

Total ingredient profiling of human and animal foods

Padmini Ramachandran¹, Andrea Ottesen², Elizabeth Reed¹, Mark Mammel³, and Monica Pava-Ripoll⁴

¹ Center for Food Safety and Applied Nutrition, Office of Regulatory Science, College Park, MD, ² Center for Veterinary Medicine, Office of Research, Laurel, MD, ³ Center for Food Safety and Applied Nutrition, Office of Applied Research and Safety Assessment, Laurel, MD, ⁴ Center for Food Safety and Applied Nutrition, Office of Food Safety, College Park, MD.



Abstract

DNA of mitochondrial and chloroplast genomes are useful for molecular identification of target and non-target organisms, and ingredients in foods, supporting molecular epidemiology for source attribution of pathogens and contaminants. In-house databases (MitochonTrakr and ChloroplasTrakr) and k-mer based pipelines (MitoK-mer and ChloroK-mer) targeting identification of species composition in food and feed matrices were developed and are publicly accessible on www.galaxytrakr.com. The purpose of this study was to apply in-house k-mer pipelines (ChloroK-mer and MitoK-mer) to metagenomic sequence data from single and multi-ingredient foods to identify total composition including intended and unintended contaminants. DNA was extracted from single-ingredient (dried fish maw, ground oregano, marjoram and paprika) and multi-ingredient (white bread, papads, sandwich cookies, garam masala and raw pet food). Libraries (Illumina DNA Prep) were sequenced on the Illumina Miseq. Sequence reads were annotated using MitoK-mer and ChloroK-mer pipelines with databases of mitochondrial and chloroplast genomes. K-mer tools successfully provided species level identification of ingredient(s) in accordance with food labeling. For single-ingredient foods using MitoK-mer, dried fish maw and paprika were accurately identified as *Alopius pelagicus* (shark) and *Capsicum annuum* (chili pepper), respectively. For low complexity multi-ingredient food such as white bread, main ingredients identified were *Saccharomyces cerevisiae* (yeast) and *Triticum aestivum* (wheat). *Vigna radiata* (mung bean), *Vigna angularis* (lentil) were identified in accordance with the label in papads. At the time of analysis, the database did not contain *Cumin cyminum* (cumin) and *Coriander sativum* (coriander) (Apiaceae) the two major ingredients of garam masala. Thus, the pipeline returned an annotation of the closest relative in the database; *Daucus carota*. Raw pet food ingredients were identified in accordance with the label as *Gallus gallus* (chicken). ChloroK-mer correctly identified all plant ingredients in accordance with labels for all foods examined.

Results

We created MitochonTrakr, a publicly accessible, high-quality reference collection containing ~11K eukaryote mitochondrial genomes of animal and plant origin. We also created a reference collection of ~6000 publicly available chloroplast genomes. Using sequence fragments from these reference collections, we developed an in-house analytical pipeline with species specific K-mers to aid the identification of plant and animal species in food samples.

MitoK-mer provided species level identification of the ingredient(s) listed on the labeling for all products evaluated. However, accurate species identification was dependent of the particular plant or animal species included in the database. For example, at the time of analysis, the database did not contain *Cumin cyminum* (Apiaceae; cumin) or *Coriander sativum* (Apiaceae; coriander) the two major ingredients of garam masala. Thus, the pipeline returned an annotation of the closest relative in the database *Daucus carota* (Apiaceae; carrot) (Figure 1 and Table 1). In multi-ingredient products, the major ingredients were correctly identified. For example, species identification for white bread *Triticum aestivum* (wheat) and *Saccharomyces cerevisiae* (yeast). Raw pet food ingredients were identified in accordance with the label as *Gallus gallus* (chicken). ChloroK-mer correctly identified all plant ingredients in accordance with labels for all foods examined (Figure 2).

