Mapping the potential distribution of high artemisinin-yielding Artemisia annua L. (Qinghao) in China with a geographic information system

Linfang Huang, Caixiang Xie, Baozhong Duan and Shilin Chen*

Abstract
Background: Artemisia annua L. is an important source for artemisinin, a potent drug for treating malaria. This study aims to map and predict the potential geographic distribution of A. annua L. in China.

Methods: The Geographic Information System for traditional Chinese medicine (TCM-GIS) was developed and used to map the potential geographic distribution of A. annua L.

Results: Climatic, edaphic and topographic characteristics of A. annua L. microhabitats in Youyang County were mapped to find distribution patterns. The maps identified that certain habitats in the Chongqing region and some potential regions, especially in Guizhou Province, possess similarity indices of ≥98%. In particular, high quality microhabitats A. annua L. were found in the Wuling mountains region.

Conclusion: The present study demonstrates a GIS approach to predict potential habitats for A. annua L. TCM-GIS is a powerful tool for assessing bioclimatic suitability for medicinal plants.

Background
Artemisia annua L. (Qinghao, Annual Wormwood) is a strongly fragrant, annual herbaceous plant used in Chinese medicine [1]. A. annua L. is the only natural botanical source for artemisinin (Qinghaosu) [2,3] and a potential source for essential oils for the perfume industry [4]. A. annua L. is now cultivated in China, Vietnam, India, Romania, Kenya and Tanzania [5]. Artemisinin, an endoperoxide sesquiterpene lactone in the aerial parts of A. annua L., is more efficacious, faster and less toxic than chloroquine in treating malaria. In addition, artemisinin is a potent anti-cancer agent, a possible antibacterial agent as well as a natural pesticide [6,7]. Chemical and biological synthesis of artemisinin is still under development due to poor yields [8-11]. Therefore, wild or cultivated A. annua L. is a major source for artemisinin [2,3,12].

The artemisinin content is highly dependent on plant ecotypes, ecological interactions, seasonal and geographic variations [13-18]. In fact, artemisinin is absent in some A. annua L. Artemisinin was first isolated in China and some Chinese germplasm has relatively higher artemisinin levels than those of Europe, North America, East Africa and Australia [2,13,16,17,19,20]. In Youyang County, Chongqing, China, the hometown of A. annua L., the plants have high (0.9%) levels of artemisinin. In 2006 the county became a national protected geographic area recognized by the General Administration of Quality Supervision, Inspection and Quarantine of China [21]. As the demand for artemisinin remains high around the world, finding suitable geographic regions for A. annua L. is a critical research area for the World Health Organization [22].

The geographic information system (GIS) technology manages geographic information with applications for various fields such as natural resources, transportation planning, environmental studies and vegetation distribution studies [23-26]. Recently updated, the geographic information system for traditional Chinese medicine (TCM-GIS) captures, stores, analyzes and displays geographically referenced information to analyze genetic, ecological and geographic patterns of the spatial distribution of a target species. Using the TCM-GIS, our previous

* Correspondence: slchen@implad.ac.cn
1 Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100193, China
Full list of author information is available at the end of the article

© 2010 Huang et al; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
studies analyzed the potential habitats and distributions of Chinese medicinal plants such as Glycyrrhiza uralensis Fisch., Panax quinquefolium and Panax ginseng [27-29]. The present study aims to characterize the eco-environmental conditions in the A. annua L. production areas in Youyang County and predict the potential distributions of A. annua L. with a high artemisinin-yielding potential.

Methods

Data collection

The spatial distribution of A. annua L. was based on the following four sources: (1) the flora of China [30]; (2) scientific literature concerning the geographic distribution of A. annua L. in China [31]; (3) the Chinese Virtual Herbarium (CVH) [32], (4) germplasm accessions from the Sharing Information System for Chinese Medicinal Plant Germplasm Resources [33]; (5) field data of wild A. annua L. and interviews in Youyang County in 2008. Due to the excellent quality of A. annua L. from the habitats in Youyang County [31,34-36], a total of 180 accessions of A. annua L. germplasm were collected and used in the present study.

