

Drug Discovery and Development



# Accurate and Rapid DNA Barcoding Technique for Herbal and Botanical Products

• High Accuracy • Fast Turn Around Time • Streamlined Processes

Jane Luo and Handy Yowanto SCIEX, Brea, California, USA

### Introduction

The global herbal supplement market is estimated to reach \$107 billion dollars by 2017.1 Surveys indicated that commonly used medicinal plants belong to 11,146 species from 2,309 genera of 383 families.<sup>2</sup> Accurate and rapid authentication of these plants and their adulterants is very important for international trade and safe use of medicinal plants and adherence to regulatory compliance. While the conventional chemical analysis can not identify the genus and species of plants; compounding to the challenge, contaminants, mislabeling and adulteration of processed raw materials could post safety risks to consumers and economic loss to manufacturers. Certain conserved regions such as internal (ITS) and external (ETS) ribosomal DNA transcribed spacer regions have been used as markers within the nuclear genome for plant population and evolution studies. In this technical note, we describe a process for sequencing ITS and ETS regions of herb samples using the GenomeLab GeXP™ Genetic Analysis System that offers fast turnaround time (8 hrs) and high identification accuracy. Upon isolation of nucleic acid from each herb sample, the gene target is amplified using polymerase chain reaction (PCR), followed by Dye Terminator Cycle Sequencing (DTCS) on the GeXP. Results are compared against the National Center for Biotechnology Information (NCBI) plant sequence database to identify the genus and species of the plant.

## **Materials and Methods**

**Genomic DNA Isolation:** Dried herb samples were soaked in nuclease free water overnight and grounded with a small mortar and pestle. About 100 mg of the grounded sample was used for isolation of genomic DNA using Qiagen DNeasy Plant Mini Kit. DNA concentration was determined by a UV spectrophotometer.

*Target Gene Amplification and Post PCR Cleanup:* Amplification of ITS and ETS regions were performed as described by Sonnante et al.<sup>3</sup> Amplicons were purified using Agencourt AMPure XP or Qiaquick Gel Extraction Kit.

**Sequencing and Data Analysis:** Sanger Sequencing (Figure 2) was carried out per instructions for the DTCS Quick Start Kit from SCIEX. Sequencing fragments were purified using Agencourt CleanSeq. Separation and data analysis are fully automated on the GenomeLab GeXP Genetic Analysis System (Figure 1). Sequences were submitted to NCBI Database for BLAST search to identify the herb.

*Work Flow for Herb Identification:* The entire work flow for herb identification using GeXP is shown in Figure 3.



Figure 1. Basic operator steps for using the GenomeLab GeXP Genetic Analysis System.





Figure 2. Sanger Dye Terminator Cycle Sequencing.



Figure 3. Work Flow for Herb Identification using GenomeLab GeXP Genetic Analysis System.



# Results

Identification of an herbal plant sample as Momordica charantia (bitter melon): The ITS region was amplified from the genomic DNA isolated from the sample; and sequenced using DTCS Quick Start Kit. Sequencing fragments were purified by CleanSeq and separated on the GeXP instrument (Figure 4).



Figure 4. Sequencing results showing analyzed data, Quality Values and DNA sequence. The Quality Value (QV) is a measure of confidence for base calling. It relates to estimated error rate. The lower the error rate, the higher the Quality Value. QVs are useful in identifying the highest quality portions of reads for use in sequence assembly. Before the assembly, trimming using QV is preferred than other trimming methods since it provides longer readlength. During assembly, QV can be used to guide determining the consensus sequence.

After quality based trimming, sequence results were exported in .scf format; imported to GeneStudio for contig assembly

to generate consensus sequence (Figure 5).



Figure 5. Contig Assembly in GeneStudio to create consensus sequence through alignment of DNA sequences obtained in two independent sequencing reactions with the same sample. Consensus sequence was exported as a .txt file



Consensus sequence was entered in BLASTN as "Query Sequence" for BLAST search (Figure 6A). The search result (Figure 6B) indicated that the sample is identified as "Momordica charantia" (bitter melon).



