



Research article

Converting *Panax ginseng* DNA and chemical fingerprints into two-dimensional barcode



Yong Cai^{1,2,☆}, Peng Li^{1,☆}, Xi-Wen Li^{1,3}, Jing Zhao¹, Hai Chen², Qing Yang⁴, Hao Hu^{1,4,*}

¹ State Key Laboratory of Quality Research in Chinese Medicine, Institute of Chinese Medical Sciences, University of Macau, Macao

² Information Technology College of Beijing Normal University Zhuhai Campus, Zhuhai City, China

³ Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing, China

⁴ State Key Laboratory of Hydraulics and Mountain River Engineering, Sichuan University, Chengdu, Sichuan, China

ARTICLE INFO

Article history:

Received 7 April 2016

Received in Revised form

22 June 2016

Accepted 29 June 2016

Available online 21 July 2016

Keywords:

internal transcribed spacer 2

Panax ginseng

quality traceability of traditional Chinese medicine

quick response code

traditional Chinese medicine chemical fingerprints

ABSTRACT

Background: In this study, we investigated how to convert the *Panax ginseng* DNA sequence code and chemical fingerprints into a two-dimensional code. In order to improve the compression efficiency, GATC2Bytes and digital merger compression algorithms are proposed.

Methods: HPLC chemical fingerprint data of 10 groups of *P. ginseng* from Northeast China and the internal transcribed spacer 2 (ITS2) sequence code as the DNA sequence code were ready for conversion. In order to convert such data into a two-dimensional code, the following six steps were performed: First, the chemical fingerprint characteristic data sets were obtained through the inflection filtering algorithm. Second, precompression processing of such data sets is undertaken. Third, precompression processing was undertaken with the *P. ginseng* DNA (ITS2) sequence codes. Fourth, the precompressed chemical fingerprint data and the DNA (ITS2) sequence code were combined in accordance with the set data format. Such combined data can be compressed by Zlib, an open source data compression algorithm. Finally, the compressed data generated a two-dimensional code called a quick response code (QR code). **Results:** Through the abovementioned converting process, it can be found that the number of bytes needed for storing *P. ginseng* chemical fingerprints and its DNA (ITS2) sequence code can be greatly reduced. After GTCA2Bytes algorithm processing, the ITS2 compression rate reaches 75% and the chemical fingerprint compression rate exceeds 99.65% via filtration and digital merger compression algorithm processing. Therefore, the overall compression ratio even exceeds 99.36%. The capacity of the formed QR code is around 0.5k, which can easily and successfully be read and identified by any smartphone.

Conclusion: *P. ginseng* chemical fingerprints and its DNA (ITS2) sequence code can form a QR code after data processing, and therefore the QR code can be a perfect carrier of the authenticity and quality of *P. ginseng* information. This study provides a theoretical basis for the development of a quality traceability system of traditional Chinese medicine based on a two-dimensional code.

© 2016 The Korean Society of Ginseng, Published by Elsevier Korea LLC. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Panax ginseng is a worldwide popular herbal medicine, which has been used in traditional Asian medicine for thousands of years. Since the quality of *P. ginseng* depends on many conditions and factors, the most fundamental problem to be solved is the authenticity and evaluation model of quality level issue, which involves a lot of information. Therefore, it is essential to design a

very simple and easy information carrier to carry the related quality information. The current use of DNA sequences to identify true and false methods has been widely recognized [1,2]. Meanwhile, chemical fingerprints form the industry's widely accepted evaluation model of quality level [3–7]. If these two kinds of *P. ginseng* information can be represented by the current popular two-dimensional code information carrier, the final consumer can easily trace the quality of *P. ginseng* using a smartphone, which will

* Corresponding author. State Key Laboratory of Quality Research in Chinese Medicine, Institute of Chinese Medical Sciences, University of Macau, Macao.

E-mail address: haohu@umac.mo (H. Hu).

☆ Yong Cai and Peng Li contributed equally to this work.

promote the development and application of a *P. ginseng* quality traceability system.

More convenient traditional Chinese medicine quality information carrier has been studied. Chen et al [8] found that the internal transcribed spacer 2 (ITS2) sequence code is more appropriate for identifying medicinal plant species. Liu et al [9] studied the conversion of the Chinese herbal medicine DNA sequence into a quick response code (QR code). Kumar et al [10] studied the conversion of DNA barcoding into PDF417. Cai et al [11] studied the conversion of Chinese medicine chemical fingerprints into a QR code. However, the two-dimensional code storage space is very limited; the original data of the normal *P. ginseng* DNA sequence and chemical fingerprints far exceed the capacity of the current two-dimensional codes. Therefore, compression algorithms for *P. ginseng* DNA sequence and chemical fingerprint data are very important. Many methods are used for compression of DNA sequences; some effective methods are BioCompress-2 [12], Gen Compress [13], CTW + LZ [14], DNA Compress [15], DNAPack [16], DNADP [17], GeNML [18], and so on. Currently, the main processing step for chemical fingerprints is to find the inflection point in the filtering process [11]. The subsequent compression algorithm is to be studied in the future.

Based on the existing studies of conversion of traditional Chinese medicine chemical fingerprints into two-dimensional codes [11], in this study, we explored a novel way to further convert the *P. ginseng* DNA (ITS2) sequence code and chemical fingerprints into a QR code to record authenticity information and quality information of ginseng.

