Metabolomic strategies for aquaculture: A primer

Tim Young and Andrea C. Alfaro
Institute for Applied Ecology New Zealand, Auckland University of Technology, New Zealand

Abstract

Metabolomics is a fast-evolving field that provides qualitative and quantitative analyses of metabolites within cells, tissues or biofluids. Recent applications of metabolomics approaches in aquaculture research have highlighted the huge potential for solving problems within all aspects of the production line, from hatchery production to post-harvest quality control. To assist with the growing application of metabolomics in aquaculture research, this contribution provides a review of techniques and steps necessary to conduct metabolomics research, from experimental design to data interpretation. Specifically, we target scientists who are new to the field of metabolomics, and we offer simple, but comprehensive steps and strategies to conduct this type of research. We conclude this primer with some advice on how to access relevant expertise and facilities for metabolomics-based aquaculture research.

Keywords: Metabolomics · Profiling · Fingerprinting · Aquaculture · Fish · Shellfish · Bivalves · Molluscs · Marine Bioinformatics · Biomarkers · Metabolism · Physiology · NMR · GC-MS · LC-MS · FT-IR · Multivariate data analysis · Chemometrics

Introduction

Within the last decade, the field of metabolomics (the study of metabolites within cells, tissues or biofluids) has expanded, with a number of applications across the life sciences. In aquaculture alone, metabolite patterns have been successfully used to identify and resolve issues related to hatchery production (Young et al. 2015a, 2015b), nutrition and diet (Castro et al. 2015; Cheng et al. 2015), disease and immunology (Liu et al. 2015; Peng et al. 2015), and post-harvest quality control (e.g. Melis 2014; Chen et al. 2015), among others (reviewed by Alfaro & Young 2016). There are some reasons why this approach has been so successful in such a relatively short amount of time. To begin with, metabolomics is an approach that can generate comprehensive datasets of metabolites to describe complex biological systems. Furthermore, the same analytical and computational tools used to generate and interpret data can be performed on any living organism, since metabolites are highly conserved in structure and function across species (in contrast to genes). With recent advances in analytical techniques and computational analysis, complete datasets that describe changes and/or differences in biological systems can be carried out in a rapid and cost-effective manner. However, results stemming from this approach do not necessarily provide mechanistic and/or causal information regarding the patterns observed. In other words, the exploratory nature of this approach is likely to generate new hypotheses, and further targeted experiments may lead to validation of the resultant biological markers. This process allows for unexpected information to be revealed, leading to innovation and discovery in a very efficient manner. Compared to other areas of research, such as agriculture, food science, and medical science, the
application of metabolomics to aquaculture research has only recently been realised. These applications have been reviewed in a companion paper (Alfaro & Young 2016). Thus, we limit the scope of this review to a review of the techniques and steps necessary to conduct metabolomics research in aquaculture. Specifically, the purpose of this review is to provide aquaculture researchers and other aquatic scientists who are new to the field of metabolomics with a simple, but comprehensive, primer on the various strategies that are involved in conducting a metabolomics-based investigation for the first time. This primer summarises information on experimental design, sample collection and preparation, choice of analytical platform, bioinformatics processing, statistical analyses, biological interpretation of the data, and reporting guidelines. We outline several aspects which require careful consideration, specifically for experiments involving aquatic organisms, and we direct readers to a range of specific aquaculture-related research studies to showcase the relevance of these topics.

We conclude this review with some advice on how researchers can access the relevant expertise and facilities for conducting a metabolomics-based project, and we provide some perspectives on the development of future technological strategies for assessing the health and welfare of wild and cultured aquatic organisms.

**Metabolomic strategies**

There are generally six steps involved in a metabolomics study: (i) robust experimental design, (ii) sample collection and preparation, (iii) analytical measurement and data acquisition, (iv) bioinformatics (data integrity checking and metabolite identifications), (v) statistical analyses, and (vi) biological interpretation and/or biomarker validation (Fig. 1). Due to the wide range of fields encompassed by metabolomics studies (biology, biochemistry, analytical chemistry, bioinformatics and statistics), it is highly recommended that consultation with a metabolomics specialist is carried out in the early stages of experimental design.

![General workflow involved in a metabolomics study outlining the six main steps.](Image)
Experimental design & sampling

Along with good standard experimental design practices, there are a number of special considerations to keep in mind when planning a strategy for a metabolomics investigation. There are particular requirements, which mandate that samples be taken only after an experiment has been specifically designed and performed with a metabolomic-based analysis in mind. Samples which have previously been collected and stored for another purpose will unlikely be suitable for incorporation into a metabolomics-based study.

Collected samples must reflect and represent the biology in question, and be appropriate for the particular research questions of the study. It is critical that biological, technical and experimental variability be minimized, since the metabolome can change very rapidly in response to subtle changes in the environment. For example, the metabolic signatures of aquatic organisms can be affected by handling stress and air exposure (Karakach et al. 2009; Connor & Gracey 2012; Young et al. 2015a), so this should be kept to an absolute minimum - even experimentally characterized if possible. The acute stress of transferring fish and crustaceans between culture or storage tanks is reflected in the metabolome and, if not controlled, may influence results of a study (Schock et al. 2013; Mushtaq et al. 2014a). In the case of shellfish, metabolic responses to treatments can be masked when organisms are taken from their natural environments and into the laboratory for a period of acclimatization (Hines et al. 2007). Therefore, sampling and tissue dissections should be performed in situ, when possible.

Furthermore, in the case of time-course experiments, sampling at the same time of day can be important due to inherent effects associated with circadian rhythms (Gooley 2014; Li et al. 2015). Selection of adequate control animals is crucial in all omics-based investigations. In most cases, controls and treatment groups should have the same genetic background and should be matched for gender, age, size-class and/or development stage. For example, male and female mussels from a homogenous population can easily be discriminated based on their metabolite profiles (Cubero-Leon et al. 2012), and have sex-specific physiological responses to environmental stressors, toxin exposures and pathogen infections (Ji et al. 2013; Liu et al. 2014a; Ellis et al. 2014). The metabolome is so sensitive that differences in the age of fish larvae can be detected within samples that are only a few hours apart in developmental stage (Huang et al. 2013), and marine invertebrate larvae of the same age but different size-class can be discriminated based on their metabolite profiles (Young et al. 2015b). Thus, these features should be carefully managed to avoid potential experimental bias, unless they are the specific biological aspect under investigation.

Correct selection of sample material is also important. Different tissues (e.g., muscle, gills, liver, pancreas) undergo specific metabolic processes by virtue of their distinct functional purpose. Recent studies of tissue-specific metabolism in aquatic organisms include digestive gland vs. gill response differences during pathogen infection in mussels, and differences measured under future climate change scenarios in oysters (Liu et al. 2014b; Wei et al. 2015). In the case of biofluids, the serum and plasma components of blood contain significant chemical differences due to the way in which they are prepared (Yin et al. 2015). Thus, prior knowledge of the biological system is favourable in order to assess the suitability of particular tissues or biofluids for a given experiment.

Once the sample type has been decided, protocols for sampling should be developed. While there is limited information on how the speed of sampling affects the metabolite profile, we suggest that samples be taken rapidly and in a highly reproducible manner to minimize biological and technical variation. For example, if liver samples are to be taken from a number of fish, it would be prudent to make sure that the timing and procedures used to immobilise the organisms and to dissect the tissue be very similar between each animal. Application of anaesthetics during this process should be used with caution since they may disturb the metabolic baseline signature (Bando et al. 2010). The highly dynamic state of the metabolome continues in tissues and biological fluids even after they have been extracted from the organism. Therefore, in almost all metabolomics investigations it is vital that metabolic processes within samples be stopped, or quenched, as soon as possible during collection (reviewed by van Gulik et al. 2012). While other options exist, a typical method to quench metabolism in animal tissues involves snap-freezing samples in liquid nitrogen. Special considerations may need to be made for this, especially if sampling in the field. Furthermore,
It is recommended that samples be stored at or below -80°C until metabolite extraction in order to maintain inactivation of enzymatic and chemical processes, which may influence the metabolite profile. Immediate access to appropriate facilities for sampling and storage is essential. The choice of containers in which the samples will be stored also requires attention due to potential introduction of contaminants, such as surfactants and plasticizers, which may cause severe interferences during analysis (Courant et al. 2014). See Álvarez-Sánchez et al. (2010a) for additional information regarding appropriate selection of biological samples and a review of some practical aspects, which require consideration prior to sample preparation.

Techniques for preparing samples for analysis strongly depend on the type of biological material collected, and the analytical platform to be employed. Regardless of the approach, the metabolite extraction process should be rapid and robust, while minimizing the potential for sample degradation and metabolite modification (Allwood et al. 2013). Special considerations may also be required for processing marine samples due to potential interferences from salts within the sample matrices (Keller et al. 2008), or presence of complex polysaccharides in the case of macroalgae (Goulitquer et al. 2012). Approaches are numerous and constant method development by chemists provide an array of options. These range from simple one-step solvent extraction processes to more complicated procedures involving multiple stages and/or organic synthesis reactions (derivatization).

In general, the most commonly applied solvent extraction methods include: 1) extraction of polar and/or non-polar metabolites with a mixture of methanol, water and chloroform, 2) extraction of polar metabolites with methanol alone or in combination with water, and 3) extraction of polar metabolites with perchloric acid. There are many variations as to the solvent ratios which can be used, the temperature of extraction, the extraction duration, and the mechanical techniques used to disrupt tissue samples and lyse cells. Due to the diversity of possible techniques and wealth of excellent information already available in the literature, method particulars regarding sample preparation are outside the scope of this review. However, we have provided a sizeable table (Table I) containing references to primary literature which have an aquatic metabolomics-based focus, and we highlight the various strategies employed by each study, including the extraction technique used. These studies may be useful to readers as guiding exemplars for many of the strategies discussed in this article. For further details on the preparation of biological samples prior to metabolite detection, see Álvarez-Sánchez et al. (2010b). For comprehensive information on platform-specific sample preparation techniques for general biofluids and animal tissues, see Beckonert et al. (2007), Nováková & Vlčková (2009), Liebeke & Bundy (2012), Römisch-Margl et al. (2012), Vuckovi (2012) and Mushtaq et al. (2014b). For sample preparation techniques with a particular focus on fish and marine invertebrates, see Lin et al. (2007), Wu et al. (2008), del Carmen Alvarez et al. (2010), and Fernández-Varela et al. (2015).

**Analytical platforms**

A clear understanding of the analytical platform/s to be used is necessary before starting an experiment. Certain platforms have special requirements and may or may not be able to deliver the desired data/information. For example, to obtain broad metabolite coverage, including low abundance compounds, some procedures may require a tissue sample of only 2 mg wet weight, whereas others may require >100 mg. Unfortunately, there is not yet a single platform which can analyse all metabolites within a sample, and some instruments are better-suited for the analysis of particular metabolite classes than others. Hence, multiple platforms may need to be used depending on the aims and scope of the investigation. The costs associated with employing different analytical platforms vary widely, and access to appropriate facilities for sample analysis may limit the decision making process. Therefore, selection of the most appropriate instrument for a given metabolomics-based study will depend largely on the type of sample material collected, the available sample mass, the accessibility of analytical platforms, the end-goals of the researchers, and the budget of the project.

The most commonly applied, high-throughput and high-resolution platforms to analyse samples in metabolomics studies are nuclear magnetic resonance (NMR) and mass spectrometry (MS). In certain circumstances, lower resolution vibrational spectroscopy can also be used. See Figure 2 for
usage trends of the various platforms employed over the past decade. The selection of which platform to apply for a particular metabolomics study is always a compromise between cost, sensitivity, speed, chemical selectivity, and metabolite coverage (Table II). However, realistically, the choice of platform most-often comes down to the availability of analytical facilities and technical expertise through commercial or academic collaborations.

**Nuclear magnetic resonance**

Nuclear magnetic resonance (NMR) detects the characteristic spin properties of atomic nuclei. When nuclei with particular magnetic attributes are immersed in an external magnetic field, they align themselves with (low energy state) or against (high energy state) that field. Application of very specific radio frequency pulses to the nuclei induces a change in the energy state called a ‘spin flip’ (Savorani et al. 2013). The presence of other nuclei and chemical bonds in the immediate vicinity of a nucleus changes the intensity of the applied magnetic field by a small amount called nuclear shielding. As a result of this shielding, nuclei within a metabolite will absorb energy at slightly different frequencies, known as a chemical shift. The combination of all of these different frequencies produces a characteristic spectrum, or ‘fingerprint’ of the sample (Fig. 3A–D). In addition, more complex interactions of the spins under various pulse conditions can provide rich sets of information about the chemical bonding and composition of a molecule or mixture.

All isotopes that contain an odd number of protons and/or neutrons can theoretically be assessed by NMR approaches. However, if they are not found in biological molecules, or have low NMR sensitivities or low natural abundances, they are not often used for metabolomic studies. $^1$H NMR is frequently applied in metabolomics investigations to probe the molecular arrangements of hydrogen atoms. The $^1$H isotope is highly abundant in nature (>99.98%) and has a very high NMR sensitivity. See Schock et al. (2012) for an applied example of how $^1$H NMR was employed to monitor the health of cobia in response to reduced fishmeal-based protein diets, and to identify differential regulation of metabolism
indicative of thyroid disruption and variations in the composition of gut microflora. $^{13}$C NMR can also be used, but is much less abundant (1.1%) and less sensitive. However, $^{13}$C NMR has special applications in tracer studies to investigate metabolite transformations and metabolic flux (Tikunov et al. 2014).

