Analysis of Food Pairing in Regional Cuisines of India

Anupam Jain[1☯], Rakhi N K[2☯], Ganesh Bagler[2]*

1 Center for System Science, Indian Institute of Technology Jodhpur, Jodhpur, Rajasthan, India, 2 Center for Biologically Inspired System Science, Indian Institute of Technology Jodhpur, Jodhpur, Rajasthan, India

☯ These authors contributed equally to this work.
* bagler@iitj.ac.in; ganesh.bagler@gmail.com

Abstract

Any national cuisine is a sum total of its variety of regional cuisines, which are the cultural and historical identifiers of their respective regions. India is home to a number of regional cuisines that showcase its culinary diversity. Here, we study recipes from eight different regional cuisines of India spanning various geographies and climates. We investigate the phenomenon of food pairing which examines compatibility of two ingredients in a recipe in terms of their shared flavor compounds. Food pairing was enumerated at the level of cuisine, recipes as well as ingredient pairs by quantifying flavor sharing between pairs of ingredients. Our results indicate that each regional cuisine follows negative food pairing pattern; more the extent of flavor sharing between two ingredients, lesser their co-occurrence in that cuisine. We find that frequency of ingredient usage is central in rendering the characteristic food pairing in each of these cuisines. Spice and dairy emerged as the most significant ingredient classes responsible for the biased pattern of food pairing. Interestingly while individual spices contribute to negative food pairing, dairy products on the other hand tend to deviate food pairing towards positive side. Our data analytical study highlighting statistical properties of the regional cuisines, brings out their culinary fingerprints that could be used to design algorithms for generating novel recipes and recipe recommender systems. It forms a basis for exploring possible causal connection between diet and health as well as prospection of therapeutic molecules from food ingredients. Our study also provides insights as to how big data can change the way we look at food.

Introduction

Cooking is a unique trait humans possess and is believed to be a major cause of increased brain size [1–3]. While cooking encompasses an array of food processing techniques [4], cuisine is an organized series of food preparation procedures intended to create tasty and healthy food. India has a unique blend of culturally and climatically diverse regional cuisines. Its culinary history dates back to the early Indus valley civilization [5–7]. Indian dietary practices are deeply rooted in notions of disease prevention and promotion of health.

Food perception involving olfactory and gustatory mechanisms is the primary influence for food preferences in humans. These preferences are also determined by a variety of factors such as culture, climate geography and genetics, leading to emergence of regional cuisines [4, 8–12]. Food pairing is the idea that ingredients having similar flavor constitution may taste well in a recipe. Chef Blumenthal was the first to propose this idea, which in this study we term as positive food pairing [13]. Studies by Ahn et al found that North American, Latin American and Southern European recipes follow this food pairing pattern where as certain others like North Korean cuisine and Eastern European cuisines do not [14, 15]. Our previous study of food pairing in Indian cuisine revealed a strong negative food pairing pattern in its recipes [16].

Knowing that each of the regional cuisines have their own identity, the question we seek to answer in this paper is whether the negative food pairing pattern in Indian cuisine is a consistent trend observed across all of the regional cuisines or an averaging effect. Towards answering this question, we investigated eight geographically and culturally prominent regional cuisines viz. Bengali, Gujarati, Jain, Maharashtrian, Mughlai, Punjabi, Rajasthani and South Indian. The pattern of food pairing was studied at the level of cuisine, recipes and ingredient pairs. Such a multi-tiered study of these cuisines provided a thorough understanding of its characteristics in terms of ingredient usage pattern. We further identified the features that contribute to food pairing, thereby revealing the role of ingredients and ingredient categories in determining food pairing of the regional cuisines.

Availability of large datasets in the form of cookery blogs and recipe repositories has prompted the use of big data analytical techniques in food science and has led to the emergence of computational gastronomy. This field has made advances through many recent studies [14, 15, 17, 18] which is changing the overall outlook of culinary science in recent years. Our study is an offshoot of this approach. We use statistical and computational models to analyse food pairing in the regional cuisines. Our study reveals the characteristic signature of each Indian regional cuisines by looking at the recipe and ingredient level statistics of the cuisine.

Results and Discussion

Details of recipes, ingredients, and their corresponding flavor compounds constitute the primary data required for study of food pairing in a cuisine. Much of this is documented in the form of books and recently through online recipe sources. We obtained the Indian cuisine recipes data S1 Dataset from one of the popular cookery websites *TarlaDalal.com* [19]. The flavor profiles of ingredients were compiled using previously published data [15] and through extensive literature survey S2 Dataset. Table 1 lists details of recipes and ingredients in each of the regional cuisines.

The ingredients belonged to following 15 categories: spice, vegetable, fruit, plant derivative, nut/seed, cereal/crop, dairy, plant, pulse, herb, meat, fish/seafood, beverage, animal product, and flower. Category-wise ingredient statistics of regional cuisines is provided in S1 Table.

Statistics of recipe size and ingredient frequency

We started with investigation of preliminary statistics of regional cuisines. All the eight regional cuisines under consideration showed bounded recipe-size distribution (Fig 1). While most cuisines followed uni-modal distribution, Mughlai cuisine showed a strong bimodal distribution and had recipes with large sizes when compared with the rest. This could be an indication of the fact that Mughlai is derivative of a royal cuisine. To understand the ingredient usage pattern, we ranked ingredients according to decreasing usage frequency within each cuisine. As shown in Fig 2, all cuisines showed strikingly similar ingredient usage profile reflecting the pattern of Indian cuisine (Fig 2, inset). While indicating a generic culinary growth

Table 1. Statistics of regional cuisines.

Cuisine	Recipe count	Ingredient count
Bengali	156	102
Gujarati	392	112
Jain	447	138
Maharashtrian	130	93
Mughlai	179	105
Punjabi	1013	152
Rajasthani	126	78
South Indian	474	114

Recipes of size ≥ 2 were considered for the purpose of flavor analysis.

doi:10.1371/journal.pone.0139539.t001
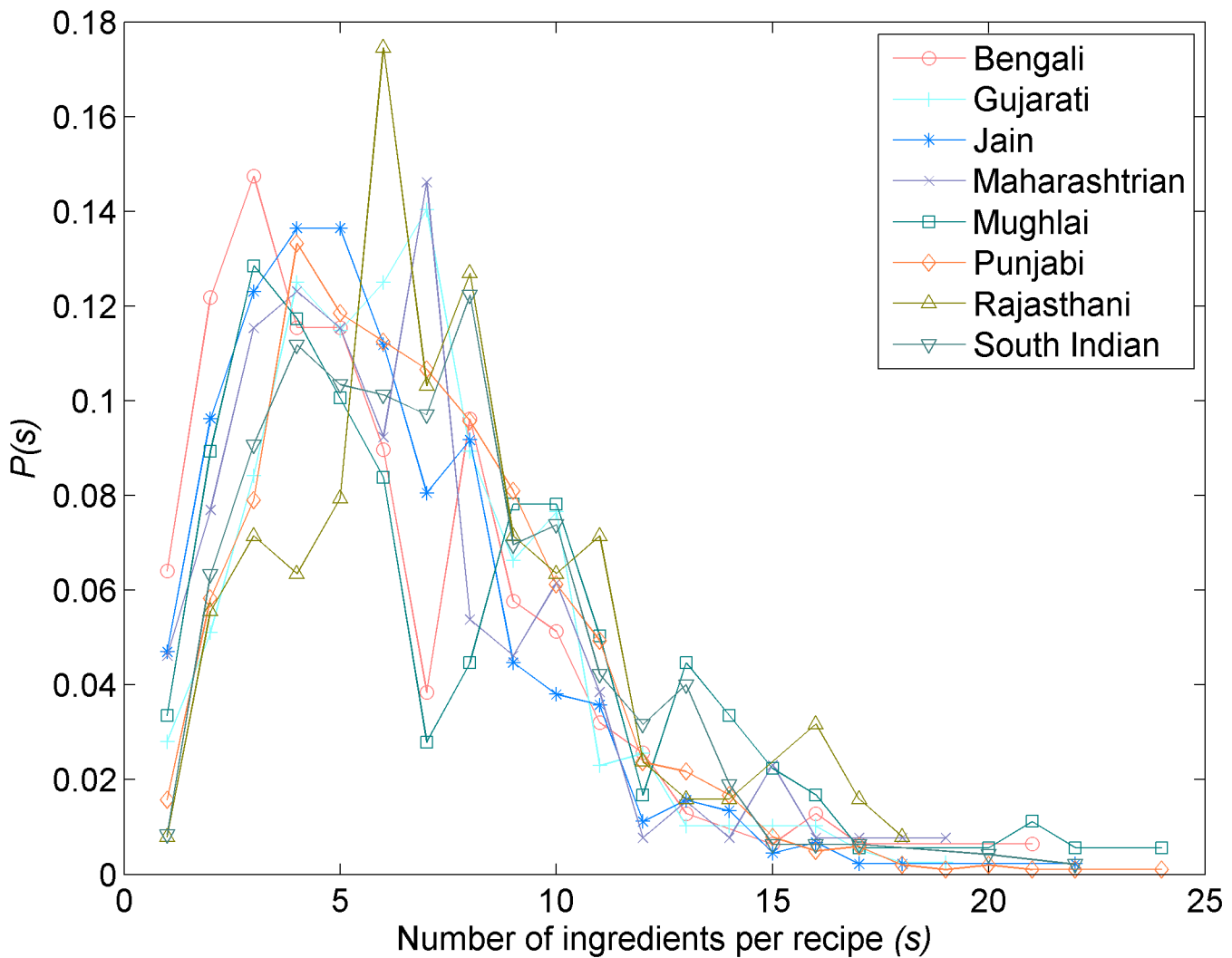
Fig 1. Recipe size distributions. Plot of probability of finding a recipe of size *s* in the cuisine. Consistent with other cuisines, the distributions are bounded. Mughlai and Punjabi cuisines have recipes of large sizes compared to other cuisines.

