

# Large-scale sequencing and characterization of silkworm full-length cDNAs

Yoshitaka Suetsugu, Keiko Kadono-Okuda, Akiya Joraku, Kimiko Yamamoto

Insect Genome Research Unit

**We performed large scale analysis of 21 full-length cDNA libraries derived from 14 tissues of the domesticated silkworm and determined the complete sequences and accurate structures of 11,104 cDNAs. These data is now opened to public on the silkworm genome database named KAIKObase.**

Keywords: silkworm, gene, genome

---

## Background

The domesticated silkworm *Bombyx mori* is a traditionally important insect for silk production and also a model species of Lepidopteran insects which include many agriculturally destructive pests. The international consortium on silkworm genome sequencing revealed a high-quality genome sequence in 2008 which led to significant progress in silkworm research. However, a more accurate information on the structure and function of the silkworm genes as well as the exact position on the genome is necessary to fully utilize the genome information for industrial applications such as genome-oriented discovery of insecticides. The full-length cDNA sequences (FL-cDNAs) represent exact copies of messenger RNAs that are transcribed from genomic DNA, and contain all information to produce proteins. The nucleotide sequence information of FL-cDNA is therefore indispensable for understanding structure and function of each gene. We analyzed the complete FL-cDNA sequences of silkworm to facilitate efficient utilization of silkworm genome information and accelerate its practical application in various industries.

## Results and Discussion

1. We constructed twenty-one FL-cDNA libraries representing 14 distinct tissues of the domesticated silkworm, *B. mori* and randomly selected approximately 250,000 cDNA clones from these libraries. The complete sequences of unique 11,104 FL-cDNAs were determined (Fig. 1).
2. We characterized the 11,104 FL-cDNA sequences as well as other available silkworm datasets including 408,172 ESTs, 2089 mRNA sequences in public databases, and 14,623 gene models. The result shows that *Bombyx mori* has more than 17,000 genes, which is much larger than a previous estimate of 14,623 genes.
3. Comparison of silkworm FL-cDNAs with the gene sets of other Lepidopteran insect such as *Danaus plexippus* revealed evidence of silkworm specific gene groups related with biological defense or structural genes.
4. The silkworm FL-cDNAs were integrated with the genome sequence information and can be accessed via KAIKObase (<http://sgp.dna.affrc.go.jp/KAIKObase/>). A system for distribution of the cDNA clones as genomic resources has also been established so that interested researchers could get access to the clones (Fig. 2).

## Future prospects

1. The newly obtained FL-cDNA sequences enabled us to annotate the genome of silkworm more accurately, enhancing functional studies of this lepidopteran model insect. Since many agricultural pests belong to lepidopteran insects, our results will facilitate analysis of important genes such as those related to resistance against specific chemicals or target genes that could lead to the development of lepidopteran-specific insecticides.
2. Our results are expected to facilitate accurate analysis of gene structure and characterization of gene functions that will eventually accelerate the identification of useful genes associated with the

production of various materials using transgenic silkworm and the promotion of their quality and performance. It will also facilitate the identification of genes that control the glycosylate system of silkworm that can be used for efficient production of human-type glycoproteins for therapeutic applications.

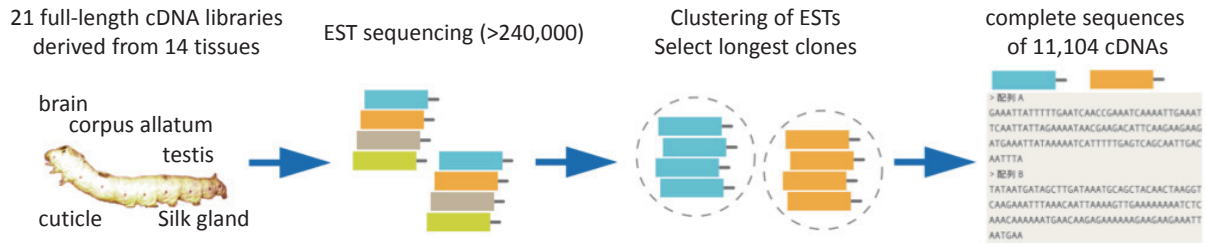


Fig. 1. Summary of construction and sequencing of silkworm FL-cDNAs. Twenty-one FL-cDNA libraries were constructed from 14 distinct tissues. Approximately 250,000 cDNA clones were sequenced and clustered. The longest clone was selected from each cluster and fully sequenced by primer-walking method.

