

General Notices

Introduction

Chinese Materia Medica Reference Sequence Library (CMMRSL) is a repository of reference DNA sequences of specimens with known taxonomic identity determined in research projects of GCMTI. To facilitate dissemination of the latest research deliverables, reference DNA sequences and information of selected DNA barcodes are being announced to the sectors and the public in the form of fact sheet in a timely manner.

Arrangement and interpretation of fact sheet

A fact sheet provides information and reference DNA sequences of one CMM species, which comprises cover page, Reference DNA Sequences and Reference DNA sequences in FASTA format.

- (a) Cover page – summarizes the information of all specimens used for data generation, the properties of all reference DNA sequences, and, if deemed necessary, additional information on experimental conditions of DNA barcode generation for particular organisms.
- (b) Reference DNA Sequence – provides the collection information and reference DNA sequences of each specimen, which is arranged in the following order:
 - DNA Information – list default DNA barcodes in FASTA format with a unique sequence identifier, if deemed appropriate. In most circumstances, the reference DNA sequences are produced by using GCMTI in-house validated methods which the entire workflow is released and freely available. The selected barcodes are adopted in national Pharmacopeias or widely accepted by the scientific communities for the purpose of species identification. Fact sheet furnishes up to three DNA barcodes for each specimen to strengthen discrimination capability and flexibility.
 - Reference Detail – provides details of each specimen which includes the description of reference material, scientific name, unique laboratory number and sample mark, locality, authenticity, and if any, corresponding voucher specimen / CMM, and remarks.

- ◆ All reference DNA sequences come from three types of reference materials of known provenience: voucher specimens of organism, authenticated parts of organisms and authenticated CMM. Reference materials may be sourced from retention samples collected in Hong Kong Chinese Materia Medica Standards (HKCMMS) Project, GCMTI Herbarium, and recognized museums, national authorities, universities or institutes if appropriate.
 - ◆ Each specimen and medicinal material have been authenticated by experts, well documented and permanently curated in GCMTI. The authenticity of specimen, validated test method and DNA barcode data in fact sheet constitutes the foundation of genetic traceability for species identification, for the sake of ensuring efficacy and safety of clinical application of Chinese medicines being used originated from proper medicinal materials.
 - ◆ Two independent measurements for each reference material will be conducted to ensure the quality of reference DNA sequences being generated.
 - ◆ If feasible, the CMMRSL provides data of related voucher specimens and their medicinal materials, where were collected from the same geographic area. Such relationship will be indicated in the field “corresponding voucher specimen /CMM”.
 - ◆ Remarks will be made upon special findings in the data is observed.
 - Image – provides the photo(s) of the specimen.
- (c) Reference DNA sequences in FASTA format– an assembly of reference DNA sequences of a DNA barcode from all specimens in FASTA format. It provides a quick way to manipulate the reference DNA sequences for subsequent analysis.

Terms

- (a) Default DNA barcodes – referring to DNA barcodes chosen for kingdoms of organism
- For plant species, the default DNA barcodes include nuclear internal transcribed spacer 2 (ITS2), chloroplast *psbA-trnH* intergenic spacer (*psbA-trnH*), and chloroplast ribulose biphosphate carboxylase large chain (*rbcL*).

- For animal species, the default DNA barcodes include mitochondrial cytochrome c oxidase subunit I (COI), mitochondrial cytochrome b (CYTB) and mitochondrial 16S ribosomal RNA (16SrRNA).
- (b) FASTA format – referring to the format for nucleotide or amino acid sequences, which is a commonly used data format in the field of bioinformatics for sequence analysis.
- (c) Laboratory number – referring to unique identification sample code of specimen.
- (d) Nucleotide code – referring to code written in International Union of Pure and Applied Chemistry (IUPAC) as tabulated below:

IUPAC Nucleotide Code	Translation
A	Adenine
C	Cytosine
G	Guanine
T	Thymine
B	C, G or T
D	A, G or T
H	A, C or T
R	A or G (purines)
Y	C or T (pyrimidines)
K	G or T
M	A or C
S	G or C
W	A or T
N	A, C, G or T
V	A, C or G

- (e) Sample mark – referring to unique identification of sample to be tested. The last letter of sample mark represents the identity of replicate of specimen. For example, RD100-A and RD201-1-A denotes the replicate A of specimen with unique sample code of RD100 and RD201-1, respectively.
- (f) Sequence identifier – referring to name of a reference sequence in FASTA files starts with a ‘greater than’ sign (i.e. ‘>’). Sequence identifier consists of the sample mark and the name of default DNA barcode. For example, sequence identifier “>RD100-A_ITS2” represents the ITS2 reference DNA sequence determined from specimen of sample mark “RD100-A”.

Abbreviations and Symbols

bp	Base pair
CMM	Chinese materia medica
CMMRSL	Chinese Materia Medica Reference Sequence Library
COI	Cytochrome c oxidase subunit I
CYTB	Mitochondrial cytochrome b
DNA	Deoxyribonucleic acid
GCMTI	Government Chinese Medicines Testing Institute
Indel	Insertion or deletion
ITS2	Internal transcribed spacer 2
psbA-trnH	psbA-trnH intergenic spacer
rbcL	Ribulose biphosphate carboxylase large chain
16SrRNA	16S ribosomal RNA

Disclaimer for use of the Fact sheet

- (a) All information presented in the fact sheet are considered the property of the Government and may be used for personal, research, or educational purposes. The fact sheet may not be copyrighted for resales or commercial purposes.
- (b) The contents of the document is subject to change without notice. The GCMTI accepts no liability for any loss or damage howsoever arising from any use or misuse of or reliance on any information in this document.