doi:10.1371/journal.pone.0139539.g001
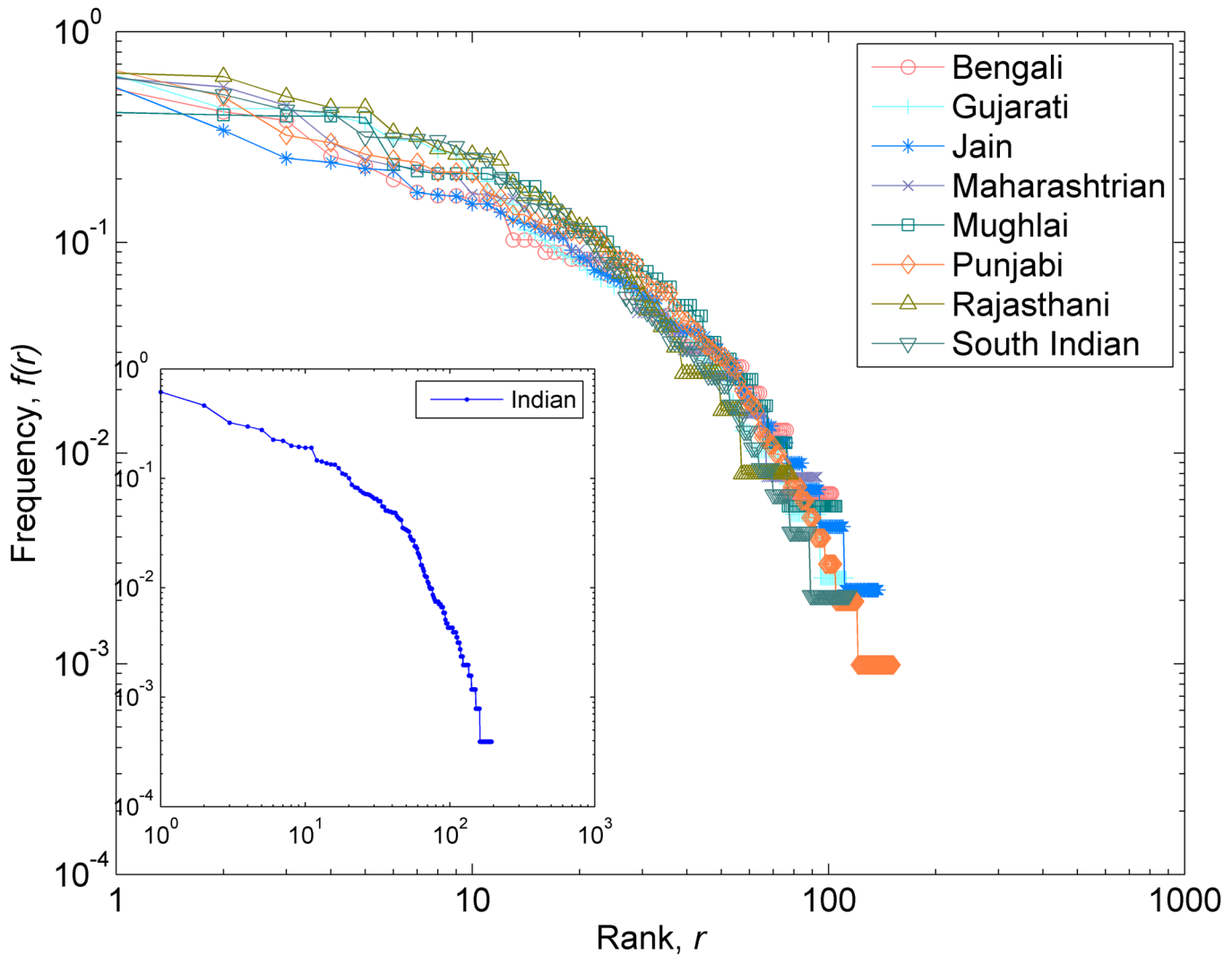
Fig 2. Frequency-Rank distributions. Ingredients ranked as per their frequency of use in the cuisine. Higher the occurrence, better the rank of the ingredient. All the cuisines have similar ingredient distribution profile indicating generic culinary growth mechanism. Inset shows the ingredient frequency-rank distribution for the whole Indian cuisine.

doi:10.1371/journal.pone.0139539.g002

mechanism, the distributions also show that certain ingredients are excessively used in cuisines depicting their inherent 'fitness' or popularity within the cuisine.

Food pairing hypothesis

Food pairing hypothesis is a popular notion in culinary science. It asserts that two ingredients sharing common flavor compounds taste well when used together in a recipe. This hypothesis has been confirmed for a few cuisines such as North American, Western European and Latin American [15]. In contrast, Korean and Southern European cuisines have been shown to deviate from positive food pairing. Our previous study of food pairing in Indian cuisine at the level of cuisine, sub-cuisines, recipes and ingredient pairs has shown that it is characterized with a strong negative food pairing [16]. We quantify food pairing with the help of flavor profiles of

ingredients. Flavor profile represents a set of volatile chemical compounds that render the characteristic taste and smell to the ingredient. Starting with the flavor profiles of each of the ingredients, average food pairing of a recipe (N_s^R) as well as that of the cuisine ($\overline{N}_s$) was computed as illustrated in Fig 3. The extent of deviation of $\overline{N}_s$ of the cuisine, when compared to that of a 'random cuisine' measures the bias in food pairing. The higher/lower the value of $\overline{N}_s$ from that of its random counterpart the more positive/negative the food pairing is.

