Chapter 4
DNA-based authentication of Triphala polyherbal formulation

4.1 Introduction

The metabarcoding for herbal formulations, introduced by Coghlan et al. (2012) was initially used to identify plant and animal raw materials in traditional Chinese medicines. They employed the p-loop region of the plastid *trnL* gene for plants and 16S mtDNA marker for animals. Subsequently, other researchers have conducted studies on metabarcoding for authenticating and identifying plant materials in herbal medicines, employing both single and multi-barcode approaches. For instance, Yao et al. (2022) and Cheng et al. (2014) utilized a multi-barcode approach combining *ITS2* and *trnL* markers in their analysis to detect plant species within various traditional Chinese medicines. Urumarudappa et al. (2020) meanwhile, used the *ITS2* and *rbcL* barcodes in their method to detect plant species in Thai herbal medicines. During the single-drug analysis of *Echinacea* species, *Hypericum perforatum*, and *Veronica officinalis* sold in the European market, significant unlabeled species were discovered at rates of 89%, 68%, and 15% respectively through their utilization of *ITS1* and *ITS2* barcodes (Raclariu et al., 2017; Raclariu, Ţebrencu, et al., 2018). Seethapathy et al. (2019) conducted a study on 12 single drugs and 27 formulations, finding that the fidelity for single-ingredient and multi-ingredient products was 67 % and 21 %, respectively.

ITS2 has proven significant due to its high interspecific variation, allowing for precise differentiation among closely related species. Its short sequence length aids in efficient DNA recovery and sequencing, making it especially useful in fields like medicinal plant research where accurate species identification is crucial. The previous Chapter 3 metabarcode studies demonstrated that the *ITS2* metabarcode exhibited greater discriminatory power to *rbcL*. This was particularly evident for TB, TC and PE, where *ITS2* facilitated species-level identification. This study aims to fill this gap by proposing a comprehensive multi-methodological DNA-based approach to authenticate ‘Triphala’ formulation. The primary objective is to evaluate the functionality of DNA-based

authentication methods by employing *ITS2*-based species-specific assays and metabarcoding techniques on this complex polyphenol-rich formulation.

4.2 Materials and Methods

4.2.1 Triphala Mock Control Preparation

Fruits of TB, TC and PE were ground and combined in equal proportions**.** Subsequently, the process involved DNA isolation, species-specific assays, and metabarcoding analyses**.**

4.2.2 Plant Collection, Market Samples, DNA Isolation, Species-Specific PCR

Plant leaves and fruits for TB, TC and PE were collected and authenticated by *rbcL* sequencing as described in Chapter 2, section 2.2.1. Six commercially available "Triphala powder" formulations were procured from local markets in Gandhinagar and Ahmedabad, Gujarat, India. DNA extraction from reference plant materials, a Triphala mock control, and the six market formulations were conducted following the optimized protocol outlined in Chapter 2, section 2.2.2. For the DNA quality check DNA quantification and PCR with *ITS2* were carried out as described in Chapter 2, section 2.2.3. The specific primer sequences and PCR reaction conditions utilized for the species-specific PCR assay targeting Triphala mock control samples and market formulations are elaborated in detail within Chapter 2, Section 2.2.1.

4.2.3 *ITS2* Metabarcoding and Bioinformatic Analysis

DNA from the Triphala mock control and the six market formulations was amplified using fusion *ITS2* metabarcode for the library preparation, following the PCR reactions and thermal cycling conditions detailed in Chapter 3, section 3.2.3. Library preparation for next-generation sequencing was performed according to the protocol outlined in Chapter 3, Section 3.2.5. The reads of resulting data were analyzed using optimized bioinformatic pipelines described in Chapter 3, section 3.2.6. to identify and quantify the constituent species within each sample.