For example, molecules can be chemically labelled, or enriched, with the $^{13}$C isotope and traced through metabolic processes, such as protein catabolism and lipid synthesis, to investigate the uptake and conversion of nutrients in fish (Conceição et al. 2007; Eckman et al. 2013).

Figure 3. Multi-platform metabolomics-based analysis of fish (Danio rerio) liver samples showing sex-specific differences in spectral fingerprints obtained from three platforms (NMR, CG-MS and LC-MS). $^1$H NMR spectra of non-polar extracts from male (A) and female (B) fish. $^1$H NMR spectra of polar extracts from male (C) and female (D) fish. GC-MS total ion chromatograms of non-polar extracts from male (E) and female (F) fish. LC-MS spectra of non-polar extracts from male (G) and female (H) fish. Numbered peaks represent reliably assigned metabolites after bioinformatics processing. Reprinted with permission from Ong et al. (2009).
Table I. A selection of studies using metabolomics-based approaches with relevance to aquaculture. Although not a complete list of all the available literature, we have provided a broad range of references which we think may be of interest to aquaculture researchers, and may be useful resources as guiding exemplars for the various strategies which can be employed, including: 1) methods for metabolite extraction in diverse organisms, tissues and biofluids; 2) use of different analytical platforms; 3) a range of primary bioinformatics software to process raw spectral data and assign metabolite identities; 4) a variety of databases for matching spectral signatures; 5) various pre-treatment techniques to prepare data for statistical analysis; 6) an array of univariate and multivariate statistical methods to identify sample group differences; and 7) some secondary bioinformatics software to aid interpretation of metabolite profiles within biologically meaningful contexts through use of global a priori knowledge stored in biochemical information databases.