2. Materials and methods

2.1. Materials

This study was carried out on HPLC chemical fingerprint data of 10 groups (Agilent rapid detection method) of roots and fibrous roots of ginseng from Northeast China, grown for 2–6 years. The ITS2 sequence code was selected as the DNA sequence code for its species. Moreover, HPLC chemical fingerprint data of *P. ginseng* were obtained from the State Key Laboratory of Quality Research in Chinese Medicine, University of Macau. The ITS2 sequence code data were acquired from the Chinese pharmacopoeia medicine DNA barcoding standard sequence [8,19]. Meanwhile, Cai et al [11] found the QR code very suitable for carrying chemical fingerprints of traditional Chinese medicine by comparing the popular two-dimensional code QR code, data matrix, PDF 417, and other formats. In this study, since we only need add DNA (ITS2) sequence code on the basis of the original, there is no need to choose any other two-dimensional code. Therefore, the QR code was suitable for use as our target two-dimensional code.

In this study, data processing was completed with Ruby language (v1.9.3 p551), the Excel spreadsheet was operated by a third-party plug-in named Gem spreadsheet 0.9.8, and the QR code was created by a third-party plug-in named Gem qrqr 0.7.0. The program could run on the operating system of Windows XP/Windows 7/Windows 8. Fig. 1 shows a flowchart of whole conversion processes.

2.2. Six steps of two-dimensional code conversion

2.2.1. Step 1: Obtaining chemical fingerprint data set of key feature points

The original data of *P. ginseng* chemical fingerprints were obtained from Agilent equipment output file, and the original

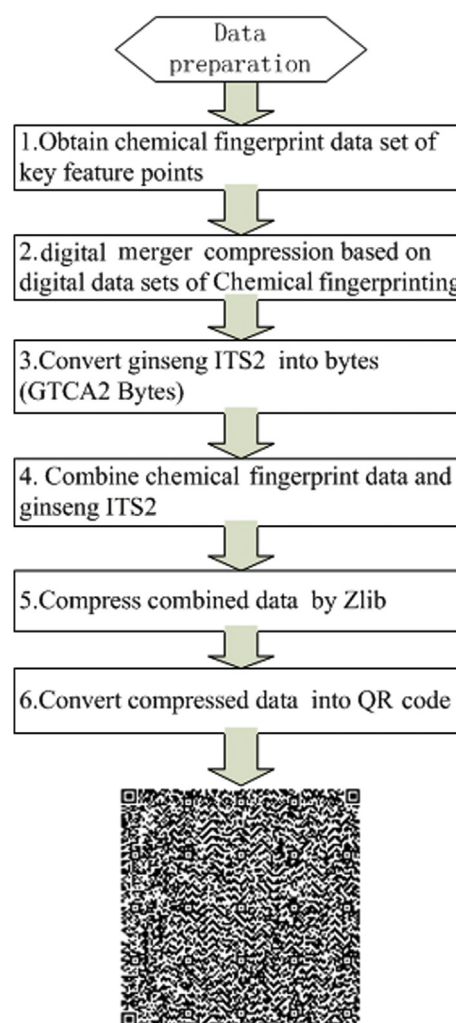


Fig. 1. Flowchart of conversion processes. ITS2, internal transcribed spacer 2; QR, quick response.

contents of the file are shown in Fig. 2 (test results of the fibrous roots of a 4-year-old ginseng).

The key data set (time–value) of chemical fingerprints in the output file was saved in an Excel file format due to the text-formatted output file. “Time” represents duration of detection in minutes, while “value” represents detected absorption values in that minute. The number of data points was about 5,000 or so (about 104K bytes), which is beyond the capacity of the current variety of any two-dimensional code format. Thus, the data needed to be filtered before being converted into a two-dimensional code. According to the selected inflection point filtering algorithm [11], we needed to collect the data of the key feature points (inflection points) of *P. ginseng* chemical fingerprints. Inflection point features are shown in Fig. 3 and inflection point judgment conditions in Fig. 4. “Time” represents the detection duration in minutes and “value” represents the absorbance value corresponding to the detected time.

Compared with the filtering algorithm proposed earlier [11], this study proposed mainly the following two improvements:

- (1) Time accuracy was up to two decimal places, while time accuracy had been up to one decimal place in the original paper.

File Path chrom://icmsqc/YYX/RS.seq/724.smp/ CAD_1_channel Channel CAD_1	Signal Parameter Information: Signal Info
Injection Information: Data Vault YYX Injection RS-4-XU Injection Number 13 Position GB3 Comment Processing Method 11 Instrument Method RS-S Type Unknown Status Finished Injection Date 2015/9/9 Injection Time 4:16:20 Injection Volume (μl) 1.000 Dilution Factor 1.0000 Weight 1.0000	Raw Data: Time (min) Step (s) Value (pA) 0.000000 n.a. -0.015466 0.003333 0.2 -0.014959 0.006667 0.2 -0.015197 0.010000 0.2 -0.014005 0.013333 0.2 -0.013648 0.016667 0.2 -0.015078 0.020000 0.2 -0.017075 0.023333 0.2 -0.018237 0.026667 0.2 -0.018297 0.030000 0.2 -0.018505 0.033333 0.2 -0.017641 0.036667 0.2 -0.017403 0.040000 0.2 -0.019340 0.043333 0.2 -0.018565 0.046667 0.2 -0.017492 0.050000 0.2 -0.018207 0.053333 0.2 -0.017969 0.056667 0.2 -0.017403 0.060000 0.2 -0.016807 0.063333 0.2 -0.015585 0.066667 0.2 -0.015227 0.070000 0.2 -0.017582 0.073333 0.2 -0.017432 0.076667 0.2 -0.015883 0.080000 0.2 -0.013767 0.083333 0.2 -0.012545 0.086667 0.2 -0.011831 0.090000 0.2 -0.011771 0.093333 0.2 -0.015466 0.096667 0.2 -0.018327 0.100000 0.2 -0.017432 0.103333 0.2 -0.015615
Raw Data Information: Time Min. (min) 0.000000 Time Max. (min) 15.002900 Data Points 4499 Detector CAD Generating Data System Chromeleon 7.2.1.5833 Exporting Data System Chromeleon 7.2.1.0 Operator Instrument Controller Signal Quantity Current Signal Unit pA Signal Min. -0.019340 Signal Max. 64.224269 Channel CAD_1 Driver Name Corona.dll Channel Type Evaluation Min. Step (s) 0.197 Max. Step (s) 0.21	

Fig. 2. Part of the original data from 4-year-old *P. ginseng* fibrous root test results (Agilent rapid detection method).

