Undersampling bias: the null hypothesis for singleton species in tropical arthropod surveys

Jonathan A. Coddington1, Ingi Agnarsson1,2, Jeremy A. Miller4, Matjaž Kuntner1,3 and Gustavo Hormiga5

1Department of Entomology, National Museum of Natural History, NHB-105, Smithsonian Institution, PO Box 37012, Washington, DC 20013-7012, USA; 2Department of Biology, University of Puerto Rico, P.O. Box 23360, San Juan, PR 00931-3360, Puerto Rico; 3Institute of Biology, Scientific Research Centre of the Slovenian Academy of Sciences and Arts, Novi trg 2, PO Box 306, SI-1001 Ljubljana, Slovenia; 4Nationaal Natuurhistorisch Museum Naturalis, Darwinweg 2, 2333 CR Leiden, The Netherlands; and 5Department of Biological Sciences, The George Washington University, 2023 G Street NW, Washington D.C., 20052, USA.

Summary

1. Frequency of singletons – species represented by single individuals – is anomalously high in most large tropical arthropod surveys (average, 32%).
2. We sampled 5965 adult spiders of 352 species (29% singletons) from 1 ha of lowland tropical moist forest in Guyana.
3. Four common hypotheses (small body size, male-biased sex ratio, cryptic habits, clumped distributions) failed to explain singleton frequency. Singletons are larger than other species, not gender-biased, share no particular lifestyle, and are not clumped at 0.25–1 ha scales.
4. Monte Carlo simulation of the best-fit lognormal community shows that the observed data fit a random sample from a community of ~700 species and 1–2 million individuals, implying approximately 4% true singleton frequency.
5. Undersampling causes systematic negative bias of species richness, and should be the default null hypothesis for singleton frequencies.
6. Drastically greater sampling intensity in tropical arthropod inventory studies is required to yield realistic species richness estimates.
7. The lognormal distribution deserves greater consideration as a richness estimator when undersampling bias is severe.

Key-words: Araneae, Guyana, neutral theory, sampling intensity, species richness estimation, spiders
Table 1. Summary of tropical arthropod surveys. Arthropod surveys from tropical forest sites reporting total abundance (abun., or species presence per sample for ants, Agosti et al. 2000), species richness (spp.), and singletons (reported, calculated from figures given, or approximated as Fisher’s α, noted in source column). Intensity is abun./spp. A search of Biological Abstracts (1986–2007) on the terms (species richness) and (Arthropoda) and (Oriental region or Australasian region or Neotropical region or Ethiopian region) produced 514 results, many of which did not provide the required inventory statistics or were not from wet tropical sites. Those meeting our criteria, in addition to those known to us personally, are listed below. References for this table are listed in the Appendix.

