




# Applying collocation and APRIORI analyses to chimpanzee diets: Methods for investigating nonrandom food combinations in primate self-medication

Elodie Freymann<sup>1</sup>  | João d'Oliveira Coelho<sup>1</sup>  | Catherine Hobaite<sup>2,3</sup>  |  
Michael A. Huffman<sup>4</sup>  | Geresomu Muhumuza<sup>2</sup> | Klaus Zuberbühler<sup>2,5</sup>  |  
Susana Carvalho<sup>1</sup> 

<sup>1</sup>Primate Models for Behavioural Evolution Lab, Department of Anthropology and Museum Ethnography, Institute of Human Sciences, University of Oxford, Oxford, UK

<sup>2</sup>Budongo Conservation Field Station, Masindi, Uganda

<sup>3</sup>Wild Minds Lab, School of Psychology and Neuroscience, University of St Andrews, St Andrews, UK

<sup>4</sup>Wildlife Research Center, Inuyama Campus, Kyoto University, Inuyama, Japan

<sup>5</sup>Department of Comparative Cognition, Institute of Biology, University of Neuchâtel, Neuchâtel, Switzerland

## Correspondence

Elodie Freymann, Primate Models for Behavioural Evolution Lab, Department of Anthropology and Museum Ethnography, Institute of Human Sciences, University of Oxford, 64 Banbury Rd, Oxford OX2 6PN, UK.  
Email: [elodie.freymann@anthro.ox.ac.uk](mailto:elodie.freymann@anthro.ox.ac.uk)

## Funding information

British Institute of Eastern Africa; Keble College, University of Oxford; Clarendon Fund, University of Oxford

## Abstract

Identifying novel medicinal resources in chimpanzee diets has historically presented challenges, requiring extensive behavioral data collection and health monitoring, accompanied by expensive pharmacological analyses. When putative therapeutic self-medicative behaviors are observed, these events are often considered isolated occurrences, with little attention paid to other resources ingested in combination. For chimpanzees, medicinal resource combinations could play an important role in maintaining well-being by tackling different symptoms of an illness, chemically strengthening efficacy of a treatment, or providing prophylactic compounds that prevent future ailments. We call this concept the self-medicative resource combination hypothesis. However, a dearth of methodological approaches for holistically investigating primate feeding ecology has limited our ability to identify nonrandom resource combinations and explore potential synergistic relationships between medicinal resource candidates. Here we present two analytical tools that test such a hypothesis and demonstrate these approaches on feeding data from the Sonso chimpanzee community in Budongo Forest, Uganda. Using 4 months of data, we establish that both collocation and APRIORI analyses are effective exploratory tools for identifying binary combinations, and that APRIORI is effective for multi-item rule associations. We then compare outputs from both methods, finding up to 60% agreement, and propose APRIORI as more effective for studies requiring control over confidence intervals and those investigating nonrandom associations between more than two resources. These analytical tools, which can be extrapolated across the animal kingdom, can provide a cost-effective and efficient method for targeting resources for further pharmacological investigation, potentially aiding in the discovery of novel medicines.

**Abbreviations:** LHS, left-hand side of an APRIORI equation; MDCA, multiple distinctive collocation analysis; MICA, mutual information collocation analysis; RHS, right-hand side of an APRIORI equation; ROI, resource of Interest.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2024 The Authors. *American Journal of Primatology* published by Wiley Periodicals LLC.

## KEYWORDS

diet, feeding ecology, food combinations, *Pan troglodytes*, zoopharmacognosy

## 1 | INTRODUCTION

When humans fall ill or get injured, our behaviors and diets are often impacted in a variety of ways (Prather, 2013). How and what we do to mediate illness or treat wounds largely depends on what culture we were raised in, which medical systems we participate in, availability of medicinal resources, and our own familial or cultural traditions (Kirmayer, 2004). However, irrespective of cultural or localized practice, it is unlikely that we rely on only one technique or resource to treat our symptoms (Che et al., 2013; Leonti & Casu, 2013; Ulrich-Merzenich et al., 2009; Verpoorte et al., 2009).

Across animal self-medication literature, therapeutic self-medicative anecdotes are often reported with focus placed on the ingestion of a single putative resource. Few studies investigate the broader feeding repertoires of sick individuals across the duration of an illness or interrogate whether other consumed resources may have relevant curative value (though see Krief et al., 2006; Struhsaker et al., 1997). Similarly, despite the known presence of preventative self-medication strategies across multiple primate species, involving regular ingestion of “medicinal foods” with bioactive properties (e.g., Huffman, 1997; Huffman et al., 2020; Petroni et al., 2017), the prophylactic benefits of these resources, when combined, remain largely unknown. This is, in part, due to the high burden of proof for establishing a feeding event as therapeutically self-medicative (Huffman, 1997), or a commonly consumed resource as a “medicinal food.”

To begin to bridge this gap in our knowledge, we propose investigation into a novel hypothesis: the self-medicative resource combination hypothesis, which posits that when ill, chimpanzees may therapeutically treat themselves with a combination of resources which cumulatively aid in the recovery process. This hypothesis could also apply to other self-medicative strategies, including passive prevention, through the combined ingestion of medicinal foods. However, for scope, nontherapeutic strategies are not further explored in this paper. This hypothesis is in part based on the evolutionary proximity between humans and chimpanzees, as well as the demonstrated overlap between the putative self-medicative repertoires of chimpanzees and reported ethnomedicinal practices of certain human communities (e.g., Huffman & Seifu, 1989; Petroni et al., 2017).

Medicinal resource combinations could benefit consumers in multiple ways. For example, different therapeutic resources, consumed sequentially, could mediate different parts of an illness or injury (i.e., symptom relief vs. combating the underlying infection). Consecutive medicinal resources could also have complementary operational mechanisms (i.e., mechanical vs. chemical) which attack the problem in the different ways (Huffman, 1997). Lastly, medicinal resource combinations could catalyze synergistic chemical interactions through different positive modes of combinatorial interactions.

These include *Reinforcement* (when herbs have similar properties which together produce a greater effect), *Potentiation* (when herbs have different properties where one strengthens the potency of the other), and *Restraint and Detoxification* (when herbs have different properties where when one detoxifies or nullifies the negative side effects of the primary acting herb) (Che et al., 2013). If nonrandom medicinal resource associations are identified, and these resources are intentionally combined by wild chimpanzees, these same modes of interaction may also be present. However, to demonstrate intentionality behind these combinations, further research would need to first rule out alternative hypotheses. One such alternative hypothesis is that identified nonrandom food combinations are the result of an opportunistic trial-and-error self-medicative strategy (Villalba & Provenza, 2007). In this case, modes of herb-herb interactions may be present but unintentional.

We propose through the self-medicative resource combination hypothesis that if ill chimpanzees use a combination of resources to aid in recovery, then we may be able to identify novel self-medicative resources through identification of nonrandom resource combinations ingested during periods of infection. These resources may have possible mechanical, chemical, or synergistic effects against the same underlying illness, other ailments with similar symptoms, or coinfections. While proof of medicinal resource combinations will require additional behavioral data collection and targeted pharmacological analyses, a logical first step to begin exploring this hypothesis is to identify nonrandom resource combinations which occur above chance in existing feeding ecology data sets. These analyses can be conducted without the need for further costly or invasive data collection. However, first we need methods and analyses to quantify these associations. Despite the vast potential of this research area, to the authors' knowledge, no methods have thus far been proposed to help identify nonrandom food combinations in primate diets.

### 1.1 | Holistic self-medication: A new paradigm for studying animal health maintenance through food combinations

The literature on food combinations in wild primate diets is surprisingly sparse, even in well-studied species such as chimpanzees. While wild chimpanzee diets have been evaluated at many of the long-term chimpanzee field sites (e.g., Gombe: Wrangham, 1975; Budongo: Newton-Fisher, 1999; Tweheyo et al., 2004; Villioth, 2018; Ngogo: Watts et al., 2012; Mahale: Itoh & Nakamura, 2015; Fongoli: Pruetz, 2006; Caiquene-Cadique: Bessa et al., 2015; Bossou: Hockings et al., 2009; Sugiyama & Koman, 1987; Goulougo: Morgan & Sanz, 2006; Tai: Goné Bi & Wittig, 2019; Bulindi: McLennan, 2013), food combinations—for medicinal use or otherwise—have yet to be systematically studied (though see Klein et al., 2008; Krief et al., 2005;

Pebsworth et al., 2019; Villalba et al., 2017). As more studies are conducted on the bioactivity of putative medicinal resources, however, it appears increasingly possible that the sequential ingestion of resources may enhance medicinal benefits for medicators through synergistic chemical interactions (Krief et al., 2005). Following observations of chimpanzees in Kibale, Uganda combining clay with *Trichilia rubescens* leaves, Klein et al. (2008) tested whether certain clay types can bioactivate botanical compounds in this species. When the interactions between these resources were modeled in gastric and intestinal compartments and assayed, the authors found that clay enhanced the antimalarial properties of *T. rubescens*. Pebsworth et al. (2019) further found that certain clays consumed by Budongo chimpanzees adsorb phenolic compounds, potentially detoxifying the consumer's diet. Ingesting clay may, therefore, allow for sequential ingestion of therapeutic plants with high concentrations of plant secondary metabolites in larger doses, without the associated high costs.