- (2) Within a given time accuracy (0.6 s) range, the time point corresponding to the greatest change of value was selected as the time point characteristic value in this region, while the first point within the range had been selected directly in the original paper.

Table 1 lists the data results after implementing the filtering algorithm.

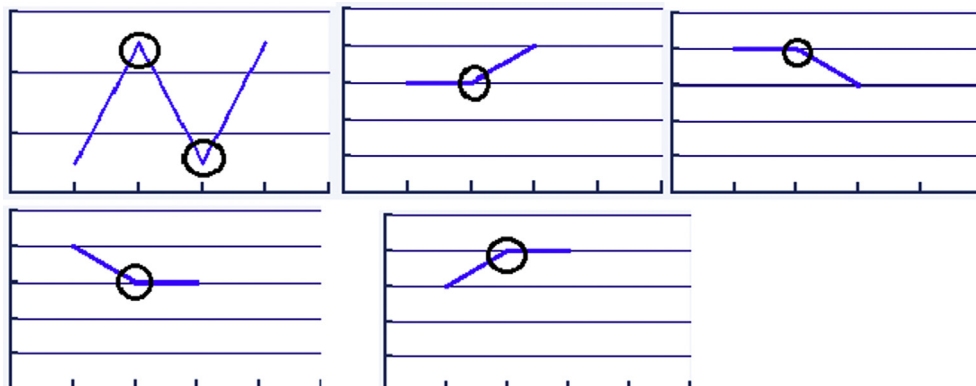


Fig. 3. Judgment conditions of inflection point (X axis → time; Y axis → value). Points inside a black circle will be considered inflection points; there are five possible conditions.

2.2.2. Step 2: Digital merger compression based on digital data sets of chemical fingerprinting

Digital merger compression (DMC) was proposed due to the digital feature of chemical fingerprints. Time data and mAu data were amplified 100 times and 10 times, respectively, in order to obtain integers. Then the data could be combined into an array of 2 bytes, 3 bytes, or 4 bytes according to its size. More details of the algorithm are shown in Fig. 5.

Using the DMC algorithm, data sets of chemical fingerprinting completed the precompression processing. Table 2 lists the data changes after running the DMC algorithm.

2.2.3. Step 3: Conversion of ginseng DNA (ITS2) sequences code into bytes (GTCA2 bytes)

The *P. ginseng* DNA (ITS2) sequence code was 230 bytes when we chose 70 characters per line breaks, as shown in Fig. 6.

Each of the four bases (G, T, C, and A) was represented by 2 bits (00, 01, 10, and 11, respectively) according to the features of *P. ginseng* DNA (ITS2) sequence code. The first two bytes were used to save the total length of the sequence code within 65,535 base sequence. Therefore, it became a byte algorithm (GTCA2Bytes) by transition, as shown in Fig. 7.

2.2.4. Step 4: Chemical fingerprint data and DNA (ITS2) sequence code data combined by a predefined data format

To ensure that the converted two-dimensional code of traditional Chinese medicine chemical fingerprints can be identified and reproduced, the compressed array of bytes needs to be combined in accordance with the set data format. The available common standard formats are XML, JSON, etc. However, it requires additional bytes, which is not proper for the two-dimensional code due to its limited data capacity. Therefore, a combination of simpler rules was adopted in this study.

The compressed byte array of DNA (ITS2) sequence and byte array of chemical footprints were combined with the following format, in which “||” was chosen to separate arrays, shown as follows:

2 bytes array of ITS2 sequence || 2 bytes array of chemical fingerprint data || 3 bytes array of chemical fingerprint data || 4 bytes array of chemical fingerprint data

2.2.5. Step 5: Compression of combined data by Zlib

The combined data needed to be compressed in advance in order to get less data. Many algorithms of compression are available, which are divided into two major categories of lossy and lossless compression. Owing to nondestructive reduction, the most common Zlib (DEFLATE RFC 1951, variation of the LZ77 algorithm [20,21]) lossless compression algorithm was suitable for this study.

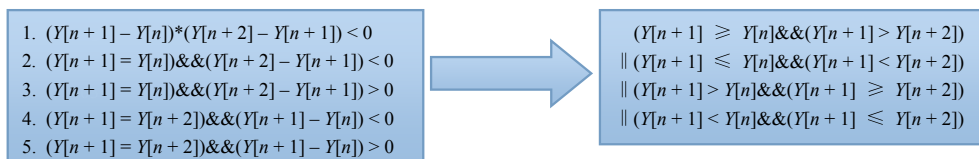


Fig. 4. Inflection point selection algorithm ($Y [] \rightarrow$ value of point; $n \rightarrow$ time of point), according to the five possible conditions shown in Fig. 2.

Table 1
Data changes of ginseng chemical fingerprint after adopting the filtering algorithm

Sample dataset	Raw source (bytes/points)	After filter (bytes/points)	Compression ratio [bytes (%)/points (%)]
xu-2.xls	79,961/4,498	1,548/166	1.93/3.69
zhu-2.xls	81,000/4,498	1,184/129	1.46/2.86
xu-3.xls	79,719/4,499	1,449/156	1.81/3.46
zhu-3.xls	80,839/4,498	1,191/130	1.47/2.89
xu-4.xls	79,128/4,499	1,515/162	1.91/3.6
zhu-4.xls	81,056/4,499	1,113/121	1.37/2.68
xu-5.xls	79,797/4,499	1,667/178	2.08/3.95
zhu-5.xls	80,274/4,498	1,303/142	1.62/3.15
xu-6.xls	79,754/4,498	1,525/163	1.91/3.62
zhu-6.xls	80,817/4,499	1,263/137	1.56/3.04

Compared with other compression algorithms, this compression algorithm has the advantage of high efficiency, free and open source, and so on. Table 3 lists the data changes after adopting the Zlib compression algorithm.

