The Bio-Geographical Regions Division of Global Terrestrial Animal by Multivariate Similarity Clustering Analysis Method

Qi Shen¹, Ji Qi Lu², Shu Jie Zhang², Zhixing You³, Ying Dan Ren⁴, Xiao Chong Shen²,4*

¹First Clinical College, Henan University of Traditional Chinese Medicine, Zhengzhou, China
²College of Life Sciences, Zhengzhou University, Zhengzhou, China
³Academy of Mathematics and Systems Science, Beijing, China
⁴Institute of Plant Protection, Henan Academy of Agricultural Sciences, Zhengzhou, China

Abstract
A novel multivariate similarity clustering analysis (MSCA) approach was used to estimate a biogeographical division scheme for the global terrestrial fauna and was compared against other widely used clustering algorithms. The faunal dataset included almost all terrestrial and freshwater fauna, a total of 4631 families, 141,814 genera, and 1,334,834 species. Our findings demonstrated that suitable results were only obtained with the MSCA method, which was associated with distinct hierarchies, reasonable structuring, and furthermore, conformed to biogeographical criteria. A total of seven kingdoms and 20 sub-kingdoms were identified. We discovered that the clustering results for the higher and lower animals did not differ significantly, leading us to consider that the analysis result is convincing as the first zoogeographical division scheme for global all terrestrial animals.

Keywords
Global Animal, Multivariate Similarity Clustering Analysis, Biogeography, Regionalization

1. Introduction
Biodiversity is important to humankind, and the significance of its protection is well recognized by both scholars and governments [1]. Of all the measures for managing and protecting biodiversity, biogeography constitutes a basic, but very useful, tool [2] [3]. The discipline of biogeography originated in 1761 when it was introduced by the French naturalist Georges Buffon [4]. Early zoogeograph-
ers were not concerned with the ecological associations of faunal distributions, but were more interested in defining areas that are characterized by certain animals. Based on this, six biogeographic regions corresponding to continents were defined [5] [6]. In 1858, the English ornithologist Philip Sclater defined six zoological regions (Aethiopian, Australasian, Indian, Palaeartic, Nearctic, and Neotropical) that are still in use today [7]. In 1876, the British zoologist Alfred Russel Wallace adopted Sclater’s scheme and, using marsupial distributions, defined the “Wallace line”; a hypothetical line that intersects Indonesia, between Kalimantan and Sulawesi, separating the Oriental (Indian) region and the Australian region [8]. His seminal work, the geographical distribution of animals, later became the foundation of modern biogeography [9], and, despite some minor revisions [10], the original map from this publication is still in use today [11].

The exploration of biogeography has continued into the 21st century. In addition to the lengthy debates regarding the rationality of the “Wallace line” [12] [13] [14] [15], the development of quantitative analytical methods for determining and refining zoogeographical regions has been central to biogeographic research [16]-[26]. There has been considerable focus on zoogeographic division schemes in the last few decades, and a variety of schemes based on different methods have been proposed between 7 - 14 divisions [27]-[42]. Being faced with numerous and disordered results, Morrone J.J. wonders that bio-geographical regionalization is a spectre haunting biogeography [43]. Unfortunately, there are three significant issues with these proposals.

Firstly, while the necessity for quantitative methods in biogeographic division schemes has been recognized, systematic comparisons of different similarity coefficient formulas and clustering algorithms are lacking. Additionally, some similarity coefficient formulas are only accurate under defined conditions, and thus their use is restricted. Secondly, researchers have used the grid method to define basic geographic units, which are typically generated using latitude and longitude coordinates or geographical distance. Although this method is acceptable, species distribution records, which have been collected and accumulated long-term by taxonomists, are not associated with grid method. The variations in the collection degrees (such as the frequency, timing, and depth) between each grid could result in discrepancies, thereby influencing the model estimation. The grid cell strategy is thus best suited to medium-scale field investigations, but is not appropriate for global-scale clustering analyses. Thirdly, there has been greater focus on higher animals, such as vertebrates, despite the fact that they only represent a small percentage of the global fauna (4% of species, 5% of genera). Lower animals should also be included in biogeographical research. Considering the concerns outlined above, the aim of this study was to develop a division scheme for the global terrestrial fauna based on a comprehensive analytical quantitative framework. To achieve this, we implemented a novel approach based on multivariate similarity clustering analysis (MSCA) combined with the similarity gen-

DOI: 10.4236/oje.2022.123014

Open Journal of Ecology
eral formula (SGF). This method was validated by comparison with widely used clustering algorithms and existing division schemes.

