

## 02-Groq.com API for test set

In this notebook, we will make API calls to Groq.com to summarize the article (up to 20k characters) in the test set.

We would use Mixtral 8x7B as it has very good performance scores, and it's open source.

```
In [1]: # !pip install groq
```

```
In [2]: # Load api key
import json
import urllib.parse

with open('./data/credentials.json') as f:
    login = json.load(f)

api_key = login["GROQ_API_KEY"]
print(len(api_key))
```

56

```
In [3]: from groq import Groq

client = Groq(
    api_key=api_key,
)
```

```
In [4]: # test pause
import time
print("Hello")
time.sleep(2)
print("World")
```

Hello  
World

```
In [5]: SLEEP_TIME = 10 # pause between requests
```

```
In [6]: def send_sumarize_request(content, model=client, min_words=250, max_words=500, max_
    """
        summarize the content
        input: context (text), model (groq_api client), max_words (int)
        output: summarized text
    """
    if not quiet:
        print("Sending request for text =", content[:100])

    result = ""
    # get 1 example from train set for the model to follow the style
    lay_summary_example = "Electrocorticography , or ECoG , is a technique that is
    # prompt = f'Simplify and summarize in minimum {min_words} to maximum {max_word
    prompt = f'One good example of lay summary is '{lay_summary_example}'. Follow t
```

```

try:
    completion = client.chat.completions.create(
        model="mixtral-8x7b-32768",
        messages=[
            {
                "role": "user",
                "content": prompt
            },
        ],
        temperature=0.8,
        max_tokens=2048,
        top_p=1,
        stream=True,
        stop=None,
    )

    for chunk in completion:
        result += chunk.choices[0].delta.content or ""
except Exception as err:
    print("Skipping, error : ", err)
    result = ""

# pause to avoid hitting bandwidth limit (~ 14K token / minute)
print(f"OK. Pausing for {SLEEP_TIME} secs...", end="")
time.sleep(SLEEP_TIME)
print("OK")

return result

```

In [7]: `import pandas as pd`

In [8]: `# Load in some from eLife test set`

```

dev_df_filename = "../data/biolaysumm2024_data/eLife_test.jsonl"
df = pd.read_json(dev_df_filename,
                  orient="records",
                  lines=True
                  )

df.head()

```

Out[8]:

	article	headings	keywords	id
0	Acylation of diverse carbohydrates occurs acro...	[Abstract, Introduction, Results and discussio...	[biochemistry and chemical biology, computatio...	elife-81547-v1
1	Honey bee ecology demands they make both rapid...	[Abstract, Introduction, Results, Discussion, ...	[computational and systems biology]	elife-86176-v2
2	Biguanides , including the world's most prescr...	[Abstract, Introduction, Results, Discussion, ...	[genetics and genomics]	elife-82210-v1
3	Ecological relationships between bacteria medi...	[Abstract, Introduction, Results, Discussion, ...	[microbiology and infectious disease, ecology]	elife-83152-v2
4	Gamma oscillations are believed to underlie co...	[Abstract, Introduction, Results, Discussion, ...	[neuroscience]	elife-83044-v2

In [9]:

```
# take some item to inspect
k = 3
item = df.iloc[k]
print("Full text article =", item["article"])
```

Full text article = Ecological relationships between bacteria mediate the services that gut microbiomes provide to their hosts . Knowing the overall direction and strength of these relationships is essential to learn how ecology scales up to affect microbiome assembly , dynamics , and host health . However , whether bacterial relationships are generalizable across hosts or personalized to individual hosts is debated . Here , we apply a robust , multinomial logistic-normal modeling framework to extensive time series data ( 5534 samples from 56 baboon hosts over 13 years ) to infer thousands of correlations in bacterial abundance in individual baboons and test the degree to which bacterial abundance correlations are ‘universal’ . We also compare these patterns to two human data sets . We find that , most bacterial correlations are weak , negative , and universal across hosts , such that shared correlation patterns dominate over host-specific correlations by almost twofold . Further , taxon pairs that had inconsistent correlation signs ( either positive or negative ) in different hosts always had weak correlations within hosts . From the host perspective , host pairs with the most similar bacterial correlation patterns also had similar microbiome taxonomic compositions and tended to be genetic relatives . Compared to humans , universality in baboons was similar to that in human infants , and stronger than one data set from human adults . Bacterial families that showed universal correlations in human infants were often universal in baboons . Together , our work contributes new tools for analyzing the universality of bacterial associations across hosts , with implications for microbiome personalization , community assembly , and stability , and for designing microbiome interventions to improve host health .

Mammalian gut microbiomes are highly diverse , dynamic communities whose members exhibit the full spectrum of ecological relationships , from strong mutualisms like syntrophy and cross-feeding , to competition , parasitism , and predation ( Faust and Raes , 2012; Foster and Bell , 2012; Dolinšek et al . , 2016; Seth and Taga , 2014 ) . These relationships mediate a variety of biological processes that have powerful effects on host health and fitness , including the metabolism of complex carbohydrates and toxins , and the synthesis of physiologically important compounds , like short-chain fatty acids , neurotransmitters , and vitamins ( Faust and Raes , 2012; Foster and Bell , 2012; Dolinšek et al . , 2016; Seth and Taga , 2014; Bäckhed et al . , 2005; Gould et al . , 2018; Pontrelli et al . , 2022; Degnan et al . , 2014 ) . Despite their importance , major gaps remain in our understanding of microbial relationships in the gut ( Faust and Raes , 2012; Loftus et al . , 2021; Bashan et al . , 2016 ) . We typically do not know if the abundance of one microbe consistently predicts the abundance of other microbes in the same host community , nor do we understand whether these correlative relationships are consistent in strength or direction across hosts ( Bashan et al . , 2016; Widder et al . , 2016; Cao et al . , 2017; Faust and Raes , 2016 ) . Knowing the overall direction and strength of these correlative relationships is important to understanding the ecological relationships that mediate gut microbial processes and shape gut microbiome assembly , stability , and productivity ( Coyte et al . , 2015; Palmer and Foster , 2022; Hu et al . , 2022 ) . For instance , sets of microbes that exhibit strong , positive relationships within hosts may represent networks of cooperating taxa that promote each other’s growth ( Bäckhed et al . , 2005; Loftus et al . , 2021; Wu et al . , 2021 ) . In turn , these strong , mutualistic interdependencies can create an ecological house of cards where microbes rise and fall together , hampering community assembly and stability ( Coyte et al . , 2015; Coyte et al . , 2021 ) . Further , understanding the degree to which correlative relationships between microbes are the same or different in different hosts can shed light on whether hosts share similar , underlying microbial ecologies ( Loftus et al . , 2021; Bashan et al . , 2016; Gao et al . , 2020; Vila et al . , 2020; San-Juan-Vergara et al . , 2018 ) . Microbial ecologies that are similar across hosts may make it possible to manipulate the microbiome’s emergent properties to improve host health ( Loftus et al . , 2021; Bashan et al . , 2016; Cao et al . , 2017; Coyte et al . , 2015; Coyte et al . , 2021; Gonze et al . , 2018 ) . To date , there are several reasons to think that correlative relationships in the gut microbiome will not

be consistent across hosts and will instead be individualized to each host . For instance , several common community and evolutionary processes—such as horizontal gene transfer and priority effects—can lead microbiome taxa to fill different ecological roles in different hosts ( Dolinšek et al . , 2016; Franzosa et al . , 2015; Faith et al . , 2013; Bik et al . , 2016; Caporaso et al . , 2011; Costello et al . , 2009 ) . Further , genotype by environment interactions and plasticity could lead some microbes to adopt context-dependent metabolisms and ecological roles depending on their microbial neighbors or other aspects of the environment ( Louca et al . , 2018; Rainey and Quistad , 2020; Martiny et al . , 2015; Debray et al . , 2022 ) . Finally , the common observation that gut microbial community compositions ( i . e . the presence and abundance of taxa ) are highly individualized is sometimes attributed to host-specific microbial ecologies and relationships ( Franzosa et al . , 2015; Faith et al . , 2013; Bik et al . , 2016; Caporaso et al . , 2011; Costello et al . , 2009; Risely et al . , 2021; Kolodny et al . , 2019; Flores et al . , 2014; Johnson et al . , 2019; Pruss et al . , 2021 ) . However , to date , the handful of studies that have tested the generalizability of gut microbial relationships across hosts suggest that microbiome community ecology is not highly individualized and is instead largely consistent ( i . e . ‘universal’ ) across hosts ( Figure 1A; Bashan et al . , 2016; Gao et al . , 2020; Vila et al . , 2020; San-Juan-Vergara et al . , 2018; Kalyuzhny et al . , 2017 ) . For instance , Bashan et al . , 2016 , inferred ‘universal’ gut microbial relationships in the human gut microbiome by applying dissimilarity overlap analysis ( DOA ) to cross-sectional samples from several human data sets . DOA infers universal microbial relationships by testing whether pairs of hosts who share many of the same microbiome taxa also have similar abundances of those taxa ( Bashan et al . , 2016; Gao et al . , 2020; Vila et al . , 2020; San-Juan-Vergara et al . , 2018; Kalyuzhny et al . , 2017 ) . This approach relies on the assumption that , when two communities share many of the same species and have similar abundances of those species , they do so because of a shared , underlying set of between-species bacterial abundance relationships ( Bashan et al . , 2016; Kalyuzhny et al . , 2017 ) . While many studies using this approach find evidence that microbial relationships are ‘universal’ ( Bashan et al . , 2016; Gao et al . , 2020; Vila et al . , 2020; San-Juan-Vergara et al . , 2018 ) , DOA’s assumptions have been questioned because environmental gradients , stochastic processes , and the presence of many non-interactive species can lead to the spurious detection of universality ( Bashan et al . , 2016; Kalyuzhny et al . , 2017; Marsland et al . , 2020 ) . An obvious alternative is to measure microbial correlations directly from microbiome time series from several hosts ( Loftus et al . , 2021; Faust et al . , 2015 ) . Unlike DOA , this approach should be able to pinpoint which microbiome taxa exhibit the most and least consistent relationships across hosts . However , measuring microbial correlations from longitudinal , multi-host microbiome time series has its own challenges: time series with adequately dense sampling are rare , and most such data sets exhibit temporal autocorrelation and irregular sampling ( Faust et al . , 2015 ) . Further , the most common , and still most feasible , way to collect microbiome community data—via high-throughput sequencing—generates noisy count data that usually can only be interpreted in terms of relative ( not absolute ) abundances ( Gloor et al . , 2017; Quinn et al . , 2017 ) . Finally , correlation cannot be used to infer causality , and in the absence of experiments , we cannot differentiate whether microbial correlation patterns arise from ecological interactions ( e . g . competition , predation , facilitation ) or shared responses to the environment . To address several of these challenges , here we combine extensive time series data on the stool-associated microbiota with a multinomial logistic-normal modeling framework ( Figure 1; n=5534 samples from 56 baboons; 75–181 samples per baboon across 6–13 . 3 years , between 2000 and 2013; Alberts and Altmann , 2012; Björk et al . , 2022; Grieneisen et al . , 2021 ) . This framework uses 16S rRNA sequencing count data to learn a smoothly evolving Gaussian process . The baboons were the subject of long-term research on individually recognized animals by the Amboseli Baboon Research Project in Kenya , which has been studying baboon ec

ology and behavior in the Amboseli ecosystem since 1971 ( Alberts and Altmann , 2012 ) . The baboons range over the same habitat and experience similar diets and sources of microbial colonization , facilitating inference about the consistency of microbial correlations across hosts ( Figure 1–figure supplement 1; Björk et al . , 2022; Grieneisen et al . , 2021 ) . To partly account for environmental drivers of microbial dynamics , our modeling approach controls for variation attributable to seasonal changes in the animals’ diets , proportionality in the count data , and irregularity in sampling to produce per-individual , per-taxon trajectories of log-ratio abundances that we used to estimate pairwise microbial correlations within each host . We pursued five main objectives . First , we characterized the overall sign and strength of pairwise correlations in bacterial abundance within each host . Second , we tested the degree to which these correlation patterns are systematically consistent across hosts or individualized by host ( Figure 1A ) . Third , we identified phylogenetic and host-related predictors of the direction and universality of bacterial correlations , including phylogenetic relationships between microbes , host age , and host genetic relatedness . Fourth , we tested whether the microbial correlations we observed were driven by shared responses to host diets or seasonality , or by synchronized microbial dynamics across hosts . Fifth , we tested the generalizability of our findings by comparing the patterns of universality in our data set to two microbiome time series from humans ( Johnson et al . , 2019; Vatanen et al . , 2016 ) . Our predictions for these analyses were influenced by ideas from community and microbial ecology . First , because strong interdependencies can hamper community assembly and destabilize community dynamics ( Coyte et al . , 2015; Palmer and Foster , 2022; Hu et al . , 2022; Coyte et al . , 2021 ) , we expected that most microbial correlations would be weak with few strong positive relationships . Second , consistent with studies that used DOA , we expected that microbial relationships would be more consistent across hosts than individualized ( see Figure 1A for a visualization of this prediction ) . This result would suggest that personalized microbiota—their compositions and dynamics—do not arise from host-specific microbiome ecologies ( Bashan et al . , 2016; Gao et al . , 2020; Vila et al . , 2020; San-Juan-Vergara et al . , 2018 ) . Third , we expected to observe positive correlations between taxa that are close phylogenetic relatives . This is because related bacteria may have similar functional properties and hence similar ecological relationships with other members of the community . They may also have dynamics that are driven by similar selective forces imposed by the host or host’s environment . Alternatively , competitive exclusion may lead closely related taxa to exhibit neutral or negative relationships . Fourth , because the environments experienced by baboons in Amboseli are far more uniform than those experienced by typical human study subjects ( Björk et al . , 2022; Grieneisen et al . , 2021 ) , we expected that the signature of ‘universality’ in baboons would be stronger than that observed in humans . We discuss the implications of these patterns for individual microbiome community assembly and dynamics , and for understanding how microbiome communities are structured across hosts—a key requirement for successful intervention to improve host health ( Bashan et al . , 2016; Widder et al . , 2016; Cullen et al . , 2020 ) .

We began by characterizing the overall sign , strength , and significance of pairwise correlations in bacterial abundance within each host . To do so , we applied the approach outlined in Figure 1A to stool-associated time series from 56 baboons ( Figure 1B ) and calculated Pearson’s correlations between pairs of bacterial taxa . To avoid biases created by zero inflation ( see Methods ) , we restricted our analysis to pairs where each member was present in at least 50% of samples across all hosts and at least 20% of samples within each host ( Supplementary file 1a-1c ) . Further , we required the joint zero abundance of a given bacterial pair to be less than 5% of all samples across hosts . After filtering , the resulting data set included ( 1 ) 1878 pairs of centered log-ratio ( CLR ) -transformed amplicon sequence variants ( ASVs; Figure 2A ) ; ( 2 ) 57 pairs of bacterial phyla ( Figure 2–figure supplement 1 ) ; and ( 3 ) 473 pairs of taxa agglomerated to the most granular possible family , or

der , or class ( Figure 2–figure supplement 1 ) . To generate an expectation of the strength of bacterial correlations possible by chance , we used a permutation procedure that randomly shuffled the taxonomic identities within each sample of the bacterial count table 10 times for each of the 56 hosts ( 560 total permutations ) . We then estimated correlations for these permuted pairs to generate an empirical null distribution of randomly generated taxon-taxon correlations . Observed correlations were judged against this reference ( Figure 2B ) . We also confirmed that the resulting correlation patterns were robust to several modeling choices and were not primarily driven by seasonal shifts in microbiome composition ( see results below ) . Consistent with the expectation that most bacterial correlations in the gut microbiome are weak ( Coyte et al . , 2015; Coyte et al . , 2021 ) , only 19% of ASV-ASV correlations in the heat map in Figure 2A were stronger than expected by chance (  $FDR \leq 0.05$ ; Figure 2–figure supplement 3; 19% of phylum-phylum; 22% of family/order/class correlations; Figure 2–figure supplement 1; Figure 2–figure supplement 2 ) . The strongest negatively correlated pair in Figure 2A included an ASV in the family Kiritimatiellaceae and another in family Lachnospiraceae which had a median correlation of  $-0.520$  (  $\pm 0.132$  s.d. ) across all baboon hosts ( Figure 2C; ASV19 and ASV23; Supplementary file 1a and d ) . The strongest positively correlated pair of ASVs included two members of the genus *Prevotella* that had a median correlation of  $0.801$  (  $\pm 0.053$  s.d. ) across all baboons ( Figure 2D; ASV2 and ASV3; Supplementary file 1a and d ) . While these two ASVs were assigned to the same genus , their V4 16S DNA sequence identity was 97.6% , indicating they are probably not simply duplicate 16S gene copies encoded in the genome of a single species ( Vetrovský and Baldrian , 2013 Supplementary file 1d ) . In support of the idea that strong , positive bacterial interdependencies are rare ( Coyte et al . , 2015; Palmer and Foster , 2022; Coyte et al . , 2021 ) , only 3.8% of ASV pairs were significantly positively correlated , and the overall bacterial correlation patterns were slightly skewed toward negative relationships . For instance , at the ASV level , the median correlation coefficient in Figure 2A was  $-0.072$  , and 60% of these correlations were negative ( binomial test  $p < 0.0001$  ) . For family/order/class-level taxa , 58% of all correlations were negative ( Figure 2–figure supplements 1A and 3A; median family/order/class-level correlation =  $-0.049$ ; binomial test  $p < 0.0001$  ) . Correlations between phyla exhibited the strongest negative skew , with 64% of phyla-phyla correlations having a negative sign ( Figure 2–figure supplements 1B and 3A; median phyla-level correlation =  $-0.100$ ; binomial test  $p < 0.0001$  ) . This bias toward negative relationships is consistent with the expectation that neutral or negative relationships between ASVs are more common than mutualisms ( Coyte et al . , 2015; Palmer and Foster , 2022; Coyte et al . , 2021 ) and that more distantly related taxa ( e.g . phyla ) respond to distinct environmental drivers due to differences in metabolic requirements and lifestyles . Next , we tested the degree to which within-host ASV-ASV correlations were consistent across hosts . We began by plotting the absolute value of each ASV pair’s median Pearson’s correlation coefficient as a function of the consistency of their correlation sign ( positive or negative ) across the 56 hosts ( Figure 3A and B ) . These plots provide two main insights into the consistency of bacterial associations . First , in support of the idea that ASVs do not exhibit vastly different correlative relationships in different hosts , no taxon pairs were strongly and inconsistently correlated across hosts ( Figure 3A and B; Figure 3–figure supplement 1A ) . Instead , the ASV pairs that had inconsistent correlation signs across hosts always had weak and often non-significant median absolute correlation coefficients within hosts ( Figure 3A and B ) . Second , the pairs with the most consistent sign agreement across hosts also exhibited the largest median absolute correlation coefficients across hosts ( Figure 3A and B; Spearman’s  $r = 0.844$  ,  $p < 0.0001$  ) . Hence , pairs of ASVs that have the strongest relationships , and are therefore likely to play the most important roles in structuring gut microbiome dynamics , also tend to have the most consistent relationships in different hosts . Indeed , for the sets of positively or negatively correlated ASV pairs that showed universal agreement in the sign of their correlation

n across all hosts ( i . e . where  $x=1$  in Figure 3A and B ) , the median absolute correlation coefficient is 0 . 359 , compared to 0 . 116 for those with no sign consistency (  $x=0 . 5$  in Figure 3A and B ) . Note , that the correlation for a given pair of ASVs was only weakly predicted by bacterial abundance (  $r=0 . 129$  and  $r=0 . 223$  for the more and less abundant partner in a pair respectively;  $p<0 . 0001$  both ) . While this effect was statistically significant , it explained only 6% of the variance in median correlation . Visual inspection of the patterns in Figures 2A , 3A and B indicate that ASV-ASV correlations are largely consistent across baboons , as opposed to individualized to each baboon . To explicitly quantify the relative strength of shared versus individualized signatures in the heat map in Figure 2A , we calculated the population mean pattern for the ASV-ASV correlation matrix ,  $m$  . For each host , we then estimated the residual difference ,  $e$  , between that individual's observed ASV-ASV correlation matrix ,  $y$  , and the population mean matrix:  $y - m$  ( see Figure 3-figure supplement 2A for a cartoon example ) . We reasoned that the observed correlation matrix for each host can be approximated by a mixture of contributions from the population mean matrix  $m$  and the host-specific residual matrix  $e$  . To identify the optimal mixture for each host ( i . e . the mixture of consistent vs . individualized correlation patterns that best explained the observed data ) , we titrated the contribution ( i . e . weight ) of  $e$  from 0% to 100% ( and correspondingly , the contribution of  $m$  from 100% to 0% ) and identified the value that minimized the Frobenius distance between the simulated combination and the observed correlation matrix ,  $y$  . In support of prior observations of 'universality' ( Bashan et al . , 2016; Gao et al . , 2020; Vila et al . , 2020; San-Juan-Vergara et al . , 2018 ) , we found that , across hosts , the optimal mixture involved contributions from the shared correlation structure ( i . e .  $m$  ) of between 60% and 80% ( median 70% ) and a host-level contribution ( i . e . from  $e$  ) of between 40% and 20% ( median 30% ) . Hence , population-level signatures contributed almost twice the weight as host-level signatures ( a median population:host ratio of 2 . 33:1; Figure 3-figure supplement 2B ) . As a result , ASV-level relationships tend to be more consistent across hosts than host-specific . One advantage of our approach , compared to DOA ( Bashan et al . , 2016 ) , is that we can identify the bacterial pairs that exhibit the most consistent relationships across hosts . Hence , we next conducted several analyses to understand why some taxon pairs are more consistent than others . To do so , we created a 'universality' score ( Figure 2-figure supplement 4 ) that could be calculated for each ASV pair . The score multiplies the pair's median absolute correlation coefficient across hosts ( y-axis of Figure 3A and B ) with its correlation consistency across hosts ( i . e . proportion of shared sign; x-axis of Figure 3A and B ) . The resulting scores range from 0 to 1 , where a score of 1 equates to perfect 'universality' ( i . e . all hosts have a correlation coefficient of 1 or all hosts have a correlation coefficient of -1 ) . Applying this score to all pairs of ASVs reveals a right-skewed distribution , reflecting the fact that most bacterial correlations are weak , with inconsistent sign directions across hosts ( Figure 3C; Figure 2-figure supplement 3B ) . However , 48% of these scores were higher than expected by chance ( permutation test;  $FDR\leq 0 . 05$ ; Figure 3C; Figure 2-figure supplement 3B ) , reflecting a signal of universality in our data . Despite the bias toward negative ASV-ASV correlations in the overall set of bacterial correlations ( Figure 2-figure supplement 3A ) , we observed no such bias in the most universal pairs . For instance , in the top 5% most universal ASV pairs (  $n=94$  pairs ) , 46 pairs exhibited net positive correlations and 48 pairs , net negative correlations , suggesting no particular bias in the direction of the strongest and most consistent associations . To visualize these highly consistent correlations , we plotted bacterial co-abundance networks connecting the top 5% most universal ASV pairs ( Figure 3C ) . A handful of ASVs were highly connected within this network , especially ASV1 ( genus *Bifidobacterium*; Supplementary file 1a and d ) , which was connected to 14 other ASVs . Three other ASVs were connected to at least 10 other ASVs , including members of families Atopobiaceae ( ASV8 ) , Eggerthellaceae ( ASV28 ) , and Muribaculaceae ( ASV86 ) ( Supplementary file 1e ) . These



e families were also enriched in this network , relative to the rest of the data , as was the family Bifidobacteriaceae ( Figure 3E ) . Pairings between members of the same family were enriched by >3-fold in this network (  $p < 0.0001$  ) , making up 32% of pairs in the most universal set versus only 9.8% of pairs outside that set . Almost two-thirds of these were Prevotellaceae-Prevotellaceae pairs ( 10 of 16 same-family pairs ) . We next asked: does the phylogenetic distance between a pair predict the nature of their relationship ? In support of the idea that homology leads closely related ASVs to respond similarly to the environment , or perhaps facilitate each other's growth ( Meehan and Beiko , 2014; Vacca et al . , 2020 ) , we found that , for positively associated ASV pairs , closely related taxa had higher universality scores than more distantly related taxa ( Pearson's  $r$  for positively correlated pairs =  $-0.232$ ;  $p < 0.0001$ ; Figure 4A ) . In contrast , when ASV pairs were negatively correlated , there was a weak positive relationship between phylogenetic distance and universality ( Pearson's  $r = 0.106$ ;  $p = 0.004$ ; Figure 4B ) . In other words , the strongest and most consistently negatively correlated taxa tend to be distantly related , whereas the strongest and most consistently positively correlated taxa were often closely related , especially members of the family Atopobiaceae ( Supplementary file 1f ) . We next asked whether host attributes , including host sex , social group membership , genetic relatedness , age , or baseline gut microbiome composition , predict host differences in patterns of bacterial correlation . Consistent with studies that use DOA to infer universality ( Bashan et al . , 2016; Gao et al . , 2020; Vila et al . , 2020; San-Juan-Vergara et al . , 2018; Kalyuzhny et al . , 2017 ) , the strongest predictor of distance in bacterial correlation patterns was distance in terms of baseline microbiome composition ( a core assumption of DOA ) . Indeed , a Mantel test correlating compositional distance of average microbial profiles ( as Aitchison distances between the per-host mean of CLR-transformed samples ) with distance in microbial correlation patterns between hosts revealed that 34% of the variation in correlation patterns was explained by baseline microbiome community composition ( Mantel:  $r^2 = 0.336$ ;  $p = 0.001$ ; Figure 5A; Supplementary file 1g ) . Consistent with prior research in our population , which finds widespread heritability for gut microbiome taxon abundances ( Grieneisen et al . , 2021 ) , we also found a weak but significant relationship between host genetic distance and the distance in microbial correlation patterns between hosts , after controlling for similarity of baseline composition across hosts . Hosts who were more closely related based on a multigenerational pedigree have slightly more similar ASV-level correlation matrices ( Figure 5B; Supplementary file 1g;  $r^2 = 0.009$ ; partial Mantel controlling for baseline similarity:  $p$ -value =  $0.002$  ) . We found no evidence that members of the same social group or sex exhibit more similar microbial correlation patterns ( social group:  $F = 2.146$ ;  $p = 0.089$ ; sex:  $F = 2.026$ ;  $p = 0.160$ ; Supplementary file 1g ) . Host age may also predict the overall strength of microbial relationships , and some studies find that gut microbial compositions become more individualized with age ( Risely et al . , 2022; Wilmanski et al . , 2021 ) . This observation suggests that host age may also be linked to individualized microbial relationships . To test these ideas , we divided our hosts into three classes: juveniles ( 0–6 years ) , prime age adults ( 6–13 years ) , and older adults ( 13+ years ) and compared ASV correlation patterns between ( 1 ) the juvenile and prime age class and ( 2 ) between prime age and older adults . Hosts were only included in these analyses if they had >35 samples in either the juvenile and prime age class or prime age and older adult class ( no host had >35 samples in all three age classes; 13 hosts were included in the juvenile and matched prime age adult groups; a separate set of 13 hosts were included in the older adult and matched prime age adult groups ) . We found no evidence that microbial correlations get stronger or weaker with age ( Figure 5—figure supplement 1 ) . Further , using the same methods described above to estimate the relative host- and population-level contributions to ASV correlation patterns , we found no strong differences in the degree of 'personalized' correlation patterns across age groups ( Figure 5—figure supplement 2 ) . Further , ASV correlation patterns were slightly more similar within ve

rsus between adjacent age classes ( juvenile vs . prime age , ANOVA  $p=0.00175$  , 1 . 3% variance explained; prime age vs . older adult , ANOVA  $p=0.0112$  , 0 . 9% variance explained ) . Principal components analysis on the microbial correlation patterns between juvenile and prime age hosts , and between prime age and older adult hosts revealed some age effects , particularly on the second principal component ( Figure 5—figure supplement 3 ) . See the legend of Figure 5—figure supplement 3 and Supplementary file 1i for information on which ASV pairs differed across age categories . Without experiments , we cannot disentangle whether our observed bacterial correlations are due to ecological interactions between bacterial species or to shared responses to environmental gradients , either inside or outside the host . While we were unable to control for many aspects of the host environment ( e . g . gut pH , hormones , or immune profiles ) , we were able to include measures of dietary variation in our models of microbial abundances . Seasonal changes in host diet therefore do not account for universality in microbial relationships across hosts . To account for additional unexplained seasonal variation , we next removed the oscillating seasonal trend from the log-ratio abundances for each ASV ( modeled as a sine wave ) and re-estimated the ASV-ASV correlation matrix ( Figure 4—figure supplement 1 ) . Removing the seasonal trend had little effect on ASV-ASV correlations , as the variance explained by seasonal oscillation was small for all ASVs ( median 1 . 1% , minimum = 0% , maximum = 6% ) . Consequently , the between-ASV correlation estimates were almost identical to those derived from our original model ( Pearson's  $r=0.982$  ,  $p<0.0001$ ; Figure 4—figure supplement 1C ) . Further , ASV pairs where one or more members was from a family that showed strong seasonal changes in a prior analysis of these data ( Björk et al . , 2022 ) , henceforth 'seasonal' families , had only slightly higher universality scores than taxon pairs where neither partner showed strong seasonal changes in abundance ( difference of 0 . 018;  $p<0.0001$ ; Figure 4—figure supplement 2 ) . Because the high level of universality we observed was not well explained by season , we also tested whether universality was explained by synchronized dynamics . We reasoned that if one member of an ASV pair shows highly synchronized dynamics across different hosts , and the other member is also strongly synchronized across hosts , then universality could be an inevitable outcome of strong , but independent synchrony in both members of the pair . We quantified synchrony as the degree to which the observed dynamics of a single , focal ASV are consistent across hosts , such that high synchrony ( near 1 ) implies that the timing and direction of shifts in log-ratio ASV abundance are identical across hosts in the population ( see Methods; Figure 4—figure supplement 3 ) . Estimates of synchrony ranged from 0 . 033 to 0 . 474 ( median=0 . 187 ) . Interestingly , ASVs in the 13 'seasonal' families are not more likely to have high synchrony than other families ( ANOVA ,  $p=0.434$ ; Figure 4—figure supplement 4; Supplementary file 1h ) . The average synchrony of an ASV-ASV pair had a statistically significant but weak relationship with that pair's universality score (  $r=0.116$  ,  $p<0.0001$ ; Figure 4—figure supplement 5 ) . Finally , to investigate parallels between baboon and human microbial communities , we turned to two publicly available gut microbial time series data sets: daily samples from 34 human adults over a 17-day span ( 483 total samples; hereafter Johnson et al . , 2019 ) , and the DIABIMMUNE cohort that consists of 285 samples , collected monthly over 3 years , from 15 infants and toddlers living in Russian Karelia ( Vatanen et al . , 2016; at the time of writing , these cohorts were the only publicly available data sets we could find that included large numbers of repeated samples from the same subjects ) . Because baboons in Amboseli experience less heterogeneity in their environments and diets than humans ( Björk et al . , 2022; Grieneisen et al . , 2021 ) , we expected they would exhibit greater consistency in microbial correlations than either human cohort . Here , we compared each host cohort's universality at the level of correlations between families , orders , and classes because these taxonomic levels offered the greatest comparative power ( 10 . 1% of families/orders/classes overlap between the cohorts compared to just 3 . 1% of genera and no ASVs ) . Contrary to our expectations , we find comparable evidence of universality in baboons and the DIABIMMUNE infant

t/toddler cohort , but weaker evidence for universality in Johnson et al . ( Figure 6A-D ) . Bacterial families in the DIABIMMUNE cohort yielded universality scores slightly higher than those observed in Amboseli ( 25th percentile=0 . 142 , median=0 . 216 , 75th percentile=0 . 321 for DIABIMMUNE; 25th percentile=0 . 088 , median=0 . 150 , 75th percentile=0 . 234 for Amboseli ) , driven by correlations between families that were stronger on average than those estimated for baboons ( median DIABIMMUNE family-family correlation strength=0 . 253; median Amboseli family-family correlation strength=0 . 170 ) . The high level of consistency between both human infants/toddlers and wild baboons is surprising and may be due to the similar sampling intervals for these cohorts . Both cohorts were sampled approximately monthly , while Johnson et al . 's subjects were sampled daily ( Coyte et al . , 2021; Guittar et al . , 2019 ) . Median correlation strengths and universality scores for the Johnson et al . , 2019 , cohort were substantially lower ( median correlation=0 . 090; 25th percentile universality=0 . 050 , median=0 . 086 , 75th percentile=0 . 113 ) than the DIABIMMUNE cohort or the baboons . Despite considerable differences in the hosts , time scales , and designs of these studies , all three data sets exhibited a positive correlation between correlation strength and sign consistency for family pairs ( Figure 6D ) . This trend was strongest in the Amboseli baboons ( exponent of power regression  $b=1.72$ ;  $p<0.0001$  ) ; weaker in the DIABIMMUNE cohort (  $b=1.51$ ;  $p<0.0001$  ) and weakest in Johnson et al . , 2019 (  $b=1.19$ ;  $p<0.0001$  ) . Further , the most universal family-family associations skewed positive in both the baboons and the infant data set . All of the top 5% most universal family pairs ( 30 of 30 pairs ) are positively associated in the DIABIMMUNE cohort , compared to 70% ( 23 of 33 pairs ) in the Amboseli baboons . Finally , we examined the relationship between universality scores for family pairs that overlapped between Amboseli and DIABIMMUNE (  $n=29$  pairs ) , and between Amboseli and Johnson et al . , 2019 ( Figure 6E;  $n=21$  pairs; only 10 family pairs overlapped between all three data sets ) . For these overlapping pairs , scores in the Amboseli data predicted scores for the same family-family pair in the DIABIMMUNE data set (  $r=0.562$  ,  $p=0.001$  ) . The association between scores in the Amboseli data and the Johnson et al . data was negative , but not statistically significant (  $r=-0.402$  ,  $p=0.071$  ) .