4.3. Results and Discussion

4.3.1 DNA Isolation and Quality Control

DNA isolation from Triphala, consisting of three polyphenol-rich dried fruits (TB, TC, and PE), presents challenges for molecular methods-based authentication processes. The inherent properties and diversity in secondary metabolites, tissue types, and processing conditions pose limitations on obtaining PCR-amplifiable, high-quality DNA for molecular assays. The total DNA yield obtained was in the range of 7.2 to 9.6 µg/g of powder used. The quality of the DNA was evaluated using PCR amplification of *ITS2* metabarcode. Results revealed positive amplification in both mock controls and market samples (Figure 4.1). The positive amplification indicated that the DNA extracted from both the mock control and market Triphala samples was suitable for downstream analysis. However, this did not confirm the presence of the labeled species in the Triphala formulations. Therefore, additional species-specific PCR tests and metabarcoding were carried out to comprehensively identify the targeted species and detect any untargeted plant-derived adulteration.

4.3.2 Species-Specific PCR Assay

Species-specific PCR was carried out using TB, TC and PE specific primer sets, with the optimized PCR condition for mock controls and market samples (Figure 4.2a; 4.2b; 4.2c). However, the PE primer did not amplify *P. emblica* in two market Triphala samples, indicating the possibility of false-negative results due to low or poor DNA quality (Figure 4.1d). Furthermore, our prior research involved the development of species-specific PCR assays for Tulsi (*O. tenuiflorum* and *O. basilicum*) (Travadi et al., 2022), Brahmi (*B. monnieri* and *C. asiatica)* (Shah et al., 2023) aimed at authenticating various market formulations. Similar investigative studies have been conducted on *Terminalia arjuna* (Sharma et al., 2017), *Senna tora* (Seethapathy et al., 2015), *Portulaca oleracea* (Zhang et al., 2023), *Viscum coloratum* (Kim et al., 2020), *Aristolochia* species (Doganay-Knapp et al., 2018), and *Isatidis Folium* (Hsieh et al., 2021). These investigations contribute to the expanding scientific knowledge, utilizing molecular authentication methods across diverse botanical products, highlighting the critical importance of rigorous authentication protocols in maintaining the quality control of herbal products.

4.3.3 Metabarcoding of Mock Controls and Herbal Products

To evaluate the applicability of the new *ITS2* metabarcode, mock controls and commercial Triphala products were subjected to an optimized metabarcoding pipeline. A total of 18,624 and 6,10,049 reads (an average of 628673 for each sample) were obtained for the blended mock control and market Triphala formulations, respectively. After quality filtering by removing reads shorter than 300 bp, 93.1 % of obtained reads were subjected to clustering, followed by BLASTn search of the representative sequence against the NCBI nr database (Table 4.1). In the in-house mock control, all three species were identified with read abundances of 75%, 22.7%, and 1.74% for TB, TC, and PE, respectively. Although PE was not detected in our previous Rasayana powder samples containing *T. cordifolia*, *T. terrestris*, and *P. emblica*, here *P. emblica* was successfully detected in mock control (Travadi et al., 2023). This improved detection is likely attributed to an enhanced DNA extraction protocol, which facilitated more effective PE fruit DNA detection. This improvement highlights the importance of using reliable and optimized DNA extraction methods for accurate detection and authentication of plant species in herbal products. The ability to detect all three target species validates the efficacy of the metabarcoding method for qualitative assessment of market formulations.

The metabarcoding analysis identified a total of 13 plant species of 12 genera belonging to 10 families from the six market samples. In the Triphala formulation sold in the market, TB had the highest percentage of reads (ranging from 78% to 35.4%), followed by TC (with a range of 34.9% to 15.8%) and PE (with a range of 6.89% to 0.01%). Notably, it was also observed 100% concordance of metabarcoding data with the species-specific PCR assay results. In the metabarcoding data for sample ID T1 and T2, the presence of PE was not observed in any of the sequencing reads. On the other hand, PE was detected in the mock control and four of the market Triphala samples. This discrepancy highlights inconsistencies in the composition and quality of commercial Triphala products, indicating potential mislabeling or variations in ingredient inclusion. Five prominent, untargeted plant species were detected, averaging 21.1% of total reads across all six samples (Figure 4.3). *Cuminum cyminum* comprised 37.5% and 7.3% of reads in T2 and T4, respectively. *Senna tora* represented 28.2%, 2.9%, and 5.9% of reads in T3, T4, and T5, respectively. *Moringa oleifera* accounted for 16.5% and 0.2%