2. Material and Method

2.1. Global Terrestrial Animal Species

The materials used in this study originated primarily from three sources: catalogs, checklists, or taxonomic monographs on global or regional fauna [44]-[66]; biodiversity materials obtained from biodiversity research institutions or websites [67]-[92]; and new species or new distribution publications by taxonomists [93]-[110]. Excluding deep marine species and fossil species, a total of 141,814 genera, and 1,334,834 species of fauna were included (Table 1). The number of genera is about 70% of total number of global animal genera, and number of species is about 75% of total number of global animal species. To maximize the utilization of the above materials and to improve the clarity of the analysis results, we selected to use genus as basic biological units (BBUs).

Table 1. Terrestrial animal of the World.

<table>
<thead>
<tr>
<th>Grade</th>
<th>Phylum</th>
<th>Class</th>
<th>Order</th>
<th>Family</th>
<th>Genus</th>
<th>Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>Higher</td>
<td>1. Chordata</td>
<td>5</td>
<td>77</td>
<td>613</td>
<td>6890</td>
<td>48,881</td>
</tr>
<tr>
<td>Lower</td>
<td>1. Acanthocephala</td>
<td>4</td>
<td>10</td>
<td>25</td>
<td>153</td>
<td>1406</td>
</tr>
<tr>
<td>Lower</td>
<td>2. Annelida</td>
<td>2</td>
<td>18</td>
<td>145</td>
<td>2105</td>
<td>16,602</td>
</tr>
<tr>
<td>Lower</td>
<td>3. Arthropoda</td>
<td>11</td>
<td>83</td>
<td>2195</td>
<td>120,397</td>
<td>1,167,110</td>
</tr>
<tr>
<td>Lower</td>
<td>4. Brachiopoda</td>
<td>5</td>
<td>19</td>
<td>167</td>
<td>460</td>
<td>3510</td>
</tr>
<tr>
<td>Lower</td>
<td>5. Bryozoa</td>
<td>3</td>
<td>7</td>
<td>229</td>
<td>987</td>
<td>9473</td>
</tr>
<tr>
<td>Lower</td>
<td>6. Cnidaria</td>
<td>3</td>
<td>5</td>
<td>22</td>
<td>63</td>
<td>491</td>
</tr>
<tr>
<td>Lower</td>
<td>7. Gastrotricha</td>
<td>2</td>
<td>17</td>
<td>66</td>
<td>911</td>
<td></td>
</tr>
<tr>
<td>Lower</td>
<td>8. Kamptozoa</td>
<td>4</td>
<td></td>
<td></td>
<td>14</td>
<td>210</td>
</tr>
<tr>
<td>Lower</td>
<td>9. Micrognathozoa</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Lower</td>
<td>10. Mollusca</td>
<td>7</td>
<td>41</td>
<td>361</td>
<td>3372</td>
<td>30,388</td>
</tr>
<tr>
<td>Lower</td>
<td>11. Nematoda</td>
<td>2</td>
<td>14</td>
<td>180</td>
<td>1982</td>
<td>16,302</td>
</tr>
<tr>
<td>Lower</td>
<td>12. Nematomorpha</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>22</td>
<td>620</td>
</tr>
<tr>
<td>Lower</td>
<td>13. Nemertea</td>
<td>3</td>
<td>3</td>
<td>42</td>
<td>305</td>
<td>1423</td>
</tr>
<tr>
<td>Lower</td>
<td>14. Onychophora</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>52</td>
<td>175</td>
</tr>
<tr>
<td>Lower</td>
<td>15. Platyheloniathes</td>
<td>4</td>
<td>41</td>
<td>423</td>
<td>3894</td>
<td>24,894</td>
</tr>
<tr>
<td>Lower</td>
<td>16. Porifera</td>
<td>4</td>
<td>32</td>
<td>146</td>
<td>786</td>
<td>9994</td>
</tr>
<tr>
<td>Lower</td>
<td>17. Rotifera</td>
<td>2</td>
<td>5</td>
<td>15</td>
<td>50</td>
<td>656</td>
</tr>
<tr>
<td>Lower</td>
<td>18. Tardigrada</td>
<td>3</td>
<td>5</td>
<td>17</td>
<td>118</td>
<td>1312</td>
</tr>
<tr>
<td>Lower</td>
<td>19. Xenacoelomorpha</td>
<td>2</td>
<td>19</td>
<td>115</td>
<td>475</td>
<td></td>
</tr>
</tbody>
</table>