Do different hosts have different microbiome 'ecologies' ? Ecological and evolutionary processes like horizontal gene flow , genotype by environment interactions , and priority effects have been predicted to lead bacterial species to occupy different niches ( with different ecological interactions ) in different communities ( Dolinšek et al . , 2016; Franzosa et al . , 2015; Faith et al . , 2013; Bik et al . , 2016; Caporaso et al . , 2011; Costello et al . , 2009; Louca et al . , 2018; Rainey and Quistad , 2020; Martiny et al . , 2015 ) . Yet contrary to these expectations , here we find that hosts in the same population exhibit pairwise bacterial correlation patterns that are predominantly shared across hosts , rather than idiosyncratic to individual hosts . If these shared correlation patterns arise from shared microbiome ecologies , this discovery has consequences for understanding the basic eco-evolutionary drivers of microbiome dynamics and for human and animal health . For instance , shared ecologies would mean that designing widely applicable microbiome interventions is a more attainable goal than personalized microbiome compositions would suggest . Shared microbiome ecologies may also enable researchers to develop microbiome interventions that leverage these interactions to manipulate the microbiome's emergent community properties to improve host health . By measuring bacterial correlations in multiple hosts , we were also able , for the first time , to pinpoint which pairs of bacterial taxa exhibit the most consistent relationships across hosts . We found that most bacterial abundance correlations—from ASV-ASV to phyla-phyla relationships—were weak and negative . Positive bacterial interactions have been the subject of recent discussion in the literature ( Loftus et al . , 2021; Palmer and Foster , 2022; Kehe et al . , 2021 ) . Ecological theory predicts that strong positive interactions should be rare in natural communities because species interdependencies can hamper community assembly and stability ( Coyte et al . , 2015; Coyte et al . , 2021 ) . This th

theory is supported by experiments that directly measure the effects of one bacterial species on another's growth ( Weiss et al . , 2022; Ortiz et al . , 2021; Carlström et al . , 2019; Venturelli et al . , 2018 ) ( but see Kehe et al . , 2021 ) . Our results suggest that strong , positive bacterial correlations are indeed uncommon in intact , unmanipulated microbiomes: significant positive relationships made up just 3 . 8% of all the pairwise correlations we observed . Hence , strong mutualisms , while key to microbiome function and dynamics , are probably rare in gut communities . While mutualisms and universal dynamics are important , the correlation patterns we observe likely arise from a combination of ecological interactions between bacteria and shared responses to the environment ( i . e . pairs of bacteria that prefer the same or different environments ) . In support of the idea that at least some of the correlations we observed are due to between-species interactions , our signature of universality was essentially unchanged after accounting for some of the strongest known drivers of microbiome composition and change in our population—host diet and season ( Björk et al . , 2022; Grieneisen et al . , 2021 ) —as well as microbial synchrony between hosts . However , our approach did not account for important environmental gradients within the gut , such as host immune profiles and intestinal pH . These factors also shape microbiome composition ( e . g . Reese et al . , 2018; Firrman et al . , 2022; de Vos et al . , 2022 ) , and can lead to shared abundance correlations between hosts even if hosts themselves differ . Ecological selection via within-host environments may explain our finding that genetic relatives share somewhat similar bacterial correlation patterns . Ecological selection is also consistent with our observation that the most consistent ASV-level correlations are between phylogenetically related taxa , and these patterns were strongest for positively associated taxon pairs . In support , phylogenetically related species have been shown to have similar environmental preferences ( Tamames et al . , 2016 ) . We note that none of the correlations we observed can be mapped directly to standard categories of pairwise ecological interactions , such as mutualism , commensalism , amensalism , exploitation , or competition . Experimental approaches that directly measure the effects of one species on another's growth in vitro are better suited to characterizing these relationships ( Kehe et al . , 2021; Weiss et al . , 2022; Ortiz et al . , 2021; Carlström et al . , 2019; Venturelli et al . , 2018 ) . The strong signal of universality we observed in bacterial abundance correlations stands in contrast to the common observation that microbiome taxonomic composition ( i . e . the presence and abundance of bacterial species ) is almost always highly personalized ( Franzosa et al . , 2015; Faith et al . , 2013; Bik et al . , 2016; Caporaso et al . , 2011; Costello et al . , 2009; Risely et al . , 2021; Kolodny et al . , 2019; Flores et al . , 2014; Johnson et al . , 2019; Pruss et al . , 2021 ) . Our own prior analyses of these data found that each baboon exhibited personalized microbiome compositions and asynchronous single-taxon dynamics ( Björk et al . , 2022 ) . These contrasting patterns—personalized compositions but shared abundance correlations—are important because personalized microbiota have been proposed to arise , at least in part , from personalized microbiome ecologies ( Franzosa et al . , 2015; Faith et al . , 2013; Bik et al . , 2016; Caporaso et al . , 2011; Costello et al . , 2009; Risely et al . , 2021; Kolodny et al . , 2019; Flores et al . , 2014; Johnson et al . , 2019; Pruss et al . , 2021 ) . We can think of at least three explanations that reconcile these observations . First , consistent with ideas discussed above , if environments in the gut shape bacterial abundances but these environments are not synchronized across hosts , this would lead to shared abundance correlations over time , but individualized microbiome compositions at any single point in time . Second , the effects of horizontal gene transfer and gene by environment interactions on microbial phenotypes may not be strong enough to substantially alter pairwise microbial associations in the gut . This may be especially true for our main unit of analysis , bacterial ASVs . Because ASVs encompass multiple species and strains , each with somewhat different functional capacities , their dynamics may be buffered against idiosyncrasies driven by horizontal gene transfer and functional redundancy , which affect single strains more strongly than

n whole species or genera . We would strain-level correlation patterns to be more individualized than those between ASVs . Third , personalized gut microbial compositions may emerge from at least two other phenomena: personalized assembly processes and interactions driven by rare , host-specific strains ( which were necessarily excluded from our analyses ) ( Costello et al . , 2012; Walter and Ley , 2011 ) . In general , a logical next step would be to confirm the microbial correlation patterns we observed using culture-based approaches , which will help reveal ( in vitro ) whether they can be attributed to direct effects of one microbe on another's growth . The observation that bacterial correlation patterns are largely shared across hosts was also apparent in one human data set , despite between-study differences in study design , host age , and time scale . Specifically , both the Amboseli baboons and the DIABIMMUNE infant/toddler cohort from Russia ( Vatanen et al . , 2016 ) exhibit comparable levels of universality of correlation patterns . This outcome surprised us: because the baboons all live in the same environment and are presumably colonized by similar bacterial strains from that environment , we expected that ecological selection and shared strain functionality should lead to stronger universality in bacterial correlation patterns compared to human infants sampled from different households and who were probably colonized by different strains . We also found that the most universal correlations between bacterial families in baboons tended to be highly universal in human infants/toddlers . Hence , some bacterial families may exhibit consistent microbial relationships within hosts , across host populations , and across host species . Finally , a recent , independent study also identified consistent bacterial correlation patterns across four different populations of human hosts ( Loftus et al . , 2021 ) . While this study lacked resolution at the level of individual hosts , it did identify a conserved network of positively associated and closely related microbes similar to those we identify in Figure 3 . The authors speculate that these conserved associations may indicate strong partner fidelity or obligate partnerships . We did , however , fail to detect universality in a second human data set reported in Johnson et al . , 2019 , in which subjects were sampled daily , rather than weekly or monthly . The lack of universality in Johnson et al . , 2019 , may be due to this difference in sampling time scale , especially if daily abundances and correlations are noisier than covariances modeled over the longer time scales in our study . In support , many fewer of the microbial correlations were stronger than random chance in Johnson et al . as compared to the baboons or children in the DIABIMMUNE cohort . However , without the ability to subsample Johnson et al . , 2019 , to monthly scales ( this data set is only 17 days long ) , it is impossible to test this prediction . The subjects in Johnson et al . , 2019 , also consumed substantially different diets from each other , perhaps more so than the children in the DIABIMMUNE cohort , and this inter-host difference in diet may reduce the universality of microbial correlations . In sum , our study indicates that microbiome personalization may not extend to microbiome community ecology . However , more work is needed to understand how relationships between microbiome taxa are explained by shared internal and external environments , direct and indirect ecological interactions , taxonomic levels ( e . g . strains to phyla ) and time scales ( days to months and years ) . Future studies should also consider how pairwise bacterial interactions scale up to affect the emergent properties of the community ( Levine et al . , 2017; Letten and Stouffer , 2019; Friedman et al . , 2017 ) . We hope that our longitudinal data set and the new methods we developed as part of this study ( e . g . the model of log-ratio dynamics , the assessment of covariation from time-ordered abundance trajectories , and the universality score ) will be useful in this enterprise .

The baboon hosts in this study were members of the Amboseli baboon population , which has been studied by the Amboseli Baboon Research Project since 1971 ( Alberts and Altmann , 2012 ) . The microbiome compositional profiles are derived from PCR amplification of an ~390-bp-long fragment that encompassed the V4 region of the 16S rRNA gene using primers 515F-806R . These microbiome profiles were previously analyzed in Björk et al . , 2022 and Grieneisen et al . , 2021 . Our analyses use 5534 of these

profiles from 56 especially well-sampled baboons , collected over a 13 . 3-year span between 2000 and 2013 ( Figure 1B ) . Each baboon host in this data set was sampled at least 75 times ( mean number of samples=99; range=75-181 samples; median number of days between samples within hosts=20 days; 25th percentile=7 days , 75th percentile=49 days ) . DNA was extracted from each sample using the MoBio and QIAGEN PowerSoil kit with a bead-beating step . All samples were sequenced on an Illumina HiSeq 2500 , with a median read count of 48 , 827 reads per sample across all 5534 samples ( range=982-459 , 315 reads per sample ) . Following recommended statistical practices ( Gloor et al . , 2017 ) , samples were not rarefied , but counts were agglomerated and transformed to additive log-ratios ( ALR ) . Variation in sampling depth and relative abundance were modeled by the method described in the subsequent section . Further details of sample collection , DNA extraction , and sequencing can be found in ( Björk et al . , 2022; Grieneisen et al . , 2021 ) . Data sets of per-sample taxonomic counts were produced at each of three taxonomic levels , from finest to coarsest: ASV , taxonomic assignments finer than phyla , but above the genus level ( e . g . class , order , family ) , and phylum . At the intermediate and coarsest levels , taxa were agglomerated using phyloseq's tax\_glom ( ) function ( McMurdie and Holmes , 2013 ) such that all sequence variants sharing taxonomic identity at that level were collapsed into a single taxon ( e . g . family Bifidobacteraceae ) . To reduce sparsity in the data set , remove 16S sequences that could represent gene duplications , and focus only on taxa that were prevalent in all 56 hosts , we further filtered as follows: ( 1 ) in each of the three taxonomically defined data sets ( i . e . ASV , taxa assigned to family/order/class , and phylum ) , we identified taxa present ( i . e . having non-zero abundance ) in at least 20% of samples from each host; ( 2 ) if a given ASV was >99% genetically similar to another ASV we removed the least abundant of the pair to minimize the risk of including duplicate 16S rRNA gene copies from the same taxa ( Vetrovský and Baldrian , 2013 ) ; and ( 3 ) counts associated with all other taxa were combined into a dummy category , hereafter referred to as 'other . ' The 'other' category therefore includes a combination of rare and host-specific gut microbes . This category was retained in the data set ( although not analyzed directly ) because 'other' counts still inform the precision of the observed relative abundances in our model . See the sub-section titled 'Filtering out taxon pairs with frequent joint absence' below for further filtering that was required to avoid biases in estimating correlations between taxon pairs . Characteristics of the filtered data at each taxonomic level are provided in Supplementary file 1a-1c . At the ASV level , these filtering steps eliminate a majority of ASVs and ASV pairs from consideration: from 22 , 097 unique ASVs before filtering to 107 after filtering . This filtering also retained 12 phyla and 35 taxa at the class/order/family level ( Supplementary file 1a-1c ) . Our modeling approach is similar to several published methods for modeling microbial time series data . There are three key features from our perspective: the use of log-ratios ( discussed above ) , the use of a state space model , and the Gaussian process component . State space models are useful for modeling a dynamic process that is observed only after the introduction of some measurement ( e . g . Joseph et al . , 2020 ) . The Gaussian process component helps contend with irregularity in the sampling of our data . Rather than evolving in discrete jumps from one time point to the next , it allowed us to model the change in microbial log-ratio abundances as smoothly flowing through interruptions in observation . Other authors have made similar choices ( Äijö et al . , 2018 ) . Specifically , estimates of taxon-taxon covariance were obtained from the basset model of the 'fido' package in R ( Silverman et al . , 2022 ) . Data for each host took the form of a  $D \times N$  count matrix , where  $D$  gives the number of taxa and  $N$  the number of samples collected for a given host . The following model was fit to each host's count matrix (  $Y$  ) , where  $Y_i$  represents the counts associated with a single sample:  $Y_i \sim \text{Multinomial}(\pi_i, \mathbf{1} = \text{ALR}^{-1}(\eta_i))$   $\eta \sim \text{Normal}(\Lambda, \Sigma, I)$   $\Lambda \sim \text{GP}(\theta[X], \Sigma, \Gamma[X])$   $\Sigma \sim \text{inv-Wishart}(\Xi, \nu)$  The observed relative abundances are considered to be drawn from a multinomial distribution parameterized by a set of proportions (  $\pi$  ) which have an analogous representation in

the ALR. The dynamics of these log-ratio abundances ( $\eta$ ) are described by what amounts to a state space model in the third and fourth lines of the specification above, where a Gaussian process models the evolution of a ‘latent’ state. The matrix  $\Sigma$  captures covariation in log-ratio abundances (the  $D$  rows of the observed count matrix  $X$ ). Sample-sample covariation arising from nearness in time (autocorrelation) is modeled by the kernel matrix  $\Gamma$ . Both the kernel matrix and the expected baseline log-ratio abundances ( $\theta$ ) are parameterized by a set of time-varying covariates  $X$  which included the day of sampling (where the date of first sample is defined as zero) and the first three principal components of diet composition, calculated following Björk et al., 2022; Grieneisen et al., 2021, as the diet of all juveniles and females living in the host’s social group in the 30 days prior to sample collection. All group members consume highly similar diets as they travel together across the habitat, encountering the same resources at the same time (Björk et al., 2022; Grieneisen et al., 2021). These data are collected via random-order behavioral observations collected two to four times per week on adult females and juveniles in each social group. The kernel matrix  $\Gamma$  was composed of two component squared exponential kernels. The first, intended to manage sample-sample autocorrelation, was selected to have a bandwidth such that this autocorrelation decayed to a minimum at 90 days. This mirrored the behavior of estimates of sample-sample autocorrelation in the raw data. The second component kernel modeled sample-sample covariance driven by similarity in composition of diet. The relative weight of these effects—autocorrelation and diet-on sample-sample variation was set at 3:1. We fit four alternative versions of our models in order to test the sensitivity of these parameter settings, varying the bandwidth of the squared exponential kernel in such a way as to give minimum sample-sample autocorrelation at either 30 or 90 days. We also varied the proportion of sample-sample covariance driven by diet from 0% to 25% to 50%, and we varied the log scale of total sample-sample variance between 1 and 2. In all cases, estimates of correlation between CLR ASVs were similar, with minimum and maximum  $r^2$  of estimates between ‘canonical’ and alternatively parameterized models of 0.993 and 0.996, respectively. This suggests our findings are reasonably robust to a range of hyperparameter settings. Posterior inference on this model is performed as described in Silverman et al., 2022, and yields estimates of the distributions of parameters necessary to reconstruct trajectories for all log-ratio taxa across sampling time. In particular, we extract the posterior estimates of one such parameter,  $\Sigma$ , the covariance of ALR taxa, from the fitted models for each host. We convert these covariance matrices over ALR taxa to the CLR form (a simple linear transformation of the matrix). We then normalize estimated CLR covariance matrices to Pearson’s correlation matrices in R using the built-in `cov2cor()` function. The ASV-level relative abundance data were sparse, even after filtering low abundance taxa (i.e. those present in <20% of samples in each host). Zeros comprised 29.7% of the ASV-level count matrix, 16.1% of the class/order/family count matrix, and 9.8% of the phylum level count matrix. This abundance of ‘missing’ observations at the ASV level led to a bias in estimates of taxon-taxon correlation: a high-frequency joint zero abundances for a pair of taxa increased the apparent positive correlation of those taxa (Figure 1–figure supplement 3). This is because abundances below some minimum level of sensitivity in sampling will be ‘flattened’ to zero, reducing the observed variation for those pairs, over those samples, to zero. This loss of variation leads to a tendency to overestimate the (positive) correlation of these pairs. This trend was observed for our basset model (Silverman et al., 2022) when estimating either Pearson’s correlation or proportionality (Quinn et al., 2017). It was also observed in ASV-ASV CLR correlations output by COAT (Cao et al., 2019), which estimates CLR correlation through a sparsity-inducing procedure intended to yield more conservative estimates, and by SparCC (Friedman and Alm, 2012, see Figure 1–figure supplement 3). To avoid this bias, we restricted our analyses to taxon pairs with strictly less than a 5% frequency of joint absence (i.e. joint zero abundance observations in less than 5% of all samples across hosts) and less t

have a 50% frequency of absence in either taxon individually across all samples. We illustrate these filtering criteria with a cartoon example involving four taxa, jointly sampled 20 times, indicating presence as an 'X' and absence (zero abundance) with a dash: Taxon A (70% present): X---XXXXXX-XXX-X-XXX Taxon B (20% present): X-----X--X---X- Taxon C (50% present): -XXX-----XX-XXXX--X- Taxon D (65% present): XX-----XX-XX-XX-XXXX In this example, taxon B would be excluded from analyses by the requirement that taxa be 'present' or non-zero in at least half of all samples. Its associations with other taxa (A x B, B x C, or B x D) would be omitted from analyses, leaving pairs A x C, A x D, and C x D. Of these, the rate of joint absence is 5%, 10%, and 15%, respectively, meaning only A x C would pass the filtering criterion of no more than 5% joint absence. After filtering, 1878 of the original 7750 ASV-ASV pairs remained in this 'high-confidence' set. It is this set that we present in all figures and results. This procedure was also carried out at the phylum and class/order/family levels. However, as the frequency of absence was generally small at these higher taxonomic levels, a higher proportion of possible taxon-taxon pairs were included in the high-confidence set: 86.4% (57 of 66 pairs) at the phylum level and 71.0% (473 of 666 pairs) at the class/order/family level. We devised a universality score for each pair of taxa intended to capture the strength and consistency of taxon-taxon correlations across hosts (Figure 2—figure supplement 4). The majority direction is negative otherwise. This score identifies the sign of the taxon-taxon correlation (positive or negative) that is most common across the 56 hosts (i.e., occurs in >50% of the 56 hosts in the data set). The direction of this sign is the 'majority correlation sign.' For a pair of taxa  $i$ , let  $n_{i,maj}$  be the number of hosts with CLR correlation over pair  $i$  with the majority correlation sign for that pair and let  $n$  be the total number of hosts. Let  $R_i$  be the subset of estimated CLR correlations for pair  $i$  across hosts with the majority sign. The universality score  $u_i$  for that taxon-taxon pair is then given by  $u_i = n_{i,maj} \times \text{median}(|R_i|)$ . This score is the product of the median CLR correlation across hosts and the proportion of hosts with the majority correlation sign, and is bounded between 0 and 1. Scores near 1 indicate strong universality and near-zero scores indicate weak universality. Strong universality can only be achieved by taxon-taxon correlations that are both large in magnitude and highly concordant across hosts. We identified correlations stronger than expected by chance using permutations of the data set to define empirical null distributions (Figure 2—figure supplement 3A). Specifically, we permuted the microbial count tables by randomly shuffling taxon identity within each sample 10 times for each of the 56 hosts. This procedure maintained relative abundance patterns within a sample but scrambled the covariance patterns of relative abundances. These randomly generated correlations were pooled into a single reference distribution. The distributions of ASV-level CLR correlations in the original and permuted data are shown in Figure 2—figure supplement 3A. We identified 'significant' correlations as those below  $\text{FDR} \leq 0.05$  (Benjamini-Hochberg), testing against the permuted data. We applied an analogous permutation test to derive a null distribution for taxon-taxon universality scores. In a single iteration of this permutation procedure, rows and columns of the observed taxon-taxon correlation matrix for each host were shuffled, maintaining the distribution over observed correlations at the host level but randomizing the identity of taxon pairs across hosts. This procedure was repeated 100 times and universality scores were calculated from each of these shuffled data sets to give a single pseudo-null distribution of universality scores. The observed and null distributions of universality scores at the ASV level are shown in Figure 2—figure supplement 3B. We used this empirical null distribution to identify universality scores significantly greater than expected ( $\text{FDR} \leq 0.05$ ). We used simulations to estimate the degree of shared 'signal' between hosts in terms of taxon-taxon correlations. Each host's 'observed correlations' were defined as the *basset* estimated maximum a posteriori estimates of CLR ASV correlations for that host. We computed the mean correlations across the population using the function `estcov()` from the *shapes* package in R (Chen, 2020) and estimated a host-s



specific contribution to the observed correlations as the residual difference between per-host observed and these mean correlations. That is, observed host correlations = mean population correlations + host residual. For each host, we simulated a hypothetical set of composite taxon-taxon correlations as a convex combination of mean and host residual:  $\text{composite correlations} = 1 - \alpha \times \text{mean population correlations} + \alpha \times \text{host residual}$ . A cartoon example of this procedure is given in Figure 3—figure supplement 2. For example, one such simulated set of taxon-taxon correlations might constitute a mixture of 90% host contribution and 10% shared population-level ‘signal’ ( $\alpha = 0.9$ ). Alternatively, a small host-level contribution might have  $\alpha = 0.1$ . For each host, we iterated over increasing proportions of host-level contribution (from 0% to 100%), generating simulated composite correlation matrices according to the formula above. We compared these simulated patterns to those observed for the same host, reasoning that simulated correlation matrices that minimize the distance between the observed correlation matrices and the simulated mixtures provide the best description of the underlying true mixture. Seasonally de-trended data was obtained in the following way. The observed ASV count matrices were CLR-transformed and linear autoregressive models were fit to each CLR-transformed ASV’s series. In these models, wet-dry season oscillation was modeled as a sine wave with a period of 365 days. The magnitude of this component was estimated during model fitting after an offset (in weeks) was estimated in a first step, in order to best align the oscillating seasonal component with the data. Per-ASV models were fit using the following syntax: `arima(x=x, xreg=model.matrix(~factor(host)+sinf(offset,days))[,-1], order=c(1,0,0))` where `x` gives CLR counts for that ASV, the ‘order’ argument of `arima` enables a single autoregressive component, and the ‘xreg’ argument specifies a covariance matrix. That covariance matrix contains a per-host label giving host-specific offsets for log-ratio abundance and an oscillating seasonal trend through ‘`sinf`’, a function that samples values corresponding to day indices (through ‘`days`’) from a sine wave with weekly offset (‘`offset`’) and a period of 365 days. The residuals from these per-ASV model fits were extracted and used as the seasonally ‘de-trended’ data (see Figure 4—figure supplement 1). Correlations across CLR ASV-ASV pairs were estimated from these residual series with the `cov()` function in R. ‘Synchrony’ was estimated by sampling aligned microbiome compositional profiles across hosts. We identified all samples collected from pairs of hosts within 1 calendar day. For instance, a sample collected from host F01 on March 14, 2011, could pair with a sample from M04 on March 15, 2011. For all possible pairs of hosts, we selected one such aligned pair of samples, yielding 1540 joint observations of gut microbiome composition. For each such paired sample, one host was arbitrarily designated as host A and the other as host B. The ‘synchrony’ of a given taxon was estimated as the correlation of a taxon’s model-inferred log-ratio abundance across the set of samples from hosts labeled A and the set of samples from hosts labeled B. The cartoon in Figure 4—figure supplement 3 illustrates this sample pairing. We performed enrichment analyses for bacterial families and family pairs in several settings. In each case we computed the frequency of ASVs belonging to a given family, or of pairs belonging to a family pair, on a subset of the data. These were compared to the overall frequencies of ASVs belonging to those families or pairs. To determine the enrichment of families and family pairs in the most universal ASV pairs (Figure 3E; Supplementary file 1e), we calculated the frequencies of ASV families and pairs in the top 5% of pairs by universality scores. Significant enrichment of families or pairs was identified using a one-sided Fisher’s exact test. Multiple test correction was applied as a Benjamini-Hochberg adjustment to observed p-values. Phylogenetic distances between ASV sequences were calculated with the `dist.ml` function in the ‘`phangorn`’ package in R (Schliep, 2011) using default settings for amino acid substitution rates. In Supplementary file 1f, low phylogenetic distance/high median correlation strength pairs were identified as those with phylogenetic distances of less than 0.2 and median correlation strengths of greater than 0.5. Again, significance of these was evaluated against overall frequencies of the same families and

d pairs . We compared our findings to those generated from two human data sets: the DIABIMMUNE project's infant/toddler cohort from Russian Karelia ( Vatanen et al . , 2016 ) and the adult diet-microbiome association study of Johnson et al . , 2019 . In both cases , count tables were obtained from the project's public website and subject identity and sampling schedules were available in the associated metadata . We compared each host cohort's universality at the family/order/class level because this taxonomic level offered the greatest comparative power ( 10 . 1% of families/orders/classes overlap between the cohorts compared to just 3 . 1% of genera and no ASVs ) . The basset model from the 'fido' R package ( Cullen et al . , 2020 ) was fit to each subject's data set using model settings analogous to those employed on the Amboseli baboon series: first , only taxa with non-zero counts in at least 20% of all subjects' series were retained; second , Gaussian process kernel bandwidth settings were chosen in such a way as to encode an expectation of minimum autocorrelation between samples at a distance in time of 90 days . We extracted CLR estimates of taxa at the family level in the same manner as described previously for the Amboseli data set .

In [10]: *# we can create a function to include headings and keywords into the article text*

```
def build_text_with_headings(item):
    """
    return the article text with heading embedded
    """
    result = ""
    paras = item["article"].split("\n")
    keywords = ', '.join(item["keywords"])
    keywords = f"## Keywords: {keywords}"
    # print(keywords)
    result += keywords + "\n"
    headings = item["headings"]
    if len(paras) != len(headings):
        print("Error, not matching length")
        return item["article"]
    else:
        for (heading, paragraph) in zip(headings, paras):
            result += f"## {heading}\n{paragraph}\n\n"

    return result
```

In [11]: processed\_text = build\_text\_with\_headings(item)  
print(processed\_text)

## Keywords: microbiology and infectious disease, ecology

## Abstract

Ecological relationships between bacteria mediate the services that gut microbiomes provide to their hosts . Knowing the overall direction and strength of these relationships is essential to learn how ecology scales up to affect microbiome assembly , dynamics , and host health . However , whether bacterial relationships are generalizable across hosts or personalized to individual hosts is debated . Here , we apply a robust , multinomial logistic-normal modeling framework to extensive time series data ( 5534 samples from 56 baboon hosts over 13 years ) to infer thousands of correlations in bacterial abundance in individual baboons and test the degree to which bacterial abundance correlations are ‘universal’ . We also compare these patterns to two human data sets . We find that , most bacterial correlations are weak , negative , and universal across hosts , such that shared correlation patterns dominate over host-specific correlations by almost twofold . Further , taxon pairs that had inconsistent correlation signs ( either positive or negative ) in different hosts always had weak correlations within hosts . From the host perspective , host pairs with the most similar bacterial correlation patterns also had similar microbiome taxonomic compositions and tended to be genetic relatives . Compared to humans , universality in baboons was similar to that in human infants , and stronger than one data set from human adults . Bacterial families that showed universal correlations in human infants were often universal in baboons . Together , our work contributes new tools for analyzing the universality of bacterial associations across hosts , with implications for microbiome personalization , community assembly , and stability , and for designing microbiome interventions to improve host health .