<table>
<thead>
<tr>
<th>Taxon</th>
<th>Study site</th>
<th>Abun.</th>
<th>Spp.</th>
<th>Singletons</th>
<th>Intensity</th>
<th>Percentage of Singletons</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arthropods</td>
<td>Australia</td>
<td>20 507</td>
<td>759</td>
<td>271</td>
<td>27</td>
<td>36</td>
<td>Basset &amp; Kitching 1991</td>
</tr>
<tr>
<td>Insecta</td>
<td>Costa Rica (Area 1)</td>
<td>488</td>
<td>142</td>
<td>91</td>
<td>3</td>
<td>64</td>
<td>Janzen &amp; Schoener 1968</td>
</tr>
<tr>
<td>Insecta</td>
<td>Costa Rica (Area 2)</td>
<td>1362</td>
<td>262</td>
<td>165</td>
<td>5</td>
<td>63</td>
<td></td>
</tr>
<tr>
<td>Insecta</td>
<td>Costa Rica (Area 3)</td>
<td>4857</td>
<td>404</td>
<td>254</td>
<td>12</td>
<td>63</td>
<td></td>
</tr>
<tr>
<td>Insecta</td>
<td>Costa Rica (Area 4)</td>
<td>1339</td>
<td>545</td>
<td>390</td>
<td>2</td>
<td>72</td>
<td></td>
</tr>
<tr>
<td>Insecta</td>
<td>New Guinea</td>
<td>80 062</td>
<td>1050</td>
<td>278</td>
<td>76</td>
<td>26</td>
<td>Novotný &amp; Basset 2000</td>
</tr>
<tr>
<td>(leaf-chewing+ sap-sucking) Insecta</td>
<td>Guyana</td>
<td>27 735</td>
<td>604</td>
<td>229-5</td>
<td>46</td>
<td>38</td>
<td>Basset et al. 2001 (singletons calculated)</td>
</tr>
<tr>
<td>Blattaria</td>
<td>Panama (BCI)</td>
<td>3224</td>
<td>79</td>
<td>15</td>
<td>41</td>
<td>19</td>
<td>Wolda 1983 (Fisher’s α)</td>
</tr>
<tr>
<td>Coleoptera</td>
<td>Australia (Queensland)</td>
<td>10 000</td>
<td>1514</td>
<td>612</td>
<td>7</td>
<td>40</td>
<td>Montefith &amp; Davies 1984 (approx. values)</td>
</tr>
<tr>
<td>Coleoptera: Curculionidae</td>
<td>Panama (BCI)</td>
<td>28 521</td>
<td>703</td>
<td>131</td>
<td>5</td>
<td>19</td>
<td>Wolda 1987 (Fisher’s α)</td>
</tr>
<tr>
<td>Coleoptera: Pselaphidae, Anthicidae</td>
<td>Panama (BCI)</td>
<td>6482</td>
<td>114</td>
<td>19-7</td>
<td>57</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>Coleoptera</td>
<td>Panama (BCI)</td>
<td>34 705</td>
<td>597</td>
<td>102-5</td>
<td>58</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>Coleoptera</td>
<td>New Guinea</td>
<td>4840</td>
<td>633</td>
<td>321</td>
<td>8</td>
<td>51</td>
<td>Allison et al. 1993</td>
</tr>
<tr>
<td>Coleoptera</td>
<td>Peru (Tambopata)</td>
<td>15 869</td>
<td>3429</td>
<td>1728</td>
<td>5</td>
<td>50</td>
<td>Erwin 1997</td>
</tr>
<tr>
<td>Coleoptera</td>
<td>Sulawesi</td>
<td>18 000</td>
<td>1355</td>
<td>623</td>
<td>13</td>
<td>46</td>
<td>Hammond et al. 1997 (approx. values)</td>
</tr>
<tr>
<td>Coleoptera</td>
<td>Brazil</td>
<td>8454</td>
<td>993</td>
<td>446-9</td>
<td>9</td>
<td>45</td>
<td>Didham et al. 1998 (singletons calculated)</td>
</tr>
<tr>
<td>Coleoptera: Curculionidae</td>
<td>Honduras</td>
<td>26 891</td>
<td>293</td>
<td>38</td>
<td>9</td>
<td>13</td>
<td>Anderson &amp; Ashe 2000</td>
</tr>
<tr>
<td>Coleoptera: Staphylinidae</td>
<td>Honduras</td>
<td>7349</td>
<td>224</td>
<td>53</td>
<td>33</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>Coleoptera</td>
<td>Malaysia</td>
<td>8028</td>
<td>1711</td>
<td>823</td>
<td>5</td>
<td>48</td>
<td>Chung et al. 2000</td>
</tr>
<tr>
<td>Coleoptera</td>
<td>Uganda</td>
<td>29 736</td>
<td>1433</td>
<td>596</td>
<td>21</td>
<td>42</td>
<td>Wagner 2000</td>
</tr>
<tr>
<td>Coleoptera</td>
<td>Ecuador</td>
<td>2329</td>
<td>318</td>
<td>91</td>
<td>7</td>
<td>29</td>
<td>Lucky et al. 2002</td>
</tr>
<tr>
<td>Coleoptera: Scarabaeinae</td>
<td>Bolivia</td>
<td>4050</td>
<td>73</td>
<td>7</td>
<td>55</td>
<td>10</td>
<td>Spector &amp; Ayzama 2003</td>
</tr>
<tr>
<td>Coleoptera: Pselaphinae, Histeridae</td>
<td>Ecuador</td>
<td>3465</td>
<td>385</td>
<td>155</td>
<td>9</td>
<td>40</td>
<td>Carlton et al. 2004</td>
</tr>
<tr>
<td>Coleoptera: Phyllophagous Coleoptera</td>
<td>Panama</td>
<td>3009</td>
<td>364</td>
<td>139</td>
<td>8</td>
<td>38</td>
<td>Ødegaard 2004</td>
</tr>
<tr>
<td>Coleoptera</td>
<td>Ecuador</td>
<td>15 181</td>
<td>2001</td>
<td>397</td>
<td>8</td>
<td>20</td>
<td>Erwin et al. 2005</td>
</tr>
<tr>
<td>Coleoptera: Sciaridae</td>
<td>Colombia</td>
<td>7894</td>
<td>101</td>
<td>20</td>
<td>78</td>
<td>20</td>
<td>Escobar et al. 2005</td>
</tr>
<tr>
<td>Coleoptera</td>
<td>Brazil (Parana)</td>
<td>1883</td>
<td>518</td>
<td>266</td>
<td>4</td>
<td>51</td>
<td>Ganho &amp; Marinoni 2005</td>
</tr>
<tr>
<td>Coleoptera</td>
<td>Brazil (Parana)</td>
<td>1891</td>
<td>106</td>
<td>32</td>
<td>18</td>
<td>30</td>
<td>Linzmeier et al. 2006</td>
</tr>
<tr>
<td>Coleoptera: Atici</td>
<td>Australia</td>
<td>29 986</td>
<td>1473</td>
<td>526</td>
<td>20</td>
<td>36</td>
<td>Stork &amp; Grimbacher 2006</td>
</tr>
<tr>
<td>Diptera: Muscidae</td>
<td>Brazil (Parana)</td>
<td>7014</td>
<td>91</td>
<td>10</td>
<td>77</td>
<td>11</td>
<td>Costacurta et al. 2003</td>
</tr>
<tr>
<td>Diptera: Phoridae</td>
<td>Costa Rica</td>
<td>3341</td>
<td>115</td>
<td>20</td>
<td>29</td>
<td>17</td>
<td>Brown 2004</td>
</tr>
<tr>
<td>Diptera: Syrphidae</td>
<td>Brazil (Parana)</td>
<td>392</td>
<td>76</td>
<td>12</td>
<td>5</td>
<td>16</td>
<td>Marinoni et al. 2004</td>
</tr>
<tr>
<td>Ephemoptera</td>
<td>Panama (Corriente Grande)</td>
<td>7178</td>
<td>27</td>
<td>4</td>
<td>266</td>
<td>15</td>
<td>Wolda &amp; Flowers 1985 (Fisher’s α)</td>
</tr>
<tr>
<td>Ephemoptera</td>
<td>Panama (Miramar)</td>
<td>29 120</td>
<td>33</td>
<td>4</td>
<td>882</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Ephemoptera</td>
<td>Zaire</td>
<td>29 892</td>
<td>21</td>
<td>2</td>
<td>1423</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Hemiptera</td>
<td>Australia</td>
<td>6004</td>
<td>98</td>
<td>35</td>
<td>61</td>
<td>36</td>
<td>Andrew &amp; Hughes 2005</td>
</tr>
<tr>
<td>Homoptera</td>
<td>Panama (BCI)</td>
<td>22 046</td>
<td>458</td>
<td>82-1</td>
<td>48</td>
<td>18</td>
<td>Wolda 1987 (Fisher’s α)</td>
</tr>
</tbody>
</table>
Table 1. Continued