Total 20 | 61 | 368 | 4626 | 141,814 | 1,334,834
2.2. Division of Basic Geographical Units (BGU) and Building Databank

Based on ecological conditions and animal distributions [111], we divided the global terrestrial surface, excluding Antarctica, into 67 BGUs (Figure 1) as the basis for the clustering analyses and biodivision. Of these BGUs, 21 are plain-based, 11 are hill-based, 12 are mountains-based, 11 are plateau-based, five are desert-based, and seven are island-based. Twenty-seven BGUs are located within the tropical zone, 34 are located within the temperate zone, and six are extended to the frigid zone.

The database was created using Microsoft Access (Microsoft Corporation, New Mexico, USA), with BBUs and BGUs in rows and columns, respectively. The distribution of each animal species of a specific genus in an administrative region was converted into a BGU, summarized as the distribution of the genus, and entered into the database. If the entry was associated with a distribution, a

![Figure 1. BGUs of the World.](image-url)
“1” was recorded in the database; otherwise nothing was recorded. These basic distributional records (BDRs) constituted the basis of the quantitative analyses. The distribution information of the major animal groups in each BGU is described in Table 2.

2.3. Clustering Methods

Although similarity formulas are more than a few dozens [112], they are only able to calculate the similarity coefficient between two regions. In this study, the similarity coefficient of multiple regions was calculated as the percentage of the average number of common species in the participating regions to the number of all species [113]. We defined the similarity general formula (SGF) as:

$$ S_{In} = \frac{\sum H_i / n S_n}{\sum (S_i - T_i) / n S_n}, $$

where $S_{In}$ is the similarity coefficient of $n$ geographical units; $S_n$, $H_n$, and $T_n$ represent the number of total species, common species, and unique species of BGU $i$, and $H_n = S_n - T_n$; $S_n$ is the total number of species in $n$ BGUs. All of these values can easily be obtained from the database, which is convenient and efficient for both manual and computational analysis. We used a combination of MSCA and SGF in this study. In MSCA, the similarity coefficient of any group of BGU is calculated directly using the raw data of the participating BGUs [114], and it is not affected by the previous similarity coefficient, and furthermore, is not limited by the sequence of the clustering analysis. The general similarity coefficient (GSC) of all 67 BGUs can even be calculated first. Final dendrogram can be generated according to the size of these similarity coefficients. This method has been validated in some fauna [115]-[124], and has been successfully used for distribution pattern analysis at large geographic scales [125] [126] [127].

The results of the above method were assessed by comparison with three common hierarchical clustering methods:

The single linkage method (SLM) [128], also known as minimum distance method, is the most basic clustering analysis method. This method uses the similarity coefficient formula proposed by Jaccard (1901) [16], where,

$$ SI = C / (A + B - C). $$

The average group linkage method (AGL), also known as the unweighted pair group means algorithm (UPGMA) [23], is a widely used clustering method. This method uses Simpson’s formula (1946) [19] proposed by Szymkiewicz [1934] (18), where $SI = C / \min (A, B)$.