## Introduction

Mammalian gut microbiomes are highly diverse , dynamic communities whose members exhibit the full spectrum of ecological relationships , from strong mutualisms like syntrophy and cross-feeding , to competition , parasitism , and predation ( Faust and Raes , 2012; Foster and Bell , 2012; Dolinšek et al . , 2016; Seth and Taga , 2014 ) . These relationships mediate a variety of biological processes that have powerful effects on host health and fitness , including the metabolism of complex carbohydrates and toxins , and the synthesis of physiologically important compounds , like short-chain fatty acids , neurotransmitters , and vitamins ( Faust and Raes , 2012; Foster and Bell , 2012; Dolinšek et al . , 2016; Seth and Taga , 2014; Bäckhed et al . , 2005; Gould et al . , 2018; Pontrelli et al . , 2022; Degnan et al . , 2014 ) . Despite their importance , major gaps remain in our understanding of microbial relationships in the gut ( Faust and Raes , 2012; Loftus et al . , 2021; Bashan et al . , 2016 ) . We typically do not know if the abundance of one microbe consistently predicts the abundance of other microbes in the same host community , nor do we understand whether these correlative relationships are consistent in strength or direction across hosts ( Bashan et al . , 2016; Widder et al . , 2016; Cao et al . , 2017; Faust and Raes , 2016 ) . Knowing the overall direction and strength of these correlative relationships is important to understanding the ecological relationships that mediate gut microbial processes and shape gut microbiome assembly , stability , and productivity ( Coyte et al . , 2015; Palmer and Foster , 2022; Hu et al . , 2022 ) . For instance , sets of microbes that exhibit strong , positive relationships within hosts may represent networks of cooperating taxa that promote each other’s growth ( Bäckhed et al . , 2005; Loftus et al . , 2021; Wu et al . , 2021 ) . In turn , these strong , mutualistic interdependencies can create an ecological house of cards where microbes rise and fall together , hampering community assembly and stability ( Coyte et al . , 2015; Coyte et al . , 2021 ) . Further , understanding the degree to which correlative relationships between microbes are the same or different in different hosts can shed light on whether hosts share similar , underlying microbial ecologies ( Loftus et al . , 2021; Bashan et al . , 2016; Gao et al . , 2020; Vila et al . , 2020; San-Juan-Vergara et al . , 2018 ) . Microbial ecologies that are similar across hosts ma

y make it possible to manipulate the microbiome's emergent properties to improve host health ( Loftus et al . , 2021; Bashan et al . , 2016; Cao et al . , 2017; Coyte et al . , 2015; Coyte et al . , 2021; Gonze et al . , 2018 ) . To date , there are several reasons to think that correlative relationships in the gut microbiome will not be consistent across hosts and will instead be individualized to each host . For instance , several common community and evolutionary processes—such as horizontal gene transfer and priority effects—can lead microbiome taxa to fill different ecological roles in different hosts ( Dolinšek et al . , 2016; Franzosa et al . , 2015; Faith et al . , 2013; Bik et al . , 2016; Caporaso et al . , 2011; Costello et al . , 2009 ) . Further , genotype by environment interactions and plasticity could lead some microbes to adopt context-dependent metabolisms and ecological roles depending on their microbial neighbors or other aspects of the environment ( Louca et al . , 2018; Rainey and Quistad , 2020; Martiny et al . , 2015; Debray et al . , 2022 ) . Finally , the common observation that gut microbial community compositions ( i . e . the presence and abundance of taxa ) are highly individualized is sometimes attributed to host-specific microbial ecologies and relationships ( Franzosa et al . , 2015; Faith et al . , 2013; Bik et al . , 2016; Caporaso et al . , 2011; Costello et al . , 2009; Risely et al . , 2021; Kolodny et al . , 2019; Flores et al . , 2014; Johnson et al . , 2019; Pruss et al . , 2021 ) . However , to date , the handful of studies that have tested the generalizability of gut microbial relationships across hosts suggest that microbiome community ecology is not highly individualized and is instead largely consistent ( i . e . 'universal' ) across hosts ( Figure 1A; Bashan et al . , 2016; Gao et al . , 2020; Vila et al . , 2020; San-Juan-Vergara et al . , 2018; Kalyuzhny et al . , 2017 ) . For instance , Bashan et al . , 2016 , inferred 'universal' gut microbial relationships in the human gut microbiome by applying dissimilarity overlap analysis ( DOA ) to cross-sectional samples from several human data sets . DOA infers universal microbial relationships by testing whether pairs of hosts who share many of the same microbiome taxa also have similar abundances of those taxa ( Bashan et al . , 2016; Gao et al . , 2020; Vila et al . , 2020; San-Juan-Vergara et al . , 2018; Kalyuzhny et al . , 2017 ) . This approach relies on the assumption that , when two communities share many of the same species and have similar abundances of those species , they do so because of a shared , underlying set of between-species bacterial abundance relationships ( Bashan et al . , 2016; Kalyuzhny et al . , 2017 ) . While many studies using this approach find evidence that microbial relationships are 'universal' ( Bashan et al . , 2016; Gao et al . , 2020; Vila et al . , 2020; San-Juan-Vergara et al . , 2018 ) , DOA's assumptions have been questioned because environmental gradients , stochastic processes , and the presence of many non-interactive species can lead to the spurious detection of universality ( Bashan et al . , 2016; Kalyuzhny et al . , 2017; Marsland et al . , 2020 ) . An obvious alternative is to measure microbial correlations directly from microbiome time series from several hosts ( Loftus et al . , 2021; Faust et al . , 2015 ) . Unlike DOA , this approach should be able to pinpoint which microbiome taxa exhibit the most and least consistent relationships across hosts . However , measuring microbial correlations from longitudinal , multi-host microbiome time series has its own challenges: time series with adequately dense sampling are rare , and most such data sets exhibit temporal autocorrelation and irregular sampling ( Faust et al . , 2015 ) . Further , the most common , and still most feasible , way to collect microbiome community data—via high-throughput sequencing—generates noisy count data that usually can only be interpreted in terms of relative ( not absolute ) abundances ( Gloor et al . , 2017; Quinn et al . , 2017 ) . Finally , correlation cannot be used to infer causality , and in the absence of experiments , we cannot differentiate whether microbial correlation patterns arise from ecological interactions ( e . g . competition , predation , facilitation ) or shared responses to the environment . To address several of these challenges , here we combine extensive time series data on the stool-associated microbiota with a multinomial logistic-normal modeling framework ( Figure 1; n=5534 samples from 56 baboons; 75-181 samples per baboon across 6-13 . 3 years , between 2000 and 2013; Alberts and

Altmann , 2012; Björk et al . , 2022; Grieneisen et al . , 2021 ) . This framework uses 16S rRNA sequencing count data to learn a smoothly evolving Gaussian process . The baboons were the subject of long-term research on individually recognized animals by the Amboseli Baboon Research Project in Kenya , which has been studying baboon ecology and behavior in the Amboseli ecosystem since 1971 ( Alberts and Altmann , 2012 ) . The baboons range over the same habitat and experience similar diets and sources of microbial colonization , facilitating inference about the consistency of microbial correlations across hosts ( Figure 1–figure supplement 1; Björk et al . , 2022; Grieneisen et al . , 2021 ) . To partly account for environmental drivers of microbial dynamics , our modeling approach controls for variation attributable to seasonal changes in the animals’ diets , proportionality in the count data , and irregularity in sampling to produce per-individual , per-taxon trajectories of log-ratio abundances that we used to estimate pairwise microbial correlations within each host . We pursued five main objectives . First , we characterized the overall sign and strength of pairwise correlations in bacterial abundance within each host . Second , we tested the degree to which these correlation patterns are systematically consistent across hosts or individualized by host ( Figure 1A ) . Third , we identified phylogenetic and host-related predictors of the direction and universality of bacterial correlations , including phylogenetic relationships between microbes , host age , and host genetic relatedness . Fourth , we tested whether the microbial correlations we observed were driven by shared responses to host diets or seasonality , or by synchronized microbial dynamics across hosts . Fifth , we tested the generalizability of our findings by comparing the patterns of universality in our data set to two microbiome time series from humans ( Johnson et al . , 2019; Vatanen et al . , 2016 ) . Our predictions for these analyses were influenced by ideas from community and microbial ecology . First , because strong interdependencies can hamper community assembly and destabilize community dynamics ( Coyte et al . , 2015; Palmer and Foster , 2022; Hu et al . , 2022; Coyte et al . , 2021 ) , we expected that most microbial correlations would be weak with few strong positive relationships . Second , consistent with studies that used DOA , we expected that microbial relationships would be more consistent across hosts than individualized ( see Figure 1A for a visualization of this prediction ) . This result would suggest that personalized microbiota—their compositions and dynamics—do not arise from host-specific microbiome ecologies ( Bashan et al . , 2016; Gao et al . , 2020; Vila et al . , 2020; San-Juan-Vergara et al . , 2018 ) . Third , we expected to observe positive correlations between taxa that are close phylogenetic relatives . This is because related bacteria may have similar functional properties and hence similar ecological relationships with other members of the community . They may also have dynamics that are driven by similar selective forces imposed by the host or host’s environment . Alternatively , competitive exclusion may lead closely related taxa to exhibit neutral or negative relationships . Fourth , because the environments experienced by baboons in Amboseli are far more uniform than those experienced by typical human study subjects ( Björk et al . , 2022; Grieneisen et al . , 2021 ) , we expected that the signature of ‘universality’ in baboons would be stronger than that observed in humans . We discuss the implications of these patterns for individual microbiome community assembly and dynamics , and for understanding how microbiome communities are structured across hosts—a key requirement for successful intervention to improve host health ( Bashan et al . , 2016; Widder et al . , 2016; Cullen et al . , 2020 ) .

## ## Results

We began by characterizing the overall sign , strength , and significance of pairwise correlations in bacterial abundance within each host . To do so , we applied the approach outlined in Figure 1A to stool-associated time series from 56 baboons ( Figure 1B ) and calculated Pearson’s correlations between pairs of bacterial taxa . To avoid biases created by zero inflation ( see Methods ) , we restricted our analysis to pairs where each member was present in at least 50% of samples across all hosts a

nd at least 20% of samples within each host ( Supplementary file 1a-1c ). Further , we required the joint zero abundance of a given bacterial pair to be less than 5% of all samples across hosts . After filtering , the resulting data set included ( 1 ) 1 878 pairs of centered log-ratio ( CLR ) -transformed amplicon sequence variants ( AS Vs; Figure 2A ) ; ( 2 ) 57 pairs of bacterial phyla ( Figure 2-figure supplement 1 ) ; and ( 3 ) 473 pairs of taxa agglomerated to the most granular possible family , or der , or class ( Figure 2-figure supplement 1 ) . To generate an expectation of the strength of bacterial correlations possible by chance , we used a permutation proced ure that randomly shuffled the taxonomic identities within each sample of the bacter ial count table 10 times for each of the 56 hosts ( 560 total permutations ) . We th en estimated correlations for these permuted pairs to generate an empirical null dis tribution of randomly generated taxon-taxon correlations . Observed correlations wer e judged against this reference ( Figure 2B ) . We also confirmed that the resulting correlation patterns were robust to several modeling choices and were not primarily driven by seasonal shifts in microbiome composition ( see results below ) . Consiste nt with the expectation that most bacterial correlations in the gut microbiome are w eak ( Coyte et al . , 2015; Coyte et al . , 2021 ) , only 19% of ASV-ASV correlation s in the heat map in Figure 2A were stronger than expected by chance (  $FDR \leq 0.05$ ; F igure 2-figure supplement 3; 19% of phylum-phylum; 22% of family/order/class correla tions; Figure 2-figure supplement 1; Figure 2-figure supplement 2 ) . The strongest negatively correlated pair in Figure 2A included an ASV in the family Kiritimatiella e and another in family Lachnospiraceae which had a median correlation of  $-0.520$  (  $\pm 0.132$  s . d . ) across all baboon hosts ( Figure 2C; ASV19 and ASV23; Supplementa ry file 1a and d ) . The strongest positively correlated pair of ASVs included two m embers of the genus Prevotella that had a median correlation of  $0.801$  (  $\pm 0.053$  s . d . ) across all baboons ( Figure 2D; ASV2 and ASV3; Supplementary file 1a and d ) . While these two ASVs were assigned to the same genus , their V4 16S DNA sequence i dentity was 97 . 6% , indicating they are probably not simply duplicate 16S gene cop ies encoded in the genome of a single species ( Vetrovský and Baldrian , 2013 Supple mentary file 1d ) . In support of the idea that strong , positive bacterial interdep endencies are rare ( Coyte et al . , 2015; Palmer and Foster , 2022; Coyte et al . , 2021 ) , only 3 . 8% of ASV pairs were significantly positively correlated , and the overall bacterial correlation patterns were slightly skewed toward negative relation ships . For instance , at the ASV level , the median correlation coefficient in Figu re 2A was  $-0.072$  , and 60% of these correlations were negative ( binomial test  $p < 0.0001$  ) . For family/order/class-level taxa , 58% of all correlations were negative ( Figure 2-figure supplements 1A and 3A; median family/order/class-level correlation =  $-0.049$ ; binomial test  $p < 0.0001$  ) . Correlations between phyla exhibited the st rongest negative skew , with 64% of phyla-phyla correlations having a negative sign ( Figure 2-figure supplements 1B and 3A; median phyla-level correlation =  $-0.100$ ; binomial test  $p < 0.0001$  ) . This bias toward negative relationships is consistent w ith the expectation that neutral or negative relationships between ASVs are more com mon than mutualisms ( Coyte et al . , 2015; Palmer and Foster , 2022; Coyte et al . , 2021 ) and that more distantly related taxa ( e . g . phyla ) respond to distinct environmental drivers due to differences in metabolic requirements and lifestyles . Next , we tested the degree to which within-host ASV-ASV correlations were consisten t across hosts . We began by plotting the absolute value of each ASV pair's median P earson's correlation coefficient as a function of the consistency of their correlati on sign ( positive or negative ) across the 56 hosts ( Figure 3A and B ) . These plo ts provide two main insights into the consistency of bacterial associations . First , in support of the idea that ASVs do not exhibit vastly different correlative relat ionships in different hosts , no taxon pairs were strongly and inconsistently correl ated across hosts ( Figure 3A and B; Figure 3-figure supplement 1A ) . Instead , the ASV pairs that had inconsistent correlation signs across hosts always had weak and o ften non-significant median absolute correlation coefficients within hosts ( Figure 3A and B ) . Second , the pairs with the most consistent sign agreement across hosts

also exhibited the largest median absolute correlation coefficients across hosts ( Figure 3A and B; Spearman's  $r=0.844$ ,  $p<0.0001$  ). Hence, pairs of ASVs that have the strongest relationships, and are therefore likely to play the most important roles in structuring gut microbiome dynamics, also tend to have the most consistent relationships in different hosts. Indeed, for the sets of positively or negatively correlated ASV pairs that showed universal agreement in the sign of their correlation across all hosts ( i.e. where  $x=1$  in Figure 3A and B ), the median absolute correlation coefficient is  $0.359$ , compared to  $0.116$  for those with no sign consistency (  $x=0.5$  in Figure 3A and B ). Note, that the correlation for a given pair of ASVs was only weakly predicted by bacterial abundance (  $r=0.129$  and  $r=0.223$  for the more and less abundant partner in a pair respectively;  $p<0.0001$  both ). While this effect was statistically significant, it explained only 6% of the variance in median correlation. Visual inspection of the patterns in Figures 2A, 3A and B indicate that ASV-ASV correlations are largely consistent across baboons, as opposed to individualized to each baboon. To explicitly quantify the relative strength of shared versus individualized signatures in the heat map in Figure 2A, we calculated the population mean pattern for the ASV-ASV correlation matrix,  $m$ . For each host, we then estimated the residual difference,  $e$ , between that individual's observed ASV-ASV correlation matrix,  $y$ , and the population mean matrix:  $y - m$  ( see Figure 3-figure supplement 2A for a cartoon example ). We reasoned that the observed correlation matrix for each host can be approximated by a mixture of contributions from the population mean matrix  $m$  and the host-specific residual matrix  $e$ . To identify the optimal mixture for each host ( i.e. the mixture of consistent vs. individualized correlation patterns that best explained the observed data ), we titrated the contribution ( i.e. weight ) of  $e$  from 0% to 100% ( and correspondingly, the contribution of  $m$  from 100% to 0% ) and identified the value that minimized the Frobenius distance between the simulated combination and the observed correlation matrix,  $y$ . In support of prior observations of 'universality' ( Bashan et al., 2016; Gao et al., 2020; Vila et al., 2020; San-Juan-Vergara et al., 2018 ), we found that, across hosts, the optimal mixture involved contributions from the shared correlation structure ( i.e.  $m$  ) of between 60% and 80% ( median 70% ) and a host-level contribution ( i.e. from  $e$  ) of between 40% and 20% ( median 30% ). Hence, population-level signatures contributed almost twice the weight as host-level signatures ( a median population:host ratio of 2.33:1; Figure 3-figure supplement 2B ). As a result, ASV-level relationships tend to be more consistent across hosts than host-specific. One advantage of our approach, compared to DOA ( Bashan et al., 2016 ), is that we can identify the bacterial pairs that exhibit the most consistent relationships across hosts. Hence, we next conducted several analyses to understand why some taxon pairs are more consistent than others. To do so, we created a 'universality' score ( Figure 2-figure supplement 4 ) that could be calculated for each ASV pair. The score multiplies the pair's median absolute correlation coefficient across hosts ( y-axis of Figure 3A and B ) with its correlation consistency across hosts ( i.e. proportion of shared sign; x-axis of Figure 3A and B ). The resulting scores range from 0 to 1, where a score of 1 equates to perfect 'universality' ( i.e. all hosts have a correlation coefficient of 1 or all hosts have a correlation coefficient of -1 ). Applying this score to all pairs of ASVs reveals a right-skewed distribution, reflecting the fact that most bacterial correlations are weak, with inconsistent sign directions across hosts ( Figure 3C; Figure 2-figure supplement 3B ). However, 48% of these scores were higher than expected by chance ( permutation test;  $FDR\leq 0.05$ ; Figure 3C; Figure 2-figure supplement 3B ), reflecting a signal of universality in our data. Despite the bias toward negative ASV-ASV correlations in the overall set of bacterial correlations ( Figure 2-figure supplement 3A ), we observed no such bias in the most universal pairs. For instance, in the top 5% most universal ASV pairs (  $n=94$  pairs ), 46 pairs exhibited net positive correlations and 48 pairs, net negative correlations, suggesting no particular bias in the direction of the strongest and most consistent associations. To visualize these highly cons

istent correlations , we plotted bacterial co-abundance networks connecting the top 5% most universal ASV pairs ( Figure 3C ) . A handful of ASVs were highly connected within this network , especially ASV1 ( genus *Bifidobacterium*; Supplementary file 1a and d ) , which was connected to 14 other ASVs . Three other ASVs were connected to at least 10 other ASVs , including members of families *Atopobiaceae* ( ASV8 ) , *Eggerthellaceae* ( ASV28 ) , and *Muribaculaceae* ( ASV86 ) ( Supplementary file 1e ) . These families were also enriched in this network , relative to the rest of the data , as was the family *Bifidobacteriaceae* ( Figure 3E ) . Pairings between members of the same family were enriched by >3-fold in this network (  $p < 0.0001$  ) , making up 32% of pairs in the most universal set versus only 9.8% of pairs outside that set . Almost two-thirds of these were *Prevotellaceae*-*Prevotellaceae* pairs ( 10 of 16 same-family pairs ) . We next asked: does the phylogenetic distance between a pair predict the nature of their relationship ? In support of the idea that homology leads closely related ASVs to respond similarly to the environment , or perhaps facilitate each other's growth ( Meehan and Beiko , 2014; Vacca et al . , 2020 ) , we found that , for positively associated ASV pairs , closely related taxa had higher universality scores than more distantly related taxa ( Pearson's  $r$  for positively correlated pairs =  $-0.232$ ;  $p < 0.0001$ ; Figure 4A ) . In contrast , when ASV pairs were negatively correlated , there was a weak positive relationship between phylogenetic distance and universality ( Pearson's  $r = 0.106$ ;  $p = 0.004$ ; Figure 4B ) . In other words , the strongest and most consistently negatively correlated taxa tend to be distantly related , whereas the strongest and most consistently positively correlated taxa were often closely related , especially members of the family *Atopobiaceae* ( Supplementary file 1f ) . We next asked whether host attributes , including host sex , social group membership , genetic relatedness , age , or baseline gut microbiome composition , predict host differences in patterns of bacterial correlation . Consistent with studies that use DOA to infer universality ( Bashan et al . , 2016; Gao et al . , 2020; Vila et al . , 2020; San-Juan-Vergara et al . , 2018; Kalyuzhny et al . , 2017 ) , the strongest predictor of distance in bacterial correlation patterns was distance in terms of baseline microbiome composition ( a core assumption of DOA ) . Indeed , a Mantel test correlating compositional distance of average microbial profiles ( as Aitchison distances between the per-host mean of CLR-transformed samples ) with distance in microbial correlation patterns between hosts revealed that 34% of the variation in correlation patterns was explained by baseline microbiome community composition ( Mantel:  $r^2 = 0.336$ ;  $p = 0.001$ ; Figure 5A; Supplementary file 1g ) . Consistent with prior research in our population , which finds widespread heritability for gut microbiome taxon abundances ( Grieneisen et al . , 2021 ) , we also found a weak but significant relationship between host genetic distance and the distance in microbial correlation patterns between hosts , after controlling for similarity of baseline composition across hosts . Hosts who were more closely related based on a multigenerational pedigree have slightly more similar ASV-level correlation matrices ( Figure 5B; Supplementary file 1g;  $r^2 = 0.009$ ; partial Mantel controlling for baseline similarity:  $p$ -value =  $0.002$  ) . We found no evidence that members of the same social group or sex exhibit more similar microbial correlation patterns ( social group:  $F = 2.146$ ;  $p = 0.089$ ; sex:  $F = 2.026$ ;  $p = 0.160$ ; Supplementary file 1g ) . Host age may also predict the overall strength of microbial relationships , and some studies find that gut microbial compositions become more individualized with age ( Risely et al . , 2022; Wilmanski et al . , 2021 ) . This observation suggests that host age may also be linked to individualized microbial relationships . To test these ideas , we divided our hosts into three classes: juveniles ( 0–6 years ) , prime age adults ( 6–13 years ) , and older adults ( 13+ years ) and compared ASV correlation patterns between ( 1 ) the juvenile and prime age class and ( 2 ) between prime age and older adults . Hosts were only included in these analyses if they had >35 samples in either the juvenile and prime age class or prime age and older adult class ( no host had >35 samples in all three age classes; 13 hosts were included in the juvenile and matched prime age adult groups; a separate set of 13 hosts were included in the older adult and



matched prime age adult groups ). We found no evidence that microbial correlations get stronger or weaker with age ( Figure 5—figure supplement 1 ). Further , using the same methods described above to estimate the relative host- and population-level contributions to ASV correlation patterns , we found no strong differences in the degree of ‘personalized’ correlation patterns across age groups ( Figure 5—figure supplement 2 ). Further , ASV correlation patterns were slightly more similar within versus between adjacent age classes ( juvenile vs . prime age , ANOVA  $p=0.00175$  , 1 . 3% variance explained; prime age vs . older adult , ANOVA  $p=0.0112$  , 0 . 9% variance explained ). Principal components analysis on the microbial correlation patterns between juvenile and prime age hosts , and between prime age and older adult hosts revealed some age effects , particularly on the second principal component ( Figure 5—figure supplement 3 ). See the legend of Figure 5—figure supplement 3 and Supplementary file 1i for information on which ASV pairs differed across age categories . Without experiments , we cannot disentangle whether our observed bacterial correlations are due to ecological interactions between bacterial species or to shared responses to environmental gradients , either inside or outside the host . While we were unable to control for many aspects of the host environment ( e . g . gut pH , hormones , or immune profiles ) , we were able to include measures of dietary variation in our models of microbial abundances . Seasonal changes in host diet therefore do not account for universality in microbial relationships across hosts . To account for additional unexplained seasonal variation , we next removed the oscillating seasonal trend from the log-ratio abundances for each ASV ( modeled as a sine wave ) and re-estimated the ASV-ASV correlation matrix ( Figure 4—figure supplement 1 ). Removing the seasonal trend had little effect on ASV-ASV correlations , as the variance explained by seasonal oscillation was small for all ASVs ( median 1 . 1% , minimum = 0% , maximum = 6% ) . Consequently , the between-ASV correlation estimates were almost identical to those derived from our original model ( Pearson’s  $r=0.982$  ,  $p<0.0001$ ; Figure 4—figure supplement 1C ). Further , ASV pairs where one or more members was from a family that showed strong seasonal changes in a prior analysis of these data ( Björk et al . , 2022 ) , henceforth ‘seasonal’ families , had only slightly higher universality scores than taxon pairs where neither partner showed strong seasonal changes in abundance ( difference of 0 . 018;  $p<0.0001$ ; Figure 4—figure supplement 2 ). Because the high level of universality we observed was not well explained by season , we also tested whether universality was explained by synchronized dynamics . We reasoned that if one member of an ASV pair shows highly synchronized dynamics across different hosts , and the other member is also strongly synchronized across hosts , then universality could be an inevitable outcome of strong , but independent synchrony in both members of the pair . We quantified synchrony as the degree to which the observed dynamics of a single , focal ASV are consistent across hosts , such that high synchrony ( near 1 ) implies that the timing and direction of shifts in log-ratio ASV abundance are identical across hosts in the population ( see Methods; Figure 4—figure supplement 3 ). Estimates of synchrony ranged from 0 . 033 to 0 . 474 ( median=0 . 187 ) . Interestingly , ASVs in the 13 ‘seasonal’ families are not more likely to have high synchrony than other families ( ANOVA ,  $p=0.434$ ; Figure 4—figure supplement 4; Supplementary file 1h ). The average synchrony of an ASV-ASV pair had a statistically significant but weak relationship with that pair’s universality score (  $r=0.116$  ,  $p<0.0001$ ; Figure 4—figure supplement 5 ). Finally , to investigate parallels between baboon and human microbial communities , we turned to two publicly available gut microbial time series data sets: daily samples from 34 human adults over a 17-day span ( 483 total samples; hereafter Johnson et al . , 2019 ) , and the DIABIMMUNE cohort that consists of 285 samples , collected monthly over 3 years , from 15 infants and toddlers living in Russian Karelia ( Vatanen et al . , 2016; at the time of writing , these cohorts were the only publicly available data sets we could find that included large numbers of repeated samples from the same subjects ) . Because baboons in Amboseli experience less heterogeneity in their environments and diets than humans ( Björk et al . , 2022; Grieneisen et al . , 2021 ) , we expected th

ey would exhibit greater consistency in microbial correlations than either human cohort. Here, we compared each host cohort's universality at the level of correlations between families, orders, and classes because these taxonomic levels offered the greatest comparative power (10% of families/orders/classes overlap between the cohorts compared to just 3% of genera and no ASVs). Contrary to our expectations, we find comparable evidence of universality in baboons and the DIABIMMUNE infant/toddler cohort, but weaker evidence for universality in Johnson et al. (Figure 6A-D). Bacterial families in the DIABIMMUNE cohort yielded universality scores slightly higher than those observed in Amboseli (25th percentile=0.142, median=0.216, 75th percentile=0.321 for DIABIMMUNE; 25th percentile=0.088, median=0.150, 75th percentile=0.234 for Amboseli), driven by correlations between families that were stronger on average than those estimated for baboons (median DIABIMMUNE family-family correlation strength=0.253; median Amboseli family-family correlation strength=0.170). The high level of consistency between both human infants/toddlers and wild baboons is surprising and may be due to the similar sampling intervals for these cohorts. Both cohorts were sampled approximately monthly, while Johnson et al.'s subjects were sampled daily (Coyte et al., 2021; Guittar et al., 2019). Median correlation strengths and universality scores for the Johnson et al., 2019 cohort were substantially lower (median correlation=0.090; 25th percentile universality=0.050, median=0.086, 75th percentile=0.113) than the DIABIMMUNE cohort or the baboons. Despite considerable differences in the hosts, time scales, and designs of these studies, all three data sets exhibited a positive correlation between correlation strength and sign consistency for family pairs (Figure 6D). This trend was strongest in the Amboseli baboons (exponent of power regression  $b=1.72$ ;  $p<0.0001$ ); weaker in the DIABIMMUNE cohort ( $b=1.51$ ;  $p<0.0001$ ) and weakest in Johnson et al., 2019 ( $b=1.19$ ;  $p<0.0001$ ). Further, the most universal family-family associations skewed positive in both the baboons and the infant data set. All of the top 5% most universal family pairs (30 of 30 pairs) are positively associated in the DIABIMMUNE cohort, compared to 70% (23 of 33 pairs) in the Amboseli baboons. Finally, we examined the relationship between universality scores for family pairs that overlapped between Amboseli and DIABIMMUNE ( $n=29$  pairs), and between Amboseli and Johnson et al., 2019 (Figure 6E;  $n=21$  pairs; only 10 family pairs overlapped between all three data sets). For these overlapping pairs, scores in the Amboseli data predicted scores for the same family-family pair in the DIABIMMUNE data set ( $r=0.562$ ,  $p=0.001$ ). The association between scores in the Amboseli data and the Johnson et al. data was negative, but not statistically significant ( $r=-0.402$ ,  $p=0.071$ ).

