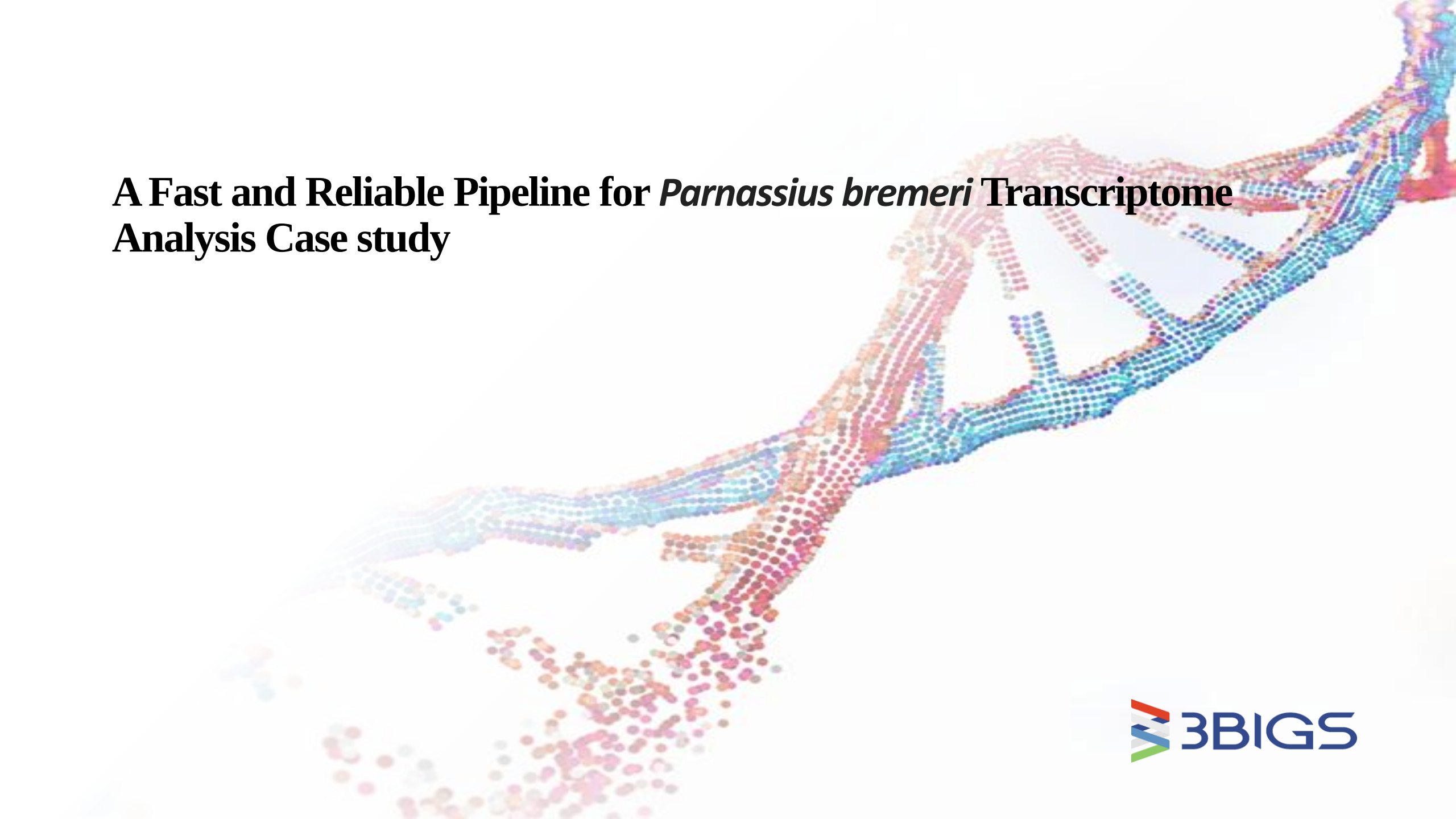
A Fast and Reliable Pipeline for *Parnassius bremeri* Transcriptome Analysis Case study