The sum of squares method (SSM), also known as Ward’s method (21), usually provides better results than the above models, but involves more complex calculations. In this method, we used the similarity coefficient formula proposed by Czekanowski (1913) (17), which is also called the Sørensen formula (1948) (20), where, $SI = 2C(A + B)$.

All three similarity coefficient formulae can only perform pairwise comparisons between regions. $A$ and $B$ represented the number of species in two regions, while $C$ represented the common species shared by two regions.
Table 2. Number of genera of higher and lower of global terrestrial animal in every BGUs.

<table>
<thead>
<tr>
<th>BGU</th>
<th>Higher</th>
<th>Lower</th>
<th>All</th>
<th>BGU</th>
<th>Higher</th>
<th>Lower</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>343</td>
<td>7845</td>
<td>8188</td>
<td>44</td>
<td>848</td>
<td>3712</td>
<td>4560</td>
</tr>
<tr>
<td>02</td>
<td>454</td>
<td>9983</td>
<td>10,437</td>
<td>45</td>
<td>666</td>
<td>2274</td>
<td>2940</td>
</tr>
<tr>
<td>03</td>
<td>507</td>
<td>8676</td>
<td>9183</td>
<td>46</td>
<td>736</td>
<td>4809</td>
<td>5545</td>
</tr>
<tr>
<td>04</td>
<td>589</td>
<td>11,557</td>
<td>12,146</td>
<td>47</td>
<td>663</td>
<td>4363</td>
<td>5026</td>
</tr>
<tr>
<td>05</td>
<td>293</td>
<td>3389</td>
<td>3682</td>
<td>48</td>
<td>674</td>
<td>6386</td>
<td>7060</td>
</tr>
<tr>
<td>06</td>
<td>298</td>
<td>2770</td>
<td>3068</td>
<td>49</td>
<td>338</td>
<td>4492</td>
<td>4830</td>
</tr>
<tr>
<td>11</td>
<td>512</td>
<td>4568</td>
<td>5080</td>
<td>51</td>
<td>465</td>
<td>5319</td>
<td>5784</td>
</tr>
<tr>
<td>12</td>
<td>317</td>
<td>1731</td>
<td>2048</td>
<td>52</td>
<td>389</td>
<td>3604</td>
<td>3993</td>
</tr>
<tr>
<td>13</td>
<td>329</td>
<td>1614</td>
<td>1943</td>
<td>53</td>
<td>400</td>
<td>2919</td>
<td>3319</td>
</tr>
<tr>
<td>14</td>
<td>527</td>
<td>3241</td>
<td>3768</td>
<td>54</td>
<td>548</td>
<td>9060</td>
<td>9608</td>
</tr>
<tr>
<td>15</td>
<td>286</td>
<td>2843</td>
<td>3129</td>
<td>55</td>
<td>530</td>
<td>8077</td>
<td>8607</td>
</tr>
<tr>
<td>16</td>
<td>163</td>
<td>1702</td>
<td>1865</td>
<td>56</td>
<td>442</td>
<td>5490</td>
<td>5932</td>
</tr>
<tr>
<td>17</td>
<td>273</td>
<td>4886</td>
<td>5159</td>
<td>57</td>
<td>273</td>
<td>3464</td>
<td>3737</td>
</tr>
<tr>
<td>18</td>
<td>221</td>
<td>3016</td>
<td>3237</td>
<td>58</td>
<td>168</td>
<td>3847</td>
<td>4015</td>
</tr>
<tr>
<td>19</td>
<td>248</td>
<td>1355</td>
<td>1603</td>
<td>59</td>
<td>359</td>
