Reviewer report for Bulah Chia-hsinang Wu's thesis entitled "Comparative Analysis of Silk Proteins and Discovery of Novel Sericin Gene in Lepidopteran Moths"

Chapter 1 of this thesis aims to provide a detailed profile of silk protein components from two *Pyralidae* species by omics analyses of silk genes and proteins. It also discusses the characteristic amino acid sequences of fibroin heavy chains from *Pyralidae* in comparison with those from other lepidopteran families from an evolutionary point of view.

Some silk proteins, especially fibroins, are poorly soluble in aqueous solvents, because they form rigid secondary structure. This sometimes causes problems for standard protein characterization using SDS-PAGE. Furthermore, the considerable repetitive structure of silk encoding genes often leads to imprecise genome assembly. Therefore, Wu used proteomic analysis using LC-MS for silk protein characterization along with transcriptome supported by other techniques such as quantitative RT-PCR, northern blotting, synteny analysis using *G. mellonella* silk encoding region.

Chapter 2 reports the *B. mori* homolog of *P150* in detail. Through characterization of this newly identified sericin gene, designated as *sericin 6*, its similarity with *mucin-12* was recognized. Phylogenetic analysis of the two genes of the silkworm and other lepidopteran insects suggested that these genes are distinct but closely related.

I think that this study is properly organized and worth reading. The results are reliable, because several techniques are combined to compensate each other. The data obtained are available online, and the databases and programs used for analysis are referred properly. Writing and display of tables and images are accessible, though some images include too small marks or letters (Figure S1 on page 31, Figure 4 on page 64 etc.). This manuscript includes a publication and its supplementary material as part of Chapter 1. This probably makes this thesis a little confusing, such as numbering of figures and tables and variable gene and protein names.

It is hard to evaluate Wu's contribution to this study, because I am not an expert of bioinformatics and cannot evaluate his skill. However, as I mentioned above, this study provides reliable information about *E. kuehniella* silk components and the most part involves database analysis conducted by Wu. Regarding novelty of the work, this is the first study to identify and characterize silk components of this species. Silks collect attentions as protein materials for their sustainability and diverse properties, though very limited information is available other than silkworm and spider silks. Data accumulation of silk protein sequences and properties is strongly desired for both basic and applied studies. This thesis demonstrates an omics-based strategy for silk gene and protein characterization, which will accelerate silk studies.

Some specific points to be considered are listed below.

1. Names of genes and proteins are variable in this thesis, for example, *FibH*, *H-fib*, and *fibH* for the *fibroin heavy chain* gene, and *ser-1* and *Ser1* for the *sericin 1* gene (e.g., page 2). I recommend terms be unified throughout the thesis.

2. Figure 2 indicates that fibroin light chain binds to the N-terminal region of fibroin heavy chain,

which is not consistent with the description in the text (page 3).

3. The width of *E. kuehniella* cocoon fibers does not seem approximately 1 μm in FIGURE 2 (page 18).

4. What, for example, is considered for specific function of Ser1-like proteins which directly cover fibroin fiber?

5. Wu speculates that *E. kuehniella* Ser4 protein may be lubricant as anticipated for *B. mori* Ser3 because of their similarity in serine content (page 26). Why are sericins with 40% serine content considered to serve as lubricant?

6. Previous studies reported that *B. mori Ser2* is expressed in the anterior and middle parts of middle silk glands until early fifth instar, and therefore Ser2 proteins are considered to constitute the outer layer of silk fibers (Figure 3 on page 5). However, Wu mentions that Ser2 protein(s) constitutes the inner layer of silkworm cocoons based on proteomic analysis (page 53). Which is more reliable and why?

7. Duplication of the sericin encoding region in *G. mellonella* is reported in Chapter 1 and it is repeated again in Chapter 2. Are there any specific features in the arrangement of *B. mori sericin* genes (page 53)?

8. Although sericins are not thought to contribute to the mechanical property of silkworm silk fibers, Wu mentions that sericins contribute to silk strength and toughness (page 55). How do sericins contribute to these properties?

9. What do "simple sericins" which are digestible by cocoonase mean? I could not find relevant part in the cited publication (page 55).

10. The importance of the conserved cysteine knot motif observed in P150/ser6 and Muc-12 is not clearly described (page 56). Does it involve the function of this motif or simply the genetic similarity of these proteins?

11. I recommend to indicate the sampling stage of the silkworm larvae in the legend of Supplementary Figure S1 (page 76), because the expression of *B. mori sericin* genes changes in levels and locations with the developmental stage (Couble et al, Dev Biol 124, 431-440, 1987).

Lastly, I recognize the value of this thesis and recommend it for defense.

Yoko TAKASU

Institute of Agrobiological Sciences

National Agriculture and Food Research Organization