2.2.6. Step 6: Conversion of the data compressed into QR code

After Zlib compressed, data generated QR code by rqr code gem, where the parameters are as follows:

- (1) Error correcting code: L-smallest
- (2) Size: 25

3. Results

Through the above six steps, all batches of *P. ginseng* DNA (ITS2) sequence and its HPLC chemical fingerprints were successfully transformed into a QR code. A very important purpose of the conversion process is to reduce the number of bytes of the original data; in this study, four processes were used to reduce the number of data bytes. The results of the four processes are analyzed as follows.

The first process was a screening process, corresponding to Step 1. The chemical fingerprint features of data sets were screened through the inflection point filtering algorithm. Screening results are given in Table 1. After screening, the length of the obtained data string was in the range of 1,113–1,667, a reduction of 1.37–2.08% compared with the original data string length. The data points were in the range of 121–178, corresponding to a reduction of 2.68–3.95% in the number of the original data points.

The second process involved the DMC algorithm, corresponding to Step 2. The characteristics of *P. ginseng* chemical fingerprint data set were compressed. Table 2 lists the differences of the *P. ginseng* chemical fingerprint characteristics before and after implementing the DMC algorithm. The compression rate of byte numbers was between 30.23% and 31.73%, which was between 0.42% and 0.65% of the original data.

The third process involving the GTCA2Bytes algorithm, which corresponded to Step 3, was a precompression algorithm for the *P. ginseng* DNA (ITS2) sequence code. Through the algorithm, the

P. ginseng DNA (ITS2) sequence code could be compressed easily; the byte count could be reduced to nearly 75%.

The fourth process involved the Zlib compression algorithm, corresponding to Step 5. Table 3 shows the changes in the number of bytes after the implementation of the Zlib compression algorithm. Compared with the number of data bytes before compression, the compression ratio was between 87.2% and 94.85%, while compared with the original data, the compression rate was between 0.48% and 0.64%.

After the six-step process, the *P. ginseng* DNA (ITS2) sequence code and chemical fingerprints were been successfully converted into a QR code. The results of the operation are shown in Figs. 8 and 9 for the selected four groups of data.

By converting the 10 groups of data into a QR code, we could find that the number of bytes for ginseng chemical fingerprints and DNA (ITS2) sequence code was significantly reduced after treatment. The DNA (ITS2) sequence code compression rate could reach 75%. After filtering the chemical fingerprints and DMC algorithm processing, the compression rate could be more than 99.65%. Finally, the overall compression ratio was over 99.36%, and the capacity of the QR code was about 0.5 kilobytes (KB).

4. Discussion

In this study, on the basis of the conversion of the DNA (ITS2) sequence code [9,10] as well as the chemical fingerprints into a two-dimensional code [11], further research was carried out. The *P. ginseng* DNA (ITS2) sequence code and HPLC chemical fingerprinting were converted into a two-dimensional code in accordance with certain procedures and formats. GATC2Bytes and DMC algorithms were used in this conversion process. After compression by Zlib, QR code bytes were about 0.5 KB, which is far less than converting chemical fingerprint QR code method proposed by Cai et al [11]. Using their method, the number of bytes in the two-dimensional code was found to be between 1 KB and 2 KB. Compared the QR code conversion method with theirs, DMC and Zlib compression processing made the number of bytes of QR code less nearly between 50% and 75%.

Many DNA sequences compression algorithm are available [12–18]; most of them have been proposed to solve the issue of DNA sequence, which is very large—more than 100 million petabytes [22]. In this study, the simple GATC2Bytes compression algorithm was adopted due to the small DNA sequence code of *P. ginseng*, which is less than 1 kb. After processing, the number of bytes can be reduced by 75%, which can be improved further. We promote a novel method to compress the chemical fingerprint code by the DMC algorithm. Since no similar literature reports are available, further research is required.

Furthermore, the converted two-dimensional code can be recognized by a computer and be restored to the original DNA (ITS2) sequence and chemical fingerprint data, the process of which has been proved to be reversible.

Traditional Chinese medicine chemistry (composition) fingerprinting methods are mainly based on spectroscopy and