<table>
<thead>
<tr>
<th>Taxon</th>
<th>Study site</th>
<th>Abun.</th>
<th>Spp.</th>
<th>Singletons</th>
<th>Intensity</th>
<th>Percentage of Singletons</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homoptera</td>
<td>Panama (Pipeline Rd.)</td>
<td>1324</td>
<td>332</td>
<td>126</td>
<td>4</td>
<td>38</td>
<td>Wolda 1979</td>
</tr>
<tr>
<td>Hymenoptera:</td>
<td>Sulawesi</td>
<td>700</td>
<td>293</td>
<td>179</td>
<td>2</td>
<td>61</td>
<td>Noyes 1989</td>
</tr>
<tr>
<td>Hymenoptera:</td>
<td>Costa Rica (Monteverde)</td>
<td>3998</td>
<td>53</td>
<td>6</td>
<td>75</td>
<td>11</td>
<td>Longino &amp; Nadkarni 1990</td>
</tr>
<tr>
<td>Hymenoptera:</td>
<td>Costa Rica</td>
<td>7904</td>
<td>437</td>
<td>51</td>
<td>18</td>
<td>12</td>
<td>Longino et al. 2002</td>
</tr>
<tr>
<td>Hymenoptera:</td>
<td>Brazil (Minas Gerais)</td>
<td>1183</td>
<td>20</td>
<td>6</td>
<td>59</td>
<td>30</td>
<td>Nemesio &amp; Silveira 2006</td>
</tr>
<tr>
<td>Lepidoptera:</td>
<td>Malaysia</td>
<td>9031</td>
<td>620</td>
<td>118</td>
<td>15</td>
<td>19</td>
<td>Corbet 1942</td>
</tr>
<tr>
<td>Lepidoptera:</td>
<td>Malaysia</td>
<td>9461</td>
<td>1048</td>
<td>538</td>
<td>9</td>
<td>51</td>
<td>Barlow &amp; Woiwod 1989</td>
</tr>
<tr>
<td>Lepidoptera:</td>
<td>Ecuador</td>
<td>6690</td>
<td>130</td>
<td>20</td>
<td>5</td>
<td>15</td>
<td>DeVries et al. 1997</td>
</tr>
<tr>
<td>Lepidoptera:</td>
<td>Ecuador</td>
<td>883</td>
<td>91</td>
<td>22</td>
<td>10</td>
<td>24</td>
<td>DeVries et al. 1999</td>
</tr>
<tr>
<td>Lepidoptera:</td>
<td>Ecuador</td>
<td>11361</td>
<td>128</td>
<td>18</td>
<td>93</td>
<td>14</td>
<td>DeVries &amp; Walla 2001</td>
</tr>
<tr>
<td>Lepidoptera:</td>
<td>Borneo</td>
<td>485</td>
<td>53</td>
<td>16</td>
<td>9</td>
<td>30</td>
<td>Schulze et al. 2001</td>
</tr>
<tr>
<td>Lepidoptera:</td>
<td>Thailand</td>
<td>1936</td>
<td>53</td>
<td>4</td>
<td>37</td>
<td>8</td>
<td>Ghaouzi 2002</td>
</tr>
<tr>
<td>Lepidoptera:</td>
<td>Ecuador</td>
<td>23270</td>
<td>868</td>
<td>161</td>
<td>27</td>
<td>19</td>
<td>Hilt et al. 2006</td>
</tr>
<tr>
<td>Lepidoptera:</td>
<td>Peru (BCI)</td>
<td>1537</td>
<td>136</td>
<td>31</td>
<td>11</td>
<td>23</td>
<td>Louton et al. 1996</td>
</tr>
<tr>
<td>Lepidoptera:</td>
<td>Panama (BCCI)</td>
<td>1566</td>
<td>73</td>
<td>15</td>
<td>21</td>
<td>22</td>
<td>Wolda 1987 (Fisher’s α)</td>
</tr>
<tr>
<td>Lepidoptera:</td>
<td>Panama (BCI)</td>
<td>10092</td>
<td>148</td>
<td>20</td>
<td>68</td>
<td>14</td>
<td>Broadhead &amp; Wolda 1985 (Fisher’s α)</td>
</tr>
<tr>
<td>Lepidoptera:</td>
<td>Panama (Fortuna)</td>
<td>4301</td>
<td>84</td>
<td>10</td>
<td>15</td>
<td>12</td>
<td>*</td>
</tr>
<tr>
<td>Araneae</td>
<td>Bolivia (50 m)</td>
<td>875</td>
<td>191</td>
<td>89</td>
<td>5</td>
<td>47</td>
<td>Coddington et al., 1991, 1996</td>
</tr>
<tr>
<td>Araneae</td>
<td>Bolivia (1200 m)</td>
<td>1109</td>
<td>329</td>
<td>147</td>
<td>3</td>
<td>45</td>
<td>*</td>
</tr>
<tr>
<td>Araneae</td>
<td>Bolivia (2200 m)</td>
<td>654</td>
<td>158</td>
<td>70</td>
<td>4</td>
<td>44</td>
<td>*</td>
</tr>
<tr>
<td>Araneae</td>
<td>Brazil (Manaus)</td>
<td>75</td>
<td>62</td>
<td>32</td>
<td>1</td>
<td>52</td>
<td>Höfer et al. 1994</td>
</tr>
<tr>
<td>Araneae</td>
<td>Tobago</td>
<td>1777</td>
<td>98</td>
<td>27</td>
<td>18</td>
<td>28</td>
<td>Hormiga &amp; Coddington 1994</td>
</tr>
<tr>
<td>Araneae</td>
<td>Peru (Samiria)</td>
<td>5895</td>
<td>1140</td>
<td>520</td>
<td>5</td>
<td>46</td>
<td>Silva 1996</td>
</tr>
<tr>
<td>Araneae</td>
<td>Peru (Pakita)</td>
<td>2616</td>
<td>498</td>
<td>207</td>
<td>5</td>
<td>42</td>
<td>Silva &amp; Coddington 1996</td>
</tr>
<tr>
<td>Araneae</td>
<td>Costa Rica (understorey)</td>
<td>7144</td>
<td>86</td>
<td>11</td>
<td>83</td>
<td>13</td>
<td>Bodner 2002</td>
</tr>
<tr>
<td>Araneae</td>
<td>Tanzania (canopy)</td>
<td>9096</td>
<td>170</td>
<td>32</td>
<td>54</td>
<td>19</td>
<td>Sørensen et al. 2002</td>
</tr>
<tr>
<td>Araneae</td>
<td>Tanzania (Canopy)</td>
<td>5233</td>
<td>149</td>
<td>35</td>
<td>35</td>
<td>23</td>
<td>Sørensen 2003</td>
</tr>
<tr>
<td>Araneae</td>
<td>Malaysia</td>
<td>6999</td>
<td>578</td>
<td>145</td>
<td>12</td>
<td>25</td>
<td>Floren &amp; Deelman-Reinhold 2005, personal communication</td>
</tr>
<tr>
<td>Araneae</td>
<td>Mt. Cameroon (500 m)</td>
<td>573</td>
<td>231</td>
<td>93</td>
<td>2</td>
<td>40</td>
<td>Coddington et al., unpublished</td>
</tr>
<tr>
<td>Araneae</td>
<td>Mt. Cameroon (3000 m)</td>
<td>1555</td>
<td>55</td>
<td>14</td>
<td>28</td>
<td>25</td>
<td>*</td>
</tr>
<tr>
<td>Araneae</td>
<td>Peru (Tambopata)</td>
<td>1821</td>
<td>635</td>
<td>341</td>
<td>3</td>
<td>54</td>
<td>Coddington &amp; Silva, unpublished</td>
</tr>
<tr>
<td>Araneae</td>
<td>Peru (Manu)</td>
<td>222</td>
<td>123</td>
<td>78</td>
<td>2</td>
<td>63</td>
<td>Erwin &amp; Coddington, unpublished</td>
</tr>
<tr>
<td>Araneae</td>
<td>Guyana</td>
<td>5964</td>
<td>351</td>
<td>101</td>
<td>17</td>
<td>29</td>
<td>This study</td>
</tr>
<tr>
<td>Averages</td>
<td></td>
<td>9372</td>
<td>464</td>
<td>176</td>
<td>61·6</td>
<td>31·6</td>
<td></td>
</tr>
</tbody>
</table>

