Biogeography of Medically Important Insects using Quantitative Analysis

By Qi Shen, Zhixing You, Xiaojing Ma & Xiaocheng Shen

Henan University of Traditional Chinese Medicine

Abstract- We summarized distributional information of medically important insects from 76 families and 4531 genera occurring worldwide. The continents were divided into 67 basic geographical units. Using a new similarity formula and a new clustering method for quantitative analysis, 67 basic geographical units were clustered into 7 large unit groups and 20 small unit groups. The results were superior to the traditional single linkage method, average group linkage method, or sum of squares method. The cluster results were similar with the result of mainly phytophagous insects 104,344 genera in the world, but were different from the Wallace’s mammal geographical division scheme. Based on these seemingly contradictory results, we infer that animals, insect and plants may have the same distribution pattern and that it is necessary to conduct precise quantitative analysis for animals and plants worldwide.

Keywords: biogeography; medical important insect; similarity general formula; multivariate similarity clustering analysis.

GJMR-G Classification: NLMC Code: WC 900
Biogeography of Medically Important Insects using Quantitative Analysis

Qi Shen*, Zhixing You*, Xiaojing Ma* & Xiaocheng Shen*

Abstract: We summarized distributional information of medically important insects from 76 families and 4531 genera occurring worldwide. The continents were divided into 67 basic geographical units. Using a new similarity formula and a new clustering method for quantitative analysis, 67 basic geographical units were clustered into 7 large unit groups and 20 small unit groups. The results were superior to the traditional single linkage method, average group linkage method, or sum of squares method. The cluster results were similar with the result of mainly phytophagous insects 104,344 genera in the world, but were different from the Wallace’s mammal geographical division scheme. Based on these seemingly contradictory results, we infer that animals, insect and plants may have the same distribution pattern and that it is necessary to conduct precise quantitative analysis for animals and plants worldwide.

Keywords: biogeography; medical important insect; similarity general formula; multivariate similarity clustering analysis.

I. Introduction

There are three categories of medically important insects: insects that feed on the blood of warm blooded animals (humans, mammals, and birds) and can transmit disease; insects that feed on the fur, feathers, and skin secretions of animals and birds, irritating the host; insects that live in the habitats of humans, mammals and birds, causing irritation and sometimes transmitting diseases. Since these insects have a close relationship with mammals and birds, they may have the same geographical distribution pattern with the mammals described by Wallace (1876). Computer and Internet technology has made it possible to collect and analyze large data sets and re-evaluate previous Wallace’s scheme which are based on qualitative analysis (Olson et al., 2001; Procheş, 2005; Cox, 2010; Krefet et al., 2010; Procheș et al., 2012; Rueda et al., 2013; Whittaker et al., 2013; Holt et al., 2013; Peixotoet al., 2017). Krefet et al. (2010) and Holt et al. (2013) both used Simpson similarity formula and the UPGMA method for quantitative analysis of mammalian species distribution but they obtained different results.

Different geographical division schemes for some insect orders and families have been proposed (Herman et al., 2001; Evans, 2007; Ballanet et al., 2008; Moor et al., 2008; Morse et al., 2011; Taegeret al., 2010). The results of most of these studies did not support the “Wallace line”, that is Wallace’s great contribution to the field. Among them, as medical important insects, Culinidae and Siphonaptera geographic division settings are also proposed the same questions (Siver, 2004; Vashchonovet al., 2013). The extensive attention and in-depth discussion in biogeography interpretation provides an exciting opportunity for evaluating insect distributions and geographical division plans.

We used the similarity general formula (SGF) proposed by Shen et al. (2008a) and multivariate similarity clustering analysis (MSCA) (Shen et al., 2008b) for quantitative analysis of the medically important insects in China (Shen, 2014). The results were unexpectedly similar to the results of all (93661) insect species in China (Shenet al., 2013a; 2013b; 2015), but different from the results of a Chinese mammalian species geographical division, made using qualitative analysis (Zhang, 2011). To study the relationship between the global distributions of medically important insects, phytophagous insects, and mammals, we used a variety of quantitative methods for this analysis.