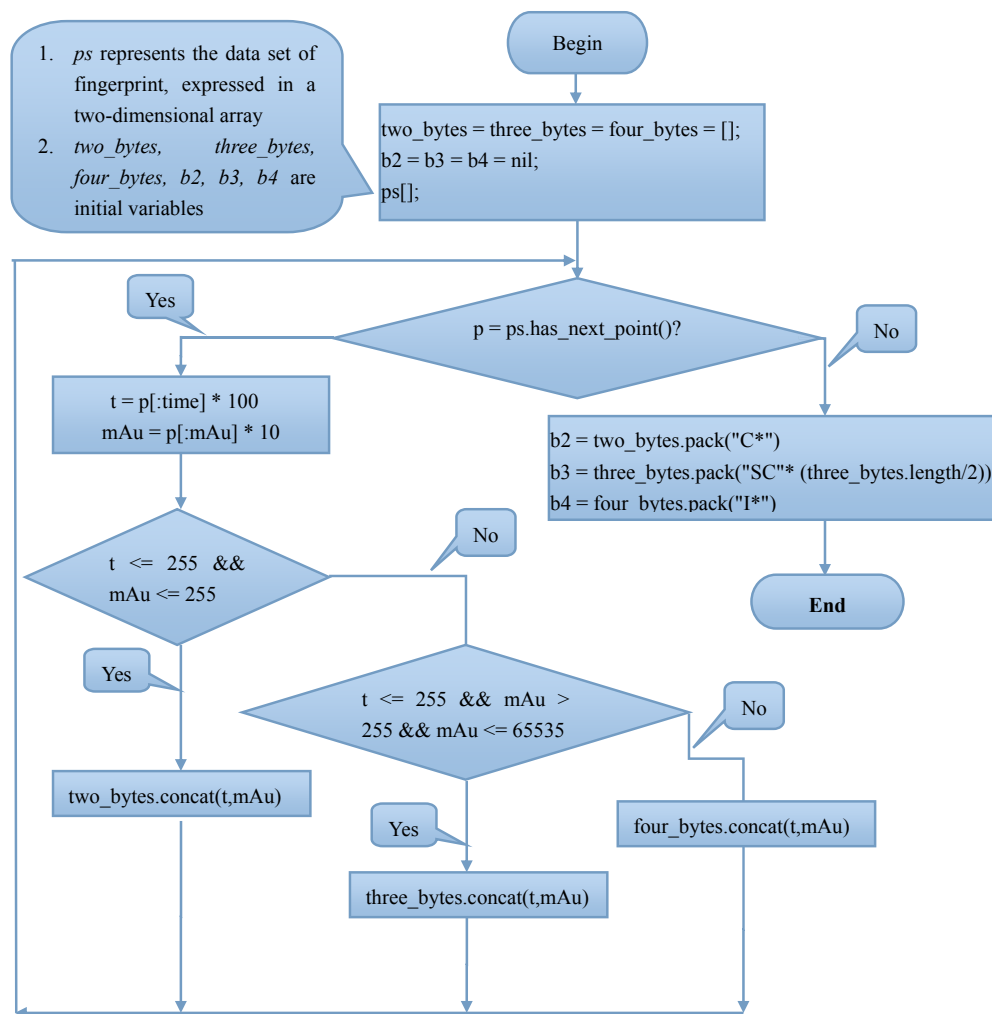


Fig. 5. DMC algorithm, in which time data and mAu data are amplified 100 times and 10 times, respectively, in order to obtain integers. Then the data can be combined into an array of 2 bytes, 3 bytes, or 4 bytes according to its size. With this algorithm, data sets of chemical fingerprinting complete the precompression processing. DMC, digital merger compression.

Table 2
Data changes after adopting the DMC algorithm

Sample dataset	Raw source (bytes)	After filter (bytes)	After DMC (bytes)	Compression ratio [bytes (%)]
xu-2.xls	79,961	1,548	476	30.75
zhu-2.xls	81,000	1,184	364	30.74
xu-3.xls	79,719	1,449	448	30.92
zhu-3.xls	80,839	1,191	371	31.15
xu-4.xls	79,128	1,515	458	30.23
zhu-4.xls	81,056	1,113	342	30.73
xu-5.xls	79,797	1,667	520	31.19
zhu-5.xls	80,274	1,303	407	31.24
xu-6.xls	79,754	1,525	462	30.3
zhu-6.xls	80,817	1,263	389	30.8

DMC, digital merger compression.

chromatography: UV, IR, MS, NMR, TLC, GC, HPLC, capillary electrophoresis, and so on. Among them, HPLC has become the primary means of fingerprint research currently. This study focuses on the conversion of HPLC and DNA (ITS2) sequence, due to its chemical fingerprint data form and DNA (ITS2) sequence have many similarities, so six steps conversion method should be applicable to some of the traditional Chinese medicine.

```
CGCATCGCGTCGCCCCCAACCCATCACTCCCTTGCGGGAGTTGAGGCGGAGGGGGCGGATAATGGCCTCC
CGTGTCTACCCGCGGGTTGGCCAAATGCGAGTCTTGCGCATGGACGTCACGACAGTGGTGGTTGTA
AAAAGCCCTCTTCTCATGTCTGTGCGGTGACCCGTCGCCAGCAAAAAGCTCTCATGACCTGTGCGCGGTC
CTCGACGTGCGCTCCGACCG
```

Fig. 6. *P. ginseng* DNA (ITS2) sequence code. ITS2, internal transcribed spacer 2.

During the experiment, we found that how to compress the space was an important challenge. DNA (ITS2) may be replaced by other ways (for example, SNP) to reduce space. In addition, recording only the major chemical fingerprints is a very good idea for space compression. However, it will cause problems with incomplete information [23,24]. Therefore, we finally chose to use DNA (ITS2) and complete chemical fingerprints for this experiment.

Time complexity should be considered within this experiment. If the time complexity cannot be reduced, the experiment does not have practical significance. As shown in Table 4, 10 sets of data conversion processes were tested (Computer Environment: Sony VAIO with window XP sp3, 3G memory, dual-core, Intel CPU: 1.8GHz). We found that the average total time for the conversion of 10 sets of data was 11.675 s, which was within the acceptable range.