But are chimpanzees capable of intentional sequential or combinatorial self-medication? Janmaat et al. (2014) suggested that wild chimpanzees in Tai National Park, Ivory Coast appear to plan their breakfast time, type, and location in advance, proposing that wild chimpanzees can premeditate future dietary decisions, and in so doing, meet their nutritional needs. Trapanese et al. (2019) suggest in their review that primates likely use mental maps to track where high-quality resources are located, and plan fitness-enhancing foraging strategies. Chimpanzees have also been found to travel far distances, taking detours from normal travel routes, to access medicinal plants in their environment (Huffman, 1997). The question, however, remains whether chimpanzees can apply dietary planning to multi-resource self-meditative practices when ill or wounded. As many bioactive plants are costly to consume in high doses or if the consumer is healthy (due to the presence of toxic plant secondary metabolites), remedy-seeking individuals would need to develop an appropriate response to these costs and toxicity budgets (Villalba et al., 2017). If chimpanzees do employ medicinal combinations for self-meditative purposes this would open substantial new questions in the field of zoopharmacognosy, providing evidence for the intentionality of primate self-medication, and expediting the discovery of novel self-meditative resources.

## 2 | DESCRIPTION

We use two methods, never before employed in self-meditative contexts, to explore nonrandom food combinations in wild chimpanzee diets, as part of an early investigation into the self-meditative resource combination hypothesis. These methods include collocation analysis (Bosshard et al., 2021; Gries, 2014; Leroux et al., 2021) and APRIORI analysis (Agrawal & Srikant, 1994). Collocation analysis looks for binary pairs that occur above chance in a data set. As collocation analysis only works with binary pairs (hereafter bigrams), we also employed APRIORI analysis to identify rule associations in the data set which exceeded two food items. While these analyses

are usually used with much larger data sets, in this study, we analyzed 4 months of feeding ecology data from the Sonso chimpanzee community living in the Budongo Forest, Uganda as a case study to identify possible resource combinations that warrant further investigation. Our example demonstrates the advantages and limitations of applying these methods to feeding ecology data. While this study aims to demonstrate the potential of these analyses in a novel context, future studies would benefit from employing larger, long-term data sets to maximize sample size, reduce seasonal or methodological biases, and increase overall accuracy.

## 3 | EXAMPLE

### 3.1 | Methods

#### 3.1.1 | Study site and subjects

The Budongo Conservation Field Station, established by Vernon Reynolds in 1990, covers 793 km<sup>2</sup>, 482 km<sup>2</sup> of which is populated by continuous, semi-deciduous forest cover (Eggeling, 1947). This study was conducted with the Sonso community, a community that has been studied continuously since 1990 (Reynolds, 2005), and which had ~65 individuals at the time of data collection. The small size of the Sonso home territory (5.33 km<sup>2</sup>) (Badihi et al., 2022) enables efficient ecological surveys and focal follows and, as this community has been studied continuously for over 30-years, ages, social relationships, and demographics are well documented. The diet of the Sonso community is also well established (Tweheyo et al., 2004) as is the available flora in the Reserve (Synnott, 1985).

#### 3.1.2 | Data collection

Data were collected over a 4-month field season (62 field days). Behavioral data, including all feeding data, were collected between 07:00 and 16:30 using observational, day-long focal follows (*sensu* Altmann, 1974), and recorded using the program Animal Observer, designed for iPad, using a custom coding scheme (see Supporting Information S1: Tables 1 and 2). As chimpanzee focal follows at Budongo end each day before the group nests, multiple day follows are not always possible. All feeding events were filmed on a Sony Handycam CX250. We also included feeding data collected from two Bushnell Trophy No-Glow camera traps during the data collection period in our analysis. These cameras were located at a clay pit site and a *Cleistopholis patens* dead wood site in the Sonso home range. To monitor health states of individuals in the community, we used internal parasite load as a metric, opportunistically collecting fecal samples and microscopically analyzing them using the McMaster Method (WHO's Bench Aid Protocols, 2019) following the methods of Huffman et al. (1993, 1997). For information on the "Preparation of Feeding Data for Analysis," see Supporting Information Materials.

## 3.2 | Data analysis

### 3.2.1 | Multiple distinctive collocation analysis (MDCA) of resource combinations

Collocation analysis was originally created by linguists to analyze lexical features and grammatical structures of natural language (e.g., Bartsch, 2004; Lehecka, 2015; Stefanowitsch & Gries, 2003; Xiao & McEnery, 2006). More recently, this approach has been adopted by primatologists (Bosshard et al., 2021; Leroux et al., 2021) to empirically identify nonrandom chimpanzee call combinations. For our study, we used collocation to explore whether any resource combinations in chimpanzee diets were more common than would be expected given an assumed random baseline.

Collocation analysis is a constructional-based technique which compares the cooccurrence of specific items with one another (most frequently individual words in linguistic analyses) (Gale et al., 1991; Gries & Stefanowitsch, 2004; Kennedy, 1991; Nesselhauf, 2005). To compare cooccurrences in the data set, collocation analysis identifies and extracts exclusive dyadic combinations, and measures the relative exclusivity of their relationship within a data set. Put simply, the analysis quantifies each bigram's relative attraction. Leroux et al. (2021) provide a useful example for understanding this technique: for any data set with A–E elements, if A and B are both elements that could combine into a bigram, collocation analysis compares the frequency of the A–B bigram with the frequencies of all other possible bigrams in which A and B independently appear in the data set (i.e., A–C, A–D, A–E, B–C, B–D, B–E).

There are two types of collocation analysis, MDCA and mutual information collocation analysis (MICA) (see Bosshard et al., 2021; Church & Hanks, 1990; Leroux et al., 2021 for more information on MICA). MICA was not used in this study as the sample size is relatively large for this analysis ( $n = 1409$  for V1 and  $n = 466$  V2). MDCA tests the attraction between units using one-tailed exact binomial tests applied to each possible bigram combination (Gries, 2014). Results are generated as log-transformed values indicating both the strength and direction of an attraction between two units in the bigram. Positive values indicate an attraction (nonrandom cooccurrence in the corpus), while negative values indicate a repulsion (nonrandom absence of cooccurrence in the corpus). The absolute value of the pbin values indicate the strength of these relationships. Pbin values  $>|1.3|$  indicate a nonrandom relationship: the higher the absolute value, the stronger the relationship between the units in the bigram (see Bosshard et al., 2021). More information on “Interpreting pbin Outputs from Collocation Analysis” can be found in the Supporting Information Materials.

To explore food combinations using this method, we ran two versions of MDCA (hereby referred to as V1 and V2). V1 evaluates all eligible resource combinations from feeding data across a 4-month period (V1 = 1409 distinct bigrams with 98 possible food items). V2 evaluates combinations from this period that include one or more unusual/potentially self-medicative resource (V2 = 466 distinct bigrams with  $n = 98$  possible food items). As our focus was to identify

meaningful food cooccurrences, we only extracted combinations with pbins  $>1.3$ :  $p < 0.05$  from analyses outputs.

#### Collocation MDCA version 1 analysis

To create the data set used in V1 and V2, chimpanzee feeding data from the whole group was first ordered by date. Next, within each day-cluster, we extracted available feeding data from each individual and compiled them into separate feeding lists, with consumed resources ordered chronologically. Cases in which individuals consumed only one food item in a day were removed, as this excluded the possibility of food combinations. When an individual consumed the same resource in consecutive order, the second observation was removed to prevent double counting feeding events which may have been disrupted and subsequently resumed. We kept cases in which the same resource was eaten multiple times in a day, if one or more different resources were eaten between these events. In total, we excluded 512 events from the original 1324 feeding events. This left 812 usable feeding events from 52 different individuals. Next, we subdivided the complete data set into two distinct subsets for each version of analysis (V1 and V2). For V1, we calculated every permutation of resource pairing for each individual's daily diet, preserving the order of ingestion (see Table 1 for example).

Once permutations were produced, V1 consisted of 1409 resource pairs which were then run through collocation analysis on R (Gries, 2014). This produced a total output of 8352 distinct bigrams. Of these bigrams, 208 had nonrandom pbins  $>1.3$  ( $p < 0.05$ ).

#### Collocation MDCA version 2 analysis

While V1 included all resources in the group's diet, even those which are popularly consumed, V2 attempted to control for frequency/availability biases. The aim of V2 was to provide a more streamlined and efficient method for identifying possible synergistic resources and/or novel medicinal resources. Only combinations which included a specified, putatively self-medicative resource, hereafter referred to as a resource of interest (ROI), were included. ROI is specifically defined here as a resource consumed by Sonso chimpanzees which

**TABLE 1** Example permutations from an individual's daily diet for collocation MDCA V1.

Example of an individual's ordered daily diet <i>Ficus exasperata</i> ripe fruit, <i>F. mucoso</i> unripe fruit, <i>Ficus variifolia</i> young fruit, and <i>Cubitermes</i> mound soil	
Food 1	Food 2
<i>Ficus exasperata</i> ripe fruit	<i>Ficus mucoso</i> unripe fruit
<i>Ficus exasperata</i> ripe fruit	<i>Ficus variifolia</i> young fruit
<i>Ficus exasperata</i> ripe fruit	<i>Cubitermes</i> mound soil
<i>Ficus mucoso</i> unripe fruit	<i>Ficus variifolia</i> young fruit
<i>Ficus mucoso</i> unripe fruit	<i>Cubitermes</i> mound soil
<i>Ficus variifolia</i> young fruit	<i>Cubitermes</i> mound soil

Abbreviation: MDCA, multiple distinctive collocation analysis.

either possesses established medicinal properties, is consumed using uncommon food processing techniques, and/or was observed being ingested by an individual with a high or diverse parasite load. A comparison of uncommon and common processing techniques can be found in Supporting Information S1: Table 3. To identify foods with known medicinal properties, we searched preexisting ethnomedical and bioactivity literature before data collection (Iwu, 2013; Kokwaro, 2009; PROTA4U database). A literature review of known medicinal properties for selected ROIs can be found in the Supporting Information S1: Table 4. Based on these factors, the resources in Table 2 were selected as ROIs for this study. Water was considered an ROI as a random factor, as it is known to combine with a "leaf sponging" behavior that was also coded as a dietary resource. While leaf sponges are not ingested, bioactive or nutritional compounds from the leaves could be ingested by chimpanzees during drinking events. Different types of water (depending on the water source) were also included.