OVERVIEW

- *Parnassius bremeri* (*P. bremeri*), a member of the genus Snow Apollo in the swallowtail family (Papilionidae), is a high alpine butterfly that lives in Russia, Korea, and China. It is an endangered wildlife (Class I) in South Korea and is a globally endangered species. The lack of transcriptomic and genomic resources of *P. bremeri* significantly hinders the study of its population genetics and conservation. The detailed information of the developmental stage-specific gene expression patterns of *P. bremeri* is of great demand for its conservation. However, the molecular mechanism underlying the metamorphic development of *P. bremeri* is still unknown. In the present study, the differentially expressed genes (DEGs) across the metamorphic developmental stages were compared using high-throughput transcriptome sequencing. We identified a total of 72,161 DEGs from eight comparisons. GO enrichment analysis showed that a range of DEGs were responsible for cuticle development and the melanin biosynthetic pathway during larval development. Pathway analysis suggested that the signaling pathways, such as the Wnt signaling pathway, hedgehog signaling pathway and Notch signaling pathway, are regulated during the developmental stages of *P. bremeri*. Furthermore, sensory receptors were also activated, especially during the larval to adult transition stage. Collectively, the results of this study provide a preliminary foundation and understanding of the molecular mechanism in their transcriptomes for further research on the metamorphic development of *P. bremeri*.

CHALLENGES

i) This organism is very rare, sample collection is big challenge. *P. bremeri,* a high-altitude butterfly, is found in Russia, Korea and China. It belongs to the snow Apollo genus (*Parnassius*) of the swallowtail family (*Papilionidae*) but none of the *Parnassius* species have tails. The insect is found in flat, open landscapes, on slopes with forests up to the alpine zone (1500 m) and in forest steppes. It flies in May and June.

ii) We don’t have a reference genome for the genome, we must build the reference with the raw fastq file.

iii) Validation of Final fasta file generated. Transcripts from the denovo assembly.