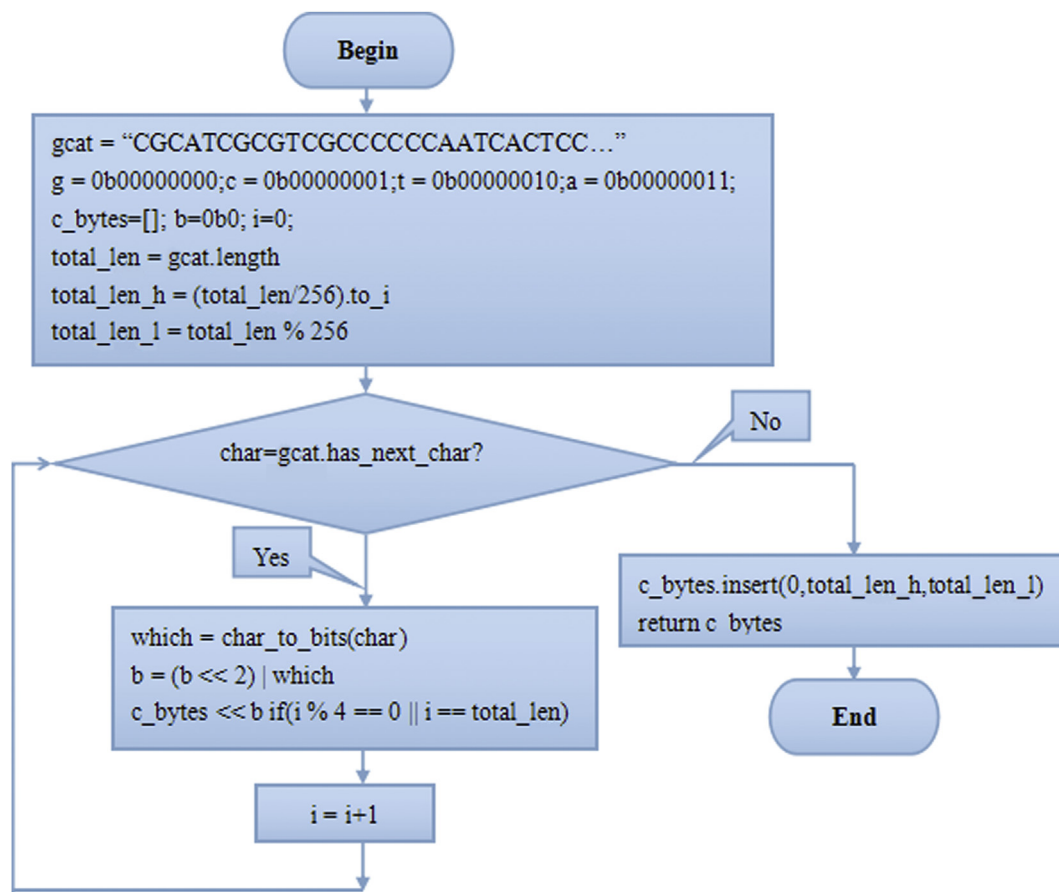


Fig. 7. GTCA2Bytes algorithm, to compress *P. ginseng* DNA (ITS2) sequence code to bytes. ITS2, internal transcribed spacer 2.

Table 3
Data changes after adopting the Zlib compression algorithm

Sample	Raw source (bytes)	Before Zlib (bytes)	After Zlib (bytes)	Zlib compression ratio [bytes (%)]	Total compression ratio [bytes (%)]
xu-2	80,194	542	483	89.11	0.60
zhu-2	81,233	430	395	91.86	0.49
xu-3	79,952	514	471	91.63	0.59
zhu-3	81,072	437	400	91.53	0.49
xu-4	79,361	524	480	91.60	0.6
zhu-4	81,289	408	387	94.85	0.48
xu-5	80,030	586	511	87.20	0.64
zhu-5	80,507	473	423	89.43	0.53
xu-6	79,987	528	479	90.72	0.60
zhu-6	81,050	455	416	91.43	0.51

Conversion of *P. ginseng* DNA (ITS2) sequence and chemical fingerprints into a two-dimensional code can be explored further. While a QR code can carry a maximum of 2,953 bytes, our experiment used only about 500 bytes, leaving a large amount space to carry more related information. If the *P. ginseng* DNA (ITS2) sequence and chemical fingerprints can be converted into a two-dimensional code successfully and easily, then it will be simple and convenient for a customer to trace the *P. ginseng* quality by a smartphone. Therefore, this method is a solid foundation of a *P. ginseng* quality traceability system.

4.1. Limitation and future research

This paper presents only the conversion of the *P. ginseng* DNA (ITS2) sequence and the HPLC chemical fingerprint data of 10

groups detected by Ailent Quick test into a two-dimensional code. Whether the DNA and chemical fingerprints of other medicines detected by other instruments can be smoothly converted into a two-dimensional code requires further research.

By analyzing the above discovery process, *P. ginseng* fingerprints and Chinese medicine DNA (ITS2) sequences must be filtered and must undergo compression processing before being converted into a two-dimensional code. In the present study, there should be room for optimization in the second-step HPLC chemical fingerprint compression (DMC) algorithm and third-step DNA (ITS2) sequence compression algorithm (GTCA2Bytes). Furthermore, the fourth-step selection of the data format combinations “||” for separation, there need to be perfected.

5. Conclusions

This study presents six steps to convert the *P. ginseng* DNA (ITS2) sequence code and chemical fingerprints into a QR code. In order to improve the compression ratio, this study proposes a simple GATC2Bytes compression algorithm for the compression of DNA (ITS2) sequence code and a DMC preprocessing algorithm for early treatment of chemical fingerprints. This study shows that the *P. ginseng* DNA (ITS2) and the chemical fingerprint information can be stored in a two-dimensional code. This study provides a theoretical basis for building a *P. ginseng* quality tracing system based on the two-dimensional code that contains quality information.

