



Heriot-Watt University  
Research Gateway

## Metabolomics

### Citation for published version:

Foito, A & Stewart, D 2018, 'Metabolomics: A high-throughput screen for biochemical and bioactivity diversity in plants and crops', *Current Pharmaceutical Design*, vol. 24, no. 19, pp. 2043-2054.  
<https://doi.org/10.2174/1381612824666180515125926>

### Digital Object Identifier (DOI):

[10.2174/1381612824666180515125926](https://doi.org/10.2174/1381612824666180515125926)

### Link:

[Link to publication record in Heriot-Watt Research Portal](#)

### Document Version:

Peer reviewed version

### Published In:

Current Pharmaceutical Design

### General rights

Copyright for the publications made accessible via Heriot-Watt Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

### Take down policy

Heriot-Watt University has made every reasonable effort to ensure that the content in Heriot-Watt Research Portal complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [open.access@hw.ac.uk](mailto:open.access@hw.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.

1 **Metabolomics: a high-throughput screen for biochemical and bioactivity diversity in**  
2 **plants and crops**

3 Alexandre Foito<sup>\*a</sup>, Derek Stewart<sup>a,b,c</sup>

4 *<sup>a</sup>Environmental and Biochemical Sciences Group, The James Hutton Institute,*  
5 *Invergowrie, Scotland, UK; <sup>b</sup>Institute of Mechanical Process and Energy Engineering,*  
6 *School of Engineering and Physical Sciences, Heriot Watt University, Edinburgh,*  
7 *Scotland, UK; <sup>b</sup>NIBIO, Norwegian Institute of Bioeconomy Research, Pb115, NO-1431*  
8 *Ås, Norway.*

9 **Abstract:** Plants and crops contain a staggering diversity of compounds, many of  
10 which have pharmacological activity towards a variety of diseases. These properties  
11 have been exploited by traditional and modern medicine providing important sources of  
12 healthcare to this day. The contribution of natural products (such as plant-derived) to  
13 the modern pharmacopeia is indeed significant; however, the process of identifying  
14 novel bioactive compounds from biological sources has been a central challenge in the  
15 discovery of natural products. The resolution of these challenges relied extensively on  
16 the use of hyphenated mass spectrometry (MS) and nuclear magnetic resonance  
17 (NMR)-based analytical technologies for the structural elucidation and annotation of  
18 novel compounds. Technical developments in instrumentation and data processing have  
19 fostered the development of the field of metabolomics which provides a wealth of tools  
20 with huge potential for application in the process of drug/bioactive discovery from plant  
21 tissues. This manuscript provides an overview of the metabolomics toolbox available for  
22 the discovery of novel bioactive compounds and the integration of these tools in the  
23 bioprospection and drug discovery workflows.

24 **Keywords:** Metabolomics, Plants, Bioprospection, Drug Discovery, Bioactivity, Chemical  
25 Diversity

26

27 \*Address correspondence to this author at the Environmental and Biochemical Sciences  
28 Group, The James Hutton Institute, Invergowrie, Dundee, DD2 5DA Scotland, UK; Tel: ++  
29 44 1382 568893; E-mails: Alexandre.Foito@hutton.ac.uk

30

31

## 32 **1. INTRODUCTION**

33 The discovery of pharmacologically active compounds has been a long standing goal of  
34 medicine since its inception. Our hunter-gatherer ancestors sought to use the diverse products  
35 found in nature to provide relief and cure from their ailments and this approach remained  
36 unchanged for millennia. Indeed, early pharmacologists focused on natural substances,  
37 mainly extracts derived from plants. Interestingly, many pharmacologically active natural  
38 products (NPs) used in modern medicine began as complex mixtures of compounds similar to  
39 our ancestor's approaches. However, modern approaches aimed at identifying and purifying  
40 the active NP(s) present within these mixtures. This process has often resulted in the  
41 development of novel pharmacological ingredients from a wide variety of biological  
42 materials. An analysis of the rate of novel chemical entity (NCE) approvals from the period  
43 of 1981 to 2014 demonstrates that the field of NPs is still highly represented with a mean of  
44  $34\pm 9\%$  in the last 15 years [1]. From the number of NPs developed a significant proportion  
45 were derived from plant extracts [2]. Indeed, it has been repeatedly shown that plants contain  
46 a diverse set of compounds and that some of these have shown pharmacological activity  
47 towards a variety of diseases [2]. Finding a set of bioactive compounds among the sheer  
48 diversity of the compounds present in plant tissues, estimated at 200,000 different metabolites  
49 [3], comprises a significant challenge. Furthermore, purification may be hindered by the  
50 limitation of biological material available and/or synergistic/antagonistic effects that might be  
51 lost as extracts attain higher degrees of purity. Developments in the field of metabolomics in  
52 the past 20 years have facilitated improvements in the accuracy, sensitivity, throughput and  
53 data analysis capacity and have contributed significantly towards the field of discovery of  
54 new bioactive biomolecules from plants. This review addresses the developments in  
55 metabolomics in the context of bioactive discovery.

## 56 **2. SOURCES OF NEW DRUGS AND BIOACTIVE COMPOUNDS**

57 Novel bioactive compounds can be classified according to their origin. For example, novel  
58 compounds are often characterised as NPs, botanical drugs (defined mixture), NP derivatives,  
59 synthetic drugs and mimics of NPs [1]. The early 2000s saw the development of high-  
60 throughput screening (HTS) based on biochemical targets which allowed many pure  
61 compounds to be screened in a relatively short time. The combination of HTS with modern  
62 combinatorial chemistry opened up the possibility of generating libraries comprising millions  
63 of compounds and screening these for efficacy in a relatively short period of time. However,

64 the enormous scale of this approach did not translate into a significant increase in the number  
65 novel drugs developed. In fact, the number of *de novo* FDA approved compounds developed  
66 has been fairly modest, with Sorafenib, an anticancer drug, the most successful output [1].  
67 The limited success has been attributed to the randomness of the approaches which may  
68 result in a lack of NCEs with high diversity [4].

69 This resulted in a shift toward smaller-scale compound collections based on NP scaffolds and  
70 more focused in chemical diversity [2]. With the growing realisation that the chemical  
71 diversity present in NPs has had enormous success with respect to the development of novel  
72 drugs, in comparison with high-throughput synthetic chemistry approaches, there has been  
73 resurgence in the application of natural chemical diversity to the field of drug discovery [5-  
74 7]. Indeed, NP molecules tend to have a higher number of chiral centers, oxygen atoms and a  
75 lower number of sulfur, nitrogen and halogen-containing groups [6] as well as a lower ratio  
76 of aromatic ring atoms to total heavy atoms in comparison with synthetic compounds [8].  
77 Since its inception in the middle of the 19th century, the history of modern synthetic  
78 chemistry represents a small fraction of time compared with the millennia associated with  
79 natural chemical evolution and diversity. The production of some of these substances has  
80 often been a determinant factor whether organisms would survive in hostile environments  
81 often providing a significant advantage over competing organisms. This explains the  
82 relatively higher percentage of biologically active substances from natural sources when  
83 compared with substances obtained from purely synthetic sources [4]. Approximately 25% of  
84 about 1 million NPs are biologically active (i.e. show toxicity or a positive activity) with  
85 approximately 60% of these products derived from plants and the remaining by microbes and  
86 animals to lesser extent [9]. NPs to this day remain important sources of biologically active  
87 compounds and during the period of 1981-2014 approximately 60% of NCEs approved are  
88 represented by NPs, NP derivatives and mimics of NPs [1]. Indeed, NPs may be isolated from  
89 diverse biological sources such as, animal, fungi, bacteria, marine organisms and plants and  
90 there has been an increasing realisation of the contribution of global biodiversity towards  
91 affordable therapeutical solutions for the majority of the population in the world [7].

### 92 **3. BIODIVERSITY AND NATURAL PRODUCT DISCOVERY**

93 Global biodiversity has great potential for providing novel sources of NPs, and indeed it has  
94 provided a significant contribution to the repertoire of pharmaceutical products currently  
95 available. Despite many successful examples of NPs used for the development of novel  
96 pharmaceuticals, the enormous biodiversity present remains largely untapped. Many

97 successful pharmaceutical products such as penicillin and statins derive from Fungi but these  
98 comprise a total of approximately 100,000 known species although the total number is  
99 estimated to be ten-fold higher [10]. Prokarya and Animalia have also provided important  
100 novel antimicrobial agents such as chloramphenicol and streptomycin from prokaryotes [9]  
101 and Hadruirin from animal sources [11]. Animalia species numbers are estimated at  
102 7,770,000 and 2 150,000 species present in terrestrial and marine environments, respectively,  
103 [12] whereas estimates for number of Prokaryote species remain unknowable in any scale in  
104 the environment [13]. Marine eukaryotic organisms are also being increasingly seen as a  
105 potential source of NPs with sponges, macroalgae and seaweeds reported to provide a wealth  
106 of NPs with anticancer and antiviral properties [14, 15]. The total number of marine species  
107 is estimated at 0.7-1.0 million species including 226,000 Eukariotic species which inhabit an  
108 extreme diversity of environments with respect to pressure, salinity and temperature [16].

109 Interestingly, the total number of described plant species pales in comparison with the  
110 aforementioned groups with the estimate of the total number of plant species at about  
111 320,000, although some studies estimate this number to be as high as 450,000 [17]. Despite  
112 this, the distribution of drugs based on NPs across different stages of drug development and  
113 according to their biological origin (Plant, Bacterial, Fungal, Animal or semi-synthetic)  
114 highlights that plants are important sources of NPs for biomedical applications [2]. Indeed,  
115 out of the 225 drugs derived from NPs undergoing development in 2008, 108 were from plant  
116 origin alone which significantly outnumbers any of the other origin classes (25, 7, 24 and 61  
117 for bacterial, fungal, animal and semisynthetic origin, respectively) [2]. Indeed there are  
118 many well-known examples of plant-derived drugs which include anticancer-drugs most  
119 notably paclitaxel (taxol), a diterpenoid compound isolated from the bark of *Taxus brevifolia*  
120 etoposide (vepesid) partly synthesized from the a lignin isolated from *Podophyllum peltatum*  
121 (podophyllotoxin) and irinotecan which is a derivative from an alkaloid extracted from  
122 *Camptotheca acuminata* [1, 18].

#### 123 **4. FROM TRADITIONAL MEDICINE TO THE DEVELOPMENT OF NOVEL** 124 **DRUGS**

125 Plants have been seen not only as a source of energy and nutrition but have been exploited in  
126 traditional medicine throughout humanity's history. The oldest records of usage of medicinal  
127 plants dates back to 2400 B.C. from Mesopotamia but more recent records are also available  
128 from Ancient Egypt and China [7]. Indeed, NPs particularly from plants have been at the  
129 forefront of medicine since ancient times until the isolation of morphine by Sertürner around

130 1804 [7]. During that period, a large wealth of ethnobotanical knowledge was acquired over  
131 countless generations and constituted the foundation of traditional and indigenous peoples  
132 medicinal knowledge. Only with the advent of the scientific revolution in the 18th century  
133 has medicine shifted from the application of herbal crude preparations to the isolation and  
134 development of the active ingredients within the plant tissues. Despite unprecedented  
135 developments in modern medicine it is still worth noting that a large proportion of the world  
136 population still relies on traditional remedies (often plant-based) as the main source of  
137 healthcare [19, 20]. With the continuous growth in the use of medicinal herbal products,  
138 issues concerning their safe use are being increasingly highlighted [20] with the lack of  
139 pharmacological and clinical data of such products seen as a major impediment to their  
140 integration into conventional medical practices [21]. Therefore, a number of studies have  
141 been focused on providing such data such as that of the clinical studies of *Rhodiola rosea* as a  
142 therapeutic agent for depression related conditions [22]. These examples highlight not only  
143 the therapeutic potential of botanical products but also the need to investigate further the  
144 mechanisms of action, the bioactive ingredients and the safe use of such products. The  
145 discovery of novel ingredients in traditional medicine often involved direct experimentation  
146 of prospective medicines in humans utilising crude preparations of biological tissues with  
147 minimal considerations regarding scientific methodology. This approach raises unacceptable  
148 safety and ethical issues, and promises very little with respect to significant success rates by  
149 modern standards. However, there were no other alternatives available until the last 200  
150 years.

## 151 **5. BIOPROSPECTION OF NATURAL PRODUCTS**

152 Several technological, medical and scientific advances eventually led to a shift towards  
153 isolation of pure compounds. Many such botanical products have been targeted with the goal  
154 of discovering commercially valuable bioactive NPs in a process known as natural product  
155 bioprospection [23]. With an estimate of plant/crop diversity of 200,000 compounds [3], many  
156 of which possess bioactive properties, the teasing out of specific bioactivity becomes  
157 incredibly difficult. Traditionally, the discovery of bioactive NPs from biological material  
158 involved the collection and screening of plant material utilising biological assays. From this  
159 step onwards there are two leading approaches utilised in the discovery of bioactivity relying  
160 on bioassay-guided fractionation and pure compound screening [24, 25] (Figure 1). The  
161 bioassay-guided fractionation approach seeks to reduce the complexity of the extracts down  
162 to its single bioactive components by utilising repeated series of fractionation and bioassay

163 testing until the successful isolation of an active NP [24]. Several fractionation  
164 methodologies are available for separating crude extracts into fractions of reduced  
165 complexity. Solvent partitioning is a popular approach in which samples are extracted  
166 sequentially using solvents of differing polarity and, while it requires minimal capital  
167 investment, it provides limited separation capacity, is labour intensive and requires the  
168 extensive use of organic solvents [25] which may be problematic for bioassays. Liquid-  
169 chromatography (LC) separation methods which include LC, High performance LC (HPLC),  
170 Ultra performance LC (UPLC) and counter-current chromatography are popular methods of  
171 offline and online separation of complex mixtures [26]. Although challenging, it is possible  
172 to couple LC techniques to post-column biochemical assays and parallel chemical analysis  
173 which may aid the identification of bioactive compounds from crude extracts [27]. Gel-  
174 chromatography has also been modestly used in applications where targeted molecules  
175 include proteins and other biopolymers and while separation efficiency is modest, the  
176 aqueous composition is often compatible with bioassays [26]. Thin layer chromatography  
177 (TLC) is also a popular separation technique particularly when coupled to bioassays and  
178 allows the separation and bioassay to occur under the same experimental conditions within  
179 the same plate thereby allowing several samples to be analysed simultaneously utilising less  
180 solvent. On the other hand when compared with liquid chromatographic methods it does have  
181 lower separation efficiency [28]. Affinity chromatographic methods provide an additional  
182 targeted method which relies on the binding of biomolecules with artificially generated  
183 antibodies or synthetic ligands allowing the enrichment of biomolecules with targeted traits in  
184 the fractions [26]. Bioassays are often categorised as isolated molecular target assays, cell-  
185 free multicomponent assays or cell-based assays. The former is usually based on assessing  
186 specific interactions between purified enzymes (eg. proteases, kinases) and pure compounds  
187 whereas the second includes assays on activities derived from cell extracts, cell membranes  
188 or reconstituted signalling pathways [29]. The later utilises cell-models and is generally based  
189 on the assays which monitor reporter genes or phenotypic responses resulting from intact  
190 cellular processes [29]. Regardless of the bioassay strategy being utilised, the bioassay-  
191 guided “offline” fractionation process is an iterative process that ultimately depends on the  
192 availability of preparative-scale analytical methods that allow for the resolution of complex  
193 mixtures of primary and secondary metabolites that are typical from the source organisms  
194 and result in a purified natural compound ideally as a single chemical entity with a specific  
195 activity [30]. This approach allows fractions with no bioactivity to be excluded and will  
196 generate bioactive fractions of reduced chemical complexity, however it requires an abundant

197 source of biological material and may often lead to the isolation and identification of known  
198 bioactive compounds in a process known as replication [25]. A literature survey indicates that  
199 the average number of isolation steps involved in the purification of NPs is less than three  
200 [30] and the authors suggested that compounds that are often present in low concentrations  
201 are rarely pursued due to their arduous isolation process. Indeed, the effort required to purify  
202 a single chemical entity depends on the concentration of the NP, the physico-chemical  
203 properties (eg. solubility and crystallisation), the use of appropriate selectivity characteristic  
204 of the separation technique chosen for the target NP, and the nature of the matrix of the crude  
205 extract [30].

206 An alternative approach aims to isolate and elucidate the structure of the majority of the  
207 secondary metabolites present in the crude extract and subsequently test pure compounds (ie  
208 standards) for bioactivity (Figure 1). The nature of this approach allows the selection of  
209 extracts which contain compounds which are not already present in the libraries of pure  
210 compounds with recorded bioactivities [24]. Both the cited approaches should aim to identify  
211 as many NPs present in the extracts at the earliest opportunity in order to avoid redundant  
212 work in a process known dereplication [24, 25]. The most common methodologies used for  
213 this purpose often combine a separation step, usually a chromatography-based separation  
214 which attempts to resolve crude extracts into separate metabolites, linked to a structure  
215 elucidation step which would typically involve mass-spectrometry (MS) or nuclear magnetic  
216 resonance (NMR) technologies [24]. This type of approach is well aligned with the prevalent  
217 paradigm of drug development of the past few decades in which single proteins whose  
218 inhibition is likely to be involved in the treatment of a targeted disease are identified and  
219 subsequently are tested against large libraries of small-molecules in order to identify lead  
220 molecules which will be further tested in appropriate and complex model (*ex vivo* an *in vivo*)  
221 [31]. The discovery of novel drugs according to this approach often aims to identify  
222 compounds that would act/bind specific targets. However, advances in genomics and  
223 proteomics highlight natural variation present in individuals, which might be associated with  
224 lack of global drug effectiveness due to individual mutations/polymorphisms that are not  
225 present in the model target protein screened. In addition, individual variation in gut  
226 microbiome content and composition can modulate the response to drug therapies [32].  
227 Furthermore, when addressing antimicrobial/antiviral drugs a single target approach is  
228 generally more vulnerable to the development of resistance to novel medicines. Indeed, a  
229 combination of antiretroviral drugs that target different viral proteins is often found to be



230 more effective [33]. As a result, novel paradigms are arising, which include system-level  
231 thinking approaches such as computational multitarget screening [31] and the development of  
232 high-throughput combination screening platforms [34]. The rekindled interest in system-level  
233 thinking has been aided by the development of high-throughput ‘omics’ platforms that are  
234 able to generate large amounts of reproducible molecular data and large curated databases  
235 integrating data from different disciplines and the large development of interdisciplinary  
236 research with the goal of integrating and mining these datasets [35]. As the paradigm in drug  
237 discovery shifts toward system-level approaches with respect to the screening of multiple  
238 targets and multiple effectors, so must the process of bioactive NP discovery from plant  
239 material. Indeed, in many cases, it is the combination of metabolites rather than a unique  
240 chemical entity that is responsible for the bioactive properties of crude preparations of  
241 biological material. For example, it has been observed that testing a mixture of the two major  
242 bioactive alkylphthalides (*Z*-ligustilide and *n*-butylidenephthalide) identified through bioassay-  
243 guided fractionation of *Angelica sinensis* roots yields lower anti-endothelial activities when  
244 compared to the crude volatile oil obtained which indicates potential synergistic effects with  
245 unknown compounds [36]. Furthermore, Skirycz et al (2016) cites the example of the use of  
246 flowers and fruits of *Psychotria colorata* in native American traditional medicine to treat  
247 pain. The analgesic effect was confirmed in rodents [37], and several alkaloids in  
248 combination were identified as the active ingredients [23]. Interestingly, the major  
249 compounds identified include hodgkinsine and psychotridine, which are opioid agonist and a  
250 N-methyl-D-aspartate receptor antagonist, respectively [38, 39]. The complex mixture of  
251 agonist-antagonist compounds in crude extracts may allow additive, synergistic and  
252 antagonistic interactions between metabolites that underpin the medicinal properties of plant  
253 extracts [23]. The complexity of these metabolite interactions and their contribution towards  
254 the bioactivity of plant extracts is in contrast to the standard aforementioned approach used in  
255 NP discovery (e.g. bioassay guided isolation of pure compounds). Indeed, it is not uncommon  
256 for the fractionation of bioactive crude extracts to generate non-bioactive extracts [36, 40].  
257 Consequently, many advocate that a shift away from less reductionist approaches towards  
258 systems biology approaches is warranted [23, 41-43]. The fields of genomics, proteomics and  
259 metabolomics have experienced great advances in the last 20 years which have allowed  
260 system biology approaches to become a reality. While the application of genomics and  
261 proteomics to the field of drug discovery has been a relatively recent event, the use of some  
262 analytical tools utilised in metabolomics such as hyphenated MS and NMR have been central  
263 in the discovery of NPs since the 1970s. However, technical developments in instrumentation

264 and post-separation and acquisition data analysis in the past two decades have increased the  
265 throughput of such analytical tools and resulted in naming the field ‘metabolomics’.

## 266 **6. METABOLOMICS: BRAVE NEW (DATA) WORLD**

267 The term metabolomics was coined in 1998 by Oliver et al and refers to the study of the  
268 quantitative and qualitative collection of all the metabolites in a given cell or tissue [44].  
269 Since then, there has been a growing interest in applying metabolomics methods particularly  
270 in the field of plant sciences [45-48]. Applications of metabolomics usually include NMR-  
271 and MS-based approaches often coupled to chromatographic separations and are diverse in  
272 scope. These include studies with the goals of understanding how metabolic networks  
273 respond in plants exposed to environmental stresses [49-51], monitoring unintended  
274 metabolic effects in genetically-modified crops [52-54], discovery of novel NPs [55, 56]  
275 amongst many others. Metabolomics approaches have been reviewed extensively elsewhere  
276 [23, 45, 46, 57] and the nature of these approaches has often been categorised either as a  
277 targeted, metabolite profiling or metabolite fingerprinting. The first aims to extract and  
278 quantify a limited predetermined set of metabolites by optimising linear ranges, efficiency  
279 and robustness of extraction, and stability of the target compounds in detriment of non-target  
280 compounds. Examples of this type of approach include the quantification of polyphenols in  
281 berries [58], identification and quantification of loline-type alkaloids in endophyte-infected  
282 grasses of the genera *Lolium* and *Festuca* [59] and quantification of glucosinolates from  
283 *Camelina sativa* seeds [60]. However, the main limitation of the targeted approach is that it  
284 provides no information regarding non-targeted metabolites which may be pathway related or  
285 relevant for the objective of the study [61].

286 Metabolite profiling, which can be defined as analysis of a group of selected pre-defined  
287 metabolites (e.g. polyphenols, carbohydrates or fatty acids), is also affected by the limitations  
288 cited above, however it differs from targeted analysis due to including a broader range of  
289 metabolites and may include known and unknown metabolites and often providing semi-  
290 quantitative rather than absolute quantitative information [23,62,63]. There is an extensive  
291 body of literature documenting the application of this type of approach but some examples  
292 include the monitoring of primary metabolites under environmental stress conditions  
293 [50,64,65] monitoring primary and secondary metabolites in genetically modified crops [54].  
294 Conversely, truly untargeted approaches attempt to be as comprehensive as possible while  
295 minimising bias towards any classes of compounds. It is generally regarded that there is no  
296 “one method to rule them all” that allows a comprehensive coverage of the entire

297 metabolome, and that the choice of sample preparation, extraction and analytical tool  
298 inevitably introduces bias towards certain classes of metabolites [66]. This limitation is often  
299 overcome by combining multiple strategies for analysis that complement each other's  
300 weaknesses.

301 Metabolite fingerprinting is an untargeted approach that focuses in the recognition of patterns  
302 in spectra (usually NMR and MS) collected from the analysis of sample extracts without  
303 attempting to quantify or identify specific compounds [62]. This is often achieved by  
304 applying multivariate pattern recognition tools such as principal component analysis (PCA)  
305 and discriminant function analysis (DFA) in order to identify metabolic features that  
306 differentiate groups of samples [62]. Multivariate analysis has been associated with studies of  
307 biomarker discovery and disease diagnostic [67] and it has also been used to distinguish  
308 different plant species and ecotypes [68, 69]. Despite the global, unbiased nature of this  
309 approach (as discussed above) the initial workflows did not include automated metabolite  
310 annotation, and therefore, annotation of metabolites was usually performed manually in a  
311 time-intensive process, which was usually reserved to the compounds of interest such as  
312 potential biomarkers. This approach is still currently relevant; however, advances in  
313 metabolomics workflows are blurring the once distinct division between profiling and  
314 fingerprinting methods. Indeed the past 15 years have seen great advancements in pre-  
315 processing and processing of data particularly from MS-based technologies coupled to  
316 chromatographic separation. This can be illustrated by the wealth of pre-processing and  
317 processing tools available both from commercial and instrument vendors but also as open-  
318 source tools. Some examples of these tools include XCMS [70], XCMS online [71],  
319 TAGfinder [72], Sieve (Thermo Fisher Scientific Inc) [73], MassHunter Profinder (Agilent  
320 Technologies) MarkerLynx (Waters), Progenesis QI (Nonlinear Dynamics) [74], Marker  
321 View (AB Sciex), MZmine 2 [75], Metalign [76] amongst many others.

322 These tools are often incorporated in workflows which allow detected features to be  
323 processed, molecular formulas to be generated, and subsequently, queried against metabolite  
324 databases from various sources (e.g. online, commercial, in-house databases). Indeed the  
325 integration of automated annotation in untargeted analysis can be seen as a hybrid of  
326 metabolite fingerprinting and metabolite profiling particularly when tailored databases are  
327 utilised. For example, Skogerson et al (2011) developed an automated peak annotation and  
328 database system for the analysis of complex volatile mixtures by gas-chromatography  
329 coupled to mass spectrometry (GC-MS) allowing the annotation of large datasets comprising

330 of hundreds to thousands of samples [77]. The process of automated annotation is highly  
331 dependent on the quality of source database utilized, and although targeted biological  
332 databases often provide promising results, it is not unusual for compounds to be wrongly  
333 annotated. Therefore, the metabolomics community encourages studies to categorise the  
334 certainty level of compound identification according to a set of guidelines based on the  
335 degree of evidence provided [78, 79].

### 336 **6.1. NMR-based metabolomics**

337 NMR spectroscopy allows molecules containing one or more atoms with a non-zero  
338 magnetic field to be detected that include  $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{14}\text{N}$ ,  $^{15}\text{N}$  and  $^{31}\text{P}$  [80, 81]. This technique  
339 allows the characterisation of metabolites with at least one NMR signal with respect to their  
340 frequency (chemical shift), intensity and magnetic relaxation properties depending on the  
341 chemical environment that the nucleus occupies [80]. NMR spectra, via varied pulse  
342 techniques, often provide structural information which is essential for the identification of  
343 unknown compounds and it is used routinely in NP discovery. Additionally, it is possible to  
344 obtain quantitative data from NMR spectra which may be of interest to several types of  
345 metabolomics approaches. Indeed, NMR has been used in many metabolomics approaches  
346 which include the metabolite profiling of opium poppy [56] and metabolic flux analysis of  
347 linseed embryos [82]. NMR has been extensively used in fingerprinting approaches as spectra  
348 are often too convoluted [61]; however, many groups have addressed convolution by  
349 coupling liquid-chromatography separation prior NMR analysis [83]. Despite its applications,  
350 LC-NMR approaches have inherent limitations such as significant capital costs, solvent  
351 suppression [83, 84] and perhaps, more importantly, the relatively low sensitivity of NMR  
352 which often limits the analysis to the most abundant compounds [61, 80, 81, 83] (Figure 2).

### 353 **6.2. Mass spectrometry-based metabolomics**

354 Mass spectrometry has become the *de rigueur* approach used in metabolomics due to its high  
355 sensitivity and wide metabolite coverage. Overall, this technique relies on the ionisation of  
356 chemical compounds, most commonly utilising Electron Ionisation (EI), Electrospray  
357 Ionisation (ESI) and Atmospheric pressure chemical ionization (APCI), generating charged  
358 molecules or molecule fragments which are subsequently measured with respect to their  
359 mass-to-charge ratio ( $m/z$ ) and abundance. Often crude extracts are analysed through direct  
360 infusion-mass spectrometry (DI-MS) generating a single mass spectrum representing the  
361 global metabolite profile of the sample. Utilising this approach it is possible to achieve a very

362 high-throughput metabolite fingerprinting approach (first pass) which can be applied in a  
363 variety of plant tissues [85, 86]. As there is no separation prior to analysis, this type of  
364 approach is usually carried out on high mass accuracy instruments as it allows the distinction  
365 between some compounds with the same nominal mass [66]. However, this technique does  
366 not allow isomers to be differentiated due to their identical molecular masses [87].  
367 Furthermore, as multiple compounds are analysed simultaneously in the mass spectrometer  
368 this technique is particularly vulnerable to ion co-suppression effects in which non-volatile  
369 compounds influence the ionization and transfer of metabolites from liquid to gas phase in an  
370 ESI source leading to a misrepresentation of the signals present in the mass spectrum [66]. In  
371 order to prevent and minimize these effects, metabolomics approaches often resort to  
372 coupling mass spectrometers with separation techniques such as GC, LC or electrophoretic  
373 separation such as capillary electrophoresis [61] (Figure 2).

374 GC-MS has seen extensive use in plant metabolite profiling of primary metabolites ranging  
375 from studying plant metabolic responses to abiotic stress [50], analysis of the volatile profiles  
376 of *Allium* samples [88], to the mapping quantitative trait loci for metabolites in tomato fruits  
377 [89]. This approach requires that the target compounds are volatile in the operating range of  
378 the GC usually 100-320 °C which in many cases may require compounds to be extracted  
379 from biological material and subsequently derivatised to ensure their volatility. This  
380 requirement for volatility comes at a cost as only the compounds amenable to derivatisation  
381 may be analysed [90]. Despite this limitation, GC-MS applied in the context of metabolomics  
382 allows the simultaneous detection of several hundred different chemical compounds ranging  
383 from organic acids, sugars to amino acids fatty acids and aromatic amines [61]. Indeed, GC  
384 provides a high separation efficiency allowing compounds with mass spectral similarities  
385 such as isomers and enantiomers to be distinguished and produces reproducible retention  
386 times that allow comparison of results derived from different labs. Upon chromatographic  
387 separation, metabolites are usually ionised by electron ionisation (EI), which is regarded as a  
388 reproducible approach generally unaffected by ion co-suppression [87]. The reproducible  
389 nature of this approach allows mass spectrum and kovat retention indexes to be used in  
390 querying readily available MS databases containing hundreds of thousands of compounds  
391 such as the NIST standard reference database 14 ([https://www.nist.gov/srd/nist-standard-  
392 reference-database-1a-v14](https://www.nist.gov/srd/nist-standard-reference-database-1a-v14)), Golm Metabolome database [91] and FiehnLib [92].

393 Liquid-chromatography coupled to mass spectrometry (LC-MS) takes advantage of the  
394 separating power of liquid chromatography techniques that allow the study of high molecular

395 weight compounds that cannot be analysed by GC-MS [87]. When coupled to high resolution  
396 MS instruments this constitutes a powerful technique for analysis of a broad range of  
397 secondary metabolites. It is used broadly not only in metabolomics studies but also in natural  
398 product chemistry studies. Indeed, LC does not require analytes to be volatile or derivatised  
399 prior to LC analysis, and allows for a wide range of analyte polarity and mass range [90],  
400 consequently, sample preparation is generally less complex. Following chromatographic  
401 separation the metabolites are ionised (generally by ESI or APCI) and mass spectrometers  
402 generally allow for positive and negative modes of operation. Metabolites often ionise more  
403 efficiently in one mode over another so the monitoring of data in both positive and negative  
404 ionisation modes allows a more comprehensive coverage of the metabolome [66]. Generally,  
405 ionisation techniques utilised in LC-MS are softer ionisation techniques when compared to  
406 EI, which is commonly utilised in GC-MS approaches, and provide spectra much less  
407 dominated by fragments. The limited number of fragments observed in LC-ESI-MS often  
408 limits the discrimination of isomeric compounds so mass spectrometrists often design  
409 tandem-MS (MS<sub>n</sub>) methods which aim to fragment selected ions utilising collision-induced  
410 dissociation [93]. The fragmentation patterns observed can then be compared against  
411 fragmentation databases of known compounds. However, instrument variability and matrix  
412 effects (e.g. presence of co-eluting compounds or ionic salts) can impact the ionisation  
413 efficiency of the instrument and may also interfere with the chromatography which ultimately  
414 leads to a lack of reproducibility which contributes to the great challenge of comparing data  
415 generated from difference sources.

### 416 **6.3. Combining multiple analytical tools**

417 Despite the high sensitivity of MS-based methods, they are ultimately limited with respect to  
418 metabolite structure elucidation and it is largely impossible to determine unambiguously the  
419 structure of a metabolite utilising MS-based methodologies alone regardless of the accuracy  
420 of the instrument [94]. Indeed, MS-based metabolomics methods provided relatively low  
421 discovery rates and often result in false identifications with a relatively low number of  
422 metabolites which can be assigned an identity with a high degree of confidence [95].  
423 Ultimately it is the availability of authentic standards or the requirement to purify individual  
424 metabolites followed by NMR analysis for structure elucidation that remain the only  
425 strategies that allow high-confidence level in the identification of metabolites. Therefore it is  
426 not unusual for bioprospecting studies to incorporate results from these two analytical  
427 approaches in order to provide unambiguous identification of metabolites.

428

429 Interestingly, in face of the limitations associated with both MS and NMR-spectroscopy  
430 approaches there have been efforts to combine these techniques in an hyphenated approach.  
431 By interfacing LC with NMR and MS analysis (e.g. LC-NMR-MS) it is possible to capitalise  
432 on the strengths and mitigate the weaknesses of each of techniques providing comprehensive  
433 structural data that can aid in addressing the bottleneck of compound identification in extracts  
434 [96, 97]. This approach was applied to *E. coli* extracts and permitted the correct identification  
435 of a wide range of metabolites including amino acids, nucleic acids and carbohydrates [98].  
436 Currently, there is no single procedure that allows the quantification and identification of the  
437 entire metabolite complement in any tissue. The choice of sampling and extraction procedure  
438 as well as the analytical tool applied will inadvertently introduce bias in the analysis, thus the  
439 methodologies should be adjusted in order to obtain the most relevant and comprehensive  
440 metabolite coverage possible for the study.

## 441 **7. METABOLOMICS FOR SCREENING BIOCHEMICAL DIVERSITY AND** 442 **CHEMOTAXONOMY**

443 The biodiversity present in the plant kingdom is often reflected in large qualitative and  
444 quantitative differences in metabolite composition in plant extracts. Indeed a plant population  
445 may display not only presence-absence polymorphism with regards to secondary metabolite  
446 composition but also quantitative differences between the concentrations of secondary  
447 metabolites [99]. There is indeed enormous qualitative diversity present in plant tissues and a  
448 large fraction is accounted from differential modification of common backbone structures  
449 which may have resulted from the evolution of enzymes with the same product specificity but  
450 catalysing different reactions [100]. A notable example of this includes flavonoids such as  
451 quercetin and their respective different glycoside derivatives which number over 130 [101].  
452 Conversely, there are enzymes with poor substrate specificity that catalyse a similar reaction  
453 in more than a single extract, for example recombinant *Vitis labrusca* flavonoid 3-O-  
454 glucosyltransferase was found to glycosylate both peonidin, malvidin and isorhamnetin [102].  
455 Plant secondary metabolites are often involved in response to (a)biotic stresses and it has  
456 been recently demonstrated that plants can undergo rapid qualitative and quantitative  
457 evolution of secondary metabolites as a response to herbivore pressure [103, 104].  
458 Interestingly, it can be observed that when plants are introduced into new habitats that it may  
459 favour the evolution of new secondary metabolite compositions compared with its native  
460 habitat [105, 106]. It has been suggested that synthesis of secondary plant metabolites are

461 among the most evolvable traits as unlike primary metabolites their presence is not essential  
462 for plant survival under favourable ecological conditions [99]. The general lack of pleiotropic  
463 effects of mutations in the genes underlying the biosynthesis of secondary metabolites may  
464 allow these biochemical systems to be relatively free to evolve [107]. Under unfavourable  
465 ecological conditions on the other hand secondary metabolites may play important roles in  
466 chemical defense against natural plant enemies. According to the screening hypothesis, the  
467 high degree of diversity and evolvability present in some classes of compounds is explained  
468 by the increased chance to evolve the production of novel active compounds if they maintain  
469 a diverse and rapidly mutating complex of compounds [108], thus explaining the presence of  
470 a large complex of compounds with no known activity which could act as precursors for  
471 novel metabolites [99]. The variation present in the metabolite complement of plant tissues is  
472 ideally suited for analysis using metabolomics approaches which facilitates the development  
473 of chemotaxonomic studies with potential applications in drug discovery processes.

474 The potential to categorise different accession/species according to chemical composition  
475 provides tools enabling the selection of plant raw materials, which display the broadest  
476 chemical diversity for bioprospection [109]. Additionally, the combination of  
477 chemotaxonomic studies with bioactivity data might aid the identification of chemical classes  
478 qualitatively and quantitatively overrepresented in plant species with medicinal properties  
479 [110]. A chemotaxonomic approach based on metabolite profiling of a collection of diverse  
480 indigenous Korean plant species showed clear separation according to family [111]. The  
481 authors subsequently combined the metabolite dataset with a bioactivity dataset to mine for  
482 correlations between individual metabolite levels and bioactivity which resulted in the  
483 identification of 5 metabolites with statistically significant correlations with bioactivity [111].  
484 A similar strategy was employed to analyse nearly 90 species from the genus *Rhododendron*  
485 and correlate the metabolite profiles with plants extracts with antimicrobial and cytotoxicity  
486 datasets resulting in the identification of seven metabolites with potential antimicrobial  
487 properties [112].

## 488 **8. METABOLOMICS TOOLS FOR BIOPROSPECTING AND DEREPLICATION**

489 The analytical chemistry toolbox available for metabolomics approaches have historically  
490 been used in traditional NP discovery approaches particularly with respect to targeted  
491 metabolite profiling, structural elucidation, qualitative and quantitative assessment of purified  
492 NPs. Indeed there are numerous examples of bioprospection studies of plant extracts using  
493 MS and/or NMR tools [113-117]. However, the discovery of novel bioactive NPs faces



494 challenges with respect to increasing rates of rediscovery of known compounds as active  
495 ingredients (replication) [118-120], high rates of attrition (lost to follow-up) in the translation  
496 of bioactivity due to unknown modes of action [121, 122] and difficulties in screening  
497 synergistic/antagonistic effects between multiple active compounds. As result “business as  
498 usual” in the field in bioprospecting is no longer an option if one is to address these  
499 challenges efficiently. Many are now turning to system-level approaches where the high-  
500 throughput chemical characterization of complex compounds plays a central role and as result  
501 metabolomics-approaches are increasingly being used in multidisciplinary approaches to  
502 address these challenges. Metabolomics approaches are particularly suited for addressing the  
503 challenge of increasing rates of rediscovery of known compounds in the process of  
504 dereplication, which is the process of identifying already known bioactive molecules in  
505 extracts allowing researchers to focus their purification efforts on novel lead molecules and  
506 avoid redundant work [24]. As untargeted metabolomics approaches have gained attention  
507 due to high-throughput methods for the chemical characterisation of bioactive extracts, there  
508 has been increased development of workflows that take advantage of MS<sup>2</sup> spectral  
509 information and high mass accuracy for rapid annotation of metabolites aligned with  
510 multivariate statistical analysis allowing for quick dereplication of complex extracts [24, 118,  
511 119, 123].

512 When applied to the analysis of extracts from plants [124] or other organisms [120] these  
513 strategies can be successful and result in the identification of novel structures of the lead  
514 compounds thereby facilitating the pure compound screening approach (Figure 1) and avoid  
515 redundant work. The automated annotation of metabolomics datasets has been a central  
516 challenge in the field of metabolomics and has seen extensive development in the past  
517 decade. Therefore, continuing developments in this field and integration with known active  
518 ingredient libraries will ensure more efficient dereplication processes that can accelerate the  
519 process of bioprospection.

## 520 **9. METABOLOMICS FOR MODE OF ACTION AND DRUG METABOLISM**

521 The use of metabolomics tools has also seen application in the field of drug metabolism and  
522 toxicology and the elucidation the mode-of-action of lead compounds. The innovations  
523 observed in HTS have allowed a significant improvement in the level of automation of *in*  
524 *vitro* biochemical or cellular assays resulting in the systematic use in the early stages of drug  
525 discovery [96]. Isolated molecular target assays have been extensively employed in HTS  
526 approaches and often rely on the interaction between small molecules and protein targets.

527 NMR methods in particular have seen extensive application not only for small molecule  
528 structure determination but also analysis of target-ligand complexes [125, 126]. Indeed, the  
529 analysis of protein ligand-interactions is of central importance for understanding the  
530 regulation of biological functions of target proteins, cooperativity of ligand mechanisms and  
531 the development of novel drugs that modulate protein functions or inhibit protein interactions  
532 [126].

533 Alternatively, bioassays based on the use of microbial, animal and human cell culture  
534 metabolic models of human metabolism can provide an alternative source to animal  
535 experimentation in order to assess the effect drugs in putative human metabolism. In these  
536 instances metabolomics can be utilised as a tool to monitor cellular drug metabolism and  
537 assess the putative risks (or benefits) of drug metabolites in model organisms [96]. The use of  
538 metabolomics may be applied to rapidly evaluate the intracellular and extracellular  
539 metabolites and consequently evaluate cellular uptake, accumulation and metabolism of the  
540 tested drug/bioactive. The metabolomics approach when complemented with transcriptomic  
541 and proteomic data, can provide detailed mechanistic information of altered cellular  
542 processes [127, 128].

543 It was highlighted that the aforementioned reductionist approaches should be avoided when  
544 attempting to obtain evidence for the bioactivity of traditional medicines and that holistic  
545 approaches are advocated [41] which include *in vivo* approaches. Indeed, understanding the  
546 *in vivo* metabolic fate and mode of action of any potential drug candidate is of paramount  
547 importance for drug discovery and design [96]. The use of metabolomic-based approaches in  
548 the analysis of biofluids such as plasma or urine, has been explored widely within the context  
549 of the metabolic fate of potential drugs/bioactives [129, 130]. For example Sun et al (2009)  
550 investigated the excretion kinetic profile of acetaminophen in rats by profiling major drug  
551 metabolites (acetyl-L-cysteine acetaminophen) in urine utilising LC-MS and NMR  
552 metabolomics approaches. These results were correlated with endogenous oxidative stress-  
553 related metabolites leading the authors to postulate that the toxicity of the drug can be  
554 monitored by utilising metabolomics tools [131]. A combined LC-MS and NMR  
555 metabolomics approach has also been applied for profiling urine, plasma and tissue samples  
556 from penicillin-treated rats revealing a decrease in many host-gut microbiota metabolites and  
557 urinary conjugated metabolites (sulphate, glucuronide and glycine conjugates) suggesting a  
558 strong interaction between gut-microbiota and drug metabolism [132]. These approaches  
559 have strong potential for application in screening drug metabolism, toxicity and compliance

560 in the pre-clinical phases of the drug discovery process that are often associated with high  
561 levels of attrition.

562 Metabolomics approaches have been routinely applied in the field of biomarker discovery  
563 with particular application in the field of cancer research [133-135] with an extensive body of  
564 literature reporting the identification of candidate metabolic biomarkers [134,136,137].  
565 Indeed during cancer initiation and progression the cellular metabolism is altered allowing  
566 cells to increase anabolic synthesis, avoid apoptosis and adapt to low nitrogen and oxygen  
567 availability [138]. These extensive metabolic changes allow the discovery of multiple cancer  
568 biomarkers which can be monitored utilising metabolomics-based approaches [137,138]. In  
569 contrast with preclinical cancer studies where tissue and tumour samples are routinely  
570 collected, the regular monitoring of biomarkers in clinical setting often requires the use of  
571 minimally invasive low-cost methods for sample collection and analysis as they provide  
572 minimal discomfort to the patient and can be collected from control individuals [138].  
573 Metabolomics-based methodologies could therefore provide robust low-cost solutions for the  
574 monitoring of multiple biomarkers with in various tissues but also extracellular fluids leading  
575 to potential applications in the assessment of disease progression, prognosis, treatment  
576 efficacy and drug metabolic effects in clinical settings [134]. As result, there has been an  
577 increasing interest in the application of metabolomics in the realm of cancer research and  
578 treatment. For example, a global untargeted mass spectrometry-based metabolomics approach  
579 was used in a preoperative window clinical trial of metformin for the treatment of  
580 endometrial carcinoma with the goal of monitoring of the metabolic effects of the drug  
581 treatment in serum (pre- and post-treatment) and tumour tissue [139]. Samples were collected  
582 from twenty obese patients and 65% were found to respond to the treatment, reducing  
583 proliferation of tumours based on pre- and post-treatment evaluation. The application of  
584 metabolite profiling tools to the analysis of serum samples indicated that responders to  
585 metformin treatment experienced a greater induction of lipolysis compared to non-responders  
586 and this correlated with increased fatty acid oxidation and glycogen metabolism observed in  
587 the patient's tumour tissue [139]. This led the authors to suggest the potential use of  
588 metformin in the treatment of endometrial cancer in obese individuals. Subsequent work has  
589 been carried out in order to understand the mechanism of action behind the effect of  
590 metformin on cancer proliferation by applying a combined metabolomics and transcriptomics  
591 approach to study its effect over time in human-derived colon cancer LoVo cells. A total of  
592 47, 45 and 66 metabolites, ranging from carbohydrates to amino acids and lipids, were found

593 to be differentially expressed at three time points. The majority of the metabolites  
594 experienced an upregulation 8h following exposure and were subsequently downregulated  
595 after 24h of exposure [140]. In combination with the transcriptome analysis the authors  
596 concluded that metformin is likely to modulate the cell energy metabolism at both  
597 metabolomics and transcriptomics level leading to the suppression of LoVo cell proliferation  
598 [140]. This illustrates the potential of metabolomics tools not only for providing insight into  
599 the drug mechanisms of action in models during the pre-clinical stages of drug development,  
600 but also for the monitoring of metabolic responses to the lead compound in patients during  
601 the clinical stages of drug development.

602

## 603 **10. METABOLOMICS OF COMPLEX COMPOUND MIXTURES AND** 604 **INTEGRATION OF BIOACTIVITY DATA**

605 The use of complex mixtures of compounds such as typically found in extract libraries has  
606 traditionally been a challenge in NP discovery due to the diversity of small molecule present  
607 in varying levels some of which may interact and interfere in bioassays. As this is  
608 problematic for the identification of bioactive components in a single extract, the standard  
609 (reductionist) approach has been to generate fractions of reduced chemical complexity or  
610 testing pure compounds. However, with high-throughput screening and high-throughput  
611 chemical characterisation it is possible to generate a wealth of information regarding  
612 qualitative and quantitative differences in compounds mixtures for all the complex mixtures  
613 tested bioactivity. These multidisciplinary resources could then be data-mined in order to find  
614 correlations between chemical composition and bioactivity phenotype [141,142]. Compound  
615 activity mapping was developed by integrating high-content biological screening and  
616 untargeted metabolomics using a correlation of individual mass signals with specific  
617 phenotypes from a cell-based assay [141]. The authors combined 10 977 mass spectral  
618 features and 58 032 biological measurements from a library of 234 NP extracts which  
619 resulted in the identification of 13 clusters of fractions containing 11 known compound  
620 families and four new compounds. The authors found this methodology suitable for finding  
621 clusters containing single bioactives indicating a single mode of action. Furthermore clusters  
622 containing multiple bioactives were also reported which highlighted potential different modes  
623 of action although false positive compounds were also reported. Ultimately, in order to  
624 prevent false positives it is necessary that the extract libraries provide wide quantitative and

625 qualitative chemical differences and minimize the occurrence of compounds that consistently  
626 share quantitative trends. It is proposed that screening larger libraries of extracts of similar  
627 organisms could aid in solving this problem [141].

## 628 **CONCLUSION**

629 Metabolomics provides a wealth of tools with huge potential for aiding the process of  
630 drug/bioactive discovery from plant tissues (Figure 3). This process starts with the selection of  
631 target raw plant material and is often guided by ethnopharmacological leads. In addition,  
632 metabolomics can be used in the context of chemotaxonomy in order to identify sources of novel  
633 chemical entities, classes of compounds overrepresented in medicinal plants and novel plant  
634 sources for bioactive components. Furthermore, traditional bioprospection approaches rely on  
635 analytical tools integral to metabolomics such as NMR- and MS-based technologies. However,  
636 metabolomics aims to provide high-throughput chemical screening tools and these can be used  
637 for dereplication and identification of novel chemical entities with promising medicinal properties  
638 from complex plant extracts. When utilised as part of multidisciplinary approaches it provides the  
639 opportunity for mining the datasets in order to identify chemical diversity and potential novel  
640 leads for drug development as well as identifying the relevant biosynthetic pathways underlying  
641 the accumulation of bioactive components. Furthermore, metabolomics approaches can be  
642 applied to cellular, animal and human models in order to assess drug metabolism and  
643 mechanisms of action. Metabolomics can also be applied throughout the development of  
644 biotechnology-based approaches in order to identify promising metabolic engineering strategies  
645 and monitor unintended metabolic effects. Finally, analysis of biofluids utilising metabolomics-  
646 based tools has great application potential for monitoring the drug metabolism response in model  
647 species. In summary, the metabolomics toolbox holds great potential for application throughout  
648 the process of drug and bioactive discovery, ranging from the screening of promising plant raw  
649 material for NPs to the elucidation of drug metabolism through the analysis of biofluids.

650

## 651 **CONFLICT OF INTEREST**

652 The authors declare no conflict of interest.

## 653 **ACKNOWLEDGEMENTS**

654 AF and DS acknowledge support from the Scottish Government's Rural and Environment  
655 Science and Analytical Services Division (RESAS) and EU-PF7 BachBerry (Project No.  
656 FP7-613793). DS also acknowledges support from DISCO (Project No. FP7-613513).

## 657 **REFERENCES**

658 [1] Newman DJ, Cragg GM. Natural products as sources of new drugs from 1981 to 2014. *J*  
659 *Nat Prod* 2016; 79:629-661

660 [2] Harvey AL. Natural products in drug discovery. *Drug Discov Today* 2008; 13:894-901.

661 [3] Weckwerth W. Metabolomics in systems biology. *Annu Rev Plant Biol* 2003; 54:669-689

662 [4] Tietze LF, Bell HP, Chandrasekhar S. Natural product hybrids as new leads for drug  
663 discovery. *Angew Chem Int Ed* 2003; 42:3996-4028

664 [5] Galm U, Shen B. Natural product discovery: the times have never been better. *Chem Biol*  
665 2007; 14:1098-1104

666 [6] Yuliana ND, Jahangir M, Verpoorte R, Choi YH. Metabolomics for the rapid  
667 dereplication of bioactive compounds from natural sources. *Phytochem Rev* 2013; 12:292-  
668 304.

669 [7] David B, Wolfender JL, Dias DA. The pharmaceutical industry and natural products:  
670 historical status and new trends. *Phytochem Rev* 2015; 14:299-315.

671 [8] Feher M, Schimdt JM. Property distributions: differences between drugs, natural  
672 products, and molecules from combinatorial chemistry. *J Chem Inf Comput Sci* 2003;  
673 43:218-227.

674 [9] Demain AL. Importance of microbial natural products and the need to revitalize their  
675 discovery. *J Ind Microbiol Biotechnol* 2014; 41:185.

676 [10] Schueffler A, Anke T. Fungal natural products in research and development. *Nat Prod*  
677 *Rep* 2014; 31:1425-1448.

678 [11] Torres-Larios A, Gurrola GB, Zamudio FZ, Possani LD. Hadrurin, a new antimicrobial  
679 peptide from the venom of the scorpion *Hadrurus aztecus*. *Eur J Biochem* 2000; 267:5023-  
680 5031.

681 [12] Mora C, Tittensor DP, Adl S, Simpson AGB, Worm B. How many species are there on  
682 earth and in the ocean? *PLOS Bio.* 2011. 9:e1001127.

683 [13] Curtis TP, Sloan WT, Scannell JW. Estimating prokaryotic diversity and its limits. *Proc*  
684 *Natl Acad Sci* 2002; 99:10494-10499.

- 685 [14] Mayer AMS, Glaser KB, Cuevas C, *et al.* The odyssey of marine pharmaceuticals: a  
686 current pipeline perspective. *Trends Pharmacol Sci* 2010; 31:255-265.
- 687 [15] Newman DJ, Cragg GM. Drugs and drug candidates from marine sources: an assessment  
688 of the current “state of play”. *Planta Med* 2016; 82:775-789.
- 689 [16] Appletans W, Ahyong ST, Anderson G, *et al.* The magnitude of global marine species  
690 diversity. *Curr Biol* 2012; 22:2189-2202.
- 691 [17] Pimm SL, Joppa LN. How many plant species are there, where are they, and at what rate  
692 are they going extinct. *Ann Missouri Bot Gard* 2015; 100:170-176.
- 693 [18] Proksch P, Edrada RA, Ebel R. Drugs from the seas – current status and microbiological  
694 implications. *Appl Microbiol Biotechnol* 2002; 59:125-134.
- 695 [19] Gilani AH, Rahman AU. Trends in ethnopharmacol. *J Ethnopharmacol* 2005; 100:43-49.
- 696 [20] Ekor M. The growing use of herbal medicines: issues relating to adverse reactions and  
697 challenges in monitoring safety. *Front. Pharmacol* 2014; 4:177.
- 698 [21] Mao JJ, Li QS, Soeller I, Xie SX, Amsterdam JD. *Rhodiola rosea* therapy for major  
699 depressive disorder: a study protocol for a randomized, double-blind, placebo- controlled  
700 trial. *J Clin Trials* 2014; 20:170.
- 701 [22] Amsterdam JD, Panossian AG. *Rhodiola rosea* L. as a putative botanical antidepressant.  
702 *Phytomedicine* 2016; 23:770-783.
- 703 [23] Skirycz A, Kierszniowska S, Meret M, Willmitzer L, Tzotzos G. Medicinal  
704 Bioprospecting of the Amazon rainforest: a modern Eldorado. *Trends Biotechnol.* 2016;  
705 34:781-790.
- 706 [24] Wolf D, Siems K. Burning the Hay to find the needle – Data mining strategies in Natural  
707 product dereplication. *Chimia* 2007; 61:339-345.
- 708 [25] Duarte K, Rocha-Santos TAP. Analytical techniques for discovery of bioactive  
709 compounds from marine fungi. *Trends Anal Chem* 2012; 34:97-110.
- 710 [26] Weller MG. A unifying review of Bioassay-guided fractionation, effect-directed analysis  
711 and related techniques. *Sensors* 2012; 12:9181-9209.
- 712 [27] Shi SY, Zhang YP, Jiang XY, *et al.* Coupling HPLC to on-line, post-column  
713 (bio)chemical assays for high-resolution screening of bioactive compounds from complex  
714 mixtures. *Trends Anal Chem* 2009; 28:865-877.

- 715 [28] Ciesla LM, Waksmundzka-Hajnos M, Wojtunik KA, Haknos M. Thin-layer  
716 chromatography coupled with biological detection to screen natural mixtures for potential  
717 drug leads. *Phytochem Lett* 2015; 11:445-454.
- 718 [29] Inglese J, Shamu CE, Guy RK. Reporting data from High-throughput screening of small-  
719 molecule libraries. *Nat Chem Biol* 2007; 438-441.
- 720 [30] Pauli GF, Chen SN, Friesen JB, McAlpine JB, Jaki BU. Analysis and Purification of  
721 Bioactive Natural Products: The AnaPurNa Study. *J Nat Prod* 2012; 75:1243-1255.
- 722 [31] Jenwitheesuk E, Horst JA, Rivas KL, Van Voorhis WC, Samudrala R. Novel paradigms  
723 for drug discovery: computational multitarget screening. *Trends Pharmacol Sci* 2008; 29:62-  
724 71.
- 725 [32] Alexander JL, Wilson ID, Teare J, Marchesi JR, Nicholson JK, Kinross JM. Gut  
726 microbiota modulation of chemotherapy efficacy and toxicity. *Nat Rev Gastroenterol Hepatol*  
727 2017; doi: 10.1038/nrgastro.2017.20. [Epub ahead of print]
- 728 [33] Metzner KJ, Allers K, Rauch P, Harrer T. Rapid selection of drug-resistant HIV-1 during  
729 the first months of suppressive ART in treatment-naïve patients. *AIDS* 2007; 21:703-711.
- 730 [34] Griner LAM, Guha R, Shinn P, *et al.* High-throughput combinatorial screening identifies  
731 drugs that cooperate with ibrutinib to kill activated B-cell-like diffuse large B-cell lymphoma  
732 cells. *PNAS* 2014; 111:2349-2354.
- 733 [35] Arrell DK, Terzic A. Network systems biology for drug discovery. *Clin Pharmacol Ther*  
734 2010; 88:120-125.
- 735 [36] Yeh JC, Garrard IJ, Cho CWC, *et al.* Bioactivity-guided fractionation of the volatile oil  
736 of *Angelica sinensis radix* designed to preserve the synergistic effects of the mixture followed  
737 by identification of the active principles. *J Chromatogr A* 2012; 1236:132-138.
- 738 [37] Elisabetsky E, Amador TA, Albuquerque RR, Nunes DS, Carvalho ACT. Analgesic  
739 activity of *Psychotria colorata* (Willd. Ex R. & S.) Muell. Arg. Alkaloids. *J Ethnopharmacol*  
740 1995; 48:77-83.
- 741 [38] Amador TA, Elisabetsky E, Souza DO. Effects of *Psychotria colorata* alkaloids in brain  
742 opioid system. *Neurochem Res* 1996; 21:97-102.
- 743 [39] Amador TA, Verotta L, Nunes DS, Elisabetsky E. Involvement of NMDA receptors in  
744 the analgesic properties of psychotridine. *Phytomedicine* 2001; 8:202-206.
- 745 [40] Byers JA. Optimal fractionation and bioassay plans for isolation of synergistic  
746 chemicals: the subtractive-combination method. *J Chem Ecol* 1992; 18:1603-1621.



- 747 [41] Verpoorte R, Choi YH, Kim HK. Ethnopharmacology and systems biology: a perfect  
748 holistic match. *J Ethnopharmacol* 2005; 100:53-56.
- 749 [42] Vicini P, van der Graaf PH. Systems pharmacology for drug discovery and development:  
750 Paradigm shift or flash in the Pan. *Clin Pharmacol Ther* 2013; 93:379-381.
- 751 [43] Wu C, Kim HK, van Wezel GP, Choi YH. Metabolomics in the natural products field –  
752 a gateway to novel antibiotics. *Drug Discov Today: Technol* 2015; 13:11-17.
- 753 [44] Oliver SG, Winson MK, Kell DB, Baganz F. Systematic functional analysis of the yeast  
754 genome. *Trends Biotechnol* 1998; 16:373-378.
- 755 [45] Fiehn O. Metabolomics – the link between genotypes and phenotypes. *Plant Mol Biol*  
756 2002; 48:155-171.
- 757 [46] Shulaev V, Cortes D, Miller G, Mittler R. Metabolomics for plant stress response.  
758 *Physiol Plant* 2008; 132:199-208.
- 759 [47] Wolfender JL, Rudaz S, Choi YH, Kim HK. Plant metabolomics: from holistic data to  
760 relevant biomarkers. *Curr Med Chem* 2013; 20:1056-1090.
- 761 [48] Hong J, Yang L, Zhang D, Shi J. Plant metabolomics: an indispensable system biology  
762 tool for plant science. *Int J Mol Sci* 2016; 17:767.
- 763 [49] Kaplan F, Kopka J, Haskell DW, *et al.* Exploring the temperature-stress metabolome of  
764 *Arabidopsis*. *Plant Phys* 2004; 136:4159-4168.
- 765 [50] Foito A, Byrne SL, Shepherd T, Stewart D, Barth S. Transcriptional and metabolic  
766 profiles of *Lolium perenne* L. genotypes in response to a PEG-induced water stress. *Plant*  
767 *Biotechnol J* 2009; 7:719-732.
- 768 [51] Moradi P, Ford-Lloyd B, Pritchard J. Metabolomic approach reveals the biochemical  
769 mechanisms underlying drought stress tolerance in thyme. *Anal Biochem* 2017; Doi:  
770 10.1016/j.ab.2017.02.006.
- 771 [52] Barros E, Lezar S, Anttonen MJ, *et al.* Comparison of two GM maize varieties with  
772 near-isogenic non-GM variety using transcriptomics, proteomics and metabolomics. *Plant*  
773 *Biotechnol J* 2010; 8:436-451.
- 774 [53] Chang Y, Zhao C, Zhu Z, *et al.* Metabolic profiling based on LC/MS to evaluate  
775 unintended effects of transgenic rice with cry1Ac and sck genes. *Plant Mol Biol* 2012;  
776 78:477-487.
- 777 [54] Shepherd LVT, Hackett CA, Alexander CJ, *et al.* Modifying glycoalkaloid content in  
778 transgenic potato – metabolome impacts. *Food Chem* 2015; 187:437-443.

- 779 [55] Joshi L, Van Eck JM, Mayo K, *et al.* Metabolomics of plant saponins: Bioprospecting  
780 triterpene glycoside diversity with respect to mammalian cell targets. *OMICS* 2002. 6:235-  
781 246.
- 782 [56] Hagel JM, Mandal R, Han B, *et al.* Metabolome analysis of 20 taxonomically related  
783 benzyloisoquinoline alkaloid-producing plants. *BMC Plant Biol* 2015; 15:220.
- 784 [57] Shepherd LVT, Fraser P, Stewart D. Metabolomics: a second-generation platform for  
785 crop and food analysis. *Bioanalysis* 2011; 3, 1143-1159.
- 786 [58] Mazur SP, Sonstebly A, Nes A, *et al.* Effects of post-flowering environmental variation  
787 along an altitudinal gradient on chemical composition of ‘Glen Ample’ red raspberry (*Rubus*  
788 *idaeus* L.). *Eur J Hortic Sci* 2014; 79:267-277.
- 789 [59] Adhikari KB, Boelt B, Fomsgaard IS. Identification and quantification of loline-type  
790 alkaloids in endophyte-infected grasses by LC-MS/MS. *J Agric Food Chem* 2016. 64:6212-  
791 6218.
- 792 [60] Berhow MA, Polat U, Glinski JA, *et al.* Optimized analysis and quantification of  
793 glucosinolates from *Camelina sativa* seeds by reverse-phase liquid chromatography. *Ind*  
794 *Crops Prod* 2013. 43:119-125.
- 795 [61] Shulaev V. Metabolomics technology and bioinformatics. *Brief Bioinform* 2006; 7:128-  
796 139.
- 797 [62] Nielsen J, Oliver SG. The next wave in metabolome analysis. *Trends Biotechnol* 2005;  
798 23:544-546.
- 799 [63] Courant F, Antignac JP, Dervilly-Pinel G, Bizec BL. Basics of mass spectrometry based  
800 metabolomics. *Proteomics* 2014. 14:21-22.
- 801 [64] Byrne SL, Foito A, Hedley PE, Morris JA, Stewart D, Barth S. Early response  
802 mechanisms of perennial ryegrass (*Lolium perenne*) to phosphorus deficiency. *Ann Bot* 2011.  
803 107:243-254.
- 804 [65] Obata T, Witt S, Lisek J, *et al.* Metabolite profiles of maize leaves in drought, heat and  
805 combined stress field trials reveal the relationship between metabolism and grain yield. *Plant*  
806 *Phys* 2015; Doi: 10.1104/pp.15.01164.
- 807 [66] Dunn WB, Ellis DI. Metabolomics: Current analytical platforms and methodologies.  
808 *Trends Anal Chem* 2005; 24:285-294.
- 809 [67] Hantao LW, Ribeiro FAL, Passador MM, *et al.* Metabolic profiling by ultra-performance  
810 liquid chromatography-mass spectrometry and parallel factor analysis for the determination  
811 of disease biomarkers in *Eucalyptus*. *Metabolomics* 2014. 10:1318-1325.

- 812 [68] Lesiak AD, Cody RB, Dane AJ, Musah RA. Plant seed species identification form  
813 chemical fingerprints: a high-throughput application of direct analysis in real time mass  
814 spectrometry. *Anal Chem* 2015; 87:8748-8757.
- 815 [69] Cottet K, Kouloura E, Kritsanida M, et al. Comparative metabolomic study between  
816 African and Amazonian *Symphonia globulifera* by tandem LC-HRMS. *Phytochem Lett* 2017.  
817 (*in press*)
- 818 [70] Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak. XCMS: Processing Mass  
819 spectrometry data for metabolite profiling using non-linear peak alignment, matching and  
820 identification. *Anal Chem* 2006; 78:779-787.
- 821 [71] Tautenhahn R, Patti GJ, Rinehart D, Siuzdak. XCMS: Online: a web-based platform to  
822 process untargeted metabolomics data. *Anal Chem* 2012; 84:5035-5039.
- 823 [72] Luedemann A, Strassburg K, Erban A, Kopka J. TagFinder for the quantitative analysis  
824 of gas-chromatography-mass-spectrometry (GC-MS)-based metabolite profiling experiments.  
825 *Bioinformatics* 2008; 24:732-737.
- 826 [73] Kamleh A, Barrett MP, Wildridge D, Burchmore RJS, Scheltema RA. Metabolomic  
827 profiling using Orbitrap Fourier transform mass spectrometry with hydrophilic interaction  
828 chromatography: a method with wide applicability to analysis of biomolecules. *Rapid*  
829 *Commun Mass Spectrom* 2008; 22:1912-1918.
- 830 [74] Burch TC, Isaac G, Booher CL, *et al.* Comparative metabolomics and lipidomic analysis  
831 of phenotype stratified prostate cells. *PLOS One* 2015; Doi: 10.1371/journal.pone.0134206.
- 832 [75] Pluskal T, Castillo S, Villar-Briones A, Oresic M. MZmine 2: Modular framework for  
833 processing, visualizing and analysing mass spectrometry-based molecular profile data. *BMC*  
834 *Bioinformatics* 2010; 11:395.
- 835 [76] Lommen A. MetAlign: Interface driven, versatile metabolomics tool for hyphenated full-  
836 scan mass spectrometry data preprocessing. *Anal Chem* 2009; 81:3079-3086.
- 837 [77] Skogerson K, Wohlgemuth G, Barupal DK, Fiehn O. The volatile compound BinBase  
838 mass spectral database. *BMC Bioinformatics* 2011; 12:321.
- 839 [78] Sumner LW, Amberg A, Barrett D *et al.* Proposed minimum reporting standards for  
840 chemical analysis. *Metabolomics* 2007; 3:211-221.
- 841 [79] Dunn WB, Erban A, Weber RJM, *et al.* Mass Appeal: metabolite identification in mass-  
842 spectrometry –focused untargeted metabolomics. *Metabolomics* 2013; 9:44-66.
- 843 [80] Krishnan P, Kruger NJ, Ratcliffe RG. Metabolite fingerprinting and profiling in plants  
844 using NMR. *J Exp Bot* 2005; 56:255-265.

- 845 [81] Verpoorte R, Choi YH, Kim HK. NMR-based metabolomics at work in phytochemistry.  
846 *Phytochem Rev* 2007; 6:3-14.
- 847 [82] Troufflard S, Roscher A, Thomasset B, Barbotin JN, Rawsthorne S, Portais JC. In vivo  
848 <sup>13</sup>C NMR determines metabolic fluxes and steady state in linseed embryos. *Phytochemistry*  
849 2007; 68:2341-2350.
- 850 [83] Wolfender JL, Ndjoko K, Hostettmann K. The potential of LC-NMR in phytochemical  
851 analysis. *Phytochem Anal* 2001; 12:2-22.
- 852 [84] Hiller W, Sinha P, Hehn M, Pasch H. Online LC-NMR – From an expensive toy to a  
853 powerful tool in polymer analysis. *Prog Polym Sci* 2013; 39:979-1016.
- 854 [85] Aharoni A, de Vos CHR, Verhoeven HA, *et al.* Nontargeted metabolome analysis by use  
855 of Fourier transform ion cyclotron mass spectrometry. *OMICS* 2004; 6:217-234.
- 856 [86] McDougall G, Martinussen I, Stewart D. Towards fruitful metabolomics: high  
857 throughput analyses of polyphenol composition in berries using direct infusion mass  
858 spectrometry. *J Chromatogr b* 2008; 871:362-369.
- 859 [87] Kopka J, Fernie AR, Weckwerth W, Gibon Y, Stitt M. Metabolite profiling in plant  
860 biology: platforms and destinations. *Genome Biol* 2004; 5:109.
- 861 [88] Kusano M, Kobayashi M, Iizuka Y, Fukushima A, Saito K. Unbiased profiling of  
862 volatile organic compounds in the headspace of *Allium* plants using nd in-tube extraction  
863 device. *BMC Res Notes* 2016; 9:133.
- 864 [89] Schauer N, Semel Y, Roessner U, *et al.* Comprehensive metabolic profiling and  
865 phenotyping of interspecific introgression lines for tomato improvement. *Nature Biotechnol*  
866 2006; 24:447:454.
- 867 [90] Halket JM, Waterman D, Przyborowska AM, Patel RKP, Fraser PD, Bramley PM.  
868 Chemical Derivatization and mass spectral libraries in metabolite profiling by GC/MS and  
869 LC/MS/MS. *J Exp Bot* 2005; 56:219-243.
- 870 [91] Kopka J, Schauer N, Krueger S, *et al.* GMD@CSB.DB: the Golm metabolome database.  
871 *Bioinformatics* 2005; 21:1635-1638.
- 872 [92] Kind T, Wohlgemuth G, Lee DY, *et al.* FiehnLib – mass spectral and retention index  
873 libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass  
874 spectrometry. *Anal Chem* 2009; 81:10038-10048.
- 875 [93] van der Hoof JJJ, Vervoot J, Bino RJ, Beekwilder J, de Vos RCH. Polyphenol  
876 identification based on systematic and robust high-resolution accurate mass spectrometry  
877 fragmentation. *Anal Chem* 2011; 83:409-413.

- 878 [94] Iijima Y, Nakamura Y, Ogata Y, *et al.* Metabolite annotations based on the integration  
879 of mass spectral information. *Plant J* 2008; 54:949-962.
- 880 [95]. Matsuda F, Nakabayashi R, Sawada Y, *et al.* Mass spectra-based framework for  
881 automated structural elucidation of metabolome data to explore phytochemical diversity.  
882 *Front Plant Sci* 2011; doi. 10.3389/fpls.2011.00040.
- 883 [96] Corcoran O, Spraul M. LC-NMR-MS in drug discovery. *Drug Discov Today* 2003;  
884 8:626-631.
- 885 [97] Bingol K, Brüsweiler R Knowns and unknowns in metabolomics identified by  
886 multidimensional NMR and hybrid MS/NMR methods. *Curr Opin Biotechnol* 2017; 43:17-  
887 24.
- 888 [98] Bingol K, Brüsweiler-Li L, Yu C, Somogyi A, Zhang F, Brüsweiler R.  
889 Metabolomics beyond spectroscopic databases: a combined MS/NMR strategy for the rapid  
890 identification of new metabolites in complex mixtures. *Anal Chem* 2015; 87:3864-3870.
- 891 [99] Moore BD, Andrew RL, Külheim C, Foley WJ. Explaining intraspecific diversity in  
892 plant secondary metabolites in an ecological context. *New Phytol* 2014; 201:733-750.
- 893 [100] Schwab W. Metabolome diversity: too few genes, too many metabolites?  
894 *Phytochemistry* 2003; 62:837-849.
- 895 [101] Harborne JB, *The Flavonoids Advances in Research since 1980*. In: Harborne JB, *The*  
896 *Flavonoids Advances in Research since 1980*. 1<sup>st</sup> ed. Chapman and Hall Ltd 1988; pp. 576-  
897 578.
- 898 [102] Hall D, Yuan XX, Murate J, De Luca V. Molecular cloning and biochemical  
899 characterization of the UDP-glucose:Flavonoid 3-O-glucosyltransferase from Concord grape  
900 (*Vitis labrusca*). *Phytochemistry* 2012; 74:90-99.
- 901 [103] Agrawal AA, Hastings AP, Johnson MTJ, Maron JJ, Salminen JP. Insect herbivores  
902 drive real-time ecological and evolutionary change in plant populations. *Science* 2012;  
903 338:113-116.
- 904 [104] Züst T, Heichinger C, Grossniklaus U, Harrington R, Kliebenstein DJ, Turnbull LA.  
905 Natural enemies drive geographic variation in plant defences. *Science* 2012; 338:116-119.
- 906 [105] Alba C, Bowers MD, Blumenthal D, Hufbauer R. Evolution of growth but not  
907 structural or chemical defence in *Verbascum Thapsus* (common mullein) following  
908 introduction to North America. *Biol Invasions* 2011; 13:2379-2389.

- 909 [106] Franks SJ, Wheeler GS, Goodnight C. Genetic variation and evolution of secondary  
910 compounds in native and introduced populations of the invasive plant *Melaleuca*  
911 *quinquenervia*. *Evolution* 2012; 66:1398-1412.
- 912 [107] Carmona D, Lajeunesse MJ, Johnson MTJ. Plant traits that predict resistance to  
913 herbivores. *Funct Ecol* 2011; 25:358-367.
- 914 [108] Jones CG, Firm RD. On the evolution of plant secondary chemical diversity. *Philos*  
915 *Trans R Soc Lond B Biol Sci* 1991; 333:273-280.
- 916 [109] Singh R. Chemotaxonomy: A tool for plant classification. *J Med Plants* 2016; 4:90-93.
- 917 [110] Xie G, Plumb R, Su M, *et al.* Ultra-performance LC/TOF MS analysis of medicinal  
918 *Panax* herbs for metabolomics research. *J Sep Sci* 2008; 31:1015-1026.
- 919 [111] Lee S, Oh D-G, Lee S, *et al.* Chemotaxonomic metabolite profiling of 62 indigenous  
920 plant species and its correlations with bioactivities. *Molecules* 2015; 20:19719-19734.
- 921 [112] Grimbs A, Shresta A, Resk ASD, *et al.* Bioactivity in *Rhododendron*: a systemic  
922 analysis of antimicrobial and cytotoxic activities and their phylogenetic and phytochemical  
923 origins. *Front Plant Sci* 2017; Doi: 10.3389/fpls.2017.00551.
- 924 [113] Freiburghaus F, Steck A, Pfander H, Brun R. Bioassay-guided isolation of a  
925 diastereoisomer of kolavenol from *Entada abyssinica* active on *Trypanosoma brucei*  
926 *rhodesiense*. *J Ethnopharmacol.* 1998; 61:179-183.
- 927 [114] Inayatullah S, Prenzler PD, Obied HK, Rehman A-U, Mirza B. Bioprospecting  
928 traditional Pakistani medicinal plants for potent antioxidants. *Food Chem* 2012; 132:222-229.
- 929 [115] Kainz KP, Krenn L, Erdem Z, *et al.* 2-Deprenyl-rheediaxanthone B isolated from  
930 *Metaxya rostrate* induces active cell death in colorectal tumor cells. *PLOS One* 2013; doi:  
931 10.1371/journal.pone.0065745.
- 932 [116] Gogna N, Hamid N, Dorai K. Metabolomic profiling of the phytomedicinal  
933 constituents of *Carica papaya* L. leaves and seeds by <sup>1</sup>H NMR spectroscopy and multivariate  
934 statistical analysis. *J Pharm Biomed Anal* 2015; 115:74-85.
- 935 [117] Geng C-A, Chen J-J. A fragmentation study on four unusual securiridoid trimers,  
936 swerilactones H-K, by the electrospray tandem mass-spectrometry. *Nat Prod Bioprospect*  
937 2016; 6:297-303.
- 938 [118] Nielsen KF, Mansson M, Rank C, Frisvad JC, Larsen TO. Dereplication of microbial  
939 natural products by LC-DAD-TOFMS. *J Nat Prod* 2011; 74:2338-2348.

- 940 [119] Yang JY, Sanchez LM, Rath CM, *et al.* Molecular Networking as a dereplication  
941 strategy. *J Nat Prod* 2013; 76:1686-1699.
- 942 [120] McIntyre L, Zhang T, Viegelmann C, *et al.* Metabolomic tools for secondary  
943 metabolite discovery from marine microbial symbionts. *Mar Drugs* 2014; 12:3416-3448.
- 944 [121] Gerwick WH, Moore BS. Lessons from the past and charting the future of marine  
945 natural products drug discovery and chemical biology. *Chem Biol* 2012; 19:85-98.
- 946 [122] Kurita KL, Glassey E, Linington RG. Integration of high-content screening and  
947 untargeted metabolomics for comprehensive functional annotation of natural product  
948 libraries. *PNAS* 2015; 112:11999-12004.
- 949 [123] Kind T, Fiehn O. Strategies for dereplication of natural compounds using high-  
950 resolution tandem mass spectrometry. *Phytochem Lett* 2016; doi:  
951 10.1016/j.phytol.2016.11.006.
- 952 [124] Jiang Y, Liu F-J, Wang Y-M, Li H-J. Dereplication-guided isolation of novel  
953 hepatoprotective triterpenoid saponins from *Celosiae Semen* by high-performance liquid  
954 chromatography coupled with electrospray ionization tandem quadrupole-time-of-flight mass  
955 spectrometry. *J Pharm Biomed Anal* 2017; 132:148-155.
- 956 [125] Moore JM. NMR screening in drug discovery. *Curr Opin Biotechnol* 1999; 10:54-58.
- 957 [126] Cala O, Guilliere F, Krimm I. NMR-based analysis of protein-ligand interactions. *Anal*  
958 *Bioanal Chem* 2014; 406:943-956.
- 959 [127] Wilme A, Limonciel A, Aschauer L, *et al.* Application of integrated transcriptomic,  
960 proteomic and metabolomics profiling for the delineation of mechanisms of drug induced cell  
961 stress. *J Proteomics* 2013; 79:180-194.
- 962 [128] Van den Hof WFPM, Ruiz-Aracama A, Summeren AV, *et al.* Integrating multiple  
963 omics to unravel mechanisms of Cyclosporin A induced hepatotoxicity in vitro. *Toxicol In*  
964 *Vitro* 2015; 29:489-501.
- 965 [129] Pimpão, R.C., Dew, T., Figueira, M.E., *et al.* Urinary metabolite profiling identifies  
966 novel colonic metabolites and conjugates of phenolics in healthy volunteers. *Mol Nutr Food*  
967 *Res* 2014; 58:1414-1425.
- 968 [130] Zhang L, Hatzakis E, Patterson AD. NMR-based metabolomics and its application in  
969 drug metabolism and cancer research. *Curr Pharm Rep* 2016; 2:231-240.
- 970 [131] Sun J, Schnackenberg LK, Beger RD. Studies of acetaminophen and metabolites in  
971 urine and their correlations with toxicity using metabolomics. *Drug Metab Lett* 2009; 3:130-  
972 136.

- 973 [132] Sun J, Schnackenberg LK, Khare S, *et al.* Evaluating effects of penicillin treatment on  
974 the metabolome of rats. *J Chromatogr B* 2013; 932:134-143.
- 975 [133] Armitage EG, Barbas C. Metabolomics in cancer biomarker discovery: Current trends  
976 and future perspectives. *J Pharm Biomed Anal* 2014. 87:1-11.
- 977 [134] Patel S, Ahmed S. Emerging field of metabolomics: big promise for cancer biomarker  
978 identification and drug discovery. *J Pharm Biomed Anal* 2015. 107:63-74.
- 979 [135] Johnson CH, Ivanisevic J, Siuzdak G. Metabolomics: Beyond biomarkers and towards  
980 mechanisms. *Nat Revs Mol Cell Biol* 2016. 17:451-459.
- 981 [136] Jin X, Yun SJ, Jeong P, *et al.* Diagnosis of bladder cancer and prediction of survival by  
982 urinary metabolomics. *Oncotarget* 2014. 30:1635-1645.
- 983 [137] Ren S, Shao Y, Zhao X, *et al.* Integration of metabolomics and transcriptomics reveals  
984 major metabolic pathways and potential biomarker involved in prostate cancer. *Mol Cell*  
985 *Proteomics* 2016. 15:154-163.
- 986 [138] Armitage EG, Southam AF. Monitoring cancer prognosis, diagnosis and treatment  
987 efficacy using metabolomics and lipidomics. *Metabolomics* 2016. 12:146.
- 988 [139] Schuler KM, Rambally BS, DiFurio MJ, *et al.* Antiproliferative and metabolic effects  
989 of metformin in a preoperative window clinical trial for endometrial cancer. *Cancer Med*  
990 2015. 4:161-173.
- 991 [140] He J, Wang K, Zheng N, *et al.* Metformin suppressed the proliferation of LoVo cells  
992 and induced a time-dependent metabolic and transcriptional alteration. *Sci Rep* 2015.  
993 5:17423.
- 994 [141] Kurita KL, Glassey E, Linington RG. Integration of high-content screening of  
995 untargeted metabolomics for comprehensive functional annotation of natural product  
996 libraries. *Proc Natl Acad Sci* 2015; 112:11999-12004.
- 997 [142] Aliagiannis N, Halabalaki M, Chaita E, *et al.* Heterocovariance based metabolomics as  
998 a powerful tool accelerating bioactive natural product identification. *ChemistrySelect* 2016;  
999 1:2531-2535.

1000

1001 **FIGURES**



1002 **Figure 1** – Diagram representing the two leading approaches utilised in bioprospection of  
1003 bioactive compound: pure compound screening (left) and bioassay-guided fractionation  
1004 (right).

1005 **Figure 2** – Comparison of relative analyte coverage with respect to molecular mass, polarity  
1006 and concentration ranges across different analytical platforms.

1007

1008 **Figure 3** – Potential applications of metabolomics tools in the drug discovery workflow

1009