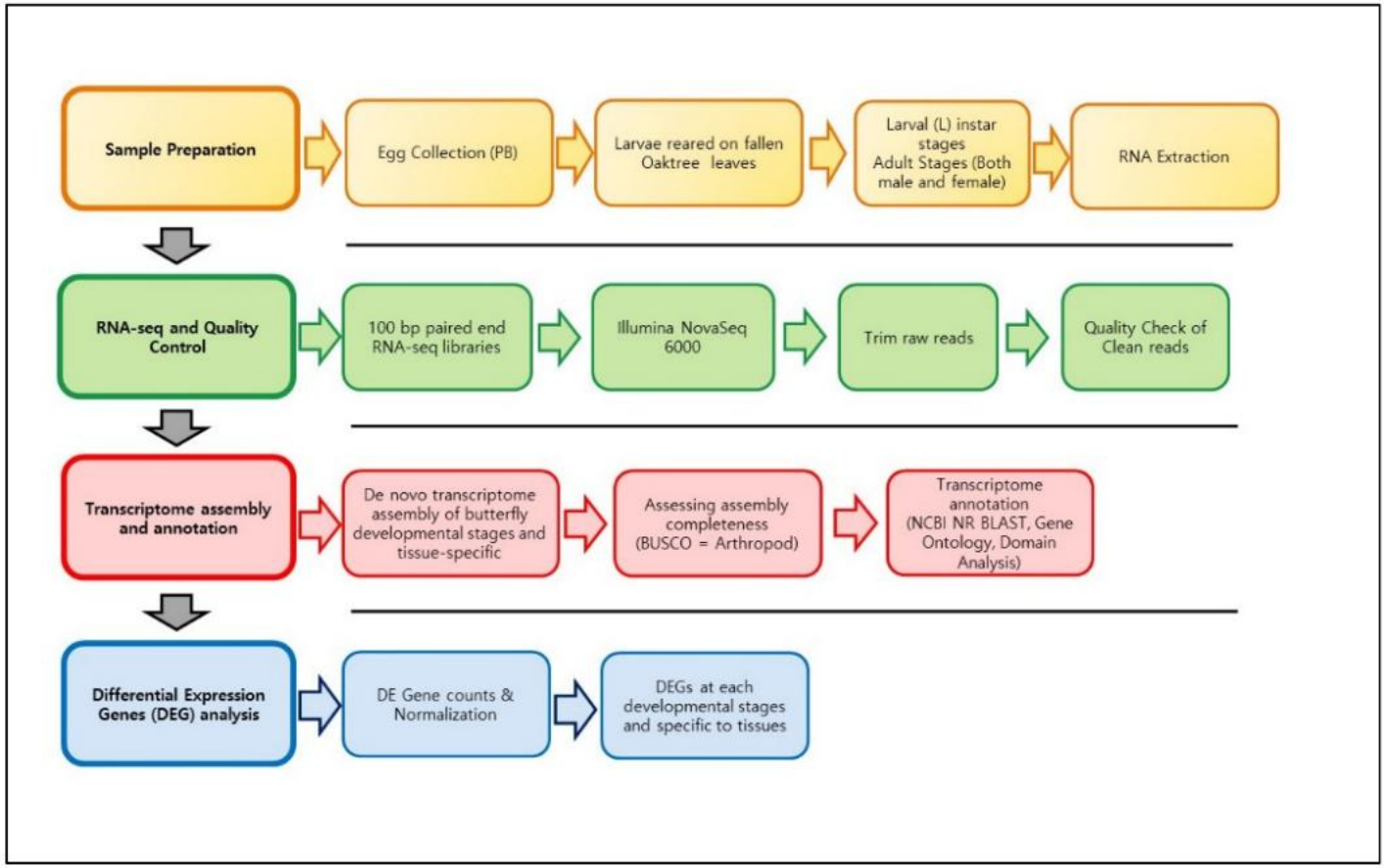
Approach

- Most previous studies on *P. bremeri* focused mainly on the organization of its mitogenome, population studies, ecological characteristics, sex pheromones and polyol profiles. Few reports illustrated the importance of population genetics and conservation biology through development of microsatellite markers. Another study, which was based on *P. bremeri* transcriptomes, reported the importance of glycerol accumulation in *P. bremeri* and its role in yielding cold tolerance to the insect. However, still more research is needed on the genomic and transcriptomic aspects for a thorough understanding of the population genetics and species conservation.

- Recently, RNA-seq technology has been used as an effective molecular tool to study the evolution of species, determine differentially expressed genes, and examine the population dynamics of organisms over time. With the help of RNA-seq technology numerous studies on arthropods have been done, especially in the order Lepidoptera. Developmental studies have been done on several butterfly species, including *Vanessa cardui*, *Pieris rapae*, *Bicyclus anynana*, and many others. The genome of the nearest species of *P. bremeri, P. apollo, Papilio xutus, Papilio machaon* and *P. bianor* have been studied recently. In addition, in recent times we have identified antimicrobial peptide candidates in *P. bremeri* against the causative agent of periodontitis, *Porphyromonas gingivalis*. However, the mechanisms involved in the development of *P. bremeri* from egg stage to adult stage remain unclear.

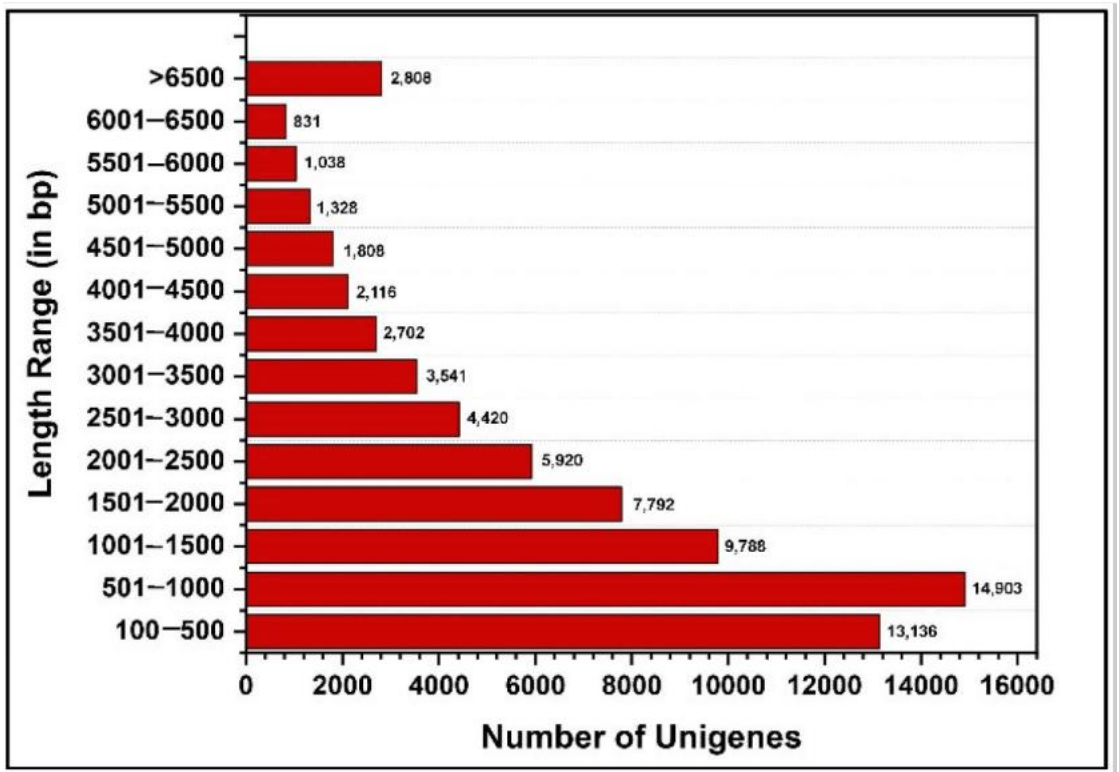
- In this study, performed high-throughput transcriptome profiling of eight developmental stages of *P. bremeri*, including the egg (PB), first to fifth instar larva (L1–L5, respectively), adult male (AM) and adult female (AF), and of six adult tissues, namely, antennae (F), head (H), leg (L), wing (W), reproductive organs (R), and body (B). A high-quality transcriptome assembly and annotated transcripts were provided to allow a comprehensive comparison between the eight stages and further elucidate their differences at the gene level. Transcripts related to genes involved in cuticle expression, phenol oxidase expression, juvenile hormone signaling, ecdysone hormone signaling, Wnt signaling, hedgehog signaling, notch signaling, and sensory development were identified and their expressions across developmental stages were investigated. Our transcriptome data expand knowledge of the molecular components of the developmental aspects of *P. bremeri*. As far as we know, this is the first study to analyze transcriptome data from the different developmental stages of *P. bremeri*.