of reads in T1 and T2, respectively. *Ipomoea batatas* made up 23.9% of reads in T5, while *Eclipta alba* constituted 6.7% of reads in T4.

A total of 25.35% of the reads across the six market samples corresponded to non-targeted plant species (Table 4.3). There was substantial variation in read abundance among families, ranging from 73% for Combretaceae to 1.64% for Phyllanthaceae (Figure 4.4a). Out of these readings, 8.1% belonged to Apiaceae, 6.2% to Moringaceae, 4% to Convolvulaceae, and 3.5% to Fabaceae. Fabaceae was present in 50% of the samples (3 out of 6), while both Moringaceae and Apiaceae appeared in 33% (2 out of 6) (Figure 4.4b).

Potential sources of contamination range from cultivation and harvesting stages to transport, storage, production, and packing, inadequately cleaned containers, conveyors, and equipment could introduce trace contamination, complicating accurate quantitative assessments (Liu et al., 2018; Raclariu & de Boer 2023; Raclariu, Heinrich, et al.,2018; Wu & Shaw 2022). Although DNA metabarcoding is one of the highly sensitive methods that can detect even trace amounts of contamination, the quantitative aspects are complicated. DNA degradation during drying, storage, and DNA extraction from botanical products such as Triphala, a blend of three dried fruits rich in polyphenols, is an important consideration for validation procedures. Several research studies have indicated that the quality of DNA extraction from plant tissues with varying types and ages can significantly impact PCR amplification and the success of sequencing (Ivanova et al., 2016; Raclariu et al., 2021; Raclariu, Heinrich, et al., 2018). Overall, these biases are attributed to various factors including relative DNA quality and quantity to plant species, inheritance-related biases linked to complex metabolites; potential degradation or loss of DNA through processing; PCR amplification bias due to poor primer fit variations with different plant species; parameters set for data analysis; molecular identification algorithm and reference databases (Arulandhu et al., 2019; Staats et al., 2016).

4.4 Conclusion

In conclusion, species-specific PCR and metabarcoding analyses revealed a discrepancy in six out of two of the commercial Triphala samples (T1 and T2), where *Phyllanthus emblica* DNA was not detected. This finding highlights the enhanced specificity and sensitivity of DNA-based methods in verifying the authenticity and

complete composition of complex herbal formulations like Triphala. Species-specific PCR assays only detect the targeted species, while metabarcoding provides a broader view of all plant species present in a sample, including nontargeted ones. Thus, using species-specific PCR and metabarcoding simultaneously could play vital roles in identifying and monitoring the presence of target species and adulterants, providing valuable insights into the authenticity and quality of herbal formulations like Triphala. Further investigation into various DNA isolation protocols and PCR dynamics holds the potential for refining these molecular approaches.

Table 4.1 Details of reads obtained through *ITS2* metabarcoding for mock controls and Triphala market formulations.

Control of mock description	Sample ID	Reads before filtering	Cumulative reads per control	Reads after filtering	Percentage of analyzed reads from filtered reads[#]
Mock control	MC	18624		8915	94.1
Triphala market formulation	T1	150184	628673	55430	89.7
	T2	67687		43335	89.2
	T3	197306		51474	92.7
	T4	35286		26962	98.1
	T5	134715		58007	93.4
	T6	24871		13042	95.5

[#]Analysed reads (from filtered reads) after discarding OTU clusters having <5 reads.