The potential distribution mapping program TCM-GIS and geo-referenced datasets were used to develop eco-adaptation models. The TCM-GIS package included three databases, namely (1) a basic geographic information database including digital line graphics and a digital elevation model (scale: 1:1,000,000), (2) a soil database (scale: 1:4,000,000), (3) and a climate database (mean values between 1971 and 2000). All three databases were used for spatial analysis and model calibration.

Raster and vector are two main data models in the TCM-GIS. Raster layers (1 × 1 km² resolution) were used for spatial analysis and model calibration. Vector layers were used to derive and identify the spatial extent and location of suitable habitats through overlay analysis. Moreover, global positioning system data on the locations of the 180 accessions were obtained for villages such as Banqiao, Zhongduo, Mawang and Nanmu and used in the TCM-GIS analysis (Figure 1).

In the present study, 14 eco-environmental variables were chosen for the prediction of spatial distribution in Youyang County. These variables, namely (1) average temperature in January (ATJA), (2) average temperature in February (ATF), (3) average temperature in March (ATM), (4) average temperature in April (ATAP), (5) average temperature in May (ATMA), (6) average temperature in June (ATJ), (7) average temperature in July (ATJU), (8) average temperature in August (ATA), (9) average annual temperature (AAT), (10) annual sunshine time (AST), (11) total annual precipitation (TAP), (12) relative humidity (RH), (13) altitude (AL), (14) and soil properties (SP), were classified into three categories: topography, climate and edaphology (Table 1).

Data analysis

An optimal range was established by identifying minima and maxima for eco-environmental variables (e.g. elevation and temperature) at sample collection sites. The A. annua L. macro-habitats were characterized by examining the mean, minimal and maximal values, standard deviation (SD), standard error (SE), and coefficient of variation (CV) of these variables (Table 2). Prior to distance analysis, we normalized the raster grid data representing each variable. We derived the mean absolute deviation using the following equation:

\[ s_f = \frac{1}{n} \sum_{k=1}^{n} |x_{ikf} - m_f| \]

where \( x_{ikf} \) was the measured values of the variable \( f \) and \( m_f \) is the mean for the variable \( f \). For the determination of similarity between grid data and eco-factor ranges, the statistical distance was calculated with the Minkowski distance equation [37]:

\[ d_{ij} = \left[ \sum_{k=1}^{n} |x_{ik} - x_{jk}|^2 \right]^{1/2} \]

which is a generalization of the Euclidean distance and Manhattan distance; in general the shorter the distance, the greater the similarity. The comprehensive similarity index (SI) of each factor layer was calculated with an overlay analysis with various weighting values. Finally, maps with two ranks of predictive distributions were generated, followed by a grid-based spatial cluster analysis, vector-based overlaying, intersection analysis and an area calculation (Figures 2, 3, 4, Table 3).

The most favorable region for A. annua L. growth is one that has an SI range of 99%-100%, while the second-most favorable region is one that has an SI range of 98%-99%.

Results and Discussion

Eco-environmental preferences

The climatic, edaphic and topographic characteristics of known A. annua L. habitats are listed in Table 2. While low CV values for RH (CV: 0.33), TAP (1.28), AST (3.33), ATJU (4.60), ATAT (4.69), ATJ (6.23), ATJ (6.77) and ATMA (6.81) suggested that these could be the major limiting factors affecting the distribution of high quality A. annua L., high CV values for AL (29.79), ATJU (21.46) and ATF (21.43) suggested otherwise. According to the CV values, weighting value for each parameter was divided into levels I (0.15), II (0.08), III (0.06) and IV
(0.03) and weighting values should add up to one. In addition, datasets of eco-factors from known habitats in Youyang County were as follows: ATJA = 1.2-5.6°C, ATF = 2.0-6.0°C, ATM = 4.0-10.0°C, ATAP = 10.0-16.0°C, ATMA = 14.0-20.0°C, ATJ = 18.0-24.0°C, ATJU = 21.6-27.3°C, ATA = 20.0-26.0°C, AAT = 15.9-21.0°C, AST = 1048-1200 h, TAP = 1169-1267 mm, RH = 79.2-80.6%, AL = 498-1010 mm. Soil types were mainly yellow soil, yellow sandy soil, limestone soil, paddy soil and brown soil with pH value at 6-7 and organic matter content ≥1.3%. Thus, we assumed that these conditions were optimal for the growth of high artemisinin-yielding *A. annua* L.