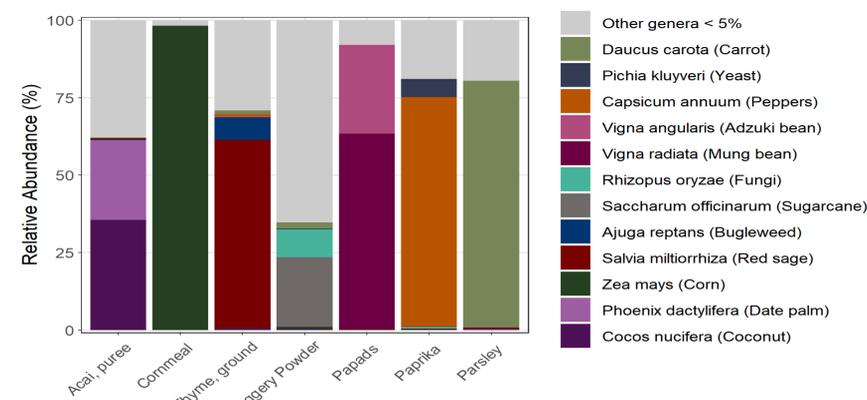


Figure 1. Identification of plant and animal species using MitochonTrakr and MitoK-mer DB.

Table 1. Identification of species of some single ingredient products tested.

| Product | Scientific name [Genus, species (Family)] | Included in Mitok-mer DB | Species identified using MitochonTrakr | Included in ChloroK-mer DB | Species identified using Chloroplast database |
|----------------|---|--------------------------|---|----------------------------|---|
| Cornmeal | <i>Zea mays</i> (Poacea) | Yes | <i>Zea mays</i> (Poacea) | Yes | <i>Zea mays</i> (Poacea) |
| Paprika | <i>Capsicum annuum</i> (Solanaceae) | Yes | <i>Capsicum annuum</i> (Solanaceae) | Yes | <i>Capsicum annuum</i> (Solanaceae) |
| Jaggery powder | <i>Saccharum officinarum</i> (Poacea) | Yes | <i>Saccharum officinarum</i> (Poacea) | Yes | <i>Saccharum officinarum</i> (Poacea) |
| Oregano | <i>Origanum vulgare</i> (Lamiaceae) | No | <i>Salvia miltiorrhiza</i> (Lamiaceae) | Yes | <i>Origanum vulgare</i> (Lamiaceae) |
| Marjoram | <i>Origanum majorana</i> (Lamiaceae) | No | <i>Salvia miltiorrhiza</i> (Lamiaceae) | No | <i>Origanum sp.</i> (Lamiaceae) |
| Acai | <i>Euterpe oleracea</i> (Arecaceae) | No | <i>Cocos nucifera</i> (Arecaceae) <i>Phoenix dactylifera</i> (Arecaceae) | Yes | <i>Euterpe oleracea</i> (Arecaceae) |
| Dried fish maw | Unknown | - | <i>Alopius pelagicus</i> (Alopiidae; shark) | - | - |

