Soil and Biota of Serpentine: A World View
2009 *Northeastern Naturalist* 16(Special Issue 5):285–296

"Serpentinomics"—An Emerging New Field of Study

Jessica W. Wright[1] and Eric von Wettberg[2]

Abstract - "Serpentinomics" is an emerging field of study which has the potential to greatly advance our understanding of serpentine ecology. Several newly developing –omic fields, often using high-throughput tools developed for molecular biology, will advance the field of serpentine ecology, or, "serpentinomics." Using tools from the fields of ionomics, metabolomics, proteomics, transcriptomics and genomics, researchers will be able to address new (and old) ecological questions in powerful and creative ways. In particular, "serpentinomics" has the potential to uncover the mechanistic and genetic basis of the complexities of tolerance of and adaptation to serpentine soils, including the biochemistry of hyperaccumulation. Here we outline each of these –omic fields and describe possible applications to the field of serpentine ecology.

Introduction

In a time of an ever-growing number of –omic fields of study, we offer one more term for consideration: "Serpentinomics." Loosely, we define this field of study as the application of any of the –omic techniques and technologies to further the study of organisms occurring on serpentine soils. In general, a field described by the suffix –omic is defined as the study encompassing information from the entire field, often accomplished by applying high-throughput tools developed for molecular biology. For example, genomics is the study of the full sequence of DNA, or the genome, of an organism. There are many –omic fields, each of which can advance the field of serpentine ecology, or, "serpentinomics." While "serpentinomics" is an ever-growing field, we suggest that serpentinophiles should consider some of the new and developing tools available in many of these –omic study areas. Here we discuss the primary –omic fields and highlight some of the questions relevant to the study of serpentine ecology they could be used to address.

Rather than present a comprehensive review of each of the –omic fields, our hope is to inspire some new thinking among serpentine ecologists, and provide the necessary references to begin to pursue those ideas. High-throughput technologies used for –omics have the potential to illuminate the mechanistic and genetic basis of complex ecological traits, such as tolerance of serpentine soils, in ways previously unimaginable. Because of the cost and computational intensity required to implement many –omics technologies, the use of –omics technology in research will differ from that

[1]Institute of Forest Genetics, USDA Forest Service, Pacific Southwest Research Station, 1731 Research Park Drive, Davis, CA 95618. [2]Center for Population Biology, University of California, 1 Shields Avenue, Davis, CA 95616. [*]Corresponding author - jessicawwright@fs.fed.us.

traditionally used in much of ecology and evolution. The actual instruments used in many –omics fields are well beyond the means of single laboratory groups, if not entire biological science departments; however, many of them are available through research centers such as genomics core facilities. These are often run on a very affordable per sample basis. As a result, many of the –omics fields we review here are (or soon will be) accessible to ecologists, who can simply submit samples, often with minimal preparation, to a core facility for analysis. The resulting data can come in files with millions of lines of data, resulting in the need for substantial data analysis. This analysis frequently requires computing power well beyond that of a standard desktop computer. However, many of the analytical issues will seem simple to most ecologists, as ecological experiments tend to have much more complex experimental designs than most –omic ones. Moreover, many core facilities offer computing resources for initial processing of –omic data, and free open-source software resources exist within the R-based bioconductor project (www.Bioconductor.org and www.R-project.org) to handle much of the final analysis. Our hope is to encourage greater numbers of serpentine ecologists to explore emerging –omics technologies, through technological outsourcing and creative collaborations beyond the traditional bounds of ecology. "Serpentinomics" is an opportunity to revolutionize the questions posed about serpentine ecological systems.

The current list of –omic fields is long and ever-growing. Here we touch upon ones with a focus ranging from the smallest molecules to the larger ones to messenger RNA to the DNA genome itself: ionomics, metabolomics, proteomics, transcriptomics, and genomics (see glossary for definitions of terms used).