*A. annua* L. is a short-day plant. Non-juvenile plants are very responsive to short photoperiodic stimuli and flower about two weeks after induction. They require about 1000 hours of sunlight per year. Our results suggest that annual sunlight time is a critical factor for the growth of *A. annua* L., which is consistent with previous studies [5,38]. Previous findings that *A. annua* L. requires a strict watering regime during the preliminary growth stages [5,39] are also consistent with our results.

**Predictive maps**

Figures 2 and 3 are the maps derived from the TCM-GIS analyses. The predicted areas were primarily located in the Wuling Mountain region in central China, covering Guizhou, Chongqing, Hunan, Hubei and Sichuan (25°14'-31°38' N to 104°31'-111°51'E). The predicted habitat density was high in northeastern Guizhou, southeastern Chongqing, northwestern Hunan, southwestern Hubei and parts of southern Sichuan.

The total favorable regions (SI 98%-99%) made up 1.60% of China’s total land area covering 162 counties and cities (a total of 60,292 km²), among which Guizhou took the lead with 31,150 km² including 68 counties and cities. The most favorable region for *A. annua* L. (SI 99%-100%) was in the 58 counties and cities in Guizhou Province with a predicted area of 54,350 km². The second largest predicted area (14,330 km²) was in the 12 counties and cities in Chongqing, followed by Hunan, Hubei and Sichuan (Figure 4). The counties and cities with significant areas of potential habitat are listed in Table 3. The data indicated that Youyang County contained the largest
favorable area with more than 4000 km². Unexpectedly, the total predicted areas in Wuchuan and Zunyi Counties in Guizhou exceeded 2000 km².

One of the world’s largest artemisinin manufacturers and its affiliates operate *A. annua* L. farms in the Chongqing Wulingshan Mountain Range [40,22]. Apart from this, Guizhou may be another important region for *A. annua* L. cultivation, particularly in the northeastern part of the province. Our model predicted that 13% of this area is potential *A. annua* L. habitats [41,42]. Our model did not predict Guangxi Province, known for its habitats of *A. annua* L. of relatively low quality, as a region for *A. annua* L. cultivation possibly due to the subtropical climate, low altitude and red soil in Guangxi which are very different from those in other *A. annua* L. regions in China [9].

Interviews with the locals suggest that the Guizhou region and Youyang County have comparative advantages

---

**Table 1: Environmental factors used in this study.**

<table>
<thead>
<tr>
<th>Category</th>
<th>Variables</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Climate</td>
<td>Average temperature in January (°C)</td>
<td>ATJA</td>
</tr>
<tr>
<td></td>
<td>Average temperature in February (°C)</td>
<td>ATF</td>
</tr>
<tr>
<td></td>
<td>Average temperature in March (°C)</td>
<td>ATM</td>
</tr>
<tr>
<td></td>
<td>Average temperature in April (°C)</td>
<td>ATAP</td>
</tr>
<tr>
<td></td>
<td>Average temperature in May (°C)</td>
<td>ATMA</td>
</tr>
<tr>
<td></td>
<td>Average temperature in June (°C)</td>
<td>ATJ</td>
</tr>
<tr>
<td></td>
<td>Average temperature in July (°C)</td>
<td>ATJU</td>
</tr>
<tr>
<td></td>
<td>Average temperature in August (°C)</td>
<td>ATA</td>
</tr>
<tr>
<td></td>
<td>Average Annual temperature (°C)</td>
<td>AAT</td>
</tr>
<tr>
<td></td>
<td>Annual sunshine time (h)</td>
<td>AST</td>
</tr>
<tr>
<td></td>
<td>Total annual precipitation (mm)</td>
<td>TAP</td>
</tr>
<tr>
<td></td>
<td>Relative humidity (%)</td>
<td>RH</td>
</tr>
<tr>
<td>Topography</td>
<td>Altitude (m)</td>
<td>AL</td>
</tr>
<tr>
<td>Edaphology</td>
<td>Soil properties</td>
<td>SP</td>
</tr>
</tbody>
</table>