Regional cuisines of India exhibit negative food pairing

We found that all regional cuisines are invariantly characterized by average food pairing lesser than expected by chance. This characteristic negative food pairing, however, varied in its extent across cuisines. Mughlai cuisine, for example, displayed the least inclination towards negative pairing ($\Delta N_s = \overline{N}_s^{Mughlai} - \overline{N}_s^{Rand} = -0.758$ and Z-score of -10.232). Whereas, Maharashtrian cuisine showed the most negative food pairing ($\Delta N_s = \overline{N}_s^{Maharashtrian} - \overline{N}_s^{Rand} = -4.523$ and Z-score of -52.047). Fig 4 depicts the generic food pairing pattern observed across regional cuisines of India. We found that the negative food pairing is independent of recipe size as shown in Fig 5. This indicates that the bias in food pairing is not an artifact of averaging over recipes of all sizes and is a quintessential feature of all regional cuisines of India. Note that, across cuisines, majority of recipes are in the size-range of around 3 to 12. Hence the significance of food pairing statistics is relevant below the recipe size cut-off of ~ 12.
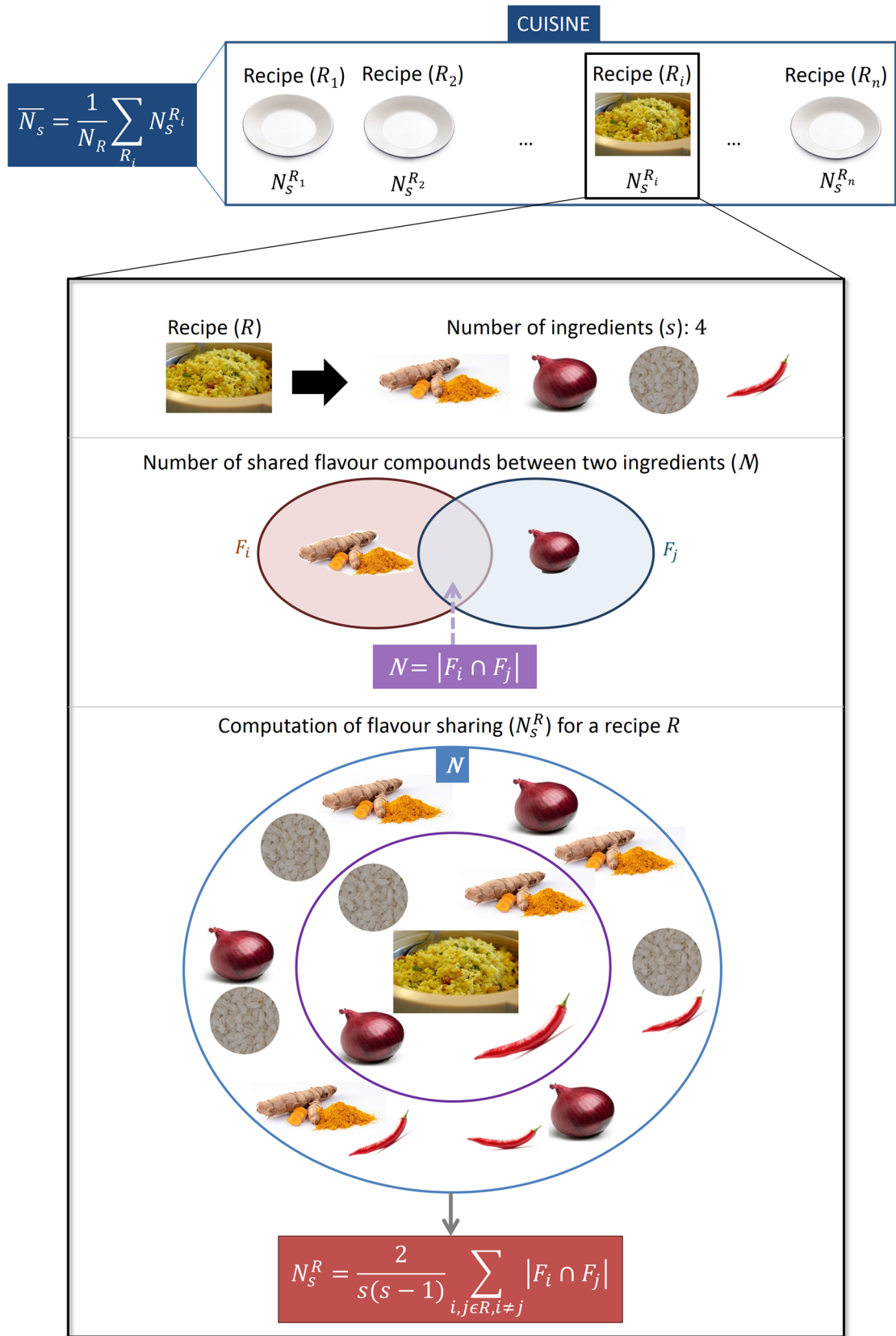
We further investigated for possible factors that could explain negative food pairing pattern observed in regional cuisines. We created randomized controls for each regional cuisine to explore different aspects that may contribute to the bias in food pairing. In the first control, frequency of occurrence of each ingredient was preserved at the cuisine level ('Ingredient frequency'). In the second control, category composition of each recipe was preserved ('Ingredient category'). A third composite control was created by preserving both category composition of each recipe as well as frequency of occurrence of ingredients ('Category + Frequency').

Interestingly, ingredient frequency came out to be a critical factor that could explain the observed bias in food pairing as reflected in $\overline{N}_s$ (Fig 4). The pattern of food pairing across different size-range of recipes is also consistent with this observation (Fig 5). On the contrary, category composition itself turned out to be irrelevant and led to food pairing that was similar to that of a randomized cuisine. Further, the control implementing a composite model featuring both the above aspects recreated food pairing observed in regional cuisines. Thus frequency of occurrence of ingredients emerged as the most central aspect which is critical for rendering the characteristic food pairing.

Food pairing at recipe level

Looking into the food pairing at recipe level, we analyzed the nature of distribution of food pairing among recipes (N_s^R). Our analysis showed that the negative ΔN_s observed for cuisines was not an averaging effect. The N_s^R values tend to follow exponential distribution, indicating that number of recipes exponentially decays with increasing N_s^R. To address the noise due to small size of cuisines, we computed cumulative distribution ($P(\leq N_s^R)$) as depicted in Fig 6. The nature of cumulative distribution for an exponential probability distribution function ($P(N_s^R) \propto e^{-\alpha N_s^R}$) would be of the following form:

$$P(\leq N_s^R) = a + \frac{k-a}{1 + e^{-\alpha N_s^R}} \tag{1}$$

$$\overline{N_s} = \frac{1}{N_R} \sum_{R_i} N_s^{R_i}$$

CUISINE

Recipe (R_1)   Recipe (R_2)   Recipe (R_i)   Recipe (R_n)

$N_s^{R_1}$ $N_s^{R_2}$ ... $N_s^{R_i}$ ... $N_s^{R_n}$

Recipe (R)   Number of ingredients (s): 4

Number of shared flavour compounds between two ingredients (N)