Ionomics

Ionomics is the study of the ions found in plant tissue (i.e., N, P, K, Ca, Mg)—the mineral nutrients and trace elements. A key feature of ionomics that separates it from standard nutrient analysis is the concept that it simultaneously encompasses all of the ions found in a plant tissue. Ionomics uses advanced technologies such as mass spectrometry (MS) to evaluate levels of each plant nutrient and metal at the same time. The field has been recently reviewed (Salt et al. 2008). In general, ionomic studies focus on relative rather than absolute levels of plant nutrients. There are important issues regarding the amount and kind of tissue sampled (compared samples need to be the same size). There are also a range of technologies that can be used, each with a range of costs and availabilities; however, in general, an ionomic screen is a comparatively cheap undertaking. Moreover, no previous knowledge of the ionome (or the genome) of the study organism is needed because data are comparative, so the ionome of plants growing on serpentine soils could be compared to that of plants growing on non-serpentine soils without any need for other data.

Ionomics will be familiar to most serpentine ecologists as it has been a part of understanding the serpentine syndrome for some time (e.g., Main 1981; see Brooks 1987 for a review). However, these studies generally have focused on only a small number of ions at a time, rather than the entire ionome. By expanding the number of ions under examination at any given time, a complete picture of the cellular response to serpentine may be reached. There may be unexpected responses to serpentine soils in other ions not previously studied, or dynamics among multiple ions that would be missed by focusing on a single ion. For example, this approach may allow us to understand the differential effects on plant physiology by different outcroppings of serpentine, such as variation in heavy metal levels, Ca and Mg absolute and relative levels, and N and P availability. Furthermore, an ionomic approach can show whether uptake or exclusion of one toxic ion (e.g., Ni or Cr) depends on another more abundant ion which may itself vary within and between serpentine outcroppings (e.g., Fe or K) (see Jaffre and Boyd, this issue). Answers to these sorts of questions are essential to understanding the mechanistic basis of serpentine tolerance, and how it varies between plant species. In addition, new technologies allow for the highly specific isolation of nutrients and metals from specific tissues (e.g., trichomes; Salt et al. 2008). This isolation can make possible very specific analyses of the roots vs shoots vs leaves vs even trichomes on leaves. Moreover, these techniques are often high-throughput, which allows for a large number of samples to be analyzed at the same time. Therefore, a fully-replicated greenhouse experiment comparing the ionome of plants growing on and off of serpentine soils (perhaps under different nutrient regimes) could yield fruitful information on the changes in the ionic composition of plants under different soil conditions.

Metabolomics

Metabolomics is the study of all of the detectable metabolites found in a specific tissue. Thus, while ionomics is the study of ions found in tissues, metabolomics is the study of the small molecules. It is a growing field that uses high-tech detection techniques such as integrated gas chromatography-mass spectrometry (GC-MS), integrated liquid chromatography-MS (LC-MS), and high pressure liquid chromatography (HPLC). GC-MS is the most widely used technique for smaller molecules, often primary metabolites, while LC-MS and HPLC are more widely used for larger molecules, such as many secondary metabolites. However, because of the large number of plant metabolites, it is impossible to sample all of the metabolites in a tissue at a given time. Most studies focus on the ones that can be detected using commonly available techniques. There have been a number of recent reviews of the field of metabolomics (Allwood et al. 2008, Hall 2006, Schauer and Fernie 2006). In addition, a number of metabolomic analysis centers have been established globally, and, in general, technicians at those centers can work with serpentine ecologists to determine the best approaches available for their particular research system. Data analysis can be complex, as each assay

can yield thousands of data points. However, much work has been done to develop analytical techniques for extracting as much information as possible from these large data sets (Steinfath et al. 2008, Zou and Tolstikov 2008).