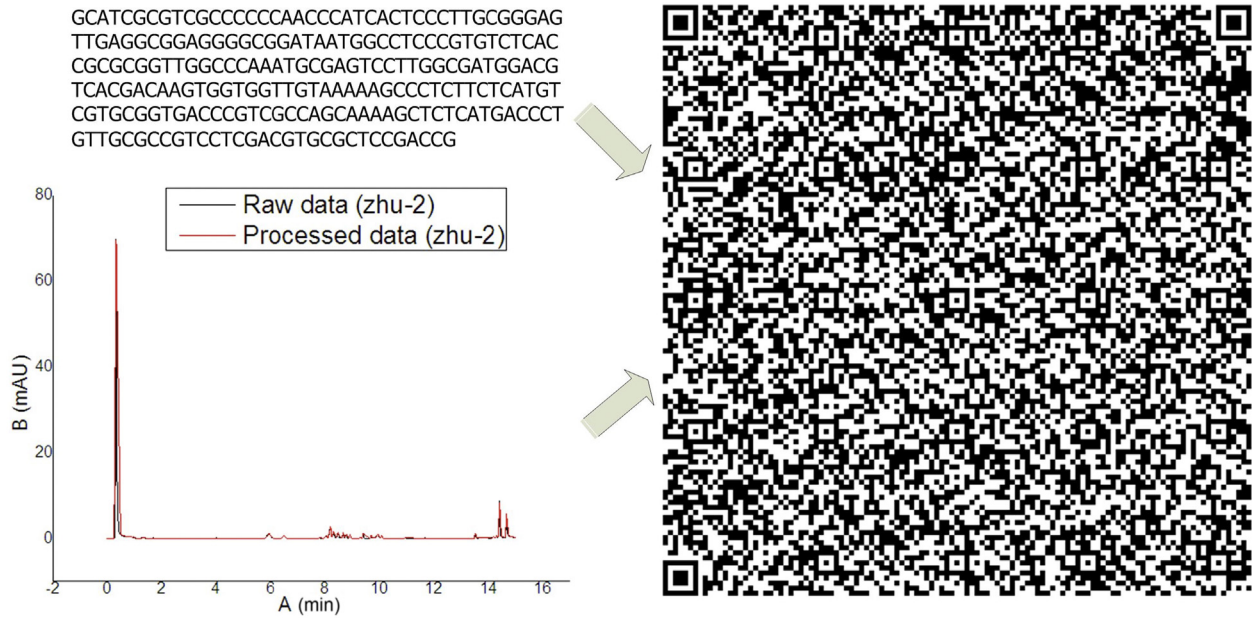


Fig. 8. Conversion result of zhu-2, *P. ginseng* DNA (ITS2), and chemical fingerprint of zhu-2 to a QR code. QR, quick response.

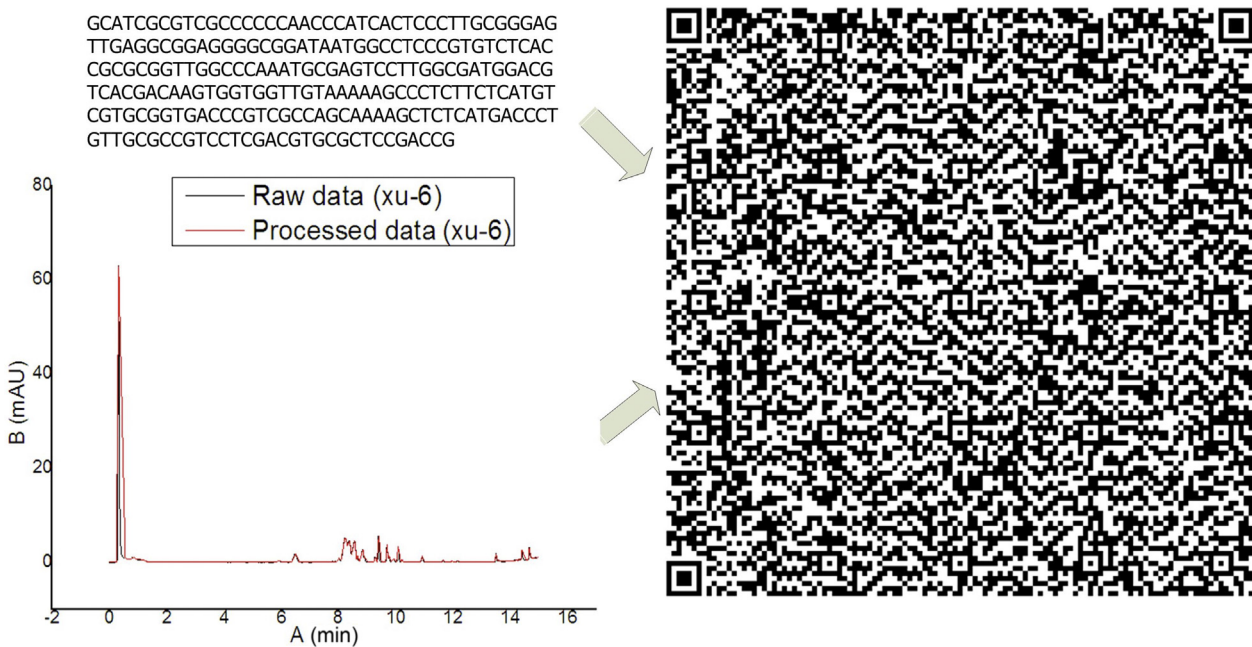


Fig. 9. Conversion result of xu-6, *P. ginseng* DNA (ITS2), and chemical fingerprint of xu-6 to a QR code. QR, quick response.

Table 4

Time spent in each step (unit: s)

Sample	Filter points	DMC	GTCA2 bytes	Data combination	Zlib	Created QR code	Total
xu-2	9.125	0.016	0.0	0.0	0.031	1.578	10.750
zhu-2	5.718	0.047	0.0	0.0	0.0	1.594	7.359
xu-3	17.063	0.047	0.0	0.0	0.0	1.594	18.704
zhu-3	6.593	0.047	0.0	0.0	0.0	1.563	7.903
xu-4	16.218	0.063	0.0	0.0	0.016	1.594	16.307
zhu-4	5.046	0.031	0.0	0.0	0.0	1.563	7.640
xu-5	15.171	0.047	0.0	0.0	0.0	1.594	16.822
zhu-5	7.375	0.016	0.0	0.0	0.0	1.578	8.969
xu-6	14.375	0.031	0.0	0.0	0.0	1.609	16.015
zhu-6	7.359	0.016	0.0	0.0	0.0	1.609	8.984
Average	10.405	0.036	0.0	0.0	0.005	1.589	11.675

Conflicts of interest

The authors declare that they have no competing interests.