<td>5819</td>
<td>6178</td>
</tr>
<tr>
<td>20</td>
<td>186</td>
<td>1354</td>
<td>1540</td>
<td>60</td>
<td>481</td>
<td>6799</td>
<td>7280</td>
</tr>
<tr>
<td>21</td>
<td>346</td>
<td>4383</td>
<td>4729</td>
<td>61</td>
<td>735</td>
<td>10,015</td>
<td>10,750</td>
</tr>
<tr>
<td>22</td>
<td>224</td>
<td>2167</td>
<td>2391</td>
<td>62</td>
<td>809</td>
<td>7433</td>
<td>8242</td>
</tr>
<tr>
<td>23</td>
<td>220</td>
<td>2737</td>
<td>2957</td>
<td>63</td>
<td>728</td>
<td>6280</td>
<td>7008</td>
</tr>
<tr>
<td>24</td>
<td>457</td>
<td>5851</td>
<td>6308</td>
<td>64</td>
<td>986</td>
<td>9327</td>
<td>10,313</td>
</tr>
<tr>
<td>25</td>
<td>777</td>
<td>8159</td>
<td>8936</td>
<td>65</td>
<td>1120</td>
<td>10,896</td>
<td>12,016</td>
</tr>
<tr>
<td>26</td>
<td>752</td>
<td>10,872</td>
<td>11,624</td>
<td>66</td>
<td>971</td>
<td>11,066</td>
<td>12,037</td>
</tr>
<tr>
<td>27</td>
<td>455</td>
<td>8382</td>
<td>8837</td>
<td>67</td>
<td>510</td>
<td>4298</td>
<td>4808</td>
</tr>
<tr>
<td>28</td>
<td>280</td>
<td>2050</td>
<td>2330</td>
<td>68</td>
<td>1294</td>
<td>3927</td>
<td>5221</td>
</tr>
<tr>
<td>29</td>
<td>343</td>
<td>7717</td>
<td>8060</td>
<td>69</td>
<td>914</td>
<td>3050</td>
<td>3964</td>
</tr>
<tr>
<td>31</td>
<td>644</td>
<td>2774</td>
<td>3418</td>
<td>70</td>
<td>1653</td>
<td>7742</td>
<td>9395</td>
</tr>
<tr>
<td>32</td>
<td>872</td>
<td>6609</td>
<td>7481</td>
<td>71</td>
<td>1200</td>
<td>5418</td>
<td>6618</td>
</tr>
<tr>
<td>33</td>
<td>776</td>
<td>4061</td>
<td>4837</td>
<td>72</td>
<td>1259</td>
<td>6463</td>
<td>7722</td>
</tr>
<tr>
<td>34</td>
<td>812</td>
<td>6201</td>
<td>7013</td>
<td>73</td>
<td>932</td>
<td>2896</td>
<td>3828</td>
</tr>
<tr>
<td>35</td>
<td>610</td>
<td>4248</td>
<td>4858</td>
<td>74</td>
<td>1024</td>
<td>4724</td>
<td>5748</td>
</tr>
<tr>
<td>36</td>
<td>1008</td>
<td>8606</td>
<td>9614</td>
<td>75</td>
<td>373</td>
<td>3196</td>
<td>3569</td>
</tr>
<tr>
<td>37</td>
<td>570</td>
<td>4866</td>
<td>5436</td>
<td>76</td>
<td>393</td>
<td>353,581</td>
<td>393,016</td>
</tr>
<tr>
<td>38</td>
<td>411</td>
<td>4912</td>
<td>5323</td>
<td>77</td>
<td>6890</td>
<td>134,924</td>
<td>141,814</td>
</tr>
<tr>
<td>39</td>
<td>463</td>
<td>4809</td>
<td>5272</td>
<td>78</td>
<td>67</td>
<td>67</td>
<td>67</td>
</tr>
<tr>
<td>42</td>
<td>838</td>
<td>4233</td>
<td>5071</td>
<td>79</td>
<td>5.72</td>
<td>2.62</td>
<td>2.77</td>
</tr>
<tr>
<td>43</td>
<td>543</td>
<td>2249</td>
<td>2792</td>
<td>80</td>
<td>589</td>
<td>5277</td>
<td>5866</td>
</tr>
</tbody>
</table>