F_i F_j

$$N = |F_i \cap F_j|$$

Computation of flavour sharing (N_s^R) for a recipe R

N

$$N_s^R = \frac{2}{s(s-1)} \sum_{i,j \in R, i \neq j} |F_i \cap F_j|$$

Fig 3. Schematic for calculation of 'average N_s' ($\overline{N}_s$). Illustration of procedure for calculating the average N_s for a given cuisine. Beginning with an individual recipe, average N_s of the recipe (N_s^R) was calculated. Averaging N_s^R over all the recipes returned $\overline{N}_s$ of the cuisine.
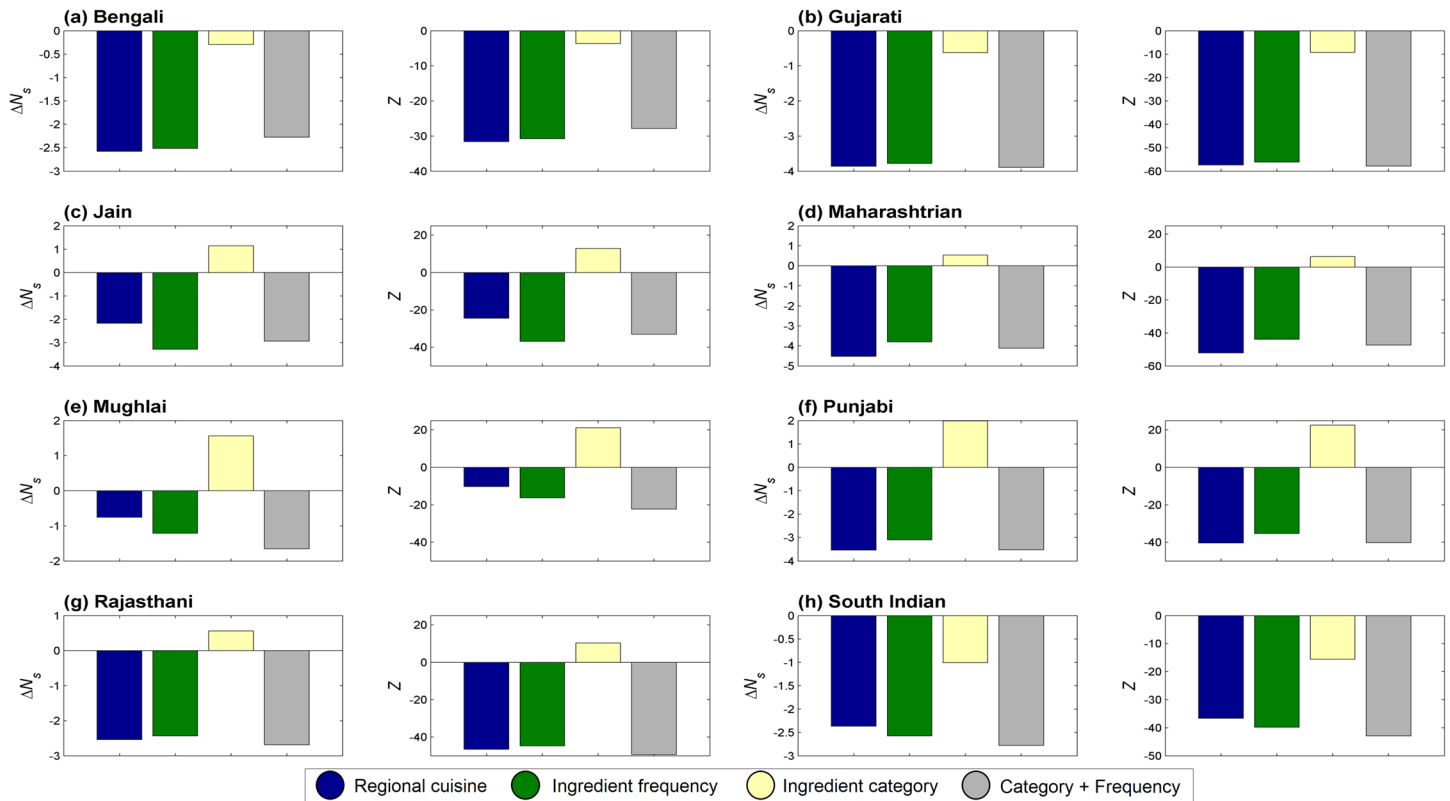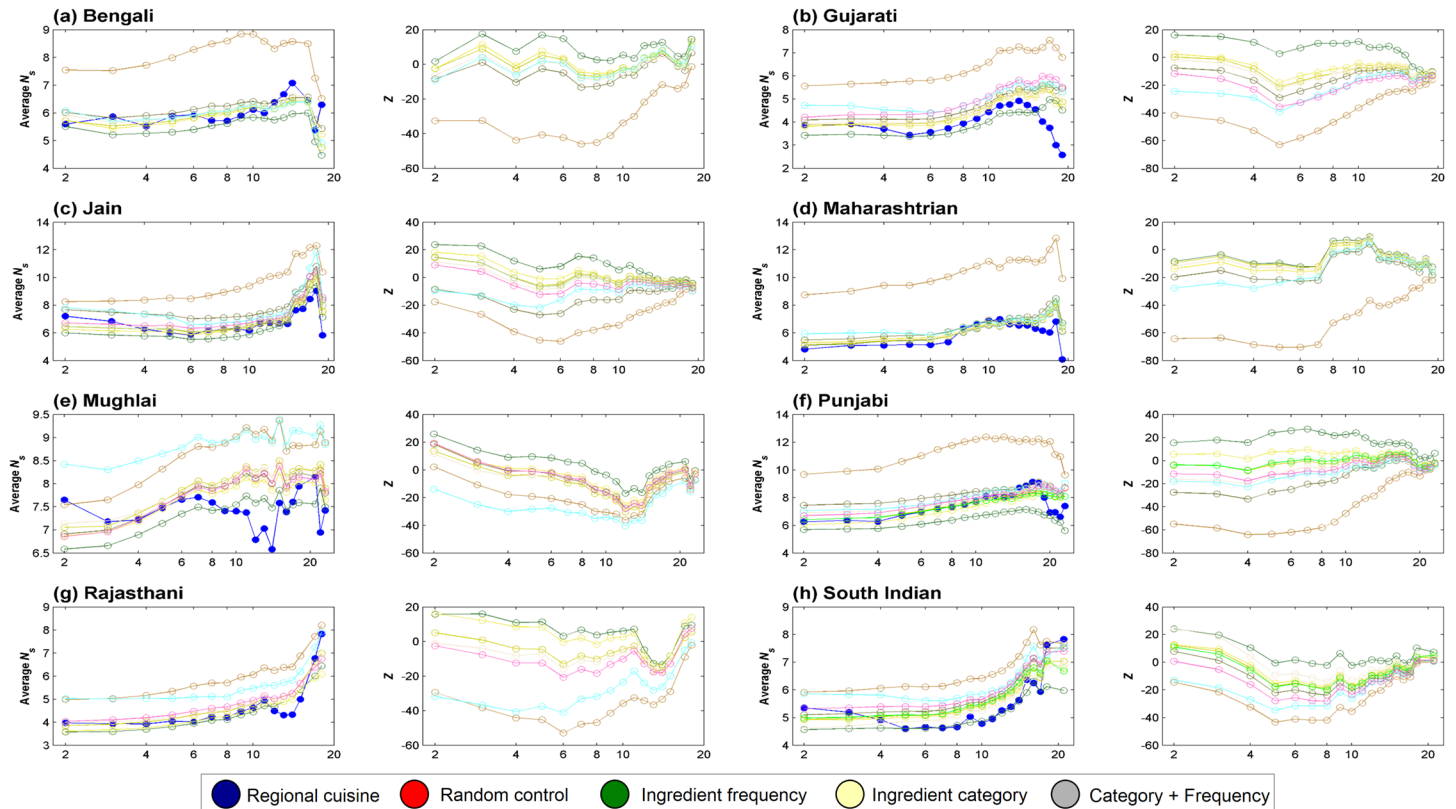
Fig 4. ΔN_s and its statistical significance. The variation in ΔN_s for regional cuisines and corresponding random controls signifying the extent of bias in food pairing. Statistical significance of ΔN_s is shown in terms of Z-score. 'Regional cuisine' refers to each of the eight cuisines analyzed; 'Ingredient frequency' refers to the frequency controlled random cuisine; 'Ingredient category' refers to ingredient category controlling random cuisine; and 'Category + Frequency' refers to random control preserving both ingredient frequency and category. Among all regional cuisines, Mughlai cuisine showed least negative food paring ($\Delta N_s = -0.758$) while Maharashtrian cuisine had most negative food pairing ($\Delta N_s = -4.523$).

We found that all regional cuisines show a strong bias towards recipes of low N_s^R values as observed in Fig 6. For each regional cuisine, the bias was accentuated in comparison to corresponding random cuisines as reflected in the exponents shown in S2 Table. Once again Mughlai cuisine emerged as an outlier, as the nature of its N_s^R distribution did not indicate a clear distinction from that of its random control. Consistent with the observation made with $\overline{N}_s$ and ΔN_s statistics (Figs 4 and 5), we found that controlling for frequency of occurrence of ingredients reproduces the nature of N_s^R distribution across all regional cuisines (barring the Mughlai cuisine). This further highlights the role of ingredient frequency as a key factor in specifying food pairing at the level of recipes as well.

Fig 5. Variation in average N_s and its statistical significance. Change in $\overline{N_s}$ with varying recipe size cut-offs reveals the nature of food pairing across the spectrum of recipe sizes. The $\overline{N_s}$ values for regional cuisines were consistently on the lower side compared to their random counterparts. Category controlled random cuisine displayed average N_s variation close to that of the 'Random control'. Frequency controlled as well as 'Category + Frequency' controlled random cuisines, on the other hand, displayed average N_s variations close to that of the real-world cuisine.

doi:10.1371/journal.pone.0139539.g005

Food pairing at the level of ingredient pairs

Beyond the level of cuisine and recipes, the bias in food pairing can be studied at the level of ingredient pairs. We computed co-occurrence of ingredients in the cuisine for increasing value of flavor profile overlap (N). We found that the fraction of pairs of ingredients with a certain overlap of flavor profiles ($f(N)$) followed a power law distribution $f(N) \propto N^{-\gamma}$ (Fig 7). This indicates that higher the extent of flavor overlap between a pair of ingredients, the lesser is its usage in these cuisines. S3 Table lists the γ values for each of the regional cuisines.

Contribution of individual ingredients towards food pairing

For each of the regional cuisines we calculated the contribution of ingredients (χ_i) towards the food pairing pattern S3 Dataset. For an ingredient whose presence in the cuisine does not lead to any bias, the value of χ_i is expected to be around zero. With increasing role in biasing food pairing towards positive (negative) side, χ_i is expected to be proportionately higher (lower). Fig 8 shows the distribution of ingredient contribution (χ_i) and its frequency of occurrence, for each regional cuisine. Ingredients that make significant contribution towards food pairing could be located, in either positive or negative side, away from the neutral vertical axis around
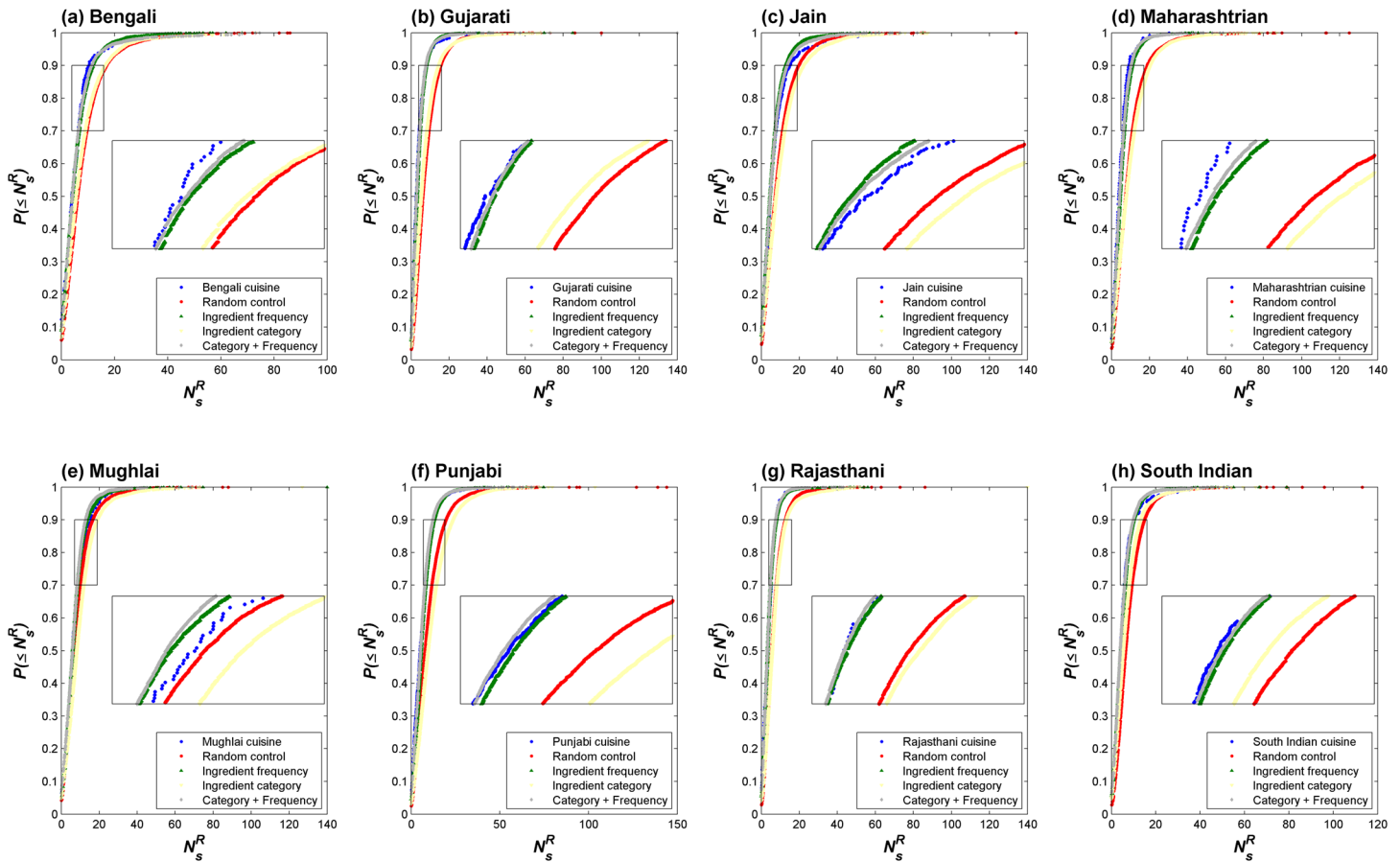
Fig 6. Cumulative probability distribution of N_s^R values for regional cuisines and their random controls. Cumulative distribution of N_s^R indicates the probability of finding a recipe having food pairing less than or equal to N_s^R. The data of regional cuisines as well as those of their controls were fitted with a sigmoid equation indicating that the $P(N_s^R)$ values fall exponentially. The exponent α Eq (1) refers to the rate of decay; larger the α more prominent is the negative food pairing in recipes of a cuisine. As evident from S2 Table, N_s^R distribution of the controls based on 'Ingredient Frequency' as well as 'Category + Frequency' displayed recipe level food pairing similar to real-world cuisines. On the other hand, as also observed at the level of cuisine (Figs 4 and 5), both the 'Random Control' as well as 'Ingredient Category' control deviate significantly.