Metabolomics is a powerful tool that could help drive understanding of the changes in plant physiology in response to serpentine soils. One of its key advantages is that a sequenced genome is not required to make sense of the data. Basically, this technique can elucidate differences in genomes without using any DNA. In a traditional view of genetics, genes code for enzymes, which work in biochemical pathways that produce metabolites. Metabolomics detects those metabolites, so, in theory, it is possible to trace backwards from observed changes in metabolite concentration to changes in gene expression and perhaps even in the genes themselves. Therefore, in study species without a sequenced genome (e.g., most of the serpentine flora), it is possible to conduct a metabolomic analysis to understand the mechanisms behind serpentine tolerance, or behind important traits such as hyperaccumulation.

What kind of questions in serpentine ecology can be addressed using metabolomics? This approach is very useful for comparisons; for example, how do serpentine and non-serpentine plants differ physiologically? It has been used to study a number of different plant stresses (reviewed in Shulaev et al. 2008), and interesting metabolic differences have been found between stress-tolerant and sensitive individuals. Given the stressful environment that serpentine soils create, metabolomics may help elucidate the physiological responses to those stresses, and perhaps even help our understanding of the underlying genetics of serpentine tolerance. An understanding of community physiological response to serpentine soils could be gained if multiple species are sampled simultaneously.

Proteomics

Moving from ions to small molecules, proteomics is the study of all of the proteins found in plant tissues. Again the term "all" is a bit of a misnomer, as isolating all of the proteins in a tissue is technologically difficult if not impossible at present. The methodology involves running the extracted proteins in a 2D gel, which allows for separation of proteins based on their molecular weight and their isoelectric point. These gels produce "spots" which can be cut out and analyzed using a MS (Huang and Xu 2008, Rampitsch and Srinivasan 2006). Because of the close association between proteins and genes, a sequenced genome is very helpful in the identification of the proteins found in the above methods (protein sequences can be converted to RNA sequences and compared using NCBI-BLAST type searches). Without genome sequence data, fewer proteins can be identified and with much more effort—though a great deal of information has been learned from proteomic studies using species without genomic data (Rampitsch and Srinivasan 2006).

Proteins are the intermediaries between genes and metabolites, and as such, a complete understanding of the biochemical processes involved in

serpentine tolerance requires an investigation into the proteins produced by plants tolerant to serpentine. While having genomic information is not required to complete a proteomics analysis, it is helpful for the identification of the proteins found in the analysis, and in establishing the underlying genetics of each protein.

Proteomic studies in serpentine habitats could reveal which proteins are associated with serpentine tolerance. Are there different mechanisms for dealing with serpentine, or do all serpentine species use similar biochemical tools to grow in their unique habitat, i.e., is there a community response to serpentine? Do hyperaccumulators have a unique set of proteins that are involved in moving and sequestering heavy metals? How does the concentration of those proteins change in different soil-metal environments?

Transcriptomics

We now move to the genetic material itself, and consider the products formed by transcription in the cells—primarily mRNA. Because the primary biochemical components of both RNA and DNA are very similar, the tools of transcriptomics and those of genomics are very similar. As with proteomics, a sequenced genome is needed to take advantage of the full range of transcriptomic tools. Many model systems have excellent transcriptomic resources that are not currently available to researchers using non-model organisms, but may be soon. That said, there are several species that grow on serpentine that are closely related to model organisms. For example, *Arabidposis thaliana* Thal, the primary plant model system is closely related to *A. lyrata* (L.) O'Kane & Al-Shehbaz, which occurs on serpentine soils (Turner et al. 2008). Many of the genomic tools available for *A. thaliana* are useful for the study of *A. lyrata* as well. As most serpentine species are not closely related to a model species, here we will focus on tools that are or soon will be available for the study of the ecology of non-model systems: cDNA-AFLPs, mRNA microarrays, and next-generation sequencing.