Table 4.2 Relative abundance (% reads) of non-targeted plant species detected within mock control and Triphala market formulations.

Description	**Sample ID**	**Non targeted reads**	**Percentage of reads from total analyzed reads#**
Mock control	MC	*Others*	0.48
Triphala market formulation	T1	*Cajanus cajan*	3.47
		Uncultured eukaryote	1.67
		Olea europaea	1.45
		Syzygium cumini	1.25
		Others	1.35
	T2	*Seseli diffusum*	2.71
		Others	1.04
	T3	*Lathyrus sativus*	2.82
		Uncultured eukaryote	1.48
		Others	6.51
	T4	*Others*	3.02
	T5	*Others*	2.21
	T6	*Others*	0.64

#Plant species that comprised <1% of reads are represented as others.

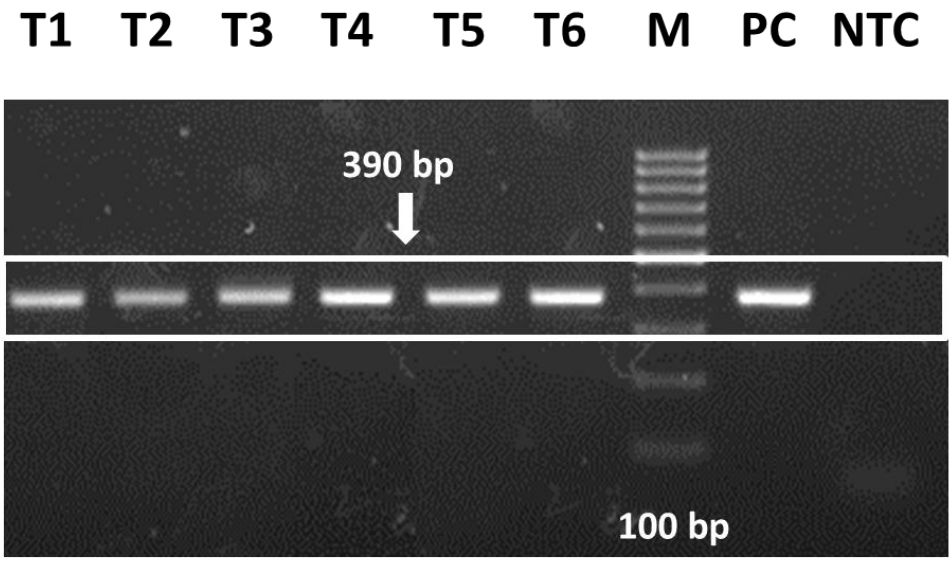

Figure 4.1 PCR using *ITS2* metabarcode with mock control (PC) and Triphala market formulation (T1 to T6). NTC = No Template DNA; M =100 bp DNA marker (100 - 1000 bp); Expected amplicon Sizes = 390 bp.