<table>
<thead>
<tr>
<th>Organism</th>
<th>Sample type</th>
<th>Experimental theme</th>
<th>Extraction method</th>
<th>Metabolite compound</th>
<th>Derivationisation method</th>
<th>Analytical Platform</th>
<th>General approach</th>
<th>Metabolites Detected</th>
<th>Data pre-treatment methods applied</th>
<th>Bioinformatics &amp; statistical software used</th>
<th>Statistical analyses &amp; data visualisation</th>
<th>Datasets used ($)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fish</td>
<td>Embryos</td>
<td>Baseline developmental metabolism: Multi-platform metabolomics</td>
<td>MeOH</td>
<td>P</td>
<td>MSTFA</td>
<td>GC-MS, LC-MS</td>
<td>Fingerprinting &amp; profiling</td>
<td>55+</td>
<td>Autoscaled</td>
<td>MZmine, SIMCA-P, MEV</td>
<td>PCA, OPLS-DA, MCA, Kruskal-Wallis test heatmap</td>
<td>NIST library, HMDB</td>
<td>Hudson et al. 2013</td>
</tr>
<tr>
<td>Larvae</td>
<td>Nutritional &amp; Thermal influence on larval physiology &amp; growth</td>
<td>MeOH/H2O</td>
<td>P</td>
<td>-</td>
<td>1H NMR</td>
<td>Fingerprinting &amp; profiling</td>
<td>28</td>
<td>Spectral area normalisation, mean centered</td>
<td>TopSpin, Chenomx NMR Suite, MATLAB, PLS Toolbox</td>
<td>PCA</td>
<td>CRl</td>
<td>Chasen et al. 2014</td>
<td></td>
</tr>
<tr>
<td>Culture conditions &amp; post-harvest storage</td>
<td>HClO4</td>
<td>P</td>
<td>-</td>
<td>1H NMR</td>
<td>Fingerprinting &amp; profiling</td>
<td>1+</td>
<td>Spectral intensity normalisation, mean centered</td>
<td>MevMixC, R</td>
<td>PCA, t-test</td>
<td>Undefined</td>
<td>Picone et al. 2011</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Effects of salmon farming on wild fish populations</td>
<td>HClO4</td>
<td>P</td>
<td>-</td>
<td>1H NMR</td>
<td>Fingerprinting &amp; profiling</td>
<td>23</td>
<td>Undefined</td>
<td>MATLAB</td>
<td>RPCA, PLS-DA</td>
<td>Chemical shift data from the literature</td>
<td>Marklund-Egna et al. 2013</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nutritional history prediction &amp; alternative feeds</td>
<td>C6H12O6</td>
<td>P, NP</td>
<td>DART-MS</td>
<td>Fingerprinting &amp; profiling</td>
<td>59+</td>
<td>Spectral area normalisation, log transformation, Pareto scaling</td>
<td>MassCenter, SIMCA, Excel</td>
<td>PCA, OPLS-DA</td>
<td>Undefined</td>
<td>Caka et al. 2013</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monitoring compositional changes in fillets during post-harvest cold storage</td>
<td>TCA</td>
<td>P</td>
<td>-</td>
<td>1H, 13C NMR</td>
<td>Fingerprinting</td>
<td>51</td>
<td>Untreated</td>
<td>TopSpin, Multivac</td>
<td>N/A</td>
<td>Chemical shift data from the literature, HMDB, BMRR, YMDB, ECOMDB</td>
<td>Shumilina et al. 2015</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enhancing disease resistance via simple metabolic modulation</td>
<td>MeOH</td>
<td>P</td>
<td>MSTFA</td>
<td>GC-MS</td>
<td>Profiling</td>
<td>60</td>
<td>Peak height normalisation, median centered, quartile range scaled, log transformed</td>
<td>AMDIS, R, SIMCA-P, SPSS, Prism, MetaboAnalyst (MetPA)</td>
<td>HCA, heatmap, PCA, ICA, MPEA</td>
<td>KEGG, NIB library</td>
<td>Peng et al. 2015</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enhancing disease resistance via simple metabolic modulation</td>
<td>MeOH</td>
<td>P</td>
<td>MSTFA</td>
<td>GC-MS</td>
<td>Profiling</td>
<td>60</td>
<td>Peak height normalisation, median centered, quartile range scaled, log transformed</td>
<td>AMDIS, R, SIMCA-P, SPSS, Prism</td>
<td>PCA, OPLS-DA, heatmap</td>
<td>KEGG, NIB library</td>
<td>Zheng et al. 2015</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enhancing disease resistance via simple metabolic modulation</td>
<td>MeOH</td>
<td>P</td>
<td>MSTFA</td>
<td>GC-MS</td>
<td>Profiling</td>
<td>58</td>
<td>Peak height normalisation, median centered, quartile range scaled, log transformed</td>
<td>AMDIS, R, SIMCA-P, SPSS, Prism, MetaboAnalyst (MetPA)</td>
<td>PCA, OPLS-DA, MFEA, HCA, heatmap</td>
<td>KEGG, NIB library</td>
<td>Ma et al. 2014</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nutrition &amp; alternative food development</td>
<td>MeOH/H2O/CHCl3</td>
<td>P</td>
<td>-</td>
<td>1H NMR</td>
<td>Fingerprinting &amp; profiling</td>
<td>12+</td>
<td>Centered (undefined), Pareto scaled</td>
<td>TopSpin, AMIX, SIMCA-P</td>
<td>PCA, OPLS-DA, ANOVA</td>
<td>CRl, BMRR, HMDB</td>
<td>Abro et al. 2014</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enhancing disease resistance via simple metabolic modulation</td>
<td>MeOH/H2O/CHCl3</td>
<td>P</td>
<td>-</td>
<td>1H NMR</td>
<td>Fingerprinting &amp; profiling</td>
<td>64+</td>
<td>Peak height normalisation, median centered, quartile range scaled, log transformed</td>
<td>AMDIS, R, Markerlex, SIMCA-P, SPSS, Prism, MetaboAnalyst (MetPA)</td>
<td>PCA, ICA, MFEA, HCA, heatmap, KEGG, NIB library</td>
<td>Zhao et al. 2015</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optimisation of extraction methods</td>
<td>MeOH/H2O/CHCl3 (various protocols)</td>
<td>P, NP</td>
<td>-</td>
<td>1H NMR, FT-ICRMS</td>
<td>Fingerprinting &amp; profiling</td>
<td>8+</td>
<td>Bin or spectral area normalisation, glog transformation, mean centered</td>
<td>TopSpin, Chenomx NMR Suite, PLS Toolbox, MATLAB</td>
<td>PCA, ANOVA</td>
<td>Chemical shift data from the literature, CRl</td>
<td>Wu et al. 2008</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Utilisation of dietary protein: Growth-metabolic interactions</td>
<td>HClO4</td>
<td>P</td>
<td>ECF</td>
<td>GC-MS</td>
<td>Profiling</td>
<td>12+</td>
<td>Peak height normalisation</td>
<td>SPSS, SIMCA-P</td>
<td>PLS-DA, t-test</td>
<td>NIST library, in-house library</td>
<td>Jin et al. 2005</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health biomarkers &amp; stress evaluation: Multi-platform, large n features</td>
<td>MeOH</td>
<td>P</td>
<td>LC-MS, LC-MS/MS, FI-MS/MS</td>
<td>Profiling</td>
<td>95+</td>
<td>Mean centered, Pareto scaled</td>
<td>MetaboAnalyst</td>
<td>PCA, PLS-DA, Mann-Whitney U test</td>
<td>N/A</td>
<td>Benks et al. 2013</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health biomarkers: Tumor diagnostics</td>
<td>MeOH/NH4O, MeOH/NH4O/CHCl3</td>
<td>P</td>
<td>-</td>
<td>FT-ICRMS</td>
<td>Fingerprinting</td>
<td>4+</td>
<td>Bin area normalisation, glog transformation, mean centered</td>
<td>MIDAS, MCSA, MATLAB, PLS Toolbox</td>
<td>PCA, PLS-R, t-test</td>
<td>N/A</td>
<td>Stentiford et al. 2005</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kidney</td>
<td>Symptoms of anaemia &amp; health biomarker identification</td>
<td>MeOH/H2O/CHCl3</td>
<td>P</td>
<td>-</td>
<td>1H NMR</td>
<td>Fingerprinting &amp; profiling</td>
<td>37+</td>
<td>Spectral area normalisation, median centered, Pareto scaled</td>
<td>TopSpin, Multivac, Chenomx NMR Suite, AMIX, SPSS</td>
<td>PCA, t-test, Mann-Whitney U test</td>
<td>HMDB, BMRL</td>
<td>Allen et al. 2015</td>
<td></td>
</tr>
<tr>
<td>Intestine</td>
<td>Feed additives to enhance growth &amp; metabolism</td>
<td>MeCN/CHCl3</td>
<td>P</td>
<td>-</td>
<td>LC-MS</td>
<td>Profiling</td>
<td>78+</td>
<td>Biomass normalised, autoscaled</td>
<td>Chemstation, EasyCIMS, SPSS, Metaboanalyst</td>
<td>PCA, HCA, heatmap, ANOVA, t-test, Undefined</td>
<td>Robles et al. 2013</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organism</td>
<td>Sample type</td>
<td>Experimental theme</td>
<td>Extraction method</td>
<td>Metabolite component</td>
<td>Derivation method(s)</td>
<td>Analytical platform(s)</td>
<td>General approach</td>
<td>Metabolites detected</td>
<td>Data pre-treatment methods applied</td>
<td>Bioinformatics &amp; statistical software used</td>
<td>Statistical analyses &amp; data visualisation</td>
<td>Databases used</td>
<td>Reference</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
<td>--------------------</td>
<td>-------------------</td>
<td>---------------------</td>
<td>----------------------</td>
<td>----------------------</td>
<td>-----------------</td>
<td>-------------------</td>
<td>---------------------</td>
<td>-----------------------------</td>
<td>-----------------------------</td>
<td>-------------------</td>
<td>----------</td>
</tr>
<tr>
<td>Humoral fluid</td>
<td>Mechanism of vaccine action against disease</td>
<td></td>
<td>H₂O</td>
<td>P</td>
<td>MISTFA</td>
<td>GC-MS</td>
<td>Profiling</td>
<td>65</td>
<td>Median centered, interquartile range scaled, Pareto scaled or autoscaled</td>
<td>XCalibur, NIST MS search, SPSS, SIMCA, MetaboAnalyst</td>
<td>OPS-L-DA, Heatmap, Mann-Whitney U test, MPEA</td>
<td>NIST library, KEGG</td>
<td>Guo et al. 2015</td>
</tr>
<tr>
<td>Haemolymph: Plasma (p) or serum (s)</td>
<td>Toxological biomarkers &amp; environmental monitoring</td>
<td>None required (p), MeOH/CHCl₃ for lipid fraction (p)</td>
<td>N, P, NP</td>
<td>-</td>
<td>¹H NMR</td>
<td>Fingerprinting (same limited profiling)</td>
<td>7+</td>
<td>Bin integral normalisation, mean centered, Pareto scaled</td>
<td>NMR Processor, VNRN, SIMCA-P</td>
<td>PCA, PL5-DA</td>
<td>Chemical shift data from the literature</td>
<td>Sammuelson et al. 2006</td>
<td></td>
</tr>
<tr>
<td>Metabolic effects of food deprivation</td>
<td>None required (p), MeOH/CHCl₃ for lipid fraction (p)</td>
<td></td>
<td>N, P, NP</td>
<td>-</td>
<td>¹H, ¹³C NMR</td>
<td>Fingerprinting (same limited profiling)</td>
<td>4+</td>
<td>Bin integral normalisation, Pareto scaled</td>
<td>NMR Processor, SIMCA-P</td>
<td>PCA, PL5-DA</td>
<td>Chemical shift data from the literature, HMDB</td>
<td>Kullgren et al. 2010</td>
<td></td>
</tr>
<tr>
<td>Utilisation of dietary protein: Growth-metabolic interactions</td>
<td>None required (p)</td>
<td></td>
<td>P</td>
<td>BSTFA</td>
<td>GC-MS</td>
<td>Profiling</td>
<td>16+</td>
<td>Peak height normalisation</td>
<td>SPSS, SIMCA-P</td>
<td>PL5-DA, t-test</td>
<td>NIST library, in-house library</td>
<td>Jin et al. 2005</td>
<td></td>
</tr>
<tr>
<td>Spawning-induced mappinette &amp; stress: High resolution platform, large n features</td>
<td>MeCN (s)</td>
<td></td>
<td>P</td>
<td>MISTFA</td>
<td>20 GCGC-MS</td>
<td>Profiling</td>
<td>137</td>
<td>Leg transformed, autoscaled, quartile range filtering, KNN (missing variables)</td>
<td>ChromoTOF, MetPP, MetaboAnalyst</td>
<td>PCA, PL5-DA, t-test, MPEA</td>
<td>In-house MS library, KEGG</td>
<td>Cipriano et al. 2015</td>
<td></td>
</tr>
<tr>
<td>Health, nutrition &amp; alternative food development</td>
<td>None required (s)</td>
<td></td>
<td>N, P, NP</td>
<td>-</td>
<td>¹H, ¹³C NMR</td>
<td>Fingerprinting &amp; profiling</td>
<td>34</td>
<td>Spectral area normalisation, mean centered, Pareto scaled</td>
<td>Chenomx NMR Suite, SIMCA-P, Statistica</td>
<td>PCA, PL5-DA, ANOVA</td>
<td>Undefined</td>
<td>Andersen et al. 2014</td>
<td></td>
</tr>
<tr>
<td>Primary cell culture: Cells (c), media (m)</td>
<td>Effects of plant derived contaminants in fish feeds: A multiplatform study (c)</td>
<td>MeOH/H₂O/CHCl₃</td>
<td>N, P, NP</td>
<td>-</td>
<td>FT-ICR-MS, ¹H NMR</td>
<td>Fingerprinting &amp; lipid profiling</td>
<td>40+</td>
<td>Probabilistic quotient normalisation, glog transformation, mean centered</td>
<td>ProMetab, MATLAB, PLS Toolbox, Mi-Pack</td>
<td>PCA, PL5-DA, ANOVA</td>
<td>KEGG</td>
<td>Saffold et al. 2014</td>
<td></td>
</tr>
<tr>
<td>Nutritional supplementation &amp; diet optimisation (m)</td>
<td>None required</td>
<td></td>
<td>N, P, NP</td>
<td>-</td>
<td>H NMR</td>
<td>Fingerprinting &amp; profiling</td>
<td>17</td>
<td>Pareto scaled</td>
<td>Chenomx NMR Suite, SIMCA-P, Statistica</td>
<td>PCA, OPLS-DA, ANOVA</td>
<td>Undefined</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Fin issue</td>
<td>Identifying animal providence</td>
<td>None required</td>
<td>N, P, NP</td>
<td>-</td>
<td>FT-IR</td>
<td>Fingerprinting</td>
<td>N/A</td>
<td>Undefined</td>
<td>Undefined</td>
<td>PCA</td>
<td>N/A</td>
<td>Nurtulla et al. 2015</td>
<td></td>
</tr>
<tr>
<td>Skin mucus</td>
<td>Minimally-invasive sampling &amp; exost: Large n features</td>
<td>MeOH/H₂O</td>
<td>N, P, NP</td>
<td>-</td>
<td>LC-MS/MS</td>
<td>Profiling</td>
<td>204</td>
<td>Peak height normalisation, autoscaled</td>
<td>SIEVE, SIMCA-P, Excel, Syntat</td>
<td>PCA, PL5-DA, ANOVA, t-test</td>
<td>HMBD, Metlin, LipidMaps</td>
<td>Ekman et al. 2015</td>
<td></td>
</tr>
<tr>
<td>Fish oil capsules</td>
<td>None required</td>
<td></td>
<td>NP</td>
<td>-</td>
<td>¹H NMR</td>
<td>Fingerprinting</td>
<td>N/A</td>
<td>Peak maximum normalisation</td>
<td>Undefined</td>
<td>PCA, KNNA, STM, PRN, GRNN</td>
<td>Chemical shift data from the literature</td>
<td>Aursand et al. 2007</td>
<td></td>
</tr>
<tr>
<td>Canned fish</td>
<td>Food authentication, forensics &amp; quality control</td>
<td>None required</td>
<td>N, P, NP</td>
<td>-</td>
<td>FT-IR</td>
<td>Fingerprinting</td>
<td>N/A</td>
<td>Undefined</td>
<td>MATLAB, PLS Toolbox</td>
<td>PCA, PL5-DA</td>
<td>N/A</td>
<td>Dominguez-Vidal et al. 2016</td>
<td></td>
</tr>
<tr>
<td>Packing oil</td>
<td>Food authentication, forensics &amp; quality control</td>
<td>None required</td>
<td>N, P, NP</td>
<td>-</td>
<td>FT-IR</td>
<td>Fingerprinting</td>
<td>N/A</td>
<td>Undefined</td>
<td>MATLAB, PLS Toolbox</td>
<td>PCA, PL5-DA</td>
<td>N/A</td>
<td>Domanz-Balcazar et al. 2016</td>
<td></td>
</tr>
<tr>
<td>Molluscs</td>
<td>Larvae</td>
<td>Identification of larval quality biomarkers during batchy culture</td>
<td>MeOH/H₂O</td>
<td>P</td>
<td>MCF</td>
<td>GC-MS</td>
<td>Profiling</td>
<td>29</td>
<td>Peak height normalisation, metabolite ratios, log transformation, autoscaled</td>
<td>R, AMOS, MetaboAnalyst</td>
<td>PCA, PL5-DA, OPLS-DA, ANOVA, in-house MS library</td>
<td>Young et al. 2015b</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P纬</td>
<td>Seaweed stress &amp; culture conditions</td>
<td>MeOH/H₂O</td>
<td>P</td>
<td>MCF</td>
<td>GC-MS</td>
<td>Profiling</td>
<td>27</td>
<td>Peak height normalisation, autoscaled</td>
<td>R, AMOS, MetaboAnalyst, SPSS</td>
<td>PCA, PL5-DA, HCA, heatmap, volcano plot, EBAM, SAM</td>
<td>In-house MS library</td>
<td>Young et al. 2015a</td>
</tr>
<tr>
<td>Adductor muscle</td>
<td>Organ function &amp; physiology: A multi organ study</td>
<td>HCl₂O</td>
<td>P</td>
<td>-</td>
<td>¹H NMR</td>
<td>Fingerprinting &amp; profiling</td>
<td>37+</td>
<td>Peak area normalisation</td>
<td>NMR Processor, SpinWorks</td>
<td>N/A</td>
<td>HMBD</td>
<td>Tikunov et al. 2010</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Optimization of extraction methods &amp; animal providence: A multi organ study</td>
<td>HCl₂O, MeCN, Ringer's solution</td>
<td>P</td>
<td>-</td>
<td>¹H NMR</td>
<td>Fingerprinting &amp; profiling</td>
<td>32</td>
<td>Bin integral normalisation, Pareto scaled</td>
<td>TopSpin, NMR Processor, Chenomx NMR Suite, JMP, Excel</td>
<td>PCA, t-test</td>
<td>HMBD, BMRR</td>
<td>Hurley-Sanders et al. 2015a</td>
<td></td>
</tr>
<tr>
<td>Mantle</td>
<td>Sex discrimination</td>
<td>MeOH/H₂O/CHCl₃</td>
<td>P</td>
<td>-</td>
<td>¹H NMR</td>
<td>Fingerprinting &amp; profiling</td>
<td>16+</td>
<td>Spectral area normalisation, biomass normalisation, glog transformation, mean centered</td>
<td>TopSpin, MATLAB, Chenomx NMR Suite, Excel</td>
<td>PCA, LDA, t-test</td>
<td>CR</td>
<td>Hines et al. 2007</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ocean acidification, disease, thermal stress &amp; sex differences</td>
<td>MeOH/H₂O/CHCl₃</td>
<td>P</td>
<td>-</td>
<td>¹H NMR</td>
<td>Fingerprinting &amp; profiling</td>
<td>25</td>
<td>Probabilistic quotient normalisation, glog transformation</td>
<td>MATLAB, PRIMER</td>
<td>UNIFIED, PERMANOVA, MDS, SIMPER</td>
<td>Undefined</td>
<td>Ellis et al. 2014</td>
<td></td>
</tr>
<tr>
<td>Hepatopancreas</td>
<td>Health &amp; immunology: Most responses to bacterial pathogens</td>
<td>MeOH/H₂O/CHCl₃</td>
<td>P</td>
<td>-</td>
<td>¹H NMR</td>
<td>Fingerprinting &amp; profiling</td>
<td>27</td>
<td>Spectral area normalisation, biomass normalisation, glog transformation, mean centered</td>
<td>TopSpin, Chenomx NMR Suite, MATLAB, PLS Toolbox, Mi-Pack</td>
<td>PCA, PL5-DA, OPLS-DA, ANOVA</td>
<td>Chemical shift data from the literature</td>
<td>Wu et al. 2013</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ocean acidification: Integrated metabolomics &amp; proteomics</td>
<td>MeOH/H₂O/CHCl₃</td>
<td>P</td>
<td>-</td>
<td>¹H NMR</td>
<td>Fingerprinting &amp; profiling</td>
<td>32</td>
<td>Spectral area normalisation, glog transformation, mean centered</td>
<td>TopSpin, Chenomx NMR Suite, MATLAB, SIMCA-P</td>
<td>PL5-DA, OPLS-DA</td>
<td>CR</td>
<td>Wei et al. 2015</td>
<td></td>
</tr>
<tr>
<td>Foot</td>
<td>Health Biomarkers for toxicity, hypoxia &amp; food limitation</td>
<td>MeCN/H₂O</td>
<td>P</td>
<td>-</td>
<td>¹H NMR</td>
<td>Fingerprinting &amp; profiling</td>
<td>20</td>
<td>Box-Cox transformation</td>
<td>SpeckManager, Excel, Metilab, GenStat</td>
<td>PCA, LDA, ANOVA</td>
<td>HMBD</td>
<td>Tufshel et al. 2009</td>
<td></td>
</tr>
<tr>
<td>Organism</td>
<td>Sample type</td>
<td>Experimental theme</td>
<td>Extraction method</td>
<td>Metabolite component</td>
<td>Derivation method</td>
<td>Analytical Platform</td>
<td>General approach</td>
<td>Metabolites Detected</td>
<td>Data pre-treatment methods applied</td>
<td>Bioinformatics &amp; statistical software used</td>
<td>Statistical analyses &amp; data visualisation</td>
<td>Databases used</td>
<td>Reference</td>
</tr>
<tr>
<td>----------</td>
<td>-------------</td>
<td>--------------------</td>
<td>------------------</td>
<td>---------------------</td>
<td>-------------------</td>
<td>-------------------</td>
<td>-----------------</td>
<td>------------------</td>
<td>----------------------</td>
<td>-------------------------------------</td>
<td>--------------------------------------</td>
<td>---------------</td>
<td>-----------</td>
</tr>
<tr>
<td><strong>Molluscs (cont.)</strong></td>
<td>Gills</td>
<td>Identification of thermal stress biomarkers</td>
<td>MoOH₂/H₂O</td>
<td>P</td>
<td>Undefined</td>
<td>GC-MS</td>
<td>Profiling</td>
<td>52</td>
<td>Biomass normalisation</td>
<td>JMP</td>
<td>DIA</td>
<td>Undefined</td>
<td>Dysph et al. 2015</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Toxocology: integrated metabolomics &amp; proteomics</td>
<td>MoOH₂/H₂O/CHCl₃</td>
<td>P</td>
<td>-</td>
<td>H NMR</td>
<td>Fingerprinting &amp; profiling</td>
<td>28</td>
<td>Spectral area normalisation, glog transformation, mean centered</td>
<td>TopSpin, Chenomx NMR Suite, MATLAB, SIMCA-P</td>
<td>PLS-DA, OPLS-DA</td>
<td>Chemical shift data from the literature</td>
<td>Ji et al. 2013</td>
</tr>
<tr>
<td></td>
<td></td>
<td>High resolution NMR: Coastal marine pollution &amp; toxocology integrated metabolomics &amp; proteomics</td>
<td>MoOH₂/H₂O/CHCl₃</td>
<td>P</td>
<td>-</td>
<td>H NMR</td>
<td>Fingerprinting &amp; profiling</td>
<td>27+</td>
<td>Spectral area normalisation, glog transformation, mean centered</td>
<td>XuXin-NMR, Chenomx NMR Suite, MATLAB, Uncrambler X, Excel</td>
<td>PCA, t-test</td>
<td>Chemical shift data from the literature</td>
<td>Cappello et al. 2013</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Central nervous system &amp; glands</td>
<td>Baseline molecular phenotyping: Multiprofile metabolomics</td>
<td>MoOH₂/H₂O/CHCl₃</td>
<td>P, NP</td>
<td>BF-MoOH (lipids)</td>
<td>GC-MS, LC-MS, RPLC &amp; HILIC</td>
<td>Fingerprinting &amp; profiling</td>
<td>73+</td>
<td>Selected peak normalisation, total protein content normalised</td>
<td>DataAnalysis, ProfilingAnalysis, SPSS, Excel, Metamapp, Cytoscape</td>
<td>Ora, BNIM, t-test</td>
<td>Reactome, DMM, KEGG, WikiPathways, Tuft et al. 2015b</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Gastrointestinal tract and/or digestive gland</td>
<td>Biodictator species for pollution monitoring</td>
<td>MoOH₂/H₂O/CHCl₃</td>
<td>P</td>
<td>-</td>
<td>H NMR</td>
<td>Fingerprinting &amp; profiling</td>
<td>19+</td>
<td>Spectral area normalisation, glog transformation, mean centered</td>
<td>MATLAB, PLS Toolbox</td>
<td>ANOVA, PCA, PLS-DA</td>
<td>Chemical shift data from the literature</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Metabolite effects of food deprivation &amp; extraction method optimisation</td>
<td>MoOH₂/H₂O/CHCl₃</td>
<td>P</td>
<td>-</td>
<td>H NMR</td>
<td>Fingerprinting &amp; profiling</td>
<td>28</td>
<td>Log transformation, median centered</td>
<td>Chenomx NMR Suite, TopSpin, R, Uncrambler</td>
<td>PLS-DA, Mann-Whitney U</td>
<td>Chemical shift data from the literature</td>
<td>Shedy et al. 2015</td>
</tr>
<tr>
<td><strong>Whole soft tissue</strong></td>
<td>Dual platform metabolomics: Toxocology mechanistic</td>
<td>MoOH₂/H₂O/CHCl₃</td>
<td>P</td>
<td>MSTFA</td>
<td>GC-MS, H NMR</td>
<td>Fingerprinting &amp; profiling</td>
<td>NMR, 1H-13C-MS, 24</td>
<td>Spectral area normalisation, autoscaled (GC-MS), mean centered &amp; Pareto scaled (NMR)</td>
<td>SpecManager, SIMCA-P</td>
<td>PCA, PLS-DA</td>
<td>Chemical shift data from the literature</td>
<td>Spann et al. 2011</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Toxocology mechanistic</td>
<td>MoOH₂/H₂O/CHCl₃</td>
<td>P</td>
<td>-</td>
<td>H NMR</td>
<td>Fingerprinting &amp; profiling</td>
<td>25+</td>
<td>Bin integral normalisation, glog transformation, mean centered</td>
<td>TopSpin, MATLAB, PLS Toolbox</td>
<td>PCA, PLS-DA, ANOVA</td>
<td>Chemical shift data from the literature</td>
<td>Wu &amp; Wang 2010</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Unique extraction &amp; platform: Method assessment</td>
<td>SPE</td>
<td>P, NP</td>
<td>-</td>
<td>2D GCxGC-MS</td>
<td>Fingerprinting &amp; profiling</td>
<td>63+</td>
<td>Fourth root transformation</td>
<td>ChromaTOF, PRIME</td>
<td>PERMANOVAN, PCA, HCA</td>
<td>Unscrambler X</td>
<td>Rocha et al. 2003</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Haemolymph: whole blood (wb) or plasma (p)</td>
<td>Mechanical shaking &amp; salinity stress</td>
<td>None required (wb)</td>
<td>P, NP</td>
<td>-</td>
<td>FT-IR</td>
<td>Fingerprinting</td>
<td>N/A</td>
<td>Undefined</td>
<td>Undefined</td>
<td>MANOVA, PCA, CVA</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Developing strategies for identifying stress</td>
<td>None required (wb)</td>
<td>P, NP</td>
<td>-</td>
<td>FT-IR</td>
<td>Fingerprinting</td>
<td>N/A</td>
<td>Undefined</td>
<td>Undefined</td>
<td>MANOVA, PCA, CVA</td>
<td>N/A</td>
<td>Goldman et al. 2007</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Health/stress biomarkers: Toxocology</td>
<td>None required (p)</td>
<td>P, NP</td>
<td>-</td>
<td>H NMR</td>
<td>Fingerprinting &amp; profiling</td>
<td>18</td>
<td>Spectral area normalisation, glog transformation, mean centered</td>
<td>MATLAB, PLS Toolbox, Chenomx NMR Suite</td>
<td>ANOVA, PCA</td>
<td>Chemical shift data from the literature</td>
<td>Zhou et al. 2015</td>
</tr>
<tr>
<td><strong>Echinoderms</strong></td>
<td>Muscle</td>
<td>Thermal stress responses</td>
<td>MoOH₂/H₂O/CHCl₃</td>
<td>P</td>
<td>-</td>
<td>H NMR</td>
<td>Fingerprinting &amp; profiling</td>
<td>31</td>
<td>Spectral area normalisation, glog transformation, mean centered</td>
<td>TopSpin, Chenomx NMR Suite, MATLAB, SIMCA-P</td>
<td>PCA, OPLS-DA</td>
<td>Chemical shift data from the literature</td>
<td>Shao et al. 2015</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Macrolegae</td>
<td>Thallass</td>
<td>Effects of food processing on nutrient composition</td>
<td>MoCN₂/H₂O</td>
<td>P</td>
<td>-</td>
<td>H NMR</td>
<td>Fingerprinting &amp; profiling</td>
<td>32+</td>
<td>Bin integral normalisation</td>
<td>AMIX, SIMCA-P, MATLAB, SPSS</td>
<td>PCA, OPLS-DA, heatmap, t-test</td>
</tr>
<tr>
<td></td>
<td></td>
<td>New Insights into metabolism</td>
<td>MoOH₂/H₂O</td>
<td>P</td>
<td>-</td>
<td>H NMR</td>
<td>Fingerprinting &amp; profiling</td>
<td>27</td>
<td>N/A</td>
<td>N/A</td>
<td>Chenomx NMR Suite</td>
<td>N/A</td>
<td>PRIMA DB, HMDB, BMRB</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Stipe &amp; blades</td>
<td>Seasonal variations in metabolism: Multiprofile metabolomics</td>
<td>None required (FT-IR), MoOH₂/H₂O (NMR)</td>
<td>P, NP</td>
<td>-</td>
<td>FT-IR, H-13C-MR</td>
<td>Fingerprinting &amp; profiling</td>
<td>51</td>
<td>Specific peak intensity normalisation (FT-IR), spectral area normalisation (NMR)</td>
<td>Excel, TopSpin, SpecAssing, OXWIN, R, Amos, Gephi, Fityk</td>
<td>PCA, ICA, SOMs, CAN, SEM, MCR-ALS</td>
<td>PRIMEa DB</td>
</tr>
<tr>
<td>Organism</td>
<td>Sample type</td>
<td>Experimental theme</td>
<td>Extraction method*</td>
<td>Metabolite components†</td>
<td>Derivation method†</td>
<td>Analytical Platform/‡</td>
<td>General approach</td>
<td>Metabolites Detected</td>
<td>Data pre-treatment methods applied\</td>
<td>Bioinformatics &amp; statistical software used**</td>
<td>Statistical analyses &amp; data visualisation†</td>
<td>Databases used††</td>
<td>Reference</td>
</tr>
<tr>
<td>-------------------</td>
<td>-------------</td>
<td>--------------------------------------------------------</td>
<td>--------------------</td>
<td>------------------------</td>
<td>-------------------</td>
<td>----------------------</td>
<td>--------------------------------------</td>
<td>----------------------</td>
<td>-------------------------------------</td>
<td>---------------------------------------------</td>
<td>--------------------------------------------</td>
<td>-----------------</td>
<td>----------------------------</td>
</tr>
<tr>
<td>Microalgae</td>
<td>Cells extracts</td>
<td>Screening microalgae to identify commercially useful mutants</td>
<td>None required (FT-IR), MeOH/H2O/CHC1 (LC-MS)</td>
<td>P, NP</td>
<td>FT-IR, LC-MS</td>
<td>Fingerprinting &amp; lipid profiling</td>
<td>11+ lipid classes</td>
<td>Biomass normalised</td>
<td>MATLAB, Xcalibur, Unscrambler</td>
<td>PCA, PC-DFA, PLS-DA, PLS-R</td>
<td>MMD, HMBD, KEGG, BioCyc, LIPIDMAPS, Drugbank</td>
<td>Bajrhyia et al. 2016</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Profiling diatoms for bioenergy &amp; feedstock</td>
<td>MTBE/MeOH/H2O</td>
<td>P, NP</td>
<td>MISTFA</td>
<td>GC-MS, LC-MS</td>
<td>Profiling</td>
<td>96+</td>
<td>Cell density normalised, median scaled, log transformed</td>
<td>Expressionist Refiner MS, ChromTOF, R</td>
<td>PCA, HCA, heatmap, t-test</td>
<td>In-house MS library</td>
<td>Brockme et al. 2015</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dual platform metabolomics: Natural products research</td>
<td>MeO/H2O/CHCl (GC-MS), MeOH/MeO (GC-MS)</td>
<td>P</td>
<td>MISTFA</td>
<td>GC-MS, LC-MS</td>
<td>Profiling</td>
<td>128</td>
<td>Peak area &amp; cell density normalisation</td>
<td>AMIDOS, SIMCA-P, MeV</td>
<td>PLS-DA, HCA, heatmap, MPEA, WGNCN</td>
<td>KEGG</td>
<td>Yu et al. 2015</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Photobiontasso culture conditions &amp; bioresource development</td>
<td>None required (FT-IR), MeOH &amp; EtOH/H2O (LC-MS)</td>
<td>P, NP</td>
<td>-</td>
<td>-</td>
<td>Fingerprinting &amp; profiling</td>
<td>13+</td>
<td>Second derivative calculation (FR-IR), log transformed, Pareto scaled</td>
<td>OpusLab, KMSL, Xcalibur, R, Spectrum Database, Statistical Data Miner, MetabolAnalyst, SIMCA-P</td>
<td>Kruskal-Wallis test, HCA, Mann-Whitney U test, heatmap, PLS-DA</td>
<td>In-house MS library</td>
<td>Courian et al. 2013</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Chemical interactions between bacteria &amp; diatoms</td>
<td>MeOH/ETOH/CHCl (GC-MS)</td>
<td>P</td>
<td>MISTFA</td>
<td>GC-MS, LC-MS</td>
<td>Profiling</td>
<td>19+</td>
<td>Cell density normalised</td>
<td>Metavlyn, AMIDOS, MET-IDEA, Sigmaplot, Excel</td>
<td>RM-ANOVA, PCG, CAP, heatmap</td>
<td>NIST library</td>
<td>Paul et al. 2013</td>
</tr>
</tbody>
</table>