cDNA-AFLPs

The earliest developed approach for identifying expressed genes of interest for non-model systems was cDNA-AFLP (Ouborg and Vriezen 2007). Currently this approach remains a viable technique for many non-model organisms, although emerging next-generation sequencing technologies should replace both cDNA-AFLPs and microarrays (discussed below). In cDNA-AFLP, a basic AFLP protocol (Vos et al. 1995) is applied to cDNA-double-stranded DNA that is synthesized in the lab from mRNA. Most cDNA-AFLP fragments will show little variation between two samples (mRNAs for basic housekeeping genes that are expressed in all plants will not show soil-specific variation in expression). However, mRNAs that are associated with serpentine tolerance might show different levels of expression in different soil environments, and that will be reflected in different frequencies of cDNA-AFLPs. These interesting cDNA-AFLP fragments can

be sequenced and then studied using complementary tools—for example, by creating a microarray of these sequences, or using qRT-PCR to quantify expression levels in different individuals. Using these techniques, genes responsible for serpentine tolerance and metal hyperaccumulation can be identified and studied across serpentine habitats. Although this technique has worked well and remains one of the easiest techniques to implement in a non-model organism, it has several shortcomings: one will not find cDNAs which lack the restriction enzyme site, nor can the technique distinguish distinct cDNAs with fragments cut to the same size (fragments of the same size will run together and be indistinguishable), one cannot use it alone to quantify expression-level differences (qRT-PCR is required to do this), and it has limited ability to find genes with very low expression level.

mRNA microarrays

The most widely used tool of transcriptomics is currently the mRNA microarray. A microarray consists of bits of cDNA attached or "spotted" onto a piece of glass or silica. In commonly used mRNA microarrays, nucleotide probes of 25 to70 base pairs representing as many genes as possible from the organism of interest are affixed to the microarray. A cDNA sample that is fragmented with a restriction enzyme and labeled with a fluorescent marker can then be hybridized to the array. A laser reads the array at each probe; expression of the gene represented by the probe is in proportion to the intensity read by the laser. As a hypothetical example, cDNA could be synthesized from mRNA collected from two plant samples, one growing in serpentine soils, and one growing in non-serpentine soils. Genes that are expressed differentially between serpentine and non-serpentine plants will show different patterns of hybridization.

Two types of mRNA microarrays are currently in widespread use: the silica, short-oligo, single channel microarrays widely sold by Affymetrix (http://www.affymetrix.com), and glass-based, long-oligo, two-channel colored microarrays. These two types of arrays differ substantially in how they are constructed, the cost of production, how they are processed and analyzed, and the sorts of organisms for which they can be obtained. Ultimately, due to the statistical concerns and the substantial rates of technical errors and artifacts, arrays have to be used as a method to find potential candidate genes, rather than verify them. Verification of expression differences requires more precise techniques such as qRT-PCR. The power to detect expression, or expression differences, of genes with low but essential expression, such as many transcription factors, will always be limited with microarrays.

Two-channel arrays may be the best option in non-model systems. In two-channel arrays, two samples labeled with red or green dye are hybridized together to the array. For each of the approximately 70 base-pair long nucleotide probes on the array, a color score will be given, determining which sample has higher expression of the gene to which the probe corresponds. Two-channel microarrays can be constructed using anonymous cDNAs (Kammenga et al. 2007) collected from the focal organism, without

prior knowledge of genome sequence information. Once the hybridization is complete, any cDNAs on the array that are of interest can be individually sequenced and compared to databases of known sequences. Although this will not get expression information about all genes, it can allow researchers to narrow their interests to likely candidate genes.