## ## Discussion

Do different hosts have different microbiome 'ecologies'? Ecological and evolutionary processes like horizontal gene flow, genotype by environment interactions, and priority effects have been predicted to lead bacterial species to occupy different niches (with different ecological interactions) in different communities (Dolinšek et al., 2016; Franzosa et al., 2015; Faith et al., 2013; Bik et al., 2016; Caporaso et al., 2011; Costello et al., 2009; Louca et al., 2018; Rainey and Quistad, 2020; Martiny et al., 2015). Yet contrary to these expectations, here we find that hosts in the same population exhibit pairwise bacterial correlation patterns that are predominantly shared across hosts, rather than idiosyncratic to individual hosts. If these shared correlation patterns arise from shared microbiome ecologies, this discovery has consequences for understanding the basic eco-evolutionary drivers of microbiome dynamics and for human and animal health. For instance, shared ecologies would mean that designing widely applicable microbiome interventions is a more attainable goal than personalized microbiome compositions would suggest. Shared microbiome ecologies may also enable researchers to develop microbiome interventions that leverage these interactions to manipulate the microbiome's emergent community properties to improve host health. By measuring bacterial correlations in multi

ple hosts , we were also able , for the first time , to pinpoint which pairs of bacterial taxa exhibit the most consistent relationships across hosts . We found that most bacterial abundance correlations—from ASV-ASV to phyla-phyla relationships—were weak and negative . Positive bacterial interactions have been the subject of recent discussion in the literature ( Loftus et al . , 2021; Palmer and Foster , 2022; Kehe et al . , 2021 ) . Ecological theory predicts that strong positive interactions should be rare in natural communities because species interdependencies can hamper community assembly and stability ( Coyte et al . , 2015; Coyte et al . , 2021 ) . This theory is supported by experiments that directly measure the effects of one bacterial species on another’s growth ( Weiss et al . , 2022; Ortiz et al . , 2021; Carlström et al . , 2019; Venturelli et al . , 2018 ) ( but see Kehe et al . , 2021 ) . Our results suggest that strong , positive bacterial correlations are indeed uncommon in intact , unmanipulated microbiomes: significant positive relationships made up just 3 . 8% of all the pairwise correlations we observed . Hence , strong mutualisms , while key to microbiome function and dynamics , are probably rare in gut communities . While mutualisms and universal dynamics are important , the correlation patterns we observe likely arise from a combination of ecological interactions between bacteria and shared responses to the environment ( i . e . pairs of bacteria that prefer the same or different environments ) . In support of the idea that at least some of the correlations we observed are due to between-species interactions , our signature of universality was essentially unchanged after accounting for some of the strongest known drivers of microbiome composition and change in our population—host diet and season ( Björk et al . , 2022; Grieneisen et al . , 2021 ) —as well as microbial synchrony between hosts . However , our approach did not account for important environmental gradients within the gut , such as host immune profiles and intestinal pH . These factors also shape microbiome composition ( e . g . Reese et al . , 2018; Firrman et al . , 2022; de Vos et al . , 2022 ) , and can lead to shared abundance correlations between hosts even if hosts themselves differ . Ecological selection via within-host environments may explain our finding that genetic relatives share somewhat similar bacterial correlation patterns . Ecological selection is also consistent with our observation that the most consistent ASV-level correlations are between phylogenetically related taxa , and these patterns were strongest for positively associated taxon pairs . In support , phylogenetically related species have been shown to have similar environmental preferences ( Tamames et al . , 2016 ) . We note that none of the correlations we observed can be mapped directly to standard categories of pairwise ecological interactions , such as mutualism , commensalism , amensalism , exploitation , or competition . Experimental approaches that directly measure the effects of one species on another’s growth in vitro are better suited to characterizing these relationships ( Kehe et al . , 2021; Weiss et al . , 2022; Ortiz et al . , 2021; Carlström et al . , 2019; Venturelli et al . , 2018 ) . The strong signal of universality we observed in bacterial abundance correlations stands in contrast to the common observation that microbiome taxonomic composition ( i . e . the presence and abundance of bacterial species ) is almost always highly personalized ( Franzosa et al . , 2015; Faith et al . , 2013; Bik et al . , 2016; Caporaso et al . , 2011; Costello et al . , 2009; Risely et al . , 2021; Kolodny et al . , 2019; Flores et al . , 2014; Johnson et al . , 2019; Pruss et al . , 2021 ) . Our own prior analyses of these data found that each baboon exhibited personalized microbiome compositions and asynchronous single-taxon dynamics ( Björk et al . , 2022 ) . These contrasting patterns—personalized compositions but shared abundance correlations—are important because personalized microbiota have been proposed to arise , at least in part , from personalized microbiome ecologies ( Franzosa et al . , 2015; Faith et al . , 2013; Bik et al . , 2016; Caporaso et al . , 2011; Costello et al . , 2009; Risely et al . , 2021; Kolodny et al . , 2019; Flores et al . , 2014; Johnson et al . , 2019; Pruss et al . , 2021 ) . We can think of at least three explanations that reconcile these observations . First , consistent with ideas discussed above , if environments in the gut shape bacterial abundances but these environments are not synchronized across hosts , this would

lead to shared abundance correlations over time , but individualized microbiome compositions at any single point in time . Second , the effects of horizontal gene transfer and gene by environment interactions on microbial phenotypes may not be strong enough to substantially alter pairwise microbial associations in the gut . This may be especially true for our main unit of analysis , bacterial ASVs . Because ASVs encompass multiple species and strains , each with somewhat different functional capacities , their dynamics may be buffered against idiosyncrasies driven by horizontal gene transfer and functional redundancy , which affect single strains more strongly than whole species or genera . We would strain-level correlation patterns to be more individualized than those between ASVs . Third , personalized gut microbial compositions may emerge from at least two other phenomena: personalized assembly processes and interactions driven by rare , host-specific strains ( which were necessarily excluded from our analyses ) ( Costello et al . , 2012; Walter and Ley , 2011 ) . In general , a logical next step would be to confirm the microbial correlation patterns we observed using culture-based approaches , which will help reveal ( in vitro ) whether they can be attributed to direct effects of one microbe on another's growth . The observation that bacterial correlation patterns are largely shared across hosts was also apparent in one human data set , despite between-study differences in study design , host age , and time scale . Specifically , both the Amboseli baboons and the DIABIMMUNE infant/toddler cohort from Russia ( Vatanen et al . , 2016 ) exhibit comparable levels of universality of correlation patterns . This outcome surprised us: because the baboons all live in the same environment and are presumably colonized by similar bacterial strains from that environment , we expected that ecological selection and shared strain functionality should lead to stronger universality in bacterial correlation patterns compared to human infants sampled from different households and who were probably colonized by different strains . We also found that the most universal correlations between bacterial families in baboons tended to be highly universal in human infants/toddlers . Hence , some bacterial families may exhibit consistent microbial relationships within hosts , across host populations , and across host species . Finally , a recent , independent study also identified consistent bacterial correlation patterns across four different populations of human hosts ( Loftus et al . , 2021 ) . While this study lacked resolution at the level of individual hosts , it did identify a conserved network of positively associated and closely related microbes similar to those we identify in Figure 3 . The authors speculate that these conserved associations may indicate strong partner fidelity or obligate partnerships . We did , however , fail to detect universality in a second human data set reported in Johnson et al . , 2019 , in which subjects were sampled daily , rather than weekly or monthly . The lack of universality in Johnson et al . , 2019 , may be due to this difference in sampling time scale , especially if daily abundances and correlations are noisier than covariances modeled over the longer time scales in our study . In support , many fewer of the microbial correlations were stronger than random chance in Johnson et al . as compared to the baboons or children in the DIABIMMUNE cohort . However , without the ability to subsample Johnson et al . , 2019 , to monthly scales ( this data set is only 17 days long ) , it is impossible to test this prediction . The subjects in Johnson et al . , 2019 , also consumed substantially different diets from each other , perhaps more so than the children in the DIABIMMUNE cohort , and this inter-host difference in diet may reduce the universality of microbial correlations . In sum , our study indicates that microbiome personalization may not extend to microbiome community ecology . However , more work is needed to understand how relationships between microbiome taxa are explained by shared internal and external environments , direct and indirect ecological interactions , taxonomic levels ( e . g . strains to phyla ) and time scales ( days to months and years ) . Future studies should also consider how pairwise bacterial interactions scale up to affect the emergent properties of the community ( Levine et al . , 2017; Letten and Stouffer , 2019; Friedman et al . , 2017 ) . We hope that our longitudinal data set and the new methods we developed as part of this study ( e . g . the model of log-ratio dynamics , the a

assessment of covariation from time-ordered abundance trajectories , and the universality score ) will be useful in this enterprise .

## ## Methods

The baboon hosts in this study were members of the Amboseli baboon population , which has been studied by the Amboseli Baboon Research Project since 1971 ( Alberts and Altmann , 2012 ) . The microbiome compositional profiles are derived from PCR amplification of an ~390-bp-long fragment that encompassed the V4 region of the 16S rRNA gene using primers 515F-806R . These microbiome profiles were previously analyzed in Björk et al . , 2022 and Grieneisen et al . , 2021 . Our analyses use 5534 of these profiles from 56 especially well-sampled baboons , collected over a 13 . 3-year span between 2000 and 2013 ( Figure 1B ) . Each baboon host in this data set was sampled at least 75 times ( mean number of samples=99; range=75-181 samples; median number of days between samples within hosts=20 days; 25th percentile=7 days , 75th percentile=49 days ) . DNA was extracted from each sample using the MoBio and QIAGEN PowerSoil kit with a bead-beating step . All samples were sequenced on an Illumina HiSeq 2500 , with a median read count of 48 , 827 reads per sample across all 5534 samples ( range=982-459 , 315 reads per sample ) . Following recommended statistical practices ( Gloor et al . , 2017 ) , samples were not rarefied , but counts were agglomerated and transformed to additive log-ratios ( ALR ) . Variation in sampling depth and relative abundance were modeled by the method described in the subsequent section . Further details of sample collection , DNA extraction , and sequencing can be found in ( Björk et al . , 2022; Grieneisen et al . , 2021 ) . Data sets of per-sample taxonomic counts were produced at each of three taxonomic levels , from finest to coarsest: ASV , taxonomic assignments finer than phyla , but above the genus level ( e . g . class , order , family ) , and phylum . At the intermediate and coarsest levels , taxa were agglomerated using phyloseq's tax\_glom ( ) function ( McMurdie and Holmes , 2013 ) such that all sequence variants sharing taxonomic identity at that level were collapsed into a single taxon ( e . g . family Bifidobacteraceae ) . To reduce sparsity in the data set , remove 16S sequences that could represent gene duplications , and focus only on taxa that were prevalent in all 56 hosts , we further filtered as follows: ( 1 ) in each of the three taxonomically defined data sets ( i . e . ASV , taxa assigned to family/order/class , and phylum ) , we identified taxa present ( i . e . having non-zero abundance ) in at least 20% of samples from each host; ( 2 ) if a given ASV was >99% genetically similar to another ASV we removed the least abundant of the pair to minimize the risk of including duplicate 16S rRNA gene copies from the same taxa ( Vetrovský and Baldrian , 2013 ) ; and ( 3 ) counts associated with all other taxa were combined into a dummy category , hereafter referred to as 'other . ' The 'other' category therefore includes a combination of rare and host-specific gut microbes . This category was retained in the data set ( although not analyzed directly ) because 'other' counts still inform the precision of the observed relative abundances in our model . See the sub-section titled 'Filtering out taxon pairs with frequent joint absence' below for further filtering that was required to avoid biases in estimating correlations between taxon pairs . Characteristics of the filtered data at each taxonomic level are provided in Supplementary file 1a-1c . At the ASV level , these filtering steps eliminate a majority of ASVs and ASV pairs from consideration: from 22 , 097 unique ASVs before filtering to 107 after filtering . This filtering also retained 12 phyla and 35 taxa at the class/order/family level ( Supplementary file 1a-1c ) . Our modeling approach is similar to several published methods for modeling microbial time series data . There are three key features from our perspective: the use of log-ratios ( discussed above ) , the use of a state space model , and the Gaussian process component . State space models are useful for modeling a dynamic process that is observed only after the introduction of some measurement ( e . g . Joseph et al . , 2020 ) . The Gaussian process component helps contend with irregularity in the sampling of our data . Rather than evolving in discrete jumps from one time point to the next , it allowed us to model the change in microbial log-rati

o abundances as smoothly flowing through interruptions in observation . Other authors have made similar choices ( Åijö et al . , 2018 ) . Specifically , estimates of taxon-taxon covariance were obtained from the *basset* model of the ‘fido’ package in R ( Silverman et al . , 2022 ) . Data for each host took the form of a  $D \times N$  count matrix , where  $D$  gives the number of taxa and  $N$  the number of samples collected for a given host . The following model was fit to each host’s count matrix (  $Y$  ) , where  $Y_i$  represents the counts associated with a single sample:  $Y_i \sim \text{Multinomial}(\pi_i)$   $\pi_i = \text{ALR-1}(\eta_i)$   $\eta \sim \text{Normal}(\Lambda, \Sigma, I)$   $\Lambda \sim \text{GP}(\theta[X], \Sigma, \Gamma[X])$   $\Sigma \sim \text{inv-Wishart}(\Xi, \nu)$  The observed relative abundances are considered to be drawn from a multinomial distribution parameterized by a set of proportions (  $\pi$  ) which have an analogous representation in the ALR . The dynamics of these log-ratio abundances (  $\eta$  ) are described by what amounts to a state space model in the third and fourth lines of the specification above , where a Gaussian process models the evolution of a ‘latent’ state . The matrix  $\Sigma$  captures covariation in log-ratio abundances ( the  $D$  rows of the observed count matrix ) . Sample-sample covariation arising from nearness in time ( autocorrelation ) is modeled by the kernel matrix  $\Gamma$  . Both the kernel matrix and the expected baseline log-ratio abundances (  $\theta$  ) are parameterized by a set of time-varying covariates  $X$  which included the day of sampling ( where the date of first sample is defined as zero ) and the first three principal components of diet composition , calculated following Björk et al . , 2022; Grieneisen et al . , 2021 , as the diet of all juveniles and females living in the host’s social group in the 30 days prior to sample collection . All group members consume highly similar diets as they travel together across the habitat , encountering the same resources at the same time ( Björk et al . , 2022; Grieneisen et al . , 2021 ) . These data are collected via random-order behavioral observations collected two to four times per week on adult females and juveniles in each social group . The kernel matrix  $\Gamma$  was composed of two component squared exponential kernels . The first , intended to manage sample-sample autocorrelation , was selected to have a bandwidth such that this autocorrelation decayed to a minimum at 90 days . This mirrored the behavior of estimates of sample-sample autocorrelation in the raw data . The second component kernel modeled sample-sample covariance driven by similarity in composition of diet . The relative weight of these effects—autocorrelation and diet—on sample-sample variation was set at 3:1 . We fit four alternative versions of our models in order to test the sensitivity of these parameter settings , varying the bandwidth of the squared exponential kernel in such a way as to give minimum sample-sample autocorrelation at either 30 or 90 days . We also varied the proportion of sample-sample covariance driven by diet from 0% to 25% to 50% , and we varied the log scale of total sample-sample variance between 1 and 2 . In all cases , estimates of correlation between CLR ASVs were similar , with minimum and maximum  $r^2$  of estimates between ‘canonical’ and alternatively parameterized models of 0 . 993 and 0 . 996 , respectively . This suggests our findings are reasonably robust to a range of hyperparameter settings . Posterior inference on this model is performed as described in Silverman et al . , 2022 , and yields estimates of the distributions of parameters necessary to reconstruct trajectories for all log-ratio taxa across sampling time . In particular , we extract the posterior estimates of one such parameter ,  $\Sigma$  , the covariance of ALR taxa , from the fitted models for each host . We convert these covariance matrices over ALR taxa to the CLR form ( a simple linear transformation of the matrix ) . We then normalize estimated CLR covariance matrices to Pearson’s correlation matrices in R using the built-in `cov2cor()` function . The ASV-level relative abundance data were sparse , even after filtering low abundance taxa ( i . e . those present in <20% of samples in each host ) . Zeros comprised 29 . 7% of the ASV-level count matrix , 16 . 1% of the class/order/family count matrix , and 9 . 8% of the phylum level count matrix . This abundance of ‘missing’ observations at the ASV level led to a bias in estimates of taxon-taxon correlation: a high-frequency joint zero abundances for a pair of taxa increased the apparent positive correlation of those taxa ( Figure 1—figure supplement 3 ) . This is because abundances below some minimum level of sensitivity in sampling will be ‘flattened’ to zero , reducing the ob-

served variation for those pairs , over those samples , to zero . This loss of variation leads to a tendency to overestimate the ( positive ) correlation of these pairs . This trend was observed for our basset model ( Silverman et al . , 2022 ) when estimating either Pearson’s correlation or proportionality ( Quinn et al . , 2017 ) . It was also observed in ASV-ASV CLR correlations output by COAT ( Cao et al . , 2019 ) , which estimates CLR correlation through a sparsity-inducing procedure intended to yield more conservative estimates , and by SparCC ( Friedman and Alm , 2012 , see Figure 1–figure supplement 3 ) . To avoid this bias , we restricted our analyses to taxon pairs with strictly less than a 5% frequency of joint absence ( i . e . joint zero abundance observations in less than 5% of all samples across hosts ) and less than a 50% frequency of absence in either taxon individually across all samples . We illustrate these filtering criteria with a cartoon example involving four taxa , jointly sampled 20 times , indicating presence as an ‘X’ and absence ( zero abundance ) with a dash: Taxon A ( 70% present ) : X---XXXXX-XXX-X-XXX Taxon B ( 20% present ) : X-----X--X---X- Taxon C ( 50% present ) : -XXX-----XX-XXXX--X- Taxon D ( 65% present ) : XX-----XX-XX-XX-XXXX In this example , taxon B would be excluded from analyses by the requirement that taxa be ‘present’ or non-zero in at least half of all samples . Its associations with other taxa ( A x B , B x C , or B x D ) would be omitted from analyses , leaving pairs A x C , A x D , and C x D . Of these , the rate of joint absence is 5% , 10% , and 15% , respectively , meaning only A x C would pass the filtering criterion of no more than 5% joint absence . After filtering , 1878 of the original 7750 ASV-ASV pairs remained in this ‘high-confidence’ set . It is this set that we present in all figures and results . This procedure was also carried out at the phylum and class/order/family levels . However , as the frequency of absence was generally small at these higher taxonomic levels , a higher proportion of possible taxon-taxon pairs were included in the high-confidence set: 86 . 4% ( 57 of 66 pairs ) at the phylum level and 71 . 0% ( 473 of 666 pairs ) at the class/order/family level . We devised a universality score for each pair of taxa intended to capture the strength and consistency of taxon-taxon correlations across hosts ( Figure 2–figure supplement 4 ) . The majority direction is negative otherwise . This score identifies the sign of the taxon-taxon correlation ( positive or negative ) that is most common across the 56 hosts ( i . e . occurs in >50% of the 56 hosts in the data set ) . The direction of this sign is the ‘majority correlation sign . ’ For a pair of taxa  $i$  , let  $n_{i,maj}$  be the number of hosts with CLR correlation over pair  $i$  with the majority correlation sign for that pair and let  $n$  be the total number of hosts . Let  $R$  be the subset of estimated CLR correlations for pair  $i$  across hosts with the majority sign . The universality score  $u_i$  for that taxon-taxon pair is then given by  $u_i = n_{i,maj} \times \text{median} ( R )$  . This score is the product of the median CLR correlation across hosts and the proportion of hosts with the majority correlation sign , and is bounded between 0 and 1 . Scores near 1 indicate strong universality and near-zero scores indicate weak universality . Strong universality can only be achieved by taxon-taxon correlations that are both large in magnitude and highly concordant across hosts . We identified correlations stronger than expected by chance using permutations of the data set to define empirical null distributions ( Figure 2–figure supplement 3A ) . Specifically , we permuted the microbial count tables by randomly shuffling taxon identity within each sample 10 times for each of the 56 hosts . This procedure maintained relative abundance patterns within a sample but scrambled the covariance patterns of relative abundances . These randomly generated correlations were pooled into a single reference distribution . The distributions of ASV-level CLR correlations in the original and permuted data are shown in Figure 2–figure supplement 3A . We identified ‘significant’ correlations as those below  $\text{FDR} \leq 0 . 05$  ( Benjamini-Hochberg ) , testing against the permuted data . We applied an analogous permutation test to derive a null distribution for taxon-taxon universality scores . In a single iteration of this permutation procedure , rows and columns of the observed taxon-taxon correlation matrix for each host were shuffled , maintaining the distribution over observed correlations at the host level but randomizing the identity of taxon pairs across hosts .

This procedure was repeated 100 times and universality scores were calculated from each of these shuffled data sets to give a single pseudo-null distribution of universality scores. The observed and null distributions of universality scores at the ASV level are shown in Figure 2—figure supplement 3B. We used this empirical null distribution to identify universality scores significantly greater than expected ( $FDR \leq 0.05$ ). We used simulations to estimate the degree of shared ‘signal’ between hosts in terms of taxon-taxon correlations. Each host’s ‘observed correlations’ were defined as the basset estimated maximum a posteriori estimates of CLR ASV correlations for that host. We computed the mean correlations across the population using the function `estcov()` from the `shapes` package in R (Chen, 2020) and estimated a host-specific contribution to the observed correlations as the residual difference between per-host observed and these mean correlations. That is,  $\text{observed host correlations} = \text{mean population correlations} + \text{host residual}$ . For each host, we simulated a hypothetical set of composite taxon-taxon correlations as a convex combination of mean and host residual:  $\text{composite correlations} = (1 - \alpha) \times \text{mean population correlations} + \alpha \times \text{host residual}$ . A cartoon example of this procedure is given in Figure 3—figure supplement 2. For example, one such simulated set of taxon-taxon correlations might constitute a mixture of 90% host contribution and 10% shared population-level ‘signal’ ( $\alpha = 0.9$ ). Alternatively, a small host-level contribution might have  $\alpha = 0.1$ . For each host, we iterated over increasing proportions of host-level contribution (from 0% to 100%), generating simulated composite correlation matrices according to the formula above. We compared these simulated patterns to those observed for the same host, reasoning that simulated correlation matrices that minimize the distance between the observed correlation matrices and the simulated mixtures provide the best description of the underlying true mixture. Seasonally de-trended data was obtained in the following way. The observed ASV count matrices were CLR-transformed and linear autoregressive models were fit to each CLR-transformed ASV’s series. In these models, wet-dry season oscillation was modeled as a sine wave with a period of 365 days. The magnitude of this component was estimated during model fitting after an offset (in weeks) was estimated in a first step, in order to best align the oscillating seasonal component with the data. Per-ASV models were fit using the following syntax: `arima(x=x, xreg=model.matrix(~factor(host)+sinf(offset,days))[, -1], order=c(1, 0, 0))` where `x` gives CLR counts for that ASV, the ‘order’ argument of `arima` enables a single autoregressive component, and the ‘xreg’ argument specifies a covariance matrix. That covariance matrix contains a per-host label giving host-specific offsets for log-ratio abundance and an oscillating seasonal trend through ‘`sinf`’, a function that samples values corresponding to day indices (through ‘`days`’) from a sine wave with weekly offset (‘`offset`’) and a period of 365 days. The residuals from these per-ASV model fits were extracted and used as the seasonally ‘de-trended’ data (see Figure 4—figure supplement 1). Correlations across CLR ASV-ASV pairs were estimated from these residual series with the `cov()` function in R. ‘Synchrony’ was estimated by sampling aligned microbiome compositional profiles across hosts. We identified all samples collected from pairs of hosts within 1 calendar day. For instance, a sample collected from host F01 on March 14, 2011, could pair with a sample from M04 on March 15, 2011. For all possible pairs of hosts, we selected one such aligned pair of samples, yielding 1540 joint observations of gut microbiome composition. For each such paired sample, one host was arbitrarily designated as host A and the other as host B. The ‘synchrony’ of a given taxon was estimated as the correlation of a taxon’s model-inferred log-ratio abundance across the set of samples from hosts labeled A and the set of samples from hosts labeled B. The cartoon in Figure 4—figure supplement 3 illustrates this sample pairing. We performed enrichment analyses for bacterial families and family pairs in several settings. In each case we computed the frequency of ASVs belonging to a given family, or of pairs belonging to a family pair, on a subset of the data. These were compared to the overall frequencies of ASVs belonging to those families or pairs. To determine the enrichment of families and family pairs in the most universal ASV pairs (Figure



3E; Supplementary file 1e ) , we calculated the frequencies of ASV families and pairs in the top 5% of pairs by universality scores . Significant enrichment of families or pairs was identified using a one-sided Fisher's exact test . Multiple test correction was applied as a Benjamini-Hochberg adjustment to observed p-values . Phylogenetic distances between ASV sequences were calculated with the `dist.ml` function in the 'phangorn' package in R ( Schliep , 2011 ) using default settings for amino acid substitution rates . In Supplementary file 1f , low phylogenetic distance/high median correlation strength pairs were identified as those with phylogenetic distances of less than 0 . 2 and median correlation strengths of greater than 0 . 5 . Again , significance of these was evaluated against overall frequencies of the same families and pairs . We compared our findings to those generated from two human data sets: the DIABIMMUNE project's infant/toddler cohort from Russian Karelia ( Vatanen et al . , 2016 ) and the adult diet-microbiome association study of Johnson et al . , 2019 . In both cases , count tables were obtained from the project's public website and subject identity and sampling schedules were available in the associated metadata . We compared each host cohort's universality at the family/order/class level because this taxonomic level offered the greatest comparative power ( 10 . 1% of families/orders/classes overlap between the cohorts compared to just 3 . 1% of genera and no ASVs ) . The basset model from the 'fido' R package ( Cullen et al . , 2020 ) was fit to each subject's data set using model settings analogous to those employed on the Amboseli baboon series: first , only taxa with non-zero counts in at least 20% of all subjects' series were retained; second , Gaussian process kernel bandwidth settings were chosen in such a way as to encode an expectation of minimum autocorrelation between samples at a distance in time of 90 days . We extracted CLR estimates of taxa at the family level in the same manner as described previously for the Amboseli data set .

```
In [12]: # test with 1 example
send_sumarize_request(processed_text, quiet=False)
```

```
Sending request for text = ## Keywords: microbiology and infectious disease, ecology
## Abstract
Ecological relationships between
OK. Pausing for 10 secs...OK
```

```
Out[12]: 'Gut microbiomes, which are communities of bacteria in our stomach, play a significant role in our health by affecting various processes like digestion and metabolism. However, there are still many unknowns when it comes to the relationships between these bacteria in the gut. This study aimed to understand these relationships by analyzing extensive time series data from 56 baboons over 13 years. The researchers found that most of the relationships between bacteria are weak, negative, and consistent across different hosts. This means that the same patterns of bacterial relationships are more common than unique patterns in each host. The researchers also compared these patterns to two human data sets and found that the universality of bacterial associations in baboons is similar to that in human infants and stronger than one data set from human adults. These findings contribute to our understanding of microbiome personalization, community assembly, stability, and designing microbiome interventions to improve host health.'
```

```
In [13]: ## apply to all rows in eval miniset
# text_cap = 20_000 # temporarily limit to 20k characters because of API restrictions

# print("Summarization process started...")
# df["groq_mistral_summary"] = df["article"].apply(lambda text: send_sumarize_request(text, quiet=False))
# print("Completed")
```

```
In [14]: # create empty column
df["mixtral_summary"] = df["article"].apply(lambda x: "")
df.head()
n = len(df)

for i in range(len(df)):
    print(f"Processing item {i}/{n} = {i*100/n:.1f} %")
    item = df.iloc[i]
    text = build_text_with_headings(item)
    print("Processing article =", text[:200])
    summary = send_sumarize_request(text)
    print("Summary =", summary[:200])
    # df.at[i, "llm_templates"] = parsed_item
    df.at[i, "mixtral_summary"] = summary
    print("-----")
```

Processing item 0/142 = 0.0 %

Processing article = ## Keywords: biochemistry and chemical biology, computational and systems biology

## Abstract

Acylation of diverse carbohydrates occurs across all domains of life and can be catalysed by proteins with

OK. Pausing for 10 secs...OK

Summary = Acylation of carbohydrates is a process that happens in all forms of life and is essential for various functions in bacteria, including symbiosis, resistance to viruses and antimicrobials, and biosynt

-----  
Processing item 1/142 = 0.7 %

Processing article = ## Keywords: computational and systems biology

## Abstract

Honey bee ecology demands they make both rapid and accurate assessments of which flowers are most likely to offer them nectar or pollen . To

OK. Pausing for 10 secs...OK

Summary = Honey bees have impressive decision-making skills when it comes to choosing flowers for nectar or pollen. They can quickly and accurately assess which flowers are most likely to offer rewards based on

-----  
Processing item 2/142 = 1.4 %

Processing article = ## Keywords: genetics and genomics

## Abstract

Biguanides , including the world's most prescribed drug for type 2 diabetes , metformin , not only lower blood sugar , but also promote longevity in prec

OK. Pausing for 10 secs...OK

Summary = Metformin, a commonly prescribed drug for type 2 diabetes, has been associated with longer lifespan and reduced risk of age-related diseases in humans. While its benefits are clear, the molecular mech

-----  
Processing item 3/142 = 2.1 %

Processing article = ## Keywords: microbiology and infectious disease, ecology

## Abstract

Ecological relationships between bacteria mediate the services that gut microbiomes provide to their hosts . Knowing the overall d

OK. Pausing for 10 secs...OK

Summary = Gut microbiomes are diverse and dynamic communities made up of various ecological relationships between bacteria. These relationships can be mutualistic, such as syntrophy and cross-feeding, or compet

-----  
Processing item 4/142 = 2.8 %

Processing article = ## Keywords: neuroscience

## Abstract

Gamma oscillations are believed to underlie cognitive processes by shaping the formation of transient neuronal partnerships on a millisecond scale . These oscilla

OK. Pausing for 10 secs...OK

Summary = Gamma oscillations are a type of brain wave that occur in the frequency range of 30-100 Hz. They have been studied in the context of cognitive functions such as sensory processing, memory, navigation,

-----  
Processing item 5/142 = 3.5 %

Processing article = ## Keywords: genetics and genomics, chromosomes and gene expression

## Abstract

Asynchronous replication of chromosome domains during S phase is essential for eukar

yotic genome function , but the mech

OK. Pausing for 10 secs...OK

Summary = Genetics and genomics are fields of study that look at how genes and genomes work, including chromosomes and gene expression. Chromosomes are structures in the cell that contain genes, and gene expres

-----  
Processing item 6/142 = 4.2 %

Processing article = ## Keywords: computational and systems biology, epidemiology and global health

## Abstract

While biological age in adults is often understood as representing general health and resilience , the concep

OK. Pausing for 10 secs...OK

Summary = Biological age refers to the overall health and resilience of an individual, which can be assessed through various biological indicators such as telomere length and DNA methylation age. In a study of

-----  
Processing item 7/142 = 4.9 %

Processing article = ## Keywords: genetics and genomics, epidemiology and global health

## Abstract

Many genes associated with asthma explain only a fraction of its heritability . Most genome-wide association studies ( GW

OK. Pausing for 10 secs...OK

Summary = This text is a medical research paper about genetics and genomics, epidemiology and global health. It reports on a study that aimed to identify genetic associates of childhood wheezing phenotypes. The

-----  
Processing item 8/142 = 5.6 %

Processing article = ## Keywords: ecology, plant biology

## Abstract

By capitalising on positive biodiversity-productivity relationships , intercropping provides opportunities to improve agricultural sustainability . Inte

OK. Pausing for 10 secs...OK

Summary = Intercropping, or growing multiple species together in the same field, can enhance agricultural sustainability by capitalizing on the positive relationship between biodiversity and productivity. Howev

-----  
Processing item 9/142 = 6.3 %

Processing article = ## Keywords: neuroscience

## Abstract

Many real-world decisions in social contexts are made while observing a partner's actions . To study dynamic interactions during such decisions , we developed a s

OK. Pausing for 10 secs...OK

Summary = This text is about a study that compares how humans and macaque monkeys make decisions in social situations. The study involved a game where two agents, either human or monkey, sit face-to-face and pl

-----  
Processing item 10/142 = 7.0 %

Processing article = ## Keywords: stem cells and regenerative medicine, developmental biology

## Abstract

An in vitro model of human ovarian follicles would greatly benefit the study of female reproduction . Ovarian devel

OK. Pausing for 10 secs...OK

Summary = This text is about the study of female reproduction, specifically the deve

lopment of ovarian follicles. Ovarian development requires the combination of germ c  
ells and several types of somatic cells. A

-----  
Processing item 11/142 = 7.7 %

Processing article = ## Keywords: evolutionary biology

## Abstract

Our interest in the genetic basis of skin color variation between populations led us  
to seek a Native American population with genetically African admixtu

OK. Pausing for 10 secs...OK

Summary = The study focused on the genetic basis of skin color variation in a Native  
American population, the Kalinago of Dominica, who have genetic admixture from Afric  
a but low frequency of European light ski

-----  
Processing item 12/142 = 8.5 %

Processing article = ## Keywords: biochemistry and chemical biology

## Abstract

Antioxidant intervention is considered to inhibit reactive oxygen species ( ROS ) an  
d alleviate hyperglycemia . Paradoxically , moderate exer

OK. Pausing for 10 secs...OK

Summary = Exercise and antioxidant interventions can improve blood glucose control i  
n diabetes by promoting redox balance in different ways. Moderate exercise upregulat  
es compensatory antioxidant capability and

-----  
Processing item 13/142 = 9.2 %

Processing article = ## Keywords: ecology, evolutionary biology

## Abstract

The escape trajectory ( ET ) of prey - measured as the angle relative to the predato  
r's approach path - plays a major role in avoiding predation

OK. Pausing for 10 secs...OK

Summary = Prey animals use escape trajectories (ET) to avoid being caught by predato  
rs. These trajectories can vary greatly between different species and even within th  
e same species. While it was previously be

-----  
Processing item 14/142 = 9.9 %

Processing article = ## Keywords: tools and resources, genetics and genomics, stem c  
ells and regenerative medicine

## Abstract

The recent development of prime editing ( PE ) genome engineering technologies has t  
he potenti

OK. Pausing for 10 secs...OK

Summary = Prime editing (PE) is a new genome editing technology that can simplify th  
e creation of human pluripotent stem cell (hPSC) disease models. PE is a system that  
uses a Cas9-nickase fused to a reverse tr

-----  
Processing item 15/142 = 10.6 %

Processing article = ## Keywords: cancer biology

## Abstract

Multiple myeloma ( MM ) accounts for ~10% of all haematologic malignancies . Little  
is known about high intratumour heterogeneities in patients stratified by th

OK. Pausing for 10 secs...OK

Summary = Multiple myeloma (MM) is a type of blood cancer that affects plasma cells,  
a type of white blood cell found in the bone marrow. The Revised International Stagi  
ng System (R-ISS) is a way to divide MM p

-----  
Processing item 16/142 = 11.3 %

Processing article = ## Keywords: biochemistry and chemical biology, structural biol

ogy and molecular biophysics

## Abstract

Cas7-11 is a Type III-E CRISPR Cas effector that confers programmable RNA cleavage and has poten

OK. Pausing for 10 secs...OK

Summary = Cas7-11 is a type of CRISPR-Cas system that can be used for programmable RNA cleavage and has potential applications in RNA interference. It is unique because it has a single polypeptide containing fo

-----  
Processing item 17/142 = 12.0 %

Processing article = ## Keywords: ecology

## Abstract

Addressing global biodiversity loss requires an expanded focus on multiple dimensions of biodiversity. While most studies have focused on the consequences of plant in

OK. Pausing for 10 secs...OK

Summary = Biodiversity loss is a major concern due to its impact on ecosystems and their functions. While many studies have focused on how the number of species affects functions like plant productivity, fewer

-----  
Processing item 18/142 = 12.7 %

Processing article = ## Keywords: structural biology and molecular biophysics, cell biology