Figure 6A. Blast search in NCBI database. Consensus sequence was copy- and-pasted as "Query Sequence". "Others (nr etc.)" was selected as the database for BLAST search. The search was initiated by clicking on "BLAST" button.

-	Alignments Download - GenBank Graphics Distance tree of results									
		Description	Max score	Total score	Query cover	E value	Ident	Accession		
٢		Momordica charantia voucher T0037 clone 2 18S ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene, and internal trans	1210	1210	100%	0.0	99%	GQ845154.1		
ſ		Momordica charantia voucher T0037 clone 3 18S ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene, and internal trans-	1205	1205	100%	0.0	99%	<u>GQ845155.1</u>		
E		Momordica charantia iselate shawpc0969118S ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene, and internal transcri	1201	1201	100%	0.0	99%	JN407451.1		
E		Momordica charantia var. abbreviata isolate HY 531 internal transcribed spacer 1, 5.8S ribosomal RNA gene, and internal transcribed spacer 2, complete sequence	1190	1190	98%	0.0	99%	AY606258.1		
Γ		Momordica charantia var. abbreviata isolate HY 5851 internal transcribed spacer 1, 5.8S ribosomal RNA gene, and internal transcribed spacer 2, complete sequence	1184	1184	98%	0.0	99%	AY606261.1		

Figure 6B. Blast search results showed the ITS sequence from sample is 99% identical to ITS of Momordica charantia.

#### Identification of an herbal plant sample as Silybum

marianum (milk thistle): Both the ITS region and the ETS

region of the sample were sequenced (Figure 7). BLAST search

(Figure 8B) both indicated that the sample is Silybum marianum

Analyzed Data

190

200

210

180



results with ITS sequence (Figure 8A) and ETS sequence (milk thistle).

160

3100 3200 3300 3400 3500 3600 3700 3800 3900

170





1	11	Alignments Download v GenBank Graphics Distance tree of results						
		Description	Max score	Total score	Query cover	E value	Ident	Accession
		Silyburn marianum 18S rRNA gene (partial), ITS1, 5.8S rRNA gene, ITS2, 25S rRNA gene (partial)	817	817	99%	0.0	99%	AJ831537.1
		Silyburn marianum internal transcribed spacer 1, 5.8 S ribosomal RNA gene, and internal transcribed spacer 2, complete sequence	795	795	95%	0.0	100%	AY914831.1

Figure 8A. BLAST search results with ITS sequence.

11	Alignments Download  GenBank Graphics Distance tree of results								
	Description	Max score	Total score	Query cover	E value	Ident	Accession		
	Silyburn marianum partial IGS and 18S rRNA gene, from Italy	606	606	100%	8e-170	98%	AM267320.1		

Figure 8B. BLAST search results with ETS sequence.

# Conclusion

DNA Sequencing targeting the ITS and ETS regions using SCIEX GenomeLab GeXP Genetic Analysis System is an accurate, simple, fast and reliable method to identify genus and species of many different herbal plants.

### References

- 1. Nutraceuticals World (2012) http://www.nutraceuticalsworld.com
- 2. Chen, et al., (2010) PLOS One 5(1) e8613.
- 3. Sonnante et., (2007) *Genet Resour Crop Evol* 54:483-495.

AB Sciex is doing business as SCIEX.

© 2016 AB Sciex. For research use only. Not for use in diagnostic procedures. The trademarks mentioned herein are the property of the AB Sciex Pte. Ltd. or their respective owners. AB SCIEX™ is being used under license.

RUO-MKT-02-3907 04/2016



Headquarters

500 Old Connecticut Path, Framingham, MA 01701, USA Phone 508-383-7800 www.sciex.com International Sales For our office locations please call the division headquarters or refer to our website at www.sciex.com/offices