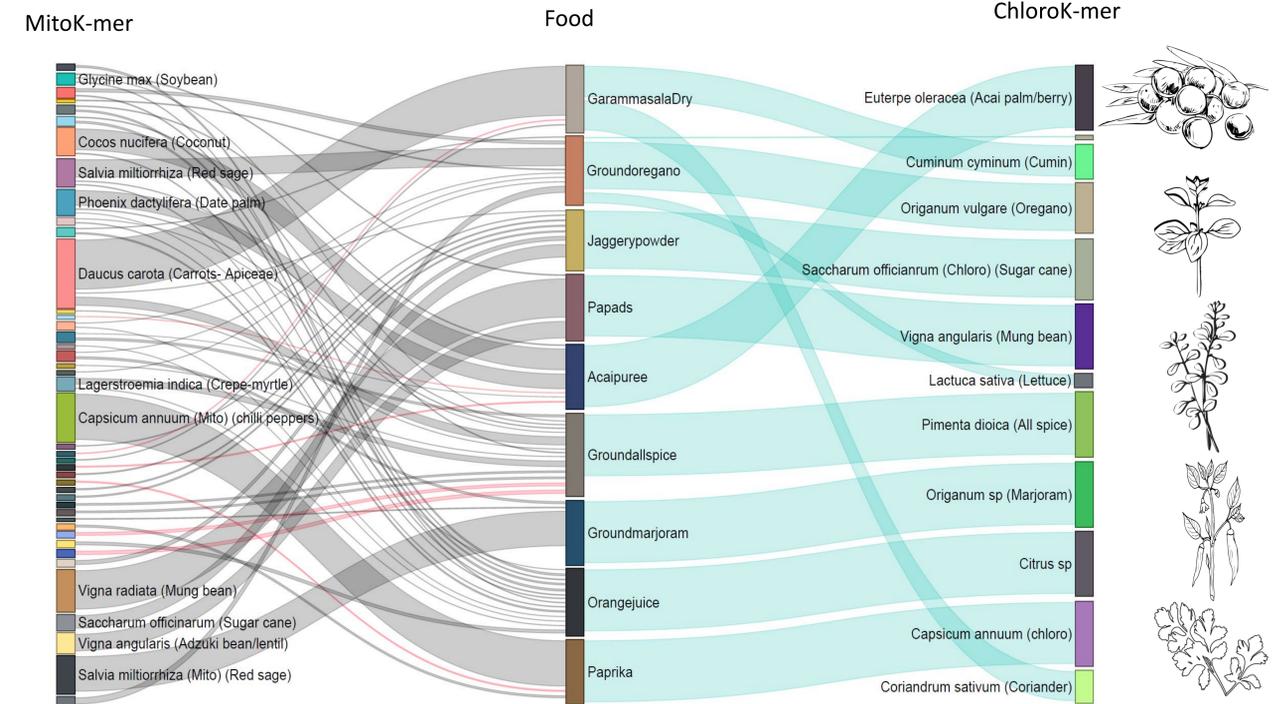


Figure 2. Sankey diagram representing some single ingredient and multi-ingredient food products tested. One specific food commodity of interest here is Papads. Both ChloroK-mer and MitoK-mer correctly identified ingredients as *Vigna radiata* (mung bean) and *Vigna angularis* (bean/lentil). Another food commodity of interest is paprika which was correctly identified as *Capsicum annuum* by both MitoK-mer and ChloroK-mer.

Materials and Methods

Single and multi-ingredient food of plant and animal origin were selected for this study. Single ingredient food included dried fish maw, corn meal, spices (ground oregano, marjoram, parsley, thyme, nutmeg, and paprika), and fruit juices, nectars, or concentrates (orange, guava, and acai). Multi ingredient food included white bread, papads, sandwich cookies, garam masala, and raw pet food.

Samples were homogenized prior to genomic DNA extraction by vigorously shaken the food container for 30-60s. DNA extraction was performed in 5-10 g sample using the Qiagen DNeasy PowerMax Soil Kit per manufacturer's instructions. Extracted gDNA was quantified using the Invitrogen Qubit™ fluorometer and Next Generation Sequencing (NGS) libraries were prepared using Illumina Nextera DNA Library Prep Kit. NGS was performed using the Illumina Miseq instrument after normalizing pooled libraries to 2 nM. Sequence reads were queried against previously constructed k-mers from references sequences from in-house mitochondria and chloroplast reference collections. A comprehensive list of all plant or animal species detected in a sample was generated by estimating the relative abundance of sequence reads.

Conclusions

- Identification of all ingredients and contaminants in human and animal food is needed for maintaining food/feed integrity and safety.
- Ingredients of a food commodity can be identified by shotgun sequencing method.
- Species not listed in the databases were matched to the nearest species within the same taxonomic family.
- The pipelines will identify additional species as more mitochondrial and chloroplast genomes are added to the respective databases.
- The collection of mitochondrial genomes and their raw data is publicly available at the National Center for Biotechnology Information (NCBI) which is under the MetagenomeTRAKR umbrella (Bioproject PRJNA423170).

Significance

The K-mer based mitochondrial and chloroplast reference databases deliver high resolution identification of plant and animal-based foods, providing a valuable tool in food integrity and safety.