---

**Table 2: Summary of eco-environmental characteristics from known *A. annua* L. habitats (n = 180).**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Mean</th>
<th>SE</th>
<th>CV%</th>
<th>SD</th>
<th>Range</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATJA(°C)</td>
<td>3.95</td>
<td>0.005</td>
<td>21.46</td>
<td>0.849</td>
<td>1.2-5.6</td>
<td>0.03</td>
</tr>
<tr>
<td>ATF (°C)</td>
<td>4.1</td>
<td>0.005</td>
<td>21.43</td>
<td>0.765</td>
<td>2.0-6.0</td>
<td>0.03</td>
</tr>
<tr>
<td>ATM(°C)</td>
<td>8.50</td>
<td>0.005</td>
<td>13.36</td>
<td>1.136</td>
<td>4.0-10.0</td>
<td>0.06</td>
</tr>
<tr>
<td>ATAP(°C)</td>
<td>13.35</td>
<td>0.007</td>
<td>10.45</td>
<td>1.39</td>
<td>10.0-16.0</td>
<td>0.06</td>
</tr>
<tr>
<td>ATMA(°C)</td>
<td>17.92</td>
<td>0.006</td>
<td>6.81</td>
<td>1.22</td>
<td>14.0-20.0</td>
<td>0.08</td>
</tr>
<tr>
<td>ATJ(°C)</td>
<td>21.23</td>
<td>0.007</td>
<td>6.77</td>
<td>1.43</td>
<td>18.0-24.0</td>
<td>0.08</td>
</tr>
<tr>
<td>ATJU(°C)</td>
<td>25.30</td>
<td>0.06</td>
<td>4.60</td>
<td>1.164</td>
<td>21.6-27.3</td>
<td>0.08</td>
</tr>
<tr>
<td>ATA(°C)</td>
<td>23.56</td>
<td>0.08</td>
<td>6.23</td>
<td>1.469</td>
<td>20.0-26.0</td>
<td>0.08</td>
</tr>
<tr>
<td>AAT(°C)</td>
<td>19.32</td>
<td>0.05</td>
<td>4.69</td>
<td>0.907</td>
<td>15.9-21.0</td>
<td>0.08</td>
</tr>
<tr>
<td>AST(h)</td>
<td>1118.00</td>
<td>0.21</td>
<td>3.33</td>
<td>37.32</td>
<td>1048-1200</td>
<td>0.08</td>
</tr>
<tr>
<td>TAP(mm)</td>
<td>1209.00</td>
<td>0.09</td>
<td>1.28</td>
<td>15.46</td>
<td>1169-1267</td>
<td>0.08</td>
</tr>
<tr>
<td>RH(%)</td>
<td>79.85</td>
<td>0.02</td>
<td>0.33</td>
<td>2.63</td>
<td>79.2-80.6</td>
<td>0.15</td>
</tr>
<tr>
<td>AL(m)</td>
<td>771.03</td>
<td>1.28</td>
<td>29.79</td>
<td>229.73</td>
<td>498-1010</td>
<td>0.03</td>
</tr>
</tbody>
</table>

*Indication of five soil types: yellow soil, yellow sandy soil, limestone soil, paddy soil and brown soil; pH: 6-7; organic matter content ≥1.3%
Figure 2 Distribution of suitable *A. annua* L. production areas in China with a similarity index (SI) of 99-100%.

Figure 3 Distribution of suitable *A. annua* L. production areas in China with a similarity index (SI) of 98-99%.
for *A. annua* L. growth with a high-yield variety and minimal pests. Furthermore, the northeastern Guizhou is home to wild populations of *A. annua* L. which may be an alternative source for artemisinin.

Using the TCM-GIS, we aimed to determine the optimal ecological factors from known habitats and the results showed that RH, TAP, AST, STJU, AAT and SP were important limiting factors. We also aimed to map the distribution of potential regions for the development of *A. annua* L. in China based on selected climatic, soil and topographical values. Using bioclimatic similarity theory and the TCM-GIS, we predicted the potential growing areas at the county level, particularly in northeastern Guizhou Province. The TCM-GIS is adequate for predicting and identifying potential areas for *A. annua* L. cultivation.

Using a higher resolution raster and vector spatial databases, we improved the resolution of species distribution considerably on the national surveys conducted in the 1960s, 1970s and 1980s. While most of the survey data were based largely on personal experiences and rough estimates, the model used in the present study is relatively objective.

