

PacBio Hi-Fi genome assembly of the cereal aphid *Sipha maydis*, a model for the study of multipartite symbiosis in insects

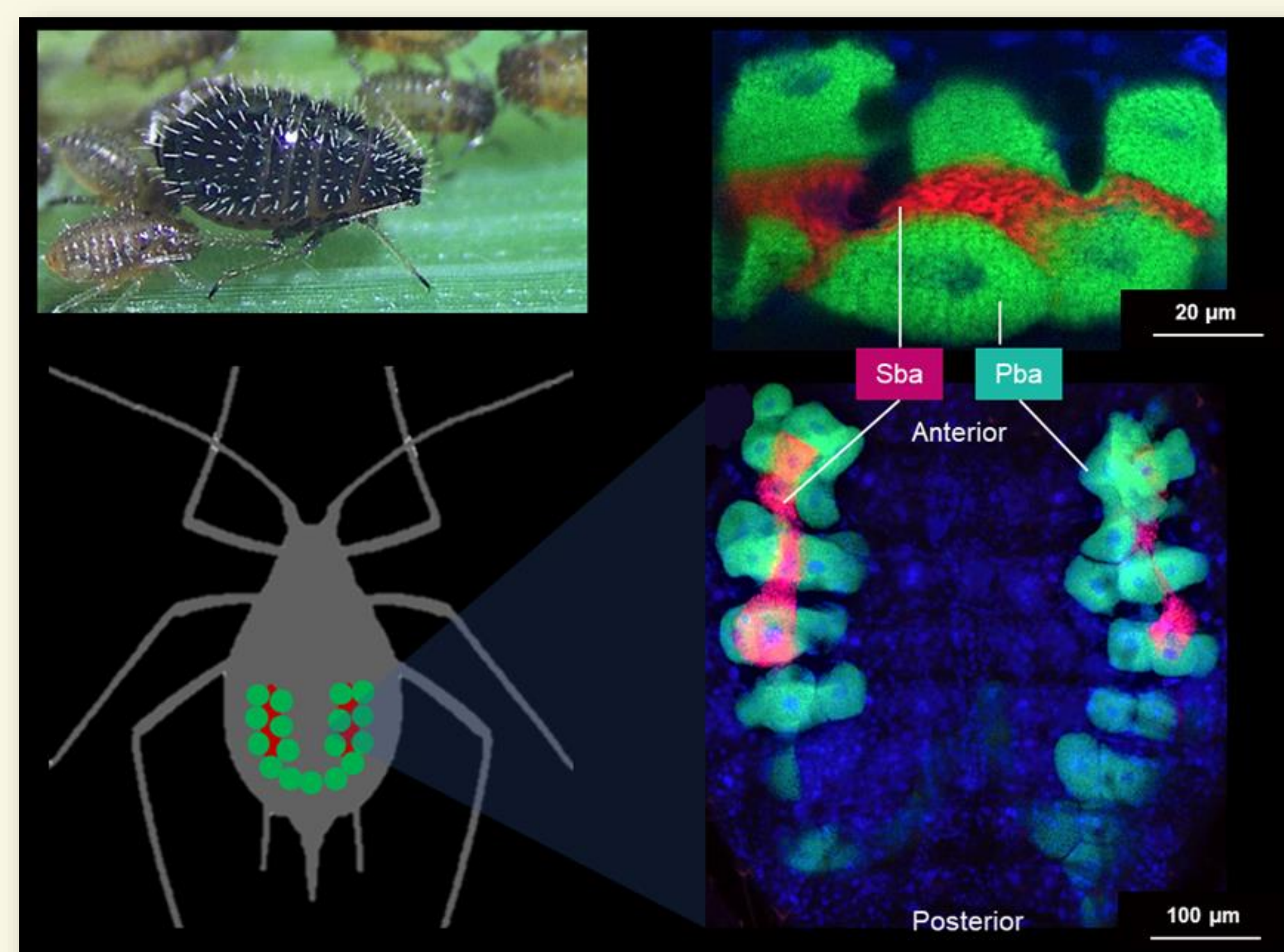
François Renoz^{1,2,3†}, Nicolas Parisot^{2†}, Patrice Baa-Puyoulet⁴, Léo Gerlin⁴, Samir Fakhour^{1,5}, Hubert Charles², Thierry Hance¹, Federica Calevro⁴

¹Biodiversity Research Centre, Earth and Life Institute, UCLouvain, Louvain-la-Neuve, 1348, Belgium, ²Univ Lyon, INSA Lyon, INRAE, BF21, UMR203, Villeurbanne, F-69621, France, ³Institute of Agrobiological Sciences, National Agriculture and Food Research Organization (NARO), Tsukuba, Ibaraki 305-8634, Japan, ⁴Univ Lyon, INRAE, INSA Lyon, BF21, UMR203, Villeurbanne, F-69621, France, ⁵Department of Plant Protection, National Institute for Agricultural Research (INRA), Béni-Mellal, 23000, Morocco

† These authors contributed equally to this work.