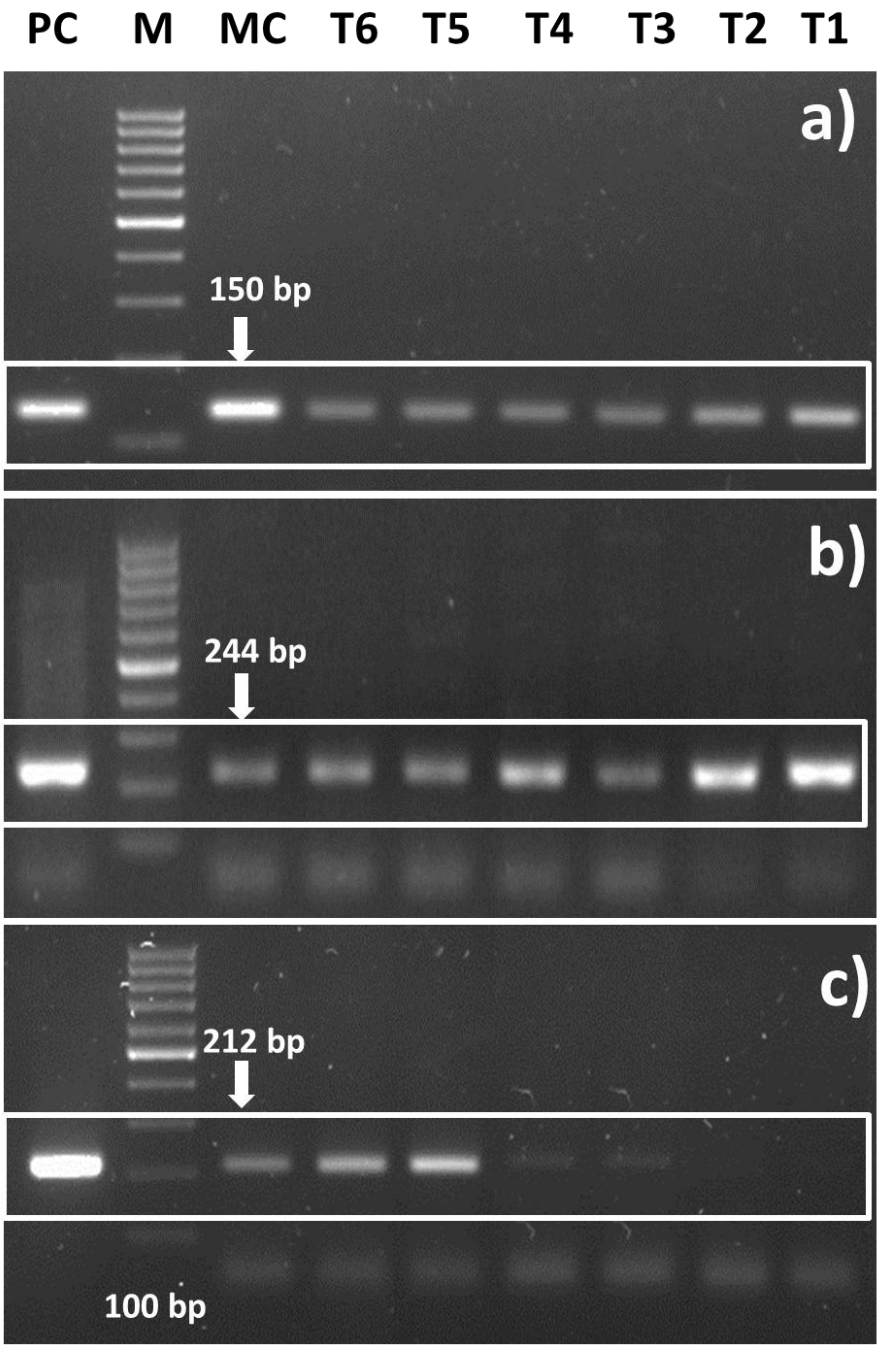

Figure 4.2 Species-specific PCR assay results with mock control (MC) and Triphala market formulation (T1 to T6) with respective primers. a) *Terminalia bellirica* (TB), b) *Terminalia chebula* (TC) and c) *Phyllanthus emblica* (PE). PC = DNA isolated from respective dried fruit tissue; M 100 bp DNA marker (100 -1000 bp); Expected amplicon Sizes: TB = 150 bp, TC = 244 bp, PE = 212 bp.