Singleton tropical arthropod species are anomalous for several reasons. First, minimum viable population sizes are conventionally at least 500 individuals (Gilpin & Soulé 1986). Second, many arthropods begin life clumped because eggs are clumped when laid – in spiders eggs are clustered within an egg sac. Most nonvolant arthropods are small and probably rarely travel hundreds or even dozens of metres to mate. Third, clumped distributions in nature are far more common than random or dispersed (Krebs 1999). While clumping certainly depends on scale, at hectare scales randomness is typical of canopy trees and jaguars, not small, nonflying, sedentary arthropods such as spiders.

Ad hoc explanations for singletons often invoke aspects of the biology of particular groups, such as host or food plant specificity (Price et al. 1995). In spiders, males of sedentary web-spinning species must wander to find females (potentially passing through atypical habitat patches, i.e., tourists), and are likely to be small and rare (Vollrath & Parker 1992). General explanations include source-sink phenomena or mass-effects (e.g. ‘ecological drift,’ Hubbell 2001) at both local (‘tourist’) and regional (‘waif’ or ‘vagrant’) scales (Schmida & Wilson 1985; Pulliam 1988; Southwood 1996; Stork & Hammond 1997; Novotný & Basset 2000; Magurran & Henderson 2003; Basset et al. 2004; Odengaard 2004). Time, space, or method ‘edge effects’ are also frequent explanations. Adults outside their breeding seasons are scarce, and if only adults are identifiable (true for spiders), will be artefactually rare (Ulrich 2001; Longino, Coddington, & Colwell 2002; Scharff et al. 2003; Basset et al. 2004). Nocturnality or seasonal migration could produce similar effects. Space edge effects are usually microhabitat preferences. Species patches just trespassing on plot boundaries might produce many ‘false’ singletons. Method edge effects are the accidental sampling of a species by an inappropriate method, such as a canopy species in a pitfall trap (Longino et al. 2002; Scharff et al. 2003). Finally, singletons may be absolutely rare, i.e. sparse with large nearest-neighbour distances throughout their range. Perhaps, as is now recognized for tropical trees (Pitman et al. 1999; Kenfack et al. 2006), we drastically underestimate the scale at which many tropical arthropod species live and ought to be sampled.

Undersampling bias and biological explanations are not mutually exclusive. However, if repeated random sampling of communities modelled on statistical parameters estimated from the sample mimic the observed results, undersampling should serve as the initial null explanation for high singleton frequencies (McGill 2003), analogous to the use of null models in other fields (Harte et al. 2001; Hubbell 2001). Variation not explained by undersampling may then be attributed to more complex causes.

Statistical methods to assess undersampling bias are relatively recent; quantitative estimates of its magnitude have been historically difficult, if not impossible, to obtain. Observed richness values are traditionally used for descriptive or comparative purposes (Groombridge 1992; Heywood & Watson 1995; Levin 2001). If high singleton frequencies indicate undersampling, however, then tropical arthropod communities are substantially larger than measured, and comparisons based on observed numbers are misleading. This has important implications for conservation biology, and also implies that typical inventories are under-resourced and/or poorly designed.

Here we use the results of an intensive 1-ha survey of spiders to test various explanations for high singleton frequency. Although spiders are typical sedentary arthropod predators and these results may apply only to that guild, high singleton frequencies also characterize inventories of other tropical arthropods (Table 1). Specifically, we test four process hypotheses and the null hypothesis of undersampling bias: singletons tend to be small and therefore missed; singletons tend to be males because as adults they travel further than females; nearest conspecific distances exceed 0.25–1 ha spatial scales (population structure is much larger than anticipated); singletons are ‘cryptic’ and hard to detect; and singletons are simply an artefact of undersampling because the scope of the survey exceeded sampling resources.

**Methods**

**STUDY SITE**

The study was carried out during 10 days, 5–14 July, 1999, in a primary lowland blackwater rainforest (1°36′46″N, 58°38′15″W) on the bank of the Essequibo River, 240 m elevation, 4.4 km south of Gunn’s landing. Upper Takutu-Upper Essequibo, Guyana. Four nested, concentric 0.25-ha subplots (total 1 ha) were established in uniform closed canopy forest (Fig. 1). Five experienced collectors worked simultaneously in the field during both day and night using a battery of collecting methods that broadly access most of the spider fauna (see Coddington et al. 1991; Sørensen, Coddington, & Scharff 2002; Scharff et al. 2003 for details).
CONSPECTIFIC DISTANCES

The nested subplot design provided a range of spatial scales. If singletons are spatial edge effects – multi-individual clumps or patches with only one individual in the subplot or plot – the outermost subplot with the largest perimeter should contain the most singletons (Fig. 1). Likewise, doubletons of spatially clumped species should occur in the same subplot. More generally, we tested for clumping by comparing the observed distribution of singletons and doubletons among subplots against the null hypothesis of equal frequency in all possible subplot combinations (e.g., A, B, C, or D for singletons, and AA, AB, AC, AD, BB, etc. for doubletons).

BODY SIZE

We measured total body length to the nearest 0·1 mm of one individual of each sex (when available) of each species. The ‘average’ tropical spider is then the mean of these lengths (assuming equal sex ratios) weighted by the relative abundance of each species. A t-test then compared mean size of singletons to nonsingletons.

SEX RATIO

Using the observed sex ratio in the total sample as the null, we compared the singleton sex ratio to it with a chi-squared test, both for singletons as a whole and for singletons of web-spinning species only, where males must wander to encounter the sedentary females.

CRYPTIC HABITS

Generally speaking, spider species within families are more alike in their biology than between (Coddington and Levi 1991). Araneids mostly spin orb-webs, mostly above the forest floor, but anapids and symphytognaths spin theirs mostly in the leaf litter. Philodromids run on leaves, and salticids jump between them (but only during the day). Because tropical 1-ha inventories usually find at least 30 families, testing the relative frequency of singletons among families against a null of the relative abundance of families in the total sample should detect whether singletons tend to have one lifestyle more frequently than another.

UNDERSAMPLING BIAS

We assessed inventory completeness by visually inspecting the average of 50 resamples of the observed species accumulation curve, as well as the singleton and doubleton curves, and four commonly used species richness estimators (Chao1, Chao2, ICE and ACE; Peterson & Slade 1998; Waltner & Martin 2001; Colwell 2005). In a complete inventory, the observed curve should asymptote and singletons should tend to zero, with doubletons lagging singletons. If incomplete but sufficient to estimate richness accurately, estimator curves should asymptote (Colwell & Coddington 1994). Constantly rising curves of all sorts imply incomplete inventories.

We also fit the data to a lognormal distribution using the method described in Scharff et al. (2003) and Longino et al. (2002). We use the lognormal as a reasonable null hypothesis (McGill 2003). Other models, such as the parameter-rich zero-sum multinomial, have been proposed as better fits to empirical data than the lognormal (e.g., Hubble 2001), however, a recent detailed test fails to support that claim, and indeed showed the opposite (McGill 2003). Given the high number of parameters in the zero-sum model, the cumbersome procedure of fitting the model, and lack of evidence for its superior fit to empirical data, McGill advised the preferential use of the lognormal as the simpler (more parsimonious) null model. Using the best-fit lognormal parameters, we generated and randomly sampled 6000 individuals from 1000 replicate communities of each of four sizes: 500, 600, 700, or 800 species, which parameters were chosen to mimic the empirical sample. We then compared the observed sample to these 4000 simulated samples on numbers of singletons, doubletons, and species. If the observed sample clearly deviated, undersampling bias alone does not explain high singleton frequency.

Statistical tests, curve fitting, modelling, resampling procedures, and species richness estimation used a combination of Systat 11 (Systat Software, Inc., Richmond, CA, USA), EstimateS 7·50 (Colwell 2005), and Microsoft Excel 2007 (Microsoft Corp., Redmond, WA, USA).

Finally, a few measures of leaf litter, shrub/subcanopy, and canopy tropical moist forest spider densities per m² have been published (Table 2). Extrapolated to 1 ha, these statistics provide crude estimates of spider abundance that can be compared to the total abundances predicted by the lognormal model.

SPECIMENS AND SORTING PROCEDURES

Each sample was labelled with plot, date, collector, method, and replicate number if two samples were otherwise identical. Team members (all arachnological taxonomists) or other experts on particular families sorted the specimens to morphospecies. All identifications of singletons and doubletons were checked and verified by at least two of the team members. Voucher specimens of each species identified in this study are deposited at the National Museum of Natural History (NMNH), Smithsonian Institution, Washington DC.

Results

The five collectors accumulated 300 samples over 10 days from the 1-ha plot (Table 3) containing a total of 5965 adults (and 6953 juveniles) of 352 species, of which 101 were singletons (29%) and 56 were doubletons. The most abundant species numbered 412. Inventory completion (observed richness/Chao1 estimate) ranged from 15% to 71% among methods, and overall was 79%. Sampling intensity (no. of ind./no. of spp.) ranged from 1·4 to 10·5 among methods and overall was 17. The survey compares favourably to other large published efforts in intensity and numbers of species encountered, considering that most spider species cannot be trapped (Table 1). However, the continually rising accumulation curves and richness estimators indicate that the inventory was still incomplete by the end of sampling (Fig. 2). The 95% upper confidence limit of the Chao2 estimator (itself only a lower-bound estimate), for example, was 520 species, but clearly had not reached a limit. True species richness in the hectare almost certainly exceeded 500 species, and probably much more.

The mean and standard deviation of the body lengths of adults collected was 2·89 ± 2·85 mm (thus an estimate of the average size of an adult lowland tropical moist forest spider). The mean singleton body length was 5·30 ± 4·67 mm. Singletons are significantly larger, not smaller, than the average species.

The overall male : female sex ratio in the sample was significantly female biased (1:1.3, \(P < 0.01\)). The overall singleton sex ratio was as biased as the total sample (1:1.7, \(P = 0.18\)). Sedentary web-spinner singletons were equally biased (1:2.8, \(P = 0.12\)). Singletons, therefore, are not disproportionately males of sedentary web-spinning species.

The distribution of doubletons across subplots (Fig. 1) was random (\(P = 0.82\)) as was the incidence of singletons from the centre to the outermost subplot (\(P = 0.80\)). Tripletons also showed no tendency to clump within subplots. Conspecific nearest neighbour distances, therefore, are not clumped at the coarse 0.25- to 1-ha scales tested here.

Singletons showed no taxonomic pattern, occurring in families in proportion to the latter’s relative abundance (\(P > 0.99\)). If undersampling bias varied according to lifestyle defined as family identity, the effect was not detectable at this level of sampling intensity.

The observed data fit the lognormal distribution well (Fig. 3, 0.9 > \(P > 0.5\)). The predicted number of species in the modal octave \(S_0\) was 76.4 ± 13 (\(\mu = 6.2562\)), the variance term

---

**Table 2.** Spider density in tropical forests. Estimated number of total and adult spiders in a hectare of primary tropical rainforest. Min, computed from minimum values; max, from maximum values. References for this table are listed in the Appendix

<table>
<thead>
<tr>
<th>Layer</th>
<th>Place</th>
<th>Method</th>
<th>Density (n/m²)</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ground</td>
<td>Brazil</td>
<td>Berlese</td>
<td>221</td>
<td>Höfer &amp; Brescovit 2001</td>
</tr>
<tr>
<td></td>
<td>Brazil</td>
<td>Berlese</td>
<td>330</td>
<td>Höfer &amp; Brescovit 2001</td>
</tr>
<tr>
<td></td>
<td>Brazil</td>
<td>Berlese</td>
<td>129</td>
<td>Morais 1985</td>
</tr>
<tr>
<td></td>
<td>Brazil</td>
<td>Berlese</td>
<td>108</td>
<td>Adis &amp; Schubart 1984</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td></td>
<td>197.0</td>
<td></td>
</tr>
<tr>
<td>Shrub</td>
<td>Brazil</td>
<td>Berlese</td>
<td>0.5</td>
<td>Höfer &amp; Brescovit 2001</td>
</tr>
<tr>
<td></td>
<td>Brazil</td>
<td>Berlese</td>
<td>2.7</td>
<td>Höfer &amp; Brescovit 2001</td>
</tr>
<tr>
<td></td>
<td>Guyana</td>
<td></td>
<td>2.7</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>Guyana</td>
<td></td>
<td>1.9</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td></td>
<td>2.2</td>
<td></td>
</tr>
<tr>
<td>Canopy</td>
<td>Brunei</td>
<td>Fog</td>
<td>4.7</td>
<td>Russell-Smith &amp; Stork 1994</td>
</tr>
<tr>
<td></td>
<td>Sulawesi</td>
<td>Fog</td>
<td>4.6</td>
<td>Russell-Smith &amp; Stork 1995</td>
</tr>
<tr>
<td></td>
<td>Australia</td>
<td>Fog</td>
<td>8.6</td>
<td>Basset 1990, 1991</td>
</tr>
<tr>
<td></td>
<td>Tanzania</td>
<td>Fog</td>
<td>4.8</td>
<td>Sørensen 2003</td>
</tr>
<tr>
<td></td>
<td>Brazil</td>
<td>Fog</td>
<td>4.8</td>
<td>Adis 1984</td>
</tr>
<tr>
<td></td>
<td>Brazil</td>
<td>Fog</td>
<td>2.0</td>
<td>Höfer et al 2001</td>
</tr>
<tr>
<td></td>
<td>Brazil</td>
<td>Fog</td>
<td>5.5</td>
<td>Höfer et al. 1994</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td></td>
<td>5.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total spiders (1 m² column)</td>
<td>204</td>
<td>71</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Min spiders (1 m² column)</td>
<td>111</td>
<td>36</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Max spiders (1 m² column)</td>
<td>342</td>
<td>129</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total spiders (ha)</td>
<td>2,041,500</td>
<td>708,333</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Min spiders (ha)</td>
<td>1,105,000</td>
<td>360,000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Max spiders (ha)</td>
<td>3,421,000</td>
<td>1,291,000</td>
<td></td>
</tr>
</tbody>
</table>

**Table 3.** Collecting methods and results. AE, BE, CR, GR, PF, SW, D and N stand for aerial, beating, cryptic, ground, pitfall, sweeping, day, and night collecting methods, respectively (see text). Sample intensity is total individuals/total species. Inventory completion is total species/Chao1 estimate

<table>
<thead>
<tr>
<th>Layer</th>
<th>Method</th>
<th>Total</th>
<th>Adults</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ground</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shrub</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Canopy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The overall male : female sex ratio in the sample was significantly female biased (1:1.3, \(P < 0.01\)). The overall singleton sex ratio was as biased as the total sample (1:1.7, \(P = 0.18\)). Sedentary web-spinner singletons were equally biased (1:2.8, \(P = 0.12\)). Singletons, therefore, are not disproportionately males of sedentary web-spinning species.

The distribution of doubletons across subplots (Fig. 1) was random (\(P = 0.82\)) as was the incidence of singletons from the centre to the outermost subplot (\(P = 0.80\)). Tripletons also showed no tendency to clump within subplots. Conspecific nearest neighbour distances, therefore, are not clumped at the coarse 0.25- to 1-ha scales tested here.

Singletons showed no taxonomic pattern, occurring in families in proportion to the latter’s relative abundance (\(P > 0.99\)). If undersampling bias varied according to lifestyle defined as family identity, the effect was not detectable at this level of sampling intensity.

The observed data fit the lognormal distribution well (Fig. 3, 0.9 > \(P > 0.5\)). The predicted number of species in the modal octave \(S_0\) was 76.4 ± 13 (\(\mu = 6.2562\)), the variance term
Undersampling bias and singleton species

Figure 2. Four species richness estimators (ACE, ICE, Chao 1, Chao 2), the 95% upper confidence limit of the Chao2 estimator, observed, singleton, and doubleton curves for data of Table 3.

Figure 3. Lognormal fit (0.9 > P > 0.5) to data of Table 3. Predicted community size is 694 species. Note the over-estimation of abundant species at the right-hand tail.

Figure 4. Singleton, doubleton, and observed species totals from the Guyana study (open circle, arrow) and 25 random samples of 6000 individuals from model lognormal communities of 500, 600, 700, and 800 species.
and 120 000 individuals (intensity 340) from the 700 species
lognormal community, thus 10 and 20 times the actual sampling
effort. At an intensity of 170, percentage of singletons was 14%,
and the Chao and coverage estimators were 595–600 species
with Chao upper confidence intervals of 636 species – still
short of the true 700 species richness. At an intensity of 340,
the Chao and coverage estimators were 650–663 species, with
a Chao upper confidence interval of 702 species – thus just
covering the true richness value – and percentage of single-
tons fell to 9%.

Figure 5 depicts the logarithmic decline in singletons with
sampling intensity for the data of Table 1, and predicts zero
singletons at sampling intensities of roughly 1100. Sampling
the model community at that intensity yielded on average 4% singletons and 658 species observed.

We present what few data exist on tropical spider densities
in Table 2. Ignoring differences due to locality and construed
as a ground to canopy vertical 1 m² column, the leaf litter
contains most individuals, the canopy/subcanopy less, and
the shrub/understorey layer least. Given the decrease in leaf
area or other substrate with height above the ground, the
decline is plausible. It predicts, extremely roughly, about 2
million total spiders per hectare of tropical forest (range 1 1–
3 4 M). The modelled lognormal populations ranged between
1·2 and 3·3 million individuals, which agrees with Table 2.

Discussion

Rare species and estimating total species richness in particu-
lar are difficult statistical problems (Bunge & Fitzpatrick 1993;
Ulrich 2001; Brose, Martinez, & Williams 2003; Magurran
2004; Ellison & Agrawal 2005; Cunningham & Lindenmayer
how many data are required to obtain robust and reliable
species richness estimates is also difficult (Keating et al. 1998;
McArdle 1990). This complicates inventory design. Model-
ing studies have suggested that nonparametric richness
estimators do not begin to cover the true value until about
two-thirds to four-fifths of the species have been observed
(Walther & Morand 1998; Mao & Colwell 2005). On the other
hand, intensely sampled communities usually are lognormal
at local scales, even if the full distribution is truncated by failure
to detect rare species (Sugihara 1980; Longino et al. 2002;
Connolly et al. 2005, but see Williamson & Gaston 2005).

In this study, the empirical sample of 6000 individuals may
have included only half the species present, with singletons
comprising 29% of species observed. Nonparametric richness
estimators suggested only 443–460 species, a shortfall of 35%
compared to the lognormal estimate. While any singleton may
have been due to any of the process explanations discussed
above, the simplest explanation for the high frequency is
undersampling. As sampling continues and singleton fre-
nencies drop, biological explanations become more plausible.

Two ‘biological’ explanations were statistically significant,
but neither in the direction hypothesized. Singletons were
significantly larger (not smaller) than the average spider.
Twenty-three singletons over 7 cm caused that difference.
These were mostly large cursorial species (including ctenids,
sparassids, and miturgids) for which absolute densities of one,
or very few individuals per hectare are plausible. Singletons
are also disproportionately females, not males, but the sample
in general was female-biased, and singletons no more so,
even among sedentary web-spinning species where the presumed
bias towards singleton wandering males should have been
most pronounced. Adult male spiders are relatively short
lived and wandering males experience exceptionally high
mortality (Vollrath & Parker 1992); both of these factors
likely contribute to a female-biased sex ratio in the inventory
data, even if the sex ratio at birth were even (as they are for
most spiders examined to date, see Avilés & Maddison 1991,
Avilés, McCormack, Cutter & Bukowski 2000).

The other explanations tested, lifestyle, spatial edge effects,
and clumping of individuals at 1-ha scales and below, were
insignificant. Novotný & Basset (2000) and Ulrich (2001) also
found that few biological explanations of singletons were
supported. Magurran & Henderson (2003) use a 21-year data
set on a temperate fish community of 80 species to show that
about a third to a half of the species accumulated over that
time-span were tourists or waifs. In any given short-term
sampling event, however, presumably few of the rare species
would have been tourists or waifs. In a spider inventory of a
‘known’ fauna, Scharff et al. (2003) hypothesized 58% of
singletons as phenological, methodological, or spatial edge
effects, but they did not test the null hypothesis of undersam-
pling bias. For relatively instantaneous events such as this
inventory, singleton frequencies are about what one would
expect from random samples of a lognormally distributed
community – in this case, of about 700 species. The null
hypothesis of undersampling bias cannot be rejected.

This was an intense, short-term inventory (300 person-hours),
designed to yield an ‘instantaneous’ richness estimate that
avoided the confounding effect of phenological change.
Especially in relatively aseasonal tropical habitats, sampling year round or for multiple years might yield a more complete inventory over and above the effect of greater sampling intensity (DeVries, Walla, & Greeney 1999; Scharff et al. 2003). Increasing the sampling area might also improve efficiency, especially if, as perhaps suggested by the significantly larger singleton size and the possibility that some true singletons occur in any given hectare, we underestimated the scale at which sedentary tropical arthropods should be sampled. Their lifetime ranges may encompass much larger areas. On the other hand, species richness increases logarithmically with area (Rozenzweig 1995), burdening species richness estimates. Regardless, the key point is that the scope of the inventory must be carefully matched to available resources.

What little we know of tropical spider communities broadly agrees with the predictions of the lognormal fit (Table 2). Our empirical sample included only nine of 23 predicted octaves, yet the implied community, when randomly sampled at the same intensity, compared well to empirical observations of total species, numbers of singletons and doubletons, maximum abundance, and total numbers of individuals (Fig. 4). None of the collecting methods used in Table 2 are completely efficient, therefore, the actual hectare abundance of spiders is probably higher.

When the modelled 700 species community was sampled at an intensity of 1100, on average 658 species and 4% singletons resulted. Lognormal distributions always predict some singletons (here on average 28 or 4%), and stochastic replicates never contain all 700 species (here on average 685). Practically speaking, sampling intensities of 1100 detects just about as many species as stochastic models provide.

For these data, a sampling intensity of 340 (10 times the actual sampling effort) was just sufficient to include the known richness within the upper bound of the Chao2 estimator. This implies that inventories, as a rule of thumb, should aim for intensities between that and 1100 to obtain realistic nonparametric estimates of species richness.

Richness estimators are relatively more efficient if they can report the true richness based on relatively few data. The efficiency of available nonparametric richness estimators is poor in the sense that roughly three quarters of the community must be observed before the estimator confidence interval actually covers the true value (Walther & Morand 1998). Chao estimators, moreover, have a maximum upper bound of about half the square of the observed richness (if the sample of n species contains n-1 singletons or uniques and one doubleton or duplicate), but in practice such efficiencies are never achieved because of the improbability of so biased a sample.

The lognormal distribution can potentially result in higher richness estimates than nonparametric approaches (given the same data) because it assumes the relative abundance distribution is symmetric around the modal octave (Sugihara 1980; Longino et al. 2002), and therefore tends to at least double the observed richness. A number of authors argue that empirical communities show an asymmetric excess of rare species (Nee, Harvey, & May 1991; Nekola & Brown 2007), and Hubbell and co-workers argue from first principles that such is expected (Hubbell 2001; Volkov et al. 2003). However, McGill (2003) suggests that this observed skew in species abundance distributions may also be a sampling artefact. One might also point out that the lognormal even less realistically overestimates the abundant tail of the distribution (Fig. 3, Longino et al. 2002; Magurran & Henderson 2003). However, even if the lognormal slightly underestimates rare species, that error is small compared to the gross negative bias of nonparametric estimators at small sample sizes.

The stochastic variation in small samples drawn from the same lognormal population is impressive (Fig. 4). For the 700 species case, 1000 draws of 6000 individuals produced singleton counts of 62–134 and observed richnesses from 121–414, which comfortably cover the observed statistics of 101 and 351. The lognormal distribution may still be a useful method to estimate species richness under circumstances where many data are available, yet not enough for nonparametric estimators to function well. Unlike the relative abundance distribution-based estimators of Ulrich (1999, 2005), it does not require an explicit ratio of sampled to total habitat area, and thus is more practical in the field.

If general, this result implies that even large survey efforts (Table 1, Fig. 5) continue to undersample tropical arthropod biodiversity by perhaps a factor of 2 if singletons average 32% of the total. In many surveys, the figure is much higher (Table 1). Undersampling is a serious issue even for large mammal and bird surveys, where singletons average 16% (Bernard & Fenton 2002; Shankar and Sukamar 2002; McCain 2004). Consequently, typical surveys will underestimate species richness, with obvious implications for our understanding of biodiversity, and for any conservation decisions based on such data.

In summary, it appears that most tropical arthropod biodiversity surveys have been severely under-resourced if their goal was to census or estimate species richness of a defined taxonomic community at a particular place and time. Reliable methods do exist to estimate how many data are required to estimate many ecological statistics (Krebs 1999; Magurran 2004), but species richness historically is an exception. One may hope that future statistical research will improve estimator efficiency, but in the meantime the use of existing estimators dramatically exposes the gap between inventory design as implemented and the minimum necessary to obtain reliable richness estimates. Here the lognormal was more efficient than nonparametric estimators, and perhaps should be used more frequently. Species richness estimators are increasingly used in basic research to detect undersampling bias; results thus far suggest that it is ubiquitous and severe. Rather than scaling back inventory goals, we suggest that inventory analyses continue to assess undersampling bias in order to justify the budgets required to obtain adequate data. Funding sources and consumers of these essential data can scarcely argue that inadequate results are acceptable. If results continue to demonstrate that much greater sampling intensities are required, such will eventually become the norm, rather than the exception.
Acknowledgements

Thanks to Vicky Funk, Carol Kelloff, David Clarke, Tom Hollowell, and Romeo Williams for logistical support, and Rick West, Gita Bodner, G. B. Edwards, Martín Ramírez, Fernando Alvarez-Padilla, Scott Lacher, and Dana DeRoche for specimen work and identification. We are grateful for the hospitality of the Wuai community of Southern Guiana. We thank Robert Colwell, Phil DeVries, Jack Longino, and Anne Magurran for comments on an earlier draft. Support for this research came from a Smithsonian ‘Biodiversity of the Guianas’ grant to J. A. Coddington, a Smithsonian Neotropical Lowlands grant to J. Coddington, and a National Science Foundation grant to G. Hormiga and J Coddington (DEB 9712353).

References

Appendix 1. References for Tables 1 and 2