DOI: 10.4236/oje.2022.123014
3. Results

3.1. Clustering Results of Terrestrial Animal

The results of the MSCA clustering analysis of 141,814 global terrestrial faunal genera are shown in Figure 4. The GSC value was 0.066, and at a similarity of 0.300, 67 BGUs were grouped into 20 smaller unit crowds (SUCs), labeled from a to t. At a similarity of 0.200, the BGUs were further grouped into seven larger unit crowds (LUCs), labeled from A to G. Each unit within a crowd was adjacent to another unit, thereby satisfying principles of geography. The ecological conditions of each crowd were relatively consistent, which met ecological principles. In addition, intra-crowd similarity was greater than inter-crowd similarity, thereby realizing statistical principles. The MSCA clustering analysis results for the higher animals and lower animals are shown in Figure 2 and Figure 3, respectively, using the same letters for the crowds to facilitate direct comparisons. The animals grouped into seven LUCs and some SUCs at specific levels and exhibited similar crowd compositions of Figure 4. Some variation in the location of a few BGUs between two regions existed, but these nevertheless still conformed to geographical principles.

We also assessed our scheme against the existing animal biogeographical division schemes proposed by some zoologists (28, 29, 30, 31, 34, 35, 36, 40, 42, 43, 67). Our results are closer aligned with the biogeographical patterns proposed by Kreft using global mammal data [34], than that of Holt [40] and Procheș [30]. Our scheme supported many aspects of these proposals, including the subdivision of the Palaearctic Realm into western and eastern halves based on the distribution patterns of Trichoptera and Aleyrodidae [31] [36]; the placement of New Guinea Island and the Pacific Islands into the Oriental Realm according to the distribution of Siphonaptera and Trichoptera [42] [43]; the removal of the Pacific Islands from the Australian Realm based on the distribution patterns of freshwater insects, Aleyrodidae, and Staphylinidae [28] [31] [67]; the reintroduction of Yemen and Oman into the Palaearctic Realm based on the distribution of Symphyta and Culicidae [29] [35]; and the transferal of Mexico into the Nearctic Realm according to the distribution of Culicidae [29]. However, our analysis did not support the establishment of new realms for New Zealand, Madagascar, and Antarctica.

The establishment of B LUC is closely related to Chinese biodiversity. Modern China has the most animal genera and very much endemic genera (Table 3). Obviously, this is the great contribution of Chinese zoologists.

3.2. Comparison with Traditional Clustering Methods

The SLM results for the same dataset were chaotic and the groups were difficult to categorize (Figure 5). Most of the BGUs were considered to be noise, and when the distance between two clusters was set at 0.730, only two crowds (D and E) could be recognized. In contrast, AGL provided significantly improved results with appreciably less noise (Figure 6). When the distance value was set at 0.740,
Figure 2. Clustering tree of 6890 genera higher animals.
Figure 3. Clustering tree of 134,924 genera lower animals.
Figure 4. Clustering tree of global 141,814 genera terrestrial animal by MSCA.

Table 3. The total genera and endemic genera of every LUC, key units and Main countries.