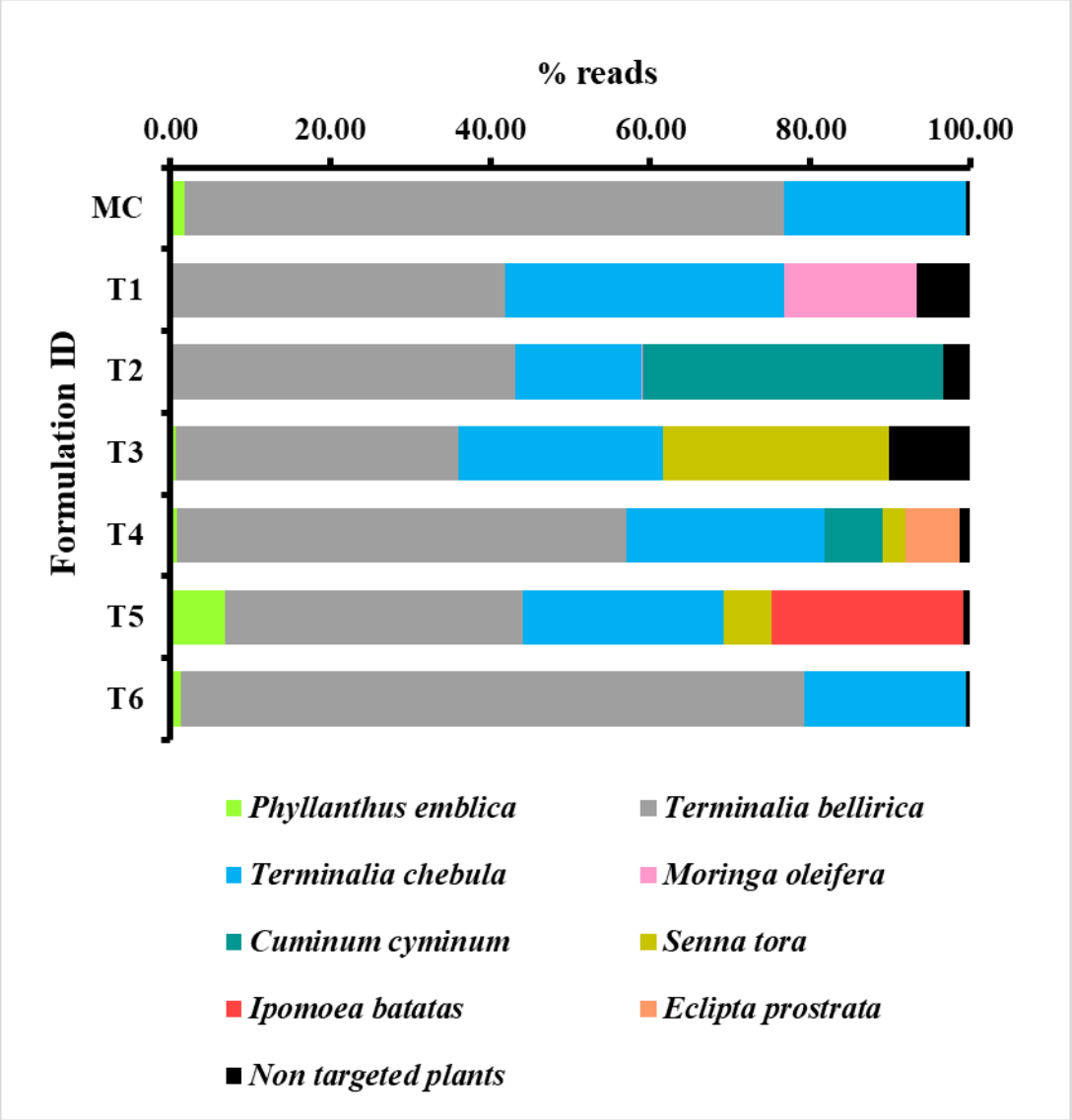

Figure 4.3 Relative abundance (% reads) of the plant species detected in Triphala in mock control (MC) and Market formulation (T1 to T6) through *ITS2* metabarcoding sequencing. The relative abundance (% reads) of nontargeted plant species is listed in Table 4.2.