From our clean data set, we only extracted feeding data on days when an individual consumed at least one ROI for our V2 subset. We

**TABLE 2** Selected ROIs for use in collocation MDCA version 2.

Resource code	Details
Cubitermes mound soil	<i>Cubitermes</i> spp. termite soil
Afm pith	<i>Aframomum</i> spp. pith
Fe bark	<i>Ficus exasperata</i> bark
Ab bark	<i>Alstonia boonei</i> bark
Fvr bark	<i>Ficus variifolia</i> bark
Mrt pith	<i>Marantachloa leucantha</i> pith
Cos pith	<i>Costus</i> spp. pith
Cp bark	<i>Cleistopholis patens</i> bark
Acp pith	<i>Acanthus polystachyus</i> pith
Clay	Clay
Clay water	Clay water
Fsu unripe fruit	<i>Ficus sur</i> unripe fruit
Fsu young fruit	<i>Ficus sur</i> young fruit
Wf young leaf	<i>Whitefeldia elongata</i> leaf
Lp young leaf	<i>Lasiodiscus pervillei</i> leaf
Water	Drunk from puddle or stream
Water Cmi	Drunk from <i>Cordia mildebreidii</i> tree hole
Water Trr	Drunk from <i>Trichia rubescens</i> tree hole
Unk root	Unknown root
Ka resin	<i>Khaya anthotheca</i> resin
Ptm root	<i>Pterygota mildebreidii</i> root
Urc flower	<i>Scepocarpus trinervis</i> flower

Abbreviations: MDCA, multiple distinctive collocation analysis; ROI, resource of interest.

**TABLE 3** Example permutations with 1 ROI (*A. polystachyus* pith) for collocation MDCA V2.

Example of an individual's ordered daily diet <i>A. polystachyus</i> pith, <i>Ficus sur</i> ripe fruit, <i>Cordia millenii</i> ripe fruit, and <i>Saba florida</i> ripe fruit	
Food 1	Food 2
<i>Acanthus polystachyus</i> pith	<i>Ficus sur</i> ripe fruit
<i>Acanthus polystachyus</i> pith	<i>Cordia millenii</i> ripe fruit
<i>Acanthus polystachyus</i> pith	<i>Saba florida</i> ripe fruit

Abbreviations: MDCA, multiple distinctive collocation analysis; ROI, resource of interest.

**TABLE 4** Example permutations with 2 ROIs (*A. polystachyus* pith & *K. anthotheca* resin) for collocation MDCA V2.

Example of an individual's ordered daily diet <i>A. polystachyus</i> pith*, <i>Ficus sur</i> ripe fruit, <i>K. anthotheca</i> resin*, and <i>S. florida</i> ripe fruit)	
Resource 1	Resource 2
<i>Acanthus polystachyus</i> pith	<i>Ficus sur</i> ripe fruit
<i>Acanthus polystachyus</i> pith	<i>Khaya anthotheca</i> resin
<i>Acanthus polystachyus</i> pith	<i>Saba florida</i> ripe fruit
<i>Ficus sur</i> ripe fruit	<i>Khaya anthotheca</i> resin
<i>Khaya anthotheca</i> resin	<i>Saba florida</i> ripe fruit

Abbreviations: MDCA, multiple distinctive collocation analysis; ROI, resource of interest.

then generated all resource pairs that included ROIs, taking the order of ingestion into account (see Table 3 for example).

On days when multiple ROIs were consumed by an individual in a 1-day period, all combinations which included at least one ROI were generated. Bigrams which included two ROIs were only generated once, following sequential order (see Table 4 for example).

### 3.2.2 | APRIORI analysis for resource rules and associations

We also used the APRIORI algorithm to evaluate chimpanzee dietary resource combinations (Agrawal & Srikant, 1994). This method reveals association rules between "items" in a large data set, by taking item combinations and generating all association rules that have support and confidence greater than the minimum support and minimum confidence intervals specified by the algorithm's user (Agrawal & Srikant, 1994; Al-Maolegi & Arkok, 2014). The original purpose of this algorithm was to use commercial transaction histories to improve information-driven marketing processes. Based on these transaction histories, the algorithm mines association rules and suggests additional products to customers (Hahsler, 2017; Hahsler & Karpjenko, 2017). In addition to e-commerce applications, APRIORI has recently been used to unravel paleoecological associations between extinct species in the fossil record across different

geological strata (Bobe et al., 2022) and to better understand facial communication systems (Mielke et al., 2022). As far as we know, APRIORI has never been used to analyze nonhuman feeding behavior, offering a novel approach for testing associations between food resources.

#### APRIORI analysis

Feeding data from the 4-month observation period were combined for all individuals and formatted identically to the collocation analysis V1 subset, making these two analyses efficient to run in parallel. Using the transactions() function from the arules package (Hornik et al., 2005), the long-form data set was transformed into a Binary Incidence Matrix, a format typically used for mining associations in transaction data. The data set was then run through APRIORI on R (version 4.0.5, R Development Core Team, 2019), and the results outputted to our accessible interactive platform: PANacea <https://osteomics.com/PANacea/>. On this platform, the scientific community can interact with and interpret results of the APRIORI algorithm applied to our data set. PANacea is an online data exploration web-app, built on top of the visualization techniques for association rules on R using the shiny (Chang et al., 2021), arules (Hahsler et al., 2011; Hornik et al., 2005), and arulesViz (Hahsler, 2017) packages. PANacea is available in the Osteomics platform.

Interpreting the results of this analysis requires an understanding of the customizable metrics: support, confidence, and lift (see Supporting Information S1: Figure 1). Support represents the number of times the association is present in the data and serves as a popularity metric. In data sets with a high diversity of item-types such as this one (i.e., resource types), support tends to be low for most associations, as the number of times each combination occurred will likely be small. Confidence (scaled between 0 and 1) can be interpreted as percentages (0 = 0% and 1 = 100%). While confidence can give a sense of association strength, it can also be affected by data set size. For example, if a combination between A and B is the only combination occurring for A, the confidence for that pair will be very high (1). The Lift metric, therefore, is crucial to consider when interpreting small data sets as it is a way of controlling for confidence. If A and B are uncommon, yet tend to be sampled in pairs, the lift will be higher. A large lift means that the confidence value is larger than the expected value and therefore this association is likely not due to chance. Lift should be >1 for confidence to be considered a usable metric. Lift can be used to indirectly control for factors like short duration of data collection. It is also useful for larger data sets that have many observations, but low frequency of occurrence for each item or combination. More detail on "Interpreting APRIORI Data Outputs on PANacea" can be found in the Supporting Information Materials.

## 4 | RESULTS

### 4.1 | MDCA version 1 results

Of the possible 1409 permutations of resource combinations, collocation MDCA V1 resulted in 8352 distinct bigrams, 208 of which had pbin values >1.3. To interpret these data, MDCA outputs

were ordered by pbin value. For brevity, we present below the 25 bigrams with the highest pbin values (Table 5; range = 3.01–17.85). The pair with the highest attraction (17.85) was *F. exasperata* (Fe) unripe fruit & *F. mucoso* (Fm) ripe fruit, both popular feeding items which were fruiting simultaneously during the study period. Another result which suggests the efficacy of this analysis is the high attraction of leaf sponge and clay water (pbin = 4.62), the former of which is a tool for accessing the latter.

### 4.2 | MDCA version 2 results

In this model, created with 22 ROIs, 466 possible resource combinations were produced from 812 total eligible feeding events. Of these, collocation analyses resulted in 4480 bigrams, 85 of which

**TABLE 5** MDCA (V1) results: 25 most significant resource bigrams from complete data set (pbins > 1.3).

Resource 1	Resource 2	pbin value
Fe unripe fruit	Fm ripe fruit	17.85
Cgp ripe fruit	Fvr young fruit	5.53
Fsu ripe fruit	Cgp ripe fruit	5.33
Fsu ripe fruit	Fvr young fruit	4.86
Unk leaf sponge	Clay water	4.62
Mie ripe fruit	Fvr young leaf	4.35
Cph young leaf	Fe young leaf	4.12
Fe young fruit	Fm ripe fruit	4.11
Cph young leaf	Avo ripe fruit	4.03
Ptm root	Dd young leaf	3.97
Cph young leaf	Cgp ripe fruit	3.83
Fm ripe fruit	Gpr ripe fruit	3.73
Myh ripe fruit	Mrt pith	3.62
Es seeds	Dd ripe fruit	3.59
Me ripe fruit	Meat colobus	3.50
Avo ripe fruit	Fe young leaf	3.32
Urc flowers	Mb mature leaf	3.30
Cli young leaf	Cze young leaf	3.27
Cgp ripe fruit	Avo ripe fruit	3.15
Blu leaf sponge	Water Fe	3.15
Water Fe	Fvr leaf bud	3.15
Fvr ripe fruit	Afm ripe fruit	3.07
Cli young leaf	Cli young leaf	3.05
Fth ripe fruit	Fsu ripe fruit	3.03
Bpy mature leaf	Bpy ripe fruit	3.01

Abbreviation: MDCA, multiple distinctive collocation analysis.