doi:10.1371/journal.pone.0139539.g006

$\chi_i = 0$. Significantly, spices were consistently present towards the negative side, while milk and certain dairy products were present on the positive side across cuisines. Prominently among the spices, cayenne consistently contributed to the negative food pairing of all regional cuisines. Certain ingredients appeared to be ambivalent in their contribution to food pairing. While cardamom contributed to the positive food pairing in Gujarati, Mughlai, Rajasthani, and South Indian cuisines, it added to negative food pairing in Maharashtrian cuisine. Green bell pepper tends to contribute to negative food pairing across the cuisines except in the case of Rajasthani cuisine. Details of χ_i values of prominent ingredients for each regional cuisine are presented in S4 Table.

Role of ingredient categories in food pairing

As discussed earlier, the random cuisine where only category composition of recipes was conserved, tends to have food pairing similar to that of the 'Random control' (Figs 4 and 5). This
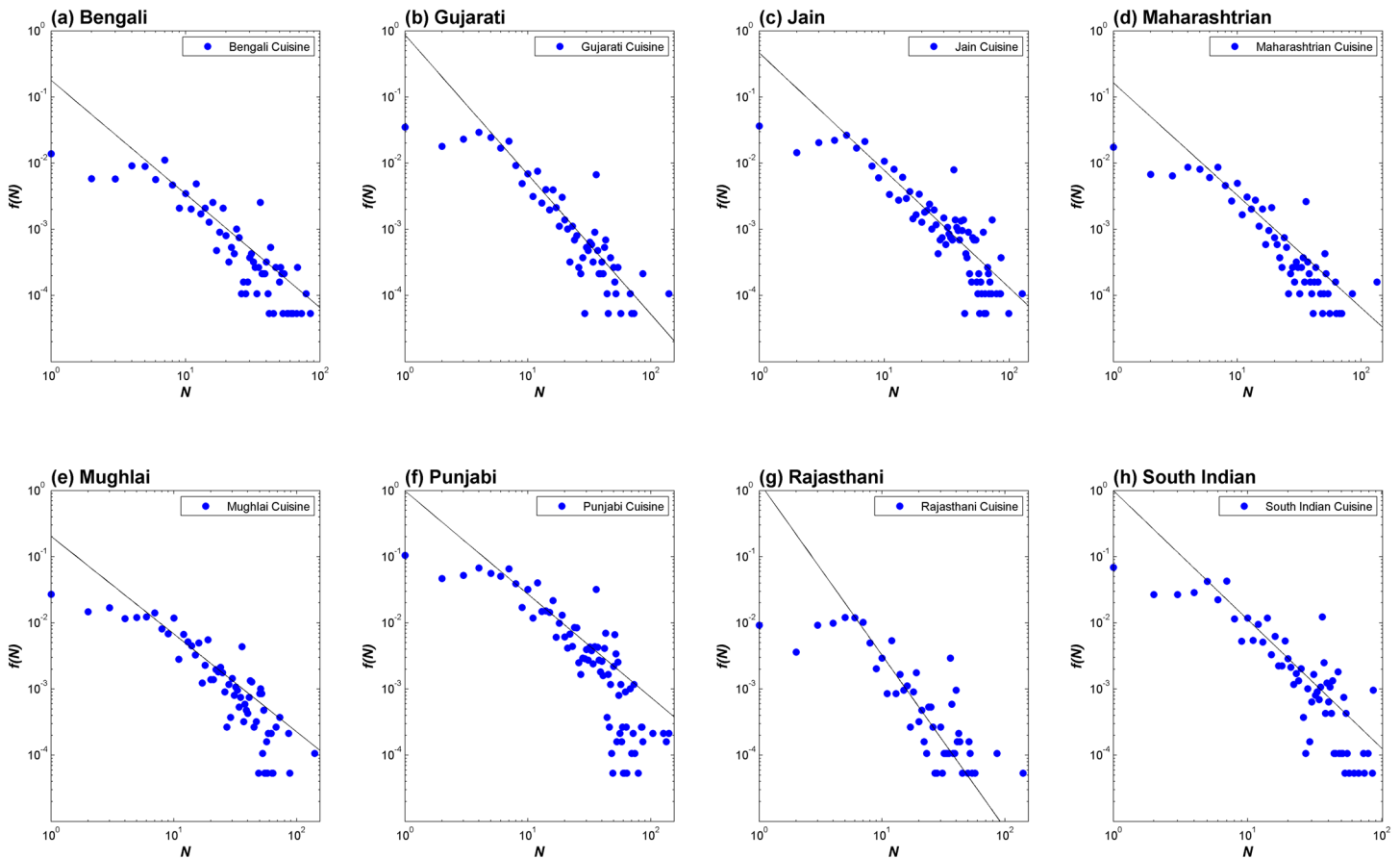
Fig 7. Co-occurrence of ingredients with increasing extent of flavor profile overlap. Fraction of ingredient pair occurrence ($f(N)$) with a certain extent of flavor profile overlap (N) was computed to assess the nature of food pairing at the level of ingredient pairs. Generically across the cuisines it was observed that, the occurrences of ingredient pairs dropped as a power law with increasing extent of flavor profile sharing. This further ascertained negative food pairing pattern in regional cuisines, beyond the coarse-grained levels of cuisine and recipes.

doi:10.1371/journal.pone.0139539.g007

raises the question whether ingredient category has any role in determining food pairing pattern of the cuisine. Towards answering this question, we created random cuisines wherein we randomized ingredients within one category, while preserving the category and frequency distribution for rest of the ingredients. The extent of contribution of an ingredient category towards the observed food pairing in the cuisine is represented by ΔN_s^{cat}. Fig 9 depicts significance of ingredient categories towards food pairing of each regional cuisine. Interestingly, the pattern of category contributions presents itself as a 'culinary fingerprint' of the cuisine.