II. Materials and Methods

a) Global medically important insect species

We used medically important insect distribution data from four resources: (1) World species and distribution data collected and summarized by entomologists, e.g. Knight et al., 1977, Durden et al., 1994, Currie et al., 2008, Adler et al., 2014; (2) Data from specific countries and regions, e.g. Seccombe et al., 1993, De Carvalho et al., 2005, Coscarón et al., 2008, Crespo et al., 2010, Chahari et al., 2015, Wolff et al., 2016, Takano et al., 2017; (3) Data from organizations studying biodiversity and academic websites, such as Beccdloni, 2014, Borkent, 2014, Pickering, 2014, Evenhuis, 2016, GBIF, 2017; (4) Recently published data for new species and new distribution records, such as Gustafsson et al., 2015, Najer et al., 2016, Szela et al., 2016, Fuenzalidaet al., 2017, Huerta et al., 2017, Natarajanet al., 2017, Vidička et al., 2017. The data cited above did not include marine species or fossil records. In total, 9 orders, 67 families, 4531 genera and 63470 species were included (Table 1). This was 4.3% of the
total number of insect genera and 6.1% of the total insect species in the world. Because insects have small bodies, their species distribution is narrower compared with higher animals and plants (Shen et al., 2018). To improve the data utilization ratio and accuracy of the analysis, the genus was used as the basic biological unit (BBU).

Table 1: Global medically important insect species used for analysis

<table>
<thead>
<tr>
<th>Orders</th>
<th>No. of families</th>
<th>No. of genera</th>
<th>No. of species</th>
<th>Main data sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anoplura</td>
<td>14</td>
<td>46</td>
<td>553</td>
<td>Durden et al., 1994, Sánchez-Montes et al., 2013</td>
</tr>
<tr>
<td>Hemiptera</td>
<td>1</td>
<td>22</td>
<td>74</td>
<td>Usinger, 1966, Iorio, 2012, GBIF, 2019c</td>
</tr>
<tr>
<td>Coleoptera</td>
<td>1</td>
<td>126</td>
<td>2480</td>
<td>GBIF, 2019a</td>
</tr>
<tr>
<td>Diptera</td>
<td>16</td>
<td>2337</td>
<td>36594</td>
<td>Knight et al., 1977, Seccombe et al., 1993, Currie, 2008,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Adler et al., 2014, Borkent, 2014, Henrques, 2016,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Evenhuis, 2016, Takaoka et al., 2017, GBIF, 2019b</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Acosta, 2003, Hastriter et al., 2006, Lewis et al.,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2013, Vashchonok et al., 2013, Beaucournet et al., 2014</td>
</tr>
<tr>
<td>Siphonaptera</td>
<td>20</td>
<td>241</td>
<td>2099</td>
<td>GBIF, 2019e</td>
</tr>
<tr>
<td>Lepidoptera</td>
<td>2</td>
<td>544</td>
<td>5969</td>
<td>GBIF, 2019d</td>
</tr>
<tr>
<td>Hymenoptera</td>
<td>5</td>
<td>440</td>
<td>6708</td>
<td>GBIF, 2019d</td>
</tr>
<tr>
<td>Total</td>
<td>76</td>
<td>4531</td>
<td>63470</td>
<td></td>
</tr>
</tbody>
</table>

b) Division of basic geographical units (BGU) and building the databank

According to the terrain, climate, and other ecological conditions, we have divided the continents (except Antarctica) into 67 basic geographical units (BGU) (Fig. 1). Of these BGUs, 21 BGUs were mainly plain, 11 were mainly hills, 12 were mainly mountain, 11 were mainly plateau, five were mainly desert and seven were mainly islands. A total of 27 BGUs were in tropical zone, 34 were in temperate zones and six were extended to the frigid zone. The names and geographical ranges of the BGUs are listed in Table 2.

We used Microsoft Access as our database software. Each BGU was listed as the column and BBU was listed as the row. The distribution of different species belonging to the same genus was transferred to the BGU and summarized as the genus distribution. During the data entry, when there was a distribution, it was marked as 1; if there was no distribution, no record was entered. These basic distributional records (BDR) were the basis of quantitative analysis. Each BGU insect genus number is listed in Table 2.