**TABLE 6** MDCA (V2) results: 25 top significant resource bigrams including at least 1 ROI (pbins > 1.3).

Resource 1	Resource 2	pbin value
Myh ripe fruit	Mrt pith	4.66
Cph young leaf	<i>Cubitermes</i> mound soil	4.16
Unk leaf sponge	Clay water	4.13
Ptm root	Dd young leaf	3.96
<i>Cubitermes</i> mound soil	Cgp ripe fruit	3.34
Water	Bpy young leaf	3.22
Bpy young leaf	Water Trr	3.18
Clay water	Avo ripe fruit	3.12
Es seeds	Water	3.12
<i>Cubitermes</i> mound soil	Fsu ripe fruit	2.99
Cli mature leaf	Cos pith	2.667
Cli young leaf	Cp bark	2.62
Fm ripe fruit	Ka resin	2.58
Avo ripe fruit	Fsu unripe fruit	2.54
Water Trr	Cgp ripe fruit	2.52
Dd ripe fruit	<i>Cubitermes</i> mound soil	2.50
Sf ripe fruit	Ka resin	2.45
Bpy ripe fruit	Ka resin	2.41
Psg unripe fruit	Ka resin	2.41
Ka resin	Fvr unripe fruit	2.39
Mie ripe fruit	<i>Cubitermes</i> mound soil	2.39
Water Cmi	Bpy young leaf	2.37
<i>Cubitermes</i> mound soil	Cph young leaf	2.35
Es seeds	Water Cmi	2.25
Afm pith	Fm ripe fruit	2.24

Abbreviations: MDCA, multiple distinctive collocation analysis; ROI, resource of interest.

had pbin values >1.3 (Table 6). The high attraction of “leaf sponge” and “clay water” (pbin = 4.13) once again indicates the test's efficacy, however, it is lower here than the pbin for this bigram in V1 (pbin = 4.62). This was expected as V2 used a smaller data set, and thus both resources are present at higher relative rates across V2's other bigrams.

### 4.3 | APRIORI results

On PANacea, we assigned the minimum thresholds of support = 0.01, confidence = 0.6, and lift = 1, rule length = 2–5, and extracted only the top 25 rules (out of 96 total generated rules). These 25 rules are presented in Table 7 and displayed as rule networks in Figure 1, including columns for support, confidence, lift, and count (number of

**TABLE 7** Example of “data exploration” results, showing top 25 results with APRIORI (ordered by lift).

Rules	Support	Confidence	Lift	Count
{Avo ripe fruit, Cgp ripe fruit, Fsu ripe fruit} ⇒ {Fe young leaf}	0.012	0.75	31.38	3
{Fe ripe fruit, Ka resin} ⇒ {Sf ripe fruit}	0.012	0.75	26.89	3
{Es seeds, Fvr ripe fruit} ⇒ {Dd ripe fruit}	0.012	0.75	20.92	3
{Es seeds, Fsu ripe fruit, Fvr ripe fruit} ⇒ {Dd ripe fruit}	0.012	0.75	20.92	3
{Unk leaf sponge} ⇒ {clay water}	0.012	0.75	17.11	3
{Avo ripe fruit, <i>Cubitermes</i> mound soil} ⇒ {Cph young leaf}	0.012	1	16.73	3
{Avo ripe fruit, Cgp ripe fruit, <i>Cubitermes</i> mound soil} ⇒ {Cph young leaf}	0.012	1	16.73	3
{Avo ripe fruit, <i>Cubitermes</i> mound soil, Fsu ripe fruit} ⇒ {Cph young leaf}	0.012	1	16.73	3
{Avo ripe fruit, Cgp ripe fruit, <i>Cubitermes</i> mound soil, Fsu ripe fruit} ⇒ {Cph young leaf}	0.012	1	16.73	3
{Acp pith} ⇒ {Myh ripe fruit}	0.012	1	14.76	3
{Fe ripe fruit, Ka resin} ⇒ {Afm pith}	0.012	0.75	14.48	3
{Cgp ripe fruit, Fe young leaf} ⇒ {Avo ripe fruit}	0.016	1	13.94	4
{Cph young leaf, Fe young leaf} ⇒ {Avo ripe fruit}	0.012	1	13.94	3
{Afm ripe fruit, Fsu ripe fruit} ⇒ {Fvr ripe fruit}	0.012	1	13.94	3
{Dd ripe fruit, Es seeds, Fsu ripe fruit} ⇒ {Fvr ripe fruit}	0.012	1	13.94	3
{Cgp ripe fruit, Cph young leaf, Fe young leaf} ⇒ {Avo ripe fruit}	0.012	1	13.94	3
{Cgp ripe fruit, Fe young leaf, Fsu ripe fruit} ⇒ {Avo ripe fruit}	0.012	1	13.94	3
{Avo ripe fruit, Fe young leaf} ⇒ {Cph young leaf}	0.012	0.75	12.55	3
{Cgp ripe fruit, Fe young leaf} ⇒ {Cph young leaf}	0.012	0.75	12.55	3
{Fm ripe fruit, Fvr young leaf} ⇒ {Bpy young leaf}	0.012	0.75	12.55	3

(Continues)

TABLE 7 (Continued)

Rules	Support	Confidence	Lift	Count
{Avo ripe fruit, Cgp ripe fruit, Fe young leaf} ⇒ {Cph young leaf}	0.012	0.75	12.55	3
{Avo ripe fruit, Cgp ripe fruit, Fsu ripe fruit} ⇒ {Cph young leaf}	0.012	0.75	12.55	3
{Cgp ripe fruit, <i>Cubitermes</i> mound soil} ⇒ {Cph young leaf}	0.020	0.625	10.46	5
{Cgp ripe fruit, <i>Cubitermes</i> mound soil, Fsu ripe fruit} ⇒ {Cph young leaf}	0.020	0.625	10.46	5
{Dd ripe fruit, Es seeds} ⇒ {Fvr ripe fruit}	0.012	0.75	10.46	3

times this combination occurred in the data set). The rule with the highest lift (31.38) was unidentified sp. (Avo) fruit, *Celtis gomphophylla* (Cgp) ripe fruit, *Ficus sur* (Fsu) ripe fruit ⇒ *F. exasperata* (Fe) young leaf, which had a 75% confidence outcome. This means that throughout the study period, when Avo ripe fruit (a large fruit from an unidentified tree species resembling the genus *Gambeya*) was eaten in combination with Cgp ripe fruit and Fsu ripe fruit, it was exceedingly likely that Fe young leaves were also eaten. This is a high confidence interval for an association with the relatively long rule length of 4, as the chances of random combinations decrease as rule length increases (see section on “Interpreting APRIORI Data Outputs on PANacea” in Supporting Information Materials). Three of these species (Cgp fruits, Fsu fruits, and Fe young leaves) are commonly eaten resources amongst Sonso chimpanzees. The unidentified spp. coded “Avo fruit,” however, was not commonly eaten.

The APRIORI algorithm can also be tailored to generate rules which include a specified ROI, using the “Find rules by food-item” under the data exploration tab. To further explore this, we selected rules which included *A. polystachyus* pith with support = 0.011, confidence = 0.6, lift = 1, and rule length between 2 and 6. The rule with highest lift (rule 1) for this set of criteria is displayed in Figure 2. This rule found that if *A. polystachyus* pith was consumed, then *M. holstii* fruit was also consumed, and that this association had a 100% confidence with the notably high lift of = 14.8.

#### 4.4 | Comparing collocation and APRIORI analyses

We demonstrate above how both collocation and APRIORI can be used to analyze and evaluate food combinations in wild chimpanzee diets. To determine whether these methods produced similar outcomes, we ran the APRIORI algorithm again, this time with a rule length of 2, to produce bigrams that could be compared to bigram outputs from collocation analysis V1. We ran APRIORI with support = 0.011, confidence = 0.6, and lift = 1, rule length = 2, and

found it produced only 17 bigrams (Table 6), although it should be noted that this minimum confidence threshold creates a bias toward rare events. Bigrams from both analyses were separately ranked and ordered and then compared to assess overall agreement between methods.

When the 17 APRIORI pairs were compared to the top 17 collocation V1 bigrams, there were four exact matches between the two algorithms' outputs, showing only 23% agreement (Supporting Information S1: Table 5). However, as these two models have different mechanisms for ranking combinations, a better assessment for agreement was how many exact matches there were between the 17 APRIORI pairs and all nonrandom collocation V1 outputs with positive attraction.

$$\text{Agreement} = \frac{(\text{total \# of exact matches across methods})}{(\text{total \# of APRIORI pairs})}$$

When we compared all exact matches outputted by both methods, agreement was notably higher (53%) (Supporting Information S1: Table 6). The clustered rule associations of these APRIORI pairs are visualized in Figure 3, which displays rule associations for the 17 bigram outputs which met our metric criteria.