Analysis Pipeline



Sequencing and De Novo Assembly of the P. bremeri Transcriptome

- Sequencing of *P. bremeri* tissue yielded 855,693,483 paired 150 bp reads with an average of 25,930,106 reads per sample. Raw reads were cleaned by a quality trim (Phred = 0.01) and adaptors were removed. The trimmed high-quality reads were subjected to de novo assembly using Trinity Release v2.14.0, which yielded 72,161 unigenes with an average length of 2063 bp. Of the 72,161 unigenes, 43,663 (60.50%) were greater than 1 kb and 26,542 (36.8%) were greater than 2000 bp



Sequencing and De Novo Assembly of the P. bremeri Transcriptome

- A BUSCO (Benchmarking Universal Single-copy Orthologs) analysis was performed to determine the completeness of the assembly. BUSCO v5.3.1 revealed a high rate of 965 (95.26%) similar orthologous genes from the Arthropoda ortholog set with few fragmented or missing BUSCOs (Complete gene representation: 94.17% [Single copy: 36%, Duplicates: 58.1%], Fragmented: 1.1%, Missing: 4.8%), indicating a high-quality transcriptome. A summary of the transcriptome composition and quality assessment by BUSCO.

Sequencing and De Novo Assembly of the P. bremeri Transcriptome

A BUSCO (Benchmarking Universal Single-copy Orthologs) analysis was performed to determine the completeness of the assembly. BUSCO v5.3.1 revealed a high rate of 965 (95.26%) similar orthologous genes from the Arthropoda ortholog set with few fragmented or missing BUSCOs (Complete gene representation: 94.17% [Single copy: 36%, Duplicates: 58.1%], Fragmented: 1.1%, Missing: 4.8%), indicating a high-quality transcriptome. A summary of the transcriptome composition and quality assessment by BUSCO.