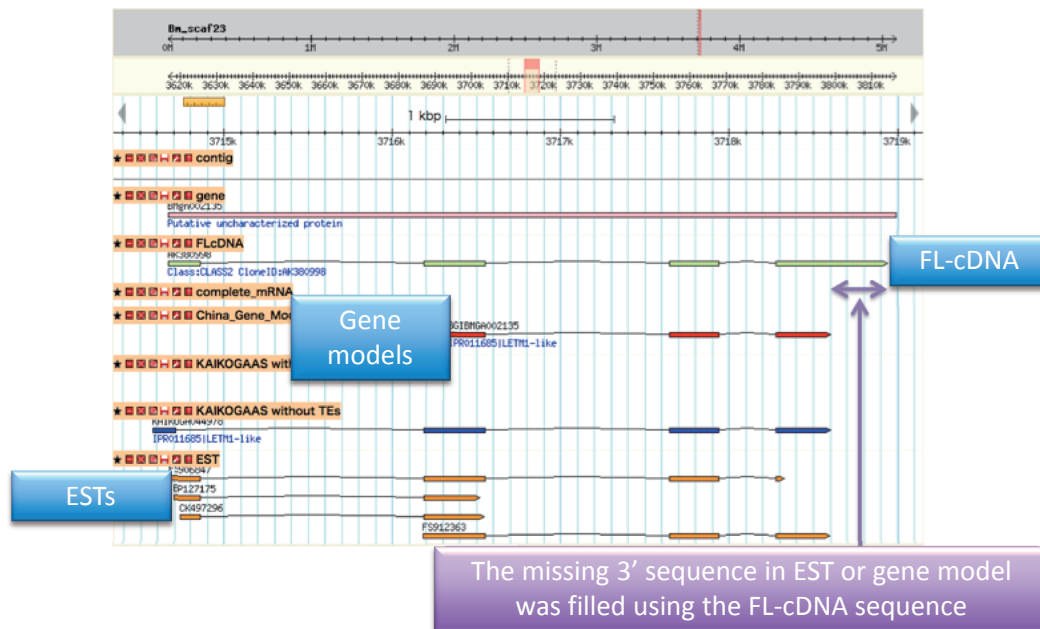


Fig. 2. Integration of FL-cDNA information in silkworm genome database. The FL-cDNA sequences can be accessed in the public domain via KAIKObase (<http://sgp.dna.affrc.go.jp/KAIKObase/>). The GBrowse features information on the nucleotide sequence, gene models and ESTs. The FL-cDNA information could be used in accurate analysis of the complete structure of the genes with missing UTR information and ESTs.

## Collaborators

Ryo Futahashi (National Institute of Advanced Industrial Science and Technology), Kazuei Mita (Southwest University), Toru Shimada (The University of Tokyo)

## Reference

1. Suetsugu Y, Futahashi R, Kanamori H, Kadono-Okuda K, Sasanuma S, Narukawa J, Ajimura M, Jouraku A, Namiki N, Shimomura M, Sezutsu H, Osanai-Futahashi M, Suzuki M.G, Daimon T, Shinoda T, Taniai K, Asaoka K, Niwa R, Kawaoka S, Katsuma S, Tamura T, Noda H, Kasahara M, Sugano S, Suzuki Y, Fujiwara H, Kataoka H, Arunkumar K.P, Tomar A, Nagaraju J, Goldsmith M.R, Feng Q, Xia Q, Yamamoto K, Shimada T, Mita K (2013) Large scale full-Length cDNA sequencing reveals a unique genomic landscape in a lepidopteran model insect, *Bombyx mori* **G3: Genes, Genomes, Genetics** 3(9):1481-1492