New Zealand Food Safety
Haumaru Kai Aotearoa

Scientific interpretive summary and peer review summary of the reassessment of the scientific basis of the regulatory export definition for mānuka honey

New Zealand Food Safety Technical Paper No: 2023/12

Prepared for New Zealand Food Safety
By Operational Research and peer-reviewers from Manaaki Whenua Landcare Research, Plant and Food Research and University of Waikato

ISBN No: 978-1-991080-96-7 (online)
ISSN No: 2624-022X (online)

June 2023

Te Kāwanatanga o Aotearoa
New Zealand Government

Ministry for Primary Industries
Manatū Ahu Matua

Disclaimer

While every effort has been made to ensure the information in this publication is accurate, the Ministry for Primary Industries does not accept any responsibility or liability for error of fact, omission, interpretation or opinion that may be present, nor for the consequences of any decisions based on this information.

This publication is also available on the Ministry for Primary Industries website at http://www.mpi.govt.nz/news-and-resources/publications/

Contents

Page

1 Background

A Ministry for Primary Industries (MPI) mānuka honey definition was developed from 2014 to provide assurances around the authenticity of New Zealand mānuka honey for our overseas markets. Following its introduction in 2018, industry raised several concerns that the definition is not fit for purpose and that some authentic mānuka honey could not meet the definition.

In response, MPI in November 2020 committed to reassessing the definition, and requested industry provide evidence to support a change. MPI scientifically analysed all data submitted, alongside data already held by MPI, and prepared individual reports with respect to the four concerns raised:

- Regionality of markers
- Blending and adulteration risks
- DNA stability in mānuka honey
- Value of mānuka DNA, marker levels and alternative markers

2 Peer review

To provide independent assurance that our reassessment processes were robust, MPI enlisted two panels of independent experts: one panel of five scientists with expertise in apicultural science; the other panel of three experts with a speciality in plant molecular genetics. MPI has also engaged two independent external statisticians to review the statistical analyses where necessary.
The panels of experts reviewed MPI's analyses and findings and delivered independent reports.

3 Data caveats

MPI provided guidance with its request for data from industry on what data and information would be useful to inform the reassessment to ensure the data submitted could be analysed appropriately to address concerns raised. This guidance included criteria such as completeness (for example, the range of marker concentrations tested for the same honey) and traceability (for example, the geographical location the honey was harvested from or the year it was harvested).

The data submitted generally did not contain crucial full traceability information such as age, known link to source plants, and storage conditions. These and other gaps were major limitations of the data submitted to the reassessment. Given this information was lacking, it was not possible to draw scientifically defensible conclusions on the issues raised with the mānuka honey definition.

4 Key findings of the reassessment

Regionality of Markers: Concerns were raised about regional variations in marker levels, particularly the concentration of 2'-methoxyacetophenone (2'-MAP). Regional variation in honey markers does exist, as would be expected with any natural product. Previous studies conducted during the development of the definition and after its implementation concluded that there were not regional differences in 2'-MAP concentration that would discriminate against a particular region's ability to produce monofloral mānuka honey, and analysis of the data provided to the reassessment did not change that conclusion.

Blending and Adulteration Risks: Industry expressed concerns about honey blending practices and the potential for adulteration. While strategic blending occurs to maximize the proportion of honey meeting the definition, extensive computer-based modelling by MPI and an independent statistician demonstrated that the risk of blending non- mānuka honeys to pass the definition is low.

Investigating the risk of adulteration was out-of-scope of the reassessment as these are managed through MPI's broader regulatory regime, including production, registration and traceability requirements, and compliance measures.

DNA Stability in Mānuka Honey: Industry expressed concerns that DNA stability in mānuka honey is compromised by honey processing, heat exposure, and long maturation periods. However, the data submitted lacked traceability information and test repetitions necessary for drawing robust conclusions. Exploratory analyses suggested a potential decline in DNA levels over time but did not establish causative mechanisms. Controlled real-time stability studies are required to inform regulatory change, but they were not feasible within the reassessment timeframe. The evidence suggests that businesses could manage this risk by (a) not blending to the limit of the DNA threshold, or by (b) cold-chain storage.

Value of Mānuka DNA, Marker Levels, and Alternative Markers: The reassessment investigated industry concerns about the usefulness of the mānuka DNA marker, inclusion of alternative markers, and the threshold levels of the five markers currently in the definition.

Comparison of models to those used to develop the current mānuka honey definition confirmed the usefulness of the mānuka DNA marker in classifying honeys, particularly for distinguishing those from New Zealand and overseas.

The proposed alternative markers, dihydroxyacetone (DHA) and methylglyoxal (MG) were deemed unsuitable due to their instability over time. Leptosperin showed potential as a marker for mānuka honey but there was not enough data with leptosperin test results to confidently support including it in the definition.

A key concern raised by industry is that some honeys have markers at levels that do not meet the monofloral mānuka honey definition or the multifloral mānuka honey definition. The results of modelling the data provided weak evidence for a multifloral mānuka honey definition without an upper limit for 3-PLA (one of the markers

responsible for separating mono- and multifloral mānuka honey). However, this weak evidence was insufficient to support a regulatory change.

There was no evidence to support any other change to the mānuka honey definition.

5 Conclusions

1. Honey is a highly complex natural product due to many factors, including natural variables and industry processing practices.
2. MPI and peer-reviewers concluded that while the data provided contains useful information around markers and their levels in honey, it was not collected to specifically address the issues raised and therefore lacks the necessary scientific robustness and traceability information to draw strong conclusions.
3. There is insufficient scientific evidence to justify changing the regulatory export definition for mānuka honey.

Manaaki Whenua
Landcare Research

Independent panel review of the MPI reassessment of the regulatory export definition for mānuka honey

Prepared for: Ministry for Primary Industries

June 2023

Independent panel review of the MPI reassessment of the regulatory export definition for mānuka honey

Contract Report: LC4284

Gary Houliston, Nikki Harcourt

Manaaki Whenua – Landcare Research

David Chagné, John van Klink

Plant & Food Research

Merilyn Manley-Harris

University of Waikato

Approved for release by:

Pete Millard
GM Science
Manaaki Whenua – Landcare Research

Contents

Executive summary

Mānuka (*Leptospermum scoparium*) honey is a premium product that commands a high price on the international market. Due to the concerns raised in overseas markets as to the authenticity of mānuka honey, the Ministry for Primary Industries (MPI) undertook research to develop a set of criteria for the identification of genuine mānuka honey. This review is of a MPI reassessment of the mānuka honey science definition following industry concerns on whether the definition was operating effectively. Four topics were addressed:

1 Regional differences in the amount of 2′-methoxyacetophenone (2′-MAP) were leading to erroneous classification of honey as non-mānuka, or multifloral rather than monofloral.