Acknowledgments

This research was supported by the funding of Hong Kong, Macao & Taiwan Science and Technology Cooperation Project (2015DFM30030), ICM basic research funding (ZZ2014029), and Open Foundation of State Key Laboratory of Hydraulics and Mountain River Engineering, Sichuan University (SKHL1419).

References

- [1] Luo K, Chen S, Chen K, Song J, Yao H, Ma X, Zhu Y, Pang X, Yu H, Li X, et al. Assessment of candidate plant DNA barcodes using the Rutaceae family. *Sci China Life Sci* 2010;53:701–8.
- [2] Selvaraj D, Sarma RK, Shanmughanandhan D, Srinivasan R, Ramalingam S. Evaluation of DNA barcode candidates for the discrimination of the large plant family Apocynaceae. *Plant Syst Evol* 2015;301:1263–73.
- [3] Yan SK, Xin WF, Luo GA, Wang YM, Cheng YY. An approach to develop two-dimensional fingerprint for the quality control of Qingkailing injection by high-performance liquid chromatography with diode array detection. *J Chromatogr A* 2005;1090:90–7.
- [4] Xie YY, Luo D, Cheng YJ, Ma JF, Wang YM, Liang QL, Luo GA. Steaming-induced chemical transformations and holistic quality assessment of red ginseng derived from *Panax ginseng* by means of HPLC–ESI–MS/MS n-based multi-component quantification fingerprint. *J Agric Food Chem* 2012;60:8213–24.
- [5] Sun TT, Liang XL, Zhu HY, Peng XL, Guo XJ, Zhao LS. Rapid separation and identification of 31 major saponins in Shizhu ginseng by ultra-high performance liquid chromatography–electron spray ionization–MS/MS. *J Ginseng Res* 2016;40:220–8.
- [6] Lu GH, Chan K, Liang YZ, Leung K, Chan CL, Jiang ZH, Zhao ZZ. Development of high-performance liquid chromatographic fingerprints for distinguishing Chinese *Angelica* from related umbelliferae herbs. *J Chromatogr A* 2005;1073:383–92.
- [7] Liu L, Wang Y, Song Q, Bao YP. Fingerprint identification system based on two-dimensional barcode and DSP. *Adv Mater Res* 2012;479:2082–5.
- [8] Chen SL, Yao H, Han JP, Liu C, Song JY, Shi LC, Zhu YJ, Ma XY, Gao T, Pang XH, et al. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PLoS One* 2010;5:e8613.
- [9] Liu C, Shi L, Xu X, Li H, Xing H, Liang D, Jiang K, Pang X, Song J, Chen S. DNA barcode goes two-dimensions: DNA QR code web server. *PLoS One* 2012;7:e35146.
- [10] Kumar NP, Rajavel A, Jambulingam P. Application of PDF417 symbology for 'DNA barcoding'. *Comput Meth Prog Biomed* 2008;90:187–9.
- [11] Cai Y, Li XW, Li M, Chen XJ, Hu H, Ni JY, Wang YT. Traceability and quality control in traditional Chinese medicine: from chemical fingerprint to two-dimensional barcode. *Evid Based Complement Altern Med* 2015, 251304. <http://dx.doi.org/10.1155/2015/251304>. 6 pages.
- [12] Grumbach S, Taheri F. A new challenge for compression algorithms: genetic sequences. *Inform Process Manag* 1994;30:875–86.
- [13] Chen X, Kwong S, Li M. A compression algorithm for DNA sequences. *IEEE Eng Med Biol* 2010;20:61–6.
- [14] Matsumoto T, Sadakane K, Imai H. Biological sequence compression algorithms. *Genome Inform* 2000;11:43–52.
- [15] Chen X, Li M, Ma B, Tromp J. DNA Compress: fast and effective DNA sequence compression. *Bioinformatics* 2002;18:1696–8.
- [16] Behzadi B, Le Fessant F. DNA compression challenge revisited: a dynamic programming approach. *Comb Pattern Match* 2005;3537:190–200.
- [17] Srinivasa KG, Jagadish M, Venugopal KR, Patnaik LM. Efficient compression of non-repetitive DNA sequences using dynamic programming. In: *Advanced Computing and Communications, ADCOM 2006, International Conference on IEEE*; 2006. p. 569–74.
- [18] Korodi G, Tabus I. Normalized maximum likelihood model of order-1 for the compression of DNA sequences. In: *Data Compression Conference, 2007. DCC'07. IEEE*; 2007. p. 33–42.
- [19] Chen SL. Standard DNA barcodes of Chinese materia medica in Chinese pharmacopoeia, Volume 3. Beijing: Science Press; 2015. p. 473–5.
- [20] Kreft S, Navarro G. Self-indexing based on LZ77. *CPM* 2011;11:41–54.
- [21] Deutsch P, Gailly JL. Zlib compressed data format specification version 3.3 (No. RFC 1950). RFC 1950, May.
- [22] Galperin MY, Cochrane GR. Petabyte-scale innovations at the European nucleotide archive. *Nucl Acids Res* 2009;37:D1–4.
- [23] Xie PS, Leung AY. Understanding the traditional aspect of Chinese medicine in order to achieve meaningful quality control of Chinese materia medica. *J Chromatogr A* 2009;1216:1933–40.
- [24] Liang Y, Xie P, Chan K. Perspective of chemical fingerprinting of Chinese herbs. *Planta Med* 2010;76:1997–2003.