

An interdependent metabolic patchwork in the nested symbiosis of mealybug

1 Citrus mealybug *Planococcus citri* represents a nested symbiosis system: a betaproteobacteria *Candi-*
2 *datum* *Tremblaya princeps* lives inside citrus mealybug, while a gammaproteobacteria *Candidatus* *Moranella*
3 *endobia* lives in cytoplasm of *Tremblaya*.

4 *Tremblaya* genome is extremely small (0.14 Mbp) and degenerated (121 proteins). A 7-kbp region of
5 *Tremblaya* genome exists two orientations within single insect host, while *Tremblaya* lacks genes involved
6 in recombination.

7 *Tremblaya* retains genes involved in essential amino acid synthesis, but does not have complete path-
8 ways of its own. *Moranella* complements several essential amino acid synthesis genes lost in *Tremblaya*.
9 However, it is unclear how transport of metabolites occurs between cosymbionts. *Tremblaya* genome
10 encodes no predicted transporters. *Moranella* encodes a handful of proteins involved in membrane trans-
11 portation, but none are specific for amino acids or their precursors. Some components of the Sec translo-
12 cation machinery are present in the *Moranella* genome, and it is possible that these are used to transport
13 some proteins across *Moranella*'s inner membrane. A search for signal peptides in the *Moranella* proteome
14 revealed 27 proteins with N-terminal secretory signal peptides; however, none was involved in essential
15 amino acid biosynthesis.

16 *Tremblaya* is missing several gene homologs for translation-related functions that are often retained in
17 other highly reduced bacteria genomes, including all aminoacyl-tRNA synthetases, translational release
18 factors. As translation machinery is significantly different in eukaryotes and bacteria, it seems unlikely
19 that the missing translation-related genes in *Tremblaya* are complemented by host. Horizontal gene
20 transfer from bacteria to host might be the solution, although no transfer of functional genes between
21 symbiont and host has been found in another two insects, pea aphid (Nikoh *et al.*, 2010) and human body
22 louse (Kirkness *et al.*, 2010).

23 The nested structure of the mealybug symbionts is likely controlled by the host. There are at least
24 two morphological forms of *Moranella*: a reproductive form in which cells were small in size and in
25 the process of dividing, and a degenerative phase in which cells became unevenly shaped and elongated

(Buchner, 1965). The particular *Moranella* form was dependent on the life stage of the insect and seemed to be synchronized within a bacteriocyte (Buchner, 1965). Furthermore, the infection levels of *Tremblaya* and *Moranella* are uncoupled in mealybugs (Kono *et al.*, 2008). During male development, the number of *Moranella* cells relative to *Tremblaya* cells drops significantly as the insects age, whereas in female insects, the levels of the two symbionts remain roughly equivalent over the entire life cycle (Kono *et al.*, 2008). Given that *Tremblaya* has an extremely limited coding capacity that is largely devoted to essential amino acid biosynthesis and translation, and given that only seven genes are of completely unknown function, it seems impossible that *Tremblaya* itself controls any structural aspect of the symbiosis. Likewise, the *Moranella* genome does not encode any genes involved in traditional infective strategies and does not indicate any obvious pathway by which it could be an active participant involved in seeking out the *Tremblaya* cytoplasm. Thus, it seems likely that the host is largely in control of the structure and organization of this bacteria-within-a-bacterium symbiosis.

Tremblaya survives with highly reduced genome with loss of genes thought to be essential for survival (*e.g.* translation). The missing activities can be complemented by several mechanisms: (1) gene products or metabolites of either host or bacterial origin imported from the host; (2) gene products or metabolites imported directly from the other symbionts if present; (3) genetic coadaptations to the loss of genes within the reduced genome itself; (4) the direct use of *Moranella* gene products as a result of a simple, passive mechanism such as *Moranella* cell lysis within the cell membrane system of *Tremblaya*.

Tremblaya genome is extremely small, but low gene dense. During the shift from a free-living to an obligate intracellular lifestyle, where the constant exposure to the stable and rich environment of the host cell combined with a severe reduction in population size (and subsequent reduction in the efficacy of purifying selection) allows large numbers of pseudogenes to accumulate (Ochman *et al.*, 2006; Andersson *et al.*, 2001). These pseudogenes are eventually purged from the genome through mutational patterns favoring deletions (Mira *et al.*, 2001), leading to small gene-dense genomes such as those from insect nutritional symbionts. A possible explanation is that *Tremblaya* undergone genome reduction after association with mealybug, and acquisition of *Moranella* leads to further genome reduction. Basal lineages of mealybugs in the same subfamily as citrus mealybug seem to contain *Tremblaya* without the intracellular gammaproteobacterial endosymbiont (Hardy *et al.*, 2008; Thao *et al.*, 2002), indicating that *Moranella* was acquired after the establishment of *Tremblaya*. The patterns of gene pseudogenization also fit this hypothesis, as most pseudogenized *Tremblaya* genes have functional *Moranella* homologs.