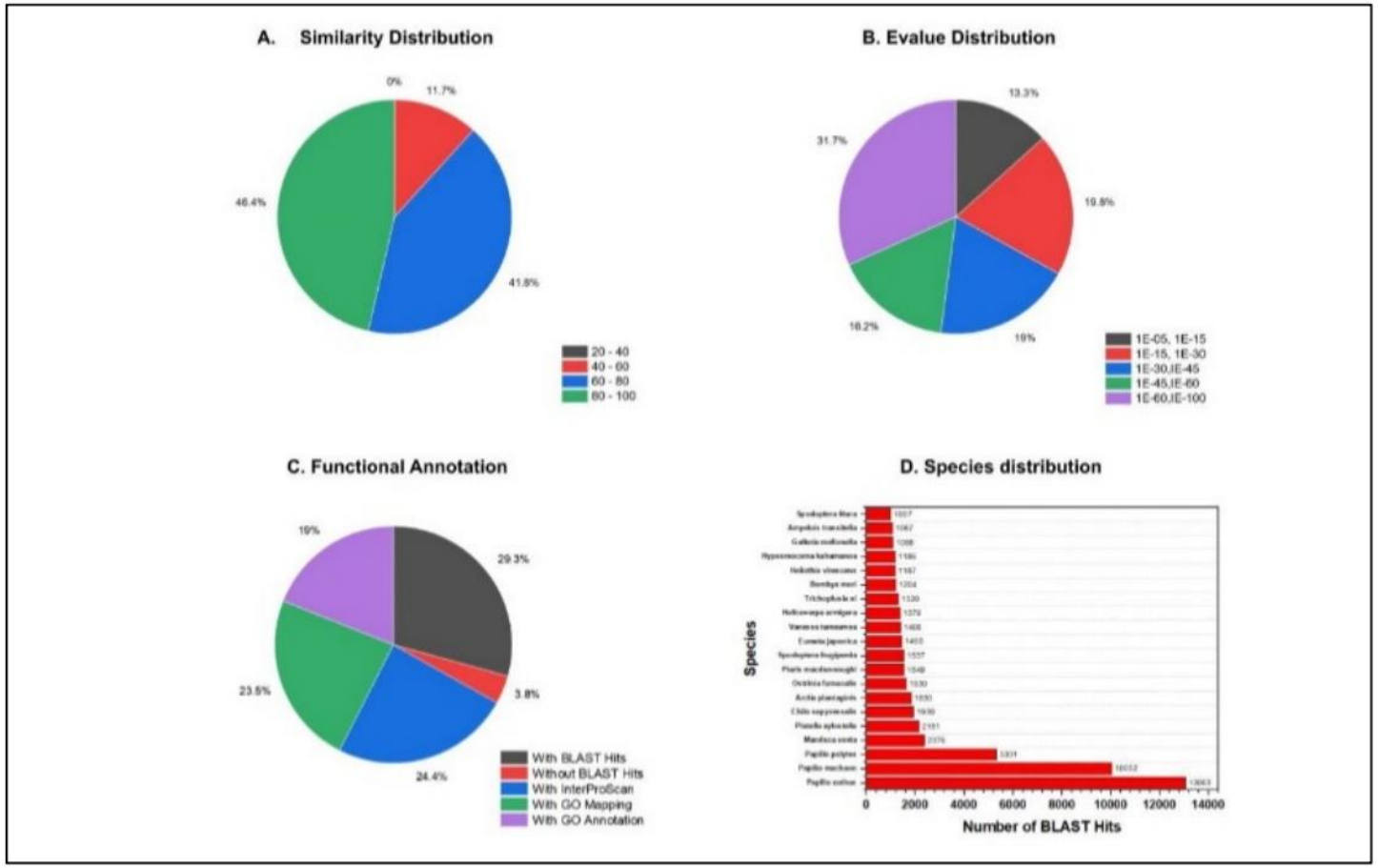
Measure	Value
Number of genes	72,161
Number of transcripts	124,158
Average contig length (bp)	1043.2
Median contig length (bp)	627
Total assembled bases	75,278,310
GC content of unigene (%)	45.43
Minimum sequence length (bp)	297
Maximum sequence length (bp)	48,967
Number of genes > 1 kb	44,218
Number of genes > 5 kb	472
N50 (bp)	1560
Features	**Results**
Complete BUSCOs (C)	954 (94.17%)
Complete + Partial	965 (95.26%)
Complete and single-copy BUSCOs (S)	365 (36%)
Complete and duplicated BUSCOs (D)	589 (58.1%)
Fragmented BUSCOs (F)	11 (1.1%)
Missing BUSCOs (M)	48 (4.8%)
Total BUSCO groups searched	1013
BUSCO version	BUSCO_v5.3.1
Selected ortholog set	Arthropoda

Functional Annotation

- Among the 72,161 assembled unigene sequences, 46.4% of unigenes showed a similarity distribution of 80–100%, 41.8% of unigenes had 60–80% similarity, and 11.7% of unigenes had 40–60% sequence similarity ($\leq 1 \times 10^{-5}$) to known proteins in the NCBI public databases (nr). The distribution of E-values also showed that most unigenes were between 1×10^{-60} and 1×10^{-100} . Of the functional annotation of the 72,161 unigenes, 63,782 unigenes yielded BLAST hits, 53,093 had InterProScan IDs (24.4%), 51,255 unigenes with GO mapping (23.5%), 41,314 unigenes with GO annotation (19%), while 8379 of unigenes yielded no BLAST hits (3.8%).