## Abstract

Single-molecule imaging provides a powerful way to study biochemical processes in live cells, yet it remains challen

OK. Pausing for 10 secs...OK

Summary = Researchers have discovered a new way to study molecular interactions in live cells using a technique called Proximity-Assisted Photoactivation (PAPA). This method uses a novel property of rhodamine d

-----  
Processing item 19/142 = 13.4 %

Processing article = ## Keywords: epidemiology and global health

## Abstract

The effect of calcium supplementation on bone mineral accretion in people under 35 years old is inconclusive. To comprehensively summarize the

OK. Pausing for 10 secs...OK

Summary = Calcium supplementation can help improve bone mineral density (BMD) and bone mineral content (BMC) in young people. A systematic review and meta-analysis found that calcium supplementation was associa

-----  
Processing item 20/142 = 14.1 %

Processing article = ## Keywords: microbiology and infectious disease, physics of living systems

## Abstract

In biological systems, liquid and solid-like biomolecular condensates may contain the same molecules but their

OK. Pausing for 10 secs...OK

Summary = This text discusses microbiology and infectious diseases, as well as physics of living systems. It focuses on how the behavior of liquid and solid-like biomolecular condensates differs based on their

-----  
Processing item 21/142 = 14.8 %

Processing article = ## Keywords: structural biology and molecular biophysics, neuroscience

## Abstract

Kinesin superfamily proteins are microtubule-based molecular motors driven by the en

ergy of ATP hydrolysis . Among th

OK. Pausing for 10 secs...OK

Summary = Kinesin superfamily proteins are molecular motors that move along microtubules, which are structures within cells that help in the organization and movement of cellular components. Kinesin-4 is a uni

-----  
Processing item 22/142 = 15.5 %

Processing article = ## Keywords: developmental biology

## Abstract

Highly potent animal stem cells either self renew or launch complex differentiation programs , using mechanisms that are only partly understood . Drosoph

OK. Pausing for 10 secs...OK

Summary = Stem cells are essential for the development and maintenance of animal bodies. The most basic type of stem cell, the totipotent stem cell, can develop into any type of cell and can even create a new o

-----  
Processing item 23/142 = 16.2 %

Processing article = ## Keywords: plant biology, developmental biology

## Abstract

Biological rhythms are ubiquitous . They can be generated by circadian oscillators , which produce daily rhythms in physiology and behavior

OK. Pausing for 10 secs...OK

Summary = The text discusses plant biology and developmental biology, particularly focusing on the circadian clock in sunflowers. The circadian clock is a biological clock found in most eukaryotes and some prok

-----  
Processing item 24/142 = 16.9 %

Processing article = ## Keywords: developmental biology, neuroscience

## Abstract

Hydrogen peroxide is the most common reactive chemical that organisms face on the microbial battlefield . The rate with which hydrogen pero

OK. Pausing for 10 secs...OK

Summary = C. elegans nematodes are often exposed to various threats in their natural habitat, such as hydrogen peroxide (H<sub>2</sub>O<sub>2</sub>) and high temperature. H<sub>2</sub>O<sub>2</sub> is a chemical threat that can be produced by bacteria an

-----  
Processing item 25/142 = 17.6 %

Processing article = ## Keywords: immunology and inflammation, chromosomes and gene expression

## Abstract

Class switch recombination generates distinct antibody isotypes critical to a robust adaptive immune system , and

OK. Pausing for 10 secs...OK

Summary = This text is about immunology and inflammation, specifically focusing on chromosomes and gene expression. The abstract explains that class switch recombination, a process that generates distinct antib

-----  
Processing item 26/142 = 18.3 %

Processing article = ## Keywords: microbiology and infectious disease, immunology and inflammation

## Abstract

Large-scale populations in the world have been vaccinated with COVID-19 vaccines , however , breakthrough infe

OK. Pausing for 10 secs...OK

Summary = A vaccine using a mosaic-type trimer of the receptor-binding domain (RBD) from the SARS-CoV-2 spike protein has been developed and tested in rats. This vaccin

e, called mos-tri-RBD, includes mutations

-----

Processing item 27/142 = 19.0 %

Processing article = ## Keywords: microbiology and infectious disease

## Abstract

Plasmodium falciparum , the causative agent of malaria , remains a global health threat as parasites continue to develop resistance to anti

OK. Pausing for 10 secs...OK

Summary = Malaria is a deadly infectious disease caused by Plasmodium parasites, with P. falciparum being the most dangerous to humans. In 2020, there were over 600,000 deaths due to malaria, mainly in sub-Saharan

-----

Processing item 28/142 = 19.7 %

Processing article = ## Keywords: tools and resources, genetics and genomics, computational and systems biology

## Abstract

Early hematopoiesis is a continuous process in which hematopoietic stem and progenitor cells (HSCs)

OK. Pausing for 10 secs...OK

Summary = This text is about a study on hematopoiesis, the process by which blood cells are formed in the body. The study used single-cell RNA sequencing (scRNA-seq) to analyze the gene expression profiles of blood

-----

Processing item 29/142 = 20.4 %

Processing article = ## Keywords: genetics and genomics, neuroscience

## Abstract

Animals must learn to ignore stimuli that are irrelevant to survival and attend to ones that enhance survival. When a stimulus regularly

OK. Pausing for 10 secs...OK

Summary = Honey bees, like other animals, need to learn to ignore irrelevant stimuli and pay attention to ones that are important for survival. One way they do this is through a process called latent inhibition

-----

Processing item 30/142 = 21.1 %

Processing article = ## Keywords: tools and resources, cell biology

## Abstract

The specific functional properties of a tissue are distributed amongst its component cell types. The various cells act coherently, as an ensemble

OK. Pausing for 10 secs...OK

Summary = The text describes a new method for identifying and tracking diverse cell types within an organ in real time, in a non-destructive way. This method uses a non-invasive imaging modality that capitalizes

-----

Processing item 31/142 = 21.8 %

Processing article = ## Keywords: stem cells and regenerative medicine, cell biology

## Abstract

Cardiovascular disease is the leading cause of death worldwide due to the inability of adult heart to regenerate after injury

OK. Pausing for 10 secs...OK

Summary = Heart disease is the leading cause of death worldwide, partly because adult hearts can't regenerate after injury. A process called methylation, specifically N6-methyladenosine (m6A) methylation, plays

-----

Processing item 32/142 = 22.5 %

Processing article = ## Keywords: epidemiology and global health, cancer biology

## Abstract



The risk of second primary cancers ( SPC ) is increasing after the first primary cancers ( FPC ) are diagnosed and treated . The

OK. Pausing for 10 secs...OK

Summary = Certain types of primary cancers can cause an increased risk of developing another type of primary cancer in the same individual. This study used data from the SEER program and the UK Biobank to analyze

-----

Processing item 33/142 = 23.2 %

Processing article = ## Keywords: neuroscience

## Abstract

In economic decision-making individuals choose between items based on their perceived value . For both humans and nonhuman primates , these decisions are often ca

OK. Pausing for 10 secs...OK

Summary = Economic decisions involve choosing between items based on their perceived value. In economic decision-making, individuals shift their gaze between the available options. These shifts in gaze can infl

-----

Processing item 34/142 = 23.9 %

Processing article = ## Keywords: ecology, evolutionary biology, short report

## Abstract

Most phytophagous insect species exhibit a limited diet breadth and specialize on a few or a single host plant . In contrast , some

OK. Pausing for 10 secs...OK

Summary = Insects that eat plants usually specialize in a few or just one type of host plant, but some can eat a wide variety of plants from many different families. It's unclear whether this broad diet is due

-----

Processing item 35/142 = 24.6 %

Processing article = ## Keywords: microbiology and infectious disease

## Abstract

Multiple genetic changes in the enteric pathogen *Yersinia pseudotuberculosis* have driven the emergence of *Yersinia pestis* , the arthropod-bo

OK. Pausing for 10 secs...OK

Summary = *Yersinia pestis*, the bacterium that causes the plague, evolved from *Yersinia pseudotuberculosis* around 6000-7000 years ago. These two species have different life cycles, with *Y. pseudotuberculosis* bei

-----

Processing item 36/142 = 25.4 %

Processing article = ## Keywords: neuroscience

## Abstract

The peer review process is a critical step in ensuring the quality of scientific research . However , its subjectivity has raised concerns . To investigate this i

OK. Pausing for 10 secs...OK

Summary = This text is a research paper about a study that used OpenAI's generative artificial intelligence, ChatGPT, to analyze language use in peer review reports from neuroscience papers published in Nature

-----

Processing item 37/142 = 26.1 %

Processing article = ## Keywords: immunology and inflammation, cancer biology

## Abstract

Hepatic metastases are a poor prognostic factor of colorectal carcinoma ( CRC ) and new strategies to reduce the risk of liver CRC

OK. Pausing for 10 secs...OK

Summary = This text discusses the use of interferon-alpha (IFN $\alpha$ ) in colorectal cancer (CRC) and its effects on liver metastasis in mouse models. The study found that continuous infusion of therapeutic doses of

-----  
Processing item 38/142 = 26.8 %

Processing article = ## Keywords: microbiology and infectious disease

## Abstract

Although vaccination is broadly used in North American swine breeding herds , managing swine influenza is challenging primarily due to the

OK. Pausing for 10 secs...OK

Summary = Influenza A viruses (IAV) are significant pathogens in humans, birds, and pigs worldwide. IAV infections in pigs can result in respiratory disease, reducing their health and profitability in the pork

-----  
Processing item 39/142 = 27.5 %

Processing article = ## Keywords: neuroscience

## Abstract

Aversive stimuli can cause hippocampal place cells to remap their firing fields , but it is not known whether remapping plays a role in storing memories of aversive

OK. Pausing for 10 secs...OK

Summary = In this study, researchers used calcium imaging to observe the behavior of hippocampal place cells in rats as they navigated a maze. Place cells are neurons in the hippocampus that fire selectively when

-----  
Processing item 40/142 = 28.2 %

Processing article = ## Keywords: immunology and inflammation

## Abstract

HIV-1 infection results in the activation of inflammasome that may facilitate viral spread and establishment of viral reservoirs . We evaluated the

OK. Pausing for 10 secs...OK

Summary = HIV-1 infection triggers an immune response that can lead to chronic inflammation and the development of reservoirs of the virus in the body. This can contribute to the progression of the disease and

-----  
Processing item 41/142 = 28.9 %

Processing article = ## Keywords: neuroscience

## Abstract

The loss of a spouse is often cited as the most traumatic event in a person's life . However , for most people , the severity of grief and its maladaptive effects

OK. Pausing for 10 secs...OK

Summary = The loss of a spouse is often considered one of the most traumatic events in a person's life, but most people eventually adapt to the loss over time. Similarly, prairie voles, which form strong pair bonds

-----  
Processing item 42/142 = 29.6 %

Processing article = ## Keywords: neuroscience

## Abstract

Motor variability is a fundamental feature of developing systems allowing motor exploration and learning . In human infants , leg movements involve a small number

OK. Pausing for 10 secs...OK

Summary = In this study, researchers explored the relationship between motor variability and the development of motor primitives in human infants. Motor primitives are basic patterns of coordination that allow

-----  
Processing item 43/142 = 30.3 %

Processing article = ## Keywords: microbiology and infectious disease, epidemiology and global health

## Abstract

The Global Typhoid Genomics Consortium was established to bring together the typhoid

research community to

OK. Pausing for 10 secs...OK

Summary = The Global Typhoid Genomics Consortium brought together researchers to analyze data on *Salmonella enterica* serovar Typhi, the bacteria causing typhoid fever. This analysis included 13,000 genome sequences

-----  
Processing item 44/142 = 31.0 %

Processing article = ## Keywords: plant biology, chromosomes and gene expression

## Abstract

To synchronize flowering time with spring, many plants undergo vernalization, a floral-promotion process triggered by exposure

OK. Pausing for 10 secs...OK

Summary = In the field of plant biology, vernalization is a process that helps flowers to bloom at the right time in spring. This is achieved by exposing the plant to long-term winter cold, which triggers a floral

-----  
Processing item 45/142 = 31.7 %

Processing article = ## Keywords: microbiology and infectious disease

## Abstract

HIV-1 reservoir cells that circulate in peripheral blood during suppressive antiretroviral therapy (ART) have been well characterized, but

OK. Pausing for 10 secs...OK

Summary = HIV-1, the virus that causes AIDS, can persist in the body even with antiretroviral therapy (ART). This is because the virus can hide in reservoirs in the body, such as in lymph nodes and the brain. It

-----  
Processing item 46/142 = 32.4 %

Processing article = ## Keywords: tools and resources, genetics and genomics, cell biology

## Abstract

Circadian clocks are highly conserved transcriptional regulators that control ~24 hr oscillations in gene expression, but

OK. Pausing for 10 secs...OK

Summary = Circadian clocks are biological timekeepers that regulate daily rhythms in gene expression, physiological functions, and behavior in many organisms, including flies and humans. These clocks are controlled by

-----  
Processing item 47/142 = 33.1 %

Processing article = ## Keywords: genetics and genomics, cell biology

## Abstract

Membrane trafficking pathways perform important roles in establishing and maintaining the endosomal network. Retrograde protein sorting from

OK. Pausing for 10 secs...OK

Summary = Membrane trafficking pathways are important for establishing and maintaining the endosomal network. Protein sorting from the endosome is promoted by SNX-BAR-containing coat complexes, including retromer

-----  
Processing item 48/142 = 33.8 %

Processing article = ## Keywords: ecology

## Abstract

The effects of temperature on interaction strengths are important for understanding and forecasting how global climate change impacts marine ecosystems; however, tracking

OK. Pausing for 10 secs...OK

Summary = The researchers studied the impact of temperature on interactions between marine fish species using environmental DNA (eDNA) metabarcoding. They collected seawater samples twice a month from 11 coastal

-----  
Processing item 49/142 = 34.5 %

Processing article = ## Keywords: evolutionary biology

## Abstract

The avian palaeognath phylogeny has been recently revised significantly due to the advancement of genome-wide comparative analyses and provides the opportunity

OK. Pausing for 10 secs...OK

Summary = Birds, or avian dinosaurs, are a diverse group with a wide variety of eggshell sizes, shapes, and colors. Eggshells are biominerals made mostly of calcium carbonate and can be well-preserved in the fossil record

-----  
Processing item 50/142 = 35.2 %

Processing article = ## Keywords: tools and resources, neuroscience

## Abstract

Ultrasonic vocalizations ( USVs ) fulfill an important role in communication and navigation in many species . Because of their social and affective

OK. Pausing for 10 secs...OK

Summary = Ultrasonic vocalizations (USVs) are sounds used by animals for communication and navigation, and are particularly important in rodents. They have social and emotional significance and are used in various contexts

-----  
Processing item 51/142 = 35.9 %

Processing article = ## Keywords: developmental biology

## Abstract

Sarcomeres are the force-producing units of all striated muscles . Their nanoarchitecture critically depends on the large titin protein , which in vertebrates

OK. Pausing for 10 secs...OK

Summary = Muscle strength and movement in both humans and animals are due to the contraction of striated muscles, which is powered by sarcomeres. These are the force-producing units of muscles and are organized into sarcomeres

-----  
Processing item 52/142 = 36.6 %

Processing article = ## Keywords: tools and resources, cell biology

## Abstract

Numerous cell functions are accompanied by phenotypic changes in viscoelastic properties , and measuring them can help elucidate higher level

OK. Pausing for 10 secs...OK

Summary = The study presents a new method for measuring the elastic and viscous properties of individual cells. The method involves suspending cells in a high-viscosity fluid and pumping them through a microfluidic device

-----  
Processing item 53/142 = 37.3 %

Processing article = ## Keywords: developmental biology

## Abstract

Cardiac muscle has the highest mitochondrial density of any human tissue , but mitochondrial dysfunction is not a recognized cause of isolated cardiomyopathy

OK. Pausing for 10 secs...OK

Summary = Mitochondria, the powerhouse of cells, are particularly important in heart function as they provide the energy needed for heart development, contraction, and repair. However, mitochondrial dysfunction

-----  
Processing item 54/142 = 38.0 %

Processing article = ## Keywords: evolutionary biology, cell biology

## Abstract

Several previous lines of research have suggested , indirectly , that mouse lifespan is particularly susceptible to endocrine or nutritional

OK. Pausing for 10 secs...OK

Summary = The study explored the impact of olfactory information during the early stages of life on the longevity of mice. Previous research suggested that mouse lifespan could be influenced by endocrine or nutrient

-----

Processing item 55/142 = 38.7 %

Processing article = ## Keywords: neuroscience

## Abstract

Designer receptors exclusively activated by designer drugs ( DREADDs ) are chemogenetic tools for remote control of targeted cell populations using chemical actuators

OK. Pausing for 10 secs...OK

Summary = Designer receptors exclusively activated by designer drugs (DREADDs) are tools used to control specific cells in the body using chemical actuators. One such actuator is clozapine-N-oxide (CNO). However

-----

Processing item 56/142 = 39.4 %

Processing article = ## Keywords: ecology, physics of living systems

## Abstract

Microorganism sensing of and responding to ambient chemical gradients regulates a myriad of microbial processes that are fundamental to ecosystems

OK. Pausing for 10 secs...OK

Summary = This text discusses a microfluidic device used for studying the chemotaxis, or movement, of microorganisms in response to chemical stimuli. The device, called a Multiplexed Chemotaxis Device (MCD), is

-----

Processing item 57/142 = 40.1 %

Processing article = ## Keywords: short report, developmental biology

## Abstract

While neural crest development is known to be transcriptionally controlled via sequential activation of gene regulatory networks ( GRNs ),

OK. Pausing for 10 secs...OK

Summary = Neural crest cells are an essential and multipotent cell population during embryonic development. They undergo a series of processes including induction, specification, and epithelial-mesenchymal transition

-----

Processing item 58/142 = 40.8 %

Processing article = ## Keywords: developmental biology, neuroscience

## Abstract

Breathing , and the motor circuits that control it , is essential for life . At the core of respiratory circuits are Dbx1-derived interneurons

OK. Pausing for 10 secs...OK

Summary = Breathing is essential for life, controlled by complex neural networks and cellular mechanisms in the brainstem and spinal cord. These networks control muscle contractions to produce respiratory-related

-----

Processing item 59/142 = 41.5 %

Processing article = ## Keywords: neuroscience

## Abstract

The powerful allure of social media platforms has been attributed to the human need for social rewards . Here , we demonstrate that the spread of misinformation on

OK. Pausing for 10 secs...OK

Summary = Social media platforms are known for their role in the spread of misinformation, which can have serious real-world consequences. This issue has been linked to the current incentive structure of social

-----

Processing item 60/142 = 42.3 %

Processing article = ## Keywords: structural biology and molecular biophysics, cell biology

## Abstract

The disassembly of microtubules can generate force and drive intracellular motility . During mitosis , for example ,

OK. Pausing for 10 secs...OK

Summary = Microtubules are vital for transporting materials within cells. They can generate force and motion by disassembling, causing protofilaments to curl outward and exert pulling force directly on kinetoch

-----

Processing item 61/142 = 43.0 %

Processing article = ## Keywords: evolutionary biology, cell biology

## Abstract

When attacked , hagfishes produce a soft , fibrous defensive slime within a fraction of a second by ejecting mucus and threads into seawater

OK. Pausing for 10 secs...OK

Summary = Hagfishes, a group of jawless fish, have a unique and effective defense mechanism where they produce a cloud of slimy, fibrous mucus when attacked. This slime is produced by specialized thread cells,

-----

Processing item 62/142 = 43.7 %

Processing article = ## Keywords: neuroscience

## Abstract

There are no pharmacological disease-modifying treatments with an enduring effect to mitigate the seizures and comorbidities of established chronic temporal lobe

OK. Pausing for 10 secs...OK

Summary = This text discusses a study about a treatment for chronic temporal lobe epilepsy (TLE) in rats. The treatment involves giving the rats sodium selenate for four weeks. The study found that this treatment

-----

Processing item 63/142 = 44.4 %

Processing article = ## Keywords: epidemiology and global health

## Abstract

Fragility fracture is associated with an increased risk of mortality , but mortality is not part of doctor-patient communication . Here , we int

OK. Pausing for 10 secs...OK

Summary = The text discusses the development of a new concept called 'Skeletal Age' to better communicate the combined risk of fracture and fracture-associated mortality to patients. Skeletal Age is defined as

-----

Processing item 64/142 = 45.1 %

Processing article = ## Keywords: tools and resources, computational and systems biology, neuroscience

## Abstract

Accurate brain tissue extraction on magnetic resonance imaging ( MRI ) data is crucial for analyzing brain

OK. Pausing for 10 secs...OK

Summary = Brain extraction, or the process of separating brain tissues from non-brain tissues in magnetic resonance imaging (MRI) data, is a crucial step in analyzing brain structure and function. Conventional

-----

Processing item 65/142 = 45.8 %

Processing article = ## Keywords: microbiology and infectious disease

## Abstract

Heme can serve as iron source in many environments , including the iron-poor animal host environment . The fungal pathobiont *Candida albica*

OK. Pausing for 10 secs...OK

Summary = Heme can be an essential source of iron for certain organisms, including the fungal pathogen *Candida albicans*. This fungus has proteins called Frp1 and Frp2 that are important for using heme as an iron source.

-----  
Processing item 66/142 = 46.5 %

Processing article = ## Keywords: biochemistry and chemical biology, structural biology and molecular biophysics

## Abstract

The seventh pandemic of the diarrheal cholera disease, which began in 1960, is caused by the G

OK. Pausing for 10 secs...OK

Summary = Cholera is a dangerous disease caused by the bacterium *Vibrio cholerae*, which has been responsible for pandemic outbreaks since 1960. The bacterium can survive in changing environments due to sensory

-----  
Processing item 67/142 = 47.2 %

Processing article = ## Keywords: neuroscience

## Abstract

The hippocampus supports multiple cognitive functions including episodic memory. Recent work has highlighted functional differences along the anterior-posterior

OK. Pausing for 10 secs...OK

Summary = The hippocampus, a brain structure essential for memory and navigation, shows functional differences along its length, with different parts linked to various cognitive functions. A recent study used M

-----  
Processing item 68/142 = 47.9 %

Processing article = ## Keywords: neuroscience

## Abstract

Stimulator of interferon genes (STING) is activated after detection of cytoplasmic dsDNA by cGAS (cyclic GMP-AMP synthase) as part of the innate immunity defense

OK. Pausing for 10 secs...OK

Summary = The text discusses a study on the role of STING, a protein involved in the body's immune response, in neurodegenerative diseases. The researchers used mice with a specific genetic variant of STING, N1

-----  
Processing item 69/142 = 48.6 %

Processing article = ## Keywords: immunology and inflammation, cancer biology

## Abstract

Despite the remarkable successes of cancer immunotherapies, the majority of patients will experience only partial response followed by

OK. Pausing for 10 secs...OK

Summary = This text is about a study on immunology and inflammation, specifically in the field of cancer biology. The study found that tumor cells can evade immunotherapy by forming unique transient cell-in-cell

-----  
Processing item 70/142 = 49.3 %

Processing article = ## Keywords: neuroscience

## Abstract

Theory of Mind (ToM) refers to the cognitive ability to attribute mental states to other individuals. This ability extends even to the attribution of mental states

OK. Pausing for 10 secs...OK

Summary = Theory of Mind (ToM) refers to the ability to attribute mental states to oneself and others. This study explores how marmosets, a highly social primate species, process Frith-Happé animations, which are

Processing item 71/142 = 50.0 %

Processing article = ## Keywords: stem cells and regenerative medicine, developmental biology

## Abstract

Coordinated regulation of gene activity by transcriptional and translational mechanisms poise stem cells for a time

OK. Pausing for 10 secs...OK

Summary = Stem cells and regenerative medicine are areas of study that focus on how stem cells can be used to create new cells, tissues, or even organs in order to replace or repair those that are damaged due to

-----  
Processing item 72/142 = 50.7 %

Processing article = ## Keywords: medicine

## Abstract

Billions of apoptotic cells are removed daily in a human adult by professional phagocytes ( e . g . macrophages ) and neighboring nonprofessional phagocytes ( e . g .

OK. Pausing for 10 secs...OK

Summary = Neutrophils, typically known for combating infection, have a unique and unexpected role in the disposal of dead cells in the liver. In the human liver, neutrophils are the main type of cell responsible

-----  
Processing item 73/142 = 51.4 %

Processing article = ## Keywords: neuroscience

## Abstract

Elaborate behaviours are produced by tightly controlled flexor-extensor motor neuron activation patterns . Motor neurons are regulated by a network of interneurons

OK. Pausing for 10 secs...OK

Summary = In simple terms, the text discusses a series of studies on how the brain controls movement. The researchers used various techniques to trace the pathways that signals take from the brain to the muscle

-----  
Processing item 74/142 = 52.1 %

Processing article = ## Keywords: neuroscience

## Abstract

Respiration can induce motion and CO2 fluctuation during resting-state fMRI ( rsfMRI ) scans , which will lead to non-neural artifacts in the rsfMRI signal . In the

OK. Pausing for 10 secs...OK

Summary = Resting-state fMRI (rsfMRI) is a powerful tool used to study brain-wide functional connectivity. However, rsfMRI signals can be affected by systemic physiological changes such as respiration. Respiration

-----  
Processing item 75/142 = 52.8 %

Processing article = ## Keywords: ecology

## Abstract

Predators that target multiple prey types are predicted to switch foraging modes according to prey profitability to increase energy returns in dynamic environments . Here

OK. Pausing for 10 secs...OK

Summary = Bats that primarily hunt for ground-dwelling insects, called greater mouse-eared bats, use two different hunting strategies. They either swoop down to the ground to grab their prey (gleaning) or they

-----  
Processing item 76/142 = 53.5 %

Processing article = ## Keywords: evolutionary biology, neuroscience

## Abstract

The evolution of human right-handedness has been intensively debated for decades . Manual lateralization patterns in non-human primates have



OK. Pausing for 10 secs...OK

Summary = Handedness, or the preference for using one hand over the other, is a universal trait in humans, with the majority being right-handed. However, the evolution of human right-handedness and its relation

-----

Processing item 77/142 = 54.2 %

Processing article = ## Keywords: medicine

## Abstract

Williams-Beuren syndrome ( WBS ) is a rare genetic multisystemic disorder characterized by mild-to-moderate intellectual disability and hypersocial phenotype , while

OK. Pausing for 10 secs...OK

Summary = Williams-Beuren syndrome (WBS) is a genetic disorder that affects both the central nervous system and the cardiovascular system. It is caused by a deletion of a region on chromosome 7 that contains 26

-----

Processing item 78/142 = 54.9 %

Processing article = ## Keywords: evolutionary biology, developmental biology

## Abstract

During development , the growing organism transits through a series of temporally regulated morphological stages to generate the ad

OK. Pausing for 10 secs...OK

Summary = In living organisms, development from embryo to adult involves passing through different stages, each with distinct characteristics. In humans, this progresses from childhood to puberty and adulthood,

-----

Processing item 79/142 = 55.6 %

Processing article = ## Keywords: computational and systems biology

## Abstract

Cycling of co-substrates , whereby a metabolite is converted among alternate forms via different reactions , is ubiquitous in metabolism . Se

OK. Pausing for 10 secs...OK

Summary = Co-substrate cycling is a common occurrence in metabolism, where a metabolite is converted among alternate forms via different reactions. This cycling creates interconnections within metabolism, spann

-----

Processing item 80/142 = 56.3 %

Processing article = ## Keywords: stem cells and regenerative medicine

## Abstract

The origin and differentiation mechanism of articular chondrocytes remain poorly understood . Broadly , the difference in developmental me

OK. Pausing for 10 secs...OK

Summary = Stem cell research is crucial for understanding regenerative medicine. Articular chondrocytes, which make up the cartilage in joints, have an unclear origin and differentiation mechanism. This study a

-----

Processing item 81/142 = 57.0 %

Processing article = ## Keywords: structural biology and molecular biophysics

## Abstract

Photosystem II ( PSII ) uses the energy from red light to split water and reduce quinone , an energy-demanding process based on chl

OK. Pausing for 10 secs...OK

Summary = Photosystem II (PSII) is a key enzyme in oxygenic photosynthesis, responsible for converting energy from sunlight into chemical energy. It typically contains 35 chlorophyll a (Chl-a) and 2 pheophytins

-----

Processing item 82/142 = 57.7 %

Processing article = ## Keywords: cell biology, developmental biology

## Abstract

The skeletal system contains a series of sophisticated cellular lineages arising from the mesenchymal stem cells ( MSCs ) and hematopoietic

OK. Pausing for 10 secs...OK

Summary = Osteocytes, the long-living cells in the bone matrix, play crucial roles in maintaining skeletal homeostasis. They regulate bone formation, bone resorption, and bone marrow hematopoiesis. During aging

-----

Processing item 83/142 = 58.5 %

Processing article = ## Keywords: genetics and genomics, evolutionary biology

## Abstract

Organisms can adapt to an environment by taking multiple mutational paths . This redundancy at the genetic level , where many mutat

OK. Pausing for 10 secs...OK

Summary = The text discusses a study that uses the Escherichia coli long-term evolution experiment (LTEE) to understand the molecular mechanisms behind complex adaptations. The LTEE involves 12 replicate popula

-----

Processing item 84/142 = 59.2 %

Processing article = ## Keywords: cell biology, cancer biology

## Abstract

CD133 ( prominin 1 ) is widely viewed as a cancer stem cell marker in association with drug resistance and cancer recurrence . Herein , we report

OK. Pausing for 10 secs...OK

Summary = CD133 is a protein commonly viewed as a marker for cancer stem cells, and is associated with drug resistance and cancer recurrence. In this study, researchers found that in mouse liver regeneration, c

-----

Processing item 85/142 = 59.9 %

Processing article = ## Keywords: biochemistry and chemical biology

## Abstract

Ciliopathies manifest from sensory abnormalities to syndromic disorders with multi-organ pathologies , with retinal degeneration a highly pen

OK. Pausing for 10 secs...OK

Summary = This text is about a study that aimed to find drug candidates for maintaining photoreceptor survival in retinal diseases, specifically Leber congenital amaurosis (LCA) type 10, which is caused by muta

-----

Processing item 86/142 = 60.6 %

Processing article = ## Keywords: microbiology and infectious disease, ecology

## Abstract

While foraging for nectar and pollen , bees are exposed to a myriad of xenobiotics , including plant metabolites , which may exert

OK. Pausing for 10 secs...OK

Summary = Bees are exposed to various plant compounds, including amygdalin, a cyanogenic glycoside found in almonds and other fruits. While amygdalin can be toxic to bees, they are still able to consume it in l

-----

Processing item 87/142 = 61.3 %

Processing article = ## Keywords: structural biology and molecular biophysics

## Abstract

Candida albicans causes severe invasive candidiasis . C . albicans infection requires the virulence factor candidalysin ( CL ) whic

OK. Pausing for 10 secs...OK

Summary = Candida albicans is a harmful fungus that causes severe invasive candidias

is, a common and deadly infection in hospitalized patients. This fungus releases a toxin called candidalysin (CL) that damages

-----  
Processing item 88/142 = 62.0 %

Processing article = ## Keywords: medicine, neuroscience

## Abstract

Tailoring interventions to patient subgroups can improve intervention outcomes for various conditions . However , it is unclear how much of this improve

OK. Pausing for 10 secs...OK

Summary = The text is a medical research paper that explores the role of contextual factors, such as personalization, in the effectiveness of interventions for various conditions. The paper describes two studie

-----  
Processing item 89/142 = 62.7 %

Processing article = ## Keywords: neuroscience

## Abstract

Shared lineage has diverse effects on patterns of neuronal connectivity . In mammalian cortex , excitatory sister neurons assemble into shared microcircuits . In

OK. Pausing for 10 secs...OK

Summary = Neuroscientists have studied how shared lineage affects the way neurons connect with each other in circuits. In mammal and fly brains, neurons that come from the same origin tend to form specific patt

-----  
Processing item 90/142 = 63.4 %

Processing article = ## Keywords: neuroscience

## Abstract

Healthy neuronal networks rely on homeostatic plasticity to maintain stable firing rates despite changing synaptic drive . These mechanisms , however , can themse

OK. Pausing for 10 secs...OK

Summary = Homeostatic plasticity is a crucial mechanism that allows the brain to maintain stable activity levels in the face of changing conditions. This process can be destabilizing if activated inappropriatel

-----  
Processing item 91/142 = 64.1 %

Processing article = ## Keywords: genetics and genomics, developmental biology

## Abstract

The tips of the developing respiratory buds are home to important progenitor cells marked by the expression of SOX9 and ID2 . Earl

OK. Pausing for 10 secs...OK

Summary = The text is about a study that investigates the role of genetics and genomics in developmental biology, specifically focusing on the development of the respiratory system in mice. The study identifies

-----  
Processing item 92/142 = 64.8 %

Processing article = ## Keywords: neuroscience

## Abstract

Working memory ( WM ) is one of the most affected cognitive domains in multiple sclerosis ( MS ) , which is mainly studied by the previously established binary mo

OK. Pausing for 10 secs...OK

Summary = Working memory (WM), a crucial cognitive domain often affected in Multiple Sclerosis (MS), has typically been studied using the slot model hypothesis. However, recent studies suggest that the resource

-----  
Processing item 93/142 = 65.5 %

Processing article = ## Keywords: neuroscience

## Abstract

Nociception is a neural process that animals have developed to avoid potentially tissue-damaging stimuli . While nociception is triggered in the peripheral nervous system .