**Conclusion**

The present study demonstrates a GIS approach to predict the potential habitats for *A. annua* L. TCM-GIS is a powerful tool for assessing bioclimatic suitability for medicinal plants.

**Table 3: Major *A. annua* L. regions with similarity index (SI) of 99%-100%**

<table>
<thead>
<tr>
<th>County/City, Province*</th>
<th>Suitable areas km²</th>
<th>Suitable areas %</th>
<th>County/City, Province</th>
<th>Suitable areas km²</th>
<th>Suitable areas %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Youyang, Chongqing</td>
<td>4386</td>
<td>92</td>
<td>Hefeng, Hubei</td>
<td>1225</td>
<td>46</td>
</tr>
<tr>
<td>Xiushan, Chongqing</td>
<td>1419</td>
<td>63</td>
<td>Enshi, Hubei</td>
<td>2038</td>
<td>55</td>
</tr>
<tr>
<td>Wulong, Chongqing</td>
<td>1290</td>
<td>48</td>
<td>Zunyi, Guizhou</td>
<td>3264</td>
<td>70</td>
</tr>
<tr>
<td>Qiangjiang, Chongqing</td>
<td>2286</td>
<td>97</td>
<td>Zhijin, Guizhou</td>
<td>1594</td>
<td>62</td>
</tr>
<tr>
<td>Pengshui, Chongqing</td>
<td>3182</td>
<td>87</td>
<td>Zheng'an, Guizhou</td>
<td>1590</td>
<td>67</td>
</tr>
<tr>
<td>Zhangjiajie, Hunan</td>
<td>1388</td>
<td>58</td>
<td>Yanhe, Guizhou</td>
<td>1471</td>
<td>65</td>
</tr>
<tr>
<td>Yongshun, Hunan</td>
<td>1863</td>
<td>52</td>
<td>Xixiu, Guizhou</td>
<td>1387</td>
<td>90</td>
</tr>
<tr>
<td>Shangzhi, Hunan</td>
<td>1966</td>
<td>61</td>
<td>Wuchuan, Guizhou</td>
<td>2119</td>
<td>82</td>
</tr>
<tr>
<td>Longshan, Hunan</td>
<td>2017</td>
<td>69</td>
<td>Tongzi, Guizhou</td>
<td>1579</td>
<td>53</td>
</tr>
<tr>
<td>Baojing, Hunan</td>
<td>1235</td>
<td>77</td>
<td>Shuiyang, Guizhou</td>
<td>1463</td>
<td>62</td>
</tr>
<tr>
<td>Xuanen, Hubei</td>
<td>1909</td>
<td>74</td>
<td>Others are omitted</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Xianfengshi, Hubei</td>
<td>2257</td>
<td>96</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lichuan, Hubei</td>
<td>2266</td>
<td>52</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Areas smaller than 1400 km² are not listed.*
Abbreviations
TCM-GIS: traditional Chinese medicine geographic information system; GIS: geographic information system; SI: similarity index; SD: standard deviation; SE: standard error; CV: coefficient of variation; ATJA: average temperature in January; ATF: average temperature in February; ATM: average temperature in March; ATAP: average temperature in April; ATMA: average temperature in May; AT: average temperature in June; ATUJ: average temperature in July; ATA: average temperature in August; AAT: average annual temperature; AST: annual sun-shine time; TAP: total annual precipitation; RH: relative humidity; AL: altitude; SP: soil properties; CVH: Chinese Virtual Herbarium.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
LFH, SLC and CXX designed the study and performed the analyses. BZD helped with data analysis. All authors wrote the manuscript. All authors read and approved the final version of the manuscript.

Acknowledgements
The authors would like to thank the National Natural Science Foundation of China for its support through project No. 3050081.

Author Details
Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100193, China

Received: 23 November 2009 Accepted: 17 May 2010
Published: 17 May 2010

References
32. The Chinese Virtual Herbarium (CVH) [http://www.cvherb.org.cn]
38. Purdue University Annual Wormwood (Artemisia annua L) [http://www.hort.purdue.edu/newcrop/cropfactsheets/artemisia.pdf]


Cite this article as: Huang et al., Mapping the potential distribution of high artemisinin-yielding *Artemisia annua* L. (Qinghao) in China with a geographic information system *Chinese Medicine* 2010, 5:18