INTRODUCTION

- Nutritional dependence on metabolically interconnected endosymbionts has evolved repeatedly in insects feeding on unbalanced diets (Douglas 2016, *Nature Reviews Microbiology*).
- However, reference genomes for species with multi-symbiotic nutritional systems are lacking, even though they are essential for understanding the processes governing cooperative life.
- We generated a **PacBio Hi-Fi assembly** of the genome of the cereal aphid *Sipha maydis*, which has evolved a nutritional dependence on *Buchnera aphidicola* and *Serratia symbiotica*, and examined the nutritional role of each partner in amino acid and vitamin biosynthesis.



The di-symbiotic system of *S. maydis* is composed of *B. aphidicola* residing in primary bacteriocytes (Pba) and *S. symbiotica* residing in secondary bacteriocytes (Sba).

RESULTS & DISCUSSION

A Comparison of the central metabolism in a monosymbiotic (*Acyrtosiphon pisum*) vs a disymbiotic (*S. maydis*) aphid:

- Assimilation pathways more limited in *S. maydis* than in *A. pisum*, but overall very few differences in metabolic capacity.
- Significant contribution of *S. symbiotica* to the metabolism of *S. maydis*.

B Amino acid metabolism in the *S. maydis* symbiotic system

- Interdependence in AA metabolism.
- B. aphidicola* and *S. symbiotica* have lost the ability to produce most non-essential AA & depend on the host.
- Redundant pathways in symbionts could boost the production of AA precursors.
- Phloem is a likely source of AA precursors (e.g. for cysteine & methionine).

C Vitamin metabolism in the *S. maydis* symbiotic system

- Small number of shared pathways.
- The system is essentially based on direct supply by the plant.
- Phloem is most likely the source of biotin.
- Riboflavin is supplied by *S. symbiotica*.

	<i>A. pisum</i>	<i>S. maydis</i>
Biosyntheses		
amino acids		
Arg	✓ Ba	✗ Aphid / Ba
Sulfurated amino acid	✓ Ba	✗
Chorismate	✓ Ba	✓ Ba ✓ Ss
DAP	✓ Ba	✓ Ba ✓ Ss
Ile, Leu, Val precursors	✓ Ba	✓ Ba ✓ Ss
Coenzyme A	✗ Ba	✗ Ss
B1	✗ Ba	✓ Ss
B2	✓ Ba	✓ Ss
FRDP (heme O precursor)	✓ Aphid ✓ Ba	✓ Aphid ✓ Ss
NAD from NADP	✓ Aphid	✗
Assimilation		
sulfur		
Sulfate	✓ Ba	✗
Sulfite	✓ Aphid ✓ Ba	✗
choline		
Phosphatidylcholine hydrolysis	✓ Aphid	✗
Glycine betaine production	✓ Aphid	✗
carbo-hydrate		
Mannose	✓ Aphid	✗
Mannitol	✓ Aphid	✗
Mellobiose	✓ Aphid	✗

✓ Achievable
✗ Codependence (↗ means codependence between diet and one organism)
✗ Unachievable, expected to be absent/undone in holobiont metabolism
✗ Unachievable, expected to be compensated by a supply from aphid diet

METHODS

Genome sequencing

DNA extraction (phenol-chloroform approach)
Sequencing on the PacBio Sequel IIe system
Raw PacBio HiFi reads

Genome survey and assembly

Genome size assesement : Jellyfish, findGSE, R and gce
Assembly: hifiasm, NCBI FCS, scaffolds of symbionts
Accuracy and completeness: QUAST, BUSCO + Insecta ODB10 database

Gene prediction & function annotation

RepeatModeler, RepeatMasker, BRAKER3, HISAT2, Arthropoda protein dataset from OrthoDB, Augustus productions, BUSCO + Insecta ODB10 database, EnTAP, UniProtKB/Swiss-Prot & NCBI RefSeq invertebrate annotated proteins databases.

Functional annotation of metabolism

KAAS-KEGG, PRIAM Blast2GO, InterProScan
CycADS (Vellozo et al. 2011, *Database*), BLAST alignments on UniProtKB/Swiss-Prot protein sequence database.