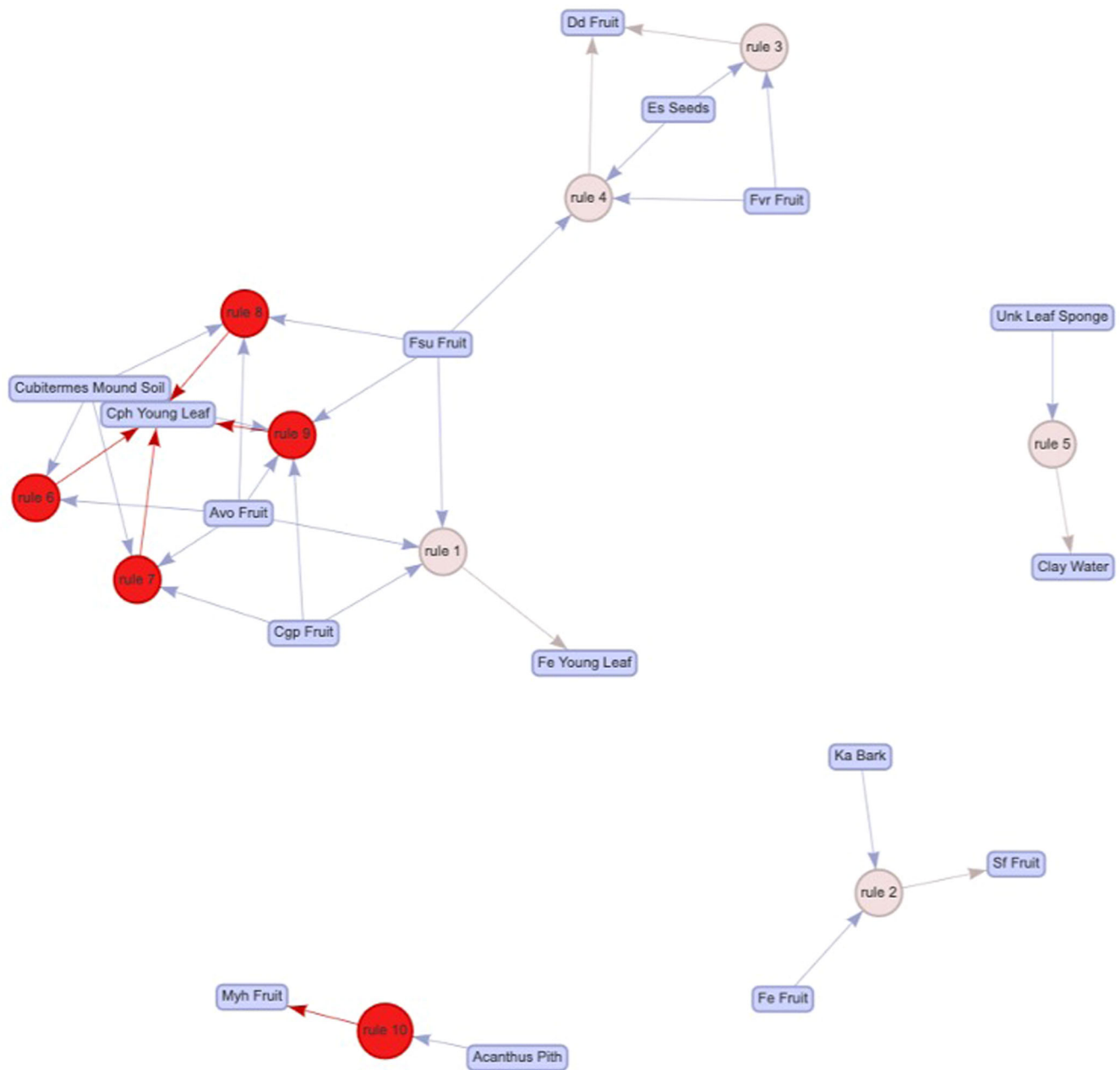
We also wanted to determine the role of confidence level on percentage of matched pairs between the two algorithms, as this metric is changeable in APRIORI but not in collocation analyses. The confidence interval was lowered from 0.6 to 0.011 on PANacea, and results were reinterpreted. Using this new minimum confidence level, PANacea produced 249 pairs, the top 25 of which were compared to the top 25 collocation analysis results, showing 28% agreement (Supporting Information S1: Table 7). To better assess agreement regardless of rank, we again found the total number of exact matches between the top 25 APRIORI pairs and all nonrandom collocation results with positive attraction ( $p_{bin} > 1.3$ ) (Supporting Information S1: Table 8). This yielded a 60% agreement between models, higher than agreement found in the 0.6 confidence interval model. As the 25 top APRIORI pairs selected represent a small sample of all produced APRIORI pairs ( $n = 249$ ), agreement here is likely an underestimation.

## 5 | DISCUSSION

### 5.1 | MDCA version 1

Collocation MDCA V1 explored all food combinations in the Sonso diet during a 4-month observation period, resulting in nonrandom resource pairs which may provide relevant interpretations for future investigation. For example, V1 produced a single bigram which contained the same resource twice: Cli young leaf & Cli young leaf ( $p_{bin} = 3.05$ ). As duplicate feeding events were excluded when the same resource was consumed consecutively, this means that on days when individuals ingested climber leaves, they nonrandomly sought this resource out again later the same day. Results of this kind may offer valuable future insight into primate dietary decisions and spatial feeding patterns.



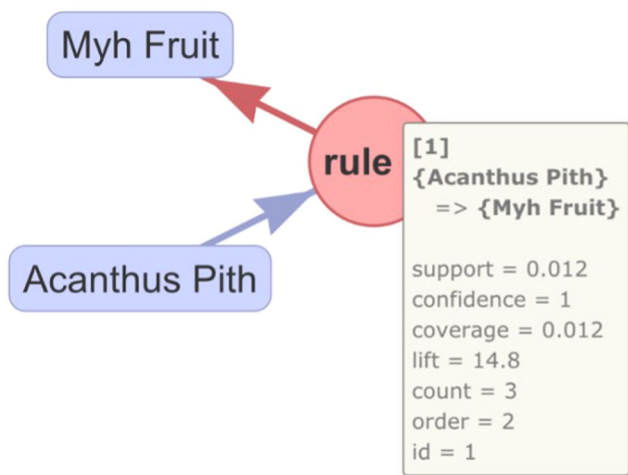


**FIGURE 1** Example of a medicinal network result on PANacea showing top 10 APRIORI results (ordered by lift) with support = 0.011, confidence = 0.6, and lift = 1, rule length = 2–5. NB: Rule circles gradated to red based on confidence. Blue arrows pointing toward rule circles indicate resources on left-hand side (LHS) of each equation. Red arrows pointing away from rule circles indicate resources on right-hand side (RHS). When “Fruit” is not modified, it can be assumed to be ripe.

Another notable bigram is *U. trinervis* (Urc) flowers & *M. bpterygocaulos* (Mb) mature leaf (pbin = 3.30). Throughout the study period, only two chimpanzees (maternal brothers MB and MZ) were observed eating Urc flowers, with feeding events occurring on two consecutive days. MZ was the only individual observed eating the leaves of Mb, with these events also occurring on the same 2 days. On the second day of ingestion, the brothers left the group to travel a far distance before consuming the resources in this bigram. MZ's fecal sample showed a high-intensity and diverse parasite load. MZ also consumed several other putative medicinal species that day.

While this remains a single anecdote at present, during this event, MZ demonstrated all but one of Huffman's self-medicative criteria (Huffman, Gotoh, et al., 1997).

Collocation MDCA V1 can provide a general overview of which resources are eaten in nonrandom combination in a given feeding data set from a study group. When researchers have hypotheses concerning the function or benefits of a particular resource, outputted bigrams can be searched for the targeted resource. This allows for easy investigation into which other resources may be combined with the targeted resource at frequencies higher than expected by chance. For self-medication studies,



**FIGURE 2** Top results of “Rule by food-item” search targeting *A. polystachyus* (ordered by lift) with support = 0.011, confidence = 0.6, lift = 1, and rules length = 2–6.

this could allow for preliminary identification of putative self-medicative resources which could then be more thoroughly evaluated. However, to produce more conclusive results about dietary choice concerning food resources in the general diet, this analysis should be rerun on long-term data sets to account for seasonality bias. Additional surveying should also be conducted to control for ecological availability and spatial distribution of resources.

## 5.2 | MDCA version 2

Collocation MDCA V2 targeted putatively self-medicative resources and aimed to identify other resources eaten in combination. Results of this analysis showed that *Cubitermes* mound soil was ingested in nonrandom combination with several other resources. Relevant bigrams include: *C. philppensis* (Cph) young leaf & *Cubitermes* mound soil (pbin = 4.16); *Cubitermes* mound soil & *C. gomphophylla* (Cgp) ripe fruit (pbin = 3.34); *Cubitermes* soil & *F. sur* (Fsu) ripe fruit (pbin = 2.99); *D. dewevrii* (Dd) ripe fruit & *Cubitermes* mound soil (pbin = 2.50); *Mildbraediendron excelsum* (Mie) ripe fruit & *Cubitermes* mound soil (pbin = 2.39); *Cubitermes* mound soil & *C. philppensis* (Cph) young leaf (pbin = 2.35). Food order may be relevant for *Cubitermes* mound soil, as the pbin of Cph young leaf & *Cubitermes* mound soil (pbin = 4.16) ( $n = 7$  occurrences) differs from the pbin of *Cubitermes* mound soil & Cph young leaf (pbin = 2.35) ( $n = 5$  occurrences). The number of nonrandom bigrams which include *Cubitermes* mound soil in V2, suggests that on days when individuals ate this resource, their diets were more predictable than on days when they ate other ROIs. In other words, ingestion of termite soil amongst Sonso chimpanzees is unlikely to be purely opportunistic. There are several possible interpretations of this pattern, one being that chimpanzees may target termite soil during periods of gastrointestinal distress (as was observed by Mahaney et al., 1996 in Mahale), potentially caused by

ingestion of other resources in these bigrams. Future research should further explore explanations for these observed patterns.

As interpretation of collocational strength is relatively subjective in V2, it is crucial that results are contextualized with behavioral observations. For example, we observed individuals consuming *A. polystachyus* three times during the 4-month study period. Each of these times, the consumer had a high parasite load. In one case, an individual consumed *A. polystachyus* pith immediately before leaf swallowing, one of the two confirmed self-medicative behaviors (Huffman & Caton, 2001; Huffman et al., 1996). In another case, an individual consuming *A. polystachyus* was observed wadging the pith of this plant with the ROI *C. patens* (Cp) dead wood, following the ingestion of another ROI (*Cubitermes* mound soil). Immediately after, the individual consumed a sequence of additional ROIs (e.g., *K. anthotheca resin* and *M. bpterygocaulos leaves*) and the bark of *S. myrtina*.

*A. polystachyus* has already been suggested as a candidate self-medicative resource at Kibale (Krief et al., 2005) and is known to be used in traditional medicine to treat skin infections, dermatosis, and sterility in Burundi (Krief et al., 2005; Pebsworth et al., 2006). A leaf decoction of this species is used for treating liver and spleen problems and stems are pounded with their leaves to treat depressive psychosis (Kokwaro, 2009). Collocation MDCA V2 analysis can be employed to identify resources for further in-depth investigation of medicinal value, when some behavioral or ethnomedicinal evidence has already been attained.