The 'spice' category was the most significant contributor to negative food pairing across cuisines with the exception of Mughlai cuisine. Another category which consistently contributed to negative food pairing was 'dairy'. On the other hand, 'vegetable' and 'fruit' categories tend to bias most cuisines towards positive food pairing. Compared to the above-mentioned categories, 'nut/seed', 'cereal/crop', 'pulse' and 'plant derivative' did not show any consistent trend. 'Plant' and 'herb' categories, sparsely represented in cuisines, tend to tilt the food pairing towards positive side. In Mughlai cuisine all ingredient categories, except 'dairy', tend to contribute towards positive food pairing. This could be a reflection of the meagre negative food pairing observed for the cuisine (Fig 4). Above observations were found to be consistent across the spectrum of recipe sizes (Fig 10).
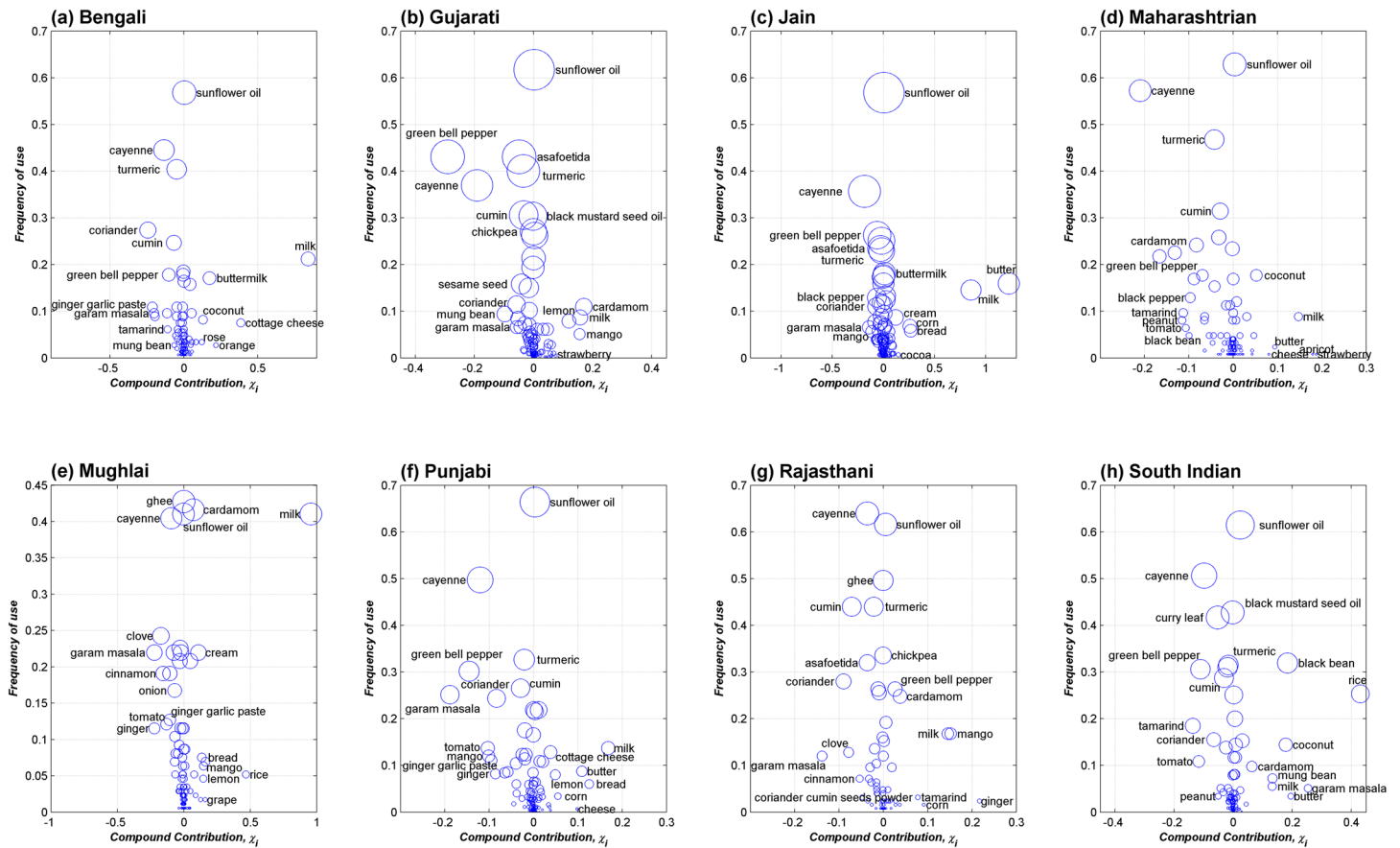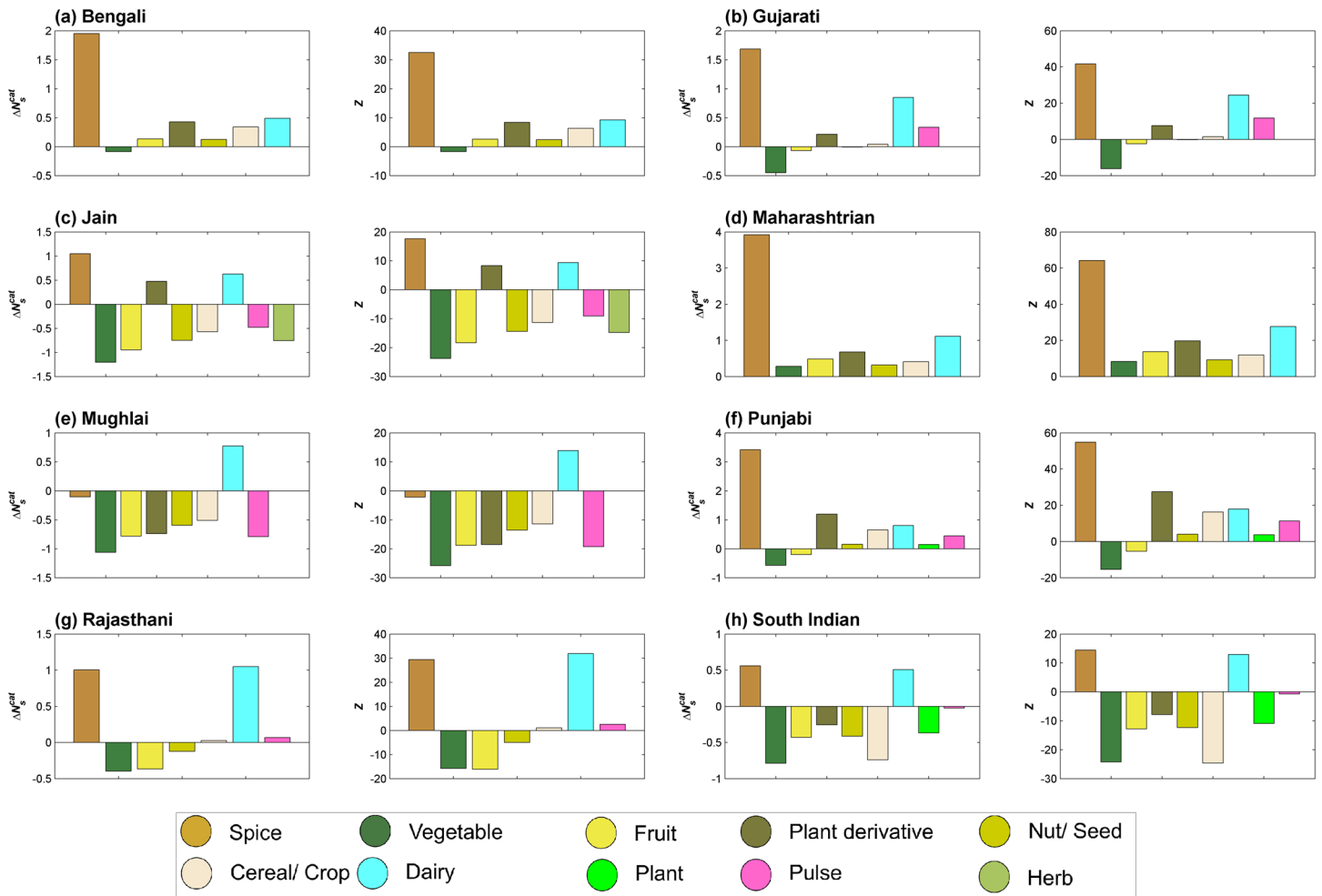
Fig 8. Contribution of ingredients (χ_i) towards flavor pairing. For all eight regional cuisines we calculated the χ_i value of ingredients that indicates their contribution to flavor pairing pattern of the cuisine and plotted them against their frequency of appearance. Size of circles are proportional to frequency of ingredients. Across cuisines, prominent negative contributors largely comprised of spices, whereas a few dairy products consistently appeared on the positive side.

doi:10.1371/journal.pone.0139539.g008

Conclusions

With the help of data analytical techniques we have shown that food pairing in major Indian regional cuisines follow a consistent trend. We analyzed the reason behind this characteristic pattern and found that spices, individually and as a category, play a crucial role in rendering the negative food pairing to the cuisines. The use of spices as a part of diet dates back to ancient Indus civilization of Indian subcontinent [5–7]. They also find mention in Ayurvedic texts such as *Charaka Samhita* and *Bhaavprakash Nighantu* [20–23]. *Trikatu*, an Ayurvedic formulation prescribed routinely for a variety of diseases, is a combination of spices viz., long pepper, black pepper and ginger [24]. Historically spices have served several purposes such as coloring and flavoring agents, preservatives and additives. They also serve as anti-oxidants, anti-inflammatory, chemopreventive, antimutagenic and detoxifying agents [23, 25]. One of the strongest hypothesis proposed to explain the use of spices is the antimicrobial hypothesis, which suggests that spices are primarily used due to their activity against food spoilage bacteria [9, 26]. A few of the most antimicrobial spices [27] are commonly used in Indian cuisines. Our recent studies have shown the beneficial role of capsaicin, an active component in cayenne which was