2 Producers can intentionally blend honey with different chemical characteristics to produce higher value multi or monofloral mānuka honey.

3 The DNA marker does not add value to the classification of mānuka honey, and that other chemical markers, should be substituted in its place.

4 DNA is unstable in honey, especially in 'high-grade' honey, and therefore decreases over time, leading to honey being misclassified as non-mānuka.

The current MPI mānuka honey definition is built on a considerable dataset, made up of two main data collection activities. The first is comprised of samples collected as part of the MPI Science programme, collected expressly to develop a definition. The second is from industry submitted data, which is strongly biased towards mānuka honey, both monofloral and multifloral, with far fewer samples of non-mānuka. This is a considerable weakness in the ability to draw statistical conclusions on the data, with the review panel finding that due to study limitations it is difficult to test the above topics with a high degree of rigour.

The review panel recommend that there is a need for a curated library of honey samples and associated metadata to be collected independently from industry submissions. This would require the type of honey, its geographical source, and concentrations of all the chemical and DNA markers, both currently used and of identified potential future use, to be collected. It should comprise equal numbers of samples of honey from multifloral mānuka, monofloral mānuka, non-mānuka and Leptospermum honey from other countries. Additionally, specific experimental manipulations should be performed to test the effects of temperature, chemistry and other factors on the persistence of DNA in mānuka and non-mānuka honey.

The review panel conclude that with current data, there is limited value in modifying the current MPI mānuka honey definition. The only exception to this is the cap on the 3-PLA level in multifloral mānuka honey could be reconsidered based on the available data.

1 Introduction

Mānuka (*Leptospermum scoparium*) honey is a premium product that commands a high price on the international market. Due to the value that this product holds, concerns about authenticity and even counterfeit products have the potential to damage this key export. The Ministry for Primary Industries (MPI) undertook research to develop a set of criteria for the identification of genuine mānuka honey originating from New Zealand. This led to the development and implementation of the mānuka honey definition in the General Export Requirements for Bee Products in 2018, based on the criteria in Table 1.

Table 1: MPI mānuka honey definition

Attribute	Monofloral mānuka	Multifloral mānuka
2-methoxyacetophenone (2'-MAP)	≥ 5 mg/kg	≥ 1 mg/kg
2-methoxybenzoic acid (2-MBA)	≥ 1 mg/kg	≥ 1 mg/kg
4-hydroxyphenyllactic acid (4-HPA)	≥ 1 mg/kg	≥ 1 mg/kg
3-phenyllactic acid (3-PLA)	≥ 400 mg/kg	≥ 20 mg/kg and < 400 mg/kg
DNA from mānuka pollen	< Cq 36 (~3 fg/µL)	< Cq 36 (~3 fg/µL)

This definition defined two types, or grades, of mānuka honey, monofloral mānuka honey, with the requirement of ≥ 5 mg/kg of 2-methoxyacetophenone (2'-MAP), and multifloral honey, with ≥ 1 mg/kg of 2'-MAP, as well as meeting the other chemical and DNA marker requirements of the definition (Table 1). While this definition was intended to provide greater certainty to producers and to protect the integrity of the product in the market, there has been considerable discussion about whether this represents a robust and appropriate method for determining the authenticity of mānuka honey.

Industry has raised several concerns about the definition, but the most common can be summarised as:

1 Regional differences in the amount of 2'-methoxyacetophenone (2-MAP) produced in the nectar of mānuka has the effect that some areas of the country are disadvantaged in that monofloral mānuka honey sourced from these regions will be erroneously classified as multifloral or non-mānuka.

2 Producers can intentionally blend honey with different chemical characteristics to meet the current definition, or that honey can be adulterated with synthetic chemicals to allow it to pass as mānuka.

3 The DNA marker does not add value to the classification of mānuka honey, and that other chemical markers, such as leptosperin, dihydroxyacetone (DHA), and/or methylglyoxal (MG) should be substituted in its place.

4 DNA is unstable in honey, especially in 'high-grade' honey, and therefore decreases over time, leading to honey being misclassified as non-mānuka.

MPI undertook a reassessment of the definition and put out a "request for data" to the industry for evidence that may be used in this reassessment process. This data was used alongside other information that MPI has collected, from the MPI funded trial to establish a honey reference collection. Further analyses were undertaken in an attempt to address the industry concerns. This report is a scientific review of the resulting reports generated by MPI during this reassessment process. To produce this review, two panels of scientists were convened, one with expertise in apiculture science, and one with expertise on plant molecular genetics. The membership of these was:

1.1.1 Panel 1

To review MPI definition based on the key industry concerns raised above and described in detail in the four sections that follow.

- Dr David Chagné, Plant & Food Research
- Dr Nikki Harcourt, Manaaki Whenua – Landcare Research
- Dr Gary Houliston, Manaaki Whenua – Landcare Research
- Dr John van Klink, Plant & Food Research
- Professor Merilyn Manley-Harris, University of Waikato
- Professor Christopher Triggs, formerly Professor of Statistics, The University of Auckland (statistical advice to the panel)

1.1.2 Panel 2

The second panel of molecular biologists was formed to review the DNA marker section of the review. This panel was made up of:

- Mr Kim Richardson – Senior Scientist, AgResearch
- Dr Vaughan Symonds – Senior Lecturer, Massey University
- Professor Richard MacKnight – Otago University

The following report addresses the points listed above based on the data and analysis provided to the panel by MPI. Conclusions on each point are noted at the end of each section, and in the concluding summary. Recommendations for changes and further work are included at the end of the report.

2 Part 1: Investigating the Regionality of 2-methoxyacetophenone (2′-MAP) in mānuka honey

2.1 Background

The Ministry for Primary Industries (MPI) have collated data to determine whether 2′-methoxyacetophenone (2′-MAP) varies in mānuka honey across New Zealand geographic regions. Data from their own study plus submissions by industry have been summarized in the following reports: Regionality report (2018), NPJ Science of Food publication (McDonald et al. 2018), MPI report for assessment panel (November 2021), MPI report for assessment panel revision (May 2022).

The Independent review panel assessed these documents and have commented on the key issues using the MPI-initiated questions in bold below to direct the evaluation and discussion.

1 Are the conclusions reached in this review robust and justifiable?

- **a How strongly do you think the evidence and data presented suggest 2′-MAP varies across regions and is discriminating against certain regions producing monofloral mānuka honey?**

2 Given the data available, are there any other avenues you think we could explore to explain the variation in 2′-MAP between regions?

3 What are the strengths of our approach, what are the weaknesses? How could we ameliorate the weaknesses?

4 From the data available, are there are steps you think we could take to address industry concerns regarding variability of 2′-MAP?

1 Are the conclusions reached in this review robust and justifiable?

A recent analysis of honey samples by MPI (MPI November 2021) showed that levels of the 2′-MAP below the definition threshold was the most common reason for a particular sample not being classified as monofloral mānuka honey. Industry is aware that this is the case and have claimed that 2′-MAP concentrations naturally vary geographically, and the definition should be revised to reflect this, to avoid honey being wrongly classified. To test this idea, MPI compared levels of 2′-MAP in honey samples from different regions to see whether there were statistically significant differences that could be consistently found.

While it was found that there was variability in the concentration of 2′-MAP, the data used was insufficient to allow definitive conclusions to be drawn. The reason for this is due to the lack of associated information about the origin of many of the samples tested, the lack of proper sampling and replication in the samples across both geographic/ecological regions, and information about the time of collection.

The review panel agree with the conclusions of the MPI report in that due to the data limitations (e.g. wide confidence intervals) and low number of overall samples used for the models, it was not possible to accurately predict the 2′-MAP concentrations in mānuka honey from different geographic regions within NZ. For a robust model, further intensive sampling is required to determine the true extent of 2′-MAP variation in NZ mānuka honey. Furthermore, this requires an appropriate statistical design with sufficient representative sampling across regions and seasons. This would require collection of corresponding sample metadata (including handling/processing information), to determine the potential influence of environmental and genetic factors.

a How strongly do you think the evidence and data presented suggest 2′-MAP varies across regions and is discriminating against certain regions producing monofloral mānuka honey?

MPI have acknowledged that there are limitations to the dataset, and this includes the restricted number of honey samples from some regions, and over-representation of honeys from single sites within a region. The use of geographic / council boundaries rather than ecological districts or some other spatial delimitation that better reflects variation in plant communities is a weakness of the approach. That much of the data has very limited information about the specific location of collection also makes it difficult to be confident of the conclusions. The lack of sufficient metadata (accurate location, time of collection) along with the unbalanced sampling (i.e. high-density sampling at some locations and no data for others), means that there is not the statistical power required to draw strong conclusions. For example, within a geographic region there will be variation in microclimatic conditions, intraspecific genetic diversity of mānuka plants, variability within seasons due to the physiology of the plant, differences in bee foraging and other factors. A scientific experiment to test something such as variation in nectar metabolites needs to be well-designed to take these variables into consideration if strong conclusions are to be drawn.

We agree with the MPI conclusion that the current data is insufficient to conclude whether regionality is unduly discriminating against producers from certain regions. Neither does the data indicate whether there is robust evidence for changing the threshold for 2′-MAP concentrations in either monofloral or multifloral mānuka honey, which will be further discussed in Part 2 below.

2 Given the data available, are there any other avenues you think we could explore to explain the variation in 2′-MAP between regions?

To draw any robust conclusions about regionality, and given the limitations of the existing data set, full metadata for each of the samples is required. Any metabolites derived from the plant could be influenced by many factors: the genetics of the plant, the growing environment including both climate and soil properties, seasonal variation, as well as other variables. These variables can all vary over space (geography) and time, so to determine which best explain this variation requires careful design of the data collection. Given that most of the samples had either limited or no information on these factors, and the data was analysed using regional council boundaries, it is unsurprising that the results were inconclusive. To improve this analysis it would be necessary to collect detailed and accurate metadata for the samples, and compare this to site information, including both

physical and environmental variables, and landscape-scale variation in plant genetics as described in a recently published paper (Koot et al. 2022). It is unclear if it would be possible to get access to this metadata for the existing data as it may or may not be held by the submitters, and there may be commercial sensitivity around sharing it.

It should be noted that from a chemistry perspective, 2′-methoxyacetophenone (2′-MAP) is the most volatile (i.e. it has the lowest boiling point) of the four markers used in the honey definition. The boiling point of 2′-MAP is 122–123 °C; boiling points for 2-MBA, 4-HPA and 3-PLA are 280, 414, and 331 °C respectively.

This means that at any given ambient temperature, the vapour pressure for 2′-MAP is going to be significantly higher than for the other three markers and therefore there will be more loss of 2′-MAP from volatilization. During processing, honey may be subject to high temperatures, albeit briefly, which could mean that there may be effects of handling on the final concentration. It is clear from the results that the concentration of 2′-MAP doesn't correlate with geographic region, i.e. if temperature has an important impact, then a concentration gradient from North to South may be expected to correlate with regional temperatures from North to South. It is possible that the temperature effect may be more localized, either via local temperature variation or in differences in processors in different parts of the country. There is insufficient metadata attached to the samples to test these ideas in the current dataset.

A further issue to be addressed is that metabolites with a ketone functionality (including 2′-MAP and DHA) undergo Maillard-type reaction sequences with amines, which has the effect of potentially lowering their concentration within a sample over time. These side reactions are attributable to the high content of undefined catalytic/reactive material in mānuka honey (which may vary with genetics and/or environment), which are not usually found in other types of honey. How this varies over the landscape is not well understood, but could well be a significant contributor to 2′-MAP variation.

3 What are the strengths of our approach, what are the weaknesses? How could we ameliorate the weaknesses?

Strengths

The large number of samples analysed showed that there is broad variability in 2′-MAP concentrations. The solid approach for the analytical and statistical analyses undertaken on the honey samples (supplied by producers) means that the data can be considered reliable. The large number of samples analysed showed there were some differences and potential patterns of chemical variation.

Weaknesses

The main weakness is inherent in the sampling design, with reliance on producers to supply samples, some with and some without the appropriate provenance provided. The lack of control over the full design and implementation of sample collection has limited the utility of the outcomes, since there was a lack of standardisation and missing meta data. The lack of suitable replication for some geographic regions (that themselves need to be better defined) and seasons also means the resolving powers are limited. That most

samples submitted were for mānuka honey rather than representing both mānuka and non-mānuka honey also makes it more difficult to run robust statistical tests. Any investigation such as this requires samples representing what is not mānuka as much as what is.

Potential methods for amelioration of the weaknesses

If it were possible to retrieve accurate and detailed metadata for the samples supplied to MPI for this work, it may be possible that further analysis could provide a more robust conclusion on 2′-MAP variation in mānuka honey. The review panel has no information on whether this would be possible or not.

4 From the data available, are there are steps you think we could take to address industry concerns regarding variability of 2′-MAP?

Limitations in the current dataset mean that further analysis is unlikely to result in robust enough results to warrant a change to the definition for 2′-MAP. The only possible way to address the question of whether the current 2′-MAP level in the definition is acceptable is via further data collection, using information designed to specifically test whether the current level is suitable. It is possible that this would not necessarily need to be carried out at a national scale, but a carefully designed smaller study may be sufficient to provide the information required on what drives 2′-MAP concentrations in mānuka honey.

2.2 Conclusion

The review panel conclude that on the basis of the MPI report, '*Regionality of 2′-methoxyacetophenone (2′-MAP) in mānuka honey*', there is no justification for revising the levels of 2′-MAP in the current MPI mānuka honey definition. The reason for this conclusion is that there is insufficient data of the correct type to robustly test if there is regional variation effecting the levels of 2′-MAP. To implement a change to the definition reflecting that different regions should be treated differently in 2′-MAP requirements there would need to be strong evidence that this was the case. This would require a carefully designed study with good replication (sites, seasons, soil type, plant origin) along with detailed information on the role of climate, potentially comparing different ecological regions or some other appropriate proxy. It is unsurprising that a comparison across regional authority boundaries, with little correlation to environment, and using data that is unbalanced and not systematically collected, that this study is inconclusive.

3 Part 2 Investigating the risk of honey blending on the MPI mānuka honey definition

3.1 Background

The Ministry for Primary Industries (MPI) have collated data to determine whether blending of honey with different chemical marker levels is a risk to industry, specifically that the integrity of the classification system could be undermined by production of monofloral mānuka honey from non-monofloral samples.

The Independent review panel assessed the evidence provided and commented on the key issues using the MPI-initiated questions in bold below to direct the evaluation and discussion.

1 Are the conclusions reached in this review robust and justifiable?

- **a Does the evidence, data, and analysis presented suggest the probability of successful blending non-mānuka honeys to create monofloral honey is minimal?**

2 What are the strengths of our BioSS approach, what are the weaknesses? How could we ameliorate the weaknesses?

3 From the data available, are there are other steps you think we could take to address industry concerns regarding the risk posed by blending?

1 Are the conclusions reached in this review robust and justifiable?

The panel agree with the findings of the MPI report on blending, especially the overarching conclusion that the issue of industry blending to meet the current MPI definition is unlikely to be practical at scale and will not have any significant effect on the volumes of mānuka honey produced in New Zealand. The blending analysis presented, indicated with good justification, that the potential to circumvent the intent of the definition by blending honey was not likely to become common practise based on the practicality and economics of this activity. There is the possibility in some cases to blend monofloral and multifloral mānuka honey to create a larger volume of product that meets the monofloral definition.

The blending analysis showed that the blending of multifloral honey was highly unlikely to produce anything other than multifloral honey, and that if anything there was a strong dilution effect as expected when non-mānuka and mānuka honey was blended. The report clearly showed that there was a decreasing probability of being classified as multifloral or monofloral when higher proportions of non-mānuka honey is blended, as would be expected.

a Does the evidence, data, and analysis presented suggest the probability of successful blending non-mānuka honeys to create monofloral honey is minimal?

The probability of producing monofloral mānuka honey by blending solely non-mānuka honeys is null at the 5 mg/kg level for 2′-MAP and vanishingly small at the 1 mg/kg level of 2′-MAP when DNA was included in the consideration. We believe that the addition of the DNA marker further adds to the robustness of the approach when blending is considered. We also note that while the DNA test appears to add robustness to the definition, we should be aware that pollen presence in honey does not always directly correlate to the nectar that is collected, due to differences in bee foraging behaviour. While the assumption of volumetric equivalence is a reasonable assumption for nectar chemistry (i.e. a blended sample will have a proportional chemistry to those honeys it is created from), we have no empirical evidence that this will apply to Cq values derived from the DNA test. This should be experimentally investigated.

2 What are the strengths of our BioSS approach, what are the weaknesses? How could we ameliorate the weaknesses?

The analysis is predicated on the assumption that blending honey samples will result in a uniform, volumetric representative of the samples in question e.g. the resulting honey will perfectly reflect the levels of each chemical marker present in the original samples by volume. There are several reasons as to why this may not necessarily be the case:

Test results for any batch of honey may vary depending on how these were sampled from the bulk/different parts of the drum may have slightly different chemical properties such that test results on a subsample may not be perfectly representative.

While this is an assumption with caveats attached, it is not an unreasonable one, and it is unlikely that anything other than following this assumption could be justified in examining this question. While the reviewers thought that it should be abundantly clear that this assumption has been made, it does not cause undue concern about the conclusions.

A more notable concern or weakness is the lack of understanding of the effect of blending on Cq values resulting from honey DNA testing. As DNA test results are likely to be highly influenced by the amount of pollen remaining in the sample, any processing including blending can potentially alter the results of this test. As almost all commercial honey is filtered, there is a high probability that this plays an important role in determining how much pollen remains in the honey. High grade medical honey is filtered to such a standard that pollen, and therefore the majority of the mānuka DNA, is removed from the sample. The effect of processing including blending on the honey DNA test should be investigated empirically.

3 From the data available, are there are other steps you think we could take to address industry concerns regarding the risk posed by blending?

The review panel recognise that blending is a customary practice in the industry and provided that no deliberate adulteration occurs, is not outside of the mānuka honey definition. Any definition that makes public the levels of marker compounds and when accurate testing is readily available, then producers will attempt to blend honeys to arrive at appropriate levels of the marker compounds to optimise profit, irrespective of what the specific requirements may be.

A possible weakness of the approach is that although MPI has an extensive database of honey from around the country, it is possible that honey with naturally high 3-PLA and low 2'-MAP concentrations may exist outside of the range of those currently tested, particularly if these have been developed or will be developed as cultivars for plantation mānuka. If such honey did exist, this could alter the results of the blending simulation, but as this is currently speculation it is not reason to revisit the definition as they currently stand.

Finally, there is of course the potential for adulteration of honey by the addition of the chemicals required in the definition. This is already outside of the definition and again is not an argument for altering the requirements of the definition.

3.2 Conclusion

The panel concludes that blending is not a substantial concern in the application of the definition, and any definition that has publicly available thresholds and the availability of independent testing will result in industry blending. It is unlikely that monofloral mānuka honey can be produced from multifloral or non-mānuka honey in any significant volume due to the design of the panel of markers used. Empirical experiments that include blending of samples and remeasuring the levels of chemical and particularly the DNA marker would be preferable to further theoretical blending analyses, but we do not expect that this would result in substantially different conclusions based on the work carried out to date.

4 Part 3 Investigation of the role of the mānuka DNA marker and alternative markers proposed by industry in the mānuka honey definition

4.1 Background

MPI aims to evaluate the role of the mānuka DNA marker and alternative markers proposed by industry in the mānuka honey definition. To achieve this aim MPI:

1 **Reviewed the submissions made by industry.**

2 **Used data submitted by the industry and data held by MPI to investigate the suggested changes to the definition in the submissions.**

3 **Built, analysed, and interpreted a range of classification models to determine markers and appropriate threshold levels that can separate mānuka honey from other honey types.**

4 **Compared the performance of potential alternative markers to the mānuka DNA marker in the definition.**

The Independent review panel assessed the MPI report (November 2022) and have comments on the key issues using the MPI-initiated questions in bold below to direct the evaluation and discussion.

1 **Are the conclusions reached in this review robust and justifiable?**

   a **Does the evidence, data, and analysis presented suggest the mānuka DNA marker should remain in the definition?**

   b **Does the evidence support the addition of alternative markers to the definition at this stage?**

   c **How does the evidence support changing the current thresholds of the markers in the definition?**

   d **How strongly do you think the MPI's approach of building the classification model to authenticate monofloral and multifloral mānuka honeys is appropriate to address the issues explored?**

2 **Given the data available, are there any other avenues you think we could explore to:**

   a **Assess alternative markers?**

   b **Assess the value of mānuka DNA in the definition?**

3 **What are the strengths of our approach, what are the weaknesses? How could we ameliorate the weaknesses?**

4 **From the data available, are there are other steps you think we could take to address industry concerns regarding the definition?**

5 **Would a period of consolidated data collection be appropriate, in order to provide an adequate database for assessing the concerns of the mānuka honey industry regarding the definition?**

1 Are the conclusions reached in this review robust and justifiable?

In general, and based on the data available, the review panel largely agree that the conclusions of this work by MPI are justifiable. However, in this part of the reassessment the limitations of the data and the approach are not ideal, but the best that can be achieved with the information to hand. Only data with traceability and supplier labels (MPI Science Programme data and samples collected for the National Honey Reference Collection more recently) were used to build the classification models, whereas the other data available was used to test the models.

An assumption of the modelling approach is that the accuracy of the definition is assessed against whether it recovers the same classification as the current method (i.e. it assumes the current definition is the correct one, and any other approach is assessed against whether it will produce the same result). While this is a clear weakness, it is difficult to design an alternative approach, as any approach will also need to be based on assumptions due to the complex nature of honey. Ideally, a comprehensive, independent sampling programme, including accurate geolocation and environmental data for the honey samples representing a balanced design of monofloral, multifloral and non-mānuka honey across the regions of production would be used to design a classification system. This would require considerable time, effort and investment.

a Does the evidence, data, and analysis presented suggest the mānuka DNA marker should remain in the definition?

Yes, particularly around the value that it adds in determining the origin of the honey samples being from New Zealand, and of *L. scoparium* origin. Given that New Zealand *L. scoparium* is being cultivated offshore for honey production, there is a limitation that even if it can detect genetic origin, this will not indicate geographic origin.

b Does the evidence support the addition of alternative markers to the definition at this stage?

No, primarily due to the lack of data available to test whether the alternative markers are suitable for use in the definition. Where most of the samples in the MPI data set and submitted by industry have information for the complete panel of markers in the current definition, the number that also include information for the alternative chemical markers are far fewer. While it cannot be concluded that other markers should be included, it is also not possible to rule them out as possible markers due to the lack of proper sampling and design in the sample collection.

c How does the evidence support changing the current thresholds of the markers in the definition?

Changing thresholds of the markers in the classification rules generated by the CART algorithm, based on the available training data, optimizes the choice of discrimination levels of each marker. The effect of changing these levels to give simpler classification rules can be directly assessed. If the change in the rate of misclassification is small such a change may be desirable, e.g. removing the upper limit of 3-PLA in the MPI multifloral mānuka honey definition.

d How strongly do you think the MPI's approach of building the classification model to authenticate monofloral and multifloral mānuka honeys is appropriate to address the issues explored?

The approach to the modelling (CART modelling) is appropriate and likely the best approach given the question. Limitations are with the data available to include in the model both as learning and testing data, and the lack of a systematic approach to data collection specifically to answer the question at hand. Development and testing of a complex and variable biological system such as honey requires careful design of the questions (hypotheses) and the data collection (statistical design and balanced sampling) to ensure that robust conclusions can be drawn. The current data available was not specifically collected to answer these questions and therefore has severe limitations for drawing statistical inferences.

2 Given the data available, are there any other avenues you think we could explore to:

a Assess alternative markers?

Due to the limitations of the data, it is unlikely more could be done to explore other markers at this point. For leptosperin, where this is some data available, it looks like further investigation and data collection is warranted, but at present there is insufficient evidence to support adding this to the definition. Other potential markers, particularly dihydroxyacetone (DHA) and methylglyoxal (MG) are known to vary over time in the honey matrix so are unlikely to be good candidates even if more data was available.

b Assess the value of mānuka DNA in the definition?

This is addressed more fully in the following section, but again, the current data has likely been explored to the extent possible and any further investigation would require new information specifically targeted at this question.

3 What are the strengths of our approach, what are the weaknesses? How could we ameliorate the weaknesses?

The strength of this approach is the statistical modelling that allows inference to be drawn despite the data limitations. The volume of data available for model construction is good, albeit with the caveats noted below. The goal was to distinguish four types of honey; sourced from monofloral mānuka, multifloral mānuka, non-mānuka, in New Zealand and non-mānuka honey sourced from outside New Zealand. The statistical algorithm required validated samples of honey from each of the four sources as its building blocks. The statistical literature contains many algorithms for the construction of classification rules to differentiate groups of samples from previously determined sources. The CART technology used by MPI has been shown to perform well. It has the additional advantage that the rules it produces have simple forms. Each rule differentiates between groups using the value of a single chemical or marker. This has the advantage that the resulting set of rules is easy to understand, test, and critique.

The main weakness as addressed above is that the sampling of honey has not been carried out in a systematic fashion, but rather around what industry has submitted for testing or for the reference library. There are two particular shortcomings in the available data. One is that for a lot of the samples there is limited metadata describing accurately the origin, environment, processing and even time of collection. The second is that the data is not balanced in that it is primarily a collection of mānuka honey whereas it is just as important to have a good representation of what is not mānuka honey to allow a robust definition to be constructed. There are also geographic biases, with some areas better represented than others. There appears to be some reluctance within the industry to share information as to the exact source of different honey samples due to concerns around commercial sensitivity. In addition, the nature of honey as a product and how nectar is collected is influenced by many factors, making it difficult to determine with high certainty what the source of the crop is even when certain species are flowering. It is well understood that mānuka is not a preferred forage species for bees, so even where mānuka is flowering there can be substantial dilution if other species are also producing nectar.

At present, the assessment of the accuracy of the classification rules depends on whether they recover the same groups as the current definition or the declared nature of the samples by submitters. This is the major underlying weakness of the current datasets, which would be mitigated by a comprehensive, independent sampling programme (see recommendations below). The panel recognises that such a programme would be costly and time consuming, but the lack of a curated library will always leave the classification rules open to criticism.

One possible difficulty is that because of commercial sensitivity, some beekeepers may be reluctant to divulge the exact location of their hives. There may also be a difficulty in that even if the source of a sample is accurately geolocated, the species composition of the pollen may not be known with the same degree of accuracy. For many samples it may only be possible to get a general location and types of other flowering species. This means that any definition will have a degree of uncertainty, resulting in a persistent, albeit low, level of honey misclassification.

4 From the data available, are there are other steps you think we could take to address industry concerns regarding the definition?

The data available has likely been used to the extent that is possible. It is unlikely that further analysis of the existing data is warranted at this point. To address the concerns of industry it is likely that further data will need to be collected (see the following point).

5 Would a period of consolidated data collection be appropriate, in order to provide an adequate database for assessing the concerns of the mānuka honey industry regarding the definition?

There is a need for a curated library of honey samples and associated metadata to be collected within a single repository. This would provide a reliable and trusted set of data to train future classification rules and to test them. A programme of consolidated data collection should be implemented, in order to provide an adequate database for assessing the concerns of the mānuka honey industry regarding the definition, and to allow conclusions to be drawn without reservation. At the very least the type of honey, its

geographical source, and concentrations of all the chemical and DNA markers, both currently used and of identified potential future use, should be collected. It is important also that approximately equal numbers of samples of honey from all four sources currently being considered be collected in each season. Collection should be an on-going process since changing climate and environmental modification will ultimately affect honey composition. The design of the collection process should be developed by a multi-disciplinary team of chemists, plant ecologists, and statisticians, in conjunction with accredited processors and beekeepers with extensive experience with mānuka honey production. As such a collection might take a protracted period of time, concerns about the stability of the DNA marker in stored honey should be tested in a separate laboratory storage study. The samples need to be obtained by independent contractors with no vested interest in the outcome of the construction of the classification rules.

4.2 Conclusion

On balance of evidence, the review panel believe that the DNA marker should be retained in the MPI mānuka honey definition, and there is insufficient evidence to recommend that it is removed. While there is some evidence that DNA does decline in the honey matrix over time, albeit at a very low rate, this is to be expected and does not appear to occur at such a rate as to cause undue concern. Further experimentation to quantify the causation and rate of decline would be relatively simple, but this has not been carried out by either MPI or industry to date. This would be necessary in the panel's opinion to provide sufficient evidence to justify a change in the definition.

There is currently insufficient quality information to assess whether alternative markers should be included in the definition. Due to known instability over time, the panel believe that DHA and MG should not be considered. There is some indication that leptosperin may be useful, but more data would be needed to properly assess this.

The panel agree the removing the cap for 3-PLA concentration in the multifloral mānuka honey definition is worth considering based on the available information. From the results of the modelling this should reduce the number of false negative results for multifloral mānuka samples, without otherwise being detrimental to the application of the definition. Any further modification to the definition would ideally be addressed following further data collection designed to specifically address the specific facet of the definition being examined.

5 Part 4 Reassessment report on DNA stability in the honey matrix

5.1 Background

Industry have expressed concerns around the inclusion of the DNA marker in the MPI definition, resulting in a total of 20 submissions being made by industry. Nine of these submissions state that the amount of mānuka DNA declines in the honey matrix over time. Many of the submissions state that the idiosyncrasies of honey processing, such as exposure to heat and long maturation periods, exacerbate the rate of DNA decline and lead to genuine monofloral mānuka honey failing the definition on the basis of low mānuka DNA levels. Further, as honey has a long shelf life, it is proposed that DNA decline may lead to authentic product failing the mānuka honey regulatory definition in market. The review panel have considered these revisions and addressed the questions in bold below in response.

1 **Are the conclusions reached in this review robust and justifiable?**

2 **How strongly do you think the evidence and data presented suggest DNA is declining in the honey matrix?**

3 **How strongly do you think the evidence and data presented suggest the ability of the ManKan™ assay to detect DNA changes over time?**

4 **Do you have any suggestions on what may be the observed DNA results? Especially given mānuka and kānuka DNA is encapsulated within the pollen grains.**

   a **Does heat seem plausible?**

   b **Does chemical degradation via MG (or another agent) seem plausible?**

   c **Could chemical adducts form between DNA and other chemicals in the honey matrix?**

   d **Do you think it likely the effectiveness of the qPCR assay used to detect DNA in honey may perform differently on aged, rather than fresh, honey?**

   e **Does some combination of the above seem plausible?**

   f **Any other ideas on potential causative factors?**

5 **Given the data available, are there any other avenues you think we could explore to explain the apparent reduction of DNA in the honey matrix?**

6 **What are the strengths of our approach, what are the weaknesses? How could we ameliorate the weaknesses?**

7 **Are there any steps that could be taken to prevent DNA degradation in the honey matrix?**

1 Are the conclusions reached in this review robust and justifiable?

The review panel agrees that the conclusions reached by MPI in this review are robust and justifiable, and that it cannot be robustly demonstrated from the existing data that DNA significantly declines in the honey matrix for any of the posited reasons provided. This is largely due to the lack of key metadata such as the age of the honey sampled, traceability to plants' provenance, and storage conditions in the industry datasets, leading to limited scope to test this idea.

a How strongly do you think the evidence and data presented suggest DNA is declining in the honey matrix?

The evidence provided by the industry and by the re-analysis of data by MPI (section 10 of the report) is not strong. While some of the data indicates that DNA is potentially declining in the honey matrix, this is indicative only as the data is not collected in such a way that this idea can be robustly tested. MPI's assessment of the industry submissions is appropriate, and they are correct to note the limitations of the information provided. A major limitation of most of the industry data provided is that it is from a single time point, whereas to determine a change in DNA concentration over time it is necessary to have data from at least two time points for a given sample.

MPI's analysis of the three datasets available (*Hills, Analytica and MPI CART)* tests four hypotheses:

- Hypothesis 1: High amounts of MG are associated with DNA decline
- Hypothesis 2: Older honeys are associated with lower amounts of DNA
- Hypothesis 3: Mānuka DNA declines over time in the honey matrix
- Hypothesis 4: Variability in ability of ManKan™ assay to detect DNA.

As none of the above datasets have repeated test results from the same samples at different time points, proxies are used, such as the reclassification rate of results when the DNA marker is excluded from the classification or using the DHA:MG ratio as a proxy for age. Neither of these approaches are acceptable replacements for repeated measures from single samples over time in determining the magnitude or cause of DNA decline in honey over time. Even if it is accepted that these are suitable proxies, with which the review panel disagree, there was limited evidence to there being a substantial effect on honey classification due to changes in the DNA marker results from the data provided.

Specifically, Hypothesis 1 about the effect of MG on DNA decline was tested using the reclassification test and the results indicated a less than 0.5% reclassification when including or not including the DNA test. Those samples that were reclassified from two of the three data sets were associated with higher MG levels, which is indicative, but cannot be interpreted as a definitive test of the role of MG on DNA concentration.

There were industry submissions that did describe retesting of different batches of honey at different time points, and there was a pattern of DNA concentrations decreasing over time. In all cases, these were not experiments designed to test specific hypotheses, but rather observations, including those from very small numbers of batches (three in one case). The explanation for the decline for these data is speculative, and it is not possible

from these results to attribute changes in Cq results to any particular factor. It is clear that DNA concentrations in any matrix will decline over time, even under optimal storage conditions, so the conclusion that Cq values decrease is unsurprising, however the actual cause of this can be due to many factors. Determining which factors are important in any given situation requires careful experimentation.

Hypotheses 2 and 3 are addressed in the section below specifically addressing the role of time on DNA concentrations in honey. Hypothesis 4 about the variability of the ManKan assay to detect mānuka DNA in honey and what variables could influence this was addressed by a conversation between MPI and the developer of the assay. This remains unresolved but is an area of possible further experimentation.

2 How strongly do you think the evidence and data presented suggest the ability of the ManKan™ assay to detect DNA changes over time?

Hypotheses 2 addresses whether the ratio of DHA:MG as a proxy for honey age influences the qPCR results. The qPCR results were separated according to the ability of the assay to detect mānuka, kānuka and plant DNA (internal control). As for hypothesis 1, the results were barely suggestive of an effect of DHA:MG on the qPCR performance and a more comprehensive study designed specifically to address this hypothesis would be more appropriate. If the actual hypothesis of interest is the effect of time, it would be best to design an experiment using this rather than DHA:MG ratio as a proxy.

Hypothesis 3 that addresses DNA decline over time was tested using a small dataset held by MPI from honeys that were kept at 4ºC for 2 years. It is important to note that such storage conditions are not representative of industry practices. The conclusion that mānuka and kānuka DNA may decline based on the quantification cycle (Cq) values of the qPCR assay, albeit with a small effect size, seem justified. This is not a surprising finding, nor does it indicate what the mechanism of degradation of the DNA in the honey may be.

3 Do you have any suggestions on what may be the observed DNA results? Especially given mānuka and kānuka DNA is encapsulated within the pollen grains.

a Does heat seem plausible?

Heat is a plausible cause for DNA decline, however the data presented here does not test this hypothesis. Designing an experiment to test the effect of heat would be a relatively simple activity, and likely useful given the use of heat in honey processing.

b Does chemical degradation via MG (or another agent) seem plausible?

Chemical degradation is a plausible cause for DNA decline, however the data presented here does not test this hypothesis.

c Could chemical adducts form between DNA and other chemicals in the honey matrix?

Yes, the formation of advanced glycation end (AGE) products in the presence of MG has been well-documented in the scientific literature. Additionally, DHA is quite reactive and could either directly, or indirectly, affect honey composition, enzyme activity and the

ManKan assay. There is also the possibility that other chemical compounds present in the honey matrix that are not tested for or quantified could play a role in reactions that affect the persistence of DNA in the honey.

d Do you think it likely the effectiveness of the qPCR assay used to detect DNA in honey may perform differently on aged, rather than fresh, honey?

It is possible that the qPCR performs more poorly with aged honey, however the data presented here are not conclusive.

e Does some combination of the above seem plausible?

Combination of the above variables is likely to be influencing the ability to detect DNA using qPCR.

f Any other ideas on potential causative factors?

There are a large number of possible factors that may influence the performance of the DNA assay. DNA variants present in the qPCR primer target sites may influence qPCR performance, PCR reaction inhibitors could be present even after steps in the protocol to remove them, or sedimentation of pollen in honey samples may make getting representative samples for testing more difficult. Specific experiments could be conducted to address some of these issues (see 5 below).

4 Given the data available, are there any other avenues you think we could explore to explain the apparent reduction of DNA in the honey matrix?

There appears to have been limited investigation into pollen content in the honey samples. If there is no pollen or if the pollen is not evenly distributed in honey samples / drums (e.g. sedimentation) then no DNA can be extracted. Experimentation introducing pollen into a honey matrix may be a solution to test the hypothesis that DNA is hard to extract from honey or that there are some inhibitors to the qPCR.

5 What are the strengths of our approach, what are the weaknesses? How could we ameliorate the weaknesses?

Strengths

The DNA test is a good indicator that the honey has some *Leptospermum scoparium* source material, and there is no chance that a positive result can be generated other than by *L. scoparium* DNA being present in the test material.

Weaknesses

There is a lack of experimental results to assess how DNA concentrations change over time in honey under different conditions. This is a relatively simple set of experiments that the review panel suggest should be undertaken.

6 Are there any steps that could be taken to prevent DNA degradation in the honey matrix?

Steps could be taken to avoid degradation; however, they may not be practical for the honey industry. Low temperature storage is the most obvious option, but this may result in other undesirable effects on the product, and make storing, processing, and packaging difficult.

5.2 Conclusion

There is currently limited evidence that DNA degrades significantly in the honey matrix over time. Most of the data available does not meet the requirements to test this, often using proxies for time such as DHA:MG ratio, or data are from storage conditions that are unlikely to represent industry practise. The panel believe that this question could be readily resolved with some simple experimentation, but this has not been carried out to date.

DNA in any matrix will decline over time, even in storage buffers at low temperature. The question is whether DNA is declining at such a rate in honey (due to temperature, chemical activity or other factors) such that it is too unstable to be included in the MPI definition. There is currently insufficient evidence to conclude that the DNA marker should be excluded. Further investigation should be carried out to better understand what factors may accelerate DNA decline in the honey matrix to determine if it is possible to avoid this by modifying industry practise, or whether this is unavoidable and to better understand if this is a substantial issue that may warrant modification of the mānuka honey definition.

6 General conclusions and recommendations

The current MPI mānuka honey definition is built on a considerable dataset, makes good use of the information available, and there is little justification for substantially altering this based on available data. While there is a considerable number of data points available to test the appropriateness of the definition, the majority of them have not been deliberately collected for this purpose and therefore are not structured ideally for drawing statistical inferences. A clear limitation of the data is that it is dominated by mānuka honey, both monofloral and multifloral. In constructing a definition for a product, it is essential that the test is developed from information on what the product is but also specifically what it is not. Ideally the amount of information for each of the classes that are to be detected should be balanced, rather than strongly biased towards one or two categories. Ideally it should also be collected by those with no vested interest in the findings, such that there is no potential bias towards assignment of classes.

The review panel recommend that there is a need for a curated library of honey samples and associated metadata to be collected within a single repository. This would provide a reliable and trusted set of data to train future classification rules and to test them. A programme of consolidated data collection should be implemented, to provide an adequate database for assessing the concerns of the mānuka honey industry regarding the definition, and to allow conclusions to be drawn without reservation. At the very least the type of honey, its geographical source, and concentrations of all the chemical and DNA markers, both currently used and of identified potential future use, should be collected. It is important also that approximately equal numbers of samples of honey from all four sources (monofloral and multifloral mānuka, non-mānuka from New Zealand, non-mānuka sourced from overseas), currently being considered be collected in each season. Collection should be an ongoing process since changing climate and environmental modification will ultimately affect honey composition. The design of the collection process should be developed by a multi-disciplinary team of chemists, statisticians, and plant ecologists, in conjunction with accredited processors and beekeepers with extensive experience with mānuka honey production. As such, a collection might take a protracted period of time, so concerns about the stability of the DNA marker in stored honey should be tested in a separate laboratory storage study. The samples need to be obtained by independent contractors with no vested interest in the outcome of the construction of the classification rules, and with a clear study design that lends itself to answering specific questions.

The review panel conclude that with current data, there is limited value in modifying the current MPI mānuka honey definition. The only exception to this is the possible removal of the cap on the 3-PLA level in multifloral mānuka honey. From the data available it appears that retention of this cap is likely undesirable in providing an accurate and robust definition for multifloral mānuka honey. Any change should be considered in light of the limitations of the current data, and ideally further information should be collected prior to revision of the definition.

7 References cited

Koot E, Arnst E, Taane M, Goldsmith K, Thrimawithana A, Reihana K, González-Martínez SC, Goldsmith V, Houliston G, Chagné D 2022. Genome-wide patterns of genetic diversity, population structure and demographic history in mānuka (*Leptospermum scoparium*) growing on indigenous Māori land. Horticulture Research 9: uhab012.

McDonald CM, Keeling SE, Brewer MJ, Hathaway SC 2018. Using chemical and DNA marker analysis to authenticate a high-value food, mānuka honey. Science of Food 2(1): 9.

MPI Technical Report (October 2018). Regional variation in mānuka honey markers.

MPI report (November 2021). Regionality of 2′-MAP in Mānuka Honey. Prepared for Overall Review Panel by Operational Research, New Zealand Food Safety, Ministry for Primary Industries.

MPI report (May 2022). Regionality of 2′-MAP in Mānuka Honey. Prepared for Overall Review Panel by Operational Research, New Zealand Food Safety, Ministry for Primary Industries.

MPI report (June 2022) Investigating the risk of honey blending on the MPI mānuka honey definition. Prepared for Overall Review Panel by Operational Research, New Zealand Food Safety, Ministry for Primary Industries.

MPI report (November 2022) Reassessment Report on DNA stability in the honey matrix. Prepared for Overall Review Panel by Operational Research, New Zealand Food Safety, Ministry for Primary Industries.

MPI report – The value of mānuka DNA in classifying honeys and alternative markers.