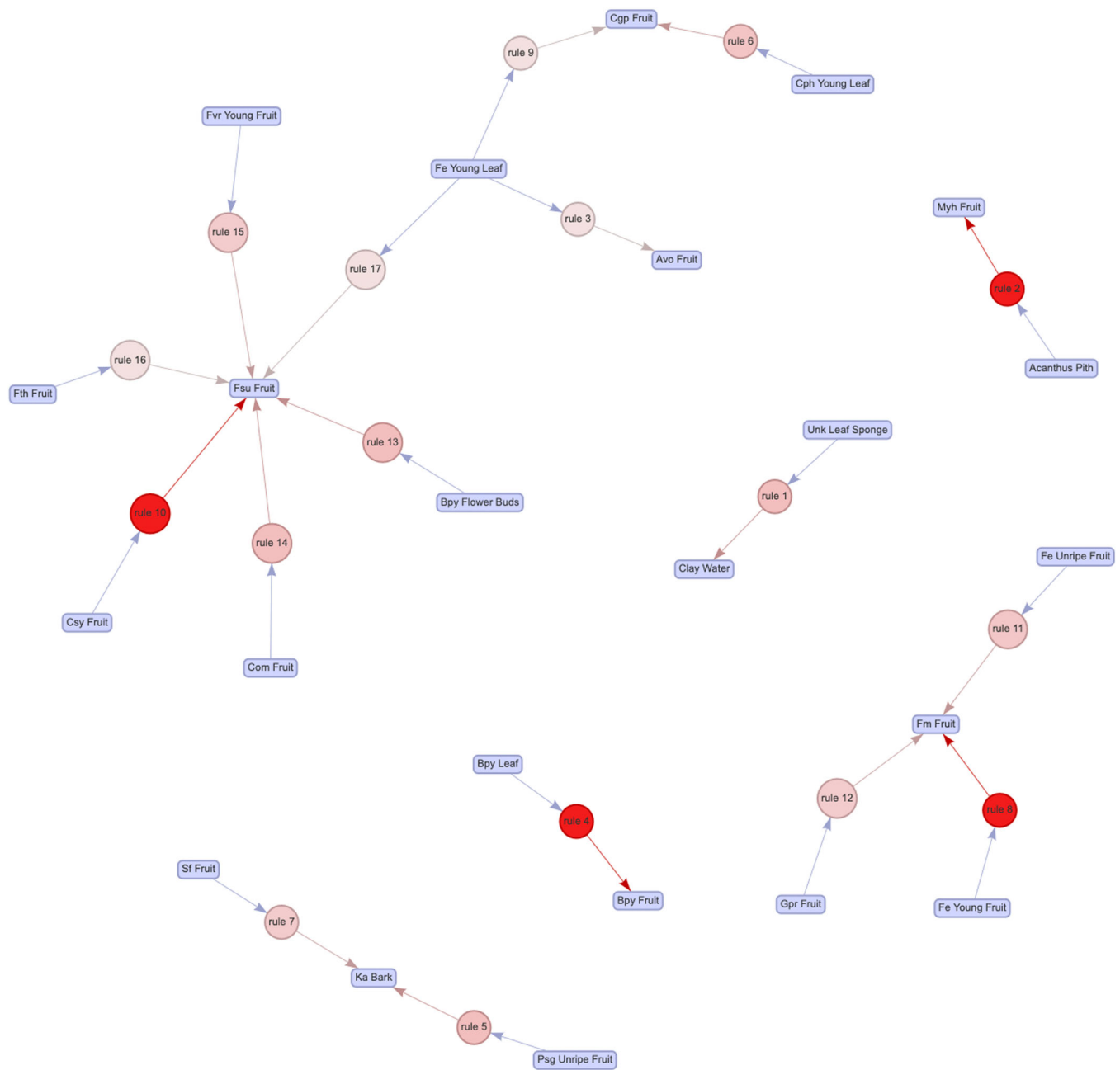
## 5.3 | APRIORI

APRIORI analysis offers the opportunity to explore more complex associations between more than two resources at a time, and to take data density and representativeness into account by adjusting for confidence and other factors. A result of interest was the high lift (16.73) and 100% confidence interval of the Avo ripe fruit, *Cubitermes* mound soil  $\Rightarrow$  Cph young leaf rule association. This rule maintained the same lift and confidence when two additional resources were added to the left-hand side of this equation (i.e., Avo ripe fruit, Cgp ripe fruit, *Cubitermes* mound soil  $\Rightarrow$  Cph young leaf; Avo ripe fruit, *Cubitermes* mound soil, Fsu ripe fruit  $\Rightarrow$  Cph young leaf). This finding supports results from Collocation MDCA V2, that *Cubitermes* mound soil combines with several other species in a nonrandom way.

Of all bigrams produced by APRIORI, the pair with the highest lift (14.76) was *A. polystachyus* (Acp)  $\Rightarrow$  *M. holstii* (Myh) (confidence = 1). This result, when interpreted in conjunction with collocation MDCA V2's results, establishes a strong case for further investigation into a synergistic relationship between *A. polystachyus* and *M. holstii*.

## 6 | COMPARISON AND CRITIQUE

While our data set was effective for a piloting these methods, collocation and APRIORI are more typically applied to larger data sets (such as linguistic corpora or e-commerce transaction histories), and



**FIGURE 3** APRIORI medicinal network visualization of 17 bigrams (ordered by lift) with support = 0.011, confidence = 0.6, and lift = 1, rule length = 2.

our study was short and seasonally biased. The specific examples of resource combinations provided here would now benefit from follow-up analyses using a multi-year data set for the Sonso community. Long-term behavioral data, however, would need to be supplemented with ecological and spatial data to determine proximity of resources and fruiting synchrony. Analyses which incorporate the whole diet of each individual (e.g., collocation MDCA V1), would benefit from additional data on feeding locations for each resource, to control for spatial and temporal factors. While collocation MDCA V2 has some measure of control for ecological noise and popularity biases toward certain species, the species and quantity of ROIs used

will substantially shape the findings and their stability. Accordingly, ROI's should be established from longer-term behavioral studies.

Collocation MDCA V1 and APRIORI also require subjective judgment when interpreting results, especially when differentiating between nonrandom resource associations caused by medicinal combinations and those caused by frequency/availability of the feeding items, seasonal synchronism, or geographic proximity. For example, during data collection, we observed the group feeding in a ripe *F. mucoso* (Fm) fruit tree each day for almost a full month. On most of these days, the group traveled from this tree to feed on a nearby *F. exasperata* (Fe) which was bearing unripe fruits. Likely due

to the proximity and synchronism of these resources, an Fe & Fm bigram is unsurprisingly present in the top 25 collocation results, ranking eighth. While this bigram is not amongst the top 25 APRIORI results when rule length was set between 2 and 5 (with confidence = 0.6), the combination is present in the top APRIORI results when rule length was adjusted to 2 (confidence = 0.6). Under these conditions, the Fe & Fm bigram was also ranked eighth in APRIORI (Supporting Information S1: Table 6). There is, therefore, a need for ecological and spatial controls in both analyses to distinguish ecological “noise” from potential medicinal combinations.

The similar data set format required for running both collocation and APRIORI makes this comparative approach easy and efficient. However, differences in result outputs must be considered. For example, we found that the level of agreement between methods increases when minimum confidence is lowered in APRIORI (thus increasing the number of generated APRIORI outputs). When running APRIORI with rule length metrics adjusted to produce bigrams, our results differed from bigrams produced during collocation analysis V1. This variation is not surprising as the algorithms and analyses being run are different. Despite variation, the algorithms showed up to 60% agreement across generated bigrams, suggesting that both models can and should be considered and compared when making robust interpretations. Comparing outputs from both methods may highlight some nonrandom bigrams which warrant further investigation.

Despite the efficacy of both approaches, APRIORI is better suited for studies which wish to customize specific metrics.

While these methods are currently insufficient for conclusively addressing the self-medication resource combination hypothesis, they offer an important exploratory first step which may prompt future research. Following preliminary investigation, the next stage would involve incorporation of health monitoring data into these analyses by separating “healthy” individuals from “unhealthy” individuals, and then statistically comparing the difference between groups. In sites where self-meditative resources have previously been established and systematically recorded, feeding data can also be extracted from days in which individuals engaged in a proven self-meditative behavior (e.g., leaf swallowing) and compared to days in which these behaviors were not observed, and individuals were healthy. Future studies should also investigate the role of food combinations in the normal diets of healthy chimpanzees, and the role bioactive medicinal food combinations may play in passive prevention strategies. While these quantitative methods may help establish meaningful medicinal resource combinations, the incorporation of behavioral, spatial, and health data will provide essential additional context before the implementation of expensive pharmacological analyses.

## 7 | CONCLUSION

We argue that MDCA and the APRIORI algorithm can be used to effectively detect potential resource combinations and association rules in wild primate diets. These methods allow for a preliminary

investigation into the self-meditative resource combination hypothesis and have the potential to change how we analyze and interpret long-term feeding data across field sites. If nonrandom medicinal combinations are identified in primate diets, this could have important implications for the discovery of novel primate self-meditative behaviors. As a cost-effective strategy for selecting natural resources for targeted bioactivity testing, it may 1 day also lead to the discovery of synergistic compounds effective in treating human pathogens. Overall, employing interdisciplinary methods, such as these, to systematically study nonhuman feeding ecologies, seems likely to soon yield fruitful outcomes. Future studies in chimpanzee self-medication and primate feeding ecology more generally should strive to consider feeding behaviors from a more holistic perspective, remaining open to the notion that nonhuman medicinal diets may be more planned than previously considered.

## AUTHOR CONTRIBUTIONS

**Eodie Freymann:** Conceptualization (lead); data curation (lead); formal analysis (lead); funding acquisition (lead); investigation (lead); methodology (lead); project administration (lead); resources (lead); software (lead); visualization (lead); writing—original draft (lead); writing—review and editing (lead). **João d'Oliveira Coelho:** Formal analysis (supporting); methodology (supporting); software (equal); writing—review and editing (supporting). **Geresomu Muhumuza:** Investigation (equal); project administration (supporting). **Catherine Hobaiter:** Resources (supporting); supervision (equal); writing—review and editing (equal). **Michael A. Huffman:** Supervision (equal); writing—review and editing (equal). **Klaus Zuberbühler:** Resources (supporting); writing—review and editing (supporting). **Susana Carvalho:** Supervision (lead); writing—review and editing (equal).

## ACKNOWLEDGMENTS

We are grateful to the BCFS field staff working with the Sonso and Waibira communities who provide invaluable instruction and guidance in the field including Chandia Bosco, Monday Mbotella Gideon, Adué Sam, Asua Jackson, Eguma Robert Yikii, Steven Mugisha, Atayo Gideon, Kizza Vincent. Daniel Sempebwa carried out all parasitological analysis and conducted systematic health monitoring critical for this study. BCFS management, including Walter Akankwasa and David Eryenyu, as well as the other researchers working at the station provided help and support. We also would like to extend our gratitude to Vernon Reynolds, founder of BCFS, and the Royal Zoological Society of Scotland which provides core funding. We are grateful to Uganda Wildlife Authority and the Uganda National Council for Science and Technology for granting permission to work in Uganda. Lastly, we would like to sincerely thank Maël Leroux for his input on the use of collocation analysis and for his valuable feedback on this paper, as well as Joon Kim, Gabrielle Melli, and Andrew Foerder for their help with data cleaning.

## CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.


## DATA AVAILABILITY STATEMENT

Data from this study are available from the corresponding author upon reasonable request. Results from APRIORI are accessible on the interactive PANacea web-app at <https://osteomics.com/PANacea/>.

## ETHICS STATEMENT

Data used in this study were collected by EF with the approval by the Uganda Wildlife Authority (Uganda Wildlife Authority permit no. COD/96/05), the Uganda National Council for Science and Technology (permit no. NS257ES). The study was purely observational and adhered to the Code of Best Practices in Field Primatology (Riley et al., 2014). All applicable international and national guidelines were followed.

## ORCID

Elodie Freymann  <https://orcid.org/0000-0002-5484-8643>  
 João d'Oliveira Coelho  <https://orcid.org/0000-0003-0871-1926>  
 Catherine Hobaiteir  <https://orcid.org/0000-0002-3893-0524>  
 Michael A. Huffman  <https://orcid.org/0000-0003-2115-7923>  
 Klaus Zuberbühler  <https://orcid.org/0000-0001-8378-088X>  
 Susana Carvalho  <https://orcid.org/0000-0003-4542-3720>

## REFERENCES

- Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules in large databases, *Proceedings of the 20th International Conference on Very Large Data Bases* (pp. 487–499). Morgan Kaufmann Publishers Inc.
- Al-Maolegi, M., & Arkok, B. (2014). An improved Apriori algorithm for association rules. ArXiv Preprint ArXiv:1403.3948.
- Altmann, J. (1974). Observational study of behavior: Sampling methods. *Behaviour*, 49(3/4), 227–266.
- Badihi, G., Bodden, K., Zuberbühler, K., Samuni, L., & Hobaiteir, C. (2022). Flexibility in the social structure of male chimpanzees (*Pan troglodytes schweinfurthii*) in the Budongo Forest, Uganda. *Royal Society Open Science*, 9(9), 220904. <https://doi.org/10.1098/rsos.220904>
- Bartsch, S. (2004). *Structural and functional properties of collocations in English: A corpus study of lexical and pragmatic constraints on lexical co-occurrence*. Gunter Narr Verlag.
- Bessa, J., Sousa, C., & Hockings, K. J. (2015). Feeding ecology of chimpanzees (*Pan troglodytes verus*) inhabiting a forest-mangrove-savanna-agricultural matrix at Caiquene-Cadique, Cantanhez National Park, Guinea-Bissau. *American Journal of Primatology*, 77(6), 651–665. <https://doi.org/10.1002/ajp.22388>
- Bobe, R., d'Oliveira Coelho, J., Carvalho, S., & Leakey, M. (2022). Early hominins and paleoecology of the Koobi Fora Formation, Lake Turkana Basin, Kenya. In S. C. Reynolds & R. Bobe (Eds.), *African paleoecology and human evolution*. Cambridge University Press.
- Bosshard, A. B., Leroux, M., Lester, N. A., Bickel, B., Stoll, S., & Townsend, S. W. (2021). From collocations to call-ocations: Using linguistic methods to quantify animal call combinations. *BioRxiv*. <https://doi.org/10.1101/2021.06.16.448679>
- Chang, W., Cheng, J., Allaire, J. J., Sievert, C., & Schloerke, B. (2021). *Shiny: Web application framework for R. R package version 1.7.1*. <https://CRAN.R-project.org/package=shiny>
- Che, C.-T., Wang, Z., Chow, M., & Lam, C. (2013). Herb-herb combination for therapeutic enhancement and advancement: Theory, practice and future perspectives. *Molecules*, 18(5), 5125–5141. <https://doi.org/10.3390/molecules18055125>
- Church, K., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics*, 16(1), 22–29.
- Eggeling, W. J. (1947). Observations on the ecology of the Budongo rain forest, Uganda. *The Journal of Ecology*, 34, 20–87.
- Gale, W., Church, K., Hanks, P., & Hindle, D. (1991). Using statistics in lexical analysis. In *Lexical Acquisition: Exploiting on-Line Resources to Build a Lexicon* (115–164).
- Goné Bi, Z. B., & Wittig, R. (2019). Long-term diet of the chimpanzees (*Pan troglodytes verus*) in Tai National Park: Interannual variations in consumption. In C. Boesch, R. Wittig, C. Crockford, L. Vigilant, T. Deschner, & F. Leendertz, (Eds.), *The chimpanzees of Tai forest 40 years of research* (pp. 242–260). Cambridge University Press.
- Gries, S. T. (2014). *Coll.analysis 3.5. A script for R to compute perform collostructional analyses*.
- Gries, S. T., & Stefanowitsch, A. (2004). Extending collocation analysis: A corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics*, 9(1), 97–129.
- Hahsler, M. (2017). arulesViz: Interactive visualization of association rules with R. *The R Journal*, 9(2), 163.
- Hahsler, M., Chelluboina, S., Hornik, K., & Buchta, C. (2011). The arules R-package ecosystem: Analyzing interesting patterns from large transaction data sets. *The Journal of Machine Learning Research*, 12, 2021–2025.
- Hahsler, M., & Karpienko, R. (2017). Visualizing association rules in hierarchical groups. *Journal of Business Economics*, 87(3), 317–335.
- Hockings, K. J., Anderson, J. R., & Matsuzawa, T. (2009). Use of wild and cultivated foods by chimpanzees at Bossou, Republic of Guinea: Feeding dynamics in a human-influenced environment. *American Journal of Primatology*, 71, 636–646.
- Hornik, K., Grün, B., & Hahsler, M. (2005). arules-A computational environment for mining association rules and frequent item sets. *Journal of Statistical Software*, 14(15), 1–25.
- Huffman, M. A. (1997). Current evidence for self-medication in primates: A multidisciplinary perspective. *American Journal of Physical Anthropology*, 104(S25), 171–200.
- Huffman, M. A., & Caton, J. M. (2001). Self-induced increase of gut motility and the control of parasitic infections in wild chimpanzees. *International Journal of Primatology*, 22(3), 329–346.
- Huffman, M. A., Gotoh, S., Izutsu, D., Koshimizu, K., & Kalunde, M. S. (1993). Further observations on the use of the medicinal plant, *Vernonia amygdalina* (Del) by a wild chimpanzee, its possible effect on parasite load, and its phytochemistry. *African Study Monographs*, 14(4), 227–240.
- Huffman, M. A., Gotoh, S., Turner, L. A., Hamai, M., & Yoshida, K. (1997). Seasonal trends in intestinal nematode infection and medicinal plant use among chimpanzees in the Mahale Mountains Tanzania. *Primates*, 38(2), 111–125.
- Huffman, M. A., Page, J. E., Sukhdeo, M. V. K., Gotoh, S., Kalunde, M. S., Chandrasiri, T., & Towers, G. H. N. (1996). Leaf-swallowing by chimpanzees: A behavioral adaptation for the control of strongyle nematode infections. *International Journal of Primatology*, 17, 475–503.
- Huffman, M. A., & Seifu, M. (1989). Observations on the illness and consumption of a possibly medicinal plant *Vernonia amygdalina* (Del.), by a wild chimpanzee in the Mahale Mountains National Park, Tanzania. *Primates*, 30, 51–63. <https://doi.org/10.1007/BF02381210>
- Huffman, M. A., Sun, B.-H., & Li, J.-H. (2020). Medicinal properties in the diet of Tibetan Macaques at Mt. Huangshan: A case for self-medication, 223–248. [https://doi.org/10.1007/978-3-030-27920-2\\_12](https://doi.org/10.1007/978-3-030-27920-2_12)
- Itoh, N., & Nakamura, M. (2015). Diet and feeding behavior. In M. Nakamura, K. Hosaka, N. Itoh, & K. Zamma (Eds.), *Mahale*

- chimpanzees: 50 years of research (pp. 227–245). Cambridge University Press.
- Iwu, M. M. (2013). *Handbook of African medicinal plants* (2nd ed.). CRC Press. <https://doi.org/10.1201/b16292>
- Janmaat, K. R., Leo, P., Dagui, B. S., & Christophe, B. (2014). Wild chimpanzees plan their breakfast time, type, and location. *Proceedings of the National Academy of Sciences*, 111(46), 16343–16348. <https://doi.org/10.1073/pnas.1407524111>
- Kennedy, G. (1991). Between and through: The company they keep and the functions they serve. *English Corpus Linguistics*, 95–110.
- Kirmayer, L. J. (2004). The cultural diversity of healing: Meaning, metaphor and mechanism. *British Medical Bulletin*, 69(1), 33–48. <https://doi.org/10.1093/bmb/ldh006>
- Klein, N., Fröhlich, F., & Krief, S. (2008). Geophagy: Soil consumption enhances the bioactivities of plants eaten by chimpanzees. *Naturwissenschaften*, 95(4), 325–331. <https://doi.org/10.1007/s00114-007-0333-0>
- Kokwaro, J. O. (2009). *Medicinal plants of East Africa*. <https://books.google.co.uk/books?id=msyHLY0dhPwC>
- Krief, S., Hladik, C. M., & Haxaire, C. (2005). Ethnomedicinal and bioactive properties of plants ingested by wild chimpanzees in Uganda. *Journal of Ethnopharmacology*, 101, 1–15. <https://doi.org/10.1016/j.jep.2005.03.024>
- Krief, S., Huffman, M. A., Sévenet, T., Hladik, C. M., Grellier, P., Loiseau, P. M., & Wrangham, R. W. (2006). Bioactive properties of plant species ingested by chimpanzees (*Pan troglodytes* schweinfurthii) in the Kibale National Park, Uganda. *American Journal of Primatology*, 68(1), 51–71. <https://doi.org/10.1002/ajp.20206>
- Lehecka, T. (2015). Collocation and colligation, *Handbook of pragmatics online*. Benjamins.
- Leonti, M., & Casu, L. (2013). Traditional medicines and globalization: Current and future perspectives in ethnopharmacology. *Frontiers in Pharmacology*, 4, 92. <https://doi.org/10.3389/fphar.2013.00092>
- Leroux, M., Bosshard, A. B., Chandia, B., Manser, A., Zuberbühler, K., & Townsend, S. W. (2021). Chimpanzees combine pant hoots with food calls into larger structures. *Animal Behaviour*, 179, 41–50. <https://doi.org/10.1016/j.anbehav.2021.06.026>
- Mahaney, W. C., Hancock, R. G. V., Aufreiter, S., & Huffman, M. A. (1996). Geochemistry and clay mineralogy of termite mound soil and the role of geophagy in chimpanzees of the Mahale Mountains, Tanzania. *Primates*, 37, 121–134.
- McLennan, M. R. (2013). Diet and feeding ecology of chimpanzees (*Pan troglodytes*) in Bulindi, Uganda: Foraging strategies at the Forest-Farm Interface. *International Journal of Primatology*, 34, 585–614. <https://doi.org/10.1007/s10764-013-9683-y>
- Mielke, A., Waller, B. M., Pérez, C., Rincon, A. V., Duboscq, J., & Micheletta, J. (2022). NetFACS: Using network science to understand facial communications systems. *Behavior Research Methods*, 54(4), 1912–1927. <https://doi.org/10.3758/s13428-021-01692-5>
- Morgan, D., & Sanz, C. M. (2006). Chimpanzee feeding ecology and comparisons with sympatric gorillas in the Goualougo Triangle, Republik of Congo. In G. Hohmann, M. M. Robbins, & C. Boesch (Eds.), *Feeding ecology in apes and other primates* (pp. 97–122). Cambridge University Press.
- Nesselhauf, N. (2005). *Collocations in a learner corpus* (Vol. 14). John Benjamins Publishing.
- Newton-Fisher, N. E. (1999). The diet of chimpanzees in the Budongo Forest Reserve, Uganda. *African Journal of Ecology*, 37, 344–354. <https://doi.org/10.1046/j.1365-2028.1999.00186.x>
- Pebsworth, P., Krief, S., & Huffman, M. A. (2006). The role of diet in self-medication among chimpanzees in the Sonso and Kanyawara communities, Uganda. *Primates of western Uganda* (pp. 105–133). Springer.
- Pebsworth, P. A., Hillier, S., Wendler, R., Glahn, R., Ta, C. A. K., Arnason, J. T., & Young, S. L. (2019). Geophagy among East African Chimpanzees: Consumed soils provide protection from plant secondary compounds and bioavailable iron. *Environmental Geochemistry and Health*, 41(6), 2911–2927. <https://doi.org/10.1007/s10653-019-00366-8>
- Petroni, L. M., Huffman, M. A., & Rodrigues, E. (2017). Medicinal plants in the diet of woolly spider monkeys (*Brachyteles arachnoides*, E. Geoffroy, 1806)—A bio-rational for the search of new medicines for human use? *Revista Brasileira de Farmacognosia*, 27(2), 135–142. <https://doi.org/10.1016/j.bjp.2016.09.002>
- Prather, A. A. (2013). Sickness behavior. In M. D. Gellman & J. R. Turner (eds), *Encyclopedia of behavioral medicine*. Springer.
- PROTA4U. Retrieved December 1, 2022, from <https://prota.prota4u.org/>
- Pruetz, J. D. (2006). Feeding ecology of savanna chimpanzees (*Pan troglodytes* verus) at Fongoli, Senegal. *Feeding Ecology in Apes and Other Primates*, 326–364.
- R Development Core Team. (2019). *R: A language and environment for statistical Computing*. R Foundation for Statistical Computing.
- Reynolds, V. (2005). *The chimpanzees of the Budongo Forest*. Oxford University Press.
- Riley, E. P., MacKinnon, K. C., Fernandez-Duque, E., Setchell, J. M., & Garber, P. A. (2014). Code of best practices for field primatology. 1–18. <https://doi.org/10.13140/2.1.2889.1847>
- Stefanowitsch, A., & Gries, S. T. (2003). Collocations: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics*, 8(2), 209–243.
- Struhsaker, T. T., Cooney, D. O., & Siex, K. S. (1997). Charcoal consumption by zanzibar red colobus monkeys: Its function and its ecological and demographic consequences. *International Journal of Primatology*, 18, 61–72.
- Sugiyama, Y., & Koman, J. A. (1987). A preliminary list of chimpanzees' alimentation at Bossou, Guinea.
- Synnott, T. (1985). A checklist of the flora of Budongo Forest Reserve, Uganda, with notes on ecology and phenology. Commonwealth Forestry Institute, University of Oxford.
- Trapanese, C., Meunier, H., & Masi, S. (2019). What, where and when: Spatial foraging decisions in primates. *Biological Reviews*, 94(2), 483–502. <https://doi.org/10.1111/bvr.12462>
- Tweheyo, M., Lye, K. A., & Weladji, R. B. (2004). Chimpanzee diet and habitat selection in the Budongo Forest Reserve, Uganda. *Forest Ecology and Management*, 188(1–3), 267–278. <https://doi.org/10.1016/j.foreco.2003.07.028>
- Ulrich-Merzenich, G., Panek, D., Zeitler, H., Wagner, H., & Vetter, H. (2009). New perspectives for synergy research with the “omic”-technologies. *Phytomedicine*, 16, 495–508. <https://doi.org/10.1016/j.phymed.2009.04.001>
- Verpoorte, R., Crommelin, D., Danhof, M., Gilissen, L. J. W. J., Schuitmaker, H., Van der Greef, J., & Witkamp, R. F. (2009). Commentary: “A systems view on the future of medicine: Inspiration from Chinese medicine?”. *Journal of Ethnopharmacology*, 121(3), 479–481.
- Villalba, J. J., Costes-Thiré, M., & Ginane, C. (2017). Phytochemicals in animal health: Diet selection and trade-offs between costs and benefits. *Proceedings of the Nutrition Society*, 76(2), 113–121. <https://doi.org/10.1017/S0029665116000719>
- Villalba, J. J., & Provenza, F. D. (2007). Self-medication and homeostatic behaviour in herbivores: Learning about the benefits of nature's pharmacy. *Animal*, 19, 1360–1370.
- Villioth, J. (2018). The foraging ecology of two neighbouring chimpanzee communities from Budongo Forest. Doctor of Philosophy (PhD) thesis, University of Kent/University of Neuchâtel. University of Kent/University of Neuchâtel.
- Watts, D. P., Potts, K. B., Lwanga, J. S., & Mitani, J. C. (2012). Diet of chimpanzees (*Pan troglodytes* schweinfurthii) at Ngogo, Kibale National Park, Uganda. 1. Diet composition and diversity. *American Journal of Primatology*, 74(2), 114–129.
- WHO's Bench Aid Protocols. (2019). *Bench aids for the diagnosis of intestinal parasites* (2nd ed.). World Health Organization.

- Wrangham, R. W. (1975). Behavioural ecology of chimpanzees in Gombe National Park, Tanzania. University of Cambridge.
- Xiao, R., & McEnery, T. (2006). Collocation, semantic prosody, and near synonymy: A cross-linguistic perspective. *Applied Linguistics*, 27(1), 103–129.

#### SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Freymann, E., d'Oliveira Coelho, J., Muhumuza, G., Hobaiter, C., Huffman, M. A., Zuberbühler, K., & Carvalho, S. (2024). Applying collocation and APRIORI analyses to chimpanzee diets: Methods for investigating nonrandom food combinations in primate self-medication. *American Journal of Primatology*, e23603.  
<https://doi.org/10.1002/ajp.23603>