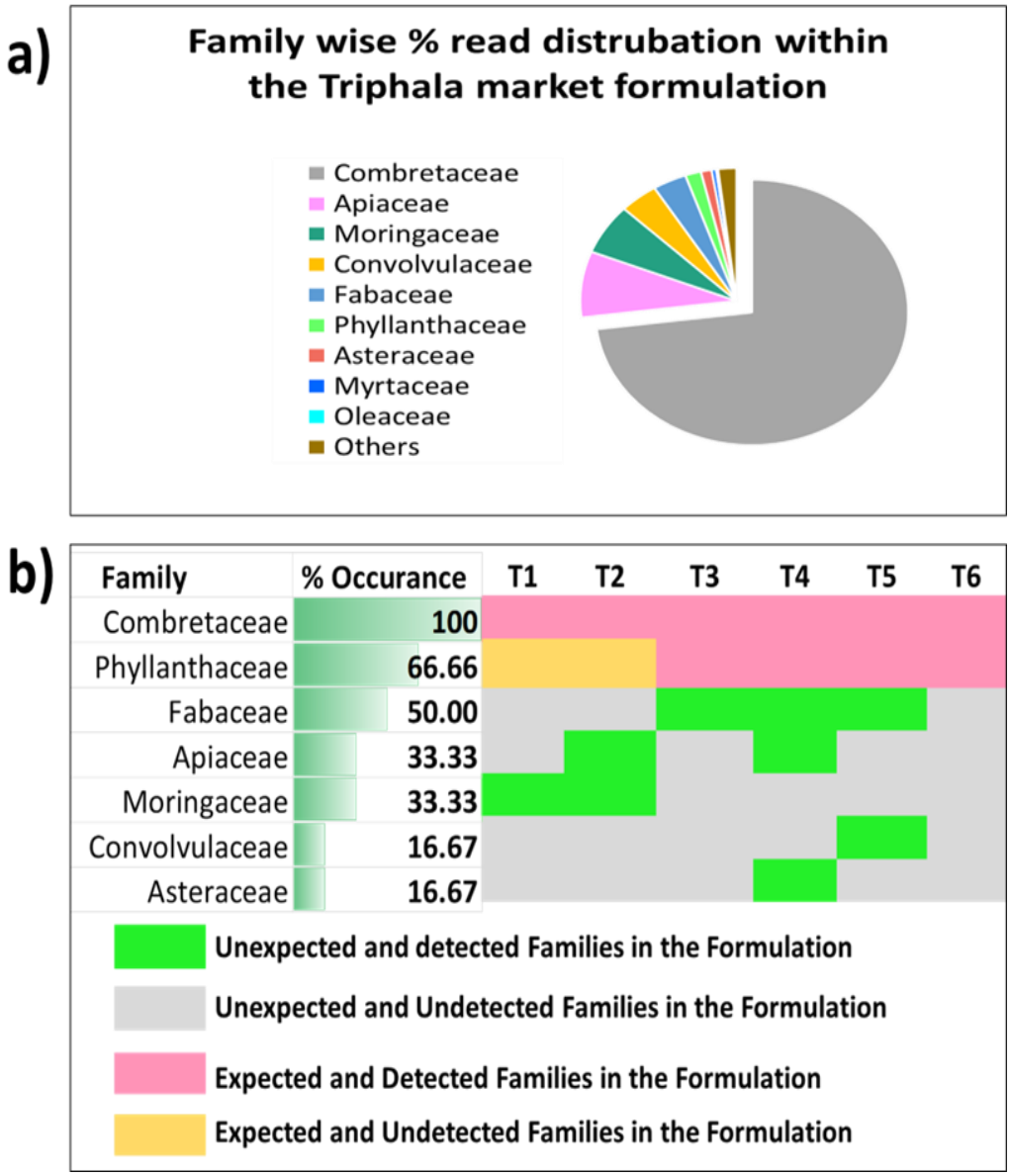

Figure 4.4 Family-wise distribution and occurrence within Triphala market formulations a) Pie chart showing family-wise % of total reads within Triphala market formulations. b) Heat map showing the presence, absence, and % occurrence of expected and unexpected families in Triphala formulations.