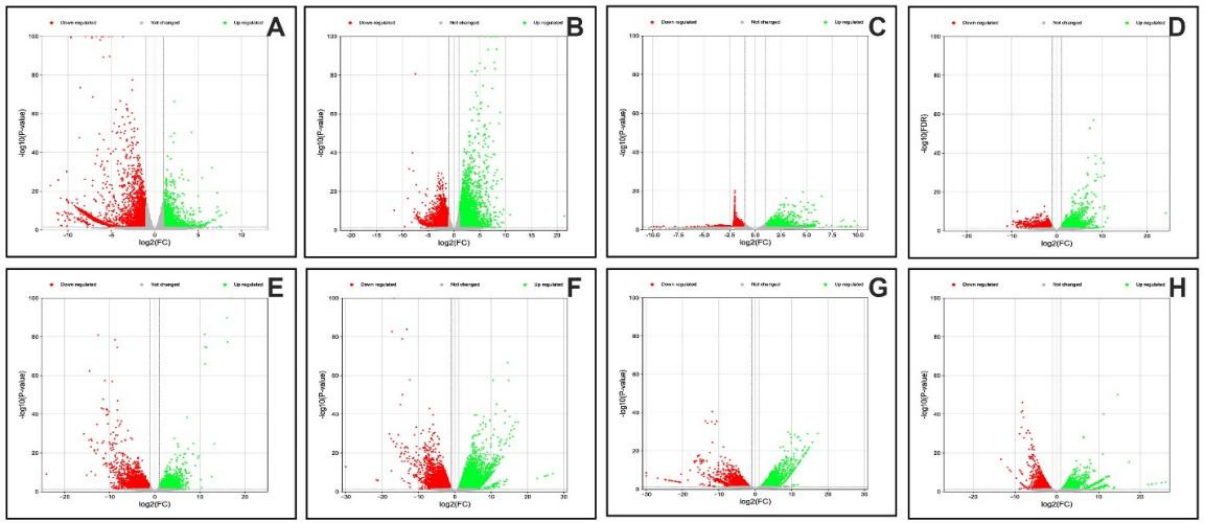
- Our results showed that the matching efficiency (i.e., sequences with hits) increased with the length of the unigene. Of the unigenes with significant blast hits, 13,063 (~21%) matched the Asian swallowtail butterfly, *Papilio xuthus*, followed by the Old-World swallowtail, Papilio machaon (~15.8%), and the common Mormon, *Papilio polytes* Blast results were used for functional categorization of the assembled unigenes.

Functional Annotation



DEG

- Among the 72,161 unigenes, differential expression of genes was analyzed in the biologically relevant comparisons. The differentially expressed genes in the different libraries were plotted in a volcano plot. The volcano plot shows significance on the *y*-axis and fold change on the x-axis. Fold change (log2FC ≥ 1) and an adjusted *p*-value (≤0.05) were used as thresholds for significance testing.



DEG

- The developmental stages of *P. bremeri* were compared in pairs. The up-regulated and down-regulated DEGs were selected if log2FC ≥ 1 and log2FC ≤ -1, respectively, with an adjusted *p*-value ≤ 0.05. DEGs in common and unambiguously regulated at all stages and tissues were analyzed using the Upset Plot in TBtools. The differentially expressed genes specific to the expression of ABC transporters and sensory receptors were identified based on Pfam domain while the members of hormone signaling pathways (Juvenile, Wnt, Hedgehog and Notch) were identified using KEGG Automatic Annotation Server (KAAS) by the assignment method of ‘bi-directional best hit’ against data of Arthropod dataset in the KEGG database.

Conclusion

- Our study on the transcriptomic regulation of *P. bremeri* metamorphosis clearly shows that certain unigenes were regulated across the developmental stages and in specific tissues. The generated resource from this study is helpful in laying foundations to understand the precise nature and development of other *Parnassius* butterflies. This study can be used to improve adaptive management plans geared towards species conservation. Pathway identification and candidate gene exploration are important steps toward a better understanding of the complex mechanisms behind the development of butterflies. Further, validation of the expression of candidate genes could improve the conservation of *P. bremeri*

https://3bigs.com | core@3bigs.com | 031-784-8611
Headquarters : URP3-3, 187, Gudeok-ro, Seo-gu, Busan, Republic of Korea (49241)
Global Center : B831, 27, Dongtancheomdansaneop 1-ro, Hwaseong-si, Gyeonggi-do, Republic of Korea (18469)