OK. Pausing for 10 secs...OK

Summary = Nociception is a fundamental process that animals, including flies and mammals, have developed to detect and avoid potentially tissue-damaging stimuli. In flies, this process is regulated by a neuropathic pain pathway

-----

Processing item 94/142 = 66.2 %

Processing article = ## Keywords: biochemistry and chemical biology, ecology

## Abstract

Plants often generate secondary metabolites as defense mechanisms against parasites . Although some fungi may potentially overcome this defense

OK. Pausing for 10 secs...OK

Summary = Plants often produce chemicals to defend against parasites, but some fungi have found ways to overcome this defense. A study found a fungus that can grow on a plant called Aglaia, despite the plant's defenses

-----

Processing item 95/142 = 66.9 %

Processing article = ## Keywords: tools and resources, neuroscience

## Abstract

Psychophysical methods are a cornerstone of psychology , cognitive science , and neuroscience where they have been used to quantify behavior

OK. Pausing for 10 secs...OK

Summary = Psychophysical methods are a crucial part of psychology, cognitive science, neuroscience, and behavioral economics, as they provide precise and reliable measurements of the relationship between external stimuli and internal responses

-----

Processing item 96/142 = 67.6 %

Processing article = ## Keywords: computational and systems biology, neuroscience

## Abstract

Models based on normative principles have played a major role in our understanding of how the brain forms decisions . However ,

OK. Pausing for 10 secs...OK

Summary = This text is a research paper in the fields of computational and systems biology and neuroscience. It discusses a model for how the brain makes decisions in dynamic environments, where conditions can change rapidly

-----

Processing item 97/142 = 68.3 %

Processing article = ## Keywords: ecology, computational and systems biology

## Abstract

Despite efforts from scientists and regulators , biodiversity is declining at an alarming rate . Unless we find transformative solutions

OK. Pausing for 10 secs...OK

Summary = Biodiversity is crucial for ecosystem services that support economic prosperity and quality of life. However, biodiversity is declining rapidly due to human activities. Freshwater ecosystems have suffered the most

-----

Processing item 98/142 = 69.0 %

Processing article = ## Keywords: neuroscience

## Abstract

When interacting with the dynamic world , the brain receives outdated sensory information , due to the time required for neural transmission and processing . In mammals

OK. Pausing for 10 secs...OK

Summary = The brain receives outdated sensory information due to delays in neural transmission and processing. To perceive motion, the brain may predict the position of moving objects using past trajectory information

-----  
Processing item 99/142 = 69.7 %

Processing article = ## Keywords: neuroscience

## Abstract

One signature of the human brain is its ability to derive knowledge from language inputs , in addition to nonlinguistic sensory channels such as vision and touch

OK. Pausing for 10 secs...OK

Summary = The human brain's unique ability to derive knowledge from language inputs is significant because it allows us to understand symbols and words beyond just sensory experiences. A study was conducted to

-----  
Processing item 100/142 = 70.4 %

Processing article = ## Keywords: neuroscience

## Abstract

Replay , the sequential reactivation within a neuronal ensemble , is a central hippocampal mechanism postulated to drive memory processing . While both rate and p

OK. Pausing for 10 secs...OK

Summary = In this study, researchers explored the role of replay, a phenomenon where neurons in the hippocampus reactivate in a sequence similar to their activation during a behavioral episode, in memory processes

-----  
Processing item 101/142 = 71.1 %

Processing article = ## Keywords: cell biology

## Abstract

Impaired spermatogenesis and male infertility are common manifestations associated with mitochondrial diseases , yet the underlying mechanisms linking these conditions

OK. Pausing for 10 secs...OK

Summary = This text is about a study of mitochondrial diseases and male infertility. Mice with a certain genetic mutation, called PARL-deficient mice, develop early testicular atrophy and a complete arrest of spermatogenesis

-----  
Processing item 102/142 = 71.8 %

Processing article = ## Keywords: medicine, immunology and inflammation

## Abstract

Analogues of the hepatokine fibroblast growth factor 21 ( FGF21 ) are in clinical development for type 2 diabetes and nonalcoholic steato

OK. Pausing for 10 secs...OK

Summary = FGF21 is a hepatokine with lipid-lowering and anti-inflammatory properties that has been identified as a promising potential therapeutic to treat NASH. It is specifically expressed in the liver and adipose tissue

-----  
Processing item 103/142 = 72.5 %

Processing article = ## Keywords: tools and resources, neuroscience

## Abstract

The neurophysiology of cells and tissues are monitored electrophysiologically and optically in diverse experiments and species , ranging from

OK. Pausing for 10 secs...OK

Summary = Neurophysiology is the scientific study of the brain's electrical activity. Neurophysiology data can be collected using various methods such as electrophysiology, optophysiology, and behavioral tracking

-----  
Processing item 104/142 = 73.2 %

Processing article = ## Keywords: immunology and inflammation, neuroscience

## Abstract

Infection with Influenza A virus ( IAV ) causes the well-known symptoms of the flu , including fever , loss of appetite , and excessive

OK. Pausing for 10 secs...OK

Summary = Influenza A virus (IAV) infection can lead to long-term neurological issues like encephalitis lethargica and narcolepsy, even after the virus is cleared from the system. A study was conducted to under

-----

Processing item 105/142 = 73.9 %

Processing article = ## Keywords: computational and systems biology, cancer biology  
## Abstract

Apicobasal cell polarity loss is a founding event in epithelial-mesenchymal transition and epithelial tumorigenesis, yet how

OK. Pausing for 10 secs...OK

Summary = This text is about a study in computational and systems biology, specifically looking at cancer biology. The researchers used single-cell RNA sequencing to study how apical-basal cell polarity loss li

-----

Processing item 106/142 = 74.6 %

Processing article = ## Keywords: neuroscience  
## Abstract

Humans learn about the environment either directly by interacting with it or indirectly by seeking information about it from social sources such as conspecifics.

OK. Pausing for 10 secs...OK

Summary = This text is about a study that explores how people learn from social and non-social information sources. The study found that people have more confidence in social information sources, like advice fr

-----

Processing item 107/142 = 75.4 %

Processing article = ## Keywords: computational and systems biology, neuroscience  
## Abstract

Synaptic communication relies on the fusion of synaptic vesicles with the plasma membrane, which leads to neurotransmitter rel

OK. Pausing for 10 secs...OK

Summary = Synaptic communication is a process that involves the release of neurotransmitters from a presynaptic terminal to transmit information to a postsynaptic terminal. This process is triggered by action p

-----

Processing item 108/142 = 76.1 %

Processing article = ## Keywords: evolutionary biology  
## Abstract

Sauropterygia was a taxonomically and ecomorphologically diverse clade of Mesozoic marine reptiles spanning the Early Triassic to the Late Cretaceous. Sa

OK. Pausing for 10 secs...OK

Summary = Sauropterygia was a diverse group of Mesozoic marine reptiles that lived from the Early Triassic to the Late Cretaceous. They are traditionally divided into two major lineages: Placodontia and Eosauro

-----

Processing item 109/142 = 76.8 %

Processing article = ## Keywords: genetics and genomics, epidemiology and global health  
## Abstract

Smoking-associated DNA methylation levels identified through epigenome-wide association studies (EWASs) are generally a

OK. Pausing for 10 secs...OK

Summary = This text is about a study that looks at the relationship between genetics, genomics, epidemiology, and global health, specifically in relation to smoking and DNA methylation. The study uses data from

-----

Processing item 110/142 = 77.5 %

Processing article = ## Keywords: neuroscience

## Abstract

The structure and function of the vertebrate retina have been extensively studied across species with an isolated, ex vivo preparation. Retinal function in vivo

OK. Pausing for 10 secs...OK

Summary = The text describes a study that compares the output of the retina in awake, anesthetized, and ex vivo states. The researchers found that the retina in awake mice had faster kinetics, a larger dynamic

-----  
Processing item 111/142 = 78.2 %

Processing article = ## Keywords: cell biology, physics of living systems

## Abstract

Fixing cells with paraformaldehyde (PFA) is an essential step in numerous biological techniques as it is thought to preserve a snapshot

OK. Pausing for 10 secs...OK

Summary = Cell fixing is a crucial step in various biological techniques, often used to preserve a snapshot of biomolecular transactions in living cells. Paraformaldehyde (PFA) is a common fixative due to its

-----  
Processing item 112/142 = 78.9 %

Processing article = ## Keywords: microbiology and infectious disease, tools and resources, physics of living systems

## Abstract

Uropathogenic Escherichia coli (UPEC) proliferate within superficial bladder umbrella cell

OK. Pausing for 10 secs...OK

Summary = Researchers have developed a bladder-chip model to study urinary tract infections (UTIs) caused by uropathogenic Escherichia coli (UPEC). UPEC forms communities called IBCs inside bladder cells during

-----  
Processing item 113/142 = 79.6 %

Processing article = ## Keywords: computational and systems biology, cell biology

## Abstract

Ammonium ( $\text{NH}_4^+$ ) is essential to generate the nitrogenous building blocks of life. It gets assimilated via the canonical bios

OK. Pausing for 10 secs...OK

Summary = The study focuses on the cellular amination network, which is essential for generating the nitrogenous building blocks of life. The network involves glutamate, a vital amino acid produced by ammonium

-----  
Processing item 114/142 = 80.3 %

Processing article = ## Keywords: microbiology and infectious disease

## Abstract

Reef-building corals thrive in oligotrophic environments due to their possession of endosymbiotic algae. Confined to the low pH interior o

OK. Pausing for 10 secs...OK

Summary = Coral reefs, home to a diverse range of marine life, depend on the relationship between corals and algae called Symbiodiniaceae. These algae live inside corals and provide them with nutrients through

-----  
Processing item 115/142 = 81.0 %

Processing article = ## Keywords: cell biology, developmental biology

## Abstract

Nephronophthisis (NPHP) is a ciliopathy characterized by renal fibrosis and cyst formation, and accounts for a significant portion of en

OK. Pausing for 10 secs...OK

Summary = Nephronophthisis (NPHP) is a genetic kidney disorder that can lead to end stage renal disease in children and young adults. It is caused by mutations in over 25 different genes, including INVS/NPHP2.

-----

Processing item 116/142 = 81.7 %

Processing article = ## Keywords: evolutionary biology

## Abstract

The extinct Steller's sea cow ( *Hydrodamalis gigas*; †1768 ) was a whale-sized marine mammal that manifested profound morphological specializations to expl

OK. Pausing for 10 secs...OK

Summary = The Steller's sea cow, an extinct marine mammal, had a unique mutation in its hemoglobin that may have helped it adapt to the cold waters of the North Pacific. This mutation, a substitution of lysine

-----

Processing item 117/142 = 82.4 %

Processing article = ## Keywords: tools and resources, cell biology, developmental biology

## Abstract

Expression of activated Ras , RasV12 , provides *Drosophila* cultured cells with a proliferation and survival advantage

OK. Pausing for 10 secs...OK

Summary = The text discusses the use of *Drosophila* cells as a model for studying biological processes. *Drosophila* cell lines are useful for this purpose because they are homogenous, scalable, and have distinct

-----

Processing item 118/142 = 83.1 %

Processing article = ## Keywords: cell biology

## Abstract

Multiciliated cells ( MCCs ) are terminally differentiated epithelia that assemble multiple motile cilia used to promote fluid flow . To template these cilia , MC

OK. Pausing for 10 secs...OK

Summary = Multiciliated cells (MCCs) are specialized cells with multiple motile cilia that help move fluids across surfaces in the body. These cilia are templated by centrioles, and MCCs produce many more centr

-----

Processing item 119/142 = 83.8 %

Processing article = ## Keywords: medicine

## Abstract

The human endometrium experiences repetitive cycles of tissue wounding characterised by piecemeal shedding of the surface epithelium and rapid restoration of tissue h

OK. Pausing for 10 secs...OK

Summary = The human endometrium, a complex tissue, goes through regular cycles of wounding and repair. During menstruation, the surface epithelium is shed and the tissue is restored without scarring. A study wa

-----

Processing item 120/142 = 84.5 %

Processing article = ## Keywords: microbiology and infectious disease, genetics and genomics

## Abstract

Adaptation to fluctuating environmental conditions is difficult to achieve . Phase variation mechanisms can overcome

OK. Pausing for 10 secs...OK

Summary = Microbiology and genomics research studies the adaptation of microbes in fluctuating environmental conditions. A study found that a phase variation system in *Burkholderia thailandensis* creates a genot



-----  
Processing item 121/142 = 85.2 %

Processing article = ## Keywords: evolutionary biology

## Abstract

The rate and accuracy of translation hinges upon multiple components - including transfer RNA ( tRNA ) pools , tRNA modifying enzymes , and rRNA molecules

OK. Pausing for 10 secs...OK

Summary = Translation, the process of creating proteins from genetic information, is a crucial process in biology. In bacteria like E. coli, this process is carried out by the translation machinery, which inclu

-----  
Processing item 122/142 = 85.9 %

Processing article = ## Keywords: tools and resources, evolutionary biology, epidemiology and global health

## Abstract

The COVID-19 pandemic has resulted in a step change in the scale of sequencing data , with more genom

OK. Pausing for 10 secs...OK

Summary = The COVID-19 pandemic has led to an explosion of SARS-CoV-2 sequencing data, with more genomes of this virus sequenced than any other organism. This data can be used to create a phylogenetic tree, whi

-----  
Processing item 123/142 = 86.6 %

Processing article = ## Keywords: genetics and genomics, chromosomes and gene expression

## Abstract

RNA molecules are localized to specific subcellular regions through interactions between RNA regulatory elements and RNA

OK. Pausing for 10 secs...OK

Summary = RNA molecules can be found in specific parts of a cell due to interactions between RNA regulatory elements and RNA binding proteins (RBPs). While we often know the details of these interactions in one

-----  
Processing item 124/142 = 87.3 %

Processing article = ## Keywords: computational and systems biology

## Abstract

Aging is often associated with a loss of cell type identity that results in an increase in transcriptional noise in aged tissues . If this ph

OK. Pausing for 10 secs...OK

Summary = Aging is often associated with a loss of cell type identity and an increase in transcriptional noise, but it's unclear if this is a fundamental property of aging. To study this, a new computational to

-----  
Processing item 125/142 = 88.0 %

Processing article = ## Keywords: chromosomes and gene expression, cancer biology

## Abstract

Tumour heterogeneity is thought to be a major barrier to successful cancer treatment due to the presence of drug resistant clon

OK. Pausing for 10 secs...OK

Summary = This text is about a study that explores the use of WILD-seq, a new platform for lineage tracing at the single-cell transcriptomic level, to understand how tumors respond to chemotherapy. The research

-----  
Processing item 126/142 = 88.7 %

Processing article = ## Keywords: genetics and genomics, developmental biology

## Abstract

DMRT1 is the testis-determining factor in several species of vertebrates , but its involvement in mammalian testes differentiation

OK. Pausing for 10 secs...OK

Summary = DMRT1 is a gene that is widely conserved in many species, including mammals, and has been identified as a testis-determining factor in some non-mammalian vertebrates. In mammals, the TDF is the SRY gene

-----

Processing item 127/142 = 89.4 %

Processing article = ## Keywords: genetics and genomics, developmental biology

## Abstract

Obesity induced by high-fat diet ( HFD ) is a multi-factorial disease including genetic , physiological , behavioral , and environmental

OK. Pausing for 10 secs...OK

Summary = Obesity, a global health issue, is caused by various factors including high-fat diets and genetics. A study using the fruit fly, *Drosophila*, found that an enzyme called CTPS, which helps make a molecule

-----

Processing item 128/142 = 90.1 %

Processing article = ## Keywords: developmental biology

## Abstract

During vertebrate embryogenesis , the germ layers are patterned by secreted Nodal signals . In the classical model , Nodals elicit signaling by binding to

OK. Pausing for 10 secs...OK

Summary = In the field of developmental biology, researchers are studying how germ layers in vertebrates are patterned by secreted Nodal signals. Nodal signals are thought to pattern the germ layers by binding to

-----

Processing item 129/142 = 90.8 %

Processing article = ## Keywords: immunology and inflammation, cell biology

## Abstract

Type I interferon ( IFN-I ) -mediated antiviral responses are central to host defense against viral infections . Crucial is the tight

OK. Pausing for 10 secs...OK

Summary = Type I interferon (IFN-I) responses are crucial for the body's defense against viral infections. These responses are tightly controlled and involve the production of IFN-Is by certain cells in the body

-----

Processing item 130/142 = 91.5 %

Processing article = ## Keywords: cell biology, developmental biology

## Abstract

Microtubules ( MTs ) are built from  $\alpha$ -/ $\beta$ -tubulin dimers and used as tracks by kinesin and dynein motors to transport a variety of cargos ,

OK. Pausing for 10 secs...OK

Summary = Microtubules (MTs) are crucial structures in a cell, made up of  $\alpha$ - and  $\beta$ -tubulin heterodimers. There are several types of tubulins, and in *Drosophila*, there are four  $\alpha$ -tubulin genes, with  $\alpha$ Tub84B and

-----

Processing item 131/142 = 92.3 %

Processing article = ## Keywords: genetics and genomics, tools and resources, evolutionary biology

## Abstract

The development of multiple chromosome-scale reference genome sequences in many taxonomic groups has yielded a

OK. Pausing for 10 secs...OK

Summary = GenESPACE is a tool used in genomics and evolutionary biology to study the patterns and processes of molecular evolution. It integrates conserved gene order and

d orthology to define the expected physic

-----

Processing item 132/142 = 93.0 %

Processing article = ## Keywords: microbiology and infectious disease

## Abstract

Most bacteria exist and interact within polymicrobial communities . These interactions produce unique compounds , increase virulence and au

OK. Pausing for 10 secs...OK

Summary = Bacteria often exist in communities made up of multiple species, and the interactions between these species can be either beneficial or harmful. One common community found in clinics is composed of Ps

-----

Processing item 133/142 = 93.7 %

Processing article = ## Keywords: microbiology and infectious disease

## Abstract

The quantity and complexity of data being generated and published in biology has increased substantially , but few methods exist for captur

OK. Pausing for 10 secs...OK

Summary = This text discusses the development of a framework for capturing knowledge about phenotypes derived from molecular interactions between different species, specifically for the Pathogen-Host Interactio

-----

Processing item 134/142 = 94.4 %

Processing article = ## Keywords: genetics and genomics, chromosomes and gene expression

## Abstract

Sequence variation in enhancers that control cell-type-specific gene transcription contributes significantly to phenotyp

OK. Pausing for 10 secs...OK

Summary = Genetic variation in enhancers, which are regions of DNA that control gene transcription, can significantly contribute to differences in human traits. A study used F1-hybrid cells derived from crosses

-----

Processing item 135/142 = 95.1 %

Processing article = ## Keywords: microbiology and infectious disease, immunology and inflammation

## Abstract

Viruses interact with the intracellular transport machinery to promote viral replication . Such host-virus int

OK. Pausing for 10 secs...OK

Summary = In this study, researchers identified a protein called NINL as a key component of the antiviral immune response in humans. NINL is a part of the dynein activating complex, which is involved in the int

-----

Processing item 136/142 = 95.8 %

Processing article = ## Keywords: biochemistry and chemical biology, structural biology and molecular biophysics

## Abstract

Actin isoforms organize into distinct networks that are essential for the normal function of euk

OK. Pausing for 10 secs...OK

Summary = Actin is a crucial protein for the normal function of eukaryotic cells, and small differences in its structure determine how it interacts with other proteins. Researchers have now determined the high-

-----

Processing item 137/142 = 96.5 %

Processing article = ## Keywords: tools and resources, genetics and genomics  
 ## Abstract  
 High-throughput transgenesis using synthetic DNA libraries is a powerful method for systematically exploring genetic function . Dive  
 OK. Pausing for 10 secs...OK  
 Summary = Transgenesis, or the introduction of foreign DNA into genomes, is a crucial tool for genetic analysis and engineering. While it's widely used in single-cell models, its application in multicellular systems is still emerging.

-----  
 Processing item 138/142 = 97.2 %  
 Processing article = ## Keywords: structural biology and molecular biophysics, computational and systems biology  
 ## Abstract  
 To reach their final destinations , outer membrane proteins ( OMPs ) of gram-negative bacteria undergo a complex journey.  
 OK. Pausing for 10 secs...OK  
 Summary = Outer membrane proteins (OMPs) of gram-negative bacteria go through a complex journey from the cytosol to their final destination in the outer membrane. This process involves multiple molecular machines.

-----  
 Processing item 139/142 = 97.9 %  
 Processing article = ## Keywords: neuroscience  
 ## Abstract  
 Midbrain dopamine ( DA ) neurons are key regulators of basal ganglia functions . The axonal domain of these neurons is highly complex , with a large subset of non-dopaminergic axons.  
 OK. Pausing for 10 secs...OK  
 Summary = Dopamine (DA) neurons are important regulators of brain functions such as movement, motivation, and reward-dependent learning. These neurons are located in the ventral tegmental area (VTA) and substantia nigra.

-----  
 Processing item 140/142 = 98.6 %  
 Processing article = ## Keywords: medicine, epidemiology and global health  
 ## Abstract  
 Mobile health ( mHealth ) interventions , which require ownership of mobile phones , are being investigated throughout Africa . We estimate the impact of these interventions.  
 OK. Pausing for 10 secs...OK  
 Summary = Mobile health (mHealth) interventions, which require mobile phone ownership, are being explored in Africa to increase access to healthcare. A study was conducted to estimate the percentage of individuals who own a mobile phone.

-----  
 Processing item 141/142 = 99.3 %  
 Processing article = ## Keywords: medicine, epidemiology and global health  
 ## Abstract  
 Severe acute respiratory syndrome coronavirus 2 ( SARS-CoV-2 ) antibody levels can be used to assess humoral immune responses following infection.  
 OK. Pausing for 10 secs...OK  
 Summary = The text is a research paper about a study on SARS-CoV-2 antibodies in a large group of people from the UK. The study measured the levels of SARS-CoV-2 antibodies in two groups of people, one group of healthcare workers and one group of the general public.

-----

In [15]: `df.head()`

Out[15]:

	article	headings	keywords	id	mixtral_summary
0	Acylation of diverse carbohydrates occurs acro...	[Abstract, Introduction, Results and discussio...	[biochemistry and chemical biology, computatio...	elife-81547-v1	Acylation of carbohydrates is a process that h...
1	Honey bee ecology demands they make both rapid...	[Abstract, Introduction, Results, Discussion, ...	[computational and systems biology]	elife-86176-v2	Honey bees have impressive decision-making ski...
2	Biguanides , including the world's most prescr...	[Abstract, Introduction, Results, Discussion, ...	[genetics and genomics]	elife-82210-v1	Metformin, a commonly prescribed drug for type...
3	Ecological relationships between bacteria medi...	[Abstract, Introduction, Results, Discussion, ...	[microbiology and infectious disease, ecology]	elife-83152-v2	Gut microbiomes are diverse and dynamic commun...
4	Gamma oscillations are believed to underlie co...	[Abstract, Introduction, Results, Discussion, ...	[neuroscience]	elife-83044-v2	Gamma oscillations are a type of brain wave th...

```
In [16]: # check how many rows have blank result (some errors)
empty_df = df.query("mixtral_summary.str.strip() == ''")
print(len(empty_df))
```

0

```
In [17]: # attempt to retry for blank rows (due to some errors)
retry = True
if retry:
    print("Retrying for empty results...")
    for i in range(len(df)):
        item = df.iloc[i]
        if item["mixtral_summary"] == "":
            print("Item =", i)
            # text = item["article"]
            text = build_text_with_headings(item)
            df.at[i, "mixtral_summary"] = send_sumarize_request(text[:text_cap], qu
            print("Completed")
```

Retrying for empty results...

Completed

```
In [18]: output_path = "./data/output/test_set/"
output_filename = "elife_groq_mixtral_summary.csv"

print("Writing to file ", output_filename)
df.to_csv(output_path+output_filename,
          index = False)
```

```
)  
print("Completed")
```

Writing to file elife\_groq\_mixtral\_summary.csv  
Completed

```
In [19]: output_filename = "elife_groq_mixtral_summary.json"  
  
print("Writing to file ", output_filename)  
df.to_json(output_path+output_filename,  
           orient="records",  
           )  
print("Completed")
```

Writing to file elife\_groq\_mixtral\_summary.json  
Completed

```
In [20]: # process the PLOS dataset:  
dev_df_filename = "../.../data/biolaysumm2024_data/PLOS_test.jsonl"  
df = pd.read_json(dev_df_filename,  
                  orient="records",  
                  lines=True  
                  )  
  
df.head()
```

Out[20]:	article	headings	keywords	id
0	Lung-resident ( LR ) mesenchymal stem and stro...	[Abstract, Introduction, Results, Discussion, ...]	[immune system, medical conditions, molecular ...]	journal.ppat.1009789
1	Visceral leishmaniasis ( VL ) is endemic in So...	[Abstract, Introduction, Methods, Results, Dis...]	[neonates, clinical laboratory sciences, trans...]	journal.pntd.0007992
2	A high burden of Salmonella enterica subspecie...	[Abstract, Introduction, Methods, Results, Dis...]	[pathogens, medical conditions, taxonomy, bact...]	journal.pntd.0010704
3	Severe Acute Respiratory Syndrome Coronavirus-...	[Abstract, Introduction, Results, Discussion, ...]	[pathogens, amniotes, medical conditions, bind...]	journal.ppat.1010691
4	Many fungal species utilize hydroxyderivatives...	[Abstract, Introduction, Results and discussio...]	[taxonomy, proteins, chemistry, genetics, enzy...]	journal.pgen.1009815

```
In [21]: # # apply to all rows in eval miniset  
# text_cap = 20_000 # temporarily limit to 20k characters due to API restriction,  
  
# print("Summarization process started...")
```

```
# df["groq_mistral_summary"] = df["article"].apply(lambda text: send_sumarize_reque
# print("Completed")
```

```
In [22]: # create empty column
df["mixtral_summary"] = df["article"].apply(lambda x: "")
df.head()
n = len(df)

for i in range(len(df)):
    print(f"Processing item {i}/{n} = {i*100/n:.1f} %")
    item = df.iloc[i]
    text = build_text_with_headings(item)
    print("Processing article =", text[:200])
    summary = send_sumarize_request(text)
    print("Summary =", summary[:200])
    # df.at[i, "llm_templates"] = parsed_item
    df.at[i, "mixtral_summary"] = summary
    print("-----")
```

Processing item 0/142 = 0.0 %

Processing article = ## Keywords: immune system, medical conditions, molecular development, health care, developmental biology, cellular types, immunology, pediatrics, respiratory physiology, cell biology, pediatric infec

OK. Pausing for 10 secs...OK

Summary = Lung-resident mesenchymal stem cells (LR-MSCs) play a crucial role in maintaining lung health and regenerating lung tissue after injury. They are found in the alveolar niche and can interact with othe

Processing item 1/142 = 0.7 %

Processing article = ## Keywords: neonates, clinical laboratory sciences, transfusion medicine, parasitic diseases, drugs, blood transfusion, developmental biology, women's health, hematology, protozoan infections, pharma

OK. Pausing for 10 secs...OK

Summary = Visceral Leishmaniasis (VL), a parasitic disease, is endemic in South Sudan and can cause severe complications for pregnant women. A study was conducted to describe the characteristics and outcomes of

Processing item 2/142 = 1.4 %

Processing article = ## Keywords: pathogens, medical conditions, taxonomy, bacteria, geography, genetics, kenya, organisms, phylogenetic analysis, data management, biogeography, pharmacology, genomics, salmonella, antibio

OK. Pausing for 10 secs...OK

Summary = Salmonella Typhi, a bacterium that causes typhoid fever, is a significant public health concern worldwide. A study was conducted in Kibera, an urban informal settlement in Nairobi, Kenya, to understand

Processing item 3/142 = 2.1 %

Processing article = ## Keywords: pathogens, amniotes, medical conditions, binding analysis, cell binding assay, model organisms, proteins, chemical characterization, rodents, animal models, mammals, research and analysis

OK. Pausing for 10 secs...OK

Summary = The COVID-19 pandemic, caused by the SARS-CoV-2 virus, has highlighted the vulnerability of humanity to novel coronaviruses. While vaccines are available, they have not been administered widely enough

Processing item 4/142 = 2.8 %

Processing article = ## Keywords: taxonomy, proteins, chemistry, genetics, enzyme chemistry, organisms, phylogenetic analysis, eukaryota, physical sciences, dna-binding proteins, enzyme metabolism, data management, bioche

OK. Pausing for 10 secs...OK

Summary = Lay Summary:

Candida parapsilosis is a type of yeast that can break down certain compounds derived from benzene and benzoic acid, using the 3-oxoadipate (3-OAP) and gentisate pathways. These pathways

Processing item 5/142 = 3.5 %

Processing article = ## Keywords: neurophysiology, systems science, working memory, engineering and technology, memory, mathematics, physical sciences, cellular types, cognitive neuroscience, cognition, synapses, anatomy,

OK. Pausing for 10 secs...OK

Summary = The text discusses a method called "oracle-supervised Neural Engineering Framework" (osNEF) used to train biologically-detailed spiking neural networks to perform cognitive tasks. The networks are tra

Processing item 6/142 = 4.2 %



Processing article = ## Keywords: entropy, white blood cells, genome analysis, thermodynamics, genetics, research and analysis methods, t cells, transcriptome analysis, departures from diploidy, mathematics, physical science

OK. Pausing for 10 secs...OK

Summary = Down syndrome (DS) is a genetic disorder caused by an extra copy of human chromosome 21 (HSA21), resulting in a genome-wide transcriptomic deregulation that impacts different tissues and cell types. A

-----  
Processing item 7/142 = 4.9 %

Processing article = ## Keywords: imaging techniques, brain mapping, neurophysiology, brain electrophysiology, clinical neurophysiology, research and analysis methods, electrophysiological techniques, neuronal dendrites,

OK. Pausing for 10 secs...OK

Summary = Brain activity can be recorded and analyzed through various imaging techniques such as EEG, MEG, and iEEG. These techniques capture the electrical activity of the brain. Neural oscillations, a type of

-----  
Processing item 8/142 = 5.6 %

Processing article = ## Keywords: signal transduction, cell signaling, network analysis, interferons, proteins, genetics, cellular types, anatomy, signaling networks, virus effects on host gene expression, biochemistry, g