Table I. Continued.

† Extraction solvents: MeCN (acetonitrile), MeOH (methanol), EtOH (ethanol), MTBE (methyl-tet-butyl ether), H2O (water), CHCl3 (chloroform), HCO2 (perchloric acid), C2H2 (cycllohexane), TCA (trichloroacetic acid), SPE (solid phase microextraction), SPE (solid phase extraction)

‡ Metabolite components: P = polar component, NP = non-polar component

† Derivation: MISTFA = silylation with N-Trimethylsilyl)Fluroacacetamide, BSTFA = silylation with N-O-Bis(trimethylsilyl)Fluroacacetamide, MCF = alkylation with methyl chloroforate, ECF = alkylation with ethyl chloroforate

‡‡ Analytical platforms: Fl-MS/MS (flow injection tandem mass spectrometry), FT-IR (Fourier transform infrared spectroscopy), 1H NMR (proton nuclear magnetic resonance), 1H-13C NMR (two dimensional proton and carbon NMR for assisting metabolite identifications), HR-MS NMR (high resolution magic angle spinning NMR), FT-ICRMS (Fourier transform ion cyclotron resonance mass spectrometry), DART-MS (direct analysis in real time mass spectrometry), GC-MS (gas chromatography mass spectrometry), LC-MS (liquid chromatography mass spectrometry), RPLC (reverse phase liquid chromatography), HILIC (hydrophilic interaction liquid chromatography)

Note: Whilst not explicitly stated within the table, all metabolomics-based investigations will also include normalisation of data to an internal standard as a data pre-treatment method to compensate for potential technical variations (e.g., variable metabolite recoveries during sample preparation and processing).

** Software: AI Trilogy [Ward Systems Group Inc., US], AMDIS (Automated Mass Deconvolution and Identification System [The National Institute of Standards and Technology, US]), Amos (IBM Corp., US), AMIX (Bruker Corp., Germany), Chemstation (Agilent, US), Chenomx (Tenon Corp., Canada), ChromAtof (LC-Corp., Germany), Compass DataAnalysis (Bruker Corp., Germany), Cytoscape (Shannon et al. 2003), DataAnalysis (Bruker Corp., Germany), EasyLCMS (Fructuoso et al. 2012), Excel (Microsoft, US), Expressionist Refiner MS (GenEdata, Switzerland), Frix (Wojdyr 2010), Gephi (Bastian et al. 2009), GenStat (VSN International, UK), IMPALA (Integrated Molecular Pathway Level Analysis (Kamburov et al. 2011)), JMP (SAS Institute Inc., US), MarkerLynx (Waters Corp., US), MassCenter (IEOL, Japan), MassLynx (Waters Corp, US), MATLAB (Mathworks, US), MestReC (Mestrelab Research, Spain), MetaGeneAlyse (Daub et al. 2003), MetaMapp (Barupal et al. 2012), MestReNova (Mestrelab Research, Spain), Metaboanalyst (Xia et al. 2015), MetaGeneAlyse (Daub et al. 2003), MetaMapp (Barupal et al. 2012), MET-IDEA (Lei et al. 2012), MetPA (Xia & Wishart 2010 [now a component of MetaboAnalyst]), MetPP (Wei et al. 2013), MUV (Multi Experiment Viewer [Howe et al. 2010]), MIDAS (Wang et al. 2014), Minitab (Minitab Inc., US), MSCalc (SoftShell International Ltd., US), MZmine (Katajamaa et al. 2006), MZmine (Katajamaa et al. 2006), NMR Processor (ACD/Labs, Canada), OMNIC (Thermo Fisher Scientific, US), PathwayScruener (Bruker Corp., Germany), PLS Toolbox (EigenResearch Inc., US), PRIME DB (Platform for RIKEN Metabolomics Database, RIKEN Yokohama Institute, Japan), PRIMER (Plymouth Routines In Multivariate Ecological Research [PRIMER-E Ltd, UK]), Prism (GraphPad Software Inc., US), ProfileAnalysis (Bruker Corp., Germany), Primo (Parson et al. 2007), R (R Core Team 2013), SAS (Statistical Analysis System [SAS Institute Inc., US]), SEVEX (Thermo Fisher Scientific, US), SIMCA-P (MKS Unimetrics, Sweden), SpecManager (ACD/Labs, Canada), Spectrum Database (ACD/Labs, Canada), SpectraStitch (Sunset Software, US), Stata (Stata Software, US), Tiberius (Tiberius Data Mining, Australia), TopSpin (Bruker Corp., Germany), Unscrambler (CAMO, Norway), VNMRS (Varian Inc., US), Xcalibur (Thermo Fisher Scientific, US), XVINNMR (Bruker Corp., Germany).

‡‡ Statistical analyses: ANOVA (analysis of variance), RM-ANOVA (repeated measures ANOVA), MANOVA (multivariate ANOVA), PERMANOVA (permutation MANOVA), BNMI (biochemical network mapping), CAP (canonical principal coordinates), CNA (correlation network reconstruction), CVA (canonical variates analysis), DFA (discriminant function analysis), EBAM (empirical Bayes analysis of metabolites), GRRN (general regression neural networks), GTM (generative topographic mapping), HCA (hierarchical cluster analysis), ICA (independent component analysis), KNNA (Kohonon neural network analysis), MDS (multi-dimensional scaling), MCR-ALS (multi-variable curve resolution-alternating least squares), MPEA (metabolite pathway enrichment analysis), ORA (over-representation analysis), PNN (probabilistic neural networks), PCA (principal components analysis), SPCA (supervised PCA), PCRA (robust PCA), PCF (principal component discriminant function analysis), PCO (principal coordinate analysis), PLS-DA (projection to latent structures discriminant analysis), OPLS-DA (orthogonal projection to latent structures discriminant analysis), PLS-LDA (projection to latent structures linear discriminant analysis), PLS-R (projection to latent structures regression), SAM (significant analysis of metabolites), SEM (structural equation modelling), SIMPER (Similarity Percentage analysis), SOMs (self-organising maps), WGCNA (weighted gene correlation network analysis).

§§ Databases: BiGG DB (Database for Biochemically and Genetically and Genomically structured genome-scale metabolic network reconstructions), BML (Birmingham Metabolite Library), BMRB (Biological Magnetic Resonance Data Bank), CRL (Chenomx Reference Library), ECDMB (E. coli Metabolome Database), EHMHN (Edinburgh Human Metabolic Network), GMDDB (Golm Metabolome Database), HMBD (Human Metabolome Database), KEGG (Kyoto Encyclopedia of Genes and Genomes), KEGG LDB (KEGG Ligan Database), NIST (The National Institute of Standards and Technology Library), MBDB (Massbank Database), MMD (Manchester Metabolomics Database), NLDB (Nature Lipidomics Database), SMPDB (The Small Molecule Pathway Database), WMSL (Wiley Mass Spectral Libraries), YMBD (Yeast Metabolome Database).
A range of other NMR-based techniques are also available which have various applications and levels of analytical sensitivity, such as two-dimensional and hyphenated platform approaches (reviewed by Simpson & Bearden 2013; Bharti & Roy 2014; Larive et al. 2014). NMR was originally the workhorse of metabolite profiling in the early days, but recent advances in mass spectrometry based approaches offer alternative methods of analyses. These platforms are often used in combination, since they have their own individual merits (Ong et al. 2009; Zhang et al. 2012). For example, NMR is a non-destructive technique and acquires highly robust and reproducible measurements. Separation of metabolites prior to detection is not necessary and minimal sample preparation is required. NMR is generally cheaper to perform, but unfortunately has comparatively low sensitivity in relation to MS-based platforms, which means only metabolites that are present in significant quantities can be detected.

Mass spectrometry

Mass spectrometry (MS) is a method which involves the measurement of molecular weights of molecules (reviewed by El-Aneed et al. 2009; Viant & Sommer 2013). There are three components to a mass spectrometer: the ion source, the mass analyser, and the detector (Glish & Vachet 2003). There are many different types of these components (see Dettmer et al. 2007; El-Aneed et al. 2009; Junot et al. 2014). At the ion source, metabolites within a sample are ionized by a variety of processes. For metabolomics work, the most commonly used ionisation techniques are electron ionisation and electrospray ionisation (Lei et al. 2011). In most cases, the molecules become sufficiently excited to fragment into a number of electrically charged ions. These ions move into the mass analyser where they are separated based on their mass to charge (m/z) ratio by accelerating them and subjecting them to various combinations of electric, magnetic or electromagnetic fields or in a ‘time of flight’ mass spectrometer which assesses how fast they are travelling. Fragments with different m/z ratios travel at different speeds and are deflected from their forward trajectory to different degrees; lighter ions deflect more than heavier ions, and the higher the ionic charge, the greater the deflection. This allows the various types of mass analysers to filter the ions. The ions are then directed into a device that counts the number of ions at each different mass. This information is plotted in a spectrum of the ion abundance as a function of the m/z ratio. The identity of a metabolite can be putatively elucidated by comparing the fragmentation patterns against open access and/or proprietary databases which contain mass spectra of known compounds. Depending on the particular instrument, some (high resolution mass spectrometers) are capable of determining the actual elemental composition of each ion, thus providing an extra dimension of information for validation of metabolite identity.

MS-based methods are becoming highly sophisticated and newly-developed platform variations are increasingly being showcased in the scientific literature. MS can be performed directly on samples without pre-separation of metabolites (reviewed by Ibáñez et al. 2014). While direct MS techniques are rapid, they also suffer from low ionization efficiencies and ion suppression. Thus, to decrease the complexity of the sample matrix and enhance the sensitivity and selectivity of the analysis, MS-based metabolomic approaches usually involve separation of metabolites via chromatography or electrophoresis prior to MS detection. The benefits of pre-separation are that a significant amount of information is available from the pre-separation process, and metabolites with the same mass can easily be distinguished since they are introduced into the MS system at different times. In addition, higher quantitative accuracies can be achieved since problems associated with ion suppression and other interferences are greatly reduced. Gas chromatography (GC), liquid chromatography (LC), and capillary electrophoresis (CE) are the most commonly applied methods for this purpose. When coupled, these instruments are called hyphenated platforms (GC-MS, LC-MS and CE-MS). Each of these platforms have their own unique advantages, and can be used in combination to obtain very broad coverage of the metabolome (Lei et al. 2011).

Gas Chromatography Mass Spectrometry

Gas Chromatography (GC) separates metabolites which are volatile and thermally stable, or which become volatile and thermally stable after functional group modifications (e.g., alkylation or
silylation via chemical derivatization [Villas-Bôas et al. 2011]) (reviewed by Garcia & Barbas 2011). Once the sample extract has been prepared, it is injected into a hot gas stream flowing through a long and very small diameter tube in the GC instrument. The inside walls of the tube, called a column for historical reasons, are coated with material that has some affinity for the various components in the mixture. The different interactions of the metabolites with the gas stream and the column walls result in differential flow speeds through the column and they exit the column at different times (producing a chromatogram [Fig. 3E,F]); thus entering the mass spectrometer in a unique sequence. This combination of unique entrance times and associated information on the physicochemical properties of metabolites provides an enhanced means of profiling and identification. See Figure 4 for an illustrated overview of the analytical processes involved using pre-separation techniques combined with MS for the metabolomic analysis of complex sample matrices. GC-MS has the advantage that it produces very stable metabolite retention times within the column, does not have drawbacks associated with ion suppression, and generates highly reproducible fragmentation patterns. These features mean that metabolite identifications can more easily be authenticated by matching spectra against those contained within numerous open-access spectral libraries. However, if samples need to be derivatized, are thermally unstable, or have too high a molecular weight, GC-MS may not be suitable. See Zhao X. et al. (2015) for an applied example of how GC-MS was used to identify biomarkers for temperature stress in tilapia, and to discover that an exogenous supply of L-proline into the culture water led to higher disease resistance against bacterial pathogens.

Figure 4. Overview of the processes involved using pre-separation techniques combined with mass spectrometry. Gas chromatography, liquid chromatograph or capillary electrophoresis is used to separate metabolites in the sample extract to produce a chromatogram. Compounds within the peaks are then sequentially analysed by mass spectrometry, and their ion m/z ratios are compared to those stored in mass spectral databases for identification. In some cases, peaks may comprise multiple metabolites with similar physicochemical properties which are unable to be separated by the pre-separation device and deconvolution of the spectra is required.
Table II. Comparisons between different analytical platforms for processing metabolomics samples.

<table>
<thead>
<tr>
<th>Platform</th>
<th>Advantages</th>
<th>Disadvantages</th>
<th>Processing Cost†</th>
</tr>
</thead>
<tbody>
<tr>
<td>NMR</td>
<td>Rapid analysis time (5–10 mins)</td>
<td>Low sensitivity</td>
<td>Cheap</td>
</tr>
<tr>
<td></td>
<td>Simple sample preparation</td>
<td>Convoluted Spectra</td>
<td>$30–100 USD per sample</td>
</tr>
<tr>
<td></td>
<td>No derivatisation needed</td>
<td>Libraries of limited use due to complex matrix</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Provides detailed structural information</td>
<td>More than one peak per component</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Low chemical bias</td>
<td>Peak overlap common</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Very reproducible</td>
<td>pH adjustment required</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Can be high resolution</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Excellent metabolite recovery (no suppression)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Highly quantitative (without standards)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GC-MS</td>
<td>Very sensitive</td>
<td>Slow analysis time (30–60 mins)</td>
<td>Expensive</td>
</tr>
<tr>
<td></td>
<td>Very robust</td>
<td>Extensive sample preparation</td>
<td>$100–200 USD per sample</td>
</tr>
<tr>
<td></td>
<td>Large linear range</td>
<td>Derivatisation required</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MS provides some structural information</td>
<td>Destructive to sample</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Many available libraries for metabolite identification</td>
<td>Some metabolites cannot be made volatile</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pre-separation provides additional information</td>
<td>Some metabolites are too large for analysis</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Does not suffer from ion suppression</td>
<td>Cannot detect some thermally unstable metabolites</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Reproducible retention times</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Quantitative (with appropriate standards)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LC-MS</td>
<td>Very sensitive</td>
<td>Analysis time can be slow (10–60 mins)</td>
<td>Very expensive</td>
</tr>
<tr>
<td></td>
<td>Can detect a very wide range of metabolites</td>
<td>Lack of comprehensive spectral libraries</td>
<td>$150–400 USD per sample</td>
</tr>
<tr>
<td></td>
<td>MS provides some structural information</td>
<td>Ion suppression &amp; adduct formation problems</td>
<td></td>
</tr>
<tr>
<td></td>
<td>High mass accuracy</td>
<td>Destructive to sample</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Many modes of pre-separation available</td>
<td>Metabolite identification is difficult</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pre-separation provides additional information</td>
<td>Low retention time reproducibility</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Quantitative (with appropriate standards)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FT-IR, NIR, Raman</td>
<td>Very rapid analysis time (10–60 secs)</td>
<td>Extremely convoluted spectra</td>
<td>Very Cheap</td>
</tr>
<tr>
<td></td>
<td>Low chemical bias</td>
<td>More than one peak per component</td>
<td>$10–50 per sample</td>
</tr>
<tr>
<td></td>
<td>Can be used directly on samples</td>
<td>Metabolite identification almost impossible</td>
<td></td>
</tr>
<tr>
<td></td>
<td>No derivatisation required</td>
<td>Often requires sample drying</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Complete fingerprint of sample composition</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Useful for identifying functional groups</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

† Per sample processing costs vary between service providers and depend highly on the number of samples to be analysed within a particular project, and the resolution of the specific platform to be employed. The price ranges displayed are not strictly defined limits but are typical of the current rates charged for commercial samples, and generally will include metabolite extraction (and derivatisation when required), instrumental analysis, and some primary bioinformatics processing. Additional statistical analysis, secondary bioinformatics processing and/or biological interpretation of the data can usually be provided as an extra service by most facilities e.g. $50–200 USD per hour for a metabolomics specialist. Academic-based metabolomics service providers may offer discounted rates for collaborative projects.

**Liquid Chromatography Mass Spectrometry**

High Performance Liquid Chromatography (HPLC or LC) is based on similar chromatographic principals as GC, but the sample is not heated to high temperatures (reviewed by Xiao et al. 2012). The most important distinction between GC and LC is that GC largely separates metabolites based on their boiling points with secondary retention by polarity, whereas metabolite size or polarity are the main mechanisms of LC. The column in this case is a tube a few millimetres wide and a few centimetres long and packed with extremely fine powder coated with material that has some affinity for the various metabolite components in the sample. A suitable solvent mixture is pumped at very high pressure through the column. The sample is introduced into the column as a solution and, like GC, the various components of the mixture travel through the column at different speeds and exit in sequence to produce a chromatogram (Fig. 3G,H). From there the outlet stream is directed into a suitable detector which, for metabolic work, is typically a high-resolution mass spectrometer. Therefore, LC-MS can analyse a wider range of metabolites since non-volatile and thermally sensitive compounds can be separated in the liquid phase. However, LC-MS suffers from greater ionization and matrix effects,
and lower chromatographic reproducibility. These features make assigning metabolites through spectral library matching considerably more difficult. Nevertheless, LC-MS is a very popular metabolomic platform and the analytical technology, spectral libraries, and software for processing spectra are continually being updated to improve metabolite identifications. See Yan et al. (2012) for an applied example of how LC-MS was used to identify species-specific metabolic stress responses of fish immediately after a tropical cyclone at various cage-farming sites, and to determine the physiological mechanisms which resulted in high mortalities during the following month of grow out.

Capillary Electrophoresis Mass Spectrometry

Capillary electrophoresis (CE) is an alternative pre-separation technique which separates metabolites based on their ionic charge characteristics, or electrophoretic mobility (reviewed by Ramautar et al. 2009). In many respects this technique is similar to LC, but molecules are separated based on their ionic affinities and size rather than on their solid phase solubilities (Hiryama et al. 2014). In CE, sample extracts enter a column which contains electrolytes. Charged metabolites migrate through the column and exit at different times under the influence of an electric field, and can be further concentrated using gradients in conductivity and pH. CE-MS is an efficient platform that does not require rigorous sample pre-treatment, is useful for small samples, is good at separating highly polar metabolites, has separation power and sensitivities which are comparable to GC-MS and LC-MS, and can quantify certain metabolites that other hyphenated MS platforms cannot. On the other hand, CE is unable to separate non-charged compounds, and suffers more than GC or LC from poor reproducibility. However, recent advances in CE-MS technologies are contributing to the increasing usage of the technique in metabolomics studies (Ramautar et al. 2015). See Koyama et al. (2015) for an applied example of how CE-MS was used to gain detailed metabolic insights into salinity adaption of brackish-water clams from four commercial fishery grounds in Japan with different water chemistries.

The sensitivity, or at least the detection limits, of MS techniques can be extremely high. As long as a substance can be separated, detection limits in parts per million or even better are possible. If pre-concentration techniques are used, molecules can easily be detected in concentrations of parts per trillion or better.

Vibrational spectroscopy

The analysis of complex sample matrices can also be performed using lower resolution instruments which measure the vibrational signatures of broad metabolite functional groups (Moore et al. 2014). Such analyses generally do not provide detailed information for identifying particular metabolites, but can still be very useful for obtaining an overall ‘metabolite fingerprint’ of a sample. This fingerprint is based on the holistic composition of functional group chemistries across all metabolites within the sample, and can be used to classify samples from different conditions when significant variations are observed. However, the drawback is that biological interpretation of spectra can be difficult because of this non-specificity. The application of vibrational-based technologies, such as Fourier transform infrared (FT-IR), near infrared (NIR), and Raman spectroscopy are growing in popularity due to their rapid and high through-put analysis capabilities, their ability to work with very small samples, and their very low cost compared to other platforms.

Infrared Spectroscopy

Infrared techniques work on the principal that when a sample is exposed to light, or electromagnetic radiation, the different chemical bonds within metabolite functional groups absorb energy at different wavelengths and vibrate in characteristic ways. A plot of the absorbance or transmittance of light at different wavelengths produces a spectrum which represents the overall metabolite composition of the sample to provide a snapshot, or fingerprint, of the organism’s metabolome (Fig 5.). Infrared platforms are categorized into Near-Infrared (NIR) 0.78–3 µm and Mid-Infrared (MIR) 3–50 µm depending on the wavelength of light used to analyse the samples. Modern instruments commonly use Fourier transform techniques (a mathematical process which converts the raw data from the instrument into a spectrum) so the expression FT-IR is often seen when discussing MIR spectroscopy. MIR analysis examines the absorptions of bond vibrations and other molecular movements, whereas NIR evaluates the overtones
and combinations of strong MIR absorptions, which, while not as specific as the sharper, stronger MIR absorptions, can be characteristic and more easily quantified for a range of biologically important functional groups such as sugars, fats, and proteins. Unlike MIR, NIR can penetrate many millimetres through water and the instruments can use glass optics. Infrared platforms have proven useful for a range of aquaculture-related purposes. For example, to identify pathogenic bacteria responsible for disease in farmed salmon (Wortberg et al. 2012), to determine the causation for post-harvest variations in shrimp quality based on the methods used for culling (Fu et al. 2014), to identify fraudulently marketed fish from different origins (Vidal et al. 2014), to assess the meat quality of various fish species (Cheng et al. 2013; Qu et al. 2015), to develop new food safety and authentication techniques for classifying shelled shrimp based on their post-harvest storage conditions (Qu et al. 2015), and to develop fast and cost-effective methods for proximate chemical analysis of cultured shellfish for the purposes of monitoring animal condition and assisting in selective breeding programs (Brown et al. 2012), among others.

**Figure 5.** An example of comparative IR and Raman spectra obtained from the analysis of blood serum (reproduced from Ellis & Goodacre 2006).

**Raman Spectroscopy**

Raman spectroscopy is a technique closely related to MIR. When a laser beam hits a molecule, approximately 1 in 10^7 photons will interact with electrons in the chemical bonds resulting in the scattered laser light having extra wavelengths (a few nm) added to it and subtracted from it which correspond to the vibration frequencies of the bonds in the molecule. These shifts in photon wavelengths are called the ‘Raman effect’. The original laser colour can be subtracted using filters and the remaining frequencies provide information about the vibrational, rotational and other low frequency transitions within metabolites. Raman spectra are closely related to MIR spectra and look very similar (Fig. 5). The principal difference is that the sorts of chemical bonds that give weak MIR absorptions are usually very strong in the Raman spectrum, and vice versa, so the two techniques are complementary. Although its big drawback is the very weak Raman signal, Raman spectroscopy has several major advantages. Glass optics can be used and since Raman spectroscopy is based on the scattering of incident light rather than on absorption, it does not suffer from interferences caused by water. Thus, measurements can be made directly on biofluids and aqueous extracts, minimal to no sample preparation is required, and spectra can be obtained very quickly. See Ishigaki et al. (2014) for an applied example of how non-invasive Raman spectroscopy was used on live fish eggs to predict and monitor their quality and viability to ensure successful fertilizations.

**Metabolite fingerprinting vs. profiling**

There are generally two approaches to generation and examination of metabolomics data – metabolite fingerprinting and metabolite profiling. The approach utilized depends largely on the objectives of the investigation and the facilities available.
Metabolite fingerprinting compares the overall nature of samples based on the entire set of signals generated by the analytical platform. These signals, or features, are analysed using statistical techniques to discern patterns in the data for the purpose of sample classification. This approach usually involves the analysis of a very large number of signals, which represents the total compositions of metabolites, and does not necessarily require metabolite identification. Data obtained from NMR and vibrational spectroscopy platforms are particularly well-suited for metabolite fingerprinting. However, interpretations of results within particular biological frameworks are limited unless further analyses of relevant features within the fingerprints are performed. Nevertheless, metabolite fingerprinting can be convenient for situations when only sample-class discrimination is required. For example, Aursand et al. (2007) used metabolite fingerprinting to reliably identify fraudulently mislabelled fish oil products to ensure food safety and develop novel techniques for food traceability and quality assurance.

Metabolite profiling evaluates all of the signals generated by the analytical platform so that they may be characterized and matched to spectra of known metabolites in reference libraries. Once identified, data analyses are then performed on the abundances of the metabolites within the samples. This approach provides data which can be more easily interpreted across various biological frameworks since features are ascribed an identity with often well-known biochemical roles. Metabolite profiling frequently leads to discovery of biomarkers and development of novel and testable hypotheses. For example, Guo et al. (2014) used metabolite profiling to identify early-warning biomarkers to predict fish health, and to better-understand the mechanisms of defence against bacterial infection.

Metabolite fingerprinting combined with profiling is sometimes used when very large numbers of signals are present within the raw spectral data so that only those features statistically different between samples, or otherwise deemed important, are subsequently identified. This approach can be used to reduce the computational and resource demands of processing noisy or large and complex datasets. In most metabolite fingerprinting applications, signals that are different between samples are usually identified to aid interpretation of the data. For example, Savorani et al. (2010) and Picone et al. (2011) used combined approaches to identify factors responsible for meat quality variation in fish reared under different culture environments, and stored under different post-harvest conditions.

NMR-based metabolomics usually involves fingerprinting as an initial step, whereas use of MS-based hyphenated platforms may involve profiling only. It is important to note that the definitions and term usage for these two approaches tend to be flexibly applied in the literature and as yet there are no standardized descriptions.

**Data Analysis**

Although metabolomic datasets are often very large and complex, recent advances in bioinformatics and streamlined statistical workflows provide simple strategies for coping with the high dimensional data (Johnson et al. 2015). Bioinformatics is an interdisciplinary field incorporating computer science, database management, mathematics and statistics. Primary bioinformatics processing involves analysis of the raw data obtained from the analytical platform and incorporates all procedures which are required to generate a list of features or metabolites. The resulting data can then be analysed by a range of classical and applied statistical procedures.

A number of steps are involved in the primary bioinformatics processing and usually includes data conversion, spectral processing (e.g., deconvolution, alignment, noise reduction), feature selection, metabolite identification via database matching, metabolite quantification, and quality control procedures. While a variety of freely or commercially available software packages exist to perform these tasks, many laboratories employ their own proprietary programs and algorithms, which have been custom designed for their unique situations and analytical set-ups. The methods used for primary bioinformatics processing vary widely and depend on data type and the analytical platform employed. Thus, it is impossible to provide general advice. However, at the end of this section we direct readers to a wide range of relevant literature, which covers these topics in more depth.

Prior to statistical analysis of metabolite profile/fingerprint data, data scaling, normalization and/or transformations are often performed to enhance extractability of biologically relevant
information from the dataset. Metabolite concentrations have huge dynamic ranges, and variance is typically larger at higher concentrations. Because many statistical procedures rely on homoscedasticity or distributional assumptions, it is important to alleviate the dependency of the variance on the concentration through variance-stabilizing transformation or transformation to normality. Furthermore, the relative abundances of different metabolites are not proportional to the biological importance that they may represent, and many data analysis techniques fail to take this into consideration. Some of the more commonly applied pre-treatment methods for metabolomics data include centering, autoscaling, pareto scaling, range scaling, log transformation, and power transformation (reviewed by van den Berg et al. 2006). The pre-treatment method chosen may vary between different metabolomics datasets; hence, a solid understanding of how the implemented method affects the outcome of subsequent statistical analyses is essential for reliable interpretations.

Statistical data analysis can be achieved using general statistical software (e.g., Minitab [Minitab Inc., PA, USA], SIMCA [Umetrics, Umea, Sweden], SPSS [IBM Corp., NY, USA], STATISTICA [Statsoft Inc., OK, USA]) or dedicated metabolomics-based data analysis packages (e.g., DeviumWeb [Grapov 2014], MeltDB [Kessler et al. 2013], Metaboanalyst [Xia et al. 2015]). A basic knowledge of programming is useful for employing, modifying, or writing script in certain data analysis environments (e.g., Matlab [Mathworks Inc., MA, USA], R [R Core Team 2014]). However, recent development of easy-to-use graphical user interfaces for these environments have substantially reduced the need for advanced programming skills. See Mishra and Van der Hoot (2016) for further information regarding the latest advances in available computational tools and resources for the analysis of metabolomics data. OMICtools (http://omictools.com/) is also a useful and growing online repository of web-accessible tools related to omics-based data analysis.

Similar to other –omics disciplines, it is common for the number of measured variables (genes, proteins or metabolites) within each sample to far exceed the number of samples analysed. Metabolomics data are by their very nature multivariate in design and lend themselves particularly well to multivariate statistical analyses. However, univariate techniques can also be employed to extract valuable information from the data. Use of both approaches in combination is routinely performed and recommended because they can expose different characteristics of the samples (Sugimoto et al. 2012).