Metabolic network reconstruction – deciphering tripartite metabolic exchanges with ArtSymbioCyc

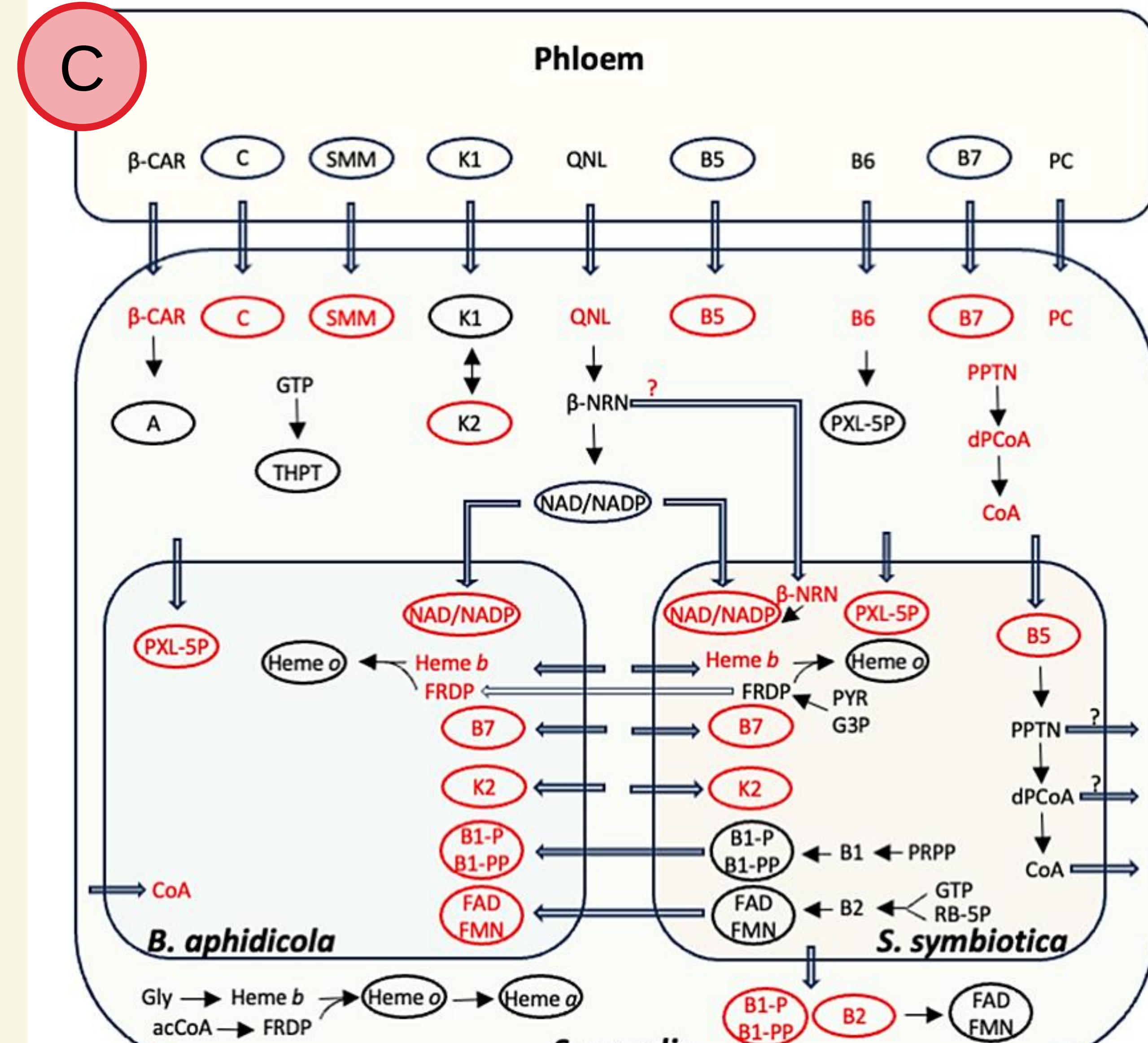
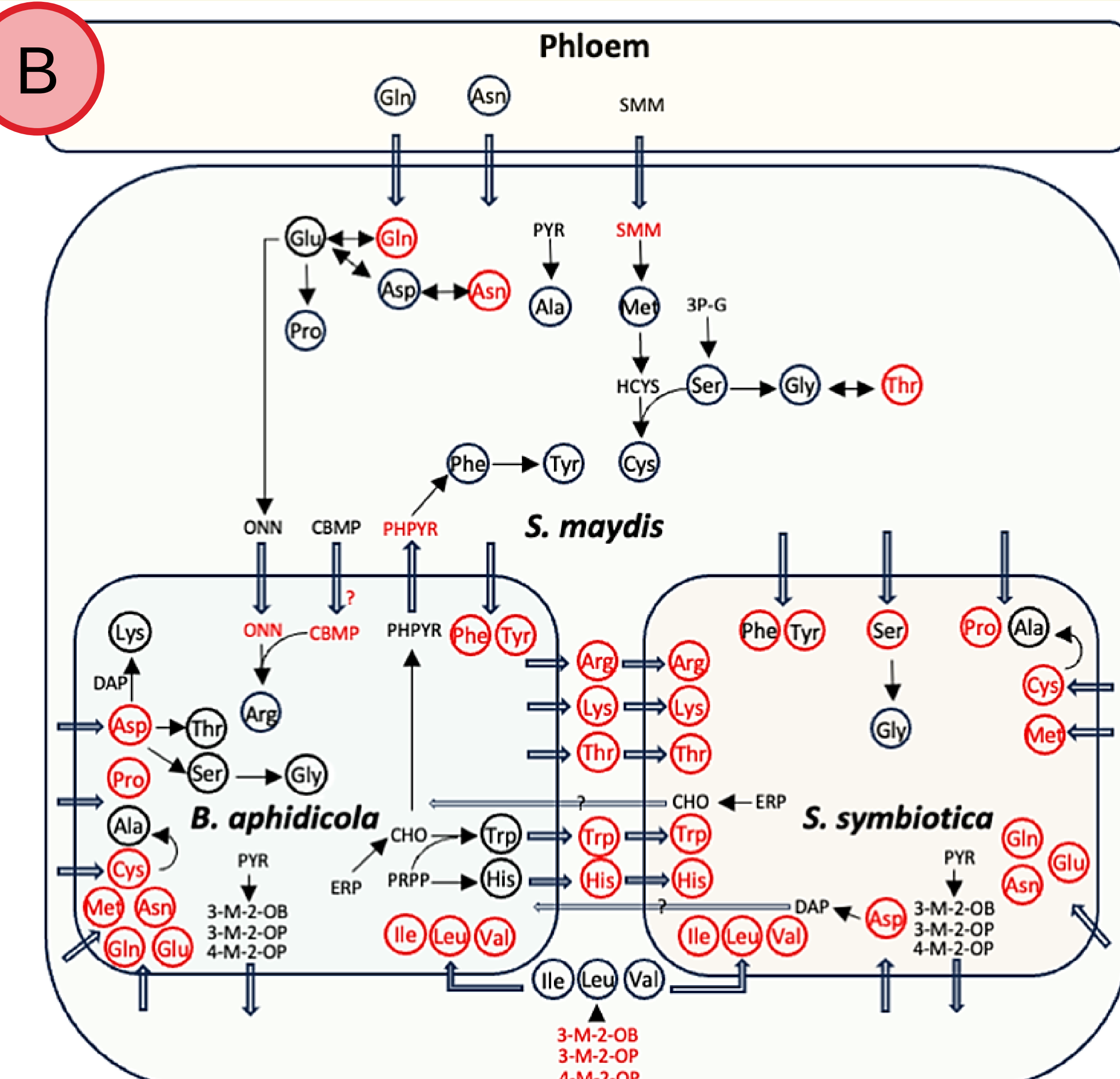
CycADS, Pathway Tools

Three metabolic databases: SipmaCyc, Bucap_smCyc & Sersy_smCyc

ArtSymbioCyc for visualization & comparison of metabolic networks

Metrics	<i>Sipha maydis</i> (this study)	<i>Acyrtosiphon pisum</i> (AL4f)
Total length (Mb)	409.54	533.6
No. of scaffolds/contigs	3,570	21,227
Scaffold/Contig N50 (kb)	187.22	132.5
	C:95.5%	C:98.9
Assembly BUSCO insecta_odb10 scores	[S:91.0%, D:4.5%]	[S:96.3%, D:2.6%]
	n:1,367	
	F:0.8%	F:0.3%
	M:3.7%	M:0.8%
Encoded genes	26,046	26,634
Encoded enzymes	2,541	3,897
Metabolic pathways	289	271

Genome assembly and annotation statistics of *S. maydis*



DATA ACCESSIBILITY

- The **sequencing data** that were used for the genome assembly and annotation have been deposited in the NCBI Sequence Read Archive with accession number **PRJNA983969**. This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession **JASXSY000000000**.
- S. maydis* **metabolic database** inferred from the genomic information can be found alone in the Arthropod collections (<https://arthropodacyc.cycadsys.org>) or associated with its obligate co-symbionts in the ArtSymbioCyc collection (<http://artsymbiocyc.cycadsys.org/>).

CONCLUSION

These genomic and metabolic data provide a unique tool for exploring in depth the molecular mechanisms governing the biology of *S. maydis*, and in particular **the mechanisms underpinning an interdependent tripartite collaborative life between an insect and its intracellular prokaryotic associates.**