OK. Pausing for 10 secs...OK

Summary = Crimean-Congo hemorrhagic fever virus (CCHFV) is a tick-borne virus that can cause severe hemorrhagic fever in humans with a high fatality rate. Although discovered over 70 years ago, our understanding

-----  
Processing item 9/142 = 6.3 %

Processing article = ## Keywords: color vision, vision, decision making, working memory, memory, schizophrenia, attention, cognitive neuroscience, cognition, cognitive psychology, perception, social sciences, sensory perception

OK. Pausing for 10 secs...OK

Summary = Schizophrenia is a severe mental disorder that affects perception and cognitive behavior, including working memory (WM) deficits. WM is a crucial function that temporally stores and manipulates information

-----  
Processing item 10/142 = 7.0 %

Processing article = ## Keywords: pathogens, lentivirus, cell membranes, hiv, immunologic techniques, proteins, amino acids, chemistry, arginine, membrane fusion, research and analysis methods, organisms, viral transmission

OK. Pausing for 10 secs...OK

Summary = This text is about a study on the HIV-1 virus and its envelope glycoprotein (Env), which plays a crucial role in the virus' entry into host cells by binding to specific receptors and undergoing conformational changes

-----  
Processing item 11/142 = 7.7 %

Processing article = ## Keywords: computer networks, cancers and neoplasms, clinical research design, research design, cluster analysis, survival analysis, research and analysis methods, applied mathematics, engineering and technology

OK. Pausing for 10 secs...OK

Summary = The CONsensus Molecular SUBtype of Cancer (COMSUC) web server is a new tool that allows users to explore consensus molecular subtypes of over 30 cancers using eight clustering methods and five types of

-----  
Processing item 12/142 = 8.5 %

Processing article = ## Keywords: antarctica, pathogens, bunyaviruses, taxonomy, tick-borne encephalitis virus, geography, genetics, organisms, phylogenetic analysis, b

ioenergetics, animal migration, mitochondria, data ma

OK. Pausing for 10 secs...OK

Summary = Ticks are blood-feeding organisms that are widely distributed and can adapt to various environments and host animals. They can carry many types of viruses, some of which can cause diseases in humans a

-----  
Processing item 13/142 = 9.2 %

Processing article = ## Keywords: nucleic acids, genome analysis, model organisms, genetics, research and analysis methods, organisms, genotyping, eukaryota, yeast and fungal models, algae, experimental organism systems,

OK. Pausing for 10 secs...OK

Summary = RecombineX is a new computational tool designed for automated, high-throughput analysis of tetrads, which are groups of four cells produced by meiosis in sexually reproducing organisms. It is a Linux-

-----  
Processing item 14/142 = 9.9 %

Processing article = ## Keywords: white blood cells, pathogens, lentivirus, amniotes, hiv, primates, proteins, transfection, t cells, research and analysis methods, genetics, mammals, organisms, hiv-1, eukaryota, cellular

OK. Pausing for 10 secs...OK

Summary = White blood cells, specifically T cells, have proteins called APOBEC3 (A3) that can block certain viruses, like HIV-1, by introducing mutations into their genetic material. There are seven types of A3

-----  
Processing item 15/142 = 10.6 %

Processing article = ## Keywords: cancers and neoplasms, genetics, research and analysis methods, cancer genetics, oncology, carcinogenesis, genomics, cancer evolution, simulation and modeling, surgical and invasive medic

OK. Pausing for 10 secs...OK

Summary = This text is about a new method for studying the evolution of tumors using genetic data. Tumors are made up of many different cells, and these cells can change and evolve over time. This can make it h

-----  
Processing item 16/142 = 11.3 %

Processing article = ## Keywords: agriculture, medical conditions, glossina, veterinary parasitology, entomology, invertebrates, protozoans, parasitic diseases, trypanosoma brucei, organisms, eukaryota, species interactio

OK. Pausing for 10 secs...OK

Summary = Trypanosomes are parasites that cause significant diseases in both humans and animals in Africa. These parasites are transmitted mainly by tsetse flies, and they are responsible for human African trypan

-----  
Processing item 17/142 = 12.0 %

Processing article = ## Keywords: paleontology, genome analysis, invertebrates, genetics, research and analysis methods, organisms, principal component analysis, eukaryota, mathematics, physical sciences, animal genomics,

OK. Pausing for 10 secs...OK

Summary = Schistosoma is a type of parasitic flatworm that can infect humans and animals, and it can cause diseases like schistosomiasis. There are different species of Schistosoma, including S. haematobium and

-----  
Processing item 18/142 = 12.7 %

Processing article = ## Keywords: imaging techniques, gastrointestinal tract, pathogens, lentivirus, lymphatic system, lymph nodes, hiv, proteins, rectum, research and analysis methods, organisms, positron emission tomogr

OK. Pausing for 10 secs...OK

Summary = The study explores the potential of human immunodeficiency virus (HIV) vaccines, focusing on dimeric IgA (dIgA) in the form of secretory IgA. The researchers used Positron Emission Tomography (PET) im

-----  
Processing item 19/142 = 13.4 %

Processing article = ## Keywords: imaging techniques, computer software, blood flow, research and analysis methods, applied mathematics, engineering and technology, software engineering, mathematics, physical sciences, an

OK. Pausing for 10 secs...OK

Summary = Iliski is a new, open-source software that can be used to calculate Transfer Functions (TFs) between two signals. This is particularly useful in the field of neurovascular imaging, where it is importa

-----  
Processing item 20/142 = 14.1 %

Processing article = ## Keywords: pathogens, neonatology, research design, qualitative studies, research and analysis methods, age groups, organisms, diet, children, women's health, food, clinical genetics, pediatrics, do

OK. Pausing for 10 secs...OK

Summary = The text is a qualitative evidence synthesis examining the values and preferences of pregnant women, mothers, families, and healthcare workers concerning infant feeding options when there are feeding

-----  
Processing item 21/142 = 14.8 %

Processing article = ## Keywords: white blood cells, pathogens, lentivirus, antigen-presenting cells, hiv, proteins, engineering and technology, organisms, nanotechnology, b cells, antibody-producing cells, cellular types

OK. Pausing for 10 secs...OK

Summary = White blood cells, specifically B cells, are crucial to our immune system as they produce antibodies to fight off pathogens. A research team has been studying the use of nanoparticles to present viral

-----  
Processing item 22/142 = 15.5 %

Processing article = ## Keywords: dna construction, proteins, transfection, research and analysis methods, immunoblotting, immunology, biochemistry, post-translational modification, molecular probe techniques, immunopreci

OK. Pausing for 10 secs...OK

Summary = Pseudorabies virus (PRV) is a harmful virus that can infect a variety of animals, including swine and humans. It has several ways to evade the immune system, one of which involves a protein called UL1

-----  
Processing item 23/142 = 16.2 %

Processing article = ## Keywords: medical conditions, entomology, invertebrates, genetics, organisms, eukaryota, saliva, species interactions, anatomy, population biology, viral replication, viral diseases, arthropoda, vi

OK. Pausing for 10 secs...OK

Summary = Mosquitoes are carriers for various diseases, including Zika virus, which is transmitted to humans through their bites. The text discusses research on how different environmental factors, such as temp

-----  
Processing item 24/142 = 16.9 %

Processing article = ## Keywords: signal transduction, sensory neurons, cell signaling, normal distribution, cell motility, probability theory, afferent neurons, mathematics, physical sciences, cellular types, cell biolog

OK. Pausing for 10 secs...OK

Summary = Olfaction, or the sense of smell, is important for survival in many animals as it helps them detect food, mates, or predators. In the case of the tiny worm Ca

enorhabditis elegans, it is attracted to c

-----

Processing item 25/142 = 17.6 %

Processing article = ## Keywords: schistosoma mansoni, graduates, medical conditions, eggs, proteins, granulomas, invertebrates, parasitic diseases, organisms, public and occupational health, preventive medicine, eukaryot

OK. Pausing for 10 secs...OK

Summary = Schistosomiasis, caused by the parasite Schistosoma mansoni, affects millions of people worldwide and has a significant impact on public health. The current treatment relies solely on the drug praziqu

-----

Processing item 26/142 = 18.3 %

Processing article = ## Keywords: pathogens, research facilities, bacteria, research and analysis methods, organisms, oncology, eukaryota, histology, anatomy, surgical and invasive medical procedures, fungi, cancer detect

OK. Pausing for 10 secs...OK

Summary = Mycetoma is a neglected tropical disease that is relatively rare in Uganda, with an estimated 3,683 people living with the disease. However, there is a lack of data on mycetoma and its complications i

-----

Processing item 27/142 = 19.0 %

Processing article = ## Keywords: amniotes, medical conditions, primates, mammals, parasitic diseases, protozoans, organisms, parasite groups, eukaryota, malaria, asia, old world monkeys, malarial parasites, vertebrates,

OK. Pausing for 10 secs...OK

Summary = Simian malaria, caused by Plasmodium parasites, is a public health concern in Malaysia. While Malaysia has successfully eliminated human malaria transmission, simian malaria parasites like Plasmodium

-----

Processing item 28/142 = 19.7 %

Processing article = ## Keywords: nucleic acids, hypoxia, dna construction, proteins, bioinformatics, sequence motif analysis, genetics, research and analysis methods, dna replication, database and informatics methods, se

OK. Pausing for 10 secs...OK

Summary = Epstein-Barr Virus (EBV) is a common virus that affects many people worldwide, and is associated with several cancers including lymphomas and epithelial tumors. EBNA1 is a protein that is consistently

-----

Processing item 29/142 = 20.4 %

Processing article = ## Keywords: plant anatomy, pistils, genetic engineering, genetics, plant science, genetically modified organisms, engineering and technology, plant biotechnology, organisms, departures from diploidy,

OK. Pausing for 10 secs...OK

Summary = In this study, researchers aimed to improve the understanding of genes regulating grain yield in wheat, a vital staple crop. They focused on a specific region of chromosome 7AL, previously identified

-----

Processing item 30/142 = 21.1 %

Processing article = ## Keywords: topography, scrub typhus, amniotes, landforms, islands, rodents, invertebrates, mammals, genetics, organisms, eukaryota, mites, geomorphology, south america, population biology, vertebrat

OK. Pausing for 10 secs...OK

Summary = Scrub typhus is an emerging disease caused by bacteria transmitted through the bites of infected chigger mites, particularly in the Asia-Pacific region. However, recent cases in the Arabian Peninsula

-----

Processing item 31/142 = 21.8 %

Processing article = ## Keywords: epidemiology, pathogens, medical conditions, zika fever, dengue fever, chikungunya virus, organisms, chikungunya infection, viral pathogens, viral diseases, flaviviruses, viruses, infecti

OK. Pausing for 10 secs...OK

Summary = During the 2015-2017 Zika epidemic in Latin America and the Caribbean, dengue and chikungunya, two other viral diseases with the same mosquito vector, were also circulating. Due to similar symptoms, i

Processing item 32/142 = 22.5 %

Processing article = ## Keywords: nucleic acids, pathogens, cell cultures, sequencing techniques, rna sequencing, research and analysis methods, genetics, dna replication, organisms, hepatitis viruses, molecular biology a

OK. Pausing for 10 secs...OK

Summary = Hepatitis B Virus (HBV) is a DNA virus that causes both acute and chronic infections, with chronic hepatitis B (CHB) affecting over 257 million people. HBV has an enveloped virion containing a partial

Processing item 33/142 = 23.2 %

Processing article = ## Keywords: nucleic acids, amniotes, bacteria, rodents, animal models, mammals, genetics, research and analysis methods, organisms, eukaryota, experimental organism systems, genomics, pharmacology, b

OK. Pausing for 10 secs...OK

Summary = Tuberculosis (TB) is a significant public health concern, worsened by the emergence of drug-resistant TB. The bacterium causing TB, Mycobacterium tuberculosis (Mtb), has a GC-rich genome, making it hi

Processing item 34/142 = 23.9 %

Processing article = ## Keywords: information theory, network analysis, research facilities, research and analysis methods, archives, mathematics, circadian rhythms, cellular types, physical sciences, biochemistry, cell b

OK. Pausing for 10 secs...OK

Summary = The suprachiasmatic nucleus (SCN), the body's "master clock," is made up of around 10,000 clock cells that vary in their circadian periods. These cells maintain coherence despite their differences, li

Processing item 35/142 = 24.6 %

Processing article = ## Keywords: imaging techniques, linguistics, brain mapping, vision, magnetic resonance imaging, research and analysis methods, phonetics, diagnostic radiology, cognitive linguistics, lexicography, co

OK. Pausing for 10 secs...OK

Summary = The text discusses a model called the Lexical Categorization Model (LCM), which explains the role of a specific part of the brain, the left-ventral occipito-temporal cortex (lvOT), in reading. The LCM

Processing item 36/142 = 25.4 %

Processing article = ## Keywords: immune system, molecular development, genetics, developmental biology, circadian rhythms, immunology, signs and symptoms, biochemistry, gene expression, circadian oscillators, cytokines,

OK. Pausing for 10 secs...OK

Summary = The text discusses the role of the immune system and circadian rhythms in inflammation. Circadian rhythms, which are biological oscillations with a period of approximately 24 hours, have been found to

Processing item 37/142 = 26.1 %

Processing article = ## Keywords: pathogens, medical conditions, human capital, deng

ue fever, forecasting, health care, research and analysis methods, population metrics, organisms, mathematics, physical sciences, asia, d

OK. Pausing for 10 secs...OK

Summary = Dengue fever is a significant public health issue in Saudi Arabia, with the first case reported in 1993. The disease has since become endemic in the western and southern regions of the country, with J

-----

Processing item 38/142 = 26.8 %

Processing article = ## Keywords: gastrointestinal tract, proteins, cytoskeletal proteins, genetics, lipids, chromosome biology, chromatin, reproductive system, lamins, cellular types, anatomy, biochemistry, cell biology,

OK. Pausing for 10 secs...OK

Summary = Nuclear lipid droplets (nLDs) are fat storage sites found inside the nuclei of cells. While they have been observed in some cells, their function is not well understood. They have been suggested to pr

-----

Processing item 39/142 = 27.5 %

Processing article = ## Keywords: imaging techniques, cell membranes, cell processes, research and analysis methods, developmental biology, staining, cellular types, cell biology, cell death, cellular neuroscience, fluore

OK. Pausing for 10 secs...OK

Summary = Bacterial meningitis, a serious inflammation of the membranes covering the brain and spinal cord, can lead to significant neurological damage. One major cause of bacterial meningitis is Streptococcus

-----

Processing item 40/142 = 28.2 %

Processing article = ## Keywords: animal physiology, mapk signaling cascades, signal transduction, vertebrate physiology, cell signaling, genome analysis, model organisms, sequencing techniques, genetics, rna sequencing,

OK. Pausing for 10 secs...OK

Summary = In this study, scientists used fighting fish (Betta splendens) to understand how fighting behavior is controlled at the physiological level. They observed that fighting in these fish proceeds in a spe

-----

Processing item 41/142 = 28.9 %

Processing article = ## Keywords: nucleic acids, pathogens, macromolecular structure analysis, research and analysis methods, organisms, viral transmission and infection, west nile virus, dengue virus, rna structure, bioc

OK. Pausing for 10 secs...OK

Summary = The text discusses the discovery and analysis of a new strain of a virus named OmRV-LZ, which is a type of totivirus. Totiviruses have a non-segmented double-stranded RNA (dsRNA) genome and are typica

-----

Processing item 42/142 = 29.6 %

Processing article = ## Keywords: signal transduction, polymerase chain reaction, artificial gene amplification and extension, cell processes, cell signaling, phenotypes, genetics, research and analysis methods, developme

OK. Pausing for 10 secs...OK

Summary = Ric-8B is a protein that interacts with certain types of G proteins and can act as a guanine nucleotide exchange factor. It is widely expressed in the brain and olfactory neurons, where it plays a rol

-----

Processing item 43/142 = 30.3 %

Processing article = ## Keywords: epidemiology, pathogens, population size, medical conditions, entomology, dengue fever, invertebrates, genetics, research and analysis methods, population metrics, organisms, eukaryota, s

OK. Pausing for 10 secs...OK

Summary = Dengue fever is a common vector-borne disease that affects millions of people each year, particularly in Southeast Asia and the Western Pacific. The disease is transmitted through the bite of infected

-----  
Processing item 44/142 = 31.0 %

Processing article = ## Keywords: leprosy, medical conditions, anthropology, research design, geography, human families, research and analysis methods, health care, language, pilot studies, human geography, cognitive psyc

OK. Pausing for 10 secs...OK

Summary = Leprosy is a chronic infectious disease that can cause damage to the skin and peripheral nerves, leading to impairments if diagnosed late. The disease, along with social myths, attitudes, cultural and

-----  
Processing item 45/142 = 31.7 %

Processing article = ## Keywords: cell processes, model organisms, chromosome structure and function, proteins, research and analysis methods, organisms, chromosome biology, chromosomes, immunoblotting, eukaryota, yeast a

OK. Pausing for 10 secs...OK

Summary = Shugoshin, a protein found in budding yeast, plays a crucial role in correct chromosome segregation during cell division. It does this by helping to correctly attach sister chromatids to the mitotic s

-----  
Processing item 46/142 = 32.4 %

Processing article = ## Keywords: white blood cells, pathogens, proteins, forecasting, t cells, research and analysis methods, applied mathematics, organisms, clinical immunology, separation processes, mathematics, physic

OK. Pausing for 10 secs...OK

Summary = Lay Summary:

White blood cells, specifically T cells, play a crucial role in identifying and attacking pathogens in the human body. To enable applications like vaccine design, cancer neoantigen identi

-----  
Processing item 47/142 = 33.1 %

Processing article = ## Keywords: dna modification, nucleic acids, genome analysis, genetics, dna replication, chromosome biology, chromatin, body mass index, cellular types, anatomy, genomics, cell biology, gene expressi

OK. Pausing for 10 secs...OK

Summary = This text is a research paper about the study of DNA modification, specifically DNA methylation, in relation to genome analysis, genetics, and cell biology. The researchers used blood and sperm sample

-----  
Processing item 48/142 = 33.8 %

Processing article = ## Keywords: cell processes, engineering and technology, cellular types, anatomy, fluidics, axons, secretory pathway, cell biology, cellular neuroscience, central nervous system, neurons, endocytosis,

OK. Pausing for 10 secs...OK

Summary = Reovirus is a type of virus that can infect the central nervous system (CNS), causing encephalitis in newborn animals. It has a broad host range and can cause age-restricted disease in humans. Reoviru

-----  
Processing item 49/142 = 34.5 %

Processing article = ## Keywords: medical conditions, proteins, crystallography, genetics, parasitic diseases, vaccine development, tropical diseases, public and occupational health, crystal structure, parasite groups, pr

OK. Pausing for 10 secs...OK

Summary = Vaccines are crucial for controlling malaria, a major global health concern with an estimated 229 million cases and 409,000 deaths in 2019. However, drug-resistance is a growing issue, particularly in

-----  
Processing item 50/142 = 35.2 %

Processing article = ## Keywords: pathogens, cell membranes, gene expression and vector techniques, proteins, amino acids, chemistry, cytoskeletal proteins, protozoans, research and analysis methods, organisms, eukaryota,

OK. Pausing for 10 secs...OK

Summary = Lysine malonylation is a post-translational modification that regulates many cellular processes in the parasite *Toxoplasma gondii*. This modification varies between different strains of the parasite, w

-----  
Processing item 51/142 = 35.9 %

Processing article = ## Keywords: signal transduction, *Caenorhabditis elegans*, nicotinic acetylcholine receptors, alkaloids, model organisms, amphibians, proteins, *Caenorhabditis*, chemistry, nicotine, invertebrates, anima

OK. Pausing for 10 secs...OK

Summary = The text provided discusses research on nematodes, which are multicellular organisms that exhibit complex behaviors controlled by a neuromuscular system. Nematodes have a greater number of nicotinic a

-----  
Processing item 52/142 = 36.6 %

Processing article = ## Keywords: white blood cells, pathogens, lentivirus, flow cytometry, hiv, cell motility, t cells, research and analysis methods, organisms, hiv-1, spectrum analysis techniques, cellular types, cytop

OK. Pausing for 10 secs...OK

Summary = HIV-1, the virus that causes AIDS, infects cells called CD4 T lymphocytes (CD4TL) by binding to chemokine receptors CCR5 or CXCR4. The CXCR4-using viruses are considered more harmful and are linked to

-----  
Processing item 53/142 = 37.3 %

Processing article = ## Keywords: amniotes, agriculture, medical conditions, leishmania, atmospheric science, entomology, vector-borne diseases, invertebrates, mammals, protozoans, parasitic diseases, organisms, eukaryota

OK. Pausing for 10 secs...OK

Summary = Zoonotic cutaneous leishmaniasis (ZCL) is a disease in Palestine, spread by sand flies that live in hyrax dens. These sand flies carry the *Leishmania* parasite, which causes ZCL. Even though control me

-----  
Processing item 54/142 = 38.0 %

Processing article = ## Keywords: nucleic acids, macromolecular structure analysis, viral genomics, microbial genomics, transfection, genetics, research and analysis methods, rna structure, genomics, biochemistry, foot an

OK. Pausing for 10 secs...OK

Summary = Pseudoknot (PK) structures in the non-coding regions of foot-and-mouth disease virus (FMDV) RNA genomes are important for gene expression regulation. The 5' untranslated region (5' UTR) of FMDV RNA is

-----  
Processing item 55/142 = 38.7 %

Processing article = ## Keywords: taste, schools, medical conditions, parasitic diseases, age groups, helminth infections, children, cognitive psychology, pediatrics, perception, sensory perception, social sciences, cogni

OK. Pausing for 10 secs...OK

Summary = This text is about a study on the palatability of two new orally disintegrating tablet (ODT) formulations of a drug called Praziquantel (PZQ), which is used t



o treat schistosomiasis, a parasitic disease

-----

Processing item 56/142 = 39.4 %

Processing article = ## Keywords: pathogens, herpes simplex virus, organisms, physical sciences, waves, physics, dna viruses, cytosol, cell biology, viral pathogens, viral packaging, viral replication, viruses, scattering

OK. Pausing for 10 secs...OK

Summary = A significant challenge in healthcare is drug resistance in viruses. A research team has discovered a new way to treat all human and animal herpesviruses, as well as other viral infections with genome

-----

Processing item 57/142 = 40.1 %

Processing article = ## Keywords: topography, medical conditions, landforms, islands, wuchereria bancrofti, artificial gene amplification and extension, polymerase chain reaction, terrestrial environments, entomology, vector

OK. Pausing for 10 secs...OK

Summary = The Nancowry group of islands, located in the Indian territory, are home to a population of 7674 people. These islands are covered in forests and are prone to mosquito-borne diseases. A study was conducted

-----

Processing item 58/142 = 40.8 %

Processing article = ## Keywords: nucleic acids, pathogens, signal transduction, cell signaling, rhabdoviruses, interferons, non-coding rna, proteins, chemistry, genetics, research and analysis methods, chemical reactions

OK. Pausing for 10 secs...OK

Summary = USP12 is a protein that interacts with another protein called CBP, which is involved in various cellular functions such as proliferation, cell cycle, differentiation, apoptosis, and DNA damage response

-----

Processing item 59/142 = 41.5 %

Processing article = ## Keywords: schools, pathogens, polymerase chain reaction, artificial gene amplification and extension, caribbean, sequencing techniques, research and analysis methods, chikungunya virus, age groups,

OK. Pausing for 10 secs...OK

Summary = A study has discovered two new viruses, Melao virus (MELV) and Oropouche virus (OROV), in children from Haiti with acute febrile illness. These viruses were previously unknown to cause illness in humans

-----

Processing item 60/142 = 42.3 %

Processing article = ## Keywords: pathogens, amniotes, medical conditions, immunologic techniques, proteins, rodents, mammals, research and analysis methods, organisms, public and occupational health, preventive medicine,

OK. Pausing for 10 secs...OK

Summary = The COVID-19 pandemic, caused by the SARS-CoV-2 virus, has had significant economic impacts and resulted in over 2.6 million deaths worldwide as of March 2021. Vaccines have been approved for emergency

-----

Processing item 61/142 = 43.0 %

Processing article = ## Keywords: public and occupational health, geographical locations, people and places, lymphedema, health services administration and management, signs and symptoms, medicine and health sciences, health

OK. Pausing for 10 secs...OK

Summary = Lymphatic filariasis (LF) is a neglected tropical disease that can cause long-term disability, including lymphedema, elephantiasis, and hydrocele. While mass drug administration (MDA) has been successful

-----

Processing item 62/142 = 43.7 %

Processing article = ## Keywords: white blood cells, pathogens, cell membranes, flow cytometry, dna construction, transfection, t cells, research and analysis methods, organisms, staining, spectrum analysis techniques, ce

OK. Pausing for 10 secs...OK

Summary = White blood cells, also known as leukocytes, are crucial for immune response in humans. They help to protect the body from harmful pathogens such as bacteria and viruses. In this text, the focus is on

Processing item 63/142 = 44.4 %

Processing article = ## Keywords: operator theory, sensory deprivation, tinnitus, recurrent neural networks, mathematics, physical sciences, cellular types, cognitive psychology, otorhinolaryngology, perception, hearing d

OK. Pausing for 10 secs...OK

Summary = Tinnitus, a common form of auditory hallucination, can be caused by hearing loss or damage to higher auditory processing pathways. It is related to brain activity in sensory networks that occurs in th

Processing item 64/142 = 45.1 %

Processing article = ## Keywords: amniotes, pathogens, model organisms, immunofluorescence staining, nasal cavity, rodents, animal models, mammals, research and analysis methods, organisms, mouse models, staining, eukaryo

OK. Pausing for 10 secs...OK

Summary = Porcine hemagglutinating encephalomyelitis virus (PHEV) is a type of coronavirus that can infect the nervous system of mice, causing symptoms such as loss of smell and taste, lethargy, and neurological

Processing item 65/142 = 45.8 %

Processing article = ## Keywords: glucose metabolism, cell processes, life cycles, chemistry, autophagic cell death, carbohydrate metabolism, developmental biology, lipids, glucose, physical sciences, cell biology, cell d

OK. Pausing for 10 secs...OK

Summary = Glucose metabolism is crucial for the health and development of animals and humans. It is especially important for tissues like the brain and red blood cells, which rely on glucose for energy. Glucose

Processing item 66/142 = 46.5 %

Processing article = ## Keywords: glycosylation, pathogens, macromolecular structure analysis, biochemical simulations, proteins, chemistry, organisms, physical sciences, immunology, biochemistry, viral pathogens, post-tr

OK. Pausing for 10 secs...OK

Summary = The SARS-CoV-2 spike (S) protein is the primary target of COVID-19 vaccines and plays a crucial role in viral entry into host cells. Researchers performed extensive molecular dynamics simulations of a

Processing item 67/142 = 47.2 %

Processing article = ## Keywords: amniotes, pathogens, medical conditions, interferons, proteins, rodents, mammals, research and analysis methods, organisms, prophylaxis, public and occupational health, eukaryota, virus t

OK. Pausing for 10 secs...OK

Summary = Type I interferons (IFNs) are important for protecting the body against viruses, including SARS-CoV-2, the virus that causes COVID-19. In a study using Syrian hamsters, researchers found that giving I

Processing item 68/142 = 47.9 %

Processing article = ## Keywords: schistosoma mansoni, medical conditions, snails, c

hemistry, invertebrates, parasitic diseases, research and analysis methods, analytical chemistry, organisms, aquatic environments, helmin

OK. Pausing for 10 secs...OK

Summary = Schistosomiasis is a parasitic disease caused by blood flukes of the genus Schistosoma, and it affects over 250 million people worldwide. It is transmitted through freshwater snails, which serve as in

-----

Processing item 69/142 = 48.6 %

Processing article = ## Keywords: fluorescence microscopy, vegetables, optical lenses, protozoans, research and analysis methods, organisms, engineering and technology, giardia, cell phones, light microscopy, parasite gro

OK. Pausing for 10 secs...OK

Summary = A study was conducted to develop a smartphone-based microscopic method for detecting Cryptosporidium and Giardia cysts in vegetables and water samples. The method involved using a sapphire ball lens a

-----

Processing item 70/142 = 49.3 %

Processing article = ## Keywords: epidemiology, amniotes, medical conditions, geography, parasitic diseases, mammals, organisms, helminth infections, public and occupational health, eukaryota, tibet, asia, china, vertebra

OK. Pausing for 10 secs...OK

Summary = Echinococcosis is a parasitic disease that affects people in China, particularly in the western region. It is caused by larval stages of tapeworms from the genus Echinococcus. The two main forms found

-----

Processing item 71/142 = 50.0 %

Processing article = ## Keywords: epidemiology, schools, network analysis, research design, research and analysis methods, public and occupational health, preventive medicine, mathematics, physical sciences, immunology, v

OK. Pausing for 10 secs...OK

Summary = This text discusses the impact of different immunization strategies on preventing or mitigating outbreaks of contagion in a population. Two commonly used strategies are degree immunization, where node

-----

Processing item 72/142 = 50.7 %

Processing article = ## Keywords: operator theory, linear algebra, neurophysiology, electrode potentials, chemistry, research and analysis methods, applied mathematics, engineering and technology, electrochemistry, mathem

OK. Pausing for 10 secs...OK

Summary = Lay Summary:

This text discusses the use of extracellular recording to study brain physiology and pathology. Extracellular potential is a result of ion motion in tissue, driven by ionic currents thr

-----

Processing item 73/142 = 51.4 %

Processing article = ## Keywords: epidemiology, medical conditions, parasitic diseases, co-infections, protozoans, organisms, parasite groups, eukaryota, cellular types, malaria, cell biology, plasmodium falciparum, malar

OK. Pausing for 10 secs...OK

Summary = Malaria is a serious disease caused by Plasmodium parasites, which are transmitted to people through the bites of infected mosquitoes. In this study, the researchers investigated the interactions betw

-----

Processing item 74/142 = 52.1 %

Processing article = ## Keywords: dna modification, epidemiology, nucleic acids, can

cers and neoplasms, chemistry, genetics, health care, chemical reactions, chromosome biology, oncology, chromatin, physical sciences, gen

OK. Pausing for 10 secs...OK

Summary = DNA modification plays a crucial role in cancers and neoplasms, and its relationship with genes and nucleic acids is a popular area of research in chemistry, genetics, and biochemistry. Aberrant chang

-----

Processing item 75/142 = 52.8 %

Processing article = ## Keywords: model organisms, caenorhabditis, proteins, invertebrates, animal models, genetics, research and analysis methods, teeth, organisms, eukaryota, animal genomics, invertebrate genomics, anat

OK. Pausing for 10 secs...OK

Summary = Developmental plasticity is the ability of an organism to develop different traits in response to environmental changes. This study investigates the genetic basis of predatory vs non-predatory dimorph

-----

Processing item 76/142 = 53.5 %

Processing article = ## Keywords: signal transduction, cell signaling, chemistry, chemical reactions, engineering and technology, biological tissue, bioenergetics, electrochemistry, mitochondria, physical sciences, cellul

OK. Pausing for 10 secs...OK

Summary = Mitochondria, which are organelles found inside cells, play a crucial role in the functioning of the cell. They are particularly important in maintaining calcium dynamics, which is the way that calciu

-----

Processing item 77/142 = 54.2 %

Processing article = ## Keywords: epidemiology, amniotes, medical conditions, geography, rodents, mammals, organisms, eukaryota, human geography, asia, urbanization, social sciences, china, vertebrates, viral diseases, in

OK. Pausing for 10 secs...OK

Summary = Hemorrhagic fever with renal syndrome (HFRS) is a disease that is transmitted to humans through infected rodents, resulting in fever, hemorrhage, and renal dysfunction. In China, HFRS is a significant

-----

Processing item 78/142 = 54.9 %

Processing article = ## Keywords: gastrointestinal tract, pathogens, taxonomy, bacteria, genetics, organisms, phylogenetic analysis, data management, anatomy, population biology, helicobacter pylori, species diversity, su

OK. Pausing for 10 secs...OK

Summary = Helicobacter pylori is a bacterial pathogen that infects the stomach of around half of the world's population and can lead to diseases like gastric and duodenal ulcers, MALT lymphoma, and gastric canc

-----

Processing item 79/142 = 55.6 %

Processing article = ## Keywords: schools, medical conditions, tanzania, invertebrates, parasitic diseases, hookworms, health care, health statistics, organisms, helminth infections, eukaryota, social sciences, soil-trans

OK. Pausing for 10 secs...OK

Summary = Soil-transmitted helminth (STH) infections, including Ascaris lumbricoides, Trichuris trichiura, and hookworms, are common in tropical and subtropical countries with inadequate sanitation. Over 1.5 bi

-----

Processing item 80/142 = 56.3 %

Processing article = ## Keywords: pathogens, cell membranes, medical conditions, endosomes, membrane trafficking, membrane fusion, chikungunya virus, microtubules, organisms, viral transmission and infection, chikungunya

OK. Pausing for 10 secs...OK

Summary = Chikungunya virus (CHIKV) is a re-emerging mosquito-borne virus that has caused millions of infections worldwide. It belongs to the Togaviridae family and enters host cells through clathrin-mediated endocytosis.

