Supplementary information

Table of contents:

Supplementary tables
Table S1. Sample size information, including the number of individuals and fecal samples used in analyses of the dataset rarefied to 3,000 reads.
Table S2. Unweighted UniFrac dissimilarity comparison within and between mammalian orders or diet types.
Table S3. Diet items included in each diet category.
Table S4. CCA analysis of environment and host factors for the 3,000-read dataset.
Table S5. Best-supported generalized linear mixed model (Poisson-link) explaining variation in abundance of the four most common bacteria phyla for the subset of 76 samples with diet data. Individual identity is a random effect in all models.
Table S6. Mantel test of correlation between sampling time interval and microbiota weighted UniFrac dissimilarity between samples that were collected from the same individual.
Table S7. Genus level abundance table of 47 adult baboon samples used in the enterotype analysis.

Supplementary figure legends

Figure S1. The distribution patterns of the unclassified and classified OTU at the phylum level across the samples. OTUs have been sorted into bins based on their prevalence in the samples (X-axis). Y-axis is the count of OTUs in each bin.

Figure S2. PCoA analysis of the weighted UniFrac dissimilarities comparing baboon gut microbiota. Each point corresponds to a sample colored by (A) individual identity, (B) sex, (C) ageclass and (D) season, (E) diet group. Baboons with diet composition information (n = 76) were divided into 3 diet groups by the relative abundance of grass, fruit and invertebrate in their diet guided by the PCoA plot of diet Bray-Curtis dissimilarity: 1. Fruit, if fruit percentage is >=20%; 2. Invertebrate, if there is invertebrate in diet; 3. Grass, if grass percentage is >=70%.

Figure S3. The first principal coordinate of variation in diet composition (diet PC1) as a function of the 11 primary diet components (Table S3). Blue lines represent lowess regression fits. PC1 explained 46% of the variation in diet composition and is associated with a tradeoff in the proportion of grass (-) versus fruit (+) in the baboons' diets.

Figure S4. The second principal coordinate of variation in diet composition (diet PC2) as a function of the 11 primary diet components (Table S3). Blue lines represent lowess regression fits. PC1 explained 23% of the variation in diet composition and is associated with a tradeoff in proportion of insects (-) versus fruit (+) in the baboons' diets.

Figure S5. The third principal coordinate of variation in diet composition (diet PC3) as a function of the 11 primary diet components (Table S3). Blue lines represent lowess
regression fits. PC1 explained 11% of the variation in diet composition and is associated with the proportion of the diet attributed to 'unknown' categories (-).
Supplementary tables

**Table S1.** Sample size information, including the number of individuals and fecal samples used in analyses of the dataset rarefied to 3,000 reads.

<table>
<thead>
<tr>
<th>Individual</th>
<th>Sex</th>
<th>Number of samples</th>
<th>Range of years samples were collected</th>
<th>Age range or age at time of sample collection (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BEAM</td>
<td>M</td>
<td>9</td>
<td>1994 - 2001</td>
<td>5.95 - 13.16</td>
</tr>
<tr>
<td>DUNLIN</td>
<td>F</td>
<td>8</td>
<td>1996 - 1997</td>
<td>0.72 - 1.56</td>
</tr>
<tr>
<td>OCEAN</td>
<td>M</td>
<td>5</td>
<td>1997 - 2000</td>
<td>0.6 - 3.78</td>
</tr>
<tr>
<td>OKOT</td>
<td>M</td>
<td>5</td>
<td>1996 - 1998</td>
<td>1.32 - 2.81</td>
</tr>
<tr>
<td>VANGA</td>
<td>M</td>
<td>5</td>
<td>1995 - 1998</td>
<td>3.05 - 6.05</td>
</tr>
<tr>
<td>LAWYER</td>
<td>M</td>
<td>3</td>
<td>2001 - 2001</td>
<td>1.19 - 1.98</td>
</tr>
<tr>
<td>HONEY</td>
<td>F</td>
<td>2</td>
<td>1999 - 2000</td>
<td>1.85 - 2.75</td>
</tr>
<tr>
<td>LEBANON</td>
<td>M</td>
<td>2</td>
<td>1998 - 2000</td>
<td>1.57 - 3.21</td>
</tr>
<tr>
<td>OXYGEN</td>
<td>F</td>
<td>2</td>
<td>2001 - 2001</td>
<td>1.62 - 2.16</td>
</tr>
<tr>
<td>VIXEN</td>
<td>F</td>
<td>2</td>
<td>1994 - 1997</td>
<td>17.06 - 19.97</td>
</tr>
<tr>
<td>DYNAMO</td>
<td>M</td>
<td>1</td>
<td>1998</td>
<td>0.94</td>
</tr>
<tr>
<td>ECHO</td>
<td>F</td>
<td>1</td>
<td>1997</td>
<td>5.88</td>
</tr>
<tr>
<td>HEKO</td>
<td>F</td>
<td>1</td>
<td>1997</td>
<td>14.27</td>
</tr>
<tr>
<td>LARK</td>
<td>F</td>
<td>1</td>
<td>1997</td>
<td>9.71</td>
</tr>
<tr>
<td>VOGUE</td>
<td>F</td>
<td>1</td>
<td>1998</td>
<td>1.08</td>
</tr>
</tbody>
</table>

**Table S2.** Unweighted UniFrac dissimilarity comparison within and between mammalian orders or diet types.

<table>
<thead>
<tr>
<th>Group</th>
<th>Average within group dissimilarity</th>
<th>Average between group dissimilarity</th>
<th>P value (Wilcoxon rank sum test)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Order</td>
<td>0.80</td>
<td>0.86</td>
<td>2E-16</td>
</tr>
<tr>
<td>Diet type</td>
<td>0.82</td>
<td>0.87</td>
<td>2E-16</td>
</tr>
<tr>
<td>Diet category</td>
<td>Diet item</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------</td>
<td>-----------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grass</td>
<td>Grass corms (all species)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Grass leaves (all species)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Grass blade bases (all species)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Grass seed head (all species)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gum</td>
<td>Gum from <em>Acacia xanthophloea</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leaves</td>
<td><em>Lyceum sp.</em> leaves</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Azima tetracantha</em> leaves</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Acacia xanthophloea</em> leaves</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Salvadora persica</em> leaves</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Suaeda monoica</em> leaves</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Tribulus terrestris</em> leaves</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fruits</td>
<td><em>Trianthema ceratosepala</em> fruits</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Azima tetracantha</em> fruits</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Abutelon</em> sp. fruits</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Lyceum</em> sp. fruits</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Ramphicarpa montana</em> fruits</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Salvadora persica</em> fruits</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Solanum dubium</em> fruits</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Tribulus terrestris</em> fruits</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blossoms</td>
<td><em>Withania</em> sp. fruits</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bark</td>
<td>Bark from <em>Acacia xanthophloea</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pods</td>
<td>Fresh, green seed pods of <em>Acacia</em> spp.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seeds</td>
<td>Dried seeds of <em>Acacia</em> spp.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Invertebrates</td>
<td>Invertebrates of unknown species</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dung</td>
<td>Liquid from or items in elephant dung</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>Unknown diet items (i.e. those that could not be seen by observers)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table S4. CCA analysis of environment and host factors for the 3,000 read dataset.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Number of samples</th>
<th>Factors tested</th>
<th>Best model at phylum level</th>
<th>Best model at genus level</th>
<th>Best model at OTU level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full dataset</td>
<td>54</td>
<td>age, rainfall, sex, individual ID, social group, natal social group, group size</td>
<td>rainfall (P=0.12), age (P=0.13)</td>
<td>rainfall (P=0.09)</td>
<td>None</td>
</tr>
<tr>
<td>Subset with diet diversity info</td>
<td>38</td>
<td>age, rainfall, sex, diet diversity (richness, Shannon’s H or PCoA axis), individual ID</td>
<td>None</td>
<td>rainfall (P=0.08), diet PC1 (P=0.05)</td>
<td>None</td>
</tr>
</tbody>
</table>
Table S5. Best-supported generalized linear mixed model (Poisson-link) explaining variation in abundance of the four most common bacteria phyla for the subset of 76 samples with diet data. Individual identity is a random effect in all models.

<table>
<thead>
<tr>
<th>Bacteria phylum</th>
<th>Fixed effect</th>
<th>estimate</th>
<th>S.E.</th>
<th>Z</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actinobacteria</td>
<td>rainfall</td>
<td>-0.005</td>
<td>0.000</td>
<td>-15.126</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td></td>
<td>diet PC1</td>
<td>0.248</td>
<td>0.041</td>
<td>6.086</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td></td>
<td>diet PC2</td>
<td>1.858</td>
<td>0.113</td>
<td>16.403</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td></td>
<td>diet PC3</td>
<td>1.981</td>
<td>0.097</td>
<td>20.365</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>Bacteroidetes</td>
<td>rainfall</td>
<td>-0.002</td>
<td>0.000</td>
<td>-5.101</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td></td>
<td>diet PC1</td>
<td>-0.849</td>
<td>0.073</td>
<td>-11.70</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td></td>
<td>diet PC2</td>
<td>-0.537</td>
<td>0.106</td>
<td>-5.048</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td></td>
<td>diet PC3</td>
<td>1.044</td>
<td>0.161</td>
<td>6.471</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>Firmicutes</td>
<td>age</td>
<td>-0.006</td>
<td>0.002</td>
<td>-2.68</td>
<td>0.007*</td>
</tr>
<tr>
<td></td>
<td>rainfall</td>
<td>0.001</td>
<td>0.000</td>
<td>13.53</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td></td>
<td>diet PC1</td>
<td>-0.578</td>
<td>0.026</td>
<td>-22.60</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td></td>
<td>diet PC2</td>
<td>-0.770</td>
<td>0.034</td>
<td>-22.57</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td></td>
<td>diet PC3</td>
<td>-0.549</td>
<td>0.056</td>
<td>-9.78</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>Proteobacteria</td>
<td>age</td>
<td>0.041</td>
<td>0.011</td>
<td>3.91</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td></td>
<td>diet PC1</td>
<td>3.270</td>
<td>0.071</td>
<td>46.00</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td></td>
<td>diet PC2</td>
<td>-0.706</td>
<td>0.089</td>
<td>-7.96</td>
<td>&lt;0.001*</td>
</tr>
</tbody>
</table>

Table S6. Mantel test of correlation between sampling time interval and microbiota weighted Unifrac dissimilarity between samples that were collected from the same individual.

<table>
<thead>
<tr>
<th>Individual</th>
<th>Number of samples</th>
<th>Mantel r statistic</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>BEAM</td>
<td>10</td>
<td>-0.04</td>
<td>0.882</td>
</tr>
<tr>
<td>DUNLIN</td>
<td>10</td>
<td>-0.03</td>
<td>0.899</td>
</tr>
<tr>
<td>OKOT</td>
<td>10</td>
<td>0.40</td>
<td>0.095</td>
</tr>
<tr>
<td>OCEAN</td>
<td>8</td>
<td>-0.27</td>
<td>0.263</td>
</tr>
<tr>
<td>VIXEN</td>
<td>8</td>
<td>0.11</td>
<td>0.696</td>
</tr>
<tr>
<td>DRONGO</td>
<td>7</td>
<td>0.23</td>
<td>0.452</td>
</tr>
<tr>
<td>ECHO</td>
<td>7</td>
<td>-0.41</td>
<td>0.180</td>
</tr>
<tr>
<td>VANGA</td>
<td>7</td>
<td>0.61</td>
<td>0.038*</td>
</tr>
<tr>
<td>GOLON</td>
<td>6</td>
<td>-0.17</td>
<td>0.692</td>
</tr>
<tr>
<td>LAWYER</td>
<td>6</td>
<td>0.25</td>
<td>0.365</td>
</tr>
<tr>
<td>OXYGEN</td>
<td>6</td>
<td>0.10</td>
<td>0.792</td>
</tr>
<tr>
<td>HONEY</td>
<td>3</td>
<td>-0.64</td>
<td>0.481</td>
</tr>
</tbody>
</table>
Fig S1

Classified

Unclassified

Prevalance

Frequency
Fig S4

- Grass
- Gum
- Leaves
- Fruit
- Blossoms
- Bark
- Pods
- Seeds
- Invertebrates
- In or under dung
- Unknown

PC2