<table>
<thead>
<tr>
<th>LUC</th>
<th>Total genera</th>
<th>Endemic genera</th>
<th>Key unit</th>
<th>Total genera</th>
<th>Endemic genera</th>
<th>Main country</th>
<th>Total genera</th>
<th>Endemic genera</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>20,855</td>
<td>5742</td>
<td>04</td>
<td>12,146</td>
<td>1004</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>20,686</td>
<td>4556</td>
<td>26</td>
<td>11,624</td>
<td>819</td>
<td>China</td>
<td>18,357</td>
<td>4290</td>
</tr>
<tr>
<td>C</td>
<td>23,596</td>
<td>7919</td>
<td>36</td>
<td>9614</td>
<td>1372</td>
<td>Indonesia</td>
<td>9614</td>
<td>1372</td>
</tr>
<tr>
<td>D</td>
<td>16,588</td>
<td>8010</td>
<td>48</td>
<td>7060</td>
<td>1074</td>
<td>S. Africa</td>
<td>7060</td>
<td>1074</td>
</tr>
<tr>
<td>E</td>
<td>17,400</td>
<td>7997</td>
<td>54</td>
<td>9608</td>
<td>1169</td>
<td>Australia</td>
<td>15,774</td>
<td>6733</td>
</tr>
<tr>
<td>F</td>
<td>28,008</td>
<td>9886</td>
<td>66</td>
<td>10,313</td>
<td>985</td>
<td>USA</td>
<td>17426</td>
<td>2534</td>
</tr>
<tr>
<td>G</td>
<td>18,529</td>
<td>7428</td>
<td>73</td>
<td>9395</td>
<td>1057</td>
<td>Brazil</td>
<td>10,669</td>
<td>1812</td>
</tr>
</tbody>
</table>
five crowds could be distinguished. Of these five crowds, four corresponded to the C, D, E, and G groups. One crowd consisting of 31 BGUs was very complex and could only be categorized into three crowds when the distance value was set at 0.550. More definitive clustering results were obtained with SSM (Figure 7). When the distance value was set at 1.40, eight crowds were obtained; among which seven were comparable to the seven crowds in the MSCA, while the remaining crowd had no geographical significance, and further categorization using SSM proved challenging. These findings indicate that these three clustering methods do not satisfy zoogeographic requirements.

3.3. A Biogeographical Division Scheme for the Global Terrestrial Fauna

Based on the clustering results, we suggest that the terrestrial world can be divided into seven kingdoms and 20 subkingdoms using an animal geographical regionalization scheme (Figure 8). This is the first geographical regionalization scheme that represents the overall global terrestrial fauna.

Our scheme showed a similar overall distribution pattern to Wallace’s scheme [9], with some notable differences. For example, in our study 1) the Palaearctic Realm is further divided into eastern and western halves; 2) New Guinea Island and the Pacific Islands are regarded as part of the Oriental kingdom as opposed
Figure 7. Clustering tree of global animal by SSM.


to the Australian Realm, and the “Wallace line” no longer exists; 3) Central America is regarded as part of the Nearctic kingdom instead of the Neotropical Realm; 4) Yemen and Oman are considered to be part of the Western Palaearctic kingdom as opposed to the Afrotropical Realm; 5) Taiwan constitutes part of the Eastern Palaearctic kingdom instead of the Oriental Realm.

4. Conclusions and Discussion

To the best of our knowledge, this constitutes the first quantitative attempt at a
division scheme for the global terrestrial fauna. The MSCA method facilitated the delineation of the global terrestrial fauna at a large geographical scale and provided more accurate clustering results than other commonly used clustering methods. Interestingly, the results obtained from the MSCA closely approached the global zoogeographic regions proposed by Wallace, but provided a quantitative validation of the scheme. MSCA can therefore be considered as a useful tool for facilitating the revision of a global biogeographical faunal division scheme.

Based on the observation that the same distribution patterns exist for higher and lower animals worldwide despite their distinct evolutionary stages, degrees of evolution, and habitats, we deduce that the same distribution patterns may also be shared among animals, plants, and microbes. Therefore, we recommend the use of quantitative analyses, such as MSCA, to establish a biogeographic division scheme for all terrestrial living organisms, including plants and microbes.

**Acknowledgements**

We thank B.C. Cox (UK), H. Kreft (Germany), J. Morse (USA), M.V. Cianciaruso (Brazil), D.R. Gustafsson (USA), R.J. Whittaker (UK), P. Vrsansky (Slovakia), J.C. Beaucournu (France), T. Najer (Czech Republic), P. Soisook (Thailand), M. Caesar (France), M.P. Valim (Brazil) for presenting references. This study was supported by the Key Laboratory Foundation of Henan (112300413221).

**Conflicts of Interest**

The authors declare no conflicts of interest regarding the publication of this paper.

**References**


https://doi.org/10.1007/s10750-007-9246-3


lobocondyla from China (Hymenoptera: Formicidae), with a Revised Key to the Known Species. Zoological Systematics, 44, 132-139.


https://doi.org/10.1111/1749-4877.12485


https://doi.org/10.2307/2346439