Fig 9. Contribution of individual categories (ΔN_s^{cat}) towards food pairing bias and its statistical significance. Randomizing ingredients within a certain category provides an insight into their contribution towards bias in food pairing. Spice and dairy category showed up as prominent categories contributing to the negative food pairing of regional cuisines.

doi:10.1371/journal.pone.0139539.g009

revealed to be the most prominent ingredient in consistently rendering the negative food pairing in all regional cuisines [28]. The importance of spices in Indian regional cuisines is also highlighted by the fact that these cuisines have many derived ingredients (such as garam masala, ginger garlic paste etc.) that are spice combinations. The key role of spices in rendering characteristic food pairing in Indian cuisines and the fact that they are known to be of therapeutic potential, provide a basis for exploring possible causal connection between diet and health as well as prospection of therapeutic molecules from food ingredients. Flavor pairing has been used as a basic principle in algorithm design for both recipe recommendation and novel recipe generation, thereby enabling computational systems to enter the creative domain of cooking and suggesting recipes [17, 18]. In such algorithms, candidate recipes are generated based on existing domain knowledge and flavor pairing plays a crucial role while selecting the best among these candidates [18].
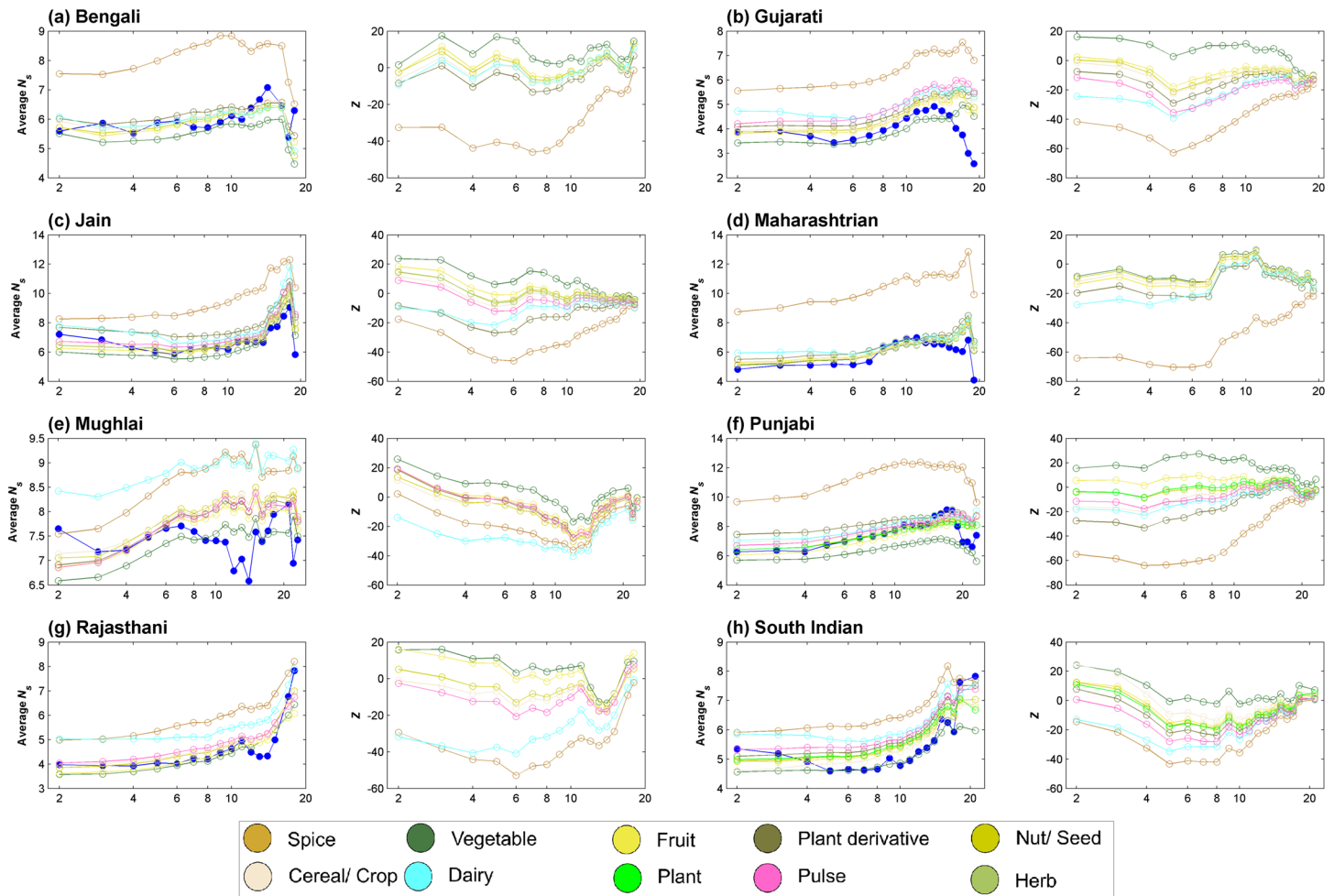
Fig 10. Variation in category contribution and its statistical significance. Across the spectrum of recipe sizes, we observed broadly consistent trend of contribution of individual categories towards food pairing bias.

doi:10.1371/journal.pone.0139539.g010

Materials and Methods

Data collection and curation

The data of regional cuisines were obtained from one of the leading cookery websites of Indian cuisine, tarladalal.com (December 2014). Among various online resources available for Indian cuisine, *TarlaDalal* [19] (http://www.tarladalal.com) was found to be the best in terms of authentic recipes, cuisine annotations and coverage across major regional cuisines. The website had 3330 recipes from 8 Indian cuisines. Among others online sources: Sanjeev Kapoor (http://www.sanjeevkapoor.com) had 3399 recipes from 23 Indian cuisines; NDTV Cooks (http://cooks.ndtv.com) had 667 Indian recipes across 15 cuisines; Manjula's Kitchen (http://www.manjulaskitchen.com) was restricted to 730 Indian vegetarian recipes across 19 food categories; Recipes Indian (http://www.recipesindian.com) had 891 recipes from around 16 food categories; All Recipes (http://www.allrecipes.com) had only 449 recipes from 6 food categories. In

comparison to these sources, Tarladalal.com was identified as a best recipe source of Indian cuisine.

The data of 3330 recipes and 588 ingredients were curated for redundancy in names and to drop recipes with only one ingredient. These ingredients belonged to 17 categories. Ingredients of 'snack' and 'additive' categories, for which no flavor compounds could be determined, were removed. The ingredients were further aliased to 339 source ingredients out of which we could determine flavor profiles for 194 of them. Aliasing involves mapping ingredients to their source ingredient. For example 'chopped potato' and 'mashed potato' were aliased to 'potato'. The final data comprised of 2543 recipes and 194 ingredients belonging to 15 categories. The statistics of regional cuisines, their recipes and ingredient counts is provided in Table 1.

The data of flavor compounds were obtained from Ahn et. al. [15], Fenaroli's Handbook of Flavor Compounds [29] and extensive literature search. All the flavor profiles were cross checked with those in 6th edition (latest) of Fenaroli's Handbook of Flavor Compounds [29] for consistency of names. Chemical Abstract Service numbers were used as unique identifiers of flavor molecules.

Flavor sharing

Flavor sharing was computed for each pair of ingredients that co-occur in recipes in terms of number of shared compounds $N = |F_i \cap F_j|$. Further, the average number of shared compounds in a recipe N_s^R having s ingredients was calculated (Eq (2)).

$$N_s^R = \frac{2}{s(s-1)} \sum_{i,j \in R, i \neq j} |F_i \cap F_j| \tag{2}$$

where F_i represents the flavor profile of ingredient i and R represents a recipe.

For a cuisine with N_R recipes, we then calculated the average flavor sharing of the cuisine $\overline{N}_s^{cuisine} (= \frac{\Sigma_R N_s^R}{N_R})$. Fig 3 illustrates this procedure graphically. We compared average N_s of the cuisine with that of corresponding randomized cuisine (Fig 4) by calculating $\Delta N_s (= \overline{N}_s^{cuisine} - \overline{N}_s^{Rand})$, where $cuisine$ and $Rand$ indicate the regional cuisine and corresponding 'random cuisine' respectively.

A total of four random controls were created viz. 'Random control', 'Ingredient frequency', 'Ingredient category' and 'Category + Frequency'. While in all random cuisines recipe size distribution of the original cuisine was preserved, 'Random control' implemented uniform selection of ingredients (1 set of 10,000 recipes for each regional cuisine); 'Ingredient frequency' control was created while maintaining the ingredient usage frequency distribution (1 set of 10,000 recipes for each regional cuisine); 'Ingredient category' control was created by randomizing ingredient usage in recipes with ingredients belonging to same categories, thus maintaining the category composition of recipes (8 sets of recipes for a total of > 10,000 recipes for each regional cuisine); and 'Category + Frequency' control preserved both the ingredient categories in recipes as well as frequency of overall ingredient usage within the cuisine (8 sets of recipes for a total of > 10,000 recipes for each regional cuisine).