-----  
Processing item 81/142 = 57.0 %

Processing article = ## Keywords: pathogens, gene types, bacteria, point mutation, genetics, research and analysis methods, frameshift mutation, organisms, suppressor genes, homeostasis, mutant strains, neurochemistry, exocytosis

OK. Pausing for 10 secs...OK

Summary = Bacteria use second messengers to respond to changing environmental conditions. One important second messenger is c-di-AMP, which is essential in many bacteria and can be toxic if it accumulates. C-di-AMP is produced by the diguanylate cyclase (DGC) enzyme.

-----  
Processing item 82/142 = 57.7 %

Processing article = ## Keywords: microhomology-mediated end joining, white blood cells, nucleic acids, pathogens, epidemiology, cancers and neoplasms, proteins, genetics, organisms, oncology, B cells, antibody-producing

OK. Pausing for 10 secs...OK

Summary = Epstein-Barr virus (EBV) is a type of virus that can cause certain types of cancer, such as lymphomas and epithelial cell cancers. EBV is prevalent in many people and is generally harmless, but in some cases it can lead to cancer.

-----  
Processing item 83/142 = 58.5 %

Processing article = ## Keywords: storage and handling, medical conditions, polymerase chain reaction, artificial gene amplification and extension, research facilities, Buruli ulcer, bacteria, research and analysis method

OK. Pausing for 10 secs...OK

Summary = Buruli ulcer is a chronic and debilitating disease affecting at least 33 tropical, subtropical, and temperate countries, with most cases in West and Central Africa. The exact mode of transmission is unknown.

-----  
Processing item 84/142 = 59.2 %

Processing article = ## Keywords: cell membranes, macromolecular structure analysis, biochemical simulations, proteins, membrane composition, crystallography, lipids, crystal structure, physical sciences, membrane protein

OK. Pausing for 10 secs...OK

Summary = CD44 is a versatile protein involved in various cellular processes, including inflammation, hematopoiesis, cell migration, and cancer invasiveness. It consists of an extracellular domain, a transmembrane domain, and a cytoplasmic tail.

-----  
Processing item 85/142 = 59.9 %

Processing article = ## Keywords: entomology, hydrology, systems science, motion, invertebrates, research and analysis methods, organisms, eukaryota, mathematics, physical sciences, physics, flooding, simulation and modeling

OK. Pausing for 10 secs...OK

Summary = In this research, the authors created a numerical model to understand how ant rafts, or groups of ants floating on water, maintain their shape and move. Ant rafts are made up of two layers: a structural layer and a functional layer.

-----  
Processing item 86/142 = 60.6 %

Processing article = ## Keywords: nucleic acids, signal transduction, medical conditions, genome analysis, Culex quinquefasciatus, entomology, olfactory receptors, proteins, invertebrates, genetics, organisms, transcriptomics

OK. Pausing for 10 secs...OK

Summary = The text is a research paper about the study of mosquitoes, specifically Culex quinquefasciatus, and its role in transmitting diseases.

ulex pipiens quinquefasciatus and Culex pipiens molestus. The researchers analyzed the genetic differences between these two s

-----  
Processing item 87/142 = 61.3 %

Processing article = ## Keywords: developmental neuroscience, neurophysiology, neuronal dendrites, cellular types, cell biology, membrane potential, cellular neuroscience, neurons, synaptic plasticity, pyramidal cells, or

OK. Pausing for 10 secs...OK

Summary = This text is a summary of a medical research paper about the development of place fields in the brain. Place fields are areas in the brain that become active when an animal is in a certain location. T

-----  
Processing item 88/142 = 62.0 %

Processing article = ## Keywords: medical conditions, research assessment, research design, entomology, invertebrates, parasitic diseases, research and analysis methods, linear regression analysis, organisms, helminth inf

OK. Pausing for 10 secs...OK

Summary = Molecular xenomonitoring (MX) is a method used to detect the presence of Onchocerca volvulus DNA in black flies as an indicator for the presence of the parasite in human populations. A study was condu

-----  
Processing item 89/142 = 62.7 %

Processing article = ## Keywords: leprosy, immune system, proteins, bacteria, genetics, research and analysis methods, organisms, lectins, mycobacterium leprae, gene mapping, complement system, immunology, introns, genomi

OK. Pausing for 10 secs...OK

Summary = Leprosy is a chronic infectious disease caused by the bacteria Mycobacterium leprae. It affects the skin and nerves, and can lead to disfigurement and disability. The disease is more prevalent in Braz

-----  
Processing item 90/142 = 63.4 %

Processing article = ## Keywords: population size, glucose metabolism, cell processes, proteins, chemistry, carbohydrate metabolism, genetics, population metrics, glucose, physical sciences, dna-binding proteins, biochemi

OK. Pausing for 10 secs...OK

Summary = This text discusses the role of population size and glucose metabolism in evolution and genetic processes. It explains that nongenetic variation, or variation that occurs between individuals even when

-----  
Processing item 91/142 = 64.1 %

Processing article = ## Keywords: nucleic acids, pathogens, rna-binding proteins, non-coding rna, proteins, transfection, genetics, research and analysis methods, organisms, dengue virus, genomics, gene expression, viral

OK. Pausing for 10 secs...OK

Summary = Nucleic acids, proteins, and RNA-binding proteins all play a role in genetics and gene expression, particularly in relation to pathogens such as the dengue virus. The dengue virus is an RNA virus that

-----  
Processing item 92/142 = 64.8 %

Processing article = ## Keywords: conservation biology, covariance, systems science, probability theory, dwell time, mathematics, physical sciences, biodiversity, species extinction, evolutionary processes, species divers

OK. Pausing for 10 secs...OK

Summary = Lay Summary:

Conservation biology and ecology involve studying the impact of temporal environment

al variations on biodiversity. In particular, the covariance between competition and environment can f

-----  
Processing item 93/142 = 65.5 %

Processing article = ## Keywords: nucleic acids, macromolecular structure analysis, non-coding rna, bioinformatics, research and analysis methods, database and informatics methods, computational techniques, sequence analy

OK. Pausing for 10 secs...OK

Summary = Functional non-coding RNAs, or fncRNAs, are nucleotide sequences that significantly influence gene expression and translation, genome stability, and human health and disease. To understand the determi

-----  
Processing item 94/142 = 66.2 %

Processing article = ## Keywords: epidemiology, medical conditions, network analysis, covariance, geography, probability theory, cholera, human geography, physical sciences, mathematics, social sciences, human mobility, i

OK. Pausing for 10 secs...OK

Summary = The text discusses a method for detecting the source of epidemics, specifically focusing on cholera outbreaks. The method is based on network analysis and probability theory, using a sensor-based appr

-----  
Processing item 95/142 = 66.9 %

Processing article = ## Keywords: imaging techniques, cancers and neoplasms, machine learning, breast tumors, immunologic techniques, forecasting, intelligence, research and analysis methods, immunofluorescence, oncology,

OK. Pausing for 10 secs...OK

Summary = Researchers have developed advanced imaging techniques that allow for detailed examination of cancer tissues by staining them with multiple markers. However, these techniques have limitations such as

-----  
Processing item 96/142 = 67.6 %

Processing article = ## Keywords: amniotes, pathogens, medical conditions, atrophy, urine, bacteria, rodents, mammals, biomarkers, organisms, leptospirosis, eukaryota, immunology, signs and symptoms, anatomy, biochemistry

OK. Pausing for 10 secs...OK

Summary = Leptospirosis is a disease caused by the bacterium Leptospira. It can affect both humans and animals, particularly rodents. The bacteria is transmitted through urine from infected animals and can caus

-----  
Processing item 97/142 = 68.3 %

Processing article = ## Keywords: white blood cells, medical conditions, machine learning, forecasting, t cells, research and analysis methods, population metrics, learning curves, mathematics, physical sciences, cellular

OK. Pausing for 10 secs...OK

Summary = Talaromycosis is a serious regional disease endemic in Southeast Asia, with high morbidity and mortality rates in HIV/AIDS patients in China, particularly in Guangxi. The disease is caused by the fung

-----  
Processing item 98/142 = 69.0 %

Processing article = ## Keywords: topography, amniotes, socioeconomic aspects of health, landforms, islands, veterinary parasitology, life cycles, united states, mammals, parasitic diseases, health care, developmental bio

OK. Pausing for 10 secs...OK

Summary = Toxocara, a type of parasitic roundworm, is a common parasite in pets and can cause a disease called toxocariasis in humans. This disease is often underdiagnosed and understudied, particularly in soci

-----  
Processing item 99/142 = 69.7 %

Processing article = ## Keywords: immunologic techniques, proteins, arachnida, scorpions, chemistry, invertebrates, research and analysis methods, organisms, molecular mass, enzymes, eukaryota, physical sciences, venoms,

OK. Pausing for 10 secs...OK

Summary = Scorpions in the genus Tityus, particularly Tityus trivittatus, are a public health issue in tropical America, including parts of Brazil, Argentina, and Paraguay. A recent study focused on a populatio

-----  
Processing item 100/142 = 70.4 %

Processing article = ## Keywords: amniotes, bats, bioacoustics, plant anatomy, mammals, plant science, organisms, echoes, eukaryota, acoustics, physical sciences, physics, vertebrates, plants, flowering plants, animals, b

OK. Pausing for 10 secs...OK

Summary = Bat-pollinated flowers have unique acoustic properties that differ from flowers pollinated by insects or birds. These acoustic properties are likely used by bats for identification and classification

-----  
Processing item 101/142 = 71.1 %

Processing article = ## Keywords: oncomelania, snails, artificial neural networks, hydrology, habitats, invertebrates, parasitic diseases, organisms, helminth infections, eukaryota, malacology, schistosomiasis, ecology an

OK. Pausing for 10 secs...OK

Summary = Schistosomiasis is a parasitic disease affecting millions of people worldwide, with China being one of the endemic areas. The disease is caused by parasitic worms of the genus Schistosoma, with Oncome

-----  
Processing item 102/142 = 71.8 %

Processing article = ## Keywords: proteins, transfection, protozoans, research and analysis methods, organisms, enzyme chemistry, giardia, parasite groups, eukaryota, enzyme metabolism, trophozoites, cloning, biochemistry

OK. Pausing for 10 secs...OK

Summary = Giardia duodenalis, a protozoan parasite, causes giardiasis, a diarrheal disease that affects humans and other mammals. The treatment of giardiasis mainly relies on metronidazole (MTZ) and albendazole

-----  
Processing item 103/142 = 72.5 %

Processing article = ## Keywords: sulfates, methionine, amino acids, chemistry, proteins, organisms, enzyme chemistry, cysteine, eukaryota, physical sciences, enzyme metabolism, biochemistry, sulfur containing amino acids

OK. Pausing for 10 secs...OK

Summary = Sulfur is an essential element for all life forms, and its utilization is a fundamental process across different domains of life. However, how cells deal with varying sulfur availability is not well u

-----  
Processing item 104/142 = 73.2 %

Processing article = ## Keywords: locusts, nucleic acids, agriculture, signal transduction, cell signaling, entomology, proteins, invertebrates, genetics, organisms, biological tissue, ova, reproductive system, insect pes

OK. Pausing for 10 secs...OK

Summary = Juvenile Hormone (JH) is a crucial hormone that stimulates insect egg production by regulating vitellogenesis, a process involving the synthesis, secretion, and transport of yolk proteins. A study on

-----  
Processing item 105/142 = 73.9 %



Processing article = ## Keywords: gastrointestinal tract, machine learning, addiction, ascites, liver fibrosis, statistical data, mathematics, physical sciences, anatomy, social sciences, liver diseases, alcoholics, gastr

OK. Pausing for 10 secs...OK

Summary = Alcohol-related liver disease (ALD) is a major cause of liver-related deaths, with alcoholic liver cirrhosis (ALC) being the most common form of ALD in registry data. ALC patients have a strong over-r

-----

Processing item 106/142 = 74.6 %

Processing article = ## Keywords: nucleic acids, developmental neuroscience, non-coding rna, life cycles, regeneration, genetics, developmental biology, neuronal dendrites, micrnas, cellular types, cognitive psychology,

OK. Pausing for 10 secs...OK

Summary = In the field of developmental neuroscience, researchers are studying how neurons in the brain change and adapt over time. One important aspect of this is how neurons alter their dendrite arbors, which

-----

Processing item 107/142 = 75.4 %

Processing article = ## Keywords: white blood cells, immune system, signal transduction, medical conditions, molecular development, proteins, chemistry, parasitic diseases, t cells, protozoans, developmental biology, orga

OK. Pausing for 10 secs...OK

Summary = Vγ9Vδ2 T cells are a type of white blood cell that can quickly respond to malaria parasites without the need for prior exposure or processing. These cells can inhibit the replication of the parasites

-----

Processing item 108/142 = 76.1 %

Processing article = ## Keywords: computer software, amniotes, pathogens, lentivirus, macromolecular structure analysis, biochemical simulations, hiv, proteins, mammals, research and analysis methods, organisms, engineeri

OK. Pausing for 10 secs...OK

Summary = DEER-PREdict is a new software program that predicts Double Electron-Electron Resonance (DEER) distance distributions and Paramagnetic Relaxation Enhancement (PRE) rates from ensembles of protein conf

-----

Processing item 109/142 = 76.8 %

Processing article = ## Keywords: pathogens, polymerase chain reaction, artificial gene amplification and extension, model organisms, bacteria, health care, research and analysis methods, organisms, public and occupationa

OK. Pausing for 10 secs...OK

Summary = Enterotoxigenic Escherichia coli (ETEC) is a significant cause of diarrhea in children under five in low-middle income countries (LMICs), but the lack of diagnostic tools results in limited data on it

-----

Processing item 110/142 = 77.5 %

Processing article = ## Keywords: cancers and neoplasms, breast tumors, research and analysis methods, chemotherapy, principal component analysis, oncology, mathematics, physical sciences, digestive system procedures, pha

OK. Pausing for 10 secs...OK

Summary = Researchers have developed a new method for analyzing data from longitudinal studies, which are experiments that take place over a period of time with repeated measurements. This new method, called Re

-----

Processing item 111/142 = 78.2 %

Processing article = ## Keywords: network analysis, chemistry, reaction dynamics, probability theory, research and analysis methods, applied mathematics, computational

techniques, mathematics, physical sciences, signaling

OK. Pausing for 10 secs...OK

Summary = This text discusses a model for disease diagnosis that uses a reaction network to represent the time evolution of disease states. The authors use a stochastic process to simulate the development of de

-----  
Processing item 112/142 = 78.9 %

Processing article = ## Keywords: viral upper respiratory tract infection, amniotes, mixtures, medical conditions, pathogens, rodents, aerosols, mammals, organisms, eukaryota, physical sciences, viral pathogens, sars cov

OK. Pausing for 10 secs...OK

Summary = Lay Summary:

The transmission of the virus that causes COVID-19, SARS-CoV-2, is dependent on the amount of infectious particles present in the environment and the proximity between the susceptible and

-----  
Processing item 113/142 = 79.6 %

Processing article = ## Keywords: medical conditions, polymerase chain reaction, artificial gene amplification and extension, parasitic diseases, protozoans, research and analysis methods, organisms, parasite groups, euka

OK. Pausing for 10 secs...OK

Summary = Lay Summary:

This text discusses methods for estimating the overlap between two communities, such as groups of species or genes, when only partial samples are available. Traditional methods have assu

-----  
Processing item 114/142 = 80.3 %

Processing article = ## Keywords: trypanosoma cruzi, epidemiology, medical conditions, immunologic techniques, protozoans, research and analysis methods, parasitic diseases, organisms, eukaryota, mathematics, physical sci

OK. Pausing for 10 secs...OK

Summary = Chagas Disease (CD) is a parasitic disease caused by the protozoan Trypanosoma cruzi, which is common in Latin America and has become a global health issue due to international migration. The disease

-----  
Processing item 115/142 = 81.0 %

Processing article = ## Keywords: finite element analysis, neostriatum, dopamine, neurophysiology, chemistry, research and analysis methods, applied mathematics, neurotransmission, mathematics, physical sciences, cellular

OK. Pausing for 10 secs...OK

Summary = Dopamine (DA) is a neurotransmitter that affects various cognitive functions and behavioral traits. It exists in two forms, phasic and tonic DA, which have distinct functions and are associated with d

-----  
Processing item 116/142 = 81.7 %

Processing article = ## Keywords: nucleic acids, pathogens, cell membranes, medical conditions, cell processes, ixodes, proteins, arachnida, autophagic cell death, invertebrates, genetics, organisms, exocrine glands, euka

OK. Pausing for 10 secs...OK

Summary = The research focuses on the degeneration of tick salivary glands, which occurs rapidly after the tick finishes feeding. This degeneration is caused by a combination of autophagy and apoptosis. The stu

-----  
Processing item 117/142 = 82.4 %

Processing article = ## Keywords: epidemiology, plant communities, terrestrial envir

onments, geography, plant science, ecosystems, plant ecology, south america, seasons, forests, geographical locations, ecology and environment

OK. Pausing for 10 secs...OK

Summary = Yellow fever (YF) has seen significant changes in its incidence and geographic extent over the past 20 years, with the largest outbreaks in South America since 1940 occurring in the previously unaffected

-----  
Processing item 118/142 = 83.1 %

Processing article = ## Keywords: decision making, human learning, health care, behavioral disorders, neuroses, public and occupational health, cognitive psychology, cognition, anxiety disorders, sensory cues, perception,

OK. Pausing for 10 secs...OK

Summary = Obsessive-Compulsive Disorder (OCD) is a condition characterized by uncontrollable, repetitive actions and thoughts. These actions and thoughts are believed to be caused by abnormalities in fundamental

-----  
Processing item 119/142 = 83.8 %

Processing article = ## Keywords: protozoans, review, research and analysis methods, organisms, parasite groups, eukaryota, pharmacology, biochemistry, pharmaceutical processing technology, malarial parasites, medicine and

OK. Pausing for 10 secs...OK

Summary = The review discusses the biosynthesis of pantothenate (Pan) and coenzyme A (CoA) in apicomplexan parasites, with a focus on Toxoplasma gondii and Plasmodium falciparum. Pan is an essential precursor for

-----  
Processing item 120/142 = 84.5 %

Processing article = ## Keywords: cell membranes, protein structure databases, macromolecular structure analysis, biochemical simulations, proteins, chemistry, forecasting, research and analysis methods, lipids, database

OK. Pausing for 10 secs...OK

Summary = Membrane proteins have unique traits, with many of their hydrophobic amino acids exposed to the lipid bilayers rather than being embedded in the protein interior. This is often not explicitly considered

-----  
Processing item 121/142 = 85.2 %

Processing article = ## Keywords: epidemiology, medical conditions, systems science, research and analysis methods, public and occupational health, virus testing, preventive medicine, physical sciences, mathematics, immunology

OK. Pausing for 10 secs...OK

Summary = The COVID-19 pandemic has been a major global challenge, and managing it requires the use of simulation models to predict and control the spread of the disease. A new study has developed a detailed agent-based

-----  
Processing item 122/142 = 85.9 %

Processing article = ## Keywords: professions, health care, medical doctors, age groups, children, immunology, histology, signs and symptoms, diarrhea, anatomy, enteropathies, families, endoscopy, surgical and invasive medicine

OK. Pausing for 10 secs...OK

Summary = Environmental enteric dysfunction (EED) is a poorly understood condition that affects children in low- and middle-income countries and is associated with undernutrition and stunted growth. Its etiology

-----  
Processing item 123/142 = 86.6 %

Processing article = ## Keywords: decision making, recreation, applied mathematics, research and analysis methods, computational techniques, games, reaction time, mathematics, physical sciences, cognitive psychology, cognition

OK. Pausing for 10 secs...OK

Summary = Transparent games are a new game-theoretic setting where players can observe each other's actions with a certain probability. This is different from classical game theory, where players either act sim

-----  
Processing item 124/142 = 87.3 %

Processing article = ## Keywords: cancers and neoplasms, chromosome structure and function, connective tissue cells, connective tissue, transfection, research and analysis methods, chromosome biology, chromosomes, biologi

OK. Pausing for 10 secs...OK

Summary = This research paper discusses the role of ATRX in the Alternative Lengthening of Telomeres (ALT) pathway, which is a recombination-based mechanism for maintaining telomere length in some cancers. The

-----  
Processing item 125/142 = 88.0 %

Processing article = ## Keywords: cell membranes, cell motility, protozoans, organisms, viral transmission and infection, host cells, eukaryota, membrane proteins, cell biology, physiological processes, secretion, medicin

OK. Pausing for 10 secs...OK

Summary = Toxoplasma gondii is a parasite that affects host cells and its survival is due to the secretion of virulent factors. These factors are stored in specific organelles and are released when the parasite

-----  
Processing item 126/142 = 88.7 %

Processing article = ## Keywords: neurophysiology, single neuron function, developmental biology, cellular types, mood disorders, synapses, anatomy, cell biology, social sciences, membrane potential, cellular neuroscience

OK. Pausing for 10 secs...OK

Summary = A single neuron's stimulation in the rat's somatosensory cortex can result in a behavioral response, according to a study. The likelihood of a behavioral response does not significantly depend on the

-----  
Processing item 127/142 = 89.4 %

Processing article = ## Keywords: nucleic acids, connective tissue cells, connective tissue, chromatophores, genetics, research and analysis methods, melanocytes, biological tissue, cellular types, anatomy, dna damage, ge

OK. Pausing for 10 secs...OK

Summary = Human skin is constantly exposed to damaging environmental factors that can cause DNA damage and mutations in skin cells. While the impact of UV radiation on skin cells and melanoma has been studied,

-----  
Processing item 128/142 = 90.1 %

Processing article = ## Keywords: epidemiology, medical conditions, invertebrates, parasitic diseases, foodborne trematodiasis, organisms, helminth infections, clonorchiasis, eukaryota, trematodes, asia, ponds, social sci

OK. Pausing for 10 secs...OK

Summary = Clonorchiasis is a neglected tropical disease caused by the liver fluke Clonorchis sinensis, which is found in China, Taiwan, Korea, eastern Russia, north Vietnam, and could spread further due to glob

-----  
Processing item 129/142 = 90.8 %

Processing article = ## Keywords: epidemiology, schools, medical conditions, chemistry, parasitic diseases, age groups, public and occupational health, children, ectoparasitic infections, physical sciences, social science

OK. Pausing for 10 secs...OK

Summary = Skin problems are a significant cause of illness in communities with poor

living conditions, but they have received less attention in global health studies due to their low mortality rates. A study co

-----  
Processing item 130/142 = 91.5 %

Processing article = ## Keywords: imaging techniques, white blood cells, nucleic acids, pathogens, model organisms, non-coding rna, granulomas, bacteria, animal models, genetics, research and analysis methods, development

OK. Pausing for 10 secs...OK

Summary = This text discusses the role of microRNA (miRNA) in the interaction between mycobacteria and host immune cells. Mycobacteria, which cause diseases like tuberculosis and leprosy, can manipulate host si

-----  
Processing item 131/142 = 92.3 %

Processing article = ## Keywords: cell processes, machine learning, ears, research and analysis methods, developmental biology, homeostasis, anatomy, cell biology, integumentary system, simulation and modeling, physiologi

OK. Pausing for 10 secs...OK

Summary = This text discusses the use of machine learning techniques, particularly graph neural networks (GNN), to understand and predict the behavior of cells in tissues. The GNN models are used to analyze the

-----  
Processing item 132/142 = 93.0 %

Processing article = ## Keywords: nucleic acids, cancers and neoplasms, cell processes, non-coding rna, genetics, research and analysis methods, oncology, permutation, micrnas, mathematics, physical sciences, biochemist

OK. Pausing for 10 secs...OK

Summary = This text is about a study using a new method called TSCCA (Tensor Sparse Canonical Correlation Analysis) to analyze the relationship between microRNA (miRNA) and gene expression in various cancers. T

-----  
Processing item 133/142 = 93.7 %

Processing article = ## Keywords: developmental neuroscience, neurophysiology, systems science, nonlinear dynamics, mathematics, physical sciences, cellular types, mood disorders, anatomy, cell biology, membrane potential

OK. Pausing for 10 secs...OK

Summary = Lay Summary:

The study investigates the interaction between short-term and long-term plasticity in the brain, which are crucial for learning and memory. Short-term plasticity lasts for a few minutes,

-----  
Processing item 134/142 = 94.4 %

Processing article = ## Keywords: amniotes, bats, genome analysis, interferons, proteins, mammals, genetics, organisms, eukaryota, immunology, genomics, biochemistry, viral replication, vertebrates, spleen, medicine and h

OK. Pausing for 10 secs...OK

Summary = Bats are known to host various viruses, including Hendra virus (HeV), without showing symptoms. This has been attributed to their innate immune response. A study was conducted to compare the immune re

-----  
Processing item 135/142 = 95.1 %

Processing article = ## Keywords: epidemiology, genome analysis, normal distribution, genetics, probability theory, schizophrenia, mathematics, physical sciences, genomics, gene expression, human genetics, mental health a

OK. Pausing for 10 secs...OK

Summary = Genetic research has identified thousands of genomic risk factors and improved our understanding of disease causes and potential treatments. Genome-Wide Assoc

iation Studies (GWAS) are commonly used an

-----  
Processing item 136/142 = 95.8 %

Processing article = ## Keywords: imaging techniques, pathogens, research and analysis methods, developmental biology, organisms, infection imaging, diagnostic radiology, dna viruses, viral pathogens, radiology and imaging

OK. Pausing for 10 secs...OK

Summary = A deep convolutional neural network (CNN) model was developed to recognize cytopathic effects (CPE) induced by influenza virus infection in Madin-Darby canine kidney (MDCK) cells. The model was trained

-----  
Processing item 137/142 = 96.5 %

Processing article = ## Keywords: microbial mutation, cell processes, amyloid proteins, necrotic cell death, protein interactions, proteins, physical sciences, physics, cell biology, cell death, biochemistry, viral packaging

OK. Pausing for 10 secs...OK

Summary = This text is a research paper about the Varicella-zoster virus (VZV) and its interaction with human cells. VZV is a virus that causes chickenpox and shingles in humans. The researchers were interested

-----  
Processing item 138/142 = 97.2 %

Processing article = ## Keywords:

## Abstract

Infectious disease risk is driven by three interrelated components: exposure, hazard, and vulnerability. For schistosomiasis, exposure occurs through contact with water,

OK. Pausing for 10 secs...OK

Summary = Schistosomiasis is a disease caused by parasitic trematodes from the genus Schistosoma, and it is the second leading cause of parasitic disease globally. The risk for schistosomiasis can be divided in

-----  
Processing item 139/142 = 97.9 %

Processing article = ## Keywords: immune system, amniotes, lassa virus, medical conditions, pathogens, molecular development, lassa fever, primates, taxonomy, rodents, animal models, mammals, research and analysis methods

OK. Pausing for 10 secs...OK

Summary = Lassa fever (LF) is a dangerous viral hemorrhagic fever prevalent in some Western African countries, particularly Nigeria, and is caused by Lassa virus (LASV). The virus is primarily transmitted to humans

-----  
Processing item 140/142 = 98.6 %

Processing article = ## Keywords: schistosoma mansoni, medical conditions, invertebrates, parasitic diseases, organisms, thrombocytopenia, helminth infections, eukaryotes, hematology, cellular types, anatomy, cell biology,

OK. Pausing for 10 secs...OK

Summary = Schistosoma mansoni is a parasitic worm that infects humans and causes schistosomiasis, a disease that affects over 260 million people worldwide. This study compares the hematological and biochemical

-----  
Processing item 141/142 = 99.3 %

Processing article = ## Keywords: white blood cells, pathogens, flow cytometry, proteins, research and analysis methods, organisms, hepatitis virus, b cells, antibody-producing cells, staining, hepatitis viruses, spectrum analysis

OK. Pausing for 10 secs...OK

Summary = White blood cells, specifically B cells, play a crucial role in the immune response to pathogens such as hepatitis C virus (HCV). These pathogens can infect an

d affect the function of B cells, which a

-----

```
In [23]: df.head()
```

```
Out[23]:
```

	article	headings	keywords	id	mixtral_summary
0	Lung-resident ( LR ) mesenchymal stem and stro...	[Abstract, Introduction, Results, Discussion, ...	[immune system, medical conditions, molecular ...	journal.ppat.1009789	Lung-resident mesenchymal stem cells (LR-MSCs)...
1	Visceral leishmaniasis ( VL ) is endemic in So...	[Abstract, Introduction, Methods, Results, Dis...	[neonates, clinical laboratory sciences, trans...	journal.pntd.0007992	Visceral Leishmaniasis (VL), a parasitic disea...
2	A high burden of Salmonella enterica subspecie...	[Abstract, Introduction, Methods, Results, Dis...	[pathogens, medical conditions, taxonomy, bact...	journal.pntd.0010704	Salmonella Typhi, a bacterium that causes typh...
3	Severe Acute Respiratory Syndrome Coronavirus-...	[Abstract, Introduction, Results, Discussion, ...	[pathogens, amniotes, medical conditions, bind...	journal.ppat.1010691	The COVID-19 pandemic, caused by the SARS-CoV-...
4	Many fungal species utilize hydroxyderivatives...	[Abstract, Introduction, Results and discussio...	[taxonomy, proteins, chemistry, genetics, enzy...	journal.pgen.1009815	Lay Summary:\nCandida parapsilosis is a type o...

```
In [24]: # check how many rows have blank result (some errors)
empty_df = df.query("mixtral_summary.str.strip() == ''")
print(len(empty_df))
```

0

```
In [25]: # attempt to retry for blank rows (due to some errors)
retry = True
if retry:
    print("Retrying for empty results...")
    for i in range(len(df)):
        item = df.iloc[i]
        if item["mixtral_summary"] == "":
            print("Item =", i)
            # text = item["article"]
            text = build_text_with_headings(item)
            df.at[i, "mixtral_summary"] = send_sumarize_request(text, quiet=False)
    print("Completed")
```

Retrying for empty results...

Completed

```
In [26]: output_path = "./data/output/test_set/"
output_filename = "plos_groq_mixtral_summary.csv"

print("Writing to file ", output_filename)
df.to_csv(output_path+output_filename,
          index = False
          )
print("Completed")
```

Writing to file plos\_groq\_mixtral\_summary.csv

Completed

```
In [27]: output_filename = "plos_groq_mixtral_summary.json"

print("Writing to file ", output_filename)
df.to_json(output_path+output_filename,
           orient="records",
           )
print("Completed")
```

Writing to file plos\_groq\_mixtral\_summary.json

Completed

In [ ]: