The selection and analysis of fatty acid ratios: A new approach for the univariate and multivariate analysis of fatty acid trophic markers in marine pelagic organisms

Martin Graeve [1]* Michael J. Greenacre [2]
[1]Alfred-Wegener Institute Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany
[2]Universitat Pompeu Fabra and Barcelona Graduate School of Economics, Barcelona, Spain

Abstract

Fatty acid (FA) compositions provide insights about storage and feeding modes of marine organisms, characterizing trophic relationships in the marine food web. Such compositional data, which are normalized to sum to 1, have values—and thus derived statistics as well—that depend on the particular mix of components that constitute the composition. In FA studies, if the set of FAs under investigation is different in two separate studies, all the summary statistics and relationships between the FAs that are common to the two studies are artificially changed due to the normalization, and thus incomparable. Ratios of FAs, however, are invariant to the particular choice of FAs under consideration—they are said to be subcompositionally coherent. Here, we document the collaboration between a biochemist (M.G.) and a statistician (M.J.G.) to determine a suitable small set of FA ratios that effectively replaces the original data set for the purposes of univariate and multivariate analysis. This strategy is applied to two FA data sets, on copepods and amphipods, respectively, and is widely applicable in other contexts. The selection of ratios is performed in such a way as to satisfy substantive requirements in the context of the respective data set, namely to explain phenomena of interest relevant to the particular species, as well as the statistical requirement to explain as much variance in the FA data set as possible. Benefits of this new approach are (1) univariate statistics that can be validly compared between different studies, and (2) a simplified multivariate analysis of the reduced set of ratios, giving practically the same results as the analysis of the full FA data set.

Marine zooplankton organisms are the main vehicles for transporting energy from the primary producers to the major consumers such as fish, marine mammals, and seabirds. Knowledge of food web structure, including the length, connectivity, and primary sources of trophic pathways, is important for our understanding of ecosystem functions (McGovern et al. 2018). Food web studies can help distinguish pathways of biogeochemical cycling, as well as elucidate the relationships between biodiversity and ecosystem functioning. Knowledge of energy flows can also provide insights into how the community may be impacted by future biotic and abiotic changes such as species introductions, altered patterns in productivity, warming temperatures, and other environmental and anthropogenic changes (Renaud et al. 2010). Biochemical tracer methods such as stable isotope analysis and

fatty acid (FA) biomarker analysis can provide a time-integrated measure of food source uses and trophic position. Lipids are important biomolecules providing energy and are forming biomembranes, which are essential building blocks for cellular structures. Major neutral lipids such as triacylglycerols or wax esters are typical energy-rich storage lipids, whereas polar lipids such as phosphoglycerolipids (e.g., phosphatidyletanolamin and phosphatidylcholin) are principal membrane lipids (Parrish 1988; Vaskovsky 1989; Sargent et al. 1993). FAs are ubiquitous components of these lipid compounds. They are useful biomarkers for trophic studies, as they tend to remain intact within tissues and can accumulate over time, representing dietary intake over longer timescales (Dalsgaard et al. 2003). FAs are biosynthesized de novo starting from the 2-carbon unit acetyl-coenzyme-A (CoA) following the fatty acid synthetase (FAS) loop, until 14:0 FA, followed by an elongation to 16:0 and 18:0 FAs. Some animal species are able to elongate the chain further and unsaturate these compounds to 20:1, 22:1, and 24:1 FAs. Only autotrophic organisms are capable of biosynthesizing (n−3) and (n−6) FAs. Following this pathway, a number of specific desaturases Δ12, Δ15, Δ5, Δ6 and elongases are essential to form end products such as 20:5(n−3), 22:6(n−3), and 20:4(n−6) polyunsaturated fatty acids (PUFAs) (Sprecher 2000;

*Correspondence: martin.graeve@awi.de

Additional Supporting Information, including a video abstract, may be found in the online version of this article.

Leonard et al. 2004). The pathways of FA biosynthesis for both zooplankton species and phytoplankton are shown in Fig. 1. These essential FAs or fatty acid trophic markers (FATMs) are transferred unchanged through the food chain from planktonic microalgae to higher trophic levels (Dalsgaard et al. 2003), such as fish, whales, and seals.

By means of FATMs the FA profiles, especially of marine organisms (e.g., Arctic zooplankton and benthic organisms), can be used to evaluate feeding history, trophic position, and life cycle strategies (Sargent et al. 1981; Falk-Petersen et al. 1990; Lee et al. 2006). For example, diatoms (Bacillariophyceae) have high amounts of the FATMs $16:1(n–7)$ and $20:5(n–3)$, along with high levels of C16 PUFAs. Dinoflagellates (Dinophyceae) have high proportions of the $22:6(n–3)$ FA and C18 PUFAs (Graeve et al. 1994a, 1994b; Dalsgaard et al. 2003). These FATMs are incorporated unchanged in storage and membrane lipids of marine zooplankton and are rapidly transferred through the food web, supplying higher trophic levels with the required energy (Falk-Petersen et al. 1990). This lipid-based flux of energy takes place in many organisms, but is essential in the lipid-driven Arctic food web (Graeve et al. 2005; Boissonnot et al. 2016, 2019). The various biochemical processes that produce lipid reserves of different compositions enable species to utilize different ecological niches, and are major determinants of biodiversity in polar zooplankton (Falk-Petersen et al. 2000, 2001). There are other studies providing information on the transfer of FAs in higher trophic level organisms such as seals and whales (Budge et al. 2008; Falk-Petersen et al. 2009)

when fed on different food resources. Especially in the case of higher trophic levels, these data sets need to be further evaluated to better understand the dynamics of FA transfer and utilization. There is a need to provide modern statistical methods that are widely applicable to a broad variety of FA data sets, for the best possible analysis of trophic relationships.

Investigation of the lipid and FA composition of marine or aquatic organisms often results in large data sets containing a high number of FA components. The reliability of the data depends on (1) lipid extraction, (2) derivatization, and (3) gas chromatography and identification of compounds. While nature determines the set of FAs, which could be up to more than 50 components, an individual data set is pretty much determined by the limit of detection of the analysis in a specific laboratory and therefore the number of detected FAs may vary between individual studies. When it comes to data analysis, the data are generally provided as mass percentages and, most commonly, are summarized by mean values and some error measure such as the standard deviation or standard error, even though these values depend on the particular subset of FAs included in the study. Some typical examples are Dalsgaard et al. (2003), Budge et al. (2008), Falk-Petersen et al. (2000, 2001, 2009), Søreide et al. (2010), and Pethybridge et al. (2014). Standard multivariate analysis plots are generally used without a discussion of the advantages/disadvantages of these methods, such as the commonly used principal component analysis (PCA) (e.g., Peterson and Klug 1994, Jolliffe et al. 2007; Petursdottir et al. 2008; Brett et al. 2009; Pethybridge et al. 2014; Tartu et al.
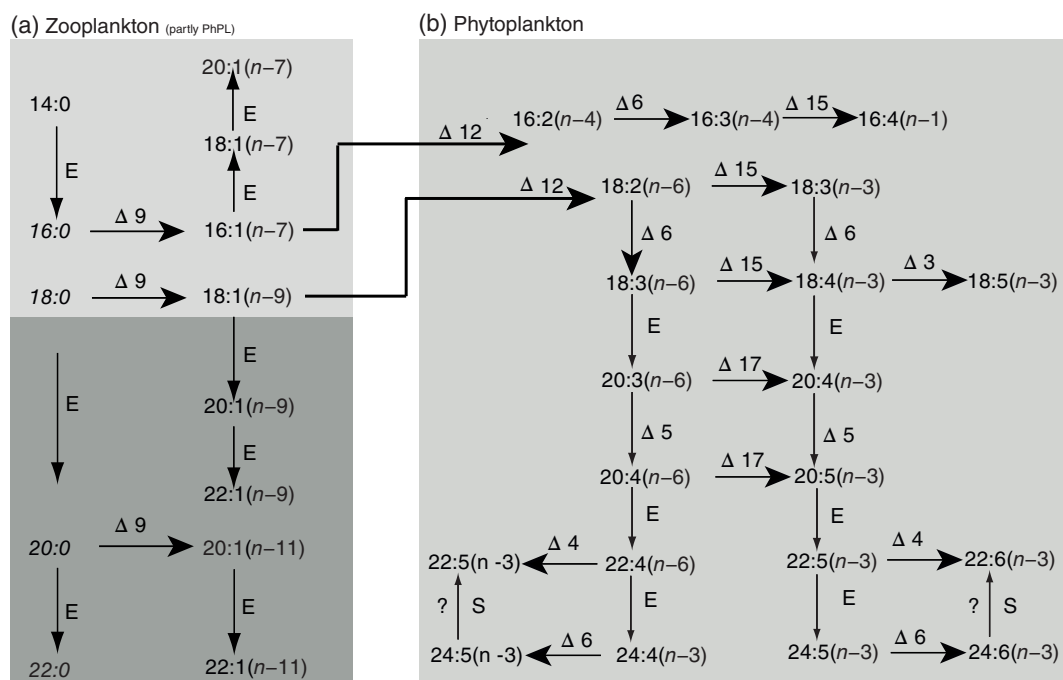

Fig. 1. Fatty acid biosynthesis of zooplankton (**a**) and phytoplankton (**b**). Δ x, desaturation at given position in the FA chain; E, chain elongation; S, chain shortage.

2016; Imamura et al. 2017). These studies all use PCA on normalized FA values (i.e., compositional data that sum to 1% or 100%), which again depend on the particular set of FAs included. The problem of spurious correlations that result from this normalization has been known for over a century (Pearson 1897), effectively ruling out PCA as an appropriate method for compositional data—see the further remarks below about the use of PCA.

When it comes to analyzing percentages such as in a typical FA data set, the compositional data analysis literature (see, e.g., the fundamental book by Aitchison 1986 and the multiauthored publication edited by Pawlowsky-Glahn and Buccianti 2011) states explicitly that conventional statistical tools should be avoided because the results depend on the subset of compositional components studied. An acceptable solution of this problem, when the components are FAs, is to rather consider FA ratios, since these are unaffected by the particular mix of FAs chosen in any particular study. The ratio 14:0/18:0, for example, remains the same whatever other FAs are included, with or without normalization—ratios are thus said to have *subcompositional coherence* and can be compared across studies. For statistical analysis such as regression, ANOVA, or PCA, ratios are analyzed on a logarithmic scale, since the logarithmic transformation converts ratio-scale data to interval-scale, hence the term "logratio."

Logratios make sense for multivariate analysis as well, as demonstrated by various publications promoting logratio analysis (LRA) as the appropriate way to ordinate compositional data (Aitchison 1986, 1990; Aitchison and Greenacre 2002; Greenacre and Lewi 2009; Pawlowsky-Glahn and Buccianti 2011; Greenacre 2018, 2019). LRA involves analyzing all the pairwise logratios in a global analysis, including optional weights for the FAs that are by default proportional to their mean percentage. Apart from theoretical advantages (*see* Greenacre and Lewi 2009), this weighting is designed to solve the practical problem that FAs with very low values can induce ratios with very high variance, while FAs with high values usually induce ratios with low variance. Weighting factors proportional to average FA percentages thus have a standardizing role, but other choices could depend on knowledge of the measurement errors in the FA values.

Correspondence analysis (CA) of compositional data analysis (e.g., Kraft et al. 2015; Meier et al. 2016; Haug et al. 2017) has been justified as being nearly equivalent to using an approach based on logratios (Greenacre 2010; Stewart 2017). This is because of the close theoretical relationship between the distance measure based on logratios and the chi-square distance that is inherent in CA (Greenacre 2010, 2011). In other words, one can say that in practice CA can be approximately subcompositionally coherent.

PCA, on the other hand, suffers severely from a lack of subcompositional coherence (Greenacre 2011). For example, a covariance or correlation between FAs changes in the presence (or absence) of other FAs, after renormalization of the percentages—for example, see Greenacre (2018). Since PCA, for example, is an analysis of the covariance or correlation matrix, this rules out PCA as a suitable multivariate method to analyze FA data. While a PCA of the FA percentages might produce a similar result and conclusion in many cases, compared to the alternative that is offered in this article, this does not justify PCA as an appropriate methodology.

Brenna et al. (2018) express similar concerns about the way FA data are analyzed. They report the wide range of numbers of FAs across many studies and make the same obvious point that "the fewer the fatty acids that are summed, the greater the apparent profile percentage of those reported." They publish a list of 21 FAs which should be included in a study, accounting for more than 95% of the total plasma FAs (their study is of human blood FA composition). Mocking et al. (2012) comment on the biased negative correlation problem between FAs, namely that "an increase in the percentage of one FA automatically results in the decrease in the relative percentage of another FA."

The main drawback of basing the statistical analysis on logratios is that no zeros are allowed. A zero value in a FA data set is not a structural zero; instead, it is a small value below the detection limit of the measurement process. A strategy is thus necessary to replace zeros in a data set with appropriate positive values, for example, half the corresponding detection limit, or another fraction of it (see, e.g., Palarea-Albaladejo et al. 2007 and references therein). Greenacre (2018) shows how a sensitivity analysis can be performed on a data set which is subjected to varying small values used to replace the zeros. Alternatively, the CA approach can be used (Greenacre 2010, 2011), adopted by Stewart (2017), because CA has no problems with analyzing data zeros—in fact, it is the ability of CA to handle large sparse data matrices (i.e., data with a very high percentage of zero values) that makes it a method of choice in ecological data analysis as well as in archeology and linguistics. The present study, however, will focus on the ideal case of logratios as the fundamental data transformation, with its property of strict subcompositional coherence, so the data set needs to have strictly positive values.

Few authors have used the logratio approach in FA analysis, although it is extensively used in the geochemical literature (e.g., see the journal *Mathematical Geosciences*, formerly *Mathematical Geology*, for many publications). In the FA literature Neubauer and Jensen (2015) consider how to select FAs that discriminate between predator diets in a controlled experiment, using centered logratios, that is, the logarithms of each FA divided by the geometric mean of all the FAs, computed for each individual (for an introduction to logratio transformations and analysis from a practitioner's viewpoint, see Greenacre 2018). Werker and Hall (2000) measure a subset of 10 FAs in an experiment and display logratios of nine of them relative to the most frequent one, 16:0, which are called additive logratios (Aitchison 1986). Similarly, Thiemann et al. (2008) use the 17 most abundant and variable FAs in ratios with respect to 18:0, also additive logratios. Our approach here is to consider the complete set of pairwise logratios in the first instance, and then to reduce it to a smaller subset with optimal properties.

The objective in the present study is twofold. First, we aim to show that by a simple stepwise procedure, a small set of FA ratios can be identified that essentially explains the major and most relevant part of the information in a FA data set (Greenacre 2019), as measured by the total variance of the logratios. This procedure comprises a statistical criterion that allows the FA ratios in a particular application to be ordered in terms of statistical relevance, from which the biochemist, who has substantive knowledge of the particular study, can make an expert choice of the FA ratio to be included at each step. Second, we aim to show that this reduced set of ratios can provide valid univariate and multivariate representations of the complete FA data set and that this considerably simplifies the interpretation and understanding of the compositional data.

Material and methods

Sample material

Two different data sets are used to illustrate the proposed approach. Although they are analyzed independently, they are chosen to show their differences in feeding behavior and thus the importance of different FAs in the selected ratios.

Calanoid copepods were collected during an extensive field study in Rijpfjorden, a high Arctic sea ice dominated ecosystem, during the International Polar Year 2007/2008. The seasonal development of the key pelagic grazer *Calanus glacialis* was investigated together with the ice algae and phytoplankton growth, see Søreide et al. (2010). This data set is composed of 42 copepods and 40 FAs.

Amphipods were sampled around Svalbard, across the eastern and central Fram Strait and the Arctic Ocean, during the ARCTOS BIO winter cruise in January 2012; the IMR Ecosystem Survey cruise in August 2011; the ARK-XXVI/2 expedition to the long-term observatory HAUSGARTEN in July and August 2011 the eastern Fram Strait and on a 78°85″N transect across the central Fram Strait (ARK-XXVI/1) in June and July 2011 (for details, refer to Kraft et al. 2015). This data set is composed of 52 amphipods and 27 FAs.

Lipid extraction

Total lipid was extracted by homogenizing animal tissues and filters in a solution of dichloromethane : methanol (2 : 1, v : v), modified after Folch et al. (1957). As internal standard, a known amount of the tricosanoic acid methyl ester (23:0) was added to each sample. A 0.88% solution of KCl (potassium chloride) was added to easily differentiate the biphasic system. Transesterification of the lipid extracts was performed by heating the samples with 3% sulfuric acid (H_2SO_4) in methanol for 4 h at 80°C under nitrogen atmosphere.

FA analysis

FA and fatty alcohol compositions were identified according to Kattner and Fricke (1986). Subsequent analyses were done by gas liquid chromatography (HP 6890N GC) on a wall-coated open tubular column (30 × 0.25 mm internal diameter; film thickness: 0.25 μm; liquid phase: DB-FFAP) using temperature programming. Standard mixtures served to identify the FA methyl esters and the fatty alcohol derivatives. If necessary, further identification was done by gas chromatography-mass spectrometry using a comparable capillary column. Detailed FA and alcohol compositions were expressed as percent of total FA and percent of total fatty alcohols, respectively. However, for doing the statistical analysis, we considered FAs only.

Statistical analysis

The objective is to define a set of FA ratios that adequately describe the FA compositional data set, which are acceptable from both a biological and a statistical point of view. The statistical methodology has been described and justified in detail in an archeometric application by Greenacre (2018, 2019). Here we give a summary of the main features of this new analytical approach and the steps involved in the selection of the FA ratios.

The ideal in compositional data analysis is to analyze the full set of FA ratios, all logarithmically transformed, that is, all pairwise logratios. However, for a set of m FAs, there are $\frac{1}{2}m(m-1)$ possible logratios, of which at most $m-1$ of them can be linearly independent (i.e., none among the $m-1$ logratios can be computed from the others). Putting this another way, given any such subset of $m-1$ linearly independent ratios, all of the others in the full set of $\frac{1}{2}m(m-1)$ logratios depend on them linearly. This is analogous to the fact that for a compositional data set of m FAs, one of them is always 1 minus the sum of the $m-1$ others—the rank, or dimensionality, of the data set is equal to $m-1$. There are very many possible choices of this subset of $m-1$ linearly independent logratios. Using a result from network theory, Greenacre (2019) reports that there are m^{m-2} possible subsets, which for only 10 FAs would give 10^8 possibilities, and it is clearly not feasible to investigate them all. Hence, a stepwise approach is adopted, which apart from being much more efficient, has the additional benefit of lending itself to a collaboration between the statistician and the biochemist at each step of the ratio selection process.

The fundamental theoretical concept in this approach is that the full set of FA logratios has a total (weighted) logratio variance, defined by Greenacre (2018, 2019), which is taken as the information "content" of the data set. A single FA logratio explains a certain percentage of this variance, which can be easily computed. What was stated previously can now be rephrased as follows: $m-1$ linearly independent logratios explain 100% of the total logratio variance. To measure the variance explained by any subset of logratios, a generalization of regression to multivariate responses, called redundancy analysis (RDA) (van den Wollenberg 1977), was employed, using the vegan package (URL: http://CRAN.R-project.org/package=vegan, last accessed 11 July 2019) in R (R Core Team 2019). RDA is generally used to relate a set of response variables (usually a high number of variables) to a set of explanatory variables (usually a small set). Here it was used to see how

well all the logratios (again, a large set) are explained by a subset of a few logratios.

The first step of our procedure, performed by the statistician, was to use RDA to calculate how much of the total variance can be explained by each FA log-ratio, and a list was made of the best ones, for example, the "top 10" or "top 20," in descending order of importance. This list was then considered by the biochemist in terms of biological relevance in the context of the specimens under study and the objective of the FA analysis. The biochemist either confirmed the best ratio or chose one near the top of the list apparently more related to biological function than the best one. This ratio was selected and the next list of top FA logratios that explained most of the residual variance was established by the statistician, again using RDA, and presented again to the biochemist, who in turn chose the most biochemically relevant ratio at the top of the list, or near the top. This iterative procedure continued until the FA ratios were becoming substantively irrelevant. This exercise was performed on the copepod and amphipod data sets, in each case resulting in a list of FA ratios that were both statistically and biologically relevant to the taxa studied as well as the objective of the research. Once the final set of ratios was established, a graph was made of the ratios, in the form of a network with vertices being the FAs and edges linking the vertices indicating the chosen ratio (Greenacre 2018, 2019). In the terminology of network analysis, this is an *acyclic graph*, since there is no closed circuit. If there were such a closed circuit, the ratios would not be independent, hence by implication a set of independent ratios is represented by an acyclic graph.

To show that the reduced set of logratios adequately described the total variance of the complete FA data set, two multivariate analyses were performed. First, an ordination was made based on the full set of logratios, constituting a weighted LRA (Greenacre and Lewi 2009), where the weights aim to compensate for the different levels of measurement error in each FA (cf. Aitchison's original definition of an unweighted LRA by Aitchison 1990 and Aitchison and Greenacre 2002)—see Greenacre (2019) for full details. This analysis gave an optimal view of the samples based on their exact intersample logratio distances.

Second, a PCA of just the selected small subset of logratios was performed to show the relative positions of the samples being almost identical to that based on the full set, thus validating the procedure taken in selecting the "best" subset of logratios. Notice that the previous criticism of PCA being applied to compositional data is not relevant here, since the data are unstandardized logratios, which are appropriate for PCA. The degree of matching of the positions of the samples between the two analyses, that is, the similarity in their multivariate structure, was measured using Procrustes analysis (see, e.g., Krzanowski 1987), specifically the Procrustes correlation—again, see Greenacre (2018) for details as well as the mathematical definition.

In all ordinations, the contribution biplot scaling of Greenacre (2013) was used, showing the major contributing variables as more outlying. This version of the biplot facilitated interpretation and justified downplaying those variables lying close to the origin of the ordination and thus contributing relatively little to the solution.

In the first data set, the samples were obtained in three different seasons. As a further illustration of the power of simple logratios to explain structure in a compositional data set, a classification tree (Breiman et al. 1984; Hastie et al. 2009) was estimated to predict the season of each sample, using the total pool of logratios as possible predictors.

All computations were performed using the R statistical system (R Core Team 2016) and extensive use was made of the new R package easyCODA, which accompanies the book by Greenacre (2018) and which includes the stepwise procedure.

The selected ratios can be validly summarized using regular univariate statistical summaries, always remembering that ratios are bound to be positively skewed. Hence, their medians were chosen as measures of centrality and their reference ranges as measures of dispersion. A reference range (Greenacre 2016) is an estimate of the interval enclosing 95% of the data values and is computed from the estimated 2.5% and 97.5% percentiles of a ratio's sample distribution, using the quantile function in R. Because ratios are subcompositionally coherent, they can be compared with the same ratios and their univariate summaries in other studies.

Results

Copepod data set

The total logratio variance in this data set was 0.2584, based on the total of 780 possible ratios formed by the 40 FAs. Each logratio explained its own part of variance as well as parts of variance in all the other logratios with which it was correlated. The following sequence of six steps identified six logratios that explained 91.0% of this variance, following which the addition of more logratios had minimal statistical and substantive relevance. The full list of ratios provided to the biochemist at each step is given as Supporting Information. The steps are summarized in Table 1, which also includes the medians and reference ranges of the respective ratios (untransformed).

Step 1: Ratio 16:0/18:4(n–3)

The logarithm of this ratio had the highest explanatory value, at 54.3%. Other ratios formed with 18:4(n–3) were close behind, for example, with 18:1(n–7) (54.2%), 22:1(n–7) (53.3%), and 16:1(n–9) (53.2%), but the best one with 16:0 was retained because it was representing a contrast of a dietary or essential FA and a typical de novo synthesized FA. In general, 18:4(n–3) is used as FATM for dinoflagellates and 16:0 is a ubiquitous FA mostly connected to membrane lipids. Eventually this pair represents both metabolic pools, storage and membrane lipids.

Step 2: Ratio 16:0/16:1(n–7)

After including 16:0/18:4(n–3), this was the seventh best in a new list of additional candidate ratios, explaining

Table 1. Sequence of FA ratios in stepwise procedure for the copepod data set of 40 FAs, showing additional and cumulative percentages of variance explained. Medians and reference ranges estimated to include 95% of the ratio's distribution are also given as examples of valid univariate statistics.

Step	Ratio	Additional % explained	Cumulative % explained	Median	Reference range
1.	16:0/18:4(n−3)	54.3	54.3	1.166	0.507–44.223
2.	16:0/16:1(n−7)	21.0	75.3	0.919	0.324–1.857
3.	20:1(n−9)/22:6(n−3)	6.6	81.9	1.750	0.590–2.518
4.	16:0/20:1(n−9)	3.2	85.1	0.611	0.425–1.382
5.	14:0/20:5(n−3)	3.3	88.4	0.731	0.417–2.591
6.	18:0/20:5(n−3)	2.6	91.0	0.143	0.068–2.603

an additional 21.0% of the variance, the best being 16:1(n−7)/22:1(n−11) which explained an additional 21.6%. The ratio 16:0/16:1(n−7) is reflecting the same structure as given for the first ratio, but the denominator represents the diatom FATM 16:1(n−7).

Step 3: Ratio 20:1(n–9)/22:6(n–3)

This ratio explained an additional 6.6% of the variance, whereas the maximum was achieved by 20:1(n−9)/24:1(n−9), explaining an additional 7.5%. The ratio 20:1(n−9)/22:6(n−3) was preferred, because 22:6(n−3) is also a major dinoflagellate FATM and is most likely connected with the de novo synthesized FA 20:1(n−9), so this ratio represents a typical *Calanus* wax ester compound.

Step 4: Ratio 16:0/20:1(n–9)

This ratio was preferred, explaining an additional 3.2% of the variance. The ratio 16:0/20:1(n−9) was explaining less variance but the occurrence of both FAs in storage or membrane lipids seemed to be more likely. As in Step 3, the ratio 20:1(n−9)/24:1(n−9) again gave a high additional explained variance of 7.5%, but in most studies 24:1(n−9) FA occurs only in traces and a contrast with 20:1(n−9) FA does not really seem obvious. Therefore we decided to eliminate 24:1(n−9) from consideration in this step and all subsequent ones.

Step 5: Ratio 14:0/20:5(n–3)

This ratio, which was the best according to the statistical criterion, explaining an additional 3.3% of the variance, was also from the biochemistry point of view a good contrast. It

showed either a combination of two phospholipid derived FA or a ratio of a typical de novo synthesized short chain FA 14:0 with a dietary FA 20:5(n−3) representing a diatom FATM.

Step 6: Ratio 18:0/20:5(n–3)

This ratio, also the best at this stage according to the statistical criterion, explained an additional 2.6% of the variance, in this case the biochemical criterion was comparable to the ratio above, 14:0/20:5(n−3).

After these six logratios entered, involving a total of only eight out of the 40 FAs, explaining 91.0% of the total variance, the entry of further logratios presented no clear substantive biological interpretation. We thus stopped the procedure at this point. Figure 2 shows the acyclic graph of the eight FAs, where each of the six edges connects the two FAs of the corresponding ratio.

The LRA of the full set of FAs is shown in Fig. 3a, representing the analysis of the full data set. Thanks to the contribution biplot scaling, the FAs contributing more are more outlying, while all those making less than average contributions to the solution are closer to the center and deemphasized by showing them in a smaller and lighter font. The LRA shows the 40 FAs but is implicitly analyzing all 780 logratios, which are the connections between all pairs of FA points. On the other hand, the PCA of the reduced set of six selected logratios is presented in Fig. 3b, showing a clear agreement with the ordination of the samples. To quantify the similarity between the two results, the Procrustes correlation is measured at 0.977, highlighting a very good concordance between the two ordinations, with the three groups of samples being separated in a similar way. The separation of these three groups reflects seasonal variations in FA composition of the copepods from summer to winter and spring population (counterclockwise, starting from the right (Fig. 3a). The only major differences are firstly, the splitting of the winter samples into two groups in Fig. 3b, and secondly, the tendency of one summer sample towards the spring group.

To show how the six selected logratios accounted for 91.0% of the total variance, Fig. 4 shows the decomposition of the variance of each FA into parts explained by the six logratios. The FAs on the left are ordered from highest to lowest contributions to logratio variance, with their percentages of variance
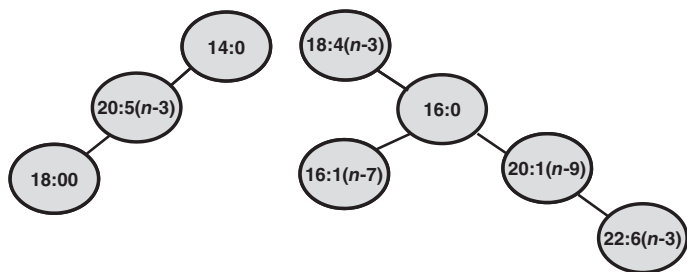

Fig. 2. Acyclic graph of eight FAs for the copepod data set, where each edge links the pair of FAs that defines a ratio.
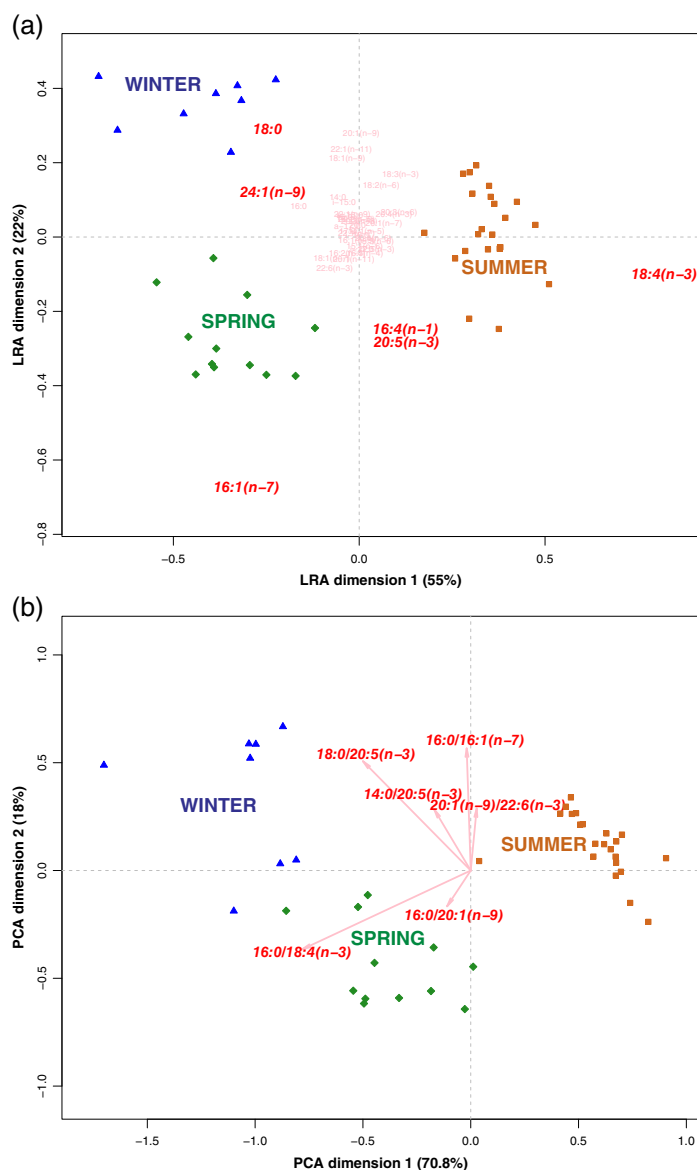
Fig. 3. (a) Contribution LRA biplot of the copepod data set. The FAs that contribute more than average to the solution are shown in red; all the others are shown in smaller font in pink. The three clusters show the influence of dietary FA in the populations during summer, winter, and spring seasons. **(b)** PCA contribution biplot of the six logratios identified by the combination of statistical and biochemical criteria, explaining 91.0% of the logratio variance of the copepod data set (the two-dimensional PCA explains 88.8% of that 91.0%, according to the percentages on the two axes).

depicted by the bar chart on the right. The logratio variance of a part is made up of the sum of variances of the logratios of that part relative to all the other parts. Then for each FA, the proportion of its logratio variance explained by each of the six logratios is shown, broken down into parts, as well as a part that is unexplained. For example, in the first and ninth rows, the logratio 16:0/18:4(n–3) is shown to explain almost all of the logratio variance of 18:4(n–3) and the major part of the

logratio variance of 16:0. The gray bars on the right indicate the unexplained parts of variance, which become very large for the FAs lower down, but these have very small logratio variances in absolute terms.

The seasonal distinction of the samples was perfectly predicted by the first two logratios selected by our procedure (Table 1), shown as a classification tree in Fig. 5. The ratio 16:0/18:4(n–3) perfectly predicted the 22 summer samples, corresponding to values of the ratio lower than 2.425. Then for higher values of that ratio, the ratio 16:0/16:1(n–7) perfectly predicted the eight winter samples for values higher than 0.9358 and the 12 spring samples for values lower than 0.9358.

Amphipod data set

The total weighted logratio variance in this data set equals 0.4528, higher than that of the copepod data set. The following sequence of steps identifies eight logratios that explain 91.6% of this variance, following which the addition of more logratios has minimal substantive relevance. Full details of the FA ratios considered at each step are given in the Supporting Information and the steps are summarized in Table 2.

Step 1: Ratio 20:5(n–3)/22:1(n–11)

This ratio was the third best, explaining 41.0% of the variance, the best ratio being 18:4(n–3)/22:1(n–11), explaining 42.1%. The ratio 20:5(n–3)/22:1(n–11) was preferred because it shows a typical membrane FA in contrast with a long-chain de novo FA. Basically, this ratio reflects a pair of FAs following different biosynthetic pathways.

Step 2: Ratio 16:0/22:1(n–11)

Having introduced 20:5(n–3)/22:1(n–11) in the first step, this ratio was the best according to the statistical criterion, explaining a maximum additional variance of 25.8%. It has the same biochemical relevance as described for the first ratio. Again, this is a biochemically relevant ratio of FAs deriving from *Calanus* dietary markers and structural FA sources.

Step 3: Ratio 18:0/18:4(n–3)

This ratio explained an additional 12.6% of the variance, whereas the maximum would be achieved by 18:1(n–9)/20:4 (n–3), explaining 13.1%, that is, only 0.5 percentage points more. The ratio 18:0/18:4(n–3) was chosen because as typical dinoflagellate FATM 18:4(n–3) and 18:0 as a de novo synthesized FA could be multiconnected, but appeared most likely in membrane lipids.

Step 4: Ratio 18:0/20:4(n–3)

This ratio was the second best one by a small fraction of a percentage point, explaining an additional 8.2% of the variance. The statistically best one was 20:4(n–3)/22:6(n–3), explaining an additional 8.3% of the variance. The ratio 18:0/20:4(n–3) was chosen because contrasting two polyunsaturated FAs will be biochemically difficult. A ratio with a numerator 18:0 and 20:4(n–3) as denominator could represent
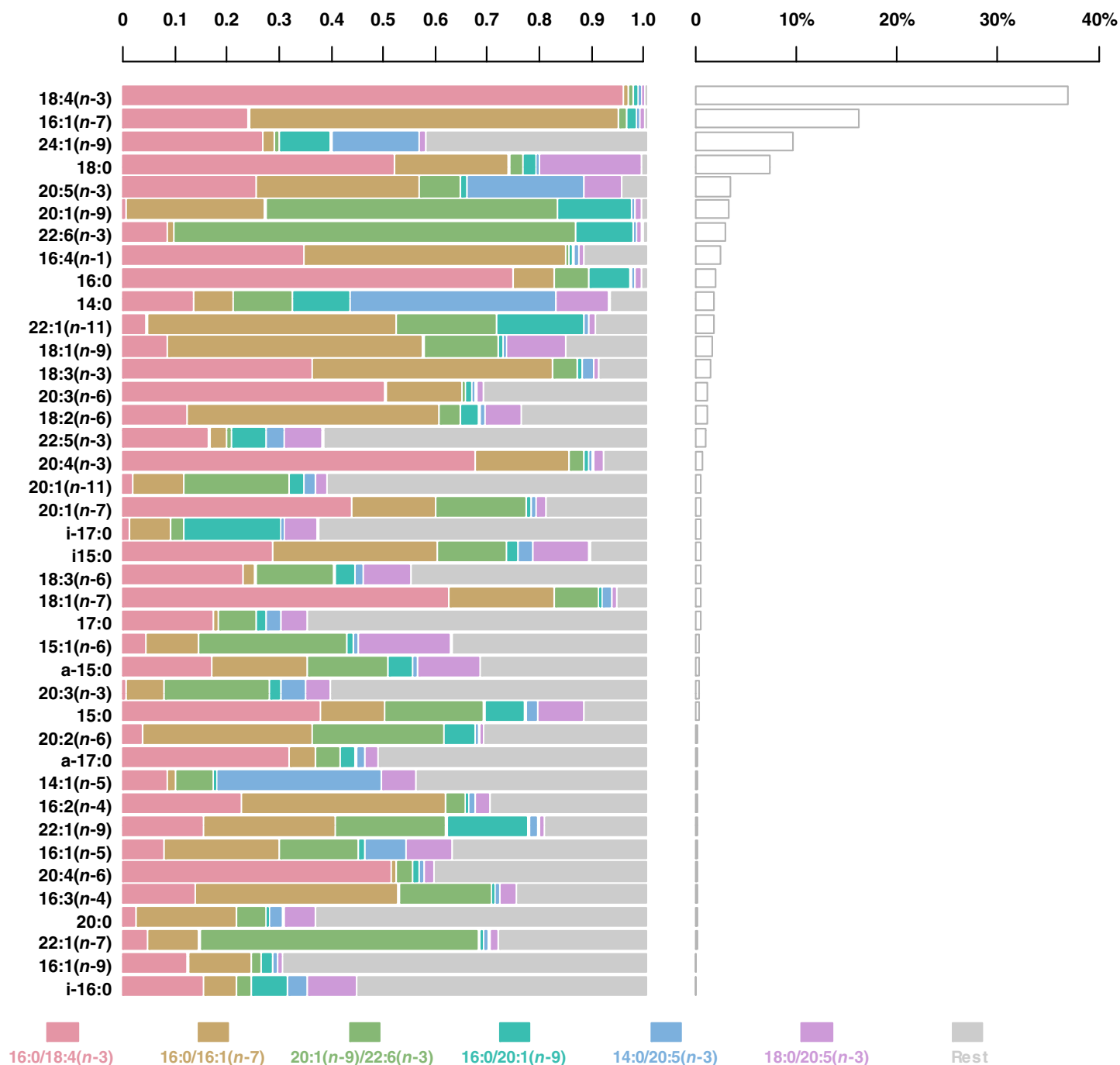
Fig. 4. Visualization of how the selected FA ratios explain the variance of the FAs for the copepod data set. The FAs are ordered from the highest to lowest variance, showing their percentages of total variance in the bar chart on the right. The proportions of each FA's variance accounted for by the six selected ratios are shown on the left, with gray indicating unexplained variance.

a typical contrast of FAs in phospholipids or reflect a moiety of storage lipids.

Step 5: Ratio 18:1(n–9)/20:1(n–9)

This was the statistically best ratio to enter at this stage, explaining an additional 6.3% of the variance. It is an entity

of a long chain de novo synthesized FA and a ubiquitous FA with membrane lipid origin.

Step 6: Ratio 16:1(n–7)/22:6(n–3)

This was again the best ratio to enter at this stage, from a statistical point of view, explaining an additional 3.6% of the variance. It
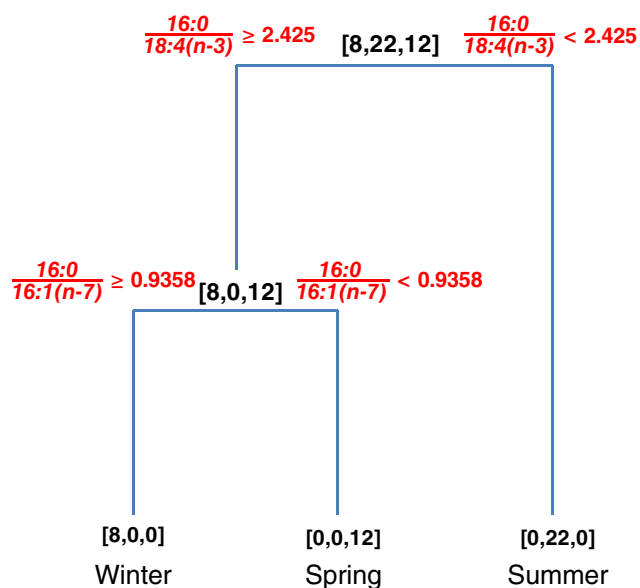
Fig. 5. Classification tree using the two best ratios of Table 1, showing a perfect prediction of the seasons of the samples. The sample sizes of winter, summer, and spring are indicated by [8, 22, 12] at the top of the classification tree, and the subsequent set of three frequencies is indicated similarly at each node of the tree, with the terminal nodes showing just one season in each. The inequality conditions sending samples left or right are given at each of the two decision nodes.

is a composite of the diatom marker 16:1(n–7) and a structural lipid derived FA 22:6(n–3), which also represents a dinoflagellate marker.

Step 7: Ratio 20:1(n–9)/22:1(n–11)

This ratio explained an additional 2.1% of the variance, whereas the maximum additional variance explained was achieved by four ratios that involve 20:1(n–11), namely with respect to 16:0 or 20:5(n–3) or 22:1(n–11) or 14:0, all explaining an additional 2.2%. The ratio 20:1(n–9)/22:1(n–11) was preferred because this contrast of long chain FAs provides, after some polar lipid dominant ratios selected previously, a typical storage lipid derived moiety, which is a good representative of *Calanus* wax esters.

Step 8: Ratio 16:1(n–7)/20:1(n–11)

This ratio represented a combination of the best choice from both points of view, statistically and biochemically. The diatom FA 16:1(n–7) was selected in combination with a long chain FA 20:1(n–11), which will be taken up by amphipods after feeding on calanoid copepods. The 20:1(n–11) does not represent the major isomer of the 20:1 FAs, but in contrast with this typical diatom FATM it explains an additional 2.0% of the variance.

After these eight logratios entered, involving a total of 11 out of the 27 FAs, having explained 91.6% of the total variance, the entry of further logratios presented no clear substantive biological interpretation. We thus stopped the procedure at this point. Figure 6 shows the acyclic graph of the 11 FAs, where each of the eight edges connects the two FAs of the corresponding ratio. Table 2 lists the chosen ratios with summaries of their explained variances along

Table 2. Sequence of FA ratios in stepwise procedure for the amphipod data set of 27 FAs, showing additional and cumulative percentages of variance explained. Medians and reference ranges estimated to include 95% of the ratio's distribution are also given as examples of valid univariate statistics.

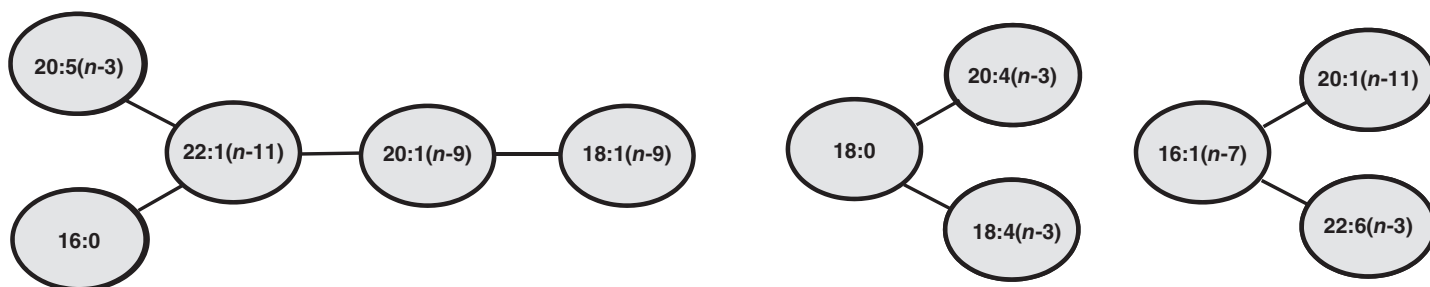| Step	Ratio	Additional % explained	Cumulative % explained	Median	Reference range
1.	20:5(n–3)/22:1(n–11)	41.0	41.0	1.084	0.123–8.295
2	16:0/22:1(n–11)	25.8	56.8	2.432	0.302–15.466
3.	18:0/18:4(n–3)	12.6	69.4	0.400	0.066–26.285
4	18:0/20:4(n–3)	8.2	77.6	1.434	0.057–32.480
5.	18:1(n–9)/20:1(n–9)	6.3	83.9	1.090	0.280–3.906
6.	16:1(n–7)/22:6(n–3)	3.6	87.5	0.818	0.189–6.830
7.	20:1(n–9)/22:1(n–11)	2.1	89.6	1.962	0.694–18.434
8.	16:1(n–7)/20:1(n–11)	2.0	91.6	2.238	0.759–13.997

Fig. 6. Acyclic graph of 11 FAs for the amphipod data set, where each edge links the pair of FAs that defines a ratio.
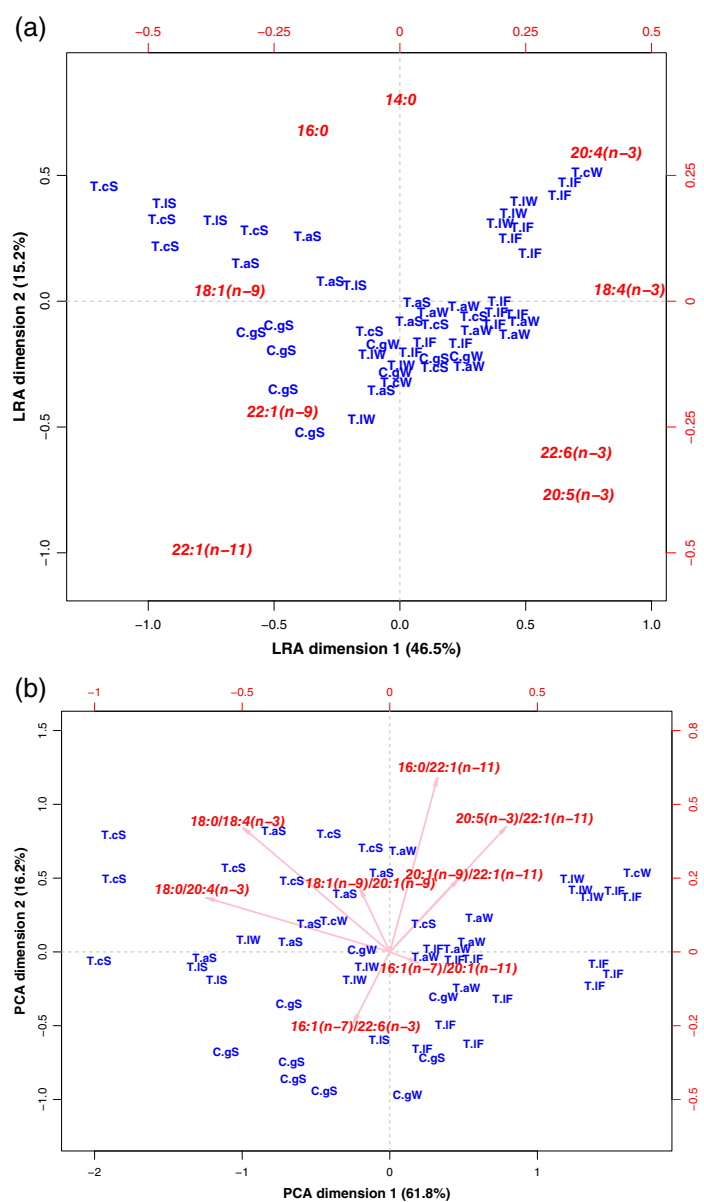
Fig. 7. (**a**) Contribution LRA biplot of the amphipod data set. For legibility, only FAs that contribute highly to the separation of the samples are shown (amphipod species: T.a., *Themisto abyssorum*; T.c., *Themisto compressa*; T.l., *Themisto libellula*; C.g., *Cyclocaris guilelmi*. In capital letters. S, F, W, summer, fall, winter). (**b**) PCA contribution biplot of the eight selected logratios, which explain 91.6% of the logratio variance of the amphipod data set (the two-dimensional PCA explains 78.0% of that 91.6%, according to the percentages on the two axes).

with the univariate measures of median and reference range.

To show how well the eight ratios approximate the original data, Fig. 7 shows first the LRA of the full set of FAs, that is, analyzing all 351 logratios, and second the PCA of the reduced set of eight selected logratios. The similarity in the ordination

of the samples according to season and amphipod species is again apparent, but not as clear as in the copepod example. To measure the concordance between the two ordinations, the Procrustes correlation equals 0.822.

Discussion

The aim of the approach presented here is to show how a small set of FA ratios, selected according to a combination of statistical and biochemical criteria, can effectively replace the complete data set, maintaining its essential multivariate structure as well as providing meaningful univariate statistics. The statistical criteria are based on considering the complete set of ratios of the compositional data set, where the ratios are logarithmically transformed (i.e., the logratios), and then identifying those that maximally explain the total variance of these logratios. A single logratio obviously accounts for its own variance but it also accounts for parts of variance of other logratios with which it is correlated. For example, the logarithm of the ratio 16:0/18:4(n–3) explained 54.3% of the total variance of the copepod data set, whereas the variance of this logratio itself constitutes only a small part, 4.5%, of this total variance (see Fig. 4). It is known that the total logratio variance can be fully explained by a set of linearly independent logratios of size one less than the number of FAs in the compositional data set.

The selection of the ratios is performed in a stepwise manner and at each step the optimal logratio is identified. In the absence of substantive knowledge of the research problem, this logratio would be chosen automatically according to statistical criteria. Our approach, however, identifies not only the best logratio but several others that are almost optimal, and the final choice at each step is given to the biochemist, who has the substantive domain knowledge to be able to select a FA ratio that satisfies the relevant biochemical criteria. Two different zooplankton species are chosen as examples to demonstrate that the ratios chosen are species-dependent and can also vary depending on the research question. For example, a study might be restricted to a specific species without any within-species group comparisons being made, or it might well be comparing species in different regions in which case the ratios would be chosen with this objective in mind. Having said that, it should be noted that the FA ratios chosen in the case of the copepod data set were chosen without taking into account the information about the three seasonal groups, yet the chosen ratios separated these groups perfectly.

Notice that the stepwise procedure is only analogous to and not the same as stepwise regression, which has been criticized in the literature—see, for example, Whittingham et al. (2006) and Mundry and Nunn (2009). The "explanatory variables" in our case are single logratios, and the "response variables" are the complete set of logratios. Both the abovementioned papers stress problems of multiple testing, which are not relevant in the present case since no testing is required. Whittingham et al. (2006)

additionally mention the problem of estimation bias—again, this is irrelevant in the present case since the model parameters are not of interest, it is rather how much variance is explained which is important. Neither is the best single model of interest, but rather the identification of a few logratios that account for almost all of the logratio variance and have substantive biochemical meaning.

The main limitation of the logratio approach is the requirement of strictly positive data. Replacing data zeros by small values can be achieved, for example, by using half the detection limit or half the smallest positive value of the respective FA in the data set. The logarithmic transformation of the ensuing ratios alleviates the effect of introducing these small values. Another potential problem might be that the stepwise procedure can present a surfeit of choice at each step, since many ratios can have the same or almost the same benefit at a particular step. The necessity for expert intervention by the biochemist is invaluable here, avoiding a purely automatic statistical selection of ratios.

Two different data sets were used to contrast the feeding behaviors of typical herbivorous copepod and carnivorous amphipod species. Their FA compositions differ in the ability of copepods to incorporate FAs unchanged from the diet (e.g., 16:1(n–7) and 18:4(n–3)) and to produce long-chain FAs (20:1(n–9), 22:1(n–11)) de novo showing significant seasonal variations with a contrasting ice-algal vs. phytoplankton derived matter. In contrast, the more opportunistic feeding behavior of amphipods revealed FA and

Table 3. Trophic markers and ratios commonly determined in FA profiles of pelagic food sources and consumers (Graeve et al. 1994*a*, 1994*b*, 1997 [a]; Falk-Petersen et al. 1987, 1999 [b]; Auel et al. 2002 [c]; Scott et al. 2002 [d]; Legezynska et al. 2012 and references therein [e]).

Fatty acid	Trophic marker
16:0	Carnivory (a, b, c)
16:1(n–7)	Spring bloom (diatoms) (a, b)
16:4(n–1)	Diatoms/ice-algae (a, b)
18:1(n–9)	Carnivory (a, b)
18:2(n–6)	Chlorophytes or cyanobacteria (a, b, c, d, e)
20:1(n–9)	*Calanus* spp. (a, b, c)
20:5(n–3)	Diatoms (a, b)
22:1(n–11)	*Calanus* spp. (a, b)
22:6(n–3)	Flagellates; e.g., presence of *Phaeocystis pouchetii* in the diet of *Calanus* (a, b, c)
	FA ratios
18:1(n–9)/18:1(n–7)	High values (>3) as indicator for increasing carnivory in marine zooplankton (d, e)
20:5(n–3)/22:6(n–3)	High ratio—Diatom-originated diet; low ratio—Flagellate-based diet (c, d, e)
PUFA/SFA	Increasing value may be used as an indicator for dominance of carnivorous vs. herbivorous feeding; however also increases under starvation conditions (c, d, e)

fatty alcohol compositions with only minor seasonal and interspecific differences in food sources of the species investigated (Søreide et al. 2010; Leu et al. 2011; Kraft et al. 2015). The biochemical criteria of the selection of FAs (Table 3) are governed by the overall animal physiology and the limits of the analytical method, where FAs are mostly separated and identified via gas chromatography. The specific FA composition of an individual is characterized by FAs deriving from the diet, de novo synthesis, degradation, and bacterial activities. The key processes of FA physiology are (1) FA synthesis which takes part in the cytosol and is being catalyzed in animals by a very large multiprotein assembly, the FAS system, and (2) the catabolic pathway which takes part in the mitochondria. Here, during ß-oxidation the long-chain FAs undergo a C2-unit breakdown until reaching acetyl coenzyme A, which will be further oxidized in the citric cycle pathway. Therefore a variety of major FAs ranging from C12 to C24 with up to six double bonds will be detected during a usual chromatographic run. An overview of typical FA synthesis pathways of marine zooplankton and phytoplankton organisms is given in Fig. 1. Since animal FA synthesis is not able to introduce a double bond between the ω9-position and the methyl end of the FA, only plants and phytoplankton with their specific desaturases are able to produce polyunsaturated FAs, which are essential for marine animals.

Copepods

Calanus copepods are playing a key role in the pelagic lipid-based Arctic food web (Falk-Petersen et al. 1990) and constitute around 80–90% of the zooplankton biomass in Arctic seas (Sargent and Henderson 1986; Conover and Huntley 1991). Their individual lipid content may be as large as 50–70% of the body weight (Lee 1975; Sargent and Falk-Petersen 1988; Scott et al. 2000) making them a major link between primary producers and higher trophic levels. Typical phytoplankton FAs are major components of Arctic copepods and are incorporated unchanged into their body lipids, for example, 18:4(n–3) FA. This FA, a typical flagellate marker, plays an important role for the life cycle of herbivorous copepods as it appears with the summer phytoplankton bloom. A combination of dietary and membrane or de novo synthesized FAs is most likely and therefore a logratio of 16:0/18:4(n–3), explaining a high percentage of the variance in the data set, seemed to be a good FA ratio selection as a starting point for further iteration of the variable selection process. Similar to 18:4(n–3) for flagellates, the 16:1(n–7) FA represents a biomarker for diatoms and/or ice algae, which normally appear in spring, when the sun is back. The ratio with 16:0 FA as numerator presented as well a contrast of a dietary FA with a FA more related to membrane structures. These two ratios, involving only three FAs, were able to explain almost 75% of the variance in the data set, as well as perfectly predict the three seasons when the samples were taken. The third ratio of the copepod data set represented a contrast of a long chain monounsaturated FA (20:1(n–9)) with an essential long chain PUFA (22:6(n–3)). Despite this ratio 20:1(n–9)/22:6(n–3) only adding a relatively small additional explained variance (6.6%), it was representing a ratio of FAs with highest likelihood in

copepods. Again, an entity of a dietary component 22:6(*n*–3) and a de novo synthesized FA was chosen. However, the PUFA would be most likely absorbed after digestion and used for the building of structural lipids. The selection of the next three ratios of the copepod data set, 16:0/20:1(*n*–9), 14:0/20:5(*n*–3), and 18:0/20:5(*n*–3), was made following the de novo biosynthesis of FAs representing three major end products of the FAS (Fig. 1), 14:0, 16:0, and 18:0 FAs, which are contrasted with the essential FA 20:5(*n*–3) and the long chain FA 20:1(*n*–9), as typical products of the copepod lipid biosynthesis.

Amphipods

A second data set of FAs was chosen to explain statistics and relationships between the FAs in animals of different feeding behavior, changing from herbivores to a mainly carnivorous animal. The pelagic amphipods are important congeners of the Arctic food web supplying lipid-based energy for higher trophic levels (Auel et al. 2002; Kraft et al. 2015). They partly feed on copepods and store *Calanus* derived lipids consisting of essential ω3 and ω6 FAs into storage and membrane lipids. Consequently, these FAs provide energy and building blocks for higher trophic levels (Clarke et al. 1985). The basic consideration for the selection of FA ratios was almost comparable to what was done for the copepod data set. However, it had to be considered that amphipods are of higher trophic position, and therefore their proportional composition of lipids and FAs are characterized by high amounts of typical long-chain FAs from copepods. The ω3 FA 20:5(*n*–3) together with the monounsaturated FA 22:1(*n*–11) represented a typical FA ratio of producers and consumers FAs. It should be noted that 20:5(*n*–3) could be an essential part of the amphipods membrane lipids, but could also derive from the prey and will be incorporated into the storage lipids, that is, triacylglycerol or wax esters. As a second pair of FAs 16:0 and 22:1(*n*–11) FA were chosen, representing a defined ratio of a typical membrane FA in combination with a dietary derived FA. The 18:4(*n*–3) FA is a typical flagellate FATM, which represents in its ratio with 18:0 FA a reasonable biochemically selection with a high FA portion. Although 18:0 FA showed a low mass percentage compared to other FAs, it is an important intermediate in FA synthesis and ß-oxidation, and therefore having a great impact for the formation of major FA end products. Furthermore, arachidonic acid or 20:4(*n*–6) FA is an important essential FA, which most of the animals need as building blocks for their phospholipids and entered as a ratio with 18:0 FA. The next ratio showed a combination of 18:1(*n*–9), a de novo biosynthesized FA and regarded as a carnivore marker, linked with the long-chain monounsaturated FA 20:1(*n*–9), most likely deriving from a copepod diet. The next two ratios were represented by diatom FATM 16:1(*n*–7) together with a membrane FA 22:6(*n*–3) and a combination of two long-chain FAs 20:1(*n*–9) and 22:1(*n*–11). Both of these ratio pairs reflected a possible biosynthesis of typical dietary FAs, deriving from algae or copepods, which are most likely stored in the triacylglycerol or wax ester to enlarge the animal's energy pool. The inclusion of the last ratio 16:1(*n*–7)/20:1(*n*–11),

a contrast of dietary derived matter, brought the percentage of variance explained to 91.6% in the amphipod data set.

In conclusion, for the investigation of an individual's FA composition by LRA, it is recommended using a reasonable number of logratios with highest biochemical impact. These logratios primarily consist of membrane related FAs, that is, 16:0, 20:5(*n*–3), and 22:6(*n*–3) or originating from dietary events and some particular de novo synthesized FA. Having identified these subsets of ratios, the same ratios can be computed for future, as well as past, data sets, as long as the FAs composing the ratios are present in these data sets. The advantage of working with ratios is that they can be validly compared between data sets, irrespective of the number of FAs included in the studies, which can range from as few as 20 to as many as 150 FAs. In fact, the present set of identified ratios could serve as a type of benchmark for comparison with other studies of copepods and amphipods. Keeping these biochemical criteria in mind, supported by clear statistical objectives, the analysis and interpretation of a complex FA data set can be simplified, by reducing the data set to a few logratios of selected FAs thanks to a combination of statistical and biochemical expertise.

References

Aitchison, J. 1986. The statistical analysis of compositional data. Chapman & Hall, Reprinted in 2003 with additional material by Blackburn Press. J. of Royal Stat. Society. **44**, No. 2 (1982), 139–177.

Aitchison, J. 1990. Relative variation diagrams for describing patterns of compositional variability. Math. Geol. **22**: 487–511. doi:10.1007/BF00890330

Aitchison, J., and M. Greenacre. 2002. Biplots of compositional data. J. R. Stat. Soc. Ser. C Appl. Stat. **51**: 375–392. doi:10.1111/1467-9876.00275

Auel, H., M. Harjes, R. d. Rocha, D. Stübing, and W. Hagen. 2002. Lipid biomarkers indicate different ecological niches and trophic relationships of the Arctic hyperiid amphipods *Themisto abyssorum* and *T. libellula*. Polar Biol. **25**: 374–383. doi:10.1007/s00300-001-0354-7

Boissonnot, L., B. Niehoff, W. Hagen, J. E. Søreide, and M. Graeve. 2016. Lipid turnover reflects life-cycle strategies of small-sized Arctic copepods. J. Plankton Res. **38**: 1420–1432. doi:10.1093/plankt/fbw076

Boissonnot, L., B. Niehoff, B. Ehrenfels, J. E. Søreide, W. Hagen, and M. Graeve. 2019. Lipid and fatty acid turnover of the pteropods *Limacina helicina*, *L. retroversa* and *Clione limacina* from Svalbard waters. Mar. Ecol. Prog. Ser. **609**: 133–149. doi:10.3354/meps12837

Breiman, L., J. Friedman, C. J. Stone, and R. A. Olshen. 1984. Classification and regression trees. Wadsworth.

Brenna, J. T., M. Plourde, K. D. Stark, P. J. Jones, and Y.-H. Lin. 2018. Best practices for the design, laboratory analysis, and reporting of trials involving fatty acids. Am. J. Clin. Nutr. **108**: 211–227. doi:10.1093/ajcn/nqy089

Brett, M. T., M. J. Kainz, S. J. Taipale, and H. Seshan. 2009. Phytoplankton, not allochthonous carbon, sustains herbivorous zooplankton production. Proc. Natl. Acad. Sci. USA **106**: 21197–21201. doi:10.1073/pnas.0904129106

Budge, S. M., A. M. Springer, S. J. Iverson, G. Sheffield, and C. Rosa. 2008. Blubber fatty acid composition of bowhead whales, *Balaena mysticetus*: Implications for diet assessment and ecosystem monitoring. J. Exp. Mar. Biol. Ecol. **359**: 40–46. doi:10.1016/j.jembe.2008.02.014

Clarke, A., A. Skadsheim, and L. J. Holmes. 1985. Lipid biochemistry and reproductive biology in two species of Gammaridae (Crustacea: Amphipoda). Mar. Biol. **88**: 247–263. doi:10.1007/BF00392587

Conover, R. J., and M. Huntley. 1991. Copepods in ice-covered seas—distribution, adaptions to seasonally limited food, metabolism, growth patterns and life cycle strategies in polar seas. J. Mar. Syst. **2**: 1–41. doi:10.1016/0924-7963(91)90011-I

Dalsgaard, J., M. A. St. John, G. Kattner, D. Müller-Navarra, and W. Hagen. 2003. Fatty acid trophic marker in the pelagic marine environment. Adv. Mar. Biol. **46**: 225–340. doi:10.1016/S0065-2881(03)46005-7

Falk-Petersen, S., C. C. E. Hopkins, and J. R. Sargent. 1990. Trophic relationships in the pelagic, arctic food web, p. 315–333. *In* R. N. Gibson and M. Barnes [eds.], Trophic relationships in the marine environment: Proceedings of the 24th European Marine Biology Symposium at Aberdeen, Scotland.

Falk-Petersen, S., W. Hagen, G. Kattner, A. Clarke, and J. Sargent. 2000. Lipids, trophic relationships, and biodiversity in Arctic and Antarctic krill. Can. J. Fish. Aquat. Sci. **57**: 178–191. doi:10.1139/f00-194

Falk-Petersen, S., J. R. Sargent, S. Kwasniewski, B. Gulliksen, and R. M. Millar. 2001. Lipids and fatty acids in *Clione limacina* and *Limacina helicina* in Svalbard waters and the Arctic Ocean: Trophic implications. Polar Biol. **24**: 163–170. doi:10.1007/s003000000190

Falk-Petersen, S. (1999). Spatial distribution and life-cycle timing of zooplankton in the marginal ice zone of the Barents Sea during the summer melt season in 1995. Journal of Plankton Research. **21**: 1249–1264. doi:10.1093/plankt/21.7.1249

Falk-Petersen, S., J. R. Sargent, and K. S. Tande. 1987. Lipid composition of zooplankton in relation to the sub-arctic food web. Polar Biology. **8**: 115–120. doi:10.1007/bf00297065

Falk-Petersen, S., T. Haug, H. Hop, K. T. Nilssen, and A. Wold. 2009. Transfer of lipids from plankton to blubber of harp and hooded seals off East Greenland. Deep-Sea Res. Part II Top. Stud. Oceanogr. **56**: 2080–2086. doi:10.1016/j.dsr2.2008.11.020

Folch, J., M. Lees, and G. H. S. Stanley. 1957. A simple method for the purification of total lipids from animal tissues. J. Biol. Chem. **226**: 497–509.

Graeve, M., G. Kattner, and W. Hagen. 1994a. Diet-induced changes in the fatty composition of Arctic herbivorous copepods: Experimental evidence of trophic markers. J. Exp. Mar. Biol. Ecol. **182**: 97–110. doi:10.1016/0022-0981(94)90213-5

Graeve, M., W. Hagen, and G. Kattner. 1994b. Herbivorous or omnivorous? On the significance of lipid compositions as trophic markers in Antarctic copepods. Deep Sea Research Part I: Oceanographic Research Papers. **41**: 915–924. doi:10.1016/0967-0637(94)90083-3

Graeve, M., G. Kattner, and D. Piepenburg. 1997. Lipids in Arctic benthos: does the fatty acid and alcohol composition reflect feeding and trophic interactions? Polar Biology**18**: 53–61. doi:10.1007/s003000050158

Graeve, M., C. Albers, and G. Kattner. 2005. Assimilation and biosynthesis of lipids in Arctic Calanus species based on feeding experiments with a ^{13}C labelled diatom. J. Exp. Mar. Biol. Ecol. **317**: 109–125. doi:10.1016/j.jembe.2004.11.016

Greenacre, M. 2010. Log-ratio analysis is a limiting case of correspondence analysis. Math. Geosci. **42**: 129–134. doi:10.1007/s11004-008-9212-2

Greenacre, M. 2011. Measuring subcompositional incoherence. Math. Geosci. **43**: 681–693. doi:10.1007/s11004-011-9338-5

Greenacre, M. 2013. Contribution biplots. J. Comput. Graph. Stat. **22**: 107–122. doi:10.1080/10618600.2012.702494

Greenacre, M. 2016. Data reporting and visualization in ecology. Polar Biol. **39**: 2189–2205. doi:10.1007/s00300-016-2047-2

Greenacre, M. 2018. Compositional data analysis in practice. Chapman & Hall.

Greenacre, M. 2019. Variable selection in compositional data analysis using pairwise logratios. Math. Geosci. **51**: 649–682. doi:10.1007/s11004-018-9754-x

Greenacre, M., and P. Lewi. 2009. Distributional equivalence and subcompositional coherence in the analysis of compositional data, contingency tables and ratio-scale measurements. J. Classif. **26**: 29–54. doi:10.1007/s00357-009-9027-y

Hastie, T., R. Tibshirani, and J. Friedman. 2009. The elements of statistical learning: Data mining, inference and prediction, 2nd Edition. Springer.

Haug, T., S. Falk-Petersen, M. Greenacre, H. Hop, U. Lindstrøm, S. Meier, K. T. Nilssen, and A. Wold. 2017. Trophic level and fatty acids in harp seals compared with common minke whales in the Barents Sea. Mar. Biol. Res. **13**: 919–932. doi:10.1080/17451000.2017.1313988

Imamura, F., and others. 2017. A combination of plasma phospholipid fatty acids and its association with incidence of type 2 diabetes: The EPIC-InterAct case-cohort study. PLoS Med. **14**: e1002409. doi:10.1371/journal.pmed.1002409

Jolliffe, I. T., J. A. Learmonth, G. J. Pierce, M. B. Santos, N. Trendafilov, A. F. Zuur, E. N. Ieno, and G. M. Smith. 2007. Principal component analysis applied to harbour porpoise fatty acid data, p. 515–528. *In* A. F. Zuur, E. N. Ieno, and G. M. Smith [eds.], Analysing ecological data. Springer.

Kattner, G., and H. S. G. Fricke. 1986. Simple gas-liquid chromatographic method for the simultaneous determination of fatty acids and alcohols in wax esters of marine organisms. J. Chromatogr. **361**: 263–268. doi:10.1016/S0021-9673(01)86914-4

Kraft, A., M. Graeve, D. Janssen, M. Greenacre, and S. Falk-Petersen. 2015. Arctic pelagic amphipods: Lipid dynamics and life strategy. J. Plankton Res. **37**: 790–807. doi:10.1093/plankt/fbv052

Krzanowski, W. J. 1987. Selection of variables to preserve multivariate data structure, using principal components. J. R. Stat. Soc. Ser. C Appl. Stat. **36**: 22–33. doi:10.2307/2347842

Lee, R. F. 1975. Lipids of Arctic Zooplankton. Comp. Biochem. Physiol. **51B**: 263–266.

Lee, R. F., W. Hagen, and G. Kattner. 2006. Lipid storage in marine zooplankton. Mar. Ecol. Prog. Ser. **307**: 273–306. doi:10.3354/meps307273

Legeżyńska, J., M. Kędra, and W. Walkusz. 2012. When season does not matter: summer and winter trophic ecology of Arctic amphipods. Hydrobiologia, **684**: 189–214. doi:10.1007/s10750-011-0982-z

Leonard, A. E., S. L. Pereira, H. Sprecher, and Y.-S. Huang. 2004. Elongation of long-chain fatty acids. Prog. Lipid Res. **43**: 36–54. doi:10.1016/S0163-7827(03)00040-7

Leu, E., J. E. Søreide, D. O. Hessen, S. Falk-Petersen, and J. Berge. 2011. Consequences of changing sea-ice cover for primary and secondary producers in the European Arctic shelf seas: Timing, quantity, and quality. Prog. Oceanogr. **90**: 18–32. doi:10.1016/j.pocean.2011.02.004

McGovern, M., J. Berge, B. Szymczycha, J. M. Węsławski, and P. E. Renaud. 2018. Hyperbenthic food-web structure in an Arctic fjord. Mar. Ecol. Prog. Ser. **603**: 29–46. doi:10.3354/meps12713

Meier, S., S. Falk-Petersen, L. A. Gade-Sørensen, M. Greenacre, T. Haug, and U. Lindstrom. 2016. Fatty acids in common minke whale (*Balaenoptera acutorostrata*) blubber reflect the feeding area and food selection, but also high endogenous metabolism. Mar. Biol. Res. **12**: 221–238. doi:10.1080/17451000.2015.111851

Mocking, R. J. T., J. Assies, A. Lok, H. G. Ruhé, M. W. J. Koeter, C. Visser, C. L. H. Bockting, and A. H. Schene. 2012. Statistical methodological issues in handling of fatty acid data: Percentage or concentration, imputation and indices. Lipids **47**: 541–547. doi:10.1007/s11745-012-3665-2

Mundry, F. R., and C. L. Nunn. 2009. Stepwise model fitting and statistical inference: Turning noise into signal pollution. Am. Nat. **173**: 119–123. doi:10.1086/593303

Neubauer, P., and O. P. Jensen. 2015. Bayesian estimation of predator diet composition from fatty acids and stable isotopes. PeerJ **3**: e920. doi:10.7717/peerj.920

Palarea-Albaladejo, J., J. A. Martín-Fernández, and J. Gómez-García. 2007. A parametric approach for dealing with compositional rounded zeros. Math. Geol. **39**: 625–645. doi:10.1007/s11004-007-9100-1

Parrish, C. C. 1988. Dissolved and particulate marine lipid classes: A review. Mar. Chem. **23**: 17–40. doi:10.1016/0304-4203(88)90020-5

Pawlowsky-Glahn, V., and A. Buccianti. 2011. Compositional data analysis: Theory and applications. Wiley.

Pearson, K. 1897. Mathematical contributions to the theory of evolution. On a form of spurious correlation which may arise when indices are used in the measurement of organs. Proc. R. Soc. Lond. **60**: 489–498. doi:10.1098/rspl.1896.0076

Petersen, S. O., and M. J. Klug. 1994. Effects of sieving, storage, and incubation temperature on the phospholipid fatty acid profile of a soil microbial community. Appl. Environ. Microbiol. **60**: 2421–2430.

Pethybridge, H. R., C. C. Parrish, B. D. Bruce, J. W. Young, and P. D. Nichols. 2014. Lipid, fatty acid and energy density profiles of white sharks: Insights into the feeding ecology and ecophysiology of a complex top predator. PLoS One **9**: e97877. doi:10.1371/journal.pone.0097877

Petursdottir, H., A. Gislason, and S. Falk-Petersen. 2008. Lipid classes and fatty acid compositions of muscle, liver and skull oil in deep-sea redfish *Sebastes mentella* over the Reykjanes Ridge. J. Fish Biol. **73**: 2485–2496. doi:10.1111/j.1095-8649.2008.02100.x

R Core Team (2019). R: A language and environment forstatistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Renaud, P. E., M.Tessmann, A. Evenset, and G. N. Christensen. 2010. Benthic food-web structure of an Arctic fjord (Kongsfjorden, Svalbard). Marine Biology Research, **7**: 13–26. doi:10.1080/17451001003671597

Sargent, J. R., R. R. Gatten, and R. J. Henderson. 1981. Lipid biochemistry of zooplankton from high latitudes. Oceanis **7**: 623–632.

Sargent, J. R., and R. J. Henderson. 1986. Lipids, p. 59–109. *In* E. D. S. Corner and S. C. M. O'Hara [eds.], The biological chemistry of marine copepods. Clarendon Press.

Sargent, J. R., and S. Falk-Petersen. 1988. The lipid biochemistry of calanoid copepods. Hydrobiologia **167–168**: 101–114. doi:10.1007/BF00026297

Sargent, J. R., J. G. Bell, M. V. Bell, R. J. Henderson, and D. R. Tocher. 1993. The metabolism of phospholipids and polyunsaturated fatty acids in fish, p. 103–124. *In* B. Lahlou and P. Vitiello [eds.], Aquaculture: Fundamental and applied research. Coastal and estuarine studies, v. **43**. American Geophysical Union.

Scott, C. L., S. Kwasniewski, S. Falk-Petersen, and J. R. Sargent. 2000. Lipids and life strategies of Calanus finmarchicus , Calanus glacialis and Calanus hyperboreus in late autumn, Kongsfjorden, Svalbard. Polar Biology. **23**: 510–516. doi:10.1007/s003000000114

Scott, C. L., S. Kwasniewski, S. Falk-Petersen, and J. R. Sargent. 2002. Species differences, origins and functions of fatty alcohols and fatty acids in the wax esters and

phospholipids of Calanus hyperboreus, C. glacialis and C. finmarchicus from Arctic waters. Marine Ecology Progress Series. **235**: 127–134. doi:10.3354/meps235127

Søreide, J. E., E. Leu, J. Berge, M. Graeve, and S. Falk-Petersen. 2010. Timing of blooms, algal food quality and *Calanus glacialis* reproduction and growth in a changing Arctic. Glob. Chang. Biol. **16**: 3154–3163. doi:10.1111/j.1365-2486.2010.02175.x

Sprecher, H. 2000. Metabolism of highly unsaturated n-3 and n-6 fatty acids. Biochim. Biophys. Acta **1486**: 219–231. doi: 10.1016/s1388-1981(00)00077-9

Stewart, C. 2017. An approach to measure distance between compositional diet estimates containing essential zeros. J. Appl. Stat. **44**: 1137–1152. doi:10.1080/02664763.2016.1193846

Tartu, S., S. Bourgeon, J. Aars, M. Andersen, D. Ehrich, G. W. Thiemann, J. M. Welker, and H. Routti. 2016. Geographical area and life history traits influence diet in an arctic marine predator. PLoS One **11**: e0155980. doi:10.1371/journal.pone.0155980

Thiemann, G. W., S. J. Iverson, and I. Stirling. 2008. Polar bear diets and arctic marine food webs: Insights from fatty acid analysis. Ecol. Monogr. **78**: 591–613. doi:10.1890/07-1050.1

van den Wollenberg, A. L. 1977. Redundancy analysis - an alternative for canonical analysis. Psychometrika **42**: 207–219. doi:10.1007/BF02294050

Vaskovsky, V. E. 1989. Phospholipids, p. 199–242. *In* R. G. Ackman [ed.], Marine biogenic lipids, fats, and oils, v. **I**. CRC Press.

Werker, A. G., and E. R. Hall. 2000. Assessment of microbial fatty acid analysis for quantifying mixed culture biomass during batch growth. Water Environ. Res. **72**: 484–489. doi:10.2175/106143000X138030

Whittingham, M. J., P. A. Stephens, R. B. Bradbury, and R. P. Freckleton. 2006. Why do we still use stepwise modelling in ecology and behaviour? J. Anim. Ecol. **75**: 1182–1189. doi:10.1111/j.1365-2656.2006.01141.x

Conflict of Interest

None declared.