The statistical significance of $\overline{N}_s$ and ΔN_s was measured with corresponding Z-scores given by Eq (3).

$$Z = \sqrt{N_{Rand}} \frac{(\overline{N}_s^{cuisine} - \overline{N}_s^{Rand})}{\sigma_{Rand}}, \tag{3}$$

where N_{Rand} and σ_{Rand} represent the number of recipes in randomized cuisine and standard deviation of N_s^R values for randomized cuisine respectively.

Ingredient contribution

For every regional cuisine, the contribution (χ_i) of each ingredient i was calculated [15] using Eq (4).

$$\chi_i = \left(\frac{1}{N_R} \sum_{i \in R} \frac{2}{n(n-1)} \sum_{j \neq i (j, i \in R)} |F_i \cap F_j| \right) - \left(\frac{2f_i}{N_R \langle n \rangle} \frac{\Sigma_{j \in c} f_j |F_i \cap F_j|}{\Sigma_{j \in c} f_j} \right) \tag{4}$$

Here, f_i is the frequency of occurrence of ingredient i.

χ_i values reflect the extent of an ingredient's contribution towards positive or negative food pairing of the cuisine.

Uniqueness of ingredient category

Despite significant flavor sharing within each category of ingredients, the uniqueness of each category, by virtue of combination of its ingredients with other ingredients, was enumerated by intra-category randomization. The average food pairing of such cuisine, randomized for a category, was compared with that of the original cuisine. Such category-randomized cuisines were created only for major categories (having 5 or more ingredients) within each regional cuisine. The deviation in $\overline{N}_s^{cat}$, that reflects the relevance of unique placements of ingredients of cat, was calculated using Eq (5).

$$\Delta N_s^{cat} = \overline{N}_s^{cat} - \overline{N}_s^{cuisine}, \forall s \geq 2 \tag{5}$$

Here, cat stands for an ingredient category and s represents recipe size. The statistical significance was again calculated using Z-score.

Supporting Information

S1 Table. Distribution of ingredients across categories. Number of ingredients in each category for all regional cuisines.
(PDF)

S2 Table. Exponents (α) of Sigmoid fits for $P(N_s^R)$ vs N_s^R distribution. Exponents (α) for regional cuisines and their random controls.
(PDF)

S3 Table. Power law exponents (γ) for $f(N)$ vs N distribution. Power law exponents (γ) of all regional cuisines.
(PDF)

S4 Table. Ingredients contributing significantly to food pairing. Details of top 10 ingredients contributing to positive and negative food pairing in each of the regional cuisines.
(PDF)

S1 Dataset. Recipes in Indian Subcuisine and their corresponding ingredients. Recipe id and aliased ingredient name.
(XLSX)

S2 Dataset. Flavor compound present in ingredients. Flavor compounds in each ingredient and their corresponding CAS number.
(XLSX)

S3 Dataset. Contribution of each Ingredient towards food pairing and its frequency of occurrence. Frequency of occurrence and corresponding χ_i values of ingredients. (XLSX)

Acknowledgments

Author Contributions

Conceived and designed the experiments: GB. Performed the experiments: AJ RNK GB. Analyzed the data: GB AJ RNK. Wrote the paper: GB AJ RNK.

References

1. Navarrete A, Schaik CPV, Isler K. Energetics and the evolution of human brain size. Nature. 2011; 480:91–93. doi: 10.1038/nature10629 PMID: 22080949

2. Richard Wrangham. Catching fire: how cooking made us human. Basic Books; 2009.

3. Fonseca-Azevedo K, Herculano-Houzel S. Metabolic constraint imposes tradeoff between body size and number of brain neurons in human evolution. Proceedings of the National Academy of Sciences of the United States of America. 2012 Nov; 109(45):18571–6. doi: 10.1073/pnas.1206390109 PMID: 23090991

4. Pollan M. Cooked: A Natural History of Transformation. Penguin Books; 2014.

5. Weber S, Kashyap A, Mounce L. Archaeobotany at Farmana: new insights into Harappan plant use strategies. In: Shinde V, Osada T, Kumar M, editors. Excavations at Farmana, District Rohtak, Haryana, India. Kyoto: Nakanish Printing; 2011. p. 808–823. Available: http://anthro.vancouver.wsu.edu/media/PDF/archaeobotany_at_Farmana_2011.pdf

6. Kashyap A, Steve W. Harappan plant use revealed by starch grains from Farmana. Antiquity. 2010; 84 (326). Available: http://antiquity.ac.uk/projgall/kashyap326/

7. Lawler A. The Ingredients for a 4000-Year-Old Proto-Curry. Science. 2012; 337(6092):288. doi: 10.1126/science.337.6092.288-a

8. Appadurai A. How to Make a National Cuisine: Cookbooks in Contemporary India. Comparative Studies in Society and History. 2009 Jun; 30(01):3–24. doi: 10.1017/S0010417500015024

9. Sherman PW, Billing J. Darwinian Gastronomy: Why we use spices. Spices taste good because they are good for us. Bioscience. 1999; 49(6):453–463. doi: 10.2307/1313553

10. Zhu YX, Huang J, Zhang ZK, Zhang QM, Zhou T, Ahn YY. Geography and similarity of regional cuisines in China. PloS one. 2013 Jan; 8(11):e79161. doi: 10.1371/journal.pone.0079161 PMID: 24260166

11. Birch LL. Development of food preferences. Annual Review of Nutrition. 1999; 19:41–62. doi: 10.1146/annurev.nutr.19.1.41 PMID: 10448516

12. Ventura AK, Worobey J. Early influences on the development of food preferences. Current biology: CB. 2013 May; 23(9):R401–8. doi: 10.1016/j.cub.2013.02.037 PMID: 23660363

13. Blumenthal H. The big fat duck cookbook. Bloomsbury Publishing PLC; 2008.

14. Ahnert SE. Network analysis and data mining in food science: the emergence of computational gastronomy. Flavour. 2013; 2(4):1–4. Available: http://www.flavourjournal.com/content/2/1/4

15. Ahn YY, Ahnert SE, Bagrow JP, Barabási AL. Flavor network and the principles of food pairing. Scientific reports. 2011 Jan; 1:196. doi: 10.1038/srep00196

16. Jain A, Rakhi N, Bagler G. Spices form the basis of food pairing in Indian cuisine. arXiv:150203815. 2015;p. 1–30. Available: arxiv:1502.03815. Accessed 2 June 2015.

17. Varshney LR, Pinel F, Varshney KR, Bhattacharjya D, Schoergendorfer A, Chee YM. A Big Data Approach to Computational Creativity. 2013 Nov;(October 2013):1–16. Available: arxiv:1311.1213. Accessed 2 June 2015.

18. Teng CY, Lin YR, Adamic LA. Recipe recommendation using ingredient networks; 2012. Available: arxiv:1111.3919v3. Accessed 2 June 2015.

19. Dalal T. Tarladalal.com; 2014. Available: http://www.tarladalal.com/ Accessed 1 February 2015.

20. Atridevji Gupt. Swami Agnivesha's Charaka Samhita. 2nd ed. Banaras: Bhargav Pustakalay; 1948.

21. Valiathan MS. The Legacy of Caraka. Universities Press; 2010.

22. Pande GS, Chunekar KC. Bhavprakash Nighantu. Varanasi: Chaukhambha Bharti Academy; 2002.

23. Tapsell LC, Hemphill I, Cobiac L, Sullivan DR, Fenech M, Patch CS, et al. Health benefits of herbs and spices: the past, the present, the future. Medical Journal of Australia. 2006; 185(4):S1–S24.

24. Johri RK, Zutshi U. An Ayurvedic formulation 'Trikatu' and its constituents. Journal of Ethnopharmacology. 1992 Sep; 37(2):85–91. doi: 10.1016/0378-8741(92)90067-2 PMID: 1434692

25. Krishnaswamy K. Traditional Indian spices and their health significance. Asia Pacific journal of clinical nutrition. 2008 Jan; 17 Suppl 1:265–268. PMID: 18296352

26. Billing J, Sherman PW. Antimicrobial functions of spices: why some like it hot. The Quarterly review of biology. 1998 Mar; 73(1):3–49. doi: 10.1086/420058 PMID: 9586227

27. Rakshit M, Ramalingam C. Screening and Comparision of Antibacterial Activity of Indian Spices. Journal of Experimental Sciences. 2010; 1(7):33–36.

28. Perumal S, Dubey K, Badhwar R, George KJ, Sharma RK, Bagler G, et al. Capsaicin inhibits collagen fibril formation and increases the stability of collagen fibers.; 2014.

29. Burdock GA. Fenaroli's Handbook of Flavor Ingredients. 6th ed. CRC Press; 2010.