Univariate methods

Univariate methods involve analysis of single variables (metabolites) at a time. T-tests and ANOVA’s (analysis of variance), or their non-parametric equivalents (e.g. Mann-Whitney U test, Kruskal-Wallis test), are the most commonly applied univariate techniques to identify differences in metabolite abundances between samples. However, due to the high number of variables, it is important to correct for multiple hypothesis testing to protect against the likelihood of identifying false-positives (Broadhurst & Kell 2006). SAM (significant analysis of microarrays/metabolites) is an example of a univariate method which is able to account for correlations between metabolites and does not assume independence, unlike the T-test and ANOVA. Volcano plots are often used for the univariate analysis of gene, protein and metabolite expression data (Li 2012). Volcano plots are scatterplots which incorporate a measure of statistical significance (T-test p-values) with information about the magnitude of metabolite change (fold-change) (Fig. 6).

**Figure 6.** Example of a volcano plot. Solid yellow circles represent metabolites which are significantly different between sample groups (P<0.05), as well as have large variation (> 1.2 fold-change) in their mean abundances.
They allow quick identification of metabolites, which are not only statistically different between two sample conditions, but which also co-display large variations in abundance. See Young et al. (2015) for an applied example of how a volcano plot and SAM was used to assist construction of a multivariate classification model for assessing the quality of hatchery-reared mussel larvae.

Univariate methods are attractive because they are generally simple to apply and the results are easily interpreted and communicated across various levels of expertise. However, they cannot detect group differences when only minor variations exist on a single molecule level. Associations between metabolites and low variations in abundance can be highly important on a systems level due to the orchestrated flux of metabolites within common biochemical networks.

**Multivariate methods**

Since univariate techniques may not account for interrelations between metabolites, multivariate methods are applied to compensate, and to provide additional and complementary information for assisting interpretation of the data. Multivariate techniques can be used to reduce complexity and identify patterns, group structure, and relationships among metabolites and samples (Worley & Powers 2013). Commonly applied procedures include Principal Components Analysis (PCA), Projection to Latent Structures Discriminant Analysis (PLS-DA), and clustering.

**Principal Components Analysis**

PCA is a mathematical procedure that aims to capture and extract most of the important information in a high-dimensional data matrix and re-express it in fewer dimensions (Abdi & Williams 2010). In doing so, the data can be more easily visualized, described, and analysed. PCA does this by combining the multiple correlated variables into a number of smaller uncorrelated variables called principal components. A different data matrix is constructed in which the first 2–3 new variables account for the vast majority of the total variance in the original data. The samples can then be projected and visualized on a 2D or 3D score-plot (Fig. 7). PCA is an unsupervised statistical technique which incorporates only the independent metabolite information. Dependant variables are not required for modelling and information of sample class membership is not included in the analysis. As an unsupervised technique, patterns among the independent variables are discerned and groups of samples are formed based solely on the structure of the metabolite data. The PCA algorithm therefore achieves unbiased dimensionality reduction and only exposes group structure when within-group variation is substantially less than between-group variation. PCA is very useful for visualizing multi-dimensional data, identifying outliers, conducting classification studies, identifying a subset of original variables which explain most of the variation between samples, and for exploratory data analysis before building predictive models. See Kokushi et al. (2015) for an applied example of how PCA was used to identify differential regulation of metabolic pathways due to insecticide exposure in freshwater carp.

![Figure 7](image-url)
Projection to Latent Structures Discriminant Analysis

Similar to PCA, PLS-DA is a technique, which can be used to reduce dimensionality and help visualize and analyze multivariate data (Worley & Powers 2013). However, PLS-DA is a supervised statistical technique, which incorporates information about the sample classes. Using this information, PLS-DA rotates the data within the newly created latent variable subspace in a way that maximizes separation between groups of samples. This can result in much clearer separations than when PCA is applied (Fig. 7). PLS-DA can be very useful for identifying and ranking metabolites which contribute most towards sample group separations and, when applied correctly, to assist construction of predictive classification models. Orthogonal PLS-DA is a related technique, which can further enhance separations due to its ability to distinguish between predictive and non-predictive (orthogonal) variation (Bylesjö et al. 2006) (Fig. 7).

PLS-DA and its extensions have a tendency to over-fit the model to the data. Therefore, validation is important when using these algorithms in predictive capacities. Model validation is the process of defining a model's performance and is a critical requirement for predictive modelling (Szymańska et al. 2012). This ensures that the model's internal variable rankings are truly informative. Commonly used methods to test a model's performance include permutation-based tests and cross validation (Worley & Powers 2013). The ideal scenario involves the use of a training dataset to build the model, and a separate validation dataset to assess its predictive capacity. See Liu et al. (2015) for an applied example of how PLS-DA and OPLS-DA were used to identify metabolites associated with white spot syndrome virus infection in shrimp, and provide preliminary information for developing biomarkers for diagnosing the pathophysiology of the disease.

Clustering

Clustering is a collection of statistical procedures, which aims to group samples together that are most similar in their metabolite profile (reviewed by Andreopoulos et al. 2009). Like PCA, most clustering techniques involve unsupervised approaches to group samples and the goal of clustering is to identify the actual groups based on the underlying structure of the data. Where PCA selects the variables with the most variation to form a reduced data matrix for partitioning samples, cluster analysis algorithms do not lose variance through dimensionality reduction in the same way and generally use all variables equally to display sample similarity/dissimilarity. Although clustering can be used to discover structures within the data irrespective of sample-class membership, it does not explain why they exist. Nevertheless, clustering is a very useful exploratory technique for uncovering patterns, finding natural groupings, confirming known groupings, identifying outliers, and discovering groups of metabolites with similar expression patterns across a wide range of biological conditions by clustering the variables rather than the samples. The most commonly applied clustering algorithms in metabolomics-based investigations are Hierarchical Cluster Analysis (HCA) and K-means clustering.

HCA is a method which seeks to construct a hierarchy of clusters and arrange them into a binary tree-structured graph called a dendrogram (Meunier et al. 2007). HCA does this by successively merging comparable groups based on the similarity/dissimilarity, or distance, between them. Visualizing this tree provides a useful summary of the data. HCA can be combined with data visualization techniques to provide new ways of looking at the data, and to enhance the extraction of important information (Fig.8). See Courant et al. (2013) for an applied example of how HCA was combined with heatmap analysis to assist visualization of metabolite-group expressions, and to identify biomarkers for fine-scale monitoring during continuous culture of microalgae under different nitrogen regimes.

K-means is a non-hierarchical, unsupervised, partitional clustering approach. Although sample class membership information is not incorporated into the analysis, the researcher must initially define how many clusters (k number of clusters) into which the samples are to be partitioned. Like other clustering techniques, the aim is to gather samples into groups so that those in the same group are most similar to one another, and those in different groups are as different as possible. Working within an n-dimensional subspace of true vectors (number of variables), the algorithm performs this task through an iterative sequence of minimizing the sample distances to a centroid point.
Figure 8. Combined heatmap and hierarchical cluster analysis of metabolites in developing zebrafish during embryogenesis via GC/MS- and LC/MS-based metabolomics (reproduced from Huang et al. 2013). Each column represents a sample (five biological replicates for each of the five development stages). Each row represents the abundance of a particular metabolite (red = high abundance, green = low abundance). Metabolites cluster naturally into groups which, in this case, have functional relationships (labelled metabolite classes in dotted boxes).

within each of the k number of clusters, and reallocating the samples to the cluster with the closest centroid so as to minimize the within-cluster sum of squares. Initially, the first centroid points are randomly placed and samples are assigned to a cluster. Then, the true centroid points of those clusters are calculated and repositioned, samples are reassigned, and the clusters are redefined. This is performed repeatedly until convergence is found. Use of k-means clustering in aquaculture-related metabolomics research is limited thus far. However, see Yu et al. (2013) for a relevant example of how k-means was used to identify groups of genes with similar expression profiles in fish which had been
fed a diet contaminated with the persistent organic pollutant BDE-47, and to determine potential enzymatic and metabolic mechanisms of toxicity defence.

While we have discussed three widely used multivariate techniques to analyse metabolomics data, one should be aware that an array of other procedures are available which may be better-suited for the analysis of particular datasets in some cases. These include: multivariate analysis of variance, linear discriminant analysis, partial least squares regression, support vector machines, k-nearest neighbour, random forests, soft independent modelling of class analogies, and self-organizing maps, among others. For further information on these alternative data analysis approaches, see Steuer et al. (2007), Liland (2011), and Xi et al. (2014). For platform-specific reviews on various bioinformatics processes see Smolinska et al. (2012), Sugimoto et al. (2012), Du & Zeisei (2013), Engel et al. (2013), and Wei et al. (2012 & 2014).

**Biomarker discovery and validation**

The aim of many metabolomics-based investigations is to discover novel metabolite biomarkers, which correlate with specific diseases or health states. These molecules can then be used as early diagnostic tools, or in conjunction with other assessments for confirmation of pathology. Metabolite biomarkers are also very useful within the aquaculture industry to assist in the evaluation of pre- and post-harvest meat quality, and for food safety and traceability purposes (Alfaro & Young 2015). Biomarkers may be single metabolites, multiple metabolites, ratios of metabolite pairs, particular features (e.g., ion fragments), or entire unannotated spectral fingerprints.

The initial step in biomarker discovery is often to perform an exploratory experiment with different treatments or animal conditions, and to identify features which are substantially different between the sample groups using one or more statistical procedures outlined in the previous section. Once determined, these features can be considered as ‘candidate biomarkers’. The purpose of initial biomarker discovery is to identify the most salient features for further investigation, and may involve low biological sample replication (n < 10), although higher replication is usually preferred. The results of these studies can be very useful for generating hypotheses, and gaining preliminary mechanistic insights into metabolic factors responsible for, or involved in, particular health states or other conditions. However, when the ultimate goal is to develop practical biomarkers with useable applications and minimal risks for Type II errors occurring, they must have extremely reproducible performances. Thus, in order to ensure that the identified candidate biomarkers have high sensitivity and specificity for the particular condition under investigation, it is important to validate them.

The process of biomarker validation was born from the medical research field where the misdiagnosis of a health condition might result in a disastrous outcome for a patient. Biomarker validation is a quality assurance process of defining the performance of a biomarker within acceptable limits, whilst understanding and minimizing the rate of false discovery. Biomarker validation usually involves one or more additional experiments where the candidate metabolite/feature is targeted more specifically for quantification using complementary or alternative analytical platforms with a high selectivity for that analyte. Such experiments will typically also involve much higher sample replication (n = 100–1000), experimental replication, a broadening of scope in some cases (e.g., incorporation of multiple sexes, development stages, and environmental conditions), and rigorously refined statistical approaches (e.g., permutation based cross-validation or using different sub-sets of the samples to construct, validate, and test the performance of a predictive model [Westerhuis et al. 2008; Szymańska et al. 2012; Xia et al. 2013]). Accordingly, biomarker validation within the framework of a high-quality clinical study can involve substantial costs and time, which may not be viable for an environmental study or commercial aquaculture exercise. However, what constitutes validation is a subjective measure and is scalable within the confines of the researcher.

In a practical scenario where time and funds are limited, an alternative approach might be to employ a particular candidate biomarker whilst accepting its potential vulnerability, and continually adding new data to the predictive model as it becomes available. In this way, quality control limits on the model’s performance can be set and monitored as it is updated. Thus, expenses are diluted over time
and performance-based milestones can be implemented based on cost-benefit analyses to guide management decisions in an empirically data driven context. Validation is an important concept when identifying and implementing new biomarkers, and should be a carefully considered component within the general strategy of a metabolomics-based biomarker discovery and development project. For further information on biomarker discovery and validation procedures, see Xia et al. (2013).

Biological interpretation & secondary bioinformatics

The discovery and identification of biomarkers do not always necessitate in-depth functional explanations for their presence and/or roles. For example, a simple metabolomics-based study using nuclear magnetic resonance and a fingerprinting approach combined with pattern-recognition tools (e.g., PCA, PLS-DA) could be used for food authentication purposes to identify biomarkers to classify an adulterated product, or determine its provenance (reviewed by Cubero-Leon et al. 2014). However, for many investigations, more detailed insights into the reasons for sample group separations are required, and the procurement of mechanistic biochemical explanations are highly desirable. In such cases, it becomes necessary to interpret the data within biologically meaningful frameworks.

The past 100 years of biological research has provided us with an amazing wealth of knowledge concerning cellular metabolism across a wide range of taxa. Rigorous empirical experimentation by a multitude of pioneers during this period has established the major biochemical pathways. Not only do we know which genes, enzymes, cofactors, substrates, products, and intermediates are involved in these pathways, in many cases we also know about individual enzyme kinetics and have detailed information about vast arrays of endogenous and exogenous factors which influence their pathway flux (German et al. 2005). Information such as this provides us with a rich source of knowledge which can be used to assist the interpretation of biochemical data. Nevertheless, interpretation of metabolite expression data can be one of the most challenging aspects of a metabolomics study.

In many cases, concentrations of particular metabolites within a tissue, biofluid, or organism may correlate very well with our current understanding of biochemical networks and the functional relationships among metabolites, enzymes, and genes within normal or perturbed systems. For example, classic signs of stress caused by pathogen or toxin exposure in aquatic animals include increased levels of reactive oxygen species (ROS), and differential co-expression of metabolites (e.g., glutathione, NADPH) and enzymes (e.g., glutathione reductase, superoxide dismutase, catalase) involved in regulating excess ROS production in order to maintain redox homeostasis (Parrilla-Taylor et al. 2013; Macías-Mayorga et al. 2015). The results of recent omics-based investigations provide data which corroborate the presence of such mechanisms in various taxa, as well as offer new information on associated regulatory pathways (Srivastava et al. 2013; Barth et al. 2014; Shi et al. 2015). On the other hand, a number of studies (particularly those containing metabolomics-based components) are providing data which are shedding light on unfamiliar biochemical associations which cannot be explained by our current theses of molecular biology and biochemistry (Steuer 2006). For example, unlike genes and proteins, it is relatively common for metabolite levels within a particular pathway to be highly correlated with metabolites from other pathways for which a mechanistic connection is not currently known. It is intriguing and unexpected results like these that are starting to deliver new information that is helping to push forward our understanding of metabolic networks at an astonishing rate, and also highlights the usefulness and efficiency of omics-based approaches for generating novel data to assist new interpretations. The continual development of metabolomic techniques to characterize larger and larger sets of metabolites requires new methods to analyse these data in order to obtain biologically meaningful information. Here, we briefly outline a few methods that can be used to help researchers interpret their metabolomics data beyond more conventional scenarios, involving assessments of single metabolite variations based on a priori biochemical knowledge.

If biological replication is sufficient, a simple method involves correlation analysis in which construction of a correlation matrix of pairwise
metabolite level comparisons are made. Such matrices can be useful for identifying potentially important relationships requiring further investigation. For example, consider the following hypothetical situation where levels of metabolite X and metabolite Y are not significantly different between control and treatment groups. However, metabolite X is positively correlated with metabolite Y in the control group, and negatively correlated in the treatment group (Fig. 9). Such a scenario would indicate that some major perturbation of the underlying network was taking place, and would have gone undetected had the correlations not been investigated. Differential non-linear correlation patterns may also be present which would require alternative methods of detection. For further information on the interpretation of linear and non-linear correlations in metabolomics data, see Camacho et al. (2005) and Steuer (2006).

While our knowledge is relatively comprehensive compared to only a few decades ago, much of our understanding to date has come from highly targeted analyses of specific pathway components, and it is increasingly becoming clear that there are many gaps to be filled. With more of a focus on the interconnections between pathway components, we are starting to uncover new insights into metabolism which are much more integrated than ever before. An alternative method for identifying metabolite association patterns is called correlation network analysis.

Correlation networks are increasingly being used in omics-based applications to visually capture the overall network of interconnections between biomolecules and to describe the correlation patterns, to identify relationships between entire biochemical pathways, to discover new modules or clusters of relationships, and to assist data interpretation (Langfelder & Horvath 2008; Hero & Rajaratnam 2015). Applied to metabolomics, correlation network analysis is a technique that maps the relationships between every metabolite pair onto a metabolite network. Lines between metabolites typically are descriptive of the relationship between them (e.g., a solid line for a positive correlation and a dotted line for a negative correlation), and may also be quantitative (e.g., defined by the width of the line). The positions of the metabolites within the network map may be placed manually to enhance visualisation, or for additional interpretive purposes they may be positioned using algorithms to identify and define metabolite modules that cluster together. For an applied example of a study involving an aquatic organism, see Southam et al. (2008), who used a combination of correlation analysis techniques to identify key metabolic differences in hepatic tumors of flatfish compared to control tissues, and to assist detection and interpretation of the underlying mechanisms involved in the diseased phenotype (Fig. 10).

Correlation network analysis can additionally be used to integrate transcriptomic, proteomic and metabolomic datasets to help identify functional roles at different biochemical levels (e.g., gene-gene/protein interactions and relationships between enzymes and metabolites) (Higashi & Saito 2013). There are a number of software packages available to perform correlation network analysis, such as DPClus (Altaf-Ul-Amin et al. 2006), Metscape (Karnovsky et al. 2012), COVAIN (Sun & Weckworth 2012), 3Omics (Kuo et al. 2013), and MetaMapR (Grapov et al. 2015). For further information on correlation network analysis and various applications, see Steuer (2006), Adourian et al. (2008), Hüning et al. 2013, and Kotze et al. (2013).
Figure 10. An example of two correlation networks constructed using NMR-based metabolomics data from samples of healthy and diseased fish livers (reproduced from Southam et al. 2008). Solid lines represent positive correlations between metabolites, and dotted lines represent negative correlations. Grey lines represent similarly shared relationships between the healthy and diseased phenotypes, and coloured lines represent those which are dissimilar. Clear differences in the underlying biochemical networks are easily visualised using this technique.

Other procedures useful for supporting data interpretation include a variety of ‘pathway enrichment analysis’ techniques. Rather than focusing on individual metabolites which may be responsible for discriminating groups of samples, pathway enrichment analysis techniques aim to discover predefined metabolic pathways or biological networks that are altered in an orchestrated manner. Such analyses make use of large amounts of biochemical information collated over decades and stored in publicly accessible depositories, such as the Kyoto Encyclopaedia of Genes and Genomes (KEGG), and can be considered as secondary bioinformatics processes. There may be cases where levels of individual compounds are not identified as being statistically different between samples using conventional statistical approaches. However, when analysed together as functional groups, or metabolite sets within their known pathways, it might be revealed that particular pathways as a whole are being differentially regulated under certain experimental conditions. The recent development of secondary bioinformatics tools (reviewed by Booth et al. 2013) to analyse biochemical data within the context of predefined metabolite sets are changing the way that the results of metabolomics projects are interpreted. Pathway Activity Profiling (PAPI) is one example of such a technique (Aggio et al. 2010).

PAPI is an algorithm developed into an R package which can be used to analyse sets of functionally-related metabolites, and quantitatively compare the activity of metabolic pathways between different groups of samples. PAPI performs this task by calculating ‘activity scores’ based on the number of metabolites identified from each pathway and their relative abundances. Pathways for which each detected metabolite is involved in is collected from KEGG, and each is given a score based on the absolute abundance/relative abundance of the metabolite to which it is linked. The pathways are ranked by the total number of metabolites they comprise, and the percentage of detected compounds within them are calculated. The sum of the scores for each pathway are then calculated and normalized (dividing by the proportion of metabolites detected from within the respective pathway) (Aggio et al. 2010). This simple yet effective method can be used to help determine the likelihood of a particular biochemical process being up- or down-regulated under certain circumstances. To our knowledge, PAPI has not yet...
been applied to studies involving aquatic organisms. However, it has been successfully applied in a number of other biological systems (Han et al. 2012; Portella et al. 2014; Zhao C. et al. 2015).

Another useful pathway analysis tool is called Metabolite Set Enrichment Analysis (MSEA) (Xia & Wishart 2010; Kankainen et al. 2011; Persicke et al. 2012). MSEA is an algorithm designed to detect subtle, but consistent changes among groups of metabolites within the same biological pathway. Using an analysis package with MSEA capabilities (e.g., MarVis-Pathway [Kaever et al. 2014], MeltDB [Kessler et al. 2013], Metaboanalyst [Xia et al. 2015]), a quantitative dataset of annotated metabolites can be cross-referenced with information in the KEGG database, and metabolite sets belonging to reference pathways from various model organisms (e.g., human, mouse, zebrafish, drosophila, nematode) can be analysed together as a group. The ability to examine biochemical information for different animal models is a key advantage of MSEA, and options also exist to use proprietary/customised background sets of data from any organism. Pathway enrichment analysis techniques, which use software to interrogate databases that contain global biochemical knowledge, are tremendously powerful data interpretation tools. For applied examples see Zhao X. et al. (2015) and Ma et al. (2015), who utilized MSEA to identify differentially enriched pathways in Tilapia infected with two pathogenic Streptococcus species, and to develop remedial strategies to enhance disease resistance.

Table III. Summary of sample-specific topics (prior to chemical analysis) which should be described in detail when reporting the results of a metabolomics project.

<table>
<thead>
<tr>
<th>Focus area</th>
<th>Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample</td>
<td>Taxonomic classifications, common name/s, genotype/s, ecotype/s, sample composition, sample type, specimen condition (phenotypic characteristics, weight, age, sex, development stage, health)</td>
</tr>
<tr>
<td>Environment</td>
<td>Any field environment: Geographic location, habitat, depth, meteorological conditions (e.g. precipitation, wind speed/direction, humidity), lunar/solar phase, other measured parameters (e.g. pollutant concentrations)</td>
</tr>
<tr>
<td>Process (biological)</td>
<td>Maintenance and acclimation of organisms: Procedure and means (e.g. cage, aquaria, static/flowthrough tanks, continuous culture), reasons for maintenance/acclimation, other parameters (e.g. feeding regime, lighting regime, tank/cage dimensions)</td>
</tr>
<tr>
<td></td>
<td>Manipulation of organisms/samples: Controlled manipulation as part of the study (e.g. exposure to a toxicant, environmental perturbation or dietary manipulation etc), dissection of a specific organ/tissue, capture/sampling means and procedures (e.g. netted, electrically stunned, anaesthetised, razor cut), reason for capture, other capture parameters (e.g. handling/stress aspects, time to capture, air exposure duration)</td>
</tr>
<tr>
<td></td>
<td>Sample handling and storage/preservation: Procedure and means (e.g. snap frozen and stored in liquid nitrogen or on dry ice, sample container material), reasons for storage/preservation, temperature and duration of storage</td>
</tr>
</tbody>
</table>

Reporting guidelines in metabolomics

The final stage of a metabolomics project is to disseminate the findings, either internally through technical reports, or externally through peer-reviewed publication. Whichever route is taken, it is advised that researcher’s follow to the best of their abilities a number of ‘minimum reporting standards’ which have been developed over the past decade by the wider metabolomics community (the Metabolomics Standards Initiative [http://www.metabolomics-msi.org/]). These readily available standards are ‘highly recommended’ guidelines for the reporting of various aspects of a metabolomics project, and provide a framework to ensure scientific rigour, allow study replication, support data sharing, and enable a better-informed process of assessment.
and interpretation. In accordance with other biological science investigations, typical areas of focus include detailed descriptions of the biological sample/s involved in the study, descriptions of the environment/s involved in the study, and descriptions of biologically-relevant processes involved in the study (Table III) (see Morrison et al. 2007). Additional aspects to consider when reporting metabolomics-derived information include prescribing to the use of specific standard terms (ontology), providing particular details of a wide range of platform-specific instrumental parameters and data processing methods (computational, bioinformatics, statistical), quality-scoring metabolite identifications, and, among others, participating in the standards initiative to advance the future of the field by reporting and exchanging various levels of metadata with others. We strongly advise that all metabolomics researchers, from aspiring to seasoned investigators, become familiar with the recommended reporting guidelines for each component of a metabolomics project (summarised in Figure 11).

![Figure 11. Overview of the metabolomics workflow showing the different components for which the Metabolomics Standards Initiative (MSI) have developed recommended minimum reporting standards (modified from Goodacre 2014).](image)

**Incorporating metabolomics**

Two main avenues exist for researchers who wish to conduct metabolomics investigations, or add a metabolomics component to an existing research project. There are a number of commercial metabolomics laboratories worldwide that offer streamlined services. Core facilities at various universities and centres house a combination of infrastructure and expertise to carry out a range of advanced metabolomics studies. These organisations can provide excellent support from consultation on experimental design to data analysis and interpretation of results. Inevitably, significant costs are usually associated with such commercial services. Alternatively, access to metabolomics facilities can be gained through academic institutions for substantially reduced charges based on collaborative agreements. For scientists wanting to conduct metabolomics research for the first time, it is important to note that running a successful metabolomics project requires an adequate experience in chemistry, statistics, bioinformatics and the advice from a

---

**Figure 11.** Overview of the metabolomics workflow showing the different components for which the Metabolomics Standards Initiative (MSI) have developed recommended minimum reporting standards (modified from Goodacre 2014).
metabolomics expert on hand. For researchers with sufficient chemistry knowledge and access to appropriate equipment and facilities, extraction and initial identification of metabolites may be relatively easy. However, there are some specific constraints in sample collection/preparation and experimental design that need to be considered. In addition, the bioinformatics required for data analysis and interpretation are significantly complex and may require the involvement of a bioinformatics expert. Regardless of the approach, we suggest that new metabolomics projects incorporate the appropriate expertise from the start. Furthermore, we urge scientists to give appropriate consideration to the expected results and implications of findings, since this approach is exploratory by nature.

Summary
In summary, metabolomics is a relatively new approach that has the potential to make a huge contribution to the field of aquaculture. With a wide range of analytical platforms available today and the rapidly evolving computational and bioinformatics capabilities, we are likely to see a growing number of studies using metabolomics in all aspects of cultivating aquatic organisms. However, it is important to be aware of the potential limitations of this approach, especially with regard to sensitivity to external influences during sample collection and complex bioinformatics procedures required to obtain meaningful biological interpretations. We are still at an early stage in the application of metabolomics in aquaculture, and it is envisioned that more streamlined procedures and strategies will be generated in the coming years to facilitate implementation of this powerful approach. Some of those advances will involve the development of extensive metabolite biomarker libraries, easy-to-use bioinformatics packages, small robust analytical platforms for use in the field, and improvements in analytical sensitivities and metabolite coverage. But more importantly, our future challenge will no doubt be to translate the clear potential of this approach into practical solutions to significantly improve the commercial aquaculture sector.

Aknowledgements
We are thankful for the valuable comments of J. Robertson, M. Rohan, and K. Campbell on an earlier version of this manuscript. We also acknowledge support from an Auckland University of Technology Vice Chancellor Doctoral Scholarship to T. Young under the supervision of A.C. Alfaro.
REFERENCES


T. Young and A. C. Alfaro