Fig. 1: BGUs of the World
01 Northern Europe, 02 Western Europe, 03 Central Europe, 04 Southern Europe, 05 Eastern Europe, 06 European Russia, 07 Middle East, 08 Saudi Arabia, 09 Yemen and Oman, 10 Plateau of Iran, 11 Central Asia, 12 Western Siberia, 13 Eastern Siberia, 14 Ussuri region, 15 Mongolia, 16 Plateau of Pamir, 17 Northeastern region of China, 18 Northwestern region of China, 19 Centre-eastern China, 21 Southern region of China, 22 Centre western China, 23 Russia, 24 Eastern Siberia, 25 Mongolia, 26 Plateau of Pamir, 27 Northeastern region of China, 28 Northwestern region of China, 29 Qinghai - Xizang region of China, 30 Southern region of China, 31 Indonesia, 32 New Guinea, 33 Islands of Pacific Ocean, 34 Northern Africa, 35 Western Africa, 36 Central Africa, 37 Congo river basin, 38 Ethiopia region, 39 Tanzania region, 40 Angola region, 41 South Africa, 42 Madagascar, 43 Western Australia, 44 Northern Territory, 45 South Australia, 46 Queensland, 47 New South Wales, 48 Victoria, 49 Tasmania, 50 New Zealand, 51 Eastern Canada, 52 Western Canada, 53 Mts. Eastern US, 54 Plain Central US, 55 Hills Central US, 56 Mts. Western US, 57 Mexico, 58 Central America region, 59 Caribbean Islands, 60 Venezuela, 61 Plateau Guyana, 62 Northern Mt. Andes, 63 Amazon Plain, 64 Plateau Brazil, 65 Bolivia, 66 Argentina, 67 South Mt. Andes

Table 2: The number of medical important insect genera of BGUs in the World

<table>
<thead>
<tr>
<th>BGU</th>
<th>Number of genera</th>
<th>BGU</th>
<th>Number of genera</th>
<th>BGU</th>
<th>Number of genera</th>
<th>BGU</th>
<th>Number of genera</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>253</td>
<td>07</td>
<td>208</td>
<td>17</td>
<td>446</td>
<td>31</td>
<td>314</td>
</tr>
<tr>
<td>02</td>
<td>316</td>
<td>08</td>
<td>208</td>
<td>18</td>
<td>314</td>
<td>32</td>
<td>315</td>
</tr>
<tr>
<td>03</td>
<td>245</td>
<td>09</td>
<td>196</td>
<td>19</td>
<td>314</td>
<td>33</td>
<td>103</td>
</tr>
<tr>
<td>04</td>
<td>271</td>
<td>10</td>
<td>314</td>
<td>20</td>
<td>127</td>
<td>34</td>
<td>127</td>
</tr>
<tr>
<td>05</td>
<td>99</td>
<td>11</td>
<td>271</td>
<td>21</td>
<td>271</td>
<td>35</td>
<td>124</td>
</tr>
<tr>
<td>06</td>
<td>143</td>
<td>12</td>
<td>245</td>
<td>22</td>
<td>245</td>
<td>36</td>
<td>103</td>
</tr>
<tr>
<td>07</td>
<td>208</td>
<td>13</td>
<td>99</td>
<td>23</td>
<td>99</td>
<td>37</td>
<td>131</td>
</tr>
<tr>
<td>08</td>
<td>81</td>
<td>14</td>
<td>143</td>
<td>24</td>
<td>143</td>
<td>38</td>
<td>207</td>
</tr>
<tr>
<td>09</td>
<td>75</td>
<td>15</td>
<td>208</td>
<td>25</td>
<td>208</td>
<td>39</td>
<td>292</td>
</tr>
<tr>
<td>10</td>
<td>211</td>
<td>16</td>
<td>81</td>
<td>26</td>
<td>81</td>
<td>40</td>
<td>457</td>
</tr>
<tr>
<td>11</td>
<td>246</td>
<td>17</td>
<td>75</td>
<td>27</td>
<td>75</td>
<td>41</td>
<td>457</td>
</tr>
<tr>
<td>12</td>
<td>165</td>
<td>18</td>
<td>211</td>
<td>28</td>
<td>211</td>
<td>42</td>
<td>125</td>
</tr>
<tr>
<td>13</td>
<td>315</td>
<td>19</td>
<td>246</td>
<td>29</td>
<td>246</td>
<td>43</td>
<td>125</td>
</tr>
<tr>
<td>14</td>
<td>124</td>
<td>20</td>
<td>165</td>
<td>30</td>
<td>165</td>
<td>44</td>
<td>125</td>
</tr>
<tr>
<td>15</td>
<td>103</td>
<td>21</td>
<td>124</td>
<td>31</td>
<td>124</td>
<td>45</td>
<td>125</td>
</tr>
<tr>
<td>16</td>
<td>127</td>
<td>22</td>
<td>103</td>
<td>32</td>
<td>103</td>
<td>46</td>
<td>125</td>
</tr>
<tr>
<td>17</td>
<td>446</td>
<td>23</td>
<td>127</td>
<td>33</td>
<td>127</td>
<td>47</td>
<td>125</td>
</tr>
<tr>
<td>18</td>
<td>314</td>
<td>24</td>
<td>446</td>
<td>34</td>
<td>446</td>
<td>48</td>
<td>125</td>
</tr>
</tbody>
</table>

*AR (average richness): BDR/BBU: **ADT (average distributional territory): BDR/BBU

c) Clustering methods

Shen’s SGF is defined as follows: similarity coefficient among multiple regions is the ratio of the average species number shared by all regions to the total species number (Shenet al., 2008a):

\[ S_{ln} = \frac{\sum H_n S_n}{\sum (S_n - T_n) / n S_n} \]

In this formula, \( S_{ln} \) is the similarity coefficient of \( n \) BGUs: \( S_n \), \( H_n \), and \( T_n \) are the BGU species number, common species number and unique species number, respectively, and \( H_n = S_n - T_n \); \( S_n \) is the total species number in \( n \) BGUs. For calculation, all values were obtained from the database search page. This was convenient for both manual and computer calculations.

The MSCA indicated that the similarity coefficient of any group could be calculated directly and not restrained by the clustering order. It even was possible the first to calculate the total similarity coefficient of the 67 BGUs.

For example, we calculated the similarity coefficient of four BGU from Europe (Fig. 2). The 4066 in the first column of the first row was the number of genera that have not distribution by all four BGUs. The first number in other each column was the unique species number of every BGU. The number 465, which was the total genus number 4531 minus 4066, was the total species number of all five BGUs. The genus numbers of four BGUs were 253, 316, 245, and 271 as shown in Table 2. Using a calculator for these steps:

\[ \frac{253 + 316 + 245 + 271 - 30 - 59 - 7 - 68}{4} \times \frac{465}{465} = 0.495 \]

The process was simple compared to the processes of other clustering methods.
To compare the analysis results, three of the most common hierarchical clustering methods were used (Kreft, 2010):

1) The single linkage method, also called nearest neighbor method, using the Jaccard (1901) similarity formula: \( SI = \frac{C}{A+B-C} \), which was the most basic clustering method;

2) Average group linkage method, which was also called unweighted pair group means algorithm (UPGMA) method, using the Szymkiewicz (1934) similarity formula, also called the Simpson (1947) formula: \( SI = \frac{C}{\min(A,B)} \). Using this method, better results can be obtained, but the calculation process is complicated.

3) Sum of squares method (Ward’s method), using the Czekanowski (1913) similarity formula (also called Sørensen (1948) formula): \( SI = \frac{2C}{A+B} \). Using this method, better results can be obtained, but the calculation process is complicated.

The three similarity formulas were subjected to pairwise comparisons. A and B were species numbers in two regions and C was the species number shared by two regions.

III. Results

MSCA clustering results (Fig. 3) showed that the 67 BGUs total similarity coefficient was 0.089. At 0.370 similarity coefficient level, 67 BGUs were clustered as approximately 20 small unit groups, at 0.250 similarity coefficient level, 20 small unit groups were clustered as approximately 7 big unit groups. Unit of each group was neighbor and connected to each other, corresponding to the geographical principles. The similarity level was greater within the group than among different groups and the ecological condition of each group was independent, corresponding to the principles of statistics and ecology.

Clustering results showed high consistency with world insects clustering results (Shen et al., 2018). The numbers of big and small unit groups were the same, components of each big and small group were almost the same, and the structures between the groups were consistent. One difference was that the total similarity coefficient and similarity level of large and small unit groups of medically important insects were higher than all insects. This may be because medically important insects have generated more attention and research. The clustering location of individual units had moved: Unit 25# moved from the g small unit group to the f small unit group. Unit 37# moved from the h small unit group to the i small unit group. Unit 74# moved from the r small unit group to the s small unit group. These movements were between neighboring groups, consistent with geographical principles.

We compared mammalian division scheme (Wallace, 1876) and except for the D big unit group that was the same in the Ethiopian realm, all other groups were different. A and B big unit groups divided the Palearctic realm into east and west sections. C and E big groups incorporated New Guinea and Pacific islands from the Australian realm into the Oriental realm. F and G big groups incorporated Central America in the Neotropic realm into the Nearctic realm.

Compared with current world plant division proposed by Cox (Cox, 2001), the C, D, E big unit groups were the same as the India–Pacific kingdom, Afrotropic kingdom, and Australian kingdom. The difference was that the A, B and F big unit groups divided the Holarctic realm into three parts. F and G big groups categorized Central America in the Neotropic kingdom into the Nearctic kingdom.

Compared with the current several insect groups division, our results support the following: the Palearctic realm is divided into two parts by Trichoptera and Aleyrodidae (Morse et al., 2011; Evans, 2007). The Siphonaptera and Trichopteracategorize New Guinea and the Pacific Islands into the Oriental realm (Vashchonok et al., 2013; Moor et al., 2008). The results also support separating Pacific Islands from the Australia realm by Staphylinidae, Aleyrodidae and aquatic insects (Herman et al., 2001; Evans, 2007; Balian et al., 2008); incorporation of Yemen and Oman into the Palearctic realm by Symphyta and Culicidae (Taeger et al., 2010; Silver, 2004); And incorporation of Mexico into Nearctic realm by the Culicidae (Silver, 2004).
However, the results did not support assigning New Zealand, Madagascar and Antarctic as separate realms.

The traditional methods did not produce accurate, sensitive, and precise results. The results of the single linkage method (Fig. 4) were chaotic with no distinct layers. Many geographical units could not be clustered, such as units 20, 31, 49, 58, 69, 78, that were called “noise”. The average group linkage method (Fig. 5) was better than the single linkage method and removed most of the “noise”. At the distant level of 0.63 the BGUs could be clustered into six unit groups, five of which had significant geographical meaning. The letters corresponding to the areas in Fig. 3, the largest group which was composed of 26 BGUs, were chaotic and lacked geographical values. More precise division did not improve this. The sum of squares method (Fig. 6) had better clustering results. At the distance of 1.2, the BGUs could be clustered into eight groups and the first seven had geographical meaning. The last group did not conform to the principle of geography and it was difficult to achieve precise clustering.

**Fig. 3:** Clustering tree of Medical insect of World by MSCA
IV. Discussion

This study demonstrated that the distribution pattern of medically important insects is consistent with that of phytophagous insects. However it is indisputable fact that the medical insects have a close relationship of food chain with higher animals. Therefore, we can speculated that most insects are phytophagous and the distribution pattern is the same as for plants. Although the feeding habits of mammals are complex and some
of them are carnivorous, the final food sources are plants. Their distribution pattern should be the same as plants. Thus, medical important insects belong to the bi-direction food chain of animals → plants ← phytophagous insects and showed the same results as the total insect distribution. Without doubt this hypothesis requires confirmation from quantitative analysis of plants and mammals and the first step would be to select and standardize the different methods. The comparisons made in this study showed that MSCA method can be useful. We look forward to establishing a consistent model of the distribution patterns of plants, mammals, and insects across the world.

Acknowledgments

This study was supported by the key laboratory foundation of Henna (112300413221). We thank also C. B. Cox (UK), H. Kreft (Germany), M.V. Cianciaruso (Brazil), D.R. Gustafsson (USA), P. Vrsansky (Slovak), J.C. Beaucournu (France), T. Najer (Czech), R.J. Whitaker (UK), M. Caesar (France), M. P. Valim (Brazil) for presenting references, or proposing suggestions, or modifying the manuscripts.

References Références Referencias


