Biochimica et Biophysica Acta 1840 (2014) 3460–3474

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbagen

Review

The utility of metabolomics in natural product and biomarker characterization

Daniel G. Cox [a,1], Joonseok Oh [a,1], Adam Keasling [a,1], Kim L. Colson [b], Mark T. Hamann [a,*]

[a] Department of Pharmacognosy, Pharmacology, Chemistry and Biochemistry and Research Institute of Pharmaceutical Sciences, School of Pharmacy, The University of Mississippi, University, MS 38677, USA

[b] R&D Division, Bruker BioSpin, 15 Fortune Drive Billerica, MA 01821, USA

ARTICLE INFO

Article history:
Received 13 August 2013
Received in revised form 13 August 2014
Accepted 14 August 2014
Available online 20 August 2014

Keywords:
Integrated approach
PCA
Plant metabolomics
Diagnostic biomarker
NMR
Targeted metabolomics

ABSTRACT

Background: Metabolomics is a well-established rapidly developing research field involving quantitative and qualitative metabolite assessment within biological systems. Recent improvements in metabolomics technologies reveal the unequivocal value of metabolomics tools in natural products discovery, gene-function analysis, systems biology and diagnostic platforms.

Scope of review: We review here some of the prominent metabolomics methodologies employed in data acquisition and analysis of natural products and disease-related biomarkers.

Major conclusions: This review demonstrates that metabolomics represents a highly adaptable technology with diverse applications ranging from environmental toxicology to disease diagnosis. Metabolomic analysis is shown to provide a unique snapshot of the functional genetic status of an organism by examining its biochemical profile, with relevance toward resolving phylogenetic associations involving horizontal gene transfer and distinguishing subgroups of genera possessing high genetic homology, as well as an increasing role in both elucidating biosynthetic transformations of natural products and detecting preclinical biomarkers of numerous disease states.

General significance: This review expands the interest in multiplatform combinatorial metabolomic analysis. The applications reviewed range from phylogenetic assignment, biosynthetic transformations of natural products, and the detection of preclinical biomarkers.

1. Introduction

Metabolomics, also referred to as metabolite profiling, is the assessment of metabolites within a biological system and has a well-established history [1–3]. Metabolomics platforms frequently combine NMR with MS [4] and when a set of defined metabolites is being analyzed a technique known as "targeted" metabolomics is performed. Typically, targeted metabolomics is utilized to detect the relative concentrations of no more than 200 predetermined analytes in a sample. Without a priori knowledge of metabolite targets, a much more comprehensive "untargeted" metabolomics approach is required. In untargeted metabolomics, biological samples are most often processed through an initial liquid chromatography (LC) or gas chromatography (GC) phase and subsequently subjected to MS analysis. Using this technique, thousands of metabolites can be detected in a single eluate, and it is this global approach that is leading the way to major revelations in our understanding of cell biology, physiology and medicine [5].

Novel MS and NMR imaging technologies and computational tools, have provided the means to obtain rapid accurate global assessment of metabolite conditions within an organism [6]. These advances have allowed us to utilize metabolomics technology to discern phylogenetic associations between a few individuals or entire populations, with greater precision than genetic analysis alone, and made possible the discovery of numerous disease biomarkers that have improved prognosis by early detection of cancer, diabetes and cardiovascular disease [7]. Metabolomics methods have provided snapshots of cellular activity used to investigate the epigenetic effects of environmental variations [8] and plant metabolite analysis has been employed to elucidate the ecological effects of climate change [9].

While metabolite analysis has been conducted for decades, recent improvements in metabolomics technologies reveal the unequivocal value of metabolomics tools in gene-function analysis, systems biology and diagnostic platforms [10]. New biological challenges are continually presenting a greater need for rapid responses from the scientific community. With cases of new drug resistant microbial infections being reported on an increasingly regular basis, there is a constant demand for

* Corresponding author. Tel.: +1 662 915 5730; fax: +1 662 915 6975
E-mail address: mthamann@olemiss.edu (M.T. Hamann).

[1] These authors contributed to this work equally.

http://dx.doi.org/10.1016/j.bbagen.2014.08.007
0304-4165/

expeditious production of novel safe and effective medicines. A substantial concerted effort is paramount if we are to obtain the knowledge required to provide the best treatments and prognosis. With advanced NMR and MS tools and state-of-the-art data analysis, rapid accurate metabolite profiles will continue to enhance metabolomics data repositories and lead to new drug therapies, more effective diagnostics and improved disease prognosis.

2. Background

The seeds of metabolomic science were officially planted when Devaux and Horning published research on "metabolite profiles" they derived from gas chromatography/mass spectrometry (GC/MS) analyses of human urine and tissue extracts [11]. An almost immediate interest in utilizing metabolite profiles to diagnose disease followed [12]. During the 1970s metabolomics studies were expanded across a broad range of activities including: novel techniques for detection and elucidation of insect hormones [13], natural products exhibiting streptomycin-like functionality extracted from marine algae [14] and chemotherapeutic agents derived from plant extracts e.g., *Hyptis tomentosa* [15]. In the early 1980s work on automating metabolite analysis began to appear [16] and in the mid-1980s metabolite profiling research using NMR and HPLC platforms were published [17–19]. In 1991, a global metabolite analysis approach was used to assess the mode of action (MOA) of herbicides in barley using GC/MS [20]. By the turn of the century, research and cooperation within the scientific community led to advances in metabolomics technologies and extraction methods promoting development of metabolite databases like METLIN [21], which are bringing us closer to a global assessment of the human metabolome [22].

3. Instrumentation and tools

Metabolomics platforms primarily fall into five general categories; (GC × MS), (LC × MS), (MS), (NMR) and (Integrated Applications). Analytical tools have been developed providing fast, comprehensive data that is annotatable, quantitative and reproducible. A general overview of some of the more prominent metabolomics methodologies relating to sample preparation, data acquisition and analysis is presented.

3.1. GC–MS approaches

The earliest research on "metabolite profiling" was published in the early 1970s by Horning and Devaux, describing their multi-component analysis of steroids and urinary drug metabolites using GC/MS techniques [11]. GC/MS technology continued to be developed from extensive research studies in the 1980s [23]. With GC/MS platforms, several hundred metabolites can be identified within a sample and some of the major advantages of GC/MS are reliable standard operating procedures (SOPs) for experimental design and sample and data analysis, which have been developed over its significant history in metabolite profiling [10].

GC/MS is highly selective; relatively inexpensive and provides reproducible analysis with an added advantage of comprehensive structural databases [24] for GC/MS metabolite identification. A wide variety of GC instruments are available in a multitude of creative combinations, each providing specific analytical qualities. While standard GC/MS platforms have been used successfully to elucidate novel metabolites [25], utilizing two-dimensional chromatography techniques further enhance metabolite quantification [26] and give metabolomics platforms even greater analytic capability [27]. Metabolomics studies require accurate, global metabolite assessment with the highest throughput possible. This need to achieve rapid comprehensive analysis, has inspired the development of novel multidimensional metabolomics platforms.

Table 1
Adapted and reprinted with permission from lecture information provided by James K. Hardy Ph.D., at the University of Akron, OH 44325.

Detector	Type	Compound applications	Sensitivity	Dynamic range
FID	D	Hydrocarbons	5 pg/s	10^7
TCD	ND	Universal	400 pg/mL	10^6
ECD	ND	Halogenated	0.1 pg/s	10^4
NPD	D	Nitrogen	0.4 pg N/s	10^4
		Phosphorus	0.2 pg P/s	
FPD	D	Sulphur	20 pg S/s	S 10^3
		Phosphorus	0.9 pg P/s	P 10^4
PID	ND	Ionizable by UV	2 pg/s	10^7

Flame Ionization Detector (FID), Thermal Conductivity Detector (TCD), Electron Capture Detector (ECD), Nitrogen Phosphorous Detector (NPD), Flame Photometric Detector (FPD), Photo Ionization Detector (FID), Hall Electro Conductivity Detector (HECD), Fourier Transform IR (FTIR), Destructive (D), and Non Destructive (ND).

GC platforms are designed according to the types of compounds being studied as well as the sensitivity required for analysis. Typical specifications for the most common types of GC detectors are provided in Table 1 below.

The most advanced GC metabolomics platforms include GC — Ion Trap/MS, GC–Time of Flight/MS (GC–TOF/MS), GC/tandem mass spectrometry (GC/MS/MS), and multidimensional GC/GC/MS. In a study of the effects of celastrol treatment in human cervical cancer cells, Hu et al. [28] combined a GC–Ion Trap/MS based metabolomics platform with multivariate statistical analysis and devised a rapid, effective methodology to simultaneously screen multiple metabolic pathways for characteristic anti-cancer drug perturbations.

3.2. LC-MS approaches

Liquid chromatography-mass spectrometry (LC-MS) is the most widely used technology in metabolomic analysis, due to its ability to separate and detect a diverse range of molecules with unparalleled sensitivity [23]. LC-MS platforms are also highly versatile, by coupling various liquid chromatography and mass spectrometry technologies, a myriad of component configurations with unique capabilities are possible [29–31].

Commonly employed liquid chromatographic methods used in metabolomic analysis include: normal-phase, reverse-phase, and hydrophilic interaction liquid chromatography. Normal-phase liquid chromatography (NPLC) readily separates highly non-polar metabolites (e.g. fatty acids, sterols, triacylglycerols, etc.), reverse-phase liquid chromatography (RPLC) is better suited for the separation of metabolites of medium to low polarity (e.g. alkaloids, flavonoids, glycosylated steroids, phenolic acids, etc.), and hydrophilic interaction liquid chromatography (HILIC) can separate a broad range of metabolites depending upon the bonded-phase of chromatographic media used (e.g. amino bonded, cationic boned, zwitterionic bonded, etc.) [23,30,31]. Chromatographic versatility extends beyond the plethoric selection of column separation media available, when utilizing ultra-high performance, capillary or multidimensional liquid chromatography. Ultra-high performance liquid chromatography (UPLC) employs a short column with small diameter particle size (<2 μm) packing material and can operate at pressures of 5000 psi, for short elution times, but at the expense of separation resolving power [23,31]. Capillary liquid chromatography incorporates much longer capillary columns, which significantly increase the resolution of chromatographic separation but necessitate increased analysis time [23]. Multidimensional liquid chromatography (MDLC) combines two or more of the preceding chromatographic modes to enhance metabolite separation [30,31].

In an LC-MS metabolomics platform, the series of metabolites produced by chromatography are subsequently exposed to an ion source which dissociates them into ionic derivatives required for mass analysis. Common ionization sources include atmospheric pressure chemical ionization (APCI), atmospheric pressure photoionization (APPI), fast

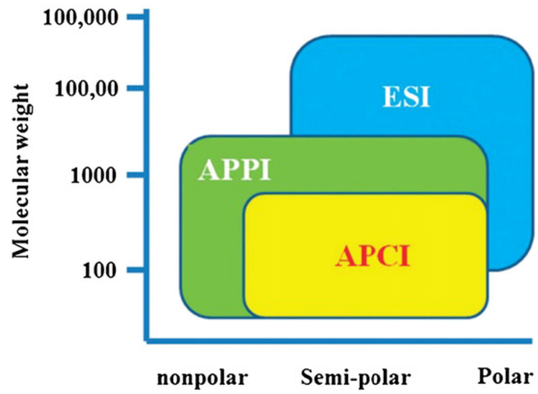

Fig. 1. Relative ionization capabilities of APCI, APPI, and ESI sources. Figure obtained/modified from reference [31]. Reprinted with permission from B. Zhou, J.F. Xiao, L. Tuli, H.W. Ressom, LC-MS-based metabolomics, Mol. Biosyst. 8 (2012) 470–481. Copyright 2012 Royal Society of Chemistry.

atom bombardment (FAB) and electrospray ionization (ESI). For metabolomics-based studies, ESI is by far the most common ionization source due to its ability to ionize both polar and semi-polar compounds ~~of~~ within a wide molecular weight range (Fig. 1) [31].

Mass spectrometers are broadly categorized as Fourier transform ion cyclotron resonance (FTICR), orbitrap, quadrupole and time of flight (TOF). FTICR mass spectrometers determine the mass-to-charge ratio (m/z) of ions, with a mass accuracy in the parts per billion range, based on the cyclotron frequency they exhibit while in a fixed magnetic field [32]. Orbitrap mass spectrometers possess a thin inner electrode and a cylindrical outer electrode [33]. A static voltage is applied to the inner electrode and ion (m/z) is assessed by measuring the frequency of harmonic oscillations observed when electrostatic attraction to the inner electrode is balanced by centrifugal forces [33]. Orbitraps are relatively inexpensive analyzers and offer mass accuracy in the low parts per million range [33]. Quadrupole analyzers are linear ion traps which utilize both quadrupole rods to radially confine ions and on-end electrodes to confine ions axially [34]. Quadrupoles tend to be the least expensive ion traps and are often employed in tandem with other analyzers [35]. TOF spectrometers determine (m/z) by measuring the specific time ions accelerated through electric fields of preset strength require to reach a detector and are far less destructive to samples than other spectrometric methods [36].

Hybrid/tandem mass spectrometers are constructed by using two or more of the previous mass analyzers in the same LC-MS system. For metabolomics studies, TOF and quadrupole mass analyzers are most commonly used, due to their ability to provide the accurate high mass resolution of MS/MS analysis [31,37].

These alternatives in LC-MS component configuration introduce a clear source of variation in the metabolomics data obtained. This is then further compounded by analytical conditions such as an elution gradient or isocratic solvent system, ionization mode used (positive or negative) and mass scan range (e.g. 50–1500 m/z vs. 100–1000 m/z) [31]. However, it is the flexible nature of LC-MS technology that is also its greatest advantage. The extreme diversity of metabolite chemo-physical properties or concentration, within a sample, precludes global assessment by any single analytical technology [38]. The ability to customize parameters to separate, detect, or even target a wide range of diverse molecules at low concentrations (e.g. pg mL^{-1}), makes LC-MS an ideal tool for metabolomic analysis [29,39].

An excellent example of system level metabolomic analysis is in the study of *Shewanella oneidensis* by Shen et al. [29], in which over 5000 metabolites were detected from a single experiment utilizing a capillary-RPLC-IT-TOF-MS/MS platform (Fig. 2). The potential for high-data output by LC-MS analytical methods makes it ideal for system level evaluation from which the data may then be analyzed by multivariate methods (e.g. principal component analysis (PCA)) or by high-order network based analysis [40].

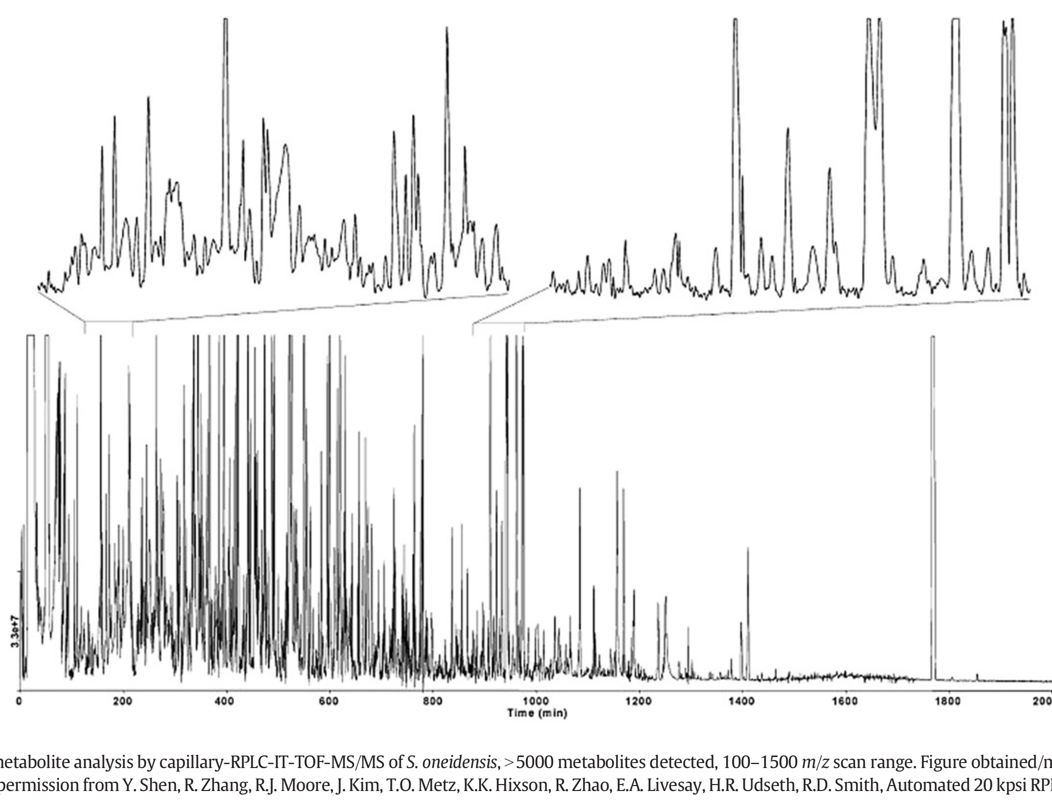

Fig. 2. System level metabolite analysis by capillary-RPLC-IT-TOF-MS/MS of *S. oneidensis*, >5000 metabolites detected, 100–1500 m/z scan range. Figure obtained/modified from reference [29]. Reprinted with permission from Y. Shen, R. Zhang, R.J. Moore, J. Kim, T.O. Metz, K.K. Hixson, R. Zhao, E.A. Livesay, H.R. Udseth, R.D. Smith, Automated 20 kpsi RPLC-MS and MS/MS with chromatographic peak capacities of 1000–1500 and capabilities in proteomics and metabolomics, Anal. Chem. 77 (2005) 3090–3100. Copyright 2005 American Chemical Society.

3.3. MSI based approaches

MS utilized in metabolomics studies is commonly based on assessing metabolite mass combined with one of several sample introduction techniques. One distinct advantage of utilizing MS analysis is the ability to obtain a mass accuracy of less than 1 ppm. In the full-scan mode, the mass spectrometer performs comprehensive analysis of all ions in a sample, within a predetermined range (e.g., 100–1000 Da). These data can include millions of data points, requiring processing into individual components. Most metabolomics MS platforms analyze samples initially processed by liquid chromatography (LC-MS) or gas chromatography (GC–MS) techniques. These interfaces are complimentary and overlapping analyte data provide corroboration when these techniques are combined.

3.4. NMR-based approaches

Recent studies have shown impressive advances in biochemical analysis with NMR. Most nuclei in a biological sample possess a nuclear spin, which allows for convenient, rapid analysis with NMR spectroscopy, and a multitude of multidimensional experiments have been designed. Given the abundant and ubiquitous presence of ^{1}H in most metabolites, one-dimensional (1D) ^{1}H NMR spectroscopy is the primary source of NMR-based metabolomics data. Unfortunately, metabolite analysis by NMR platforms alone does not match the sensitivity of MS based platforms, but the highly reproducible, quantitative results of NMR, ensure it will remain extremely valuable in biochemical studies, especially when combined with MS based techniques.

3.4.1. NMR-based diagnostic metabolomics

The informative analytical techniques of NMR and MS combined with systemic statistical analyses are a promising approach for a myriad of metabolomics-based studies. The application of NMR-based metabolomics is extremely diverse, ranging from elucidating environmental toxicology to identifying human diseases, since NMR-based research is more reproducible, nondestructive and embraces a wider dynamic range.

For example, rabies is a deadly viral infection of the central nervous system (CNS) and currently there are no antiviral drugs clinically available for treatment. O'Sullivan et al. [41] elucidated the mechanisms of rabies pathogenesis employing NMR-based metabolomics. They examined cerebrospinal fluid (CSF) collected from rabies infected patients over the course of their disease (survivors and nonsurvivors) and discovered three distinct stages of rabies progression. Several metabolites associated with energy metabolism and cell volume control, which distinguished rabies survivors from nonsurvivors, were elucidated via the quantification from NMR spectra of the CSF samples. This metabolomics approach may cast light on future therapy for this incurable disease.

Another case study employing NMR techniques examines CSF in order to diagnose leptomeningeal carcinomatosis (LMC). This metastasis is one of the most common complications of the central nervous system but no satisfactory diagnostic tool has been established. In the aforementioned study, Cho et al. [42] utilized a high-resolution magic angle-spinning (HR-MAS) probe which is optimal for obtaining a high-quality NMR spectrum with a limited CSF sample volume (~40 μL). Their procedure for establishing a metabolomic profile was to gather representative NMR spectra for normal, three-day LMC and seven-day LMC rat models, to identify several metabolites (LMC marker metabolites) reported to be present in CSF including lactate, glucose, total creatine and acetate, and to analyze the spectra utilizing the Wilcoxon rank sum tests. Analysis of the LMC marker metabolites using the normalized average spectra of the relevant peaks (See Fig. 7 in the reference 42), especially in conjunction with glucose and t-creatine levels, could readily diagnose LMC samples. This NMR-based metabolomics approach for LMC diagnosis could be meaningful because it only requires 40 μL of CSF which is offered from routine CSF sampling. Indeed, CSF samples could be further utilized to provide clinicians and patients with more accurate and fast diagnosis for LMC than current methods such as magnetic resonance imaging (MRI) which possesses low resolution for the application.

Bladder cancer diagnosis is another example where NMR-based metabolomic analysis can provide an exceptionally sensitive and specific non-invasive monitoring approach [43]. Metabolomics enables analysis of a significant number of small molecules in one step, which presents vast potential for exploring marker compounds to monitor a treatment response and recurrence in patients suffering from bladder cancer [44]. Zhang et al. [44] profiled the metabolites in urine of dogs with transitional cell carcinoma (TCC) which was verified to be similar with human invasive bladder cancer [45] and control dogs utilizing NMR and statistical analysis methods. For the NMR approach, dog urine was stabilized with phosphate buffer to produce adequate ^{1}H NMR spectra and twenty-one metabolites were subsequently de-replicated by comparison of the acquired and reported NMR spectroscopic data. Consecutively, all the NMR spectra were analyzed by orthogonal signal correction pretreated partial least squares-discriminant analysis (OSC-PLS-DA) to discover the significant difference between the urine of dogs with TCC from healthy controls. The predictions for the true class assigned models and

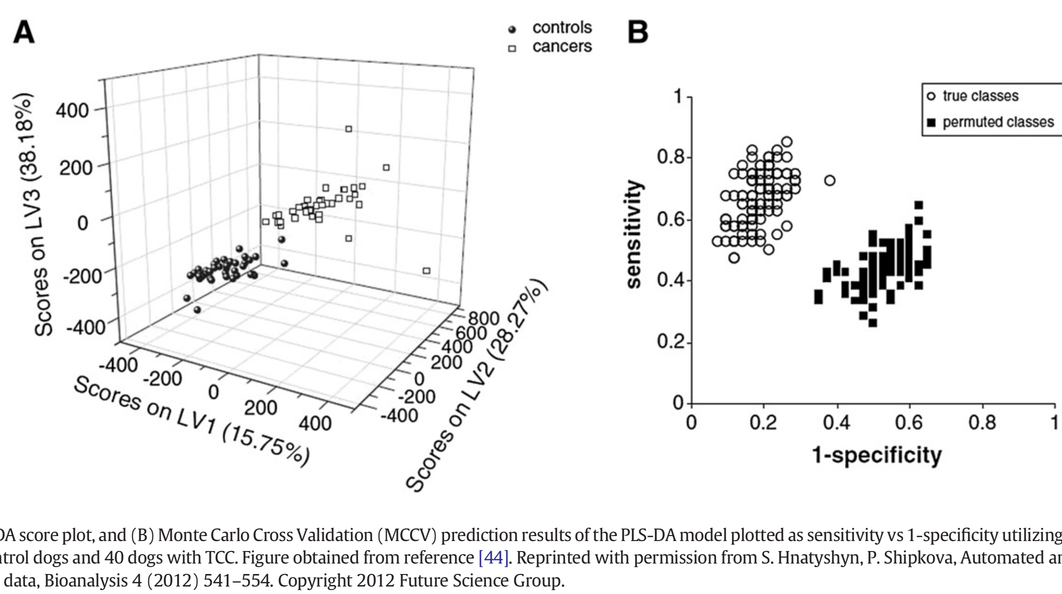

Fig. 3. (A) OSC-PLS-DA score plot, and (B) Monte Carlo Cross Validation (MCCV) prediction results of the PLS-DA model plotted as sensitivity vs 1-specificity utilizing the ^{1}H NMR spectra of samples from 42 control dogs and 40 dogs with TCC. Figure obtained from reference [44]. Reprinted with permission from S. Hnatyshyn, P. Shipkova, Automated and unbiased analysis of LC-MS metabolomic data, Bioanalysis 4 (2012) 541–554. Copyright 2012 Future Science Group.

permuted models were visualized utilizing a Receiver Operating Characteristic (ROC) space (Fig. 3B) to further validate the analysis (i.e., OSC-PLS-DA analysis). The permuted model resulted in a clustered pattern at the center of the ROC, which indicates no discrimination. The true class assigned models displayed high sensitivity and specificity, verifying the classification capacity of the employed statistical model.

A review of the patent literature reveals an increasing number of metabolomics-based applications, with the majority being focused toward clinical disease detection. In most clinical cases, the earlier the disease is diagnosed the greater the patient's prognosis improves. This is especially true of cancers arising from four major sites: lung, colorectum, breast, and prostate [46].

Targeting lung cancer Imaizumi et al. [47] patented a screening method where by a patient's blood sample would be screened for the whole amino acid profile and analyzed by multivariate analysis to discriminate between lung cancer and specific stages. Approximately half of all lung cancers are detected at such a late stage they are inoperable, but if detected at clinical stage IA the mean patient survival rate nears 90%. Their process is non-invasive and allows for the potential detection of lung cancer before any of the usual clinical markers are apparent [47].

Fedorak and Wang [48] patented a diagnostic screen for colorectal cancer and colorectal polyps by metabolomic analysis of 69 metabolites in a patient's urine sample. Their screen is non-invasive and can potentially detect colorectal cancer at a much earlier stage than current methods such as fecal occult blood tests or colonoscopy [48]. Their improved screening method has the potential to help curtail morbidity rates from colorectal cancer, a leading cause of cancer morbidity [49]. Raftery et al. [50] patented a method for the early detection of biomarkers indicative of breast cancer from a patient's serum sample. An integrated analytical approach employing two-dimensional gas chromatography -mass spectrometry and nuclear magnetic resonance was used to obtain a metabolic profile of the patent's serum which was then analyzed by multivariate statistical methods targeting a mixture of 40 metabolites [50].

Targeting prostate cancer Weinberger et al. [51] filed a patent detailing a metabolomics screening procedure in which patient blood serum was subjected to combinatorial instrument screening (e.g. HPLC-MS and NMR) to assess 221 specific metabolites as biomarkers for the diagnosis of prostate cancer [51]. The most important similarity that these methods have for the detection of the four major cancer sites are that they are non-invasive and can potentially detect at a much earlier stage in the cancer's progression than currently used methods.

Other metabolomics-based patents targeting disease diagnosis of non-cancerous origin include the screen developed by Lundin and Weinberger [52] to assess kidney disease by screening blood and urine samples for a mixture of biomarkers consisting of two amino acids, two acylcarnitines and two biogenic amines [52]. Their screen allows for disease staging, as well as early stage detection. Additionally, Cezar [53] patented a method for non-invasive analysis of multiple biofluids (e.g. amniotic fluid, blood, plasma, urine, etc.) for metabolomic signatures indicative of autism. PCA identified 6000 statistically significant metabolic features between autistic and non-autistic controls of which they further focused by use of LC-ESI-TOF-MS and HILIC (still allowing for the use of alternative instrumentation for metabolite profiling) on 98 HILIC specific metabolites and 47 C18 specific metabolites for screening [53]

3.4.2. NMR-based plant metabolomics

Plant metabolome-based drug discovery is challenging because of the inclusion of a complex matrix consisting of many uncharacterized secondary metabolites with a wide dynamic range of variations of major compounds, minor compounds, polarity, boiling and melting points, etc. To address this demand, NMR-based metabolomics could be instrumental in comprehensive qualitative and quantitative analyses of the vast metabolites originating from plants because reproducibility is thought to be the most critical prerequisite for formulating metabolomics-based drug discovery research. Furthermore, NMR-based metabolomics could be more useful in deciphering complex plant metabolisms when they are assessed in combination with multivariate data analysis. Pan et al. [54] demonstrated an application of NMR-based metabolomics with the combination of multivariate data analysis for the quantitative analyses of monoterpenoid indole alkaloid (MIA) from *Catharanthus roseus*, which has been shown to produce vinblastine through the MIA pathways. The authors designed two types of a transgenic plant overexpressing ORCA3 (Octadecanoid-derivative Responsive Catharanthus AP2-domain), a key regulator of several crucial genes in MIA biosynthesis including *DXS*, *AS*, *TDC*, *STR* and *D4H*, and co-overexpressing ORCA3 and G10H (geraniol 10-hydroxylase), which catalyzes the first step of the MIA biosynthesis pathway. Metabolites from each overexpressing plant tissue were measured and analyzed utilizing NMR and a multivariate analysis method, PLS-DA (Fig. 4B). While the ^{1}H NMR spectra of the co-overexpressing transgenic plant and the control possessed a high degree of homology (Fig. 4A), in-depth analysis of the spectra datasets employing squares-discriminant analysis (PLS-DA) revealed a distinct metabolomic separation between the co-overexpressing transgenic plant and control (Fig. 4B). Additional PLS-DA loading plot analysis found that signals of alanine, glutamic acid, arginine, glucose, sucrose, 2,3-butanediol, quercetin-3-*O*-glucoside, strictosidine, vindoline and catharanthine were up-regulated in the ORCA3- and G10 co-overexpressing transgenic plant, while the controls displayed enhanced levels of threonine, secologanin, 2,3-DHBA, chlorogenic acid, 4-*O*-caffeoyl quinic acid, malic acid and fumaric acid. These metabolic discriminations suggest that ORCA3 and G10H co-overexpression enhancement of MIA biosynthesis could affect other metabolic pathways in *C. roseus* metabolism.

3.5. Integrated approaches

The famous quote by Aristotle “The whole is greater than the sum of its parts” is uniquely exemplified when analytical technologies are combined in metabolomics studies. For example, a metabolomics study utilizing GC–MS, LC-MS, data alignment software (MetMAX Beta 1.0) and multivariate statistical analysis software (COVAIN 1.0) determined the chemical composition of a premenstrual syndrome (PMS) medication to be inconsistent between different pharmaceutical vendors. This integrated metabolomics platform generated a high resolution summary of the chemical composition of each medicine and provided data indicating medicinal values would vary between vendors [55].

3.6. Computational tools

The composite nature of the data resulting from metabolomics-based analysis is sufficiently complex as to require methods of statistical multivariate data analysis to interpret. PCA is the prevailing method for reducing data dimensionality, though other data-mining techniques such as clustering algorithms (hierarchical clustering, tree clustering, K-means, etc.) are routinely used [56,57]. With PCA, major sample components are structured to represent the data variance in a two-/three-dimensional coordinate scheme. This technique reveals grouping patterns of samples and observational events that can visually discern outliers which may indicate those samples to be potential sources for novel natural products and thus valuable targets for subsequent analysis [58]. The overall goal of PCA is to reduce the dimensionality of a dataset, while retaining data quality. In datasets containing several groups of variables, some variables often show redundant information. In mass spectra, the variables are the intensities at specific masses. PCA reduces the number of dependent variables in the set by replacing groups of inter-correlated variables with a single new variable, reducing convoluted variables into manageable principal components. Excellent PCA software tools are widely available, (e.g., ClinprotTool 2.0; Bruker Daltonics) and can substantially improve the quality and efficiency of metabolomics platforms. In a study of Ontario ginseng (*Panax quinquefolius* L.) [59], PCA analysis of ^{1}H NMR spectra elucidated unique biomarkers between closely-related landraces supporting possible improvement in quality assessment and species authentication.

4. Applications in phylogeny

Metabolomic analysis evaluates organisms at the system level with the unique ability of providing a snapshot of the functional genetic status of the organism by examining its metabolic profile. An important application of this research is its use in further resolving phylogenetic associations based solely upon classical genomic sequencing. A prominent example would be in cases of horizontal gene transfer, commonly seen as the transfer of peripheral genes [60–62]. A reported example of horizontal gene transfer is the genetic cluster coding for the production of sterigmatocystin, a highly toxic aflatoxin precursor, from *Aspergillus nidulans* being identified in *Podospora anserina* [63].

Genomically, *A. nidulans* and *P. anserina* are highly divergent species (Fig. 5), however Slot et al. [63] observed the existence of an intact 24 gene cluster encoding for the production of sterigmatocystin in *P. anserina* that was analogous to the cluster found in *A. nidulans*. In addition, the extremely high degree of genetic conservation displayed by the observed cluster in terms of both gene order and orientation was especially interesting as they did not find similar clusters beyond the *Aspergillus* lineage. By employing a phylogenetic analysis of 23 cluster genes, they revealed six topological patterns (Fig. 6) supporting their hypothesis that the ability of *P. anserina* to produce sterigmatocystin was obtained by a horizontal gene transfer from *A. nidulans*. Ultimately, this represents an example of two otherwise metabolically dissimilar species that would be phylogenetically clustered if assigned based upon all or portions this horizontally transferred genomic region.

Metabolomics can also aid in discriminating phylogenetic associations in cases where genomic sequence diversity is either too low or too similar for efficient resolution. A study by Dieckmann et al. [64] of numerous bacteria isolated from marine sponges unequivocally resolved several species problematic to resolve genetically due to extreme relatedness (e.g. 99.6% identity in 1400 bp 16S rDNA sequencing) [64, 65]. They found incorporating metabolomic analysis by use of intact cell MALDI-TOF-MS to be effective in producing additional phenotypic markers to quickly and reliably discriminate between species due to their unique metabolite profile. For example, when they compared the metabolomic fingerprints of *Pseudoalteromonas* species and *Alteromonas* species there were no overlapping biomarkers observed (Fig. 7). Furthermore, metabolomic analysis can readily differentiate between *Pseudoalteromonas* species that are both difficult and laborious to discriminate using classical genetic methods (Fig. 5).

In their study, Dieckmann et al. [64] mixed a portion of their bacterial isolates with matrix solution and spotted directly onto a target well of a sample plate. After allowing the sample/matrix mixture to air dry it was analyzed by MALDI-TOF-MS. Their mass spectra data was obtained using a linear delayed extraction mode and they averaged 100 sample sites along the width of the target well focusing on a mass range of 2000–20,000 m/z.

The MS data was then analyzed by multivariate clustering. Proceeding from baseline correction they ranked the detected metabolite signals relative to computed intensities: the most intense peak was set to 100%, 2000–4000 Da below 40%, 4000–10,000 Da below 10%, 10,000–20,000 Da below 5%. A spectral grouping was defined as metabolic signals between 4000–20,000 Da being present as five or more of the most intense signals. The resulting mass-relative intensity lists were then used in a multivariate statistical package (Kovach Computing Services) to construct the subsequent scatter plots and dendrograms (Fig. 8) based upon the MALDI-TOF-MS data.

Perhaps most importantly, metabolomic analysis allows for the examination of emergent functional genomic responses of an organism due to interactions with its environment which results in the development of complex phenotypes [66,67]. Metabolomic analysis facilitates the combined phenotypic comparison of species–environment interactions with genetic approaches by providing direct links to changes in metabolites and their originating biochemical pathways. This type of analysis ultimately aids in elucidation of gene function and determining effects of genetic alterations [68–70].

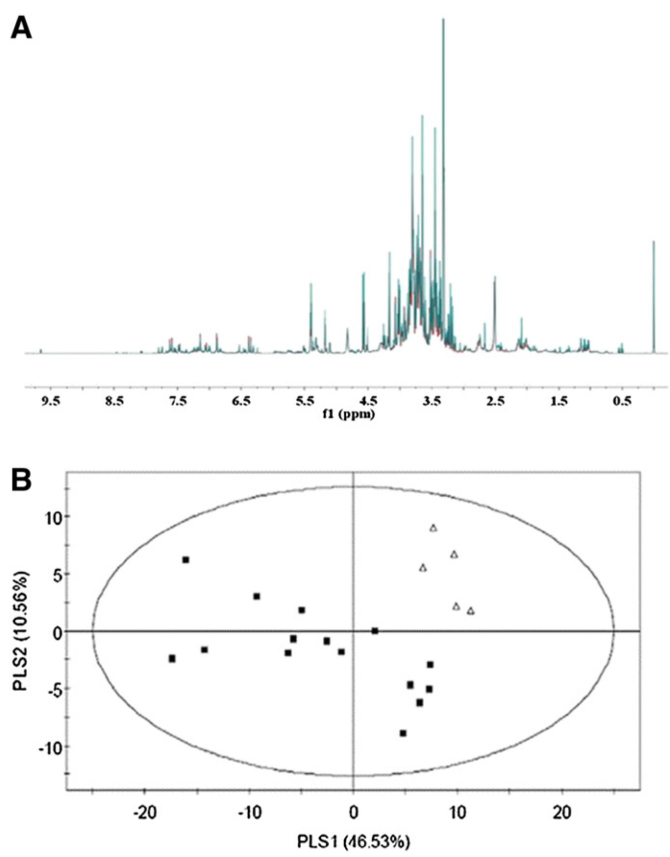

Fig. 4. A: ^{1}H NMR spectra of the co-overexpressing transgenic plant (in green) and control (in red) lines. B: Score plot of PLS-DA of the co-overexpressing transgenic plant (△) and control (■) lines. Figure obtained from reference [54]. Reprinted with permission from Q. Pan, Q.Wang, F. Yuan, S. Xing, J. Zhao, Y.H. Choi, R. Verpoorte, Y. Tian, G.Wang, K. Tang, Overexpression of ORCA3 and G10H in Catharanthus roseus plants regulated alkaloid biosynthesis and metabolism revealed by NMR-metabolomics, PLoS One 7 (2012) e43038. Copyright 2012 PLOS.

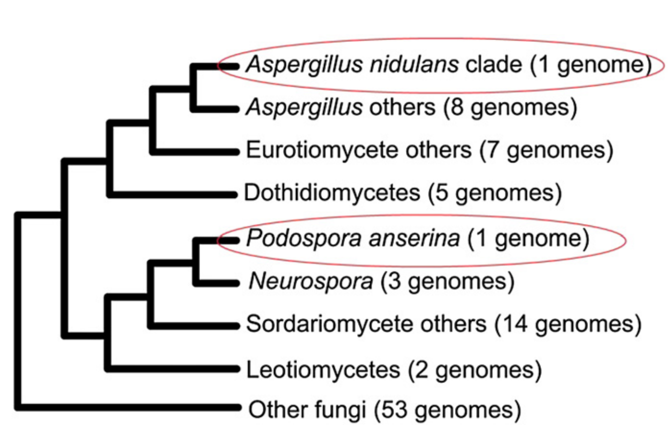

Fig. 5. Genomic phylogenies of evaluated fungal species. Figure obtained/modified from reference [63]. Adapted with permission from J.C. Slot, A. Rokas, Horizontal transfer of a large and highly toxic secondary metabolic gene cluster between fungi, Curr. Biol. 21 (2011) 134–139. Copyright 2011 Cell Press.

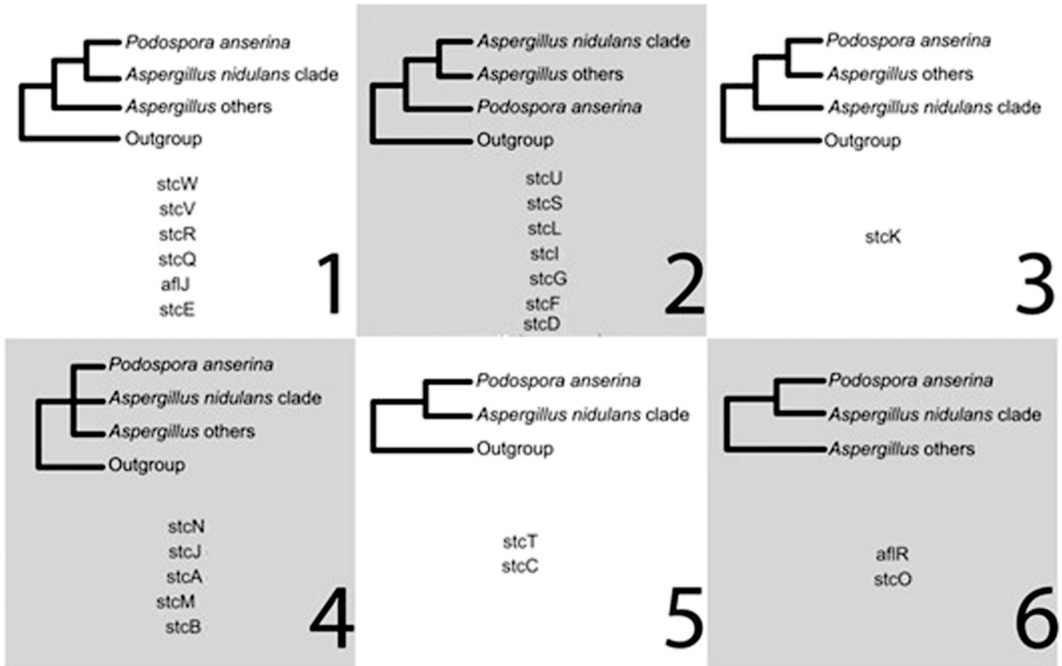

Fig. 6. Cluster gene supported phylogenies. Figure obtained/modified from reference [63]. Adapted with permission from J.C. Slot, A. Rokas, Horizontal transfer of a large and highly toxic secondary metabolic gene cluster between fungi, Curr. Biol. 21 (2011) 134–139. Copyright 2011 Cell Press.

5. Applications in disease diagnosis

Disease, regardless of its specific origin (e.g. bacterial, viral, acute, chronic, etc.), disrupts normal metabolism and thus presents a potential target for disease diagnosis. Broadly, there are two metabolomic profiling strategies for this: (1) metabolic fingerprinting, the metabolites of interest are not known a priori and diagnosis relies on pattern recognition techniques. (2) Metabolites of interest are defined a priori, verified biomarkers, and are precisely quantified. These strategies make it possible to produce metabolic profiles to differentiate signatures from diseased states from those of healthy controls [71,72].

As the metabolomic profile is the summation of transcriptional, translational, environmental interactions and operates on a measureable timescale of fractions of a second it is the most rapid and sensitive measurement of a systems, phenotype [73]. This makes metabolomic analysis an immensely useful tool in assessing the real-time biochemical changes resulting from both disease progression, as well as, therapeutic treatment. Additionally, this type of analysis can

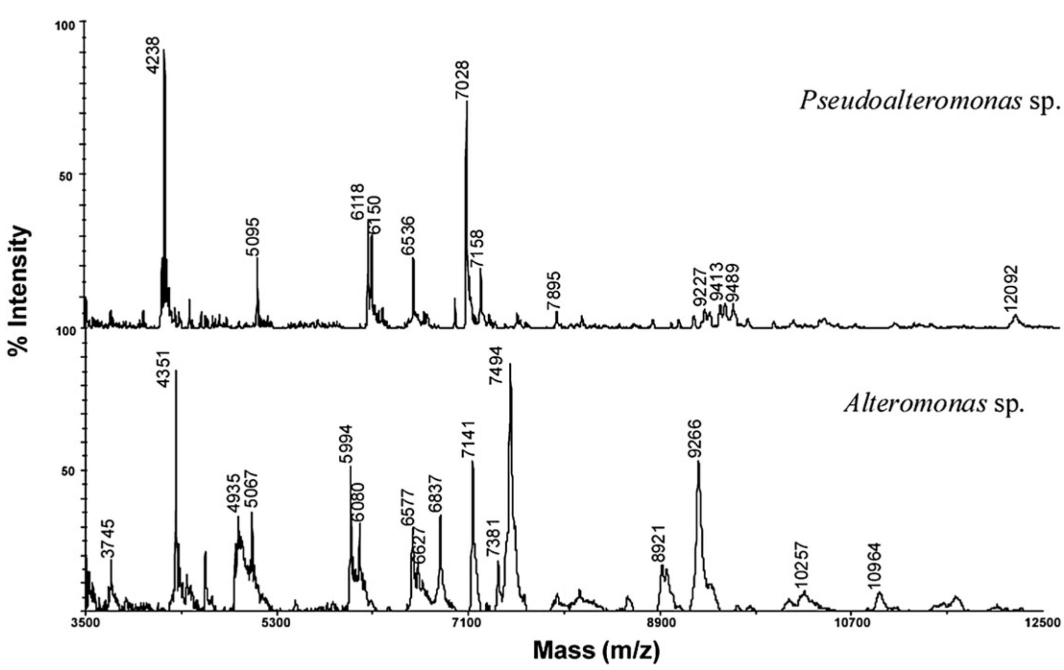

Fig. 7. MALDI-TOF-MS fingerprint analysis of *Pseudoalteromonas* sp. and *Alteromonas* sp. Figure obtained/modified from [64]. Reprinted with permission from R. Dieckmann, I. Graeber, I. Kaesler, U. Szewzyk, H. Von Döhren, Rapid screening and dereplication of bacterial isolates from marine sponges of the Sula Ridge by Intact-Cell-MALDI-TOF mass spectrometry (ICM-MS), Appl. Microbiol. Biotechnol. 67 (2005) 539-548. Copyright 2005 Springer Berlin Heidelberg.

be applied towards the diagnosis if complex disease states which do not exhibit a clear genotypic expression, an important example of which would be Parkinson's disease [74]. Along these lines, preliminary metabolomic profiling has been performed encompassing several diseased states, such as diabetes, cardiovascular disease, Alzheimer's disease and Huntington's disease [72,75–79].

Metabolomic analysis has been applied to detect early stage biomarker candidates for diabetic kidney disease, which is incurable and affects approximately 33% of type 1 diabetes mellitus patients. However, if diabetic kidney disease is detected early enough (i.e. pre-clinically) its progression may be halted. Currently diabetic kidney disease is clinically detected by abnormally high urinary albumin excretion rates.

Using a combination of GC–MS and LC-MS analysis of urine samples, van der Kloet et al. [80] identified a combination of 34 (from an initial 130) biomarker candidates from 52 patients. These metabolomic biomarkers distinguished which patients would develop abnormally high urinary albumin excretion rates from those that would not with 75% accuracy (with a 5.5 year follow-up) [80]. A key observation was that subtle metabolic changes preceded clinical presentation of diabetic kidney disease. Similar preclinical biomarkers, which may only be detectable through holistic approaches such as metabolomic analysis, may yet be discovered for numerous other disease states.

Cancer is a major disease class target for the development of rapid, reliable, non-invasive, methods to detect and differentiate benign and malignant tumors [71,72]. Metabolomic analysis using LC-MS and multivariate statistical analysis by PCA of patient serum samples allowed for the clear separation of healthy control samples and those of confirmed renal cell carcinoma, even differentiating between healthy/early/late stage renal cell carcinoma [81]. Lin et al. [81] utilized both reversed phase liquid chromatography (RPLC) and hydrophilic interaction liquid chromatography (HILIC) coupled mass spectrometry for the separation of a total of 2140 divergent metabolites from patient serum samples, 1384 from RPLC and 756 from HILIC. The detected metabolites were assessed according to variable importance in the project (VIP) values, which ultimately identified 58 biomarker candidates by RPLC and 36 by HILIC potentially applicable towards the detection of renal cell carcinoma. The data was subsequently analyzed by PCA and partial least squares-discriminant analysis producing clear separation between the control patients and those with confirmed renal cell carcinoma (Fig. 9A–C). Furthermore, their targeted metabolomic analysis proved sensitive enough to even discern the stage of cancer progression (Fig. 9D).

In a similar study targeting cancer, Sugimoto et al. [82] analyzed saliva samples from 69 patients with oral cancer and 87 control patients using capillary electrophoresis time of flight-mass spectrometry (CE-TOF-MS) to detect biomarker candidates. They detected 3041 peaks per saliva sample, ultimately discriminating 28 metabolites co-occurring in both test groups (Fig. 10). Expanding upon this work they then included 30 breast cancer patients, 18 pancreatic cancer patients, and 11 periodontal disease patients into their saliva test groups. Their CE-TOF-MS analysis identified the following number of specific metabolites as potential biomarker candidates: 28 for breast cancer, 48 for pancreatic cancer and 27 for periodontal disease [82]. Subsequently, using a multiple logistic regression model to discern those detected peaks most applicable for discrimination between the test groups settled on 57 metabolites that showed a significant difference ($p > 0.05$) between the control group and one or more disease groups (Fig. 11) [82,83].

Ultimately, the capability of metabolomic profiling to assess the real-time biochemical changes may provide earlier opportunities for intervention in disease progression, ideally modifying or even arresting its course, over the current system of assessing endpoint symptoms of disease states by their clinical manifestations [75,84].

6. Utility in natural products dereplication and drug discovery

The inherently complex nature of biological extracts resulting from the search for novel natural products gives rise to two inevitable conditions: 1) they will often contain numerous constituents and 2) many of those will be known compounds. Thus, it is extremely valuable to screen the metabolome as early as possible in the natural product discovery process to identify those extracts that are likely to possess novel metabolites from those that are not [64]. The capability of metabolomics to analyze a broad range of metabolites, at the system level, makes it a crucial tool in the de-replication process.

Due to the extreme complexity of natural product extracts, a single analytical method capable of profiling all metabolites of all extracts at once is precluded [85]. However, combinatorial chromatographic and spectroscopic techniques, such as gas and liquid chromatography,

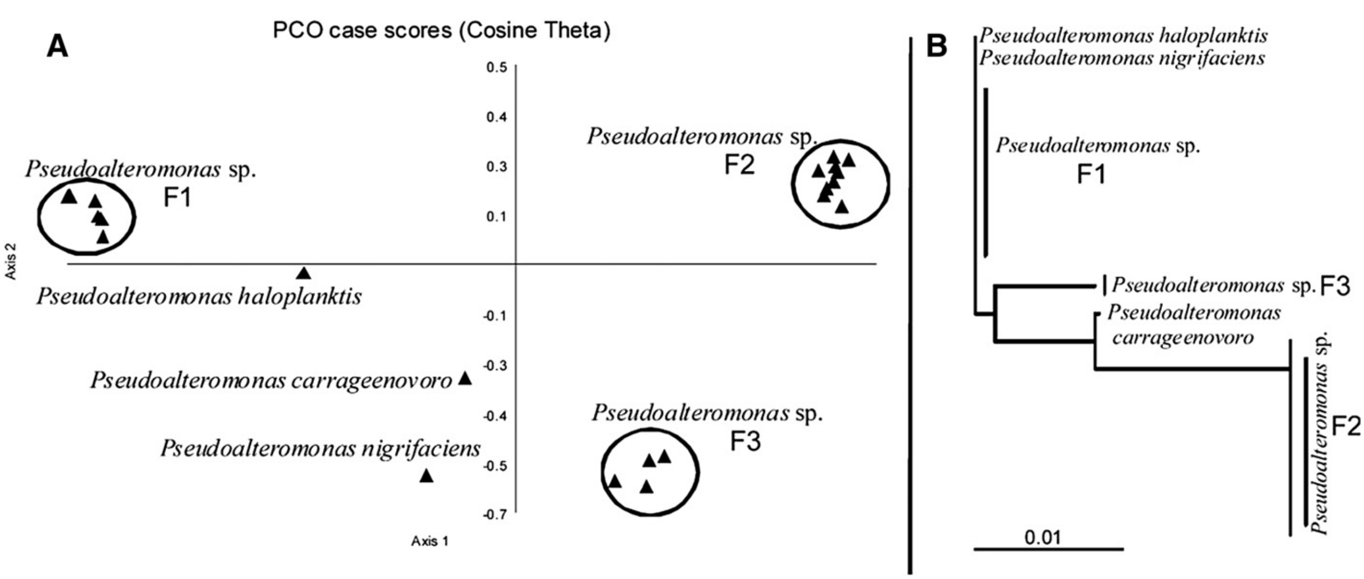

Fig. 8. (a) Principal coordinate analysis scatterplot of metabolites from *Pseudoalteromonas* sp. detected by MALDI-TOF-MS. (b) 16S rDNA based phylogenetic tree of *Pseudoalteromonas* sp. Figure was obtained/modified from [64]. Reprinted with permission from R. Dieckmann, I. Graeber, I. Kaesler, U. Szewzyk, H. Von Döhren, Rapid screening and dereplication of bacterial isolates from marine sponges of the Sula Ridge by Intact-Cell-MALDI-TOF mass spectrometry (ICM-MS), Appl. Microbiol. Biotechnol. 67 (2005) 539-548. Copyright 2005 Springer Berlin Heidelberg.

infrared (IR) spectroscopy, mass spectrometry (MS), nuclear magnetic resonance (NMR) spectroscopy and ultraviolet absorption (UV) spectroscopy, are proven efficient methods at achieving the most comprehensive analysis of complex extracts [57,86,87]. Hyphenated mass spectrometry techniques are routinely used to detect hundreds of metabolites within a single extract. Particularly suited towards metabolome-based natural products characterization is matrix-assisted laser desorption/ionization-time of flight-mass spectrometry (MALDI-TOF-MS). MALDI-TOF-MS is an imaging mass spectrometry technique that allows for the analysis of complex biological mixtures and, most excitingly, allows for the capability to obtain direct visualization of the spatial distribution of the analyte(s) within the biological matrix [88]. Moreover, this method of data acquisition allows for the unique chemo-spatial examination of, for example, the interplay between microorganism, both inter- and intra-species, and symbiotic associations [89]. This allows for the unambiguous determination of the actual physical source of isolated compounds of interest.

Of distinct value in the characterization of natural products is the use of nuclear magnetic resonance (NMR) spectroscopy, either individually or in a tandem setting. This is due to the capacity of NMR to provide detailed structural information on the individual constituents of complex biological mixtures to a greater extent than that of the preceding techniques [90]. Furthermore, an ideal feature of NMR is the nondestructive nature of its analysis; especially important when samples are derived from organisms that are difficult to culture or obtain [91]. However, access to sufficiently developed and maintained spectral databases is central to the effectiveness of these techniques [92,93]. This highlights the importance of the development of open access databases where metabolomic data can be maintained and properly annotated [85].

Hou et al. [94] employed LC-MS-PCA as a means of quickly achieving bacterial strain de-replication in their pursuit of novel natural products. Evaluating *Micromonospora* spp. and *Verrucosispora* spp. isolated from tropical ascidians they observed that cultivated microbes may appear morphologically different and possess different 16S sequences while still producing the same secondary metabolites. Conversely, strains may appear identical, with identical 16S sequence homology, and produce different secondary metabolites (Fig. 12). Ultimately, as their target was novel natural products (i.e. metabolic end-products) first-pass de-replication by use of metabolomic analysis was far more practical than indirect genetic analysis [94].

In addition to prioritizing sources of novel natural products, metabolomic analysis can also serve to identify novel mechanisms of action for drugs. Birkenstock et al. [95] investigated the antibacterial activity of triphenylbismuth dichloride (TPBC) against a series of multi-drug resistant bacterial pathogens, such as *Staphylococcus aureus*, by quantifying the extracellular metabolites of bacterial cultures by ^{1}H NMR. What they observed in cultures treated with TPBC was an alteration in the exometabolome indicating a TPBC induced metabolic dysfunction halted glucose oxidation at the pyruvate level, essentially preventing core metabolic activity. By probing the known pyruvate-pathway of *S. aureus* it was possible to identify bacterial pyruvate dehydrogenase complex as the site of action for TPBC (Fig. 13) and propose the site as a potential drug target in antibiotic-resistant bacteria [95].

The high-data output of metabolomic screening also makes it ideal for network based analysis. The holistic nature of metabolomics allows for the construction of systems level biological networks to be constructed and queried [40]. From the point of view of natural products research, metabolomics derived spectral networks can readily separate compounds by structural relatedness into discrete clusters, and allow the focus to be on nodes which representing the most chemically unique structures (Fig. 14) [96].

When metabolomic analysis is employed as part of a multi-component network with correlative data from proteomic and genomic analysis, it offers the potential to most closely model complex systems. When coupled with proteomic data, such network mapping allows for the querying of potential drugs together with potential protein targets. When coupled with genomic data, such network

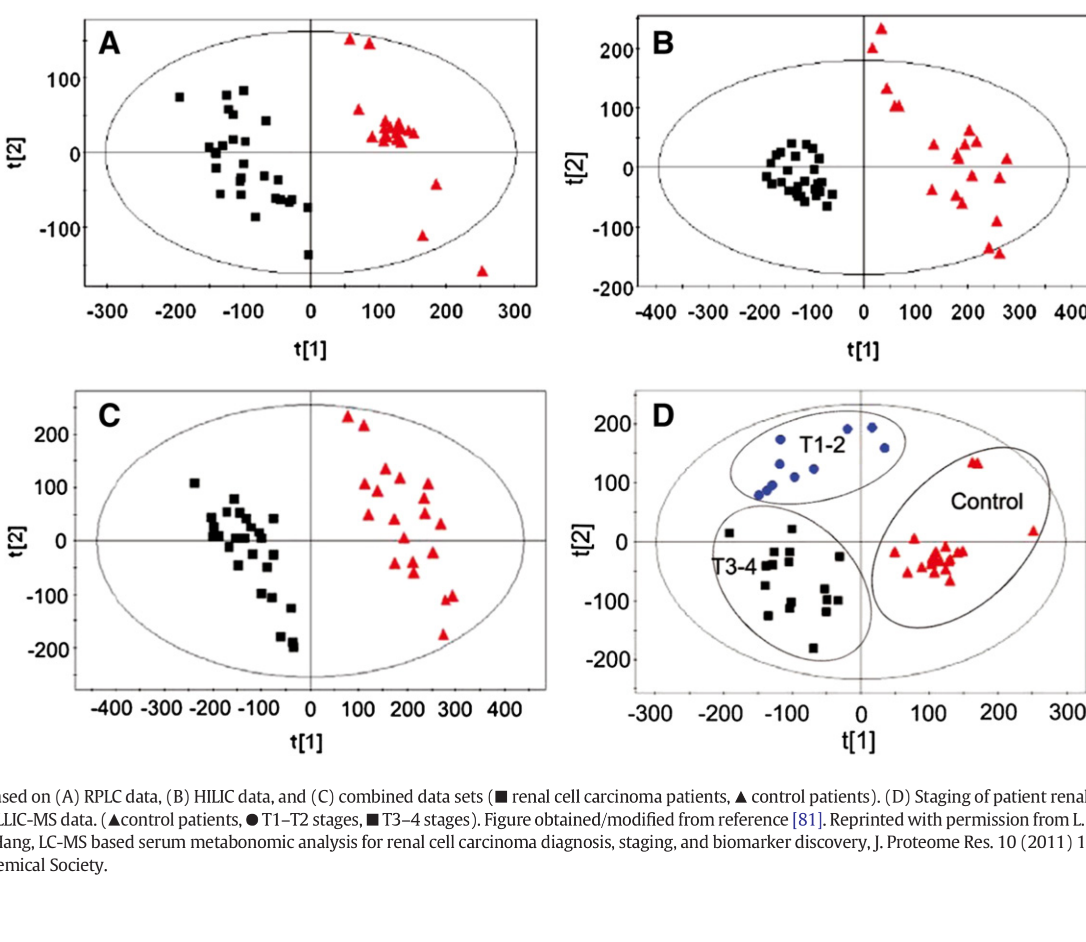

Fig. 9. Score plot based on (A) RPLC data, (B) HILIC data, and (C) combined data sets (■ renal cell carcinoma patients, ▲ control patients). (D) Staging of patient renal cell carcinoma from combined RPLC/HILLIC-MS data. (▲control patients, ● T1–T2 stages, ■ T3–4 stages). Figure obtained/modified from reference [81]. Reprinted with permission from L. Lin, Z. Huang, Y. Gao, X. Yan, J. Xing, W. Hang, LC-MS based serum metabonomic analysis for renal cell carcinoma diagnosis, staging, and biomarker discovery, J. Proteome Res. 10 (2011) 1396-1405. Copyright 2011 American Chemical Society.

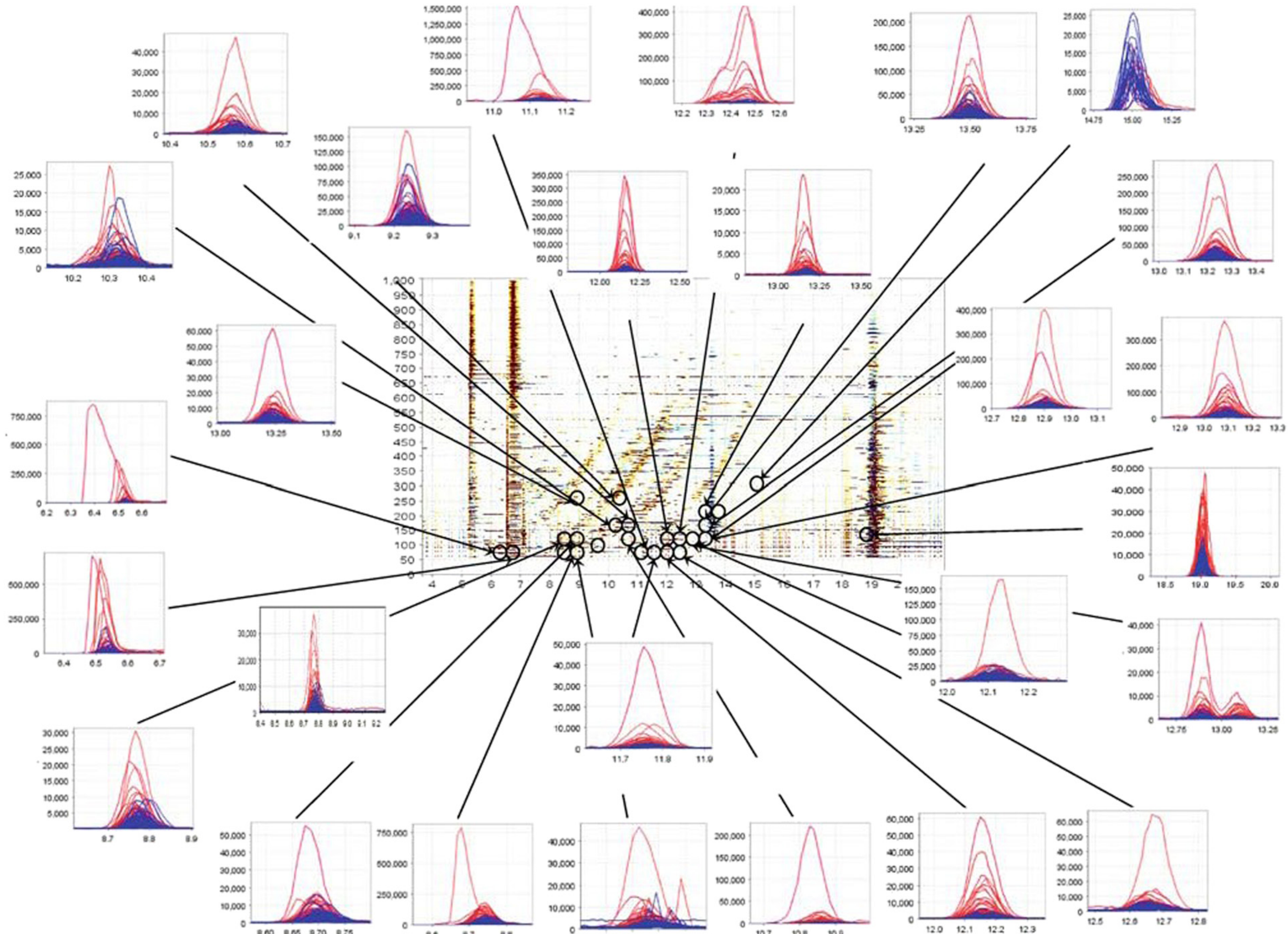

Fig. 10. CE-TOF-MS profile of salivary metabolites from patients with oral cancer (n = 69) and control samples (n = 87). X axis is migration time and Y axis is *m/z*. Circled peaks are significantly different ($p < 0.05$) between the two groups. Red corresponds to oral cancer group and blue to control. Figure obtained/modified from reference [82]. Reprinted with permission from M. Sugimoto, D. Wong, A. Hirayama, T. Soga, M. Tomita, Capillary electrophoresis mass spectrometry-based saliva metabolomics identified oral, breast and pancreatic cancer-specific profiles, Metabolomics 6 (2010) 78-95. Copyright 2010 Springer US.

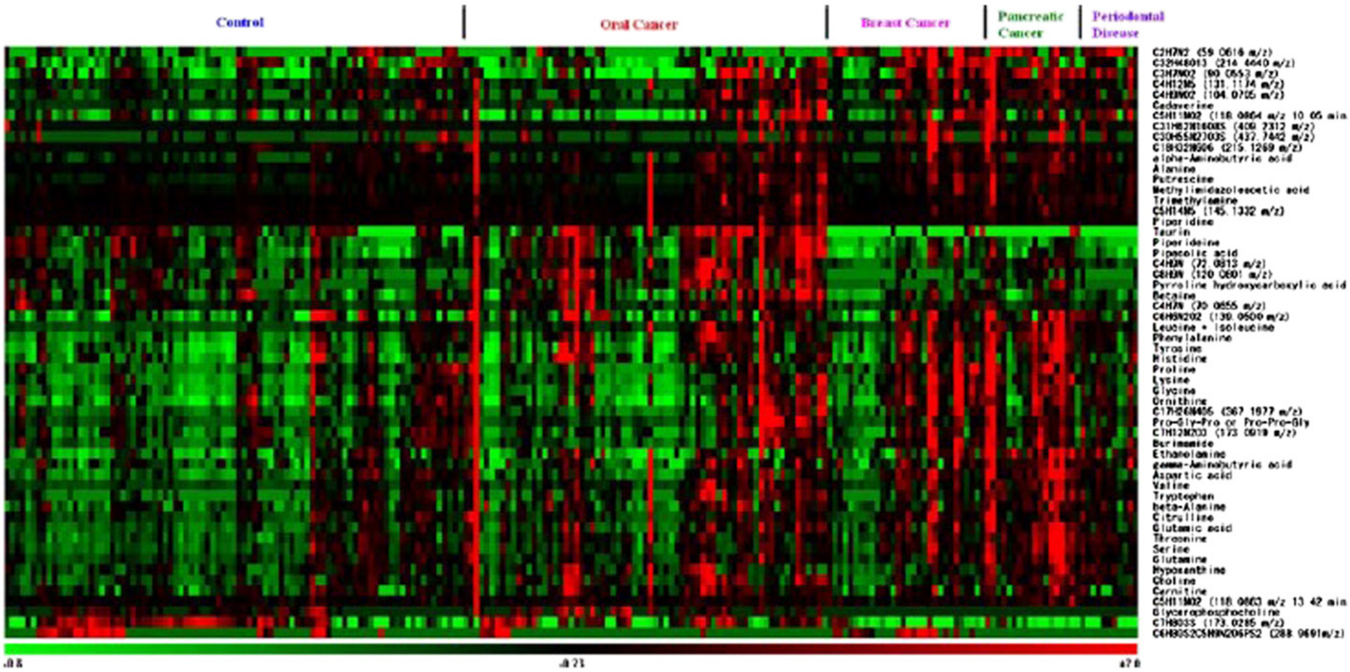

Fig. 11. Heat-map of 57 tentative biomarker candidate peaks from 215 patients (control = 85, disease = 128) saliva samples. Columns represent individual patient and rows specific metabolite. Figure obtained/modified from reference [82]. Reprinted with permission from M. Sugimoto, D. Wong, A. Hirayama, T. Soga, M. Tomita, Capillary electrophoresis mass spectrometry-based saliva metabolomics identified oral, breast and pancreatic cancer-specific profiles, Metabolomics 6 (2010) 78-95. Copyright 2010 Springer US.

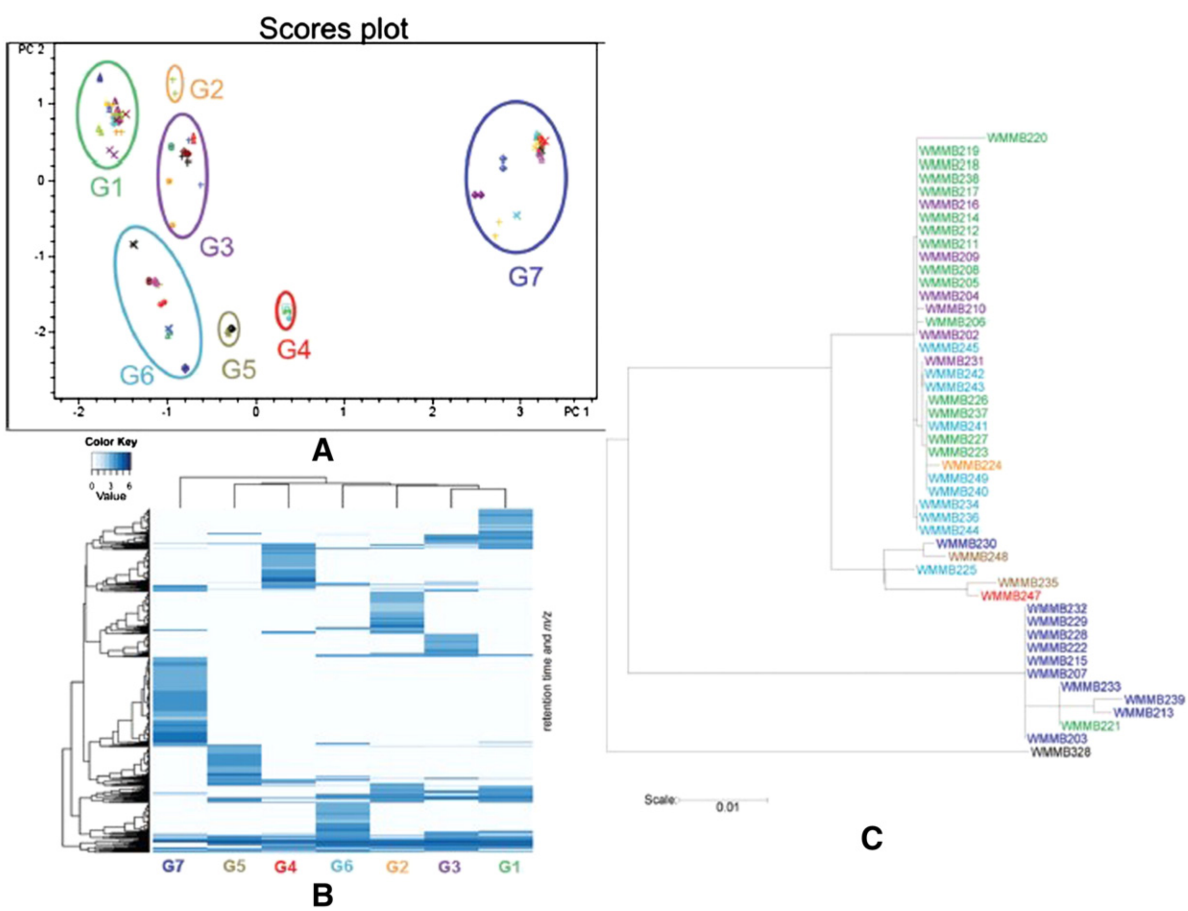

Fig. 12. Isolate de-replication by metabolomic analysis. (A) LC-MS-PCA score plot of 47 strains. (B) Heat-map illustrating metabolic profiles of clusters from PCA, groups 1–7. (C) Phylogenetic tree of strains, colors correspond to metabolomic profiles. *Streptomyces* sp. (WMMB-328) used as an outgroup. Figure obtained/modified from reference [94]. Reprinted with permission from Y. Hou, D.R. Braun, C.R. Michel, J.L. Klassen, N. Adnani, T.P. Wyche, T.S. Bugni, Microbial strain prioritization using metabolomics tools for the discovery of natural products, Anal. Chem. 84 (2012) 4277-4283. Copyright 2012 American Chemical Society.

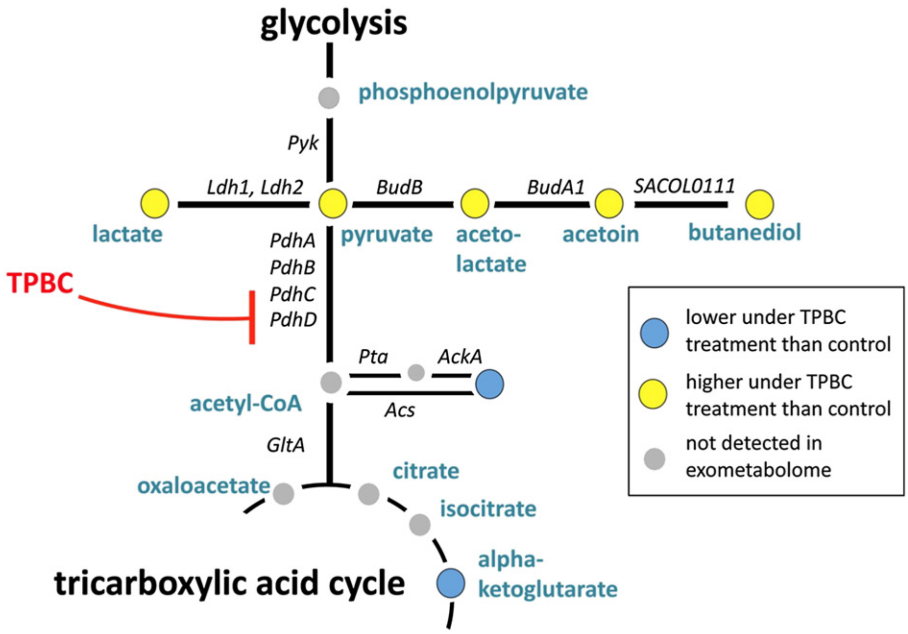

Fig. 13. Schematic representation of the pyruvate-metabolic pathways in *S. aureus*. Figure obtained/modified from reference [95]. Reprinted with permission from T. Birkenstock, M. Liebeke, V. Winstel, B. Krismer, C. Gekeler, M.J. Niemiec, H. Bisswanger, M. Lalk, A. Peschel, Exometabolome analysis identifies pyruvate dehydrogenase as a target for the antibiotic riphenylbismuthdichloride in multiresistant bacterial pathogens, J. Biol. Chem. 287 (2012) 2887-2895. Copyright 2012 American Society for Biochemistry and Molecular Biology.

mapping allows probing of potential downstream interactions by perturbation of upstream nodes (Fig. 15) [40].

An example of the potential that the combinatorial genomic–metabolomic approach has in natural products drug discovery is illustrated by the work of Xu, et al. [97] on the didemnins. The didemnins are a family of cyclic depsipeptide compounds originally isolated from the *Trididemnum* genus of Caribbean tunicates, with preliminary assays showing potent antitumor activity [98]. The most potent was didemnin B, which represents the first natural product isolated from a marine source to enter cancer clinical trials [99]. However, during Phase II clinical trials cardiac and pulmonary toxicities precluded further evaluation [100]. A subsequent didemnin B analogue possessing a modified side-chain (see Fig. 16), dehydrodidemnin B (aplidine/plitidepsin), isolated from a Mediterranean tunicate in the *Aplidium* genus was found to possess greater antitumor activity and improved safety profile [101–103].

Using the marine bacteria *Tistrella mobilis*, discovered to biosynthesize didemnin B, Xu et al. [97] utilized 454 pyrosequencing to sequence the complete genome [99]. They developed a biosynthesis model for didemnins based on a non-ribosomal peptide synthetase-polyketide synthase (NRPS-PKS) modular pathway, which was subsequently used to query the genome for the corollary gene cluster [97]. The mapped genome yielded one gene cluster consistent with their NRPS-PKS pathway model [104]. However, the product of the corresponding pathway was not didemnin B but rather the precursors didemnin X and didemnin Y (see Fig. 16).

Examining the metabolomic profile of *T. mobilis* cultures for didemnin analogues by MALDI-TOF-MS imaging it was observed that didemnin X and didemnin Y were excreted from the bacterial colonies. Moreover, during the time-course disappearance of these metabolites from the culture media they observed a concurrent rise in didemnin B. It was also noted that by coupling the identified gene cluster with the

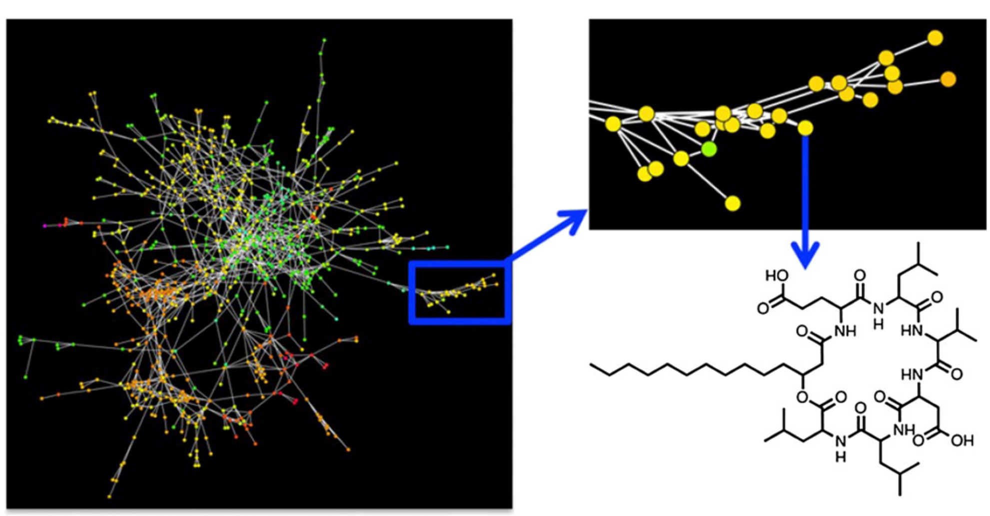

Fig. 14. Representative network map where nodes are based on spectral similarity. Figure obtained/modified from reference [96]. Reprinted with permission from M.F. Traxler, R. Kolter, A massively spectacular view of the chemical lives of microbes, Proc. Natl. Acad. Sci. U.S.A. 109 (2012) 10128-10129. Copyright 2012 National Academy of Sciences of the United States of America.

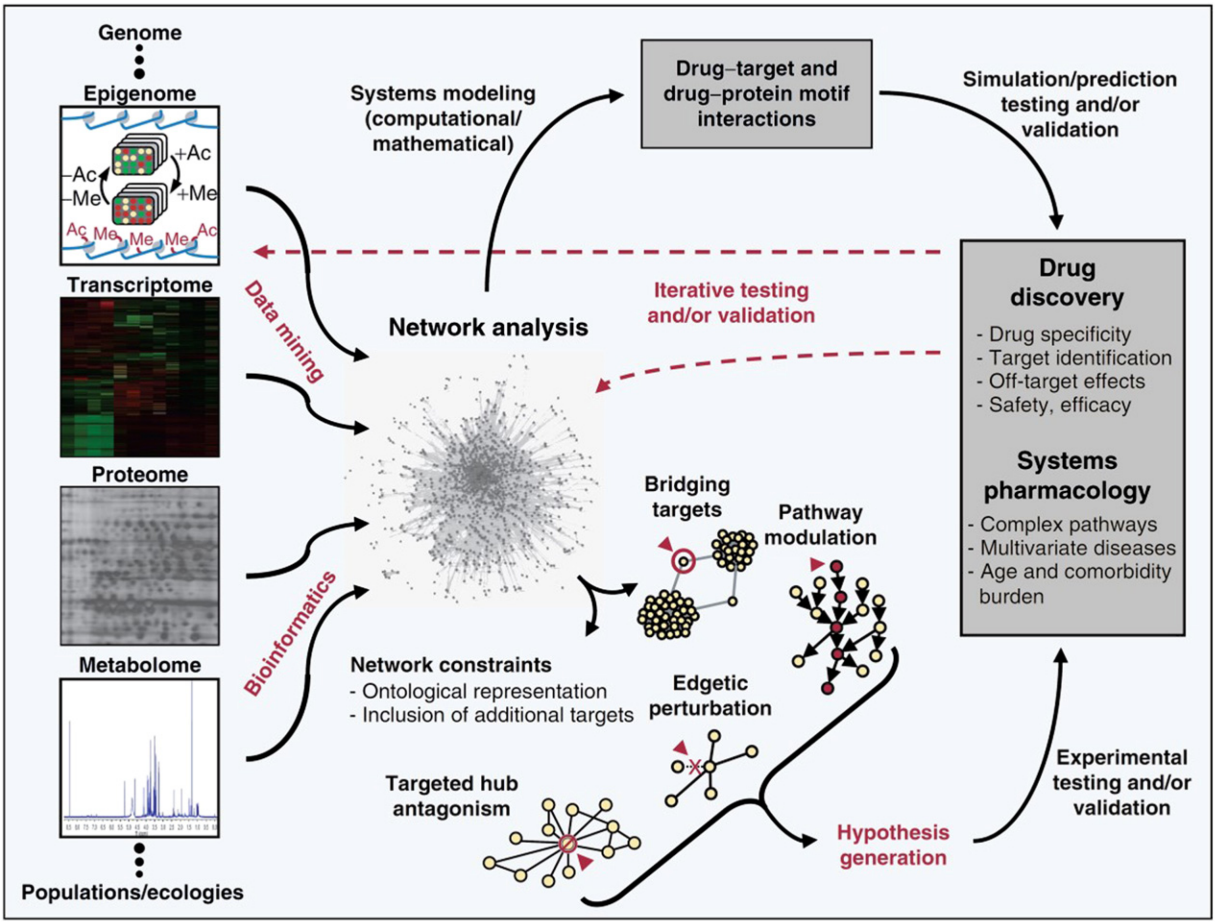

Fig. 15. Role of network analysis in systems approach to drug discovery. Figure obtained/modified from reference [40]. Reprinted with permission from D. Arrell, A. Terzic, Network systems biology for drug discovery, Clin. Pharmacol. Ther. 88 (2010) 120-125. Copyright 2010 Nature Publishing Group.

observed didemnin metabolic conversion that a single gene inactivation in the NRPS-PKS pathway would alter the downstream biosynthetic pathway from didemnin B to dehydrodidemnin B production. They suggested this could provide a novel alternative to the currently employed multi-step synthetic routes to dehydrodidemnin B, employed due to the scarcity of the natural supply [97].

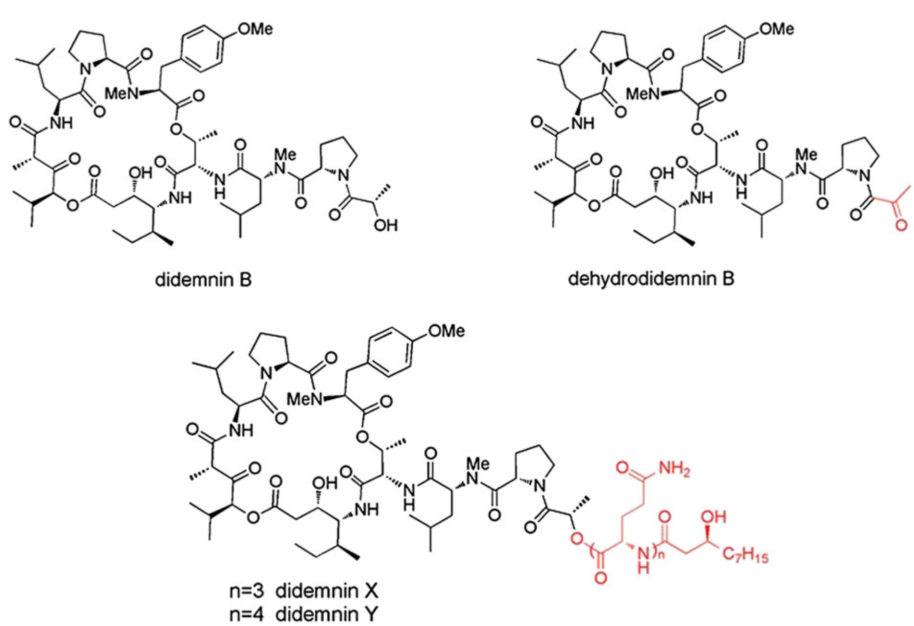

Fig. 16. Structures of didemnins with side-chain differences from didemnin B noted in red. Figure obtained/modified from reference [97]. Reprinted with permission from Y. Xu, R.D. Kersten, S.-J. Nam, L. Lu, A.M. Al-Suwailem, H. Zheng, W. Fenical, P.C. Dorrestein, B.S. Moore, P.-Y. Qian, Bacterial biosynthesis and maturation of the didemnin anti-cancer agents, J. Am. Chem. Soc. 134 (2012) 8625-8632. Copyright 2012 American Chemical Society.

In summary, metabolomics represents a developing research field involving quantitative and qualitative metabolite assessment with diverse applications. In particular, it is capable of system level metabolic analysis with direct applications towards natural products discovery and detection of disease-related biomarkers with utility in early stage diagnosis. Metabolomics has growing utility in resolving phylogenetic associations involving horizontal gene transfer and distinguishing subgroups of genera possessing high genetic homology. Additionally, metabolomics is able to provide a snapshot of the functional genetic status of an organism facilitating connections between changes in metabolites, originating biochemical pathways and gene functions. Continued advancements in metabolomic technologies will undoubtedly yield additional improvements in the utility of metabolomics tools as applied toward both scientific and medical research.

References

[1] J.E. Manson, G.A. Colditz, M.J. Stampfer, W.C. Willett, B. Rosner, R.R. Monson, F.E. Speizer, C.H. Hennekens, A prospective study of obesity and risk of coronary heart disease in women, N. Engl. J. Med. 322 (1990) 882–889.

[2] D. Hinton, C. Bacon, The distribution and ultrastructure of the endophyte of toxic tall fescue, Can. J. Bot. 63 (1985) 36–42.

[3] M. Tyree, Y. Cheung, M. MacGregor, A. Talbot, The characteristics of seasonal and ontogenetic changes in the tissue-water relations of *Acer*, *Populus*, *Tsuga*, and *Picea*, Can. J. Bot. 56 (1978) 635–647.

[4] E. Dudley, M. Yousef, Y. Wang, W. Griffiths, Targeted metabolomics and mass spectrometry, Adv. Protein Chem. Struct. Biol. 80 (2010) 45–83.

[5] G.J. Patti, O. Yanes, G. Siuzdak, Innovation: metabolomics: the apogee of the omics trilogy, Nat. Rev. Mol. Cell Biol. 13 (2012) 263–269.

[6] D.B. Kell, Metabolomics and systems biology: making sense of the soup, Curr. Opin. Microbiol. 7 (2004) 296–307.

[7] G.N. Gowda, S. Zhang, H. Gu, V. Asiago, N. Shanaiah, D. Raftery, Metabolomics-based methods for early disease diagnostics, Expert. Rev. Mol. Diagn. 8 (2008) 617–633.

[8] E. Trujillo, C. Davis, J. Milner, Nutrigenomics, proteomics, metabolomics, and the practice of dietetics, J. Am. Diet. Assoc. 106 (2006) 403–413.

[9] K. Urano, K. Maruyama, Y. Ogata, Y. Morishita, M. Takeda, N. Sakurai, H. Suzuki, K. Saito, D. Shibata, M. Kobayashi, Characterization of the ABA-regulated global responses to dehydration in *Arabidopsis* by metabolomics, Plant J. 57 (2009) 1065–1078.

[10] A.R. Fernie, R.N. Trethewey, A.J. Krotzky, L. Willmitzer, Metabolite profiling: from diagnostics to systems biology, Nat. Rev. Mol. Cell Biol. 5 (2004) 763–769.

[11] P. Devaux, M. Horning, E. Horning, Benzyloxime derivatives of steroids. A new metabolic profile procedure for human urinary steroids human urinary steroids, Anal. Lett. 4 (1971) 151–160.

[12] W. Cunnick, J. Cromie, R. Cortell, B. Wright, E. Beach, F. Seltzer, S. Miller, Value of biochemical profiling in a periodic health examination program: analysis of 1,000 cases, Bull. N. Y. Acad. Med. 48 (1972) 5–22.

[13] K.J. Judy, D.A. Schooley, L.L. Dunham, M. Hall, B.J. Bergot, J.B. Siddall, Isolation, structure, and absolute configuration of a new natural insect juvenile hormone from *Manduca sexta*, Proc. Natl. Acad. Sci. U. S. A. 70 (1973) 1509–1513.

[14] J.J. Sims, M.S. Donnell, J.V. Leary, G.H. Lacy, Antimicrobial agents from marine algae, Antimicrob. Agents Chemother. 7 (1975) 320–321.

[15] D.G. Kingston, M.M. Rao, W.V. Zucker, Plant anticancer agents. IX. Constituents of *Hyptis tomentosa*, J. Nat. Prod. 42 (1979) 496–499.

[16] J. Vrbanac, W. Braselton, J. Holland, C. Sweeley, Automated qualitative and quantitative metabolic profiling analysis of urinary steroids by a gas chromatography-mass spectrometry-data system, J. Chromatogr. A 239 (1982) 265–276.

[17] J. Nicholson, M.P. O'Flynn, P. Sadler, A. Macleod, S. Juul, P. Sonksen, Proton-nuclear-magnetic-resonance studies of serum, plasma and urine from fasting normal and diabetic subjects, Biochem. J. 217 (1984) 365–375.

[18] J.R. Bales, D.P. Higham, I. Howe, J.K. Nicholson, P.J. Sadler, Use of high-resolution proton nuclear magnetic resonance spectroscopy for rapid multi-component analysis of urine, Clin. Chem. 30 (1984) 426–432.

[19] J. Bales, J. Bell, J. Nicholson, P. Sadler, J. Timbrell, R. Hughes, P. Bennett, R. Williams, Metabolic profiling of body fluids by proton NMR: self-poisoning episodes with paracetamol (acetaminophen), Magn. Reson. Med. 6 (1988) 300–306.

[20] H. Sauter, M. Lauer, H. Fritsch, Metabolic profiling of plants: a new diagnostic technique, ACS Symp. Ser. Am. Chem. Soc. 443 (1991) 288–299.

[21] C.A. Smith, G. O'Maille, E.J. Want, C. Qin, S.A. Trauger, T.R. Brandon, D.E. Custodio, R. Abagyan, G. Siuzdak, METLIN: a metabolite mass spectral database, Ther. Drug Monit. 27 (2005) 747–751.

[22] L.W. Sumner, A.L. Duran, D.V. Huhman, J.T. Smith, Chapter three metabolomics: a developing and integral component in functional genomic studies of *Medicago truncatula*, Recent Adv. Phytochem. 36 (2002) 31–61.

[23] W.B. Dunn, D.I. Ellis, Metabolomics: current analytical platforms and methodologies, TrAC Trends Anal. Chem. 24 (2005) 285–294.

[24] A. Huenerbein, M.A. Sípoli Marques, A. dos Santos Pereira, F. Radler de Aquino Neto, Improvement in steroid screening for doping control with special emphasis on stanozolol, J. Chromatogr. A 985 (2003) 375–386.

[25] Y. Noguchi, R. Sakai, T. Kimura, Metabolomics and its potential for assessment of adequacy and safety of amino acid intake, J. Nutr. 133 (2003) 2097S–2100S.

[26] J. Dallüge, J. Beens, U.A.T. Brinkman, Comprehensive two-dimensional gas chromatography: a powerful and versatile analytical tool, J. Chromatogr. A 1000 (2003) 69–108.

[27] M. Adahchour, J. Beens, R. Vreuls, U. Brinkman, Recent developments in comprehensive two-dimensional gas chromatography (GC $\times$ GC): IV. Further applications, conclusions and perspectives, TrAC Trends Anal. Chem. 25 (2006) 821–840.

[28] Y. Hu, Y. Qi, H. Liu, G. Fan, Y. Chai, Effects of celastrol on human cervical cancer cells as revealed by ion-trap gas chromatography–mass spectrometry based metabolic profiling, Biochim. Biophys. Acta Gen. Subj. 1830 (2012) 2779–2789.

[29] Y. Shen, R. Zhang, R.J. Moore, J. Kim, T.O. Metz, K.K. Hixson, R. Zhao, E.A. Livesay, H. R. Udseth, R.D. Smith, Automated 20 kpsi RPLC-MS and MS/MS with chromatographic peak capacities of 1000–1500 and capabilities in proteomics and metabolomics, Anal. Chem. 77 (2005) 3090–3100.

[30] W. Lu, B.D. Bennett, J.D. Rabinowitz, Analytical strategies for LC–MS-based targeted metabolomics, J. Chromatogr. B 871 (2008) 236–242.

[31] B. Zhou, J.F. Xiao, L. Tuli, H.W. Ressom, LC-MS-based metabolomics, Mol. Biosyst. 8 (2012) 470–481.

[32] M. Scigelova, M. Hornshaw, A. Giannakopulos, A. Makarov, Fourier transform mass spectrometry, Mol. Cell. Proteomics 10 (2011).

[33] Q. Hu, R.J. Noll, H. Li, A. Makarov, M. Hardman, R. Graham Cooks, The Orbitrap: a new mass spectrometer, J. Mass Spectrom. 40 (2005) 430–443.

[34] R.E. March, Quadrupole ion trap mass spectrometry: a view at the turn of the century, Int. J. Mass Spectrom. 200 (2000) 285–312.

[35] D.J. Douglas, A.J. Frank, D. Mao, Linear ion traps in mass spectrometry, Mass Spectrom. Rev. 24 (2005) 1–29.

[36] S. Lacorte, A.R. Fernandez-Alba, Time of flight mass spectrometry applied to the liquid chromatographic analysis of pesticides in water and food, Mass Spectrom. Rev. 25 (2006) 866–880.

[37] J.H. Wang, J. Byun, S. Pennathur, Analytical approaches to metabolomics and applications to systems biology, Semin. Nephrol. 30 (2010) 500–511.

[38] M. Brown, D.C. Wedge, R. Goodacre, D.B. Kell, P.N. Baker, L.C. Kenny, M.A. Mamas, L. Neyses, W.B. Dunn, Automated workflows for accurate mass-based putative metabolite identification in LC/MS-derived metabolomic datasets, Bioinformatics 27 (2011) 1108–1112.

[39] R. Kaddurah-Daouk, B.S. Kristal, R.M. Weinshilboum, Metabolomics: a global biochemical approach to drug response and disease, Annu. Rev. Pharmacol. Toxicol. 48 (2008) 653–683.

[40] D. Arrell, A. Terzic, Network systems biology for drug discovery, Clin. Pharmacol. Ther. 88 (2010) 120–125.

[41] A. O'Sullivan, R.E. Willoughby, D. Mishchuk, B. Alcarraz, C. Cabezas-Sanchez, R.E. Condori, D. David, R. Encarnacion, N. Fatteh, J. Fernandez, R. Franka, S. Hedderwick, C. McCaughey, J. Ondrush, A. Paez-Martinez, C. Rupprecht, A. Velasco-Villa, C.M. Slupsky, Metabolomics of cerebrospinal fluid from humans treated for rabies, J. Proteome Res. 12 (2012) 481–490.

[42] H.R. Cho, H. Wen, Y.J. Ryu, Y.J. An, H.C. Kim, W.K. Moon, M.H. Han, S. Park, S.H. Choi, An NMR metabolomics approach for the diagnosis of leptomeningeal carcinomatosis, Cancer Res. 72 (2012) 5179–5187.

[43] Q. Van, T. Veenstra, H. Issaq, Metabolic profiling for the detection of bladder cancer, Curr. Urol. Rep. 12 (2011) 34–40.

[44] J. Zhang, S. Wei, L. Liu, G.A. Nagana Gowda, P. Bonney, J. Stewart, D.W. Knapp, D. Raftery, NMR-based metabolomics study of canine bladder cancer, Biochim. Biophys. Acta Mol. Basis Dis. 1822 (2012) 1807–1814.

[45] D.W. Knapp, N.W. Glickman, D.B. DeNicola, P.L. Bonney, T.L. Lin, L.T. Glickman, Naturally-occurring canine transitional cell carcinoma of the urinary bladder, a relevant model of human invasive bladder cancer, Urol. Oncol. 5 (2000) 47–59.

[46] A. Jemal, R. Siegel, E. Ward, Y. Hao, J. Xu, T. Murray, M.J. Thun, Cancer statistics, 2008, CA Cancer J. Clin. 58 (2008) 71–96.

[47] A. Imaizumi, N. Ono, R. Sakai, T. Ando, N. Okamoto, F. Imamura, M. Higashiyama, Lung cancer evaluating apparatus, method, system, and program recording medium therefor, US 2010/0017144A12010.

[48] R.N. Fedorak, E. Wang, Methods for the assessment of colorectal cancer and colorectal polyps by measurement of metabolites in urine, US 2013/0065320A12013.

[49] M.B. Potter, Strategies and resources to address colorectal cancer screening rates and disparities in the United States and globally, Annu. Rev. Public Health 34 (2013) 413–429.

[50] M.D. Raftery, V.M. Asiago, G.A.N. Gowda, L. Alvarado, Early detection of recurrent breast cancer using metabolite profiling, US 2013/00230562013.

[51] K. Weinberger, H. Deinger, E.I. Igwe, D. Enot, G. Dallmann, H. Klocker, Diagnosing prostate cancer relapse, US 2012/0326025A12012.

[52] U. Lundin, K. Weinberger, New biomarkers for assessing kidney disease, US 2012/0129265A12012.

[53] G.G. Cezar, Molecule biomarkers of autism, US 2012/0190055A12012.

[54] Q. Pan, Q. Wang, F. Yuan, S. Xing, J. Zhao, Y.H. Choi, R. Verpoorte, Y. Tian, G. Wang, K. Tang, Overexpression of ORCA3 and G10H in *Catharanthus roseus* plants regulated alkaloid biosynthesis and metabolism revealed by NMR-metabolomics, PLoS One 7 (2012) e43038.

[55] A. Mari, D. Lyon, L. Fragner, P. Montoro, S. Piacente, S. Wienkoop, V. Egelhofer, W. Weckwerth, Phytochemical composition of *Potentilla anserina* L. analyzed by an integrative GC–MS and LC-MS metabolomics platform, Metabolomics (2012) 1–9.

[56] K. Böröczky, H. Laatsch, I. Wagner-Döbler, K. Stritzke, S. Schulz, Cluster analysis as selection and dereplication tool for the identification of new natural compounds from large sample sets, Chem. Biodivers. 3 (2006) 622–634.

[57] G.K. Pierens, M.E. Palframan, C.J. Tranter, A.R. Carroll, R.J. Quinn, A robust clustering approach for NMR spectra of natural product extracts, Magn. Reson. Chem. 43 (2005) 359–365.

[58] D. Krug, G. Zurek, O. Revermann, M. Vos, G.J. Velicer, R. Müller, Discovering the hidden secondary metabolome of *Myxococcus xanthus*: a study of intraspecific diversity, Appl. Environ. Microbiol. 74 (2008) 3058–3068.

[59] J. Yuk, K.L. McIntyre, C. Fischer, J. Hicks, K.L. Colson, E. Lui, D. Brown, J.T. Arnason, Distinguishing Ontario ginseng landraces and ginseng species using NMR-based metabolomics, Anal. Bioanal. Chem. (2012) 1–11.

[60] L. Boto, Horizontal gene transfer in evolution: facts and challenges, Proc. R. Soc. B Biol. Sci. 277 (2010) 819–827.

[61] J.C. Clemente, K. Satou, G. Valiente, Phylogenetic reconstruction from non-genomic data, Bioinformatics 23 (2007) e110–e115.

[62] C. Pál, B. Papp, M.J. Lercher, Adaptive evolution of bacterial metabolic networks by horizontal gene transfer, Nat. Genet. 37 (2005) 1372–1375.

[63] J.C. Slot, A. Rokas, Horizontal transfer of a large and highly toxic secondary metabolic gene cluster between fungi, Curr. Biol. 21 (2011) 134–139.

[64] R. Dieckmann, I. Graeber, I. Kaesler, U. Szewzyk, H. Von Döhren, Rapid screening and dereplication of bacterial isolates from marine sponges of the Sula Ridge by Intact-Cell-MALDI-TOF mass spectrometry (ICM-MS), Appl. Microbiol. Biotechnol. 67 (2005) 539–548.

[65] G. Gauthier, M. Gauthier, R. Christen, Phylogenetic analysis of the genera *Alteromonas*, *Shewanella*, and *Moritella* using genes coding for small-subunit rRNA sequences and division of the genus *Alteromonas* into two genera, *Alteromonas* (emended) and *Pseudoalteromonas* gen. nov., and proposal of twelve new species combinations, Int. J. Syst. Bacteriol. 45 (1995) 755–761.

[66] L.-X. Duan, T.-L. Chen, M. Li, M. Chen, Y.-Q. Zhou, G.-H. Cui, A.-H. Zhao, W. Jia, L.-Q. Huang, X. Qi, Use of the metabolomics approach to characterize Chinese medicinal material Huangqi, Mol. Plant 5 (2012) 376–386.

[67] O. Fiehn, Metabolomics—the link between genotypes and phenotypes, Plant Mol. Biol. 48 (2002) 155–171.

[68] C.W. Chang, P.C. Lyu, M. Arita, Reconstructing phylogeny from metabolic substrate–product relationships, BMC Bioinforma. 12 (2011) S27.

[69] O. Fiehn, J. Kopka, P. Dörmann, T. Altmann, R.N. Trethewey, L. Willmitzer, Metabolite profiling for plant functional genomics, Nat. Biotechnol. 18 (2000) 1157–1161.

[70] R.A. Dixon, L.W. Sumner, Legume natural products: understanding and manipulating complex pathways for human and animal health, Plant Physiol. 131 (2003) 878–885.

[71] W.M. Claudino, P.H. Goncalves, A. di Leo, P.A. Philip, F.H. Sarkar, Metabolomics in cancer: a bench-to-bedside intersection, Crit. Rev. Oncol. Hematol. 84 (2012) 1–7.

[72] R. Madsen, T. Lundstedt, J. Trygg, Chemometrics in metabolomics—a review in human disease diagnosis, Anal. Chim. Acta. 659 (2010) 23–33.

[73] M. Mamas, W.B. Dunn, L. Neyses, R. Goodacre, The role of metabolites and metabolomics in clinically applicable biomarkers of disease, Arch. Toxicol. 85 (2011) 5–17.

[74] C.C. Tang, K.L. Poston, V. Dhawan, D. Eidelberg, Abnormalities in metabolic network activity precede the onset of motor symptoms in Parkinson's disease, J. Neurosci. 30 (2010) 1049–1056.

[75] J.B. German, S.M. Watkins, L.-B. Fay, Metabolomics in practice: emerging knowledge to guide future dietetic advice toward individualized health, J. Am. Diet. Assoc. 105 (2005) 1425–1432.

[76] C. Wang, H. Kong, Y. Guan, J. Yang, J. Gu, S. Yang, G. Xu, Plasma phospholipid metabolic profiling and biomarkers of type 2 diabetes mellitus based on high-performance liquid chromatography/electrospray mass spectrometry and multivariate statistical analysis, Anal. Chem. 77 (2005) 4108–4116.

[77] K. Yuan, H. Kong, Y. Guan, J. Yang, G. Xu, A GC-based metabonomics investigation of type 2 diabetes by organic acids metabolic profile, J. Chromatogr. B 850 (2007) 236–240.

[78] J.T. Brindle, H. Antti, E. Holmes, G. Tranter, J.K. Nicholson, H.W. Bethell, S. Clarke, P.M. Schofield, E. McKilligin, D.E. Mosedale, Rapid and noninvasive diagnosis of the presence and severity of coronary heart disease using ^{1}H-NMR-based metabonomics, Nat. Med. 8 (2002) 1439–1445.

[79] X. Han, D. M Holtzman, D. W McKeel, J. Kelley, J.C. Morris, Substantial sulfatide deficiency and ceramide elevation in very early Alzheimer's disease: potential role in disease pathogenesis, J. Neurochem. 82 (2002) 809–818.

[80] F. van der Kloet, F. Tempels, N. Ismail, R. van der Heijden, P. Kasper, M. Rojas-Cherto, R. van Doorn, G. Spijksma, M. Koek, J. van der Greef, Discovery of early-stage biomarkers for diabetic kidney disease using MS-based metabolomics (FinnDiane study), Metabolomics 8 (2012) 109–119.

[81] L. Lin, Z. Huang, Y. Gao, X. Yan, J. Xing, W. Hang, LC-MS based serum metabonomic analysis for renal cell carcinoma diagnosis, staging, and biomarker discovery, J. Proteome Res. 10 (2011) 1396–1405.

[82] M. Sugimoto, D. Wong, A. Hirayama, T. Soga, M. Tomita, Capillary electrophoresis mass spectrometry-based saliva metabolomics identified oral, breast and pancreatic cancer-specific profiles, Metabolomics 6 (2010) 78–95.

[83] N. Spielmann, D.T. Wong, Saliva: diagnostics and therapeutic perspectives, Oral Dis. 17 (2011) 345–354.

[84] M.P. Quinones, R. Kaddurah-Daouk, Metabolomics tools for identifying biomarkers for neuropsychiatric diseases, Neurobiol. Dis. 35 (2009) 165–176.

[85] S. Moco, R.J. Bino, O. Vorst, H.A. Verhoeven, J. de Groot, T.A. van Beek, J. Vervoort, C. R. De Vos, A liquid chromatography-mass spectrometry-based metabolome database for tomato, Plant Physiol. 141 (2006) 1205–1218.

[86] D.A. Dias, S. Urban, U. Roessner, A historical overview of natural products in drug discovery, Metabolites 2 (2012) 303–336.

[87] N.D. Yuliana, A. Khatib, Y.H. Choi, R. Verpoorte, Metabolomics for bioactivity assessment of natural products, Phytother. Res. 25 (2011) 157–169.

[88] T.A. Gulder, B.S. Moore, Chasing the treasures of the sea—bacterial marine natural products, Curr. Opin. Microbiol. 12 (2009) 252–260.

[89] E. Esquenazi, Y.-L. Yang, J. Watrous, W.H. Gerwick, P.C. Dorrestein, Imaging mass spectrometry of natural products, Nat. Prod. Rep. 26 (2009) 1521–1534.

[90] D. Staerk, J.R. Kesting, M. Sairafianpour, M. Witt, J. Asili, S.A. Emami, J.W. Jaroszewski, Accelerated dereplication of crude extracts using HPLC–PDA–MS–SPE–NMR: quinolinone alkaloids of *Haplophyllum acutifolium*, Phytochemistry 70 (2009) 1055–1061.

[91] D.S. Wishart, Quantitative metabolomics using NMR, TrAC Trends Anal. Chem. 27 (2008) 228–237.

[92] J.F. Imhoff, A. Labes, J. Wiese, Bio-mining the microbial treasures of the ocean: new natural products, Biotechnol. Adv. 29 (2011) 468–482.

[93] L. Torras-Claveria, S. Berkov, O. Jáuregui, J. Caujapé, F. Viladomat, C. Codina, J. Bastida, Metabolic profiling of bioactive *Pancratium canariense* extracts by GC–MS, Phytochem. Anal. 21 (2010) 80–88.

[94] Y. Hou, D.R. Braun, C.R. Michel, J.L. Klassen, N. Adnani, T.P. Wyche, T.S. Bugni, Microbial strain prioritization using metabolomics tools for the discovery of natural products, Anal. Chem. 84 (2012) 4277–4283.

[95] T. Birkenstock, M. Liebeke, V. Winstel, B. Krismer, C. Gekeler, M.J. Niemiec, H. Bisswanger, M. Lalk, A. Peschel, Exometabolome analysis identifies pyruvate dehydrogenase as a target for the antibiotic triphenylbismuthdichloride in multiresistant bacterial pathogens, J. Biol. Chem. 287 (2012) 2887–2895.

[96] M.F. Traxler, R. Kolter, A massively spectacular view of the chemical lives of microbes, Proc. Natl. Acad. Sci. U. S. A. 109 (2012) 10128–10129.

[97] Y. Xu, R.D. Kersten, S.-J. Nam, L. Lu, A.M. Al-Suwailem, H. Zheng, W. Fenical, P. C. Dorrestein, B.S. Moore, P.-Y. Qian, Bacterial biosynthesis and maturation of the didemnin anti-cancer agents, J. Am. Chem. Soc. 134 (2012) 8625–8632.

[98] K.L. Rinehart Jr., J.B. Gloer, R. Hughes Jr., H.E. Renis, J.P. McGovren, E.B. Swynenberg, D.A. Stringfellow, S.L. Kuentzel, L.H. Li, Didemnins: antiviral and antitumor depsipeptides from a Caribbean tunicate, Science 212 (1981) 933–935.

[99] M. Tsukimoto, M. Nagaoka, Y. Shishido, J. Fujimoto, F. Nishisaka, S. Matsumoto, E. Harunari, C. Imada, T. Matsuzaki, Bacterial production of the tunicate-derived antitumor cyclic depsipeptide didemnin B, J. Nat. Prod. 74 (2011) 2329–2331.

[100] S.K. Williamson, M.K. Wolf, M.A. Eisenberger, M. O'Rourke, W. Brannon, E.D. Crawford, Phase II evaluation of didemnin B in hormonally refractory metastatic prostate cancer, Investig. New Drugs 13 (1995) 167–170.

[101] X. Ding, M.D. Vera, B. Liang, Y. Zhao, M.S. Leonard, M.M. Joullié, Structure–activity relationships of side-chain modified didemnins, Bioorg. Med. Chem. Lett. 11 (2001) 231–234.

[102] V. Ribrag, D. Caballero, C. Fermé, E. Zucca, R. Arranz, J. Briones, C. Gisselbrecht, G. Salles, A.M. Gianni, H. Gomez, Multicenter phase II study of plitidepsin in patients with relapsed/refractory non-Hodgkin's lymphoma, Haematologica 98 (2013) 357–363.

[103] E. Zubia, M.J. Ortega, J. Salva, Natural products chemistry in marine ascidians of the genus *Aplidium*, Mini-Rev. Org. Chem. 2 (2005) 389–399.

[104] P. Qian, Y.S. Xu, P. Lai, Didemnin biosynthetic gene cluster in tistrella mobilis, WO 2013041969 A22013.