Next-generation sequencing

Emerging sequencing technologies, sometimes called next-generation sequencing, or resequencing, such as 454 from Roche (http://www.roche.com), Solexa Genome Analyzer from Illumina (http://www.illumina.com/), and SOLiD ™ from ABI (http://www3.appliedbiosystems.com), have the potential to revolutionize the kinds of approaches we take to transcriptomes. They are already being used in some non-model organisms, such as *Melitaea cinxia* L. (Glanville Fritillary Butterflies; Lepidoptera: Nymphalidae) (Ellegren 2008, Vera et al 2008**).** All of these technologies generate hundreds of thousands to billions of relatively short sequence reads from a sample in a few days time. Using cDNA (made from extracted mRNA), a transcriptome will be sequenced. This development is an important step forward from existing technologies, as the transcriptome is both sequenced and assessed for expression levels of individual genes simultaneously. In these approaches, expression level is not the intensity of a band on an acrylimide gel as in cDNA-AFLPs, or the hybridization intensity of a labeled sample to a particular probe on a microarray, but an actual count of the number of times a particular gene product is sequenced in the entire sample of sequences. This is a huge step forward in terms of accuracy and power to detect low abundance transcripts. To date, these technologies appear to have extremely low rates of technical error, making comparison between samples much easier than with other methods.

Currently, the main problems with these technologies are cost, the lack of machines, and inadequate tools for analysis. Costs per sample range from thousands to tens of thousands of dollars, depending on the particular technology, the facility used, the size of the transcriptome, and the cost of processing samples. Money can be saved if mRNA libraries are made by the researcher, using manufacturer kits, rather than having the facilities prepare those libraries. With adequate funding, all the serpentine ecologist needs to do is extract mRNA from the desired tissue and submit it to their local genomics core facility for analysis. Moreover, these costs can be expected to fall, as more machines come into service and the technology and chemistry underlying these approaches is optimized. Although open-source computation tools for analysis are not currently well developed, discussion among developers on the bioconductor webite (www.bioconductor.org) suggests that such tools will soon be available. Given the size of datasets, many researchers may find their current computing resources inadequate. Many facilities, however, offer a range of analysis services, and can do more in some cases through collaborative agreements. The actual analyses involved are not highly complex compared to most ecological experiments, but with

millions of reads per sample, the computational intensity of sorting through all the data is the challenge. Given the interest in transcriptomics for a range of research applications, more machines, facilities, and analysis tools can be expected in the coming years. When these tools are available, determining the nature of gene expression for any serpentine organism growing in any type of serpentine soil will be possible; we will be able to compare expression between serpentine and non-serpentine lineages and to determine expression patterns in response to particular aspects of serpentine soils. Once that is possible, then gene expression can be compared in different tissues by extracting mRNA from roots or shoots.

Epigenetics

There is more to the transcriptome than just mRNAs, and much is being learned about other types of RNAs and the regulation of translation and epigenetic effects through the use of transcriptomic and genomic tools. Whole-genome tiling arrays (Yazaki et al. 2007), which cover non-coding as well as coding regions, can be used to determine sequence differences between individuals or populations (appropriate for genomics), as well as for determining the methylation status of particular genes and for finding promoter regions (important for understanding gene regulation and epigenetics). As aspects of the "serpentine syndrome" may be epigenetically inherited as DNA packing rather than sequence variation (Bossdorf et al. 2008), being able to assess methylation status and other epigenetic modifications could be an interesting avenue of serpentine research. Understanding the regulatory mechanisms behind genes implicated in responses to serpentine soils could be a first step towards breeding plants for phytoremediation.

Genomics

Transcriptomic tools provide serpentine ecologists an exciting opportunity to explore the underlying genetics of serpentine tolerance and metal hyperaccumulation, even in species where a fully sequenced genome is not currently available. However, advances in genomics are happening fast, and obtaining a fully sequenced genome for any study organism will soon not be out of reach. The use of genomic tools for the study of ecology generally has been termed ecogenomics, and there have been a number of recent reviews (Kammenga et al. 2007, Ouborg and Vriezen 2007).

With traditional Sanger sequencing, it cost about $300 million and nearly four years to sequence a single human genome. Emerging technologies are reducing the cost and time towards a thousand dollars and a few days of sequencing. This is a rapidly changing field, most recently reviewed in Shendure and Ji (2008) and von Bubnoff (2008), though even these papers are likely to be out of date before ours goes to press. The quest for low-cost genome sequencing has helped spur a plethora of new resequencing technologies, all of which have the ability to generate massive amounts of sequence data in short amounts of time. No technology can currently

generate a genome for $1000, but several technologies have considerable promise, and are less than an order of magnitude away. The three technologies currently available mentioned earlier—454 from Roche, Solexa Genome Analyzer from Illumina, and SOLiD™ from ABI—all generate hundreds of thousands to billions of short sequence reads in a few days (Shendure and Ji 2008). To take the sequence information and turn it into a genome, each individual short sequence read has to be assembled into a position in a draft genome. In model organisms or their close relatives, this can be done by aligning each read to the existing genome. In non-model organisms, this can be done by assembling the reads to each other. This approach is computationally intense, but, it can work for any organism so long as read lengths are sufficiently long, duplicated sections are not too high in number, and a sufficient number of runs are used to give complete coverage of the genome. For the largest genomes, like pines, a sample may need to be read several times to get the number of reads necessary to cover the entire genome. Consequently, the only delay in having a genome for any organism of interest is cost, which will fall in time from tens of thousands dollars currently to something more within the reach of small labs.

These technological developments obviously open up a new era not only for serpentine ecology, but for all of the life sciences—an era of population and community genomics for any organism. The reasonable costs of these technologies will allow us to make genomic comparisons between lineages—be they higher taxa, species, or populations. The power of this approach to find genes linked to adaptations is unprecedented. Furthermore, because of the ability of this approach to elucidate histories of mutations (as single nucleotide polymorphisms [SNPS], indels, and copy number variations) between populations, it has the capability to give demographic history as well. For serpentine ecology, we will be able to find the unique genes in any serpentine lineage, so that we can assess their similarity to other serpentine or non-serpentine lineages. The potential this provides for understanding the genetic basis of serpentine adaptation is unprecedented.

Conclusions

"Serpentinomics" is a newly developing field with an exciting future. Advances are being made in a suite of –omic fields that will allow researchers to address long-standing questions in serpentine ecology. What exactly is the "serpentine syndrome"? How many different genetic mechanisms are involved in adaptation to serpentine soils? What changes in plant cellular chemistry are associated with tolerance to serpentine soils? What are the bases of potential costs of serpentine tolerance? How do communities respond to serpentine stresses? What are the genetic mechanisms behind hyperaccumulation? Are the same mechanisms used in different taxa? Can plants be engineered to be hyperaccumulators?

These emerging technologies have great promise to advance the field of serpentine ecology. As the technical and computational requirements of these technologies can be accomplished through core facilities and other outsourcing methods, they are well within the grasp of serpentine ecologists. They open the door to new answers to questions new and old.

Acknowledgments

The authors would like to thank N. Rajakaruna and R. Boyd for editing this volume, three anonymous reviewers for their useful comments, and S. Harrison for helping coin the term "serpentinomics."

Literature Cited

Allwood, J.W., D.I. Ellis, and R. Goodacre. 2008. Metabolomic technologies and their application to the study of plants and plant-host interactions. Physiologia Plantarum 132:117–135.

Bossdorf, O., C.L. Richards, and M. Pigliucci. 2008. Epigenetics for ecologists. Ecology Letters 11:106–115

Brooks, R.R. 1987. Serpentine and its Vegetation: A Multidisciplinary Approach. Dioscorides Press, Portland, OR, USA.

Ellegren, H. 2008. Sequencing goes 454 and takes large-scale genomics into the wild. Molecular Ecology 17:1629–1631

Hall, R.D. 2006. Plant metabolomics: From holistic hope, to hype, to hot topic. New Phytologist 169:453–468.

Huang, B., and C. Xu. 2008. Identification and characterization of proteins associated with plant tolerance to heat stress. Journal of Integrative Plant Biology 50:1230–1237.

Kammenga, J.E., M.A. Herman, N.J. Ouborg, L. Johnson, and R. Breitling. 2007. Microarray challenges in ecology. Trends in Ecology and Evolution 22:273–279.

Main, J.L. 1981. Magnesium and calcium nutrition of a serpentine endemic grass. American Midland Naturalist 105:196–199.

Ouborg, N.J., and W.H. Vriezen. 2007. An ecologist's guide to ecogenomics. Journal of Ecology 95:8–16.

Rampitsch, C., and M. Srinivasan. 2006. The application of proteomics to plant biology: A review. Canadian Journal of Botany 84:883–892.

Salt, D.E., I. Baxter, and B. Lahner. 2008. Ionomics and the study of the plant ionome. Annual Review of Plant Biology 59:709–733.

Schauer, N., and A.R. Fernie. 2006. Plant metabolomics: Towards biological function and mechanism. Trends in Plant Science 11:508–516.

Shendure, J., and H. Ji. 2008. Next-generation DNA sequencing. Nature Biotechnology 26:1135–1145

Shulaev, V., D. Cortes, G. Miller, and R. Mittler. 2008. Metabolomics for plant stress response. Physiologia Plantarum 132:199–208.

Steinfath, M., D. Groth, J. Lisec, and J. Selbig. 2008. Metabolite profile analysis: From raw data to regression and classification. Physiologia Plantarum 132:150–161.

Turner, T.L., E.J. von Wettberg, and S.V. Nuzhdin. 2008. Genomic analysis of differentiation between soil types reveals candidate genes for local adaptation in *Arabidopsis lyrata*. PLoS ONE 3(9):e3183

Vera, I.C., C.W. Wheat, H.W. Fescemyer, M.J. Frilander, D.L. Crawford, I. Hanski, and J.H. Marden. 2008. Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. Molecular Ecology 17:1636–1647

von Bubnoff, A. 2008. Next-generation sequencing: The race is on. Cell 132:721–723.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. van de Lee, M. Hornes, A. Friters, J. Pot, J. Paleman, M. Kuiper, and M. Zabeau. 1995. AFLP: A new technique for DNA fingerprinting. Nucleic Acids Research 23:4407–4414.

Yazaki, J., B.D. Gregory, and J.R. Ecker. 2007. Mapping the genome landscape using tiling array technology. Current Opinion in Plant Biology 10:534–542.

Zou, W., and V.V. Tolstikov. 2008. Probing genetic algorithms for feature selection in comprehensive metabolic profiling approach. Rapid Communication in Mass Spectrometry 22:1312–1324.

Glossary. Definitions of terms used in this paper.

Ionomics - the study of the inorganic component of tissues, including the mineral nutrients and trace elements.

Metabolomics - the study of the small molecules of organisms, including sugars, lipids, amino acids, and other small molecules with known functions in primary metabolism, as well as less well understood secondary metabolites.

Proteomics - the study of the proteins and large amino acids of organisms.

Transcriptomics - the study of the RNA of organisms. Traditionally this field includes methods to determine mRNA expression, but it encompasses other forms of RNA and DNA modification as well.

Genomics - the study of the genetic structure and composition of organisms.

cDNA - single stranded messenger RNA (mRNA) converted in double stranded DNA in the lab by reverse transcription coupled with polymerase chain reaction amplification.

qRT-PCR - quantitative reverse-transcription-polymerase chain reaction. This method quantifies the amount of particular mRNA present in a sample.

Next generation sequencing - any technology that produces vast amounts of sequence data in a short period of time (a week or less). It is sometimes called resequencing because it can be used to sequence individuals of an organism for which a completed genome exists.

Microarray - a glass slide or microscope slide-sized silica chip onto which short oligonucleotide probes representing genes of interest are affixed. If cDNA samples labeled with a fluorescent probes are hybridized to a microarray, a laser can read hybridization intensity, giving a qualitative assessment of expression level of thousands to millions of genes or gene fragments from a single sample. If genomic DNA is applied to microarrays, they can be used for genotyping. This approach can either involve a secondary use of expression arrays, or can be done with specialized arrays designed to distinguish known single